Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Eric D. Hoffer Date

Molecular characterization of ribosome-independent toxin substrate specificity

By

Eric D. Hoffer
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Biochemistry, Cell and Developmental Biology

_____
Christine M. Dunham, Ph.D.
Advisor

_____
Anita H. Corbett, Ph.D.
Committee Member

_____
Eric A. Ortlund, Ph.D.
Committee Member

_____
Daniel Reines, Ph.D.
Committee Member

_____
Richard A. Kahn, Ph.D.
Committee Member

_____
Charles P. Moran Jr., Ph.D.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

Molecular characterization of ribosome-independent toxin substrate specificity

By

Eric D. Hoffer
B.S., Kennesaw State University, 2011

Advisor: Christine M. Dunham, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Graduate Division of Biological and Biomedical Science
Biochemistry, Cell and Developmental Biology
2017

Abstract

Molecular characterization of ribosome-independent toxin substrate specificity
By Eric D. Hoffer

Bacteria have a remarkable ability to survive rapidly fluctuating environments such as nutrient deprivation and antibiotic treatment. There are several genes and pathways that contribute to bacterial survival during stress and one of the most abundant are the two-component systems known as toxin-antitoxins. Despite the nomenclature, toxin-antitoxins are beneficial to their host bacterium and are associated with the formation of non-growing, highly stress tolerance cells known as persisters. The toxin inhibits cell growth through disrupting essential cellular processes and the antitoxin inhibits the toxin under non-stress conditions. The work presented here describes the structure and substrate specificity of two toxin families. The first is from the VapC toxin family known as VapC-mt4 from *Mycobacterium tuberculosis* and was found to cleave elongator tRNAs t0 downregulate protein synthesis. To better understand how VapC recognizes tRNAs, I generated a homology model and simulated binding to tRNA and found that VapC may have different modes of binding depending on the length of substrate tRNA variable loop. In this study I also determined the structure and characterized the active site of MazF-mt6, a member of the CcdB/MazF toxin family that has an unusual specificity for ribosomal RNA. Through this research I show that MazF-mt6 has an irregular β1-β2 linker that likely plays a role in allowing structured RNA to enter the active site. Finally, I attempted to solve the crystal structures of *S. flexneri* VapC and *Mtb* MazF-mt9 bound to substrate tRNA. Though I was unable to solve the target crystal structure, I was able to optimize the purification of both toxins and tRNA substrates and show that the $VapC^{D7A}$ variant was able to stably bind $tRNA^{fMet}$. These different research projects have helped to expand our understanding of bacterial toxin structure and substrate specificity.

Molecular characterization of ribosome-independent toxin substrate specificity

By

Eric D. Hoffer
B.S., Kennesaw State University, 2011

Advisor: Christine Dunham. Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Graduate Division of Biological and Biomedical Science
Biochemistry, Cell and Developmental Biology
2017

Table of Contents

LIST OF FIGURES AND TABLES

Chapter 1

Chapter 2

Chapter 5

Appendix I

Abbreviations

aa: amino acid
A site: Aminoacyl site
APS: Advanced Photon Source
Arg: arginine
ASL: anticodon stem loop
Asp: aspartate
ATP: adenosine triphosphate
BCDB: Biochemistry, Cell and Developmental Biology
BsMazF: *B. subtilis* MazF
BZA: benzamidine
CcdA: Control of cell death and division A
CcdB: Control of cell death and division B
COOT: Crystallographic Object-Oriented Toolkit
DOE: Department of Energy
DRNC: dipeptidyl-tRNA ribosomal nascent chain
ds: double-stranded
DTT: dithiothreitol
EcMazF: *E. coli* MazF
EDTA: ethylenediaminetetraacetic acid
EF-G: elongation factor G
EF-P: elongation factor P
EF-Tu: elongation factor Tu
EMSA: electrophoretic mobility shift assay
GGQ: glycine-glycine-glutamine
Glu: glutamate
Gly: glycine
H70: Helix 70
HigA: Host inhibition of growth A
HigB: Host inhibition of growth B
$(His)_6$: hexahistidine
IDT: integrated DNA technologies
IMAC: immobilized metal affinity chromatography
IPTG: isopropyl-1-thio-β-D-galactopyranoside
IVT: *in vitro* transcription
kDa: kilodalton
LB: lysogeny broth
LC-MS/MS: liquid chromatography-tandem mass spectrometry
Lys: lysine
Min: minute
MPD: 2-methyl-2-4-pentanediol
mRNA: messenger RNA
Mtb: *Mycobacterium tuberculosis*
NE-CAT: Northeastern-Collaborative Access Team
NIH: National Institute of Health
Nt: nucleotide
PDB: Protein Data Bank
PEG: polyethylene glycol
PemK: plasmid emergency maintenance K
PIN: PilT-N-terminus
PMSF: Phenylmethylsulfonyl fluoride

(p)ppGpp: guanosine tetra/penta phosphate
PTC: peptidyl transferase center
RFs: Release Factors
RF1: Release Factor 1
RF2 Release Factor 2
RMSD: root-mean-square deviation
rRNA: ribosomal RNA
SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC: size exclusion chromatography
Ser: serine
SER-CAT: Southeast Regional Collaborative Access Team
ss: single-stranded
TA: toxin-antitoxin
TE: tris-borate EDTA
TEV: tobacco etch virus
TF: Trigger Factor
Thr: threonine
tRNA: transfer RNA
Tth: *Thermus thermophilus*
VapB: Virulence associated protein B
VapC: Virulence associated protein C
WT: wild type
XDS: x-ray Detector Software
β-Me: β-mercaptoethanol

CHAPTER 1: INTRODUCTION

Challenges of antibiotic tolerance

Since the discovery of penicillin by Alexander Fleming over 70 years ago, antibiotics are the gold standard for treating bacterial infections. Our lives were transformed and life expectancy was extended due to the effectiveness of treating bacterial infections with antibiotics. However, bacteria have evolved several resistance mechanisms to deal with antibiotic exposure including the production of β-lactamases, efflux pumps, porin defects, alterations in cell wall structure and post-transcriptional modifications (1). All of these mechanisms are examples of acquired antibiotic resistance, which make bacteria immune to specific antibiotics (1). Most commonly-used antibiotics target important cellular pathways and inhibit growth; therefore, if a cell is not actively growing, antibiotics typically have no effect (2). In addition to these well-studied resistance mechanisms, even if a bacterial population does not acquire a specific resistance gene, a small population of cells can survive antibiotic treatment for a defined amount of time (3). These cells are non-growing and this antibiotic-tolerant state is known as persistence (2-4). Unlike resistance, persistent cells do not contain resistance genes and instead stochastically are antibiotic tolerant. Persister cells are a significant problem in modern medicine as they increase the chance of relapse in bacterial infections after antibiotic treatment (5). Understanding the pathways involved in the formation of persister cells could potentially allow us to better combat the ever-present threat of bacterial infections. My dissertation is focused on the mechanisms for how bacteria control their growth that can lead to persistence.

Bacterial stringent response

A well-known and conserved pathway that leads to persister cell formation is the stringent response. The stringent response is conserved throughout bacteria and can be activated by a variety of stress and starvation signals (6). This response causes an inhibition of transcription and protein synthesis when nutrients are limiting **(Fig. 1.1)** (7, 8). However, irrespective of the type of stress that activates the stringent response, the result is always production of the second messenger, guanosine pentaphosphate or tetraphosphate ((p)ppGpp). One of the major roles of (p)ppGpp is to repress transcription by either directly binding RNA polymerase (Gram-negative bacteria) or changing the NTP pools required to initiate transcription (Gram-positive bacteria) (6, 9, 10). In some cases, (p)ppGpp has also been shown to activate the rate of protein degradation in a likely attempt to replenish the pool of available amino acids (11). The consequences of activating the stringent response not only inhibits expression of genes required for growth, such as ribosomal RNA, it also activates small protein factors called toxins that further inhibit cellular growth by halting DNA replication or translation **(Fig. 1.1)** (7, 8). Toxins are ubiquitously found in bacteria and archaea and associated with both persistence and virulence (4, 12). Toxins are encoded with antidote or antitoxin proteins in a single operon and the expression of both ensures the toxin activity is inhibited by antitoxin binding during non-stress conditions. Therefore, these toxins are not considered traditional toxins that are expressed in one bacterial cell and excreted to kill other neighboring cells like colicins or contact-dependent inhibition of growth toxins (13, 14).The focus of this dissertation is on the structure and substrate specificity of toxin proteins that target RNA for degradation in these systems.

Classification of toxin-antitoxin systems

Toxin-antitoxin genes are encoded chromosomally and on extrachromosomal DNA where they were first identified in 1983 as plasmid addiction modules (15). After 20 years and numerous toxin-antitoxins genes identified in all bacterial species, toxin-antitoxin systems have been classified based upon how the antitoxin protein binds and inhibits its cognate toxin (16). There are six known classes of toxin-antitoxin systems (type I, II, III, IV, V and VI) **(Fig. 1.2)** (17). All six classes encode a proteinaceous toxin and the antitoxin is either RNA or protein and inhibits toxin activity by direct binding (type II, III, VI), the prevention of toxin protein expression (type I, V), or by competing with the toxin for substrate binding (type IV) (17). In all cases, the ratio of antitoxin to toxin is a critical determinant of whether a cell grows or enters stasis. The system is considered dormant when antitoxin levels exceed toxin and active in the opposite scenario **(Fig. 1.3)**. There are multiple examples of type I, II and III toxins, but the most abundant class of toxin-antitoxin system is the type II and is the focus of my dissertation (18-22).

Type II toxin-antitoxins systems

As mentioned, type II toxins are proteinaceous and upon release from their cognate antitoxin, toxins inhibit cellular growth (23, 24). The activity and target of the toxin proteins varies between the six types and includes disruption of membrane potential, inhibition of DNA replication and protein synthesis (25-28). The first type II toxin-antitoxin gene identified was the *E. coli* F-plasmid *control of cell death and division AB* (*ccdAB*) (15). The *ccdAB* operon contains two protein gene products, CcdA and CcdB. The CcdB toxin protein binds and inhibits DNA gyrase subunit GyrA so while it can still cleave DNA, it is unable to relegate the DNA strands together (26, 29). The CcdA antitoxin uses two different mechanisms to counteract CcdB. The first role is that CcdA directly binds CcdB, which prevents the toxin from inhibiting gyrase. CcdA additionally acts as a transcription

factor that inhibits expression from the *ccdAB* operon by direct DNA binding (30). If bacteria are transformed with the F-plasmid which contains both CcdB and CcdA, cells survive because there is a constant supply of CcdA to inhibit CcdB. However, if the cell loses the F-plasmid, it is unable to survive, which is known as post-segregational killing (31). The cell is unable to survive because the toxin protein becomes free and inhibits growth. This is mainly because the antitoxin is labile and thus has a much shorter half-life than the toxin (31), so there is a constant need for excess antitoxin to prevent the toxin from being released. The only other type II toxin-antitoxin system that inhibits DNA replication is ParDE (32). The antitoxin ParD has the same toxin inhibition and transcriptional repression activity as CcdA, yet they share no sequence identity or homology. Likewise, the toxin ParE also inhibits gyrase, but unlike CcdB, ParE inhibits the other subunit of gyrase known as GyrB (32). CcdAB and ParDE are excellent examples of the diversity found in the type II class of toxin-antitoxins where they inhibit similar cellular functions but appear to have evolved separately (33). CcdB also belongs to a structural superfamily named CcdB/MazF, where MazF is a ribosome-independent endoribonuclease and one of the major focuses of this dissertation.

Although CcdB was the first type II toxin to be identified, the majority of type II toxins are endoribonucleases (34). There are two subclasses of type II endonucleases-ribosome-independent and ribosome-dependent RNases. The ribosome-associated toxins do not possess endoribonuclease activity on their own and instead, require the presence of a translating ribosome (34). One example of a ribosome-dependent toxin is RelE, which is found in the *delayed relaxed BEF* (*relBEF*) operon. The RelE toxin cleaves mRNA only when bound to the A site of the ribosome and has a preference for stop codons UAG and UGA (27, 35-37). RelE is structurally homologous to ParE, HigB, YoeB, YafQ and YafO toxins and make up the RelE/ParD superfamily of type II toxins (38-41).

Regulation of type II systems is thought to be mediated by a mechanism called conditional cooperativity first described for the RelBE and CcdBA toxin-antitoxin systems (42, 43). The conditional part of this mechanism refers to the variable amounts of toxins and antitoxins expressed at different states of the cell in response to the stress encountered **(Fig. 1.3)**. Cooperativity refers to the interactions between the toxin and the antitoxin. The DNA-binding antitoxin represses expression of its operon and it has been proposed that binding of the toxin to the DNA-antitoxin complex increases repression. However, if the ratio of toxin to antitoxin exceeds a critical point, the interactions between the antitoxin and DNA change which causes the antitoxin to disengage from the operon. This allows for expression from the toxin-antitoxin loci. Antitoxins are then susceptible to proteolysis thus freeing the toxin (42, 44).

Toxin structural families such as the CcdB/MazF and RelE/ParD share common tertiary folds, but contain low amino acid sequence identity. The latter makes the identification of novel toxins difficult. In this case, it is important to structurally characterize toxin proteins to identify their structural family and help predict how they function. The major aim of my thesis is to determine the structures of toxins that are predicted to contain endonuclease folds but recognize structurally diverse nucleic acid targets. These results will provide mechanistic insights into how toxins have evolved novel motifs to alter substrate specificity.

MazEF toxin-antitoxin system

The *E. coli* MazEF toxin-antitoxin system was the first toxin to be found in a bacterial genome and not encoded on a plasmid (45). The *mazEF* genes are part of the *relA* operon, which encodes the (p)ppGpp synthetase RelA and is a link between toxin-antitoxin systems and the stringent response **(Fig. 1.1)** (45) . When these genes were identified, their function was still unclear and therefore the name *mazef* was used because

“ma-ze” means “what is this?” in Hebrew (45). The MazF toxin is a ribosome-independent endoribonuclease that cleaves mRNA containing the sequence 5’-ACA-3’ (28). In this original study, the *E. coli* strain used was deficient in RelA activity and therefore caused MazF overexpression to be bactericidal (46, 47). MazF is now considered to be bacteriostatic similar to all other chromosomal type II toxins, unifying the idea that these systems are part of the stress response pathway to control growth (48, 49).

Since the discovery of *E. coli* MazF, several orthologs have been identified in other bacterial species including *B. subtilis, S. aureus* and *M. tuberculosis,* where they also cleave mRNA (50-52). Structures of *B. subtilis* MazF-MazE and *B. subtilis* MazF-RNA demonstrated how MazE inhibits MazF toxin activity and how MazF recognizes and cleaves its RNA substrate, respectively (50). Recently, the *E. coli* MazF-RNA complex allowed the authors to propose a catalytic mechanism and mode of regulation for MazF toxins (53). Most prokaryotic genomes usually contain only 1-2 *mazEF* operons, however, the *M. tuberculosis* genome contains nine MazEF pairs (called MazF-mt1-MazF-mt9) based on their position in the genome (19, 54, 55). All *M. tuberculosis* MazF toxins are RNases *in vitro* or when overexpressed in *E. coli* (52, 54, 56-58), and three of the nine MazF toxins (MazF-mt3, MazF-mt6 and MazF-mt9) promote virulence upon infection in guinea pigs (12).

M. tuberculosis MazF toxins have expanded RNA substrates not seen in other organisms. MazF-mt9 cleaves elongator tRNAs while MazF-mt3 cleaves both 16S and 23S rRNA, all resulting in the inhibition of protein synthesis (52, 54, 58). Both tRNA and rRNA are unusual substrates for MazF toxins, which typically cleave single-stranded mRNA. In addition, the sequences of MazF-mt6 and MazF-mt9 suggest that there may be significant structural changes from the rest of the toxin family including truncations and expansions of conserved regions. MazF toxins usually target unstructured single-stranded RNA, so toxins like MazF-mt6 and MazF-mt9 that cleave rRNA and tRNA are ideal for

broadening our understanding of how MazF toxins recognize complex RNA substrates. In this dissertation, I solved the structure of MazF-mt6 and characterized which residues are important for activity (Chapter 3). I also purified MazF-mt9 and $tRNA^{Pro}$ and $tRNA^{Lys}$ with the goal of solving the X-ray crystal structure of the MazF-mt9-tRNA complex (Chapter 4).

VapBC toxin-antitoxin system

The *vapB-vapC* genes were first identified from *Dichelobacter nodosus*, a bacterium that causes foot rot in sheep (59). The genes were named *Virulence associated proteins ABCD* (*vapABCD*) and two of them, VapB and VapC, were found to be a toxin-antitoxin system (59, 60). The VapC toxin is a metal-dependent endoribonuclease that belongs to the PIlt-N-terminus (PIN) structural family of proteins, which shares homology to endonuclease RNase H, an enzyme that cleaves RNA found in DNA-RNA helices (61). The VapB antitoxin is a DNA-binding protein that contains a N-terminal AbrB dimerization domain and a C-terminal toxin binding domain (62). In all studied homologs to date, both VapC and VapB form homodimers that in the context of a complex, form heterooctamers (43, 63-65).

VapC not only limits growth but has been also been shown to regulate metabolic flux, which contrasts with the global translational inhibition that is often assumed for all ribonuclease toxins (66). The first example of a VapC that tunes cellular translation was *Mycobacterium smegmatis* VapC, which cleaves mRNAs responsible for glycerol consumption, and when the gene is deleted, the rate of glycerol consumption is decoupled from growth. In other words, without VapC, *M. smegmatis* consumes 2-3 fold more glycerol than needed to generate biomass (66). If toxins like VapC are capable of changing cellular metabolism without causing growth inhibition, it could explain why some bacteria contain so many copies of toxin-antitoxin genes in their genome.

In bacteria like *Shigella flexneri* that contain a single VapC toxin, the VapC toxin cleaves initiator $tRNA^{fMet}$ that completes halts protein synthesis (67). In most organisms, there is a single *vapBC* operator while in *M. tuberculosis (Mtb),* there are 48 VapBC homologs (55). In organisms like *Mtb* that contain multiple VapBC pairs, VapC targets other RNAs including mRNA, elongator tRNAs and the sarcin ricin loop of the 50S ribosomal subunit (68-72), but most toxins are still yet to be characterized (55, 70). *Mtb* VapC-mt4 (Rv0595c, also known as VapC4) cleaves elongator $tRNA^{Ala}$, $tRNA^{Ser}$ and $tRNA^{Cys}$ (72, 73).

Despite the numerous examples of VapBC systems studied to date, it is still unclear what VapC features are important to discriminate between diverse RNA substrates. Therefore, the structural studies of VapC-RNA complexes are essential to our goals to expand our understanding of these unique toxins. In this dissertation, I performed homology modeling of *Mtb* VapC-mt4 (Chapter 2) and started crystallization trials of *S. flexneri* VapC bound to $tRNA^{fMet}$ (Chapter 4). These structures would provide insights on how VapC toxins recognize structured RNA.

Questions addressed

The *Mtb* VapC and MazF toxin-antitoxin systems are predicted to contain a common structural fold yet recognize diverse substrates. However, given their low sequence identities, it is not entirely clear if they all adopt the same tertiary fold. I approached these questions using a combination of structural, biochemical, and *in silico* techniques. In Chapter 2, in collaboration with the Woychik lab, we functionally characterized the *Mtb* VapC-mt4 toxin using RNA-seq and *in vitro* RNA cleavage assays. I performed homology modeling and simulated binding of VapC-mt4 to two tRNA substrates to predict different modes of VapC-mt4-RNA binding (69). In Chapter 3, I investigated how MazF-mt6 recognizes and cleaves rRNA of the large ribosomal 50S

subunit. I determined that the 50S was not required for RNase activity by demonstrating MazF-mt6 could cleave a small RNA containing the 23S rRNA Helix 70 sequence *in vitro*. I also solved the X-ray crystal structure of MazF-mt6 toxin and characterized which residues are essential for activity both *in vivo* and *in vitro* (Hoffer *et al.*, JBC 2017). In Chapter 4, I demonstrated that *S. flexneri* VapC^{D7A} and MazF-mt9 could be purified to homogeneity for structural studies. In addition, I purified tRNA substrates for both *S. flexneri* VapC^{D7A} and MazF-mt9 and show that the VapC^{D7A} variant binds its substrate (but does not cleave tRNA) using an electrophoretic mobility shift assay. I set up crystallization trials of both toxins and was working on optimizing crystallization trials. In Chapter 5, I discuss the structural characteristics and toxin activities of the MazF/CcdB superfamily of toxins. I also argue that the structural features of distantly related toxins, such as type II MazF and type III ToxN, share a common active site which can be used to infer novel modes of how toxins can recognize structured RNAs for example, how MazF-mt9 recognizes tRNA. Finally, I advocate that continued structural studies of toxin-antitoxin systems are critical to advancing our understanding of how toxin proteins recognize diverse RNA molecules that control growth in response to stress (Chapter 5).

Figures

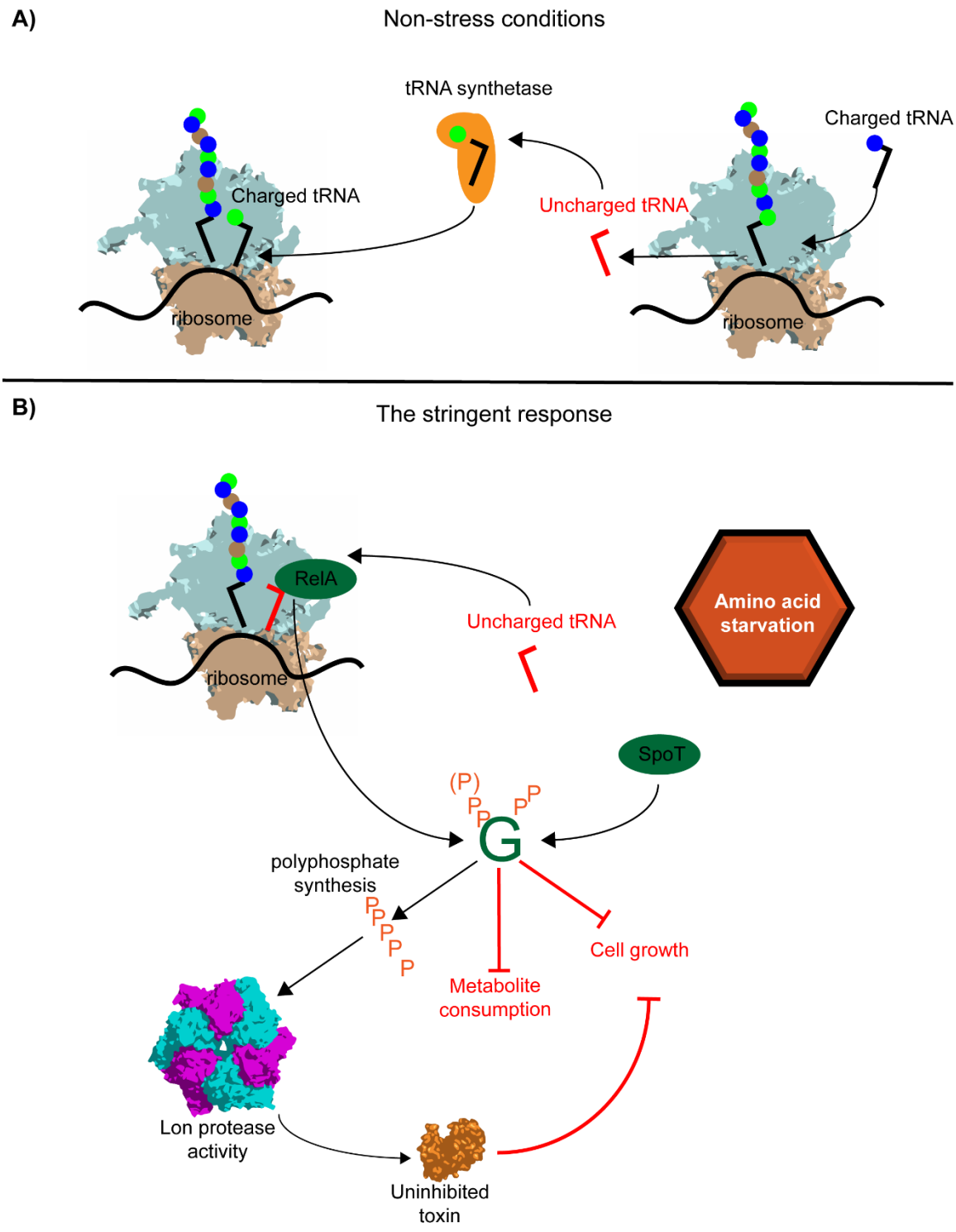
A)
Non-stress conditions
tRNA synthetase
Charged tRNA
Uncharged tRNA
Charged tRNA
ribosome
ribosome
B)
The stringent response
RelA
Uncharged tRNA
Amino acid starvation
ribosome
SpoT
(P)
P
P
P
P
G
polyphosphate synthesis
P
P
P
P
P
Cell growth
Metabolite consumption
Lon protease activity
Uninhibited toxin

Figure 1.1 The bacterial stringent response. A) Under non-stress conditions, protein synthesis is uninhibited. The 70S ribosome is shown translating mRNA into a nascent chain using charged aminoacyl-tRNAs. B) When cells encounter a stress such as amino acid starvation, the stringent response is activated. In *E. coli,* RelA recognizes uncharged tRNA bound to the ribosome and begins synthesis of the second messenger (p)ppGpp, which is also produced by the cellular protein SpoT under stress (not shown). Once synthesized, (p)ppGpp downregulates growth by repressing transcription of several growth factors and also stimulates the produces polyphosphates, which has shown to have an effect on Lon protease activity. Lon degrades antitoxin proteins that then further downregulate growth by inhibiting a variety of essential pathways. Presumably, cells exit the stringent response once they have accumulated enough metabolites to grow again and the stress has passed.

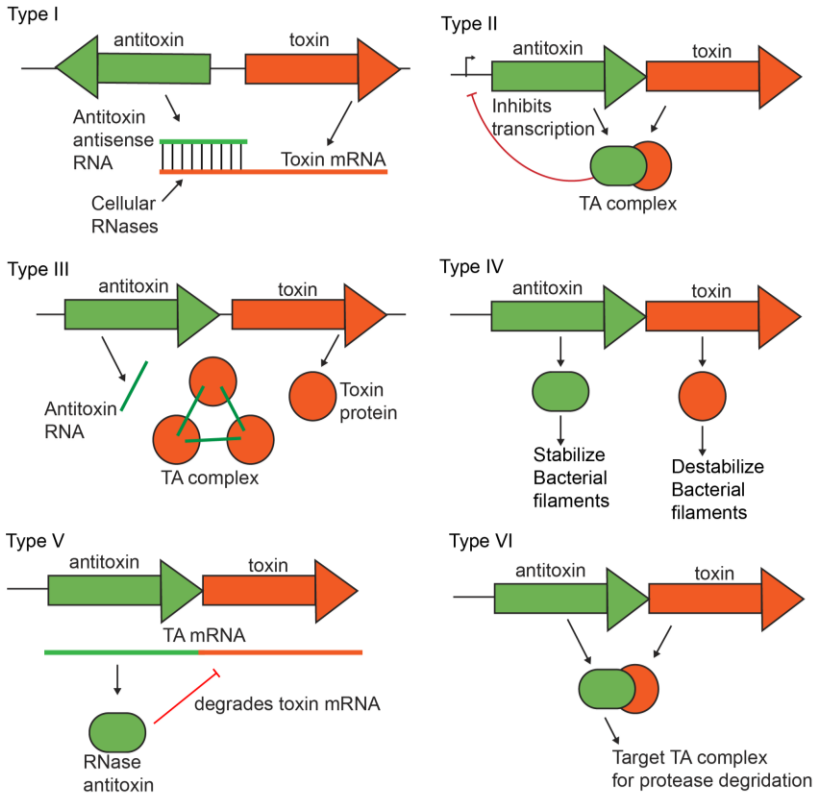

Figure 1.2 Schematics of the six toxin-antitoxin systems. Toxins are shown in orange and antitoxins in green. Type I: the antisense RNA antitoxin base pairs with toxin mRNA to inhibit translation. Type II: the antitoxin and toxin are proteins; under growth conditions, the toxin is bound to the antitoxin, which inhibits its activity and the antitoxin represses transcription from its own promoter. Type III: the antitoxin ncRNA is processed by the toxin, resulting in the formation of RNA pseudoknot–toxin complexes, which inhibit toxin activity. Type IV: the protein antitoxin stabilizes bacterial filaments, while the protein toxin destabilizes them. Type V: the antitoxin is an RNase specific for the toxin mRNA. Type VI: the antitoxin is an adaptor protein that binds the toxin to promote its degradation by cellular proteases.

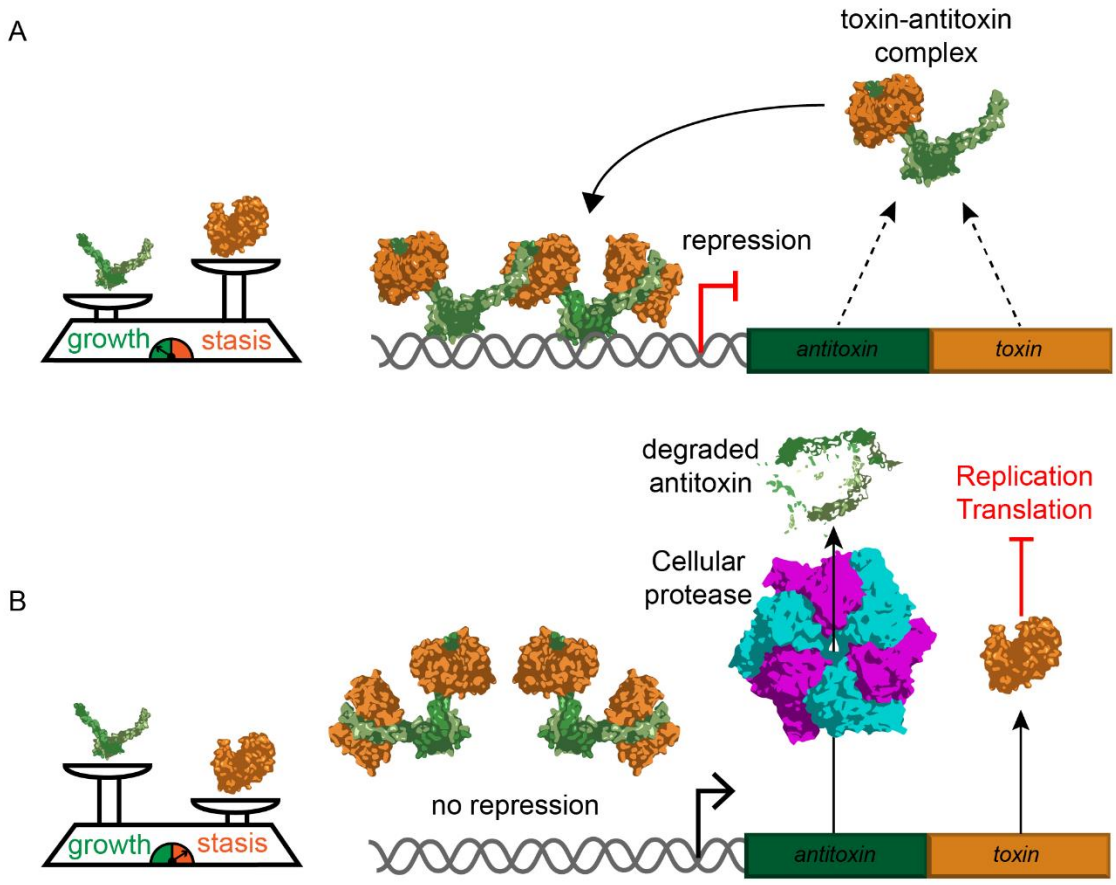

Figure 1.3 Ratios of toxin-antitoxin proteins affect cellular growth. A) When the ratio of antitoxin to toxin is high, then cell growth is uninhibited due to the repression of the toxin and the negative feedback from antitoxin transcriptional repression. B) Under stress conditions like starvation, the ratio of antitoxin to toxin decreases due to either slower rate of protein synthesis or increased protein degradation by cellular proteases. Once toxin levels hit a certain threshold, they derepress their own promoter by preventing the antitoxin from binding and any new antitoxin is rapidly turned over due to cellular protease activity.

References

1. G. Taubes, The bacteria fight back. *Science* **321**, 356-361 (2008).
2. T. K. Wood, Combatting bacterial persister cells. *Biotechnology and bioengineering* **113**, 476-483 (2016).
3. A. Harms, E. Maisonneuve, K. Gerdes, Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science* **354**, (2016).
4. E. Maisonneuve, M. Castro-Camargo, K. Gerdes, (p)ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity. *Cell* **154**, 1140-1150 (2013).
5. N. Verstraeten, W. Knapen, M. Fauvart, J. Michiels, A Historical Perspective on Bacterial Persistence. *Methods Mol Biol* **1333**, 3-13 (2016).
6. C. C. Boutte, S. Crosson, Bacterial lifestyle shapes stringent response activation. *Trends Microbiol* **21**, 174-180 (2013).
7. E. Maisonneuve, K. Gerdes, Molecular mechanisms underlying bacterial persisters. *Cell* **157**, 539-548 (2014).
8. K. Potrykus, M. Cashel, (p)ppGpp: still magical? *Annu Rev Microbiol* **62**, 35-51 (2008).
9. L. U. Magnusson, A. Farewell, T. Nystrom, ppGpp: a global regulator in Escherichia coli. *Trends Microbiol* **13**, 236-242 (2005).
10. L. Krasny, H. Tiserova, J. Jonak, D. Rejman, H. Sanderova, The identity of the transcription +1 position is crucial for changes in gene expression in response to amino acid starvation in Bacillus subtilis. *Mol Microbiol* **69**, 42-54 (2008).
11. K. Gerdes, S. K. Christensen, A. Lobner-Olesen, Prokaryotic toxin-antitoxin stress response loci. *Nat Rev Microbiol* **3**, 371-382 (2005).
12. P. Tiwari *et al.*, Corrigendum: MazF ribonucleases promote Mycobacterium tuberculosis drug tolerance and virulence in guinea pigs. *Nature communications* **6**, 7273 (2015).
13. J. L. Willett, Z. C. Ruhe, C. W. Goulding, D. A. Low, C. S. Hayes, Contact-Dependent Growth Inhibition (CDI) and CdiB/CdiA Two-Partner Secretion Proteins. *Journal of molecular biology* **427**, 3754-3765 (2015).
14. V. Braun, S. I. Patzer, Intercellular communication by related bacterial protein toxins: colicins, contact-dependent inhibitors, and proteins exported by the type VI secretion system. *FEMS Microbiol Lett* **345**, 13-21 (2013).
15. T. Ogura, S. Hiraga, Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proc Natl Acad Sci U S A* **80**, 4784-4788 (1983).
16. F. Hayes, Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science* **301**, 1496-1499 (2003).
17. D. Lobato-Marquez, R. Diaz-Orejas, F. Garcia-Del Portillo, Toxin-antitoxins and bacterial virulence. *FEMS Microbiol Rev*, (2016).
18. Y. Yamaguchi, M. Inouye, Regulation of growth and death in Escherichia coli by toxin-antitoxin systems. *Nat Rev Microbiol* **9**, 779-790 (2011).
19. D. P. Pandey, K. Gerdes, Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res* **33**, 966-976 (2005).
20. E. W. Sevin, F. Barloy-Hubler, RASTA-Bacteria: a web-based tool for identifying toxin-antitoxin loci in prokaryotes. *Genome Biol* **8**, R155 (2007).
21. K. S. Makarova, Y. I. Wolf, E. V. Koonin, Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol Direct* **4**, 19 (2009).

22. R. Leplae *et al.*, Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic Acids Res* **39**, 5513-5525 (2011).
23. K. Gerdes, E. Maisonneuve, Bacterial persistence and toxin-antitoxin loci. *Annu Rev Microbiol* **66**, 103-123 (2012).
24. Y. Yamaguchi, J. H. Park, M. Inouye, Toxin-antitoxin systems in bacteria and archaea. *Annu Rev Genet* **45**, 61-79 (2011).
25. K. Gerdes, J. E. Larsen, S. Molin, Stable inheritance of plasmid R1 requires two different loci. *J Bacteriol* **161**, 292-298 (1985).
26. S. Maki, S. Takiguchi, T. Miki, T. Horiuchi, Modulation of DNA supercoiling activity of Escherichia coli DNA gyrase by F plasmid proteins. Antagonistic actions of LetA (CcdA) and LetD (CcdB) proteins. *J Biol Chem* **267**, 12244-12251 (1992).
27. S. K. Christensen, K. Gerdes, RelE toxins from bacteria and Archaea cleave mRNAs on translating ribosomes, which are rescued by tmRNA. *Mol Microbiol* **48**, 1389-1400 (2003).
28. Y. Zhang *et al.*, MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli. *Mol Cell* **12**, 913-923 (2003).
29. S. Maki, S. Takiguchi, T. Horiuchi, K. Sekimizu, T. Miki, Partner switching mechanisms in inactivation and rejuvenation of Escherichia coli DNA gyrase by F plasmid proteins LetD (CcdB) and LetA (CcdA). *Journal of molecular biology* **256**, 473-482 (1996).
30. T. Madl *et al.*, Structural basis for nucleic acid and toxin recognition of the bacterial antitoxin CcdA. *Journal of molecular biology* **364**, 170-185 (2006).
31. L. Van Melderen, P. Bernard, M. Couturier, Lon-dependent proteolysis of CcdA is the key control for activation of CcdB in plasmid-free segregant bacteria. *Mol Microbiol* **11**, 1151-1157 (1994).
32. Y. Jiang, J. Pogliano, D. R. Helinski, I. Konieczny, ParE toxin encoded by the broad-host-range plasmid RK2 is an inhibitor of Escherichia coli gyrase. *Mol Microbiol* **44**, 971-979 (2002).
33. M. B. Kamphuis *et al.*, Structure and function of bacterial kid-kis and related toxin-antitoxin systems. *Protein and peptide letters* **14**, 113-124 (2007).
34. G. M. Cook *et al.*, Ribonucleases in bacterial toxin-antitoxin systems. *Biochim Biophys Acta* **1829**, 523-531 (2013).
35. M. Gotfredsen, K. Gerdes, The Escherichia coli relBE genes belong to a new toxin-antitoxin gene family. *Mol Microbiol* **29**, 1065-1076 (1998).
36. C. Neubauer *et al.*, The structural basis for mRNA recognition and cleavage by the ribosome-dependent endonuclease RelE. *Cell* **139**, 1084-1095 (2009).
37. K. Pedersen *et al.*, The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. *Cell* **112**, 131-140 (2003).
38. M. A. Schureck *et al.*, Structure of the Proteus vulgaris HigB-(HigA)2-HigB toxin-antitoxin complex. *J Biol Chem* **289**, 1060-1070 (2014).
39. S. Feng *et al.*, YoeB-ribosome structure: a canonical RNase that requires the ribosome for its specific activity. *Nucleic Acids Res* **41**, 9549-9556 (2013).
40. Y. Liang *et al.*, Structural and Functional Characterization of Escherichia coli Toxin-Antitoxin Complex DinJ-YafQ. *J Biol Chem* **289**, 21191-21202 (2014).
41. Y. Zhang, Y. Yamaguchi, M. Inouye, Characterization of YafO, an Escherichia coli toxin. *J Biol Chem* **284**, 25522-25531 (2009).
42. M. Overgaard, J. Borch, M. G. Jorgensen, K. Gerdes, Messenger RNA interferase RelE controls relBE transcription by conditional cooperativity. *Mol Microbiol* **69**, 841-857 (2008).

43. C. Dienemann, A. Boggild, K. S. Winther, K. Gerdes, D. E. Brodersen, Crystal structure of the VapBC toxin-antitoxin complex from Shigella flexneri reveals a hetero-octameric DNA-binding assembly. *Journal of molecular biology* **414**, 713-722 (2011).
44. M. J. Mate *et al.*, Crystal structure of the DNA-bound VapBC2 antitoxin/toxin pair from Rickettsia felis. *Nucleic Acids Res* **40**, 3245-3258 (2012).
45. E. Aizenman, H. Engelberg-Kulka, G. Glaser, An Escherichia coli chromosomal "addiction module" regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death. *Proc Natl Acad Sci U S A* **93**, 6059-6063 (1996).
46. S. Amitai, Y. Yassin, H. Engelberg-Kulka, MazF-mediated cell death in Escherichia coli: a point of no return. *J Bacteriol* **186**, 8295-8300 (2004).
47. B. C. Ramisetty, S. Raj, D. Ghosh, Escherichia coli MazEF toxin-antitoxin system does not mediate programmed cell death. *J Basic Microbiol* **56**, 1398-1402 (2016).
48. B. C. M. Ramisetty, S. Raj, D. Ghosh, Escherichia coli MazEF toxin-antitoxin system does not mediate programmed cell death. *Journal of Basic Microbiology* **56**, 1398-1402 (2016).
49. M. A. Syed, C. M. Levesque, Chromosomal bacterial type II toxin-antitoxin systems. *Can J Microbiol* **58**, 553-562 (2012).
50. D. K. Simanshu, Y. Yamaguchi, J. H. Park, M. Inouye, D. J. Patel, Structural basis of mRNA recognition and cleavage by toxin MazF and its regulation by antitoxin MazE in Bacillus subtilis. *Mol Cell* **52**, 447-458 (2013).
51. L. Zhu *et al.*, Staphylococcus aureus MazF specifically cleaves a pentad sequence, UACAU, which is unusually abundant in the mRNA for pathogenic adhesive factor SraP. *J Bacteriol* **191**, 3248-3255 (2009).
52. J. M. Schifano *et al.*, Mycobacterial toxin MazF-mt6 inhibits translation through cleavage of 23S rRNA at the ribosomal A site. *Proc Natl Acad Sci U S A* **110**, 8501-8506 (2013).
53. V. Zorzini *et al.*, Substrate Recognition and Activity Regulation of the Escherichia coli mRNA Endonuclease MazF. *J Biol Chem* **291**, 10950-10960 (2016).
54. J. M. Schifano *et al.*, tRNA is a new target for cleavage by a MazF toxin. *Nucleic Acids Res* **44**, 1256-1270 (2016).
55. H. R. Ramage, L. E. Connolly, J. S. Cox, Comprehensive functional analysis of Mycobacterium tuberculosis toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution. *PLoS Genet* **5**, e1000767 (2009).
56. L. Zhu *et al.*, The mRNA interferases, MazF-mt3 and MazF-mt7 from Mycobacterium tuberculosis target unique pentad sequences in single-stranded RNA. *Mol Microbiol* **69**, 559-569 (2008).
57. L. Zhu *et al.*, Characterization of mRNA interferases from Mycobacterium tuberculosis. *J Biol Chem* **281**, 18638-18643 (2006).
58. J. M. Schifano *et al.*, An RNA-seq method for defining endoribonuclease cleavage specificity identifies dual rRNA substrates for toxin MazF-mt3. *Nature communications* **5**, 3538 (2014).
59. M. E. Katz *et al.*, Identification of three gene regions associated with virulence in Dichelobacter nodosus, the causative agent of ovine footrot. *J Gen Microbiol* **137**, 2117-2124 (1991).
60. Y. X. Zhang *et al.*, Characterization of a novel toxin-antitoxin module, VapBC, encoded by Leptospira interrogans chromosome. *Cell Res* **14**, 208-216 (2004).
61. V. L. Arcus, P. B. Rainey, S. J. Turner, The PIN-domain toxin-antitoxin array in mycobacteria. *Trends Microbiol* **13**, 360-365 (2005).

62. V. L. Arcus, J. L. McKenzie, J. Robson, G. M. Cook, The PIN-domain ribonucleases and the prokaryotic VapBC toxin-antitoxin array. *Protein Eng Des Sel* **24**, 33-40 (2011).
63. U. Das *et al.*, Crystal structure of the VapBC-15 complex from Mycobacterium tuberculosis reveals a two-metal ion dependent PIN-domain ribonuclease and a variable mode of toxin-antitoxin assembly. *J Struct Biol* **188**, 249-258 (2014).
64. L. Miallau *et al.*, Structure and proposed activity of a member of the VapBC family of toxin-antitoxin systems. VapBC-5 from Mycobacterium tuberculosis. *J Biol Chem* **284**, 276-283 (2009).
65. A. B. Min *et al.*, The crystal structure of the Rv0301-Rv0300 VapBC-3 toxin-antitoxin complex from M. tuberculosis reveals a Mg(2)(+) ion in the active site and a putative RNA-binding site. *Protein Sci* **21**, 1754-1767 (2012).
66. J. L. McKenzie *et al.*, A VapBC toxin-antitoxin module is a posttranscriptional regulator of metabolic flux in mycobacteria. *J Bacteriol* **194**, 2189-2204 (2012).
67. K. S. Winther, K. Gerdes, Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. *Proc Natl Acad Sci U S A* **108**, 7403-7407 (2011).
68. K. S. Winther, D. E. Brodersen, A. K. Brown, K. Gerdes, VapC20 of Mycobacterium tuberculosis cleaves the Sarcin-Ricin loop of 23S rRNA. *Nature communications* **4**, 2796 (2013).
69. J. W. Cruz *et al.*, Growth-regulating Mycobacterium tuberculosis VapC-mt4 toxin is an isoacceptor-specific tRNase. *Nature communications* **6**, 7480 (2015).
70. B. A. Ahidjo *et al.*, VapC toxins from Mycobacterium tuberculosis are ribonucleases that differentially inhibit growth and are neutralized by cognate VapB antitoxins. *PLoS One* **6**, e21738 (2011).
71. A. P. Lopes *et al.*, VapC from the leptospiral VapBC toxin-antitoxin module displays ribonuclease activity on the initiator tRNA. *PLoS One* **9**, e101678 (2014).
72. J. D. Sharp *et al.*, Growth and translation inhibition through sequence-specific RNA binding by Mycobacterium tuberculosis VapC toxin. *J Biol Chem* **287**, 12835-12847 (2012).
73. K. Winther, J. J. Tree, D. Tollervey, K. Gerdes, VapCs of Mycobacterium tuberculosis cleave RNAs essential for translation. *Nucleic Acids Res* **44**, 9860-9871 (2016).

Chapter 2: Growth-regulating *Mycobacterium tuberculosis* VapC-mt4 toxin is an isoacceptor-specific tRNase

Jonathan W. Cruz*, Jared D. Sharp*, Eric D. Hoffer, Tatsuya Maehigashi, Irina O. Vvedenskaya, Arvind Konkimalla, Robert N. Husson, Bryce E. Nickels, Christine M. Dunham & Nancy A. Woychik

Mycobacterium tuberculosis is a pathogenic bacterium with many uncharacterized toxin–antitoxin (TA) systems, including the VapBC TA family. Here, in collaboration with the Bryce and Woychik labs, we developed a specialized RNA-seq approach to identify RNA targets of *Mycobacterium tuberculosis* VapC-mt4 toxin. They found that VapC-mt4 selectively cleaves three of the 45 *M. tuberculosis* tRNAs ($tRNA^{Ala2}$, $tRNA^{Ser26}$ and $tRNA^{Ser24}$) in their anticodons, resulting in the generation of tRNA halves. I used this information and generated a homology model of VapC-mt4 along with $tRNA^{Ala2}$ and $tRNA^{Ser26}$. I then performed simulated binding with VapC-mt4 and both tRNAs. The modeling suggests a possible mode of binding for VapC-mt4 to tRNA containing both short and long variable loops.

This research was originally published in *Nature communications*. Cruz J, Sharp J, Hoffer ED, Maehigashi T, Vvedenskaya I, Konkimalla A, Husson R, Nickels B, Dunham CM, Woychik N. Growth-regulating *Mycobacterium tuberculosis* VapC-mt4 toxin is an isoacceptor-specific tRNase. *Nature communications*. 2015 (6):7480.

Author contributions

J.W.C., J.D.S., I.O.V., R.N.H., B.E.N. and N.A.W. designed and performed all *in vitro* and *in vivo* experiments. E.D.H., T.M. and C.M.D designed the *in silico* modeling study. E.D.H. and T.M. performed modeling and simulated docking of VapC and tRNA **(Figure 2.8)**. J.W.C., J.D.S., E.D.H., T.M., R.N.H., B.E.N., C.M.D. and N.A.W. analyzed the data. J.W.C., E.D.H., T.M., B.E.N., C.M.D. and N.A.W. wrote the paper.

ABSTRACT

Toxin–antitoxin (TA) systems are implicated in the downregulation of bacterial cell growth associated with stress survival and latent tuberculosis infection, yet the activities and intracellular targets of these TA toxins are largely uncharacterized. Here, we use a specialized RNA-seq approach to identify targets of a *Mycobacterium tuberculosis* VapC TA toxin, VapC-mt4 (also known as VapC4), which have eluded detection using conventional approaches. Distinct from the one other characterized VapC toxin in *M. tuberculosis* that cuts 23S rRNA at the sarcin–ricin loop, VapC-mt4 selectively targets three of the 45 *M. tuberculosis* tRNAs ($tRNA^{Ala2}$, $tRNA^{Ser26}$ and $tRNA^{Ser24}$) for cleavage at, or adjacent to, their anticodons, resulting in the generation of tRNA halves. While tRNA cleavage is sometimes enlisted as a bacterial host defense mechanism, VapC-mt4 instead alters specific tRNAs to inhibit translation and modulate growth. This stress-linked activity of VapC-mt4 mirrors basic features of eukaryotic tRNases that also generate tRNA halves and inhibit translation in response to stress.

INTRODUCTION

Mycobacterium tuberculosis has adapted to survive a wide range of assaults—from our immune response to antimicrobial therapeutics—intended to eradicate the organism. However, we lack a full understanding of the molecular switches that enable *M. tuberculosis* to endure these stresses, to slow replication or to become dormant as a latent tuberculosis infection. Emerging studies on the molecular underpinnings of stress survival in Escherichia coli generally point to a major role for chromosomal toxin–antitoxin (TA) systems, which are operons comprising adjacent genes encoding two small (~10 kDa) proteins—a toxin and its cognate antitoxin that inhibits toxin activity in the TA protein–protein complex. The first described chromosomal TA system was *E. coli* mazEF, which was found to be induced by stress (1, 2). *E. coli* mazEF causes growth arrest and eventually cell death (1, 2). These studies and those on other TA systems in *E. coli* are consistent with a general role for TA systems in cell survival during periods of stress (3). Observations in *M. tuberculosis* suggest that TA systems are also important for stress survival in this organism. Several TA loci in *M. tuberculosis* are induced during heat shock (4), hypoxia (5, 6), DNA damage (7), nutrient starvation (8), macrophage infection (5, 9, 10) and antibiotic treatment (11, 12). Most recently, RNA-seq analysis of *M. tuberculosis* cells subjected to starvation revealed that the majority of TA systems, 75%, were upregulated to some degree with 25% upregulated twofold or higher (13).

The *M. tuberculosis* genome contains an estimated 48 members (5, 14) of the VapBC (virulence-associated protein) family, the highest number of VapBC TA systems among free-living bacteria. VapC toxins are characterized by the presence of a PIN (PilT amino-terminal) domain. The 48 *M. tuberculosis* VapC toxins all share protein sequence similarity and possess a PIN domain containing a conserved quartet of acidic residues and a fifth invariant serine or threonine residue, which are responsible for coordinating

divalent cation(s) in the catalytic centre (15). The presence of a PIN domain suggests a role for the VapC toxins as ribonucleases, yet the body of literature on the enzymatic activity of VapC toxins has been inconsistent and contradictory. In our opinion, the most consistent data are derived from five recent reports. First, the solitary VapC toxin in *Shigella* or *Salmonella* cleaves $tRNA^{fMet}$ at a single, identical site in the anticodon stem loop (ASL) (16). The solitary VapC in the spirochaete *Leptospira interrogans* also cleaves $tRNA^{fMet}$ (17). Second, the VapC from *M. smegmatis* cleaves synthetic ssRNAs at a short consensus sequence with some dependence on secondary structure (18). In vivo, this VapC cleaves mRNA transcripts to downregulate *M. smegmatis* glycerol uptake and metabolism (18). Third, we determined that a representative *M. tuberculosis* VapC toxin, VapC-mt4 (Rv0595c, also known as VapC4), recognizes ACGC or AC(A/U)GC in RNA (19). Fourth, another *M. tuberculosis* VapC toxin, VapC20 (VapC-mt20 using our convention), cleaves at the highly conserved sarcin–ricin loop of 23S rRNA in intact ribosomes only, with some sequence specificity (20).

A detailed understanding of the properties of the 48 VapC paralogs is essential for the interpretation of their physiological role in *M. tuberculosis* and other pathogens. In this study, we identify the primary target of VapC-mt4 using a specialized RNA-seq approach. Distinct from all other VapC toxins, and TA toxins in general, VapC-mt4 arrests growth by translation inhibition resulting from selectively targeting three of the 45 tRNAs present in *M. tuberculosis* for cleavage at a single site in their anticodon loop. This highly selective tRNA substrate discrimination is contingent on recognition of the consensus sequence in an appropriate structural context. In agreement, VapC-mt4-tRNA-simulated docking experiments place the toxin active site in proximity to the cleavage site in the tRNA anticodon loop. Overall, these studies bring to light a common theme between the role of VapC-mt4 in this pathogen and stress responses in eukaryotic cells that also engage cleaved tRNAs in unconventional roles.

RESULTS

RNA-seq reveals VapC-mt4 cleaves specific tRNA isoacceptors. Our earlier study of VapC-mt4 revealed that this toxin inhibits translation and cleaves RNA at a consensus sequence of ACGC or AC(U/A)GC and that the GC sequence within this motif is essential for cleavage (19). However, we were unable to clearly identify a primary target for this toxin. Although our data suggested that tRNA(s) may be the primary targets for *M. tuberculosis* VapC-mt4, we were unable to establish this connection using a battery of conventional biochemical, genetic and molecular biological approaches (19). To overcome this roadblock, we used a specialized RNA-seq method we recently developed to identify RNA targets of endoribonucleolytic toxins (21). This RNA-seq method was designed to differentially detect RNA cleavage products that carry a 5'-hydroxyl (OH) or a 5'-monophosphate (P). Bacterial transcripts possess a 5'-triphosphate (mRNAs), 5'-monophosphate (rRNAs and tRNAs) or 5'-hydroxyl (noncoding RNA intermediates and products cleaved by certain endoribonucleolytic toxins such as MazF or RNases such as RNase A and T1(22). Another important feature of this RNA-seq method is its use of *E. coli* as a surrogate host, which unlike *M. tuberculosis*, does not contain a 5'-to-3' exoribonuclease. Thus, the 5' ends generated by VapC-mt4 cleavage can be readily detected because they will be stable.

Using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, McKenzie *et al*. (23) reported that VapC toxins generate a 5'-P cleavage product. However, Winther and Gerdes suggested that the ends were instead marked with a 5'-OH because the 5' terminus of the 3' VapC RNA product could be phosphorylated without a prior dephosphorylation step (16). Therefore, we constructed both 5'-P and 5'-OH cDNA libraries from RNA harvested from exponentially growing *E. coli* cells expressing VapC-mt4 from an arabinose-inducible vector. We did not identify any RNAs enriched by 10-fold or more in the 5'-P library relative to the uninduced control. In contrast, in the 5'-OH library,

a cleavage product derived from three identical tRNAs—$tRNA^{Ala2}$, $tRNA^{Ala41}$ and $tRNA^{Ala58}$—exhibited a 171-fold increase in transcript abundance, relative to the uninduced control **(Fig. 2.1a)**. For simplicity, we will refer to these three identical tRNAs as $tRNA^{Ala2}$ from this point forward (numbering based on the GtRNAdb Genomic tRNA Database (24)). The sequence immediately upstream and downstream of the 5' end of the cleavage site, $A^{37}C^{38}\downarrow G^{39}C^{40}$ (where ↓ indicates the position of cleavage) was a 100% match to the recognition sequence deduced from our published in vivo and in vitro primer extension experiments (19). Generation of the putative $tRNA^{Ala2}$ cleavage product was specific to VapC-mt4 induction; we did not detect any other tRNA fragment that displayed more than a 10-fold increase in abundance in the comparison of 5'-OH libraries derived from VapC-mt4-induced cells relative to uninduced cells **(Fig. 2.9)**.

Validation of VapC-mt4 specificity for *E. coli* $tRNA^{Ala2}$. We proceeded to confirm our RNA-seq result **(plotted in Fig. 2.1a)** suggesting that $tRNA^{Ala2}$ is a major target for cleavage by VapC-mt4 using two approaches. In the first approach, we performed primer extension on the RNA samples from *E. coli* cells 0, 20, 40, 60, 80 and 100 min after induction. A strong cleavage product accumulated with time at the same site in $tRNA^{Ala2}$ identified by RNA-seq **(Fig. 2.1b and Fig. 2.11)**. The validated RNA-seq cleavage site mapped to the junction of the stem and loop in the ASL **(Fig. 2.1c)**.

For our second approach, we purified $tRNA^{Ala2}$ from total *E. coli* RNA using a biotinylated oligonucleotide complementary to a sequence in the ASL unique to this isoacceptor. Incubation of pure VapC-mt4 with this $tRNA^{Ala2}$ led to complete cleavage with increasing toxin concentration, generating products consistent with cleavage at the $A^{37}C^{38}\downarrow G^{39}C^{40}$ site in the ASL identified by RNA-seq and primer extension **(Fig. 2.1d)**. Cleavage was abolished when VapC-mt4 was preincubated with its cognate antitoxin VapB-mt4 and when EDTA was added to the reaction to chelate Mg^{2+} **(Fig. 2.1d)**.

Although none of the bases comprising the ACGC sequence in *E. coli* $tRNA^{Ala2}$ cleaved by VapC-mt4 are modified, modifications in proximal bases may enhance cleavage (for example, there is a 7-methylguanosine at position 46 of $tRNA^{Ala2}$). To address this, we compared the efficiency of cleavage of $tRNA^{Ala2}$ produced in vivo containing modified bases **(Fig. 2.1d)** with that synthesized in vitro lacking modified bases **(Fig. 2.1e)**. There was no difference in the efficiency of cleavage between the two versions of $tRNA^{Ala2}$, indicating that base modifications in this tRNA are not required for VapC-mt4 cleavage. Cleavage by VapC-mt4 was also specific for $tRNA^{Ala2}$ since non-ACGC-containing tRNAs ($tRNA^{fMet(19)}$) or other ACGC-containing tRNAs (for example, $tRNA^{Asp3}$) subjected to the same reaction as Fig. 1e were not cleaved. Therefore, our RNA-seq approach was a valid predictor of the VapC-mt4 RNA target and site of cleavage in *E. coli* cells because $tRNA^{Ala2}$, but no other tRNA **(Fig. 2.9)**, was cleaved at the same site both in vivo and in vitro.

VapC-mt4 cleaves three *M. tuberculosis* tRNAs within the ASL. Our RNA-seq analysis suggested that VapC-mt4 generally targets tRNA and not rRNA as recently reported for VapC-mt20 (20). Since VapC-mt4 cleavage of $tRNA^{Ala2}$ did not require base modifications **(Fig. 2.1d, e)**, we used synthetic tRNAs to test toxin specificity for all *M. tuberculosis* tRNAs containing a consensus sequences. ACGC was the predominant in vivo cleavage site identified by our earlier in vivo primer extension experiments(19) and was also represented in the highest ranking RNA-seq target, $tRNA^{Ala2}$. We surveyed all 45 tRNAs in *M. tuberculosis* (all but three—two $tRNA^{Met}$ and one $tRNA^{fMet}$—have unique anticodons) and identified 13 distinct tRNAs with ACGC or the related ACAGC or ACUGC cleavage consensus sequences(19). We synthesized each of these 13 *M. tuberculosis* tRNAs and subjected them to the same VapC-mt4 in vitro cleavage assay used in Fig. 1d, e. VapC-mt4 exhibited a strong preference for three of the 13 tRNAs (that is, complete cleavage of full-length tRNA), $tRNA^{Ala2}$ **(Fig. 2.2a)**, $tRNA^{Ser26}$ **(Fig. 2.2b)** and $tRNA^{Ser24}$ **(Fig.**

2.2c). Preincubation of VapC-mt4 with either its cognate antitoxin VapB-mt4 or EDTA prevented tRNA cleavage. $tRNA^{Glu33}$, representative of the 10 tRNAs that were either weakly cut or not cut at all, is shown in Fig. 2d. The entire set of 10 consensus site-containing tRNAs that were tested for cleavage and not classified as a VapC-mt4 target is shown in **Fig. 2.11**). Therefore, VapC-mt4 efficiently cleaved only 3 of 13 *M. tuberculosis* tRNAs harbouring the cleavage consensus sequence. Although VapC-mt4 cleaved tRNAs with the same name ($tRNA^{Ala2}$ from *E. coli* and *M. tuberculosis*), these tRNAs are not identical (71% identical overall among their 76 nts). However, there is high identity in the most relevant portion of these two tRNAs: the anticodon loop. The first 7 nts (which includes the anticodon) of the 9 nt loop are identical.

We then used primer extension to identify the position of VapC-mt4 cleavage in the synthetic tRNAs shown in Fig. 2. Using this approach, all three tRNA cleavage sites mapped within the anticodon loop **(Fig. 2.3a, c, e and Fig. 2.13)**. Finally, we confirmed that VapC-mt4 efficiently cleaves tRNAs from *M. tuberculosis* cells. We isolated total RNA from *M. tuberculosis* and incubated it with VapC-mt4 alone or VapC-mt4 preincubated with its cognate antitoxin VapB-mt4 as a control. We then performed northern analysis with an isoacceptor-specific oligonucleotide complementary to the ASL for $tRNA^{Ser26}$ or $tRNA^{Ser24}$ **(Fig. 2.4a, b)**; we also performed a similar experiment for *E. coli* $tRNA^{Ala2}$ as a control **(Fig. 2.4c)**. Addition of VapC-mt4 resulted in the loss of hybridization, indicating that the respective tRNAs were cleaved **(Fig. 2.4a–c)**. In the case of *M. tuberculosis* $tRNA^{Ala2}$, sequence similarity among alanine isoacceptor tRNAs was too high to enable selective hybridization for only $tRNA^{Ala2}$. In summary, VapC-mt4 cleaves three *M. tuberculosis* tRNAs within their anticodons ($tRNA^{Ser24}$ and $tRNA^{Ser26}$; when referring to both of these tRNAs they will subsequently be called $tRNA^{Ser24/6}$) or immediately 3' of the anticodon ($tRNA^{Ala2}$). As with *E. coli* $tRNA^{Ala2}$, this cleavage appears to be efficient in the presence or absence of tRNA modifications.

VapC-mt4 cleavage requires an ACGC in the proper context. VapC-mt4 appears to require more than the consensus sequence alone since only $tRNA^{Ala2}$ was cleaved even though there are 12 *E. coli* tRNAs containing a VapC-mt4 cleavage consensus sequence. Therefore, to better characterize the properties of RNA recognition and cleavage by this toxin, we first tested if VapC-mt4 cleavage at these consensus RNA sequences required single-stranded (ss) or double-stranded (ds) RNA. As with the MazF toxin (25), VapC-mt4 also requires that the cleavage consensus sequence resides within ssRNA, since an RNA fragment containing the ACGC consensus sequence could only be cleaved as a ssRNA **(Fig. 2.5, lanes 2,4)**. Increasing concentrations of a complementary RNA fragment (leading to the formation of dsRNA) resulted in a gradual decrease in cleavage **(Fig. 2.5, lanes 4–9)**.

We next enlisted *E. coli* $tRNA^{Asp3}$ as a tool to study the importance of context. $tRNA^{Asp3}$ has the ACGC sequence located in the identical position as in *E. coli* $tRNA^{Ala2}$, yet VapC-mt4 only cleaved $tRNA^{Ala2}$ **(Fig. 2.6 compare 'wild-type' tRNA red sequences)**. Interestingly, although these two tRNAs are 66% identical overall they differ at only three positions in their entire anticodon loop, two bases in the anticodon and one base just 5' of the anticodon **(Fig. 2.6 compare sequences in 'wild-type' illustrations)**. To establish if the sequence of the anticodon influenced cleavage, we mutated the $tRNA^{Ala2}$ anticodon to a GUC (the anticodon for $tRNA^{Asp3}$) and the $tRNA^{Asp3}$ anticodon to a UGC (the anticodon for $tRNA^{Ala2}$; Fig. 6 'mutant' panels). Wild-type $tRNA^{Ala2}$ was efficiently cleaved, but the anticodon mutant version was no longer cleaved **(Fig. 2.6a 'mutant' panel)**. Conversely, $tRNA^{Asp3}$ with a wild-type anticodon was not cleaved, while the mutated version was efficiently cleaved **(Fig. 2.6b 'mutant' panel)**. Therefore, the ACGC sequence is not the only determinant for RNA cleavage by VapC-mt4; the consensus sequence must be single-stranded and positioned in the tRNA such that all required conditions for proper recognition and cleavage are met.

VapC-mt4 recognizes sequence and structural determinants. Having established that sequences in the anticodon immediately adjacent to a consensus sequence influence *E. coli* tRNA cleavage by VapC-mt4, we performed more detailed studies to further interrogate the importance of sequence and secondary structure elements within a biologically relevant subset of VapC-mt4 targets. We chose *M. tuberculosis* $tRNA^{Ser24}$ from the three tRNAs cleaved by VapC-mt4 because it also contains an ACGC consensus site that bridges the junction between the anticodon stem and the anticodon loop, analogous to *E. coli* $tRNA^{Ala2}$ **(Fig. 2.7)**.

We had previously determined that the GC residues in the ACGC recognition sequence are essential for VapC-mt4 cleavage of a synthetic 20-nt RNA template[19]. To test the importance of the first two AC residues, we mutated each individually, incubated with VapC-mt4, and assessed the extent of cleavage **(Fig. 2.7a–c)**. First, we synthesized $tRNA^{Ser24}$ with the first A in the ACGC sequence mutated to a C. In comparison with the complete cleavage of wild-type $tRNA^{Ser24}$ using the highest concentration (10 pmol) of VapC-mt4, an ACGC→CCGC mutant was cleaved less efficiently **(Fig. 2.7a, b)**. Second, we synthesized an ACGC→AAGC mutant and found that it was cleaved as efficiently as wild type **(Fig. 2.7c)**. Third, we altered the ACGC consensus sequence to create minor or major changes in secondary or tertiary structure of the ASL. An ACGC→ACAC mutation, now lacking the G–C base pair at the stem of the ASL as well as one of the two critical consensus residues, severely impaired cleavage **(Fig. 2.7d)**. Therefore, three of the four bases in the VapC-mt4 ACGC cleavage consensus sequence are important for cleavage. Cleavage was less efficient when the first A of the consensus was mutated, and mutation of the G or terminal C (previously reported by Sharp *et al*. (19)) abolished cleavage. Finally, introducing mutations in the ASL that disrupt base pairing while keeping the ACGC consensus intact, yielded a marginal level of cleavage but at a different position **(Fig.**

2.7e). We conclude that recognition of specific tRNA targets by VapC-mt4 requires proper sequence and structural context.

Simulated docking of VapC-mt4 and tRNA. Interestingly, in contrast to *E. coli* $tRNA^{Ala2}$ (whose cut site is $A^{37}C^{38}\downarrow G^{39}C^{40}$), the cleavage sites for each of the three *M. tuberculosis* tRNAs are in proximity to, but not within their consensus sequences. To model how VapC-mt4 may potentially interact in distinct ways with either $tRNA^{Ala2}$ or $tRNA^{Ser24/6}$, we performed protein–RNA docking simulations using the program 3dRPC (26) **(Fig. 2.8)**. Since a structure of VapC-mt4 has not been determined, we generated a homology model of VapC-mt4 using the HHpred homology model workflow and Modeller (27, 28) from three published structures: *M. tuberculosis* VapBC5 (PDB accession code 3DBO (29)), Rickettsia VapBC2 (PDB accession code 3ZVK (30)) and FitAB (PDB accession code 2H1C (31)). On the basis of the published stoichiometry of VapC homologues (32, 33), a VapC-mt4 dimer was generated through simulated docking using GRAMM-X (27, 34) and then the interactions between the complex were optimized with RosettaDock (35). There are also no structures for any of the three VapC-mt4 targets. But $tRNA^{Phe}$ and $tRNA^{Leu}$ (PDB accession codes 4TNA, chain A and 3UZ3, chain B (36, 37)) are close homologues of $tRNA^{Ala2}$ and $tRNA^{Ser24/6}$, respectively. We used $tRNA^{Leu}$ as a representative for both serine isotypes recognized by VapC-mt4 as they all contain an expanded variable loop. The position of a divalent cation in our model was extrapolated by superpositioning the structure of *Pyrobaculum aerophilum* VapC with a coordinated divalent cation with our VapC-mt4 model generated by DaliLite (PDB accession code 2FE1 (32)) (38). On the basis of the alignment we predict that VapC-mt4 residues Asn98 and Asn116 likely coordinate a metal ion in its active site due to the proximity of the *P. aerophilum* VapC equivalent.

While both docked models indicate that one active site from each VapC-mt4 dimer is positioned adjacent to the experimentally determined tRNA cleavage site located in the

anticodon loop, the overall $(VapC\text{-}mt4)_2$ orientations differ depending on the tRNA it recognizes. In the VapC-mt4 dimer-$tRNA^{Phe/Ala2}$ model, one monomer of VapC-mt4 is predicted to interact with the base paired anticodon stem of the tRNA with potential additional contacts with the D loop **(Fig. 2.8a)**. The other monomer of VapC and its active site is modelled adjacent to the predicted cleavage site between residues 36 and 37 **(Fig. 2.8a, right panel).** However in the VapC-mt4 dimer-$tRNA^{Leu/Ser26}$ model, the large expanded variable loop likely causes the protein/toxin to recognize the anticodon stem closer toward the anticodon **(Fig. 2.8b)**. Large expanded variable loops are present in both $tRNA^{Leu}$ and $tRNA^{Ser}$; therefore, we predict this influences the location of the cleavage site. Also in this case, the active site in one VapC monomer is positioned adjacent to the tRNA backbone near cleavage site residue 36 **(Fig. 2.8b, right panel)**. Interestingly, the $tRNA^{Ala}$ model shows significant interactions with the 5'-ACUGC-3' consensus sequence in contrast to the $tRNA^{Ser}$ model, which shows minimal interactions **(Fig. 2.8c).**

DISCUSSION

VapBC modules are among the most prevalent TA systems in the genomes of a variety of pathogens, comprising >40% of the ~700 TA modules identified in 126 complete genomes of free-living bacteria (39). Therefore, when organisms carry multiple members of a TA system family, VapBC systems are inexplicably the most abundant. The multiple iterations of VapBC TA systems (though none are identical) are associated with increased virulence in the pathogens that contain them, accounting for the origin of the name Vap, for virulence-associated protein (40). A detailed understanding of the properties of *M. tuberculosis* VapC paralogs is essential for interpretation of their physiological role in *M. tuberculosis* and other pathogens. We studied a representative family member, *M. tuberculosis* VapC-mt4 (aka VapC4), to obtain a better understanding of the RNA targets and molecular mechanism of VapC toxin recognition and cleavage. In contrast to VapC-

mt20, which cleaves 23S rRNA at the sarcin–ricin loop (20), VapC-mt4 specifically targeted just three of the 45 total *M. tuberculosis* tRNA isotypes—$tRNA^{Ala2}$, $tRNA^{Ser26}$ and $tRNA^{Ser24}$. All three of these tRNAs contain one of the cleavage consensus sequences for this toxin identified in our earlier report, namely ACGC or ACUGC. Two tRNAs contained the ACUGC recognition sequence, which was the strongest cleavage site among 12 sites in the 3.5-kb bacteriophage MS2 RNA template that were cut by VapC-mt4 (19). Thus, cleavage of tRNA involves sequence recognition by VapC-mt4. This contrasts with 23S rRNA cleavage by VapC-mt20, which appears to be more dependent on recognition of the stem loop structure of the sarcin–ricin loop (20). In fact, VapC-mt4 differs from all other TA system toxins, the majority of which (1) predominantly cleave mRNA independent of the ribosome and at all cleavage consensus sequences (MazF family) or (2) cleave mRNA in a ribosome-dependent manner with no clear sequence specificity (RelE family) or some degree of sequence specificity (HigB, YafQ toxins).

Tertiary fold or recognition of determinants in the sugar phosphate backbone appears to also contribute to the selectivity of VapC-mt4, given that many other *M. tuberculosis* tRNAs also harbor the same recognition sequence yet only a few are efficiently cleaved. While tRNAs generally adopt L-shaped three-dimensional structures to enable entry into the ribosome, there may be some subtle variation in the overall structure that can be distinguished by VapC-mt4. In fact, while all tRNAs possess three stem-loops (D, anticodon, TCC), the cloverleaf secondary structures of tRNA do vary slightly. Type II tRNAs ($tRNA^{Ser26}$ and $tRNA^{Ser24}$ in our case) also contain an extended variable arm between the anticodon and TCC stem loop **(Fig. 2.3)**. Also, although the D arm usually has a 4-bp stem, $tRNA^{Ser26}$ and $tRNA^{Ser24}$, have a 3-bp stem. In fact, our simulated docking experiments indicate that the variable loop influences the positioning of the VapC-mt4 dimer on the tRNA, which we propose alters the location of the cleavage site **(Fig. 2.6)**. Independent of tertiary fold, the interaction of VapC-mt4 with the sugar-

phosphate backbone may contribute to cleavage specificity. On the basis of studies of tRNA recognition for aminoacylation of $tRNA^{Gln}$ by glutaminyl-tRNA synthetase, class II tRNA synthetase interaction with the sugar-phosphate backbone is an important recognition determinant in addition to base recognition at the anticodon loop, at the acceptor end, and with the interior base pair of the D stem (41).

Strikingly, VapC-mt4 cuts all three *M. tuberculosis* tRNA targets precisely within the anticodon loop. In the case of $tRNA^{Ala2}$, VapC-mt4 cleaves at the position immediately 3' adjacent to the anticodon **(Fig. 2.3b)**. For $tRNA^{Ser26/4}$, VapC-mt4 cleavage occurs within the anticodon itself between the second and the third nucleotides **(Fig. 2.3d,f)**. However, there was no correlation between the tRNA isoacceptors targeted for cleavage and their corresponding codon usage preferences. $tRNA^{Ala2}$ decodes the lowest percentage of alanine codons (10%) in *M. tuberculosis*, while $tRNA^{Ser26}$ (26%) and $tRNA^{Ser24}$ (21%) decode intermediate percentages of serine codons (with $tRNA^{Ser28}$ decoding the highest number of serine codons, 35%). Since the ASL of tRNA interacts with the decoding centre of the 30S subunit, cleavage at the anticodon sequence of the ASL by VapC-mt4 is consistent with the known inhibition of translation by VapC-mt4 on overexpression in *E. coli* using an arabinose inducible plasmid (~10% of wild type (19)).

In some strains of *E. coli*, colicins are secreted into the environment to kill the neighbouring bacterial cells and reduce competition[42]. Colicins D and E5 enlist cleavage of tRNA at a single site within the ASL to inhibit translation and cause cell death. The lethality of colicin D is due to specific cleavage of all four $tRNA^{Arg}$ isoacceptors between positions 38 and 39, corresponding to the junction of the loop and stem on the 3' side of the anticodon (42). Likewise, Salmonella/Shigella VapC cleaves $tRNA^{fMet}$ at the same position as colicin D (16). As with VapC-mt4, colicin E5 cleaves within the anticodon loop, at positions 34 and 35 of the anticodon (43). Also serving a defensive role, the *E. coli* endoribonuclease PrrC triggers cell suicide by cleaving $tRNA^{Lys}$ between anticodon

positions 33 and 34 in response to bacteriophage T4 infection of *E. coli* cells (44-46). Therefore, colicins D, E5 and PrrC all cause cell death. In the case of colicins D and E5, death is undoubtedly due to the absence of their cognate immunity proteins in neighbouring cells that take up the colicins alone (42, 47). For PrrC-mediated cell death, it is thought that after T4 phage infection, the Stp phage protein alters the association of the 'antitoxin' EcoprrI with PrrC, releasing free PrrC and leading to unchecked cleavage of lysine tRNA (44-46).

In contrast, the general physiological role of most TA systems, including VapC-mt4, appears to involve stress survival and not cell death due to the presence of a cognate antitoxin that dynamically regulates toxin activity in response to stress (3). In fact, as demonstrated for the MazF toxin, cell death can be precluded following toxin activation when the cognate antitoxin is also expressed (2). Cell death only occurs after prolonged activation of the toxin, when the cell reaches 'the point of no return'(2). Our earlier study showed that VapC-mt4 overexpression results in translation inhibition with accumulation of polysomes, indicative of a block at the elongation step (19). Possibly, ribosomal stalling may result from a non-productive interaction of the cleaved ASL with the A site of the ribosome. In addition, two studies have demonstrated that cleavage or alteration of the ASL leads to defects in aminoacylation. First, a break in the phosphodiester backbone of the ASL between position 34 and 35 of yeast $tRNA^{Met}$ inhibits aminoacylation (44). Second, aminoacylation by *E. coli* glutaminyl-tRNA synthetase requires recognition of the sugar-phosphate backbone at positions 34–38 (41). Therefore, disruption of the backbone by cleavage between position 36 and 37 by VapC-mt4 would also likely disrupt aminoacylation by class II synthetases, potentiating its effect on translation and growth.

VapC-mt4 targets a small subset of *M. tuberculosis* tRNAs with exquisite specificity. It not only distinguishes between tRNA isotypes, it selectively cleaves some, but not all, isoacceptors. More specifically, VapC-mt4 cleaves one isoacceptor of $tRNA^{Ala}$

($tRNA^{Ala2}$) of the three represented in *M. tuberculosis* and two of the four isoacceptors of $tRNA^{Ser}$ ($tRNA^{Ser26}$ and $tRNA^{Ser24}$) are cleaved efficiently. Although colicins D and E5 exhibit specificity for tRNAs, their tRNA targets are either limited to all four $tRNA^{Arg}$ isoacceptors (colicin D (42)) or extend to four isotypes ($tRNA^{Asn}$, $tRNA^{Asp}$, $tRNA^{Tyr}$ and $tRNA^{His}$ for colicin E5 (43). Since our data identifies the presence of the cleavage consensus sequence ACGC or ACUGC as one major determinant for tRNA cleavage, some of the selectively of VapC-mt4 is attributed to this requirement since not all isoacceptors of $tRNA^{Ser}$ harbour this essential sequence. In contrast, the sequence requirements for colicins appear to be less specific than for VapC-mt4—colicin E5 cleaves at (C/T)GU (48), while the interrogation of sequence requirements for colicin D has not been reported (42). No atomic level structures of colicins bound to tRNAs have been reported. Structural studies of VapC-mt4 bound to tRNA will be instrumental in pinpointing additional determinants (tertiary fold and/or sugar-phosphate backbone interactions) that guide toxin specificity for a handful of tRNA targets.

Although the precise physiological triggers for VapC-mt4 have not yet been identified, we know that the expression of this toxin leads to growth inhibition in *M. tuberculosis* (14, 19). In fact, with few exceptions, the unifying feature of TA systems is their role as stress responsive growth regulators. Consistent with this role, there is mounting evidence linking increased expression of TA systems with several stresses in *M. tuberculosis*, especially those encountered during latent tuberculosis (4-13). As a complement to transcript-oriented studies, recent proteomic profiles generated from parallel analyses by mass spectrometry and two-dimensional differential gel electrophoresis reinforced the role of TA systems in stress (49). Culture filtrates from *M. tuberculosis* cells after nutrient starvation contained elevated levels of 13 TA toxins, including a 410-fold increase in VapC-mt4 (49).

The need for the high number of VapC toxins in *M. tuberculosis* (48 predicted) is not yet clear. All VapC toxins are expected to function as nucleases due to the presence of a PIN domain. However, our findings reported here coupled with those of the only other well-characterized *M. tuberculosis* VapC toxin—VapC20, which targets the 23S rRNA sarcin–ricin loop (20)—suggest that the VapC family members will not be functionally redundant. Instead, each VapC may cleave unique substrates or sites within a given target, thus allowing a defined molecular response to one or more stresses. Given that *M. tuberculosis* cells can remain dormant yet viable for many years in latent tuberculosis in a granuloma, the resulting activity of each toxin may impart some degree of growth downregulation that is then lifted before cells reach the point of no return. Thus, through activation of a shifting array of VapC toxins in a calibrated, asynchronous manner, *M. tuberculosis* may be able to limit growth while never reaching the point of cell death.

Interestingly, the presence of cleaved tRNA halves and smaller fragments is also emerging as a hallmark of stress responses in many eukaryotes (reviewed in refs (50-52)). In this rapidly evolving area, tRNA halves or other tRNA-derived fragments have been associated with a range of functions, including translational control, apoptosis and RNA interference regulation. Therefore, our results uncover a new parallel between the VapC-mt4 toxin mediated stress response in *M. tuberculosis* and posttranscriptional tRNA cleavage at the ASL during stress in eukaryotes. Since the activities of only two VapC toxins have been characterized, it is highly likely that more VapC toxins will also target tRNAs. This would further expand the repertoire of tRNAs acting in a non-conventional role, shed light on the conundrum of why *M. tuberculosis* cells have so many VapBC TA systems and strengthen our understanding of how their orchestrated expression impacts stress survival, growth and virulence.

METHODS

Strains and plasmids. The *E. coli* strain BL21(DE3) (F-*omp*T *hsd*S_β(r_β^+-m_β^+) *dcm gal* (DE3) tonA) (Novagen) was used for protein expression. The *E. coli* strain BW25113Δ6 (F$^-$ *lacI*^{q} *rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ *hsdR514* Δa*raBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$ Δ*mazEF* Δ*chpBIK* Δ*relBE ΔyefM-yoeB ΔDinJ-yafQ*) (21) was used for the isolation of RNA. *E. coli* K-12 Mach1 T1 cells (ΔrecA1398 endA1 tonA Φ80ΔlacM15 Δ*lac*X74 *hsd*R($r_k^+m_k^+$); Invitrogen) were used for all cloning experiments. The pBAD33-VapC-mt4 and pET28a-His$_6$-VapC-mt4 was described in our earlier work (19). All *E. coli* liquid cultures were grown in M9 minimal media supplemented with 0.2% (w/v) casamino acids, 1 mM $MgSO_4$, 0.001% (w/v) thiamine and either 0.2% glucose or 0.2% glycerol at 37 C. 1 mM isopropyl 1-thio-b-D-galactopyranoside, 0.2% arabinose, 45 mg ml^{-1} kanamycin and 25 mg ml^{-1} chloramphenicol were added to culture medium as required.

Purification of recombinant VapC-mt4-His$_6$. *E. coli* BL21(DE3)pLysE was transformed with the pET28a-His$_6$-VapC-mt4 plasmid. Transformants were selected and grown in 1 l of M9 minimal media containing 0.2% glucose at to an OD_{600} of 0.6–0.8 and toxin expression induced with isopropyl 1-thio-b-D-galactopyranoside for 4 h. Recombinant VapC-mt4 was then purified as previously described (19). In brief, the cells were lysed by sonication. Cell lysates were centrifuged at 32,000g for 10 min to remove cell debris. Protein was then purified from the cleared cell lysates by nickel–nitrilotriacetic acid affinity chromatography (Qiagen).

Double-stranded RNA cleavage. A 20-base ACGC-containing RNA oligonucleotide, NWO1430 (5'-AGGAAGAUACGCGAUAUGAA-3'), was 5' end labelled with [γ-^{32}P]-ATP using T4 polynucleotide kinase (NEB) following the manufacturer's protocol. The labelled oligonucleotide was diluted with diethylpyrocarbonate (DEPC) -treated water to yield a

final concentration of 0.2 pmol ml ^{1}. The labelled RNA was then mixed with different amounts (at ratios of 0:1, 0.1:1, 0.2:1, 0.4:1, 0.8:1 and 1:1) of a complementary RNA oligo, NWO1541 (5'-UUCAUAUCGCGUAUCUUCCU-3'). The two oligonucleotides were incubated at 70 C for 3 min and then allowed to cool slowly to room temperature (B30 min). Samples were adjusted to 10 mM Tris pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$; 20 units of RNase inhibitor (NEB) was then added to each reaction followed by 17 pmol of VapC-mt4 (when indicated) in a final volume of 10 ml. The samples were incubated at 37 C for 1 h and reactions stopped by adding 10 ml of sequence loading buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF). Samples were then heated to 95 C for 5 min before electrophoresis on a 15% polyacrylamide, 7 M urea gel.

In vitro synthesis of tRNA. *E. coli* and *M. tuberculosis* tRNAs were synthesized in vitro following the method described by Sisido *et al*. (53) with minor modifications. In brief, a synthetic DNA oligonucleotide containing the T7 RNA polymerase promoter and the 5' end of the tRNA gene of interest was annealed to a second oligonucleotide corresponding to the 3' end of the tRNA gene. The annealed oligonucleotides were then extended using Taq DNA polymerase to create dsDNA containing the entire tRNA gene preceded by the T7 promoter. The product was then resolved on a 2% agarose gel to confirm its size and purified using the QIAquick Gel Extraction Kit (Qiagen). The sequence of the product was verified by automated DNA sequence analysis. Two-hundred ng of the tRNA gene was then transcribed in vitro using the RiboMAX Large Scale RNA Production System (Promega) as recommended by the manufacturer. The transcription reaction was separated on a 9% polyacrylamide, 7 M urea gel and visualized by staining with EtBr to confirm the size and purity of the transcribed tRNA. The tRNA transcript was then excised from the gel and incubated for 18 h at 37 C in elution buffer (1 mM EDTA, 0.5 M ammonium acetate, 10 mM magnesium acetate, 0.1% SDS). The eluate was then collected, and the

gel pieces were washed in fresh elution buffer. The elution and wash were combined, the tRNA was ethanol precipitated and resuspended in nuclease-free water.

Purification of tRNA from *E. coli* cell extracts. Total *E. coli* RNA was isolated using the hot phenol method. A culture of *E. coli* BW25113D6 was grown in M9 medium until the OD_{600} reached a value of B0.6. The cells were pelleted and resuspended in 500 ml Buffer A (0.5% SDS, 20 mM sodium acetate, pH 4.0; 10 mM EDTA) and 500 ml phenol, pH 4.0. After vortexing to completely resuspend the cell pellet, the samples were heated to 60 C for 10 min and centrifuged to separate the phases. The RNA-containing aqueous phase was removed, ethanol precipitated and resuspended in nuclease-free water. Individual tRNAs were then isolated from the total RNA using a biotinylated oligonucleotide that specifically annealed to $tRNA^{Ala2}$ adapted from the procedure of Yokogawa *et al*.(54). We first washed 450 ml of MagneSphere magnetic beads (Promega) three times (we used a magnetic stand to recover the beads after all wash and hybridization steps) with 600 ml of wash buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA and 1 M NaCl). We then added 600 pmol of the 5' biotinylated oligonucleotide for $tRNA^{Ala2}$ isolation, NWO1974 (5'-Biotin-AGACCTCCTGCGTGCAAAGCAGGC-3'), adjusted the volume to 300 ml with 10 mM Tris-HCl pH 7.5 and incubated at room temperature for 15 min with gentle inversion every 3 min. After binding of the $tRNA^{Ala2}$-specific oligonucleotide, the beads were washed three times with wash buffer. After the final wash, the beads were resuspended in 50 ml hybridization buffer (20 mM Tris-HCl pH 7.6, 1.8 M tetramethylammonium chloride (TMA-Cl), 0.2 mM EDTA). One mg of total RNA in 50 ml nuclease-free water was then added to the beads. Samples were then gently inverted, heated to 65 C for 3 min and placed at 60 C for 20 min with inversion every 4 min. The supernate containing unbound RNA was separated from the beads and transferred to a fresh tube and saved for rebinding. The beads containing the bound tRNAs were washed six times with 800 ml Tris-HCl, pH 7.6.

The target tRNA was then eluted from the beads with 150 ml of Tris-HCl, pH 7.6 heated to 75 C for 5 min and transferred to another tube containing 2 ml of 1 M $MgCl_2$ to assist in proper folding. To maximize recovery of tRNA, we repeated the last step with the saved aliquot and combined elutions.

***In vitro* tRNA cleavage.** Cleavage of tRNAs produced via T7 transcription or purified from total RNA was performed in vitro following the assay described by Winther and Gerdes (16) with slight modifications. In brief, 2 pmol of tRNA was incubated with 2.5, 5 or 10 pmol of VapC-mt4 at 37 C for 3 h in 10 mM HEPES pH 7.5, 15 mM KCl, 3 mM $MgCl_2$, 10% glycerol in a final volume of 10 ml. To check for cleavage inhibition by the cognate antitoxin, VapB-mt4 was preincubated with VapC-mt4 for 15 min at 37 C before the RNA substrate was added. In samples containing EDTA, 1 ml of 125 mM EDTA was added to the reaction before the addition of the RNA substrate. The products of these reactions were separated on a 9% polyacrylamide, 7 M urea gel and visualized by SYBR Gold (Invitrogen) staining.

***In vivo* primer extension.** *E. coli* BW25113D6 cells carrying pBAD33-VapC-mt4 were grown in M9 minimal media containing 0.2% glycerol at 37 C to an OD_{600} of 0.3–0.4. Arabinose was then added to a final concentration of 0.2%. Aliquots were removed 0, 20, 40, 60, 80 and 100 min post induction and total RNA was extracted as described in the 'Purification of tRNA from *E. coli* cell extracts' section above. The primer NWO1976 (5'-GGGATCGAACCGCAGACC-3') was 5' end labelled with [γ-^{32}P]-ATP using T4 polynucleotide kinase (NEB). Five mg of total RNA was mixed with 40 pmol of the labelled oligonucleotide in a 6.5-ml reaction containing annealing buffer for the SuperScript III reverse transcriptase (Invitrogen). The samples were heated to 80 C for 3 min and allowed to cool slowly to room temperature over 30 min. First Strand Reaction Mix (3.5 ml;

Invitrogen) and 2 ml SuperScript III Enzyme Mix (Invitrogen) were then added. The reactions were incubated at 52 C for 90 min, then heat inactivated at 70 C for 10 min followed by addition of 12 ml of sequence loading buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol). The DNA-sequencing ladders for primer extension reactions were prepared using the Sequenase Version 2.0 DNA-Sequencing Kit (Agilent) as recommended.

In vitro primer extension. T7-transcribed *M. tuberculosis* tRNAs were cleaved as described in the 'In vitro tRNA cleavage' section, purified by phenol:chloroform extraction and ethanol precipitated. Primer extension reactions were then performed as described in the previous section using 2 pmol of cleaved tRNA. Primers: $tRNA^{Ala2}$, NWO2164 (5'-GGAGCCTAGGGGACTCGAA-3'); $tRNA^{Ser26}$, NWO2165 (5'-GAGGCGAGAGGATTTGAACCTCC-3'); and $tRNA^{Ser24}$, NWO2166 (5'-GGAGGATGCGGGATTTGAACCC-3').

RNA-seq of *E. coli* RNA. *E. coli* BW25113D6 cells carrying pBAD33-VapC-mt4 were grown in M9 minimal media containing 0.2% glycerol at 37 C to OD_{600} 0.3–0.4. The culture was split in half and arabinose was then added to a final concentration of 0.2% to one of the cultures. Aliquots from induced and uninduced cultures were removed 0, 20, 40, 60, 80 and 100 min post induction. Total RNA was extracted as described in the 'Purification of tRNA from *E. coli* cell extracts'; 5' RNA-seq was performed using total RNA from the 20-min induced and uninduced cultures. The 20-min induction time was selected because it coincided with the initiation of toxicity, ensuring that the cleavage site being documented was the direct consequence of toxin induction and not due to secondary effects. Two pools of RNA, those possessing a 5'-P and those possessing a 5'-OH. After removal of 5'-P RNA from the 5'-OH pool, the resulting 5'-OH RNA was converted to 5'-P RNA via

phosphorylation by 50 U OptiKinase. A 5' SOLiD RNA adaptor was then ligated onto RNA from both pools. Reverse transcription and PCR was then performed based on the methods outlined in Schifano *et al*. (21) and Vvedenskaya *et al*. (55).

tRNA northern analysis. Total RNA was isolated from *E. coli* as described above. Total RNA from *M. tuberculosis* H37Rv was isolated as previously described[57]. In brief, *M. tuberculosis* cells were grown to exponential phase and then pelleted. Cell pellets were resuspended in TRIzol Reagent (Invitrogen). RNA was treated with TURBO DNase (Invitrogen). Total RNA (2 mg) from each organism was incubated with 20 pmol of VapC-mt4 in a 10-ml reaction containing 10 mM HEPES pH 7.5, 15 mM KCl, 3 mM $MgCl_2$, 10% glycerol at 37 C for 0, 3 or 24 h. To check for cleavage inhibition by the cognate antitoxin, 120 pmol VapB-mt4 was preincubated with VapC-mt4 for 15 min at 37 C before the RNA substrate was added and incubated at 37 C for 24 h. The products of these reactions were separated on a 9% polyacrylamide, 7 M urea gel and visualized by SYBR Gold (Invitrogen) staining. RNA was transferred to nitrocellulose and hybridized with oligonucleotides specific for a given tRNA species. The oligonucleotides used were: *E. coli* $tRNA^{Ala2}$, NWO2268, 5'-CGTGCAAAGC AGGCGCTC-3'; *M. tuberculosis* $tRNA^{Ser26}$, NWO2277, 5'-GCAGTGAGCCCCATTCG-3'; and *M. tuberculosis* $tRNA^{Ser24}$, NWO2270, 5'-CCCTTGAAGGGGGACAAC TCATTA-3'.

VapC-mt4-tRNA molecular modelling. All protein–RNA docking simulations were performed with the programme 3dRPC (26). The homology model of VapC-mt4 was generated by the HHpred homology model workflow and Modeller using the optimal multiple template function (27, 28). The structural templates used in Modeller to generate the VapC-mt4 homology model were Rickettsia felis VapC (from VapBC2 complex; PDB accession code 3ZVK, chain A(30); 19% identity, 28.7% similarity to VapC-mt4), *M.*

tuberculosis VapC (from VapBC5 complex; PDB accession code 3DBO, chain B (29); 36% identity, 53.7% similarity) and Neisseria gonorrhoeae FitB (from FitAB complex; PDB accession code 2H1C, chain A (31); 27% identity, 28.8% similarity). The VapC-mt4 homology dimer models were first generated by the global search programme GRAMM-X (27, 34), followed by further refinement using the programme RosettaDock (35). The manganese atom was positioned in the VapC-mt4 homology model using an alignment of Pyrobaculum aerophilum VapC (PDB accession code 2FE1) [(32)] generated by DaliLite[(38)]. The homologue of $tRNA^{Ala2}$ used in this study was Saccharomyces cerevisiae $tRNA^{Phe}$ (PDB accession code 4TNA, chain A)(37) and the homologue for the $tRNA^{Ser26}$ was Escherichia coli $tRNA^{Leu}$ (PDB accession code 3UZ3, chain B)(36). The expanded variable loop of *M. tuberculosis* $tRNA^{Ser26}$ was generated by the programme 3dRNA using the secondary structure predicted by NAVIEW (56, 57) and extrapolated onto the $tRNA^{Leu}$ variable loop by least-squares fitting using Coot (58). With both tRNA models, the changes in the target anticodon loop residues corresponding to $tRNA^{Ala2}$ or $tRNA^{Ser26}$ sequences were manually introduced using Coot. The proposed models were selected based on the proximity between the targeted region of the tRNA ASL and the proposed VapC catalytic residues.

ACKNOWLEDGEMENTS

This work was supported in part by the National Institutes of Health under award numbers R21 AI072399 (to N.A.W. and R.N.H.) and R01 GM095693 (to N.A.W.), R01 GM088343 (to B.E.N.), R21 AI097881 (to R.N.H.) and T32 AI007403 (to J.W.C.); a National Science Foundation CAREER award MCB 0953714 (to C.M.D.). C.M.D is a Pew Scholar in the Biomedical Sciences.

FIGURES

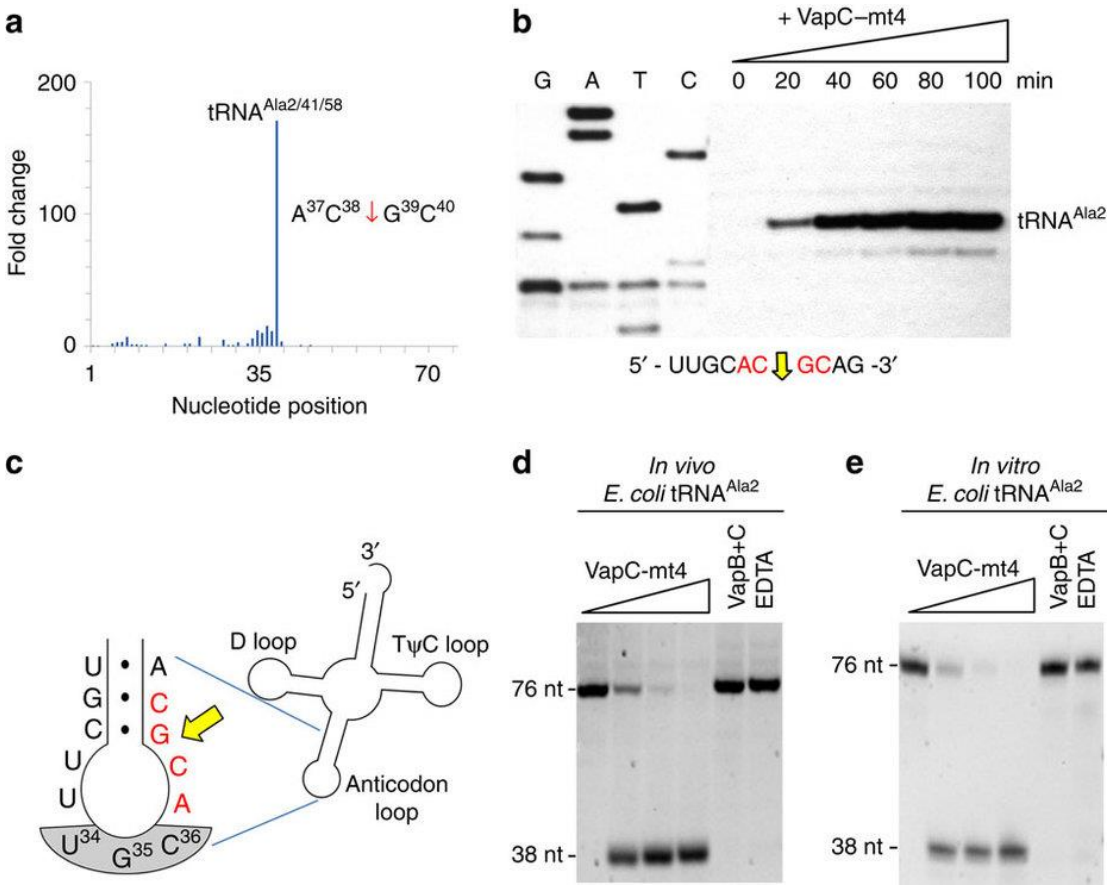

Figure 2.1 VapC-mt4 Cleaves tRNA^{Ala2} in *E. coli*. (a) Histogram representing the fold change in 5'-OH ends observed along the length of the tRNA^{Ala2} upon comparison of cells with VapC-mt4 to those without VapC-mt4. The sequence surrounding the site of cleavage in tRNA^{Ala2} is shown on the right; red arrow denotes position of cleavage. (b) Primer extension analysis with total *E. coli* RNA following the induction of VapC-mt4 for the times indicated. The RNA sequence (ACGC consensus in red) and positions of cleavage (yellow arrow) shown below the gel image. G, A, T and C lanes correspond to DNA-sequencing ladders using the same primer and a tRNA^{Ala2} DNA template. The major primer extension band migrates between the G and T residues in the sequencing ladder instead of aligning exactly to the G residue. We attribute this to the repeatable aberrant migration of the sequencing ladder below this tRNA sequence. (c) Illustration of the RNA-seq cleavage

site in the $tRNA^{Ala2}$ ASL, yellow arrow; ACGC consensus sequence in red, anticodon shaded in grey. (d, e) Cleavage assay with in vivo purified $tRNA^{Ala2}$ (d) or in vitro-synthesized $tRNA^{Ala2}$ (e) and increasing amounts of VapC-mt4 (ratios of toxin to RNA were 0:1, 1.25:1, 2.5:1 and 5:1). Control reactions on the right contained the highest concentration of VapC-mt4 preincubated with VapB antitoxin or EDTA before addition of the respective tRNAs. Reactions were incubated at 37 C for 3 h. Sizes of full-length and cleaved tRNA products on the left. Complete gel images for (b, d, e) are shown in Fig. 2.11.

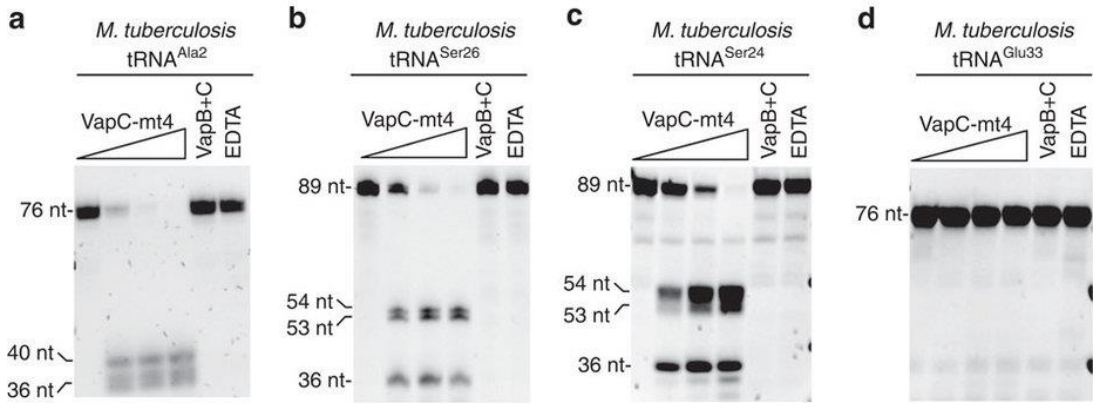

Figure 2.2 VapC-mt4 targets a specific subset of tRNAs in *M. tuberculosis*. (a–c) In vitro VapC-mt4 cleavage assays showing the three tRNAs from among the 13 consensus-containing *M. tuberculosis* tRNAs tested that are cleaved to completion: $tRNA^{Ala2}$ (a), $tRNA^{Ser26}$ (b) and $tRNA^{Ser24}$ (c). (d) $tRNA^{Glu33}$ is shown as a representative example of a tRNA not cut by VapC-mt4 even though it contains the cleavage consensus sequence. In vitro-synthesized tRNAs were incubated with increasing amounts of VapC-mt4 (ratios of toxin to RNA were 0:1, 1.25:1, 2.5:1, and 5:1). Control reactions on the right contained the highest concentration of VapC-mt4 preincubated with VapB antitoxin or EDTA before addition of the respective tRNAs. Reactions were incubated at 37 C for 3 h. Sizes of full-length and cleaved tRNA products are indicated on the left. Note that in some cases extra bands are visible for the cleavage products because the T7 RNA polymerase used to synthesize the tRNAs frequently leads to 3^0 end heterogeneity (usually ±1 nt)[61,62]. Complete gel images for a–c are shown in Fig. 2.11. Complete gel images for d along with full gels images of all 10 tRNAs that were weakly cut or not cut at all are shown in Fig. 2.12.

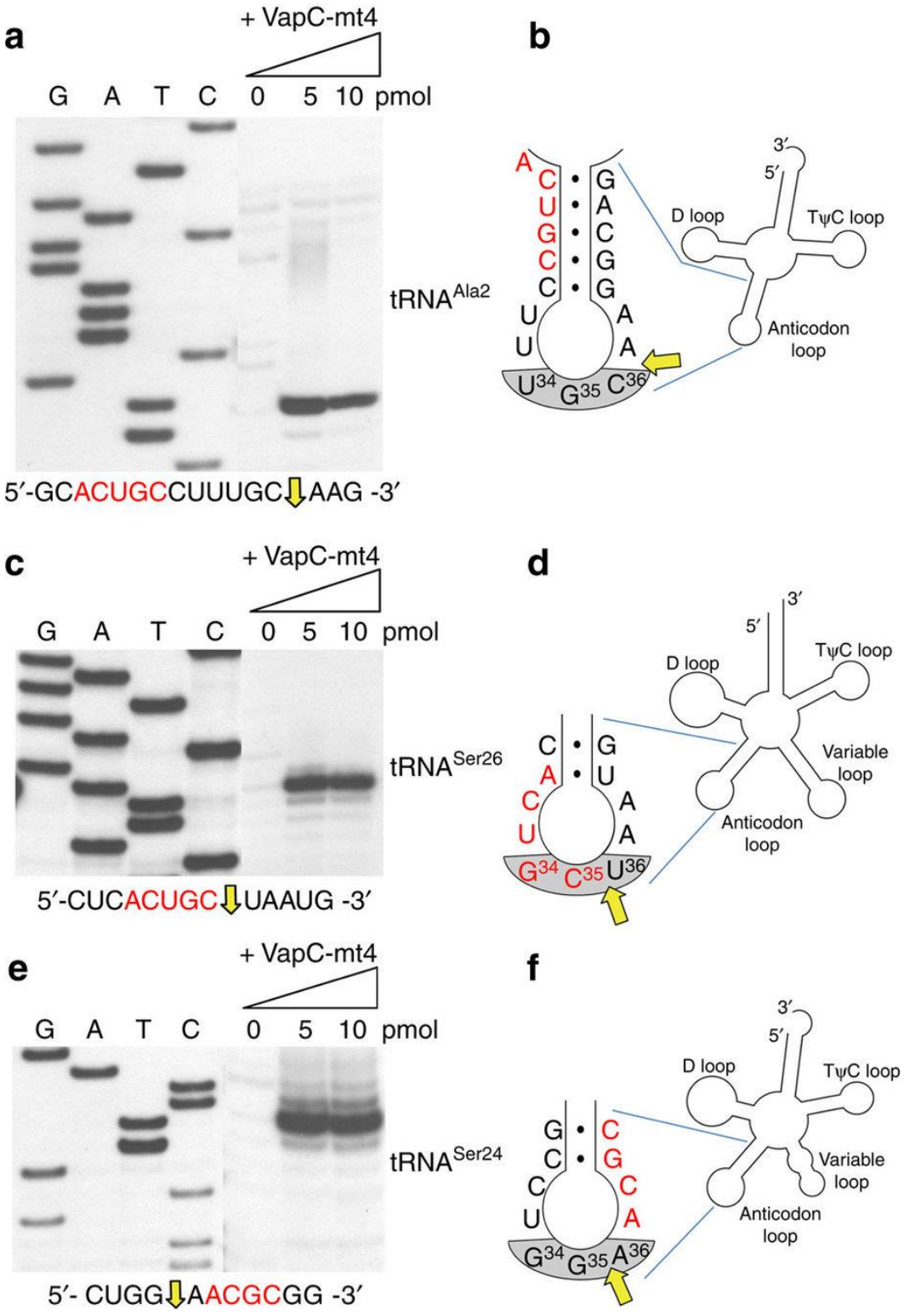
a
+ VapC-mt4
G A T C 0 5 10 pmol
tRNA^Ala2
5′-GCACUGCCUUUGC⇩AAG -3′
b
D loop
TψC loop
Anticodon loop
U34 G35 C36
c
+ VapC-mt4
G A T C 0 5 10 pmol
tRNA^Ser26
5′-CUCACUGC⇩UAAUG -3′
d
D loop
TψC loop
Variable loop
Anticodon loop
G34 C35 U36
e
+ VapC-mt4
G A T C 0 5 10 pmol
tRNA^Ser24
5′- CUGG⇩AACGCGG -3′
f
D loop
TψC loop
Variable loop
Anticodon loop
G34 G35 A36

Figure 2.3 VapC-mt4 cleaves three *M. tuberculosis* tRNAs near consensus sequences. (a–f) Primer extension analysis with in vitro-synthesized *M. tuberculosis* $tRNA^{Ala2}$ (a,b), $tRNA^{Ser26}$ (c,d) and $tRNA^{Ser24}$ (e,f). The RNA sequence (ACGC or ACUGC consensus in red) and positions of cleavage (yellow arrow) shown below gel images in left panels. G, A, T and C lanes correspond to DNA-sequencing ladders using the same primer and matched tRNA DNA template as the corresponding primer extension. To the right of each primer extension are illustrations highlighting the cleavage site (yellow arrow) relative to the consensus sequence (red) and anticodon (grey shaded). tRNAs were incubated with increasing amounts of VapC-mt4 (ratios of toxin to RNA were 0:1, 2.5:1 and 5:1) for 3 h at 37 C. Convention when numbering tRNA bases dictates that the anticodon is designated as numbered positions 34–36. This convention is followed in b,d,f. Note, however, that for $tRNA^{Ser26}$ (d) and $tRNA^{Ser24}$ (f) the actual base numbers for the anticodon are 35–37. Complete gel images are shown in Fig. 2.13.

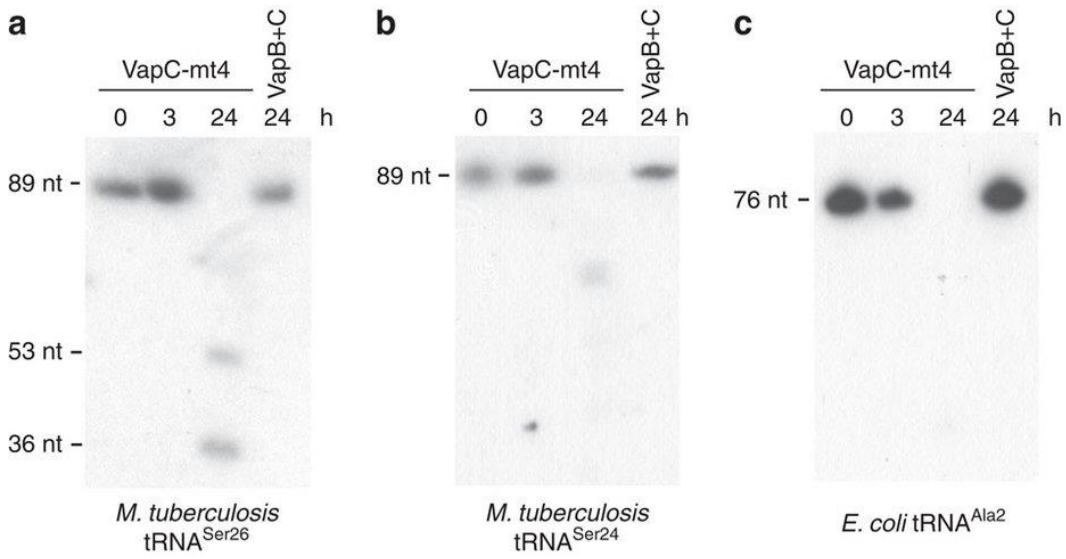

Figure 2.4 VapC-mt4 also cleaves modified *M. tuberculosis* tRNAs. (a,b) Total RNA (2 mg) was incubated with VapC-mt4 (20 pmol) for the times indicated, and northern analysis was performed using isoacceptor-specific oligonucleotides complementary to each ASL. VapB þ C samples denote preincubation of the VapC-mt4 toxin (20 pmol) with VapB-mt4 antitoxin (120 pmol) before addition of total RNA. (c) Analogous experiment for *E. coli* $tRNA^{Ala2}$ using *E. coli* total RNA and an isoacceptor-specific oligonucleotide complementary to the ASL. Complete gel images are shown in Fig. 2.14.

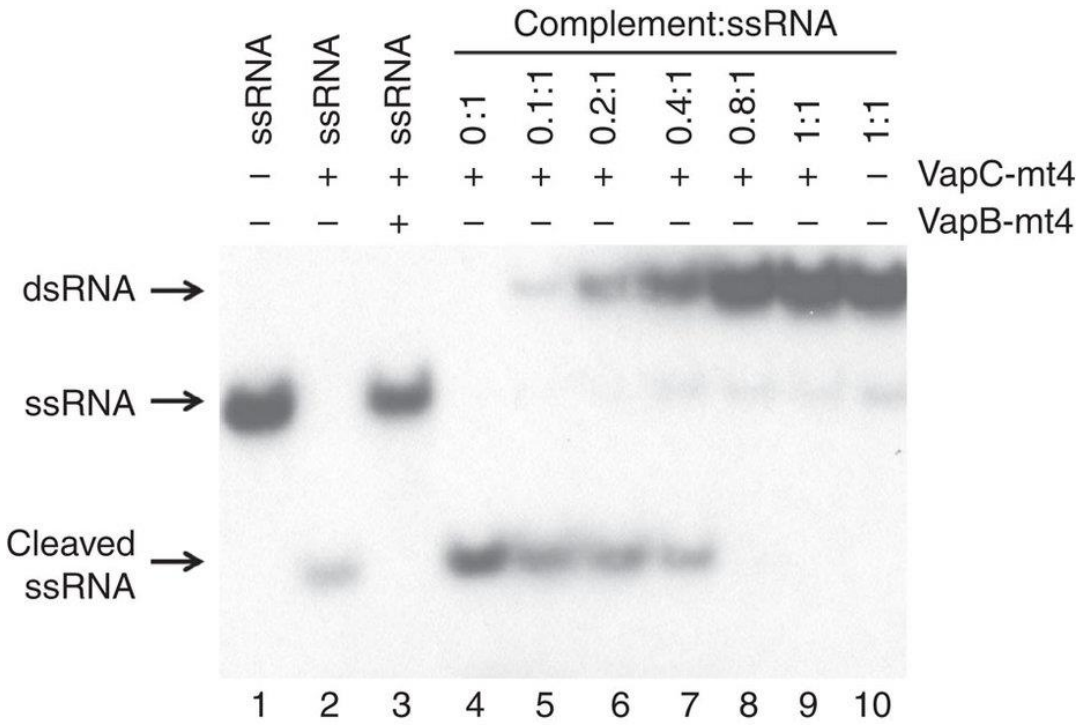

Figure 2.5 VapC-mt4 cleavage requires a single-stranded RNA template. A 20-nt 5^0 end-labelled RNA containing an ACGC consensus sequence alone (lane 1), after incubation with VapC-mt4 (lane 2), or after incubation; with VapB-mt4 and VapC-mt4 (lane 3). This 20-nt RNA was also preincubated with increasing amounts of an RNA complement lacking an ACGC consensus (lanes 4–10) followed by incubation with VapC-mt4. The positions of the dsRNA, ssRNA and cleaved ssRNA are shown on the left. The ratio of toxin to RNA was 64.2:1 and the assay was incubated at 37 C for 3 h. A complete gel image is shown in Fig. 2.15.

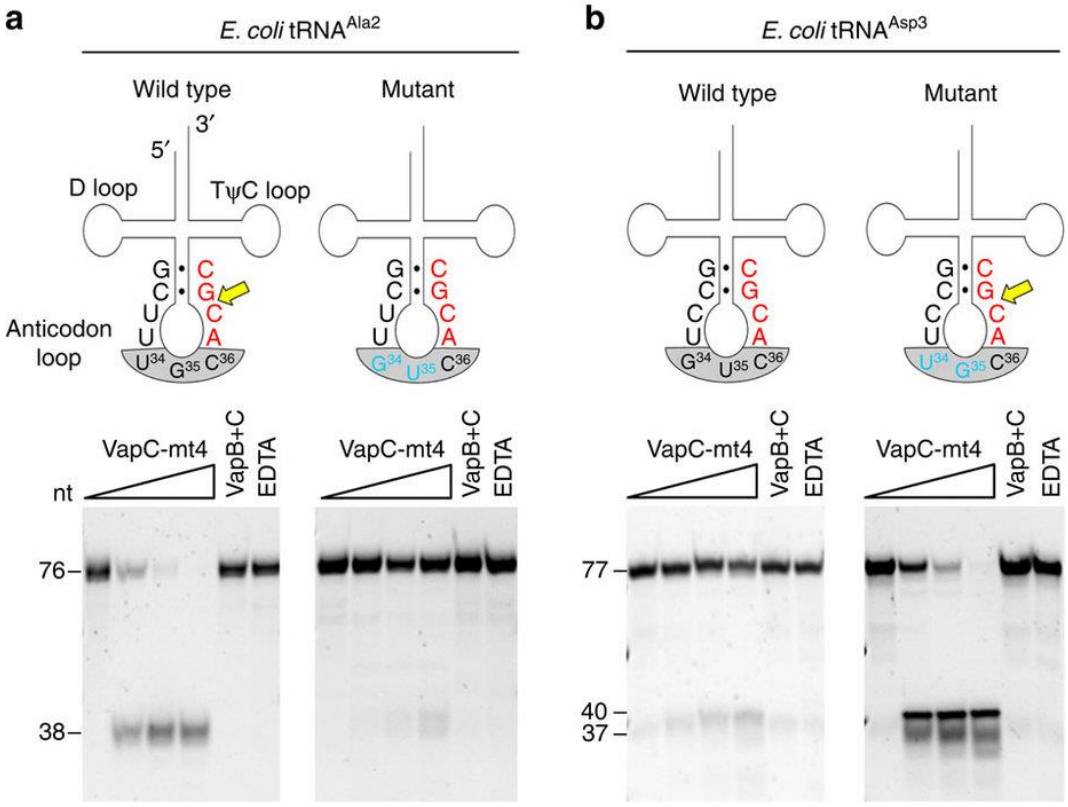
a
E. coli tRNA^Ala2
Wild type
Mutant
3′
5′
D loop
TψC loop
Anticodon loop
VapC-mt4
VapB+C
EDTA
nt
76
38
b
E. coli tRNA^Asp3
77
40
37

Figure 2.6 VapC-mt4 cleavage requires an ACGC in the proper context. (a) Left panel, complete cleavage of wild-type $tRNA^{Ala2}$ by VapC-mt4; right panel, no cleavage after mutation of two bases (blue) in the $tRNA^{Ala2}$ anticodon to match those in $tRNA^{Asp3}$. (b) Left panel, no cleavage of wild-type $tRNA^{Asp3}$ by VapC-mt4; right panel, complete cleavage after mutation of two bases (blue) in the $tRNA^{Asp3}$ anticodon to match those in $tRNA^{Ala2}$. Cleavage site (yellow arrow), consensus sequence (red), mutated bases (blue), anticodon (grey shaded), base pairing represented by black dots (). In vitro-synthesized tRNAs were incubated with increasing amounts of VapC-mt4 (ratios of toxin to RNA were 0:1, 1.25:1, 2.5:1 and 5:1). Control reactions on the right contained the highest concentration of VapC-mt4 preincubated with VapB antitoxin or EDTA before addition of the respective tRNAs. Reactions were incubated at 37 C for 3 h. Sizes of full-length and cleaved tRNA products on the left. The sizes of the $tRNA^{Asp3}$ mutant cleavage products are predicted based on the site of cleavage in $tRNA^{Ala2}$. Convention when numbering tRNA bases dictates that the anticodon bases are numbered 34–36. This convention is followed in a,b. Note, however, that in the case of $tRNA^{Asp3}$ (b) the actual base numbers for the anticodon are 35–37. Note that in some cases extra bands are visible for the cleavage products because the T7 RNA polymerase used to synthesize the tRNAs frequently leads to 3^0 end heterogeneity (usually ±1 nt)[61,62]. Complete gel images are shown in **Fig. 2.16**.

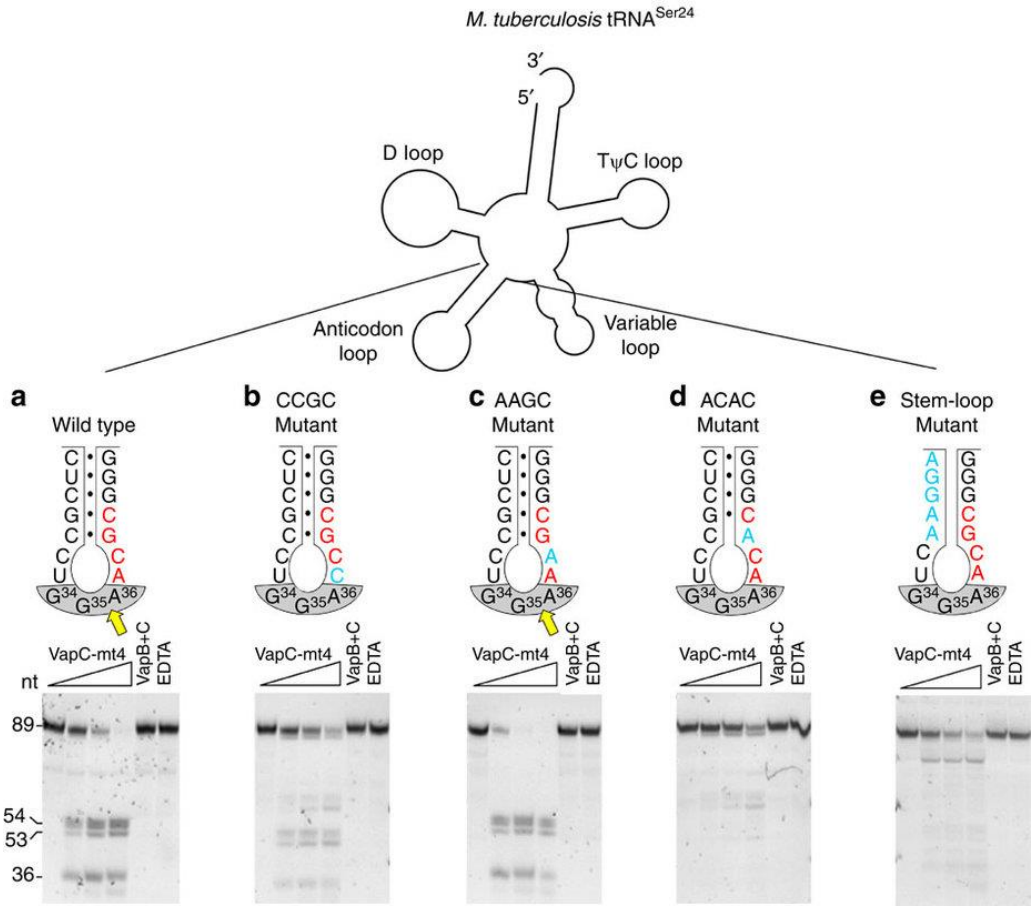
M. tuberculosis tRNASer24
3′
5′
D loop
TψC loop
Anticodon loop
Variable loop
a
Wild type
b
CCGC Mutant
c
AAGC Mutant
d
ACAC Mutant
e
Stem-loop Mutant
VapC-mt4
VapB+C
EDTA
nt
89
54
53
36

Figure 2.7 tRNA cleavage by VapC-mt4 requires both structure and sequence determinants. (a–e) Secondary structure outline of *M. tuberculosis* $tRNA^{Ser24}$ (top) and cleavage assays comparing wild-type VapC-mt4 cleavage (a) with mutants that alter the consensus sequence but not ASL base pairing (b,c), a mutant that alters the consensus and disrupts one base pair in the stem (d), a mutant that retains the consensus sequence but removes all base pairing in the stem (e). Mutated bases (blue), cleavage site (yellow arrow), consensus sequence (red), anticodon (grey shaded), base pairing represented as black dots (). In vitro-synthesized tRNAs were incubated with increasing amounts of VapC-mt4 (ratios of toxin to RNA were 0:1, 1.25:1, 2.5:1 and 5:1). Control reactions on the right contained the highest concentration of VapC-mt4 preincubated with VapB antitoxin or EDTA before addition of the respective tRNAs. Reactions were incubated at 37 C for 3 h. Sizes of full-length and cleaved tRNA products on the left. Convention when numbering tRNA bases dictates that the anticodon bases are numbered 34–36. This convention is followed in a–e. Note however that in the case of $tRNA^{Ser24}$ (a–e) the actual base numbers for the anticodon are 35–37. Complete gel images are shown in Fig. 2.17.

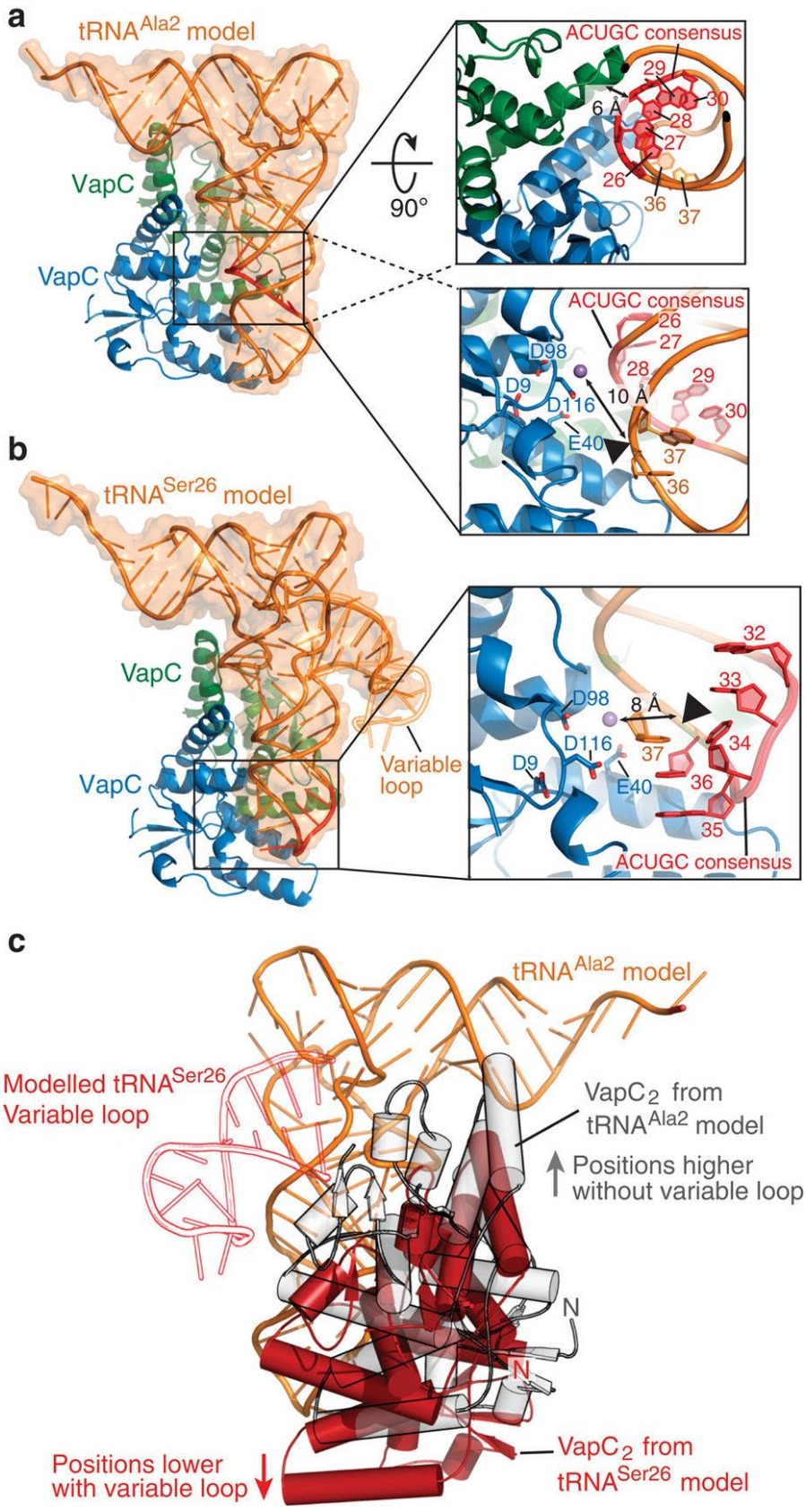
a
tRNA^Ala2 model
VapC
VapC
90°
ACUGC consensus
29
30
6 Å
28
27
26
36
37
ACUGC consensus
26
27
D98
28
29
D9
10 Å
D116
30
E40
37
36
b
tRNA^Ser26 model
VapC
VapC
Variable loop
32
33
D98
8 Å
D116
37
34
D9
E40
36
35
ACUGC consensus
c
tRNA^Ala2 model
Modelled tRNA^Ser26 Variable loop
VapC2 from tRNA^Ala2 model
Positions higher without variable loop
N
N
Positions lower with variable loop
VapC2 from tRNA^Ser26 model

Figure 2.8 VapC-mt4-tRNA docking simulations. (a) Predicted binding of a (VapC-mt4)$_2$ (blue and green) to tRNA^{Ala2} (Type I tRNA homologue, tRNA^{Phe} model shown in orange; PDB accession code 4TNA) using a protein–RNA docking simulation programme 3dRPC. The consensus sequence 5'-ACUGC-3' is highlighted in red and boxed. The upper right boxed magnification shows a 90˚ rotated view highlighting the proximity of the tRNA consensus sequence to the VapC monomer not thought to be involved in activity (green). The lower right boxed magnification is a detailed view of the cleavage site. The VapC-mt4 key residues, tRNA consensus sequence and the target nucleotides are shown as sticks, with a black triangle indicating the cleavage site. (b) Predicted binding of a (VapC-mt4)$_2$ (blue and green) to tRNA$^{Ser24/Ser26}$ (Type II tRNA homologue, tRNA^{Leu} model shown in orange; PDB accession code 3UZ3). The expanded variable loop of tRNA^{Ser26} is depicted in orange outline and the consensus sequence 5'-ACUGC-3' is highlighted in red. The boxed magnification is a detailed view of the cleavage site showing the VapC-mt4 key residues, the tRNA^{Leu} consensus sequence and the target nucleotides are shown in sticks, with a black triangle indicating the cleavage site. The extrapolated position of a divalent cation is shown as a purple sphere in both a,b. (c) A comparison of (VapC-mt4)$_2$ binding, highlighting the positional shift in tRNA binding depending on the target tRNA type (grey, tRNA^{Ala2}; red, tRNA$^{Ser24/Ser26}$). The presence of expanded variable loop in tRNA^{Ser26} (shown in red outline) influences the binding of VapC-mt4 (positions lower in anticodon loop) possibly explaining the difference observed in target cleavage sites. Convention when numbering tRNA bases dictates that the anticodon bases are numbered 34–36. This convention is followed in a–c. Note however that in the case of tRNA$^{Ser26/Ser24}$ (b,c) the actual base numbers for the anticodon are 35–37.

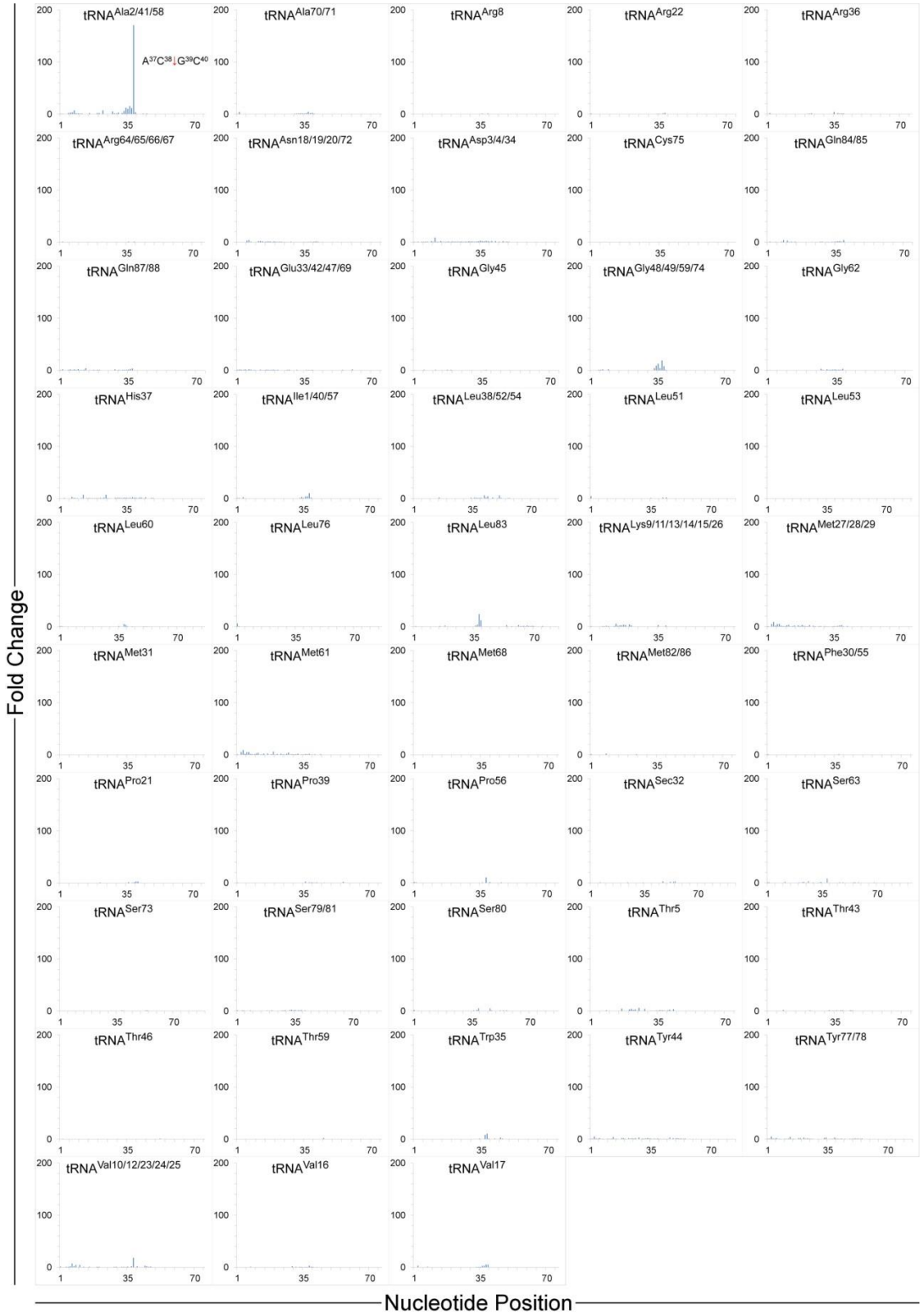
Fold Change
Nucleotide Position
tRNA^Ala2/41/58
A37C38↓G39C40
tRNA^Ala70/71
tRNA^Arg8
tRNA^Arg22
tRNA^Arg36
tRNA^Arg64/65/66/67
tRNA^Asn18/19/20/72
tRNA^Asp3/4/34
tRNA^Cys75
tRNA^Gln84/85
tRNA^Gln87/88
tRNA^Glu33/42/47/69
tRNA^Gly45
tRNA^Gly48/49/59/74
tRNA^Gly62
tRNA^His37
tRNA^Ile1/40/57
tRNA^Leu38/52/54
tRNA^Leu51
tRNA^Leu53
tRNA^Leu60
tRNA^Leu76
tRNA^Leu83
tRNA^Lys9/11/13/14/15/26
tRNA^Met27/28/29
tRNA^Met31
tRNA^Met61
tRNA^Met68
tRNA^Met82/86
tRNA^Phe30/55
tRNA^Pro21
tRNA^Pro39
tRNA^Pro56
tRNA^Sec32
tRNA^Ser63
tRNA^Ser73
tRNA^Ser79/81
tRNA^Ser80
tRNA^Thr5
tRNA^Thr43
tRNA^Thr46
tRNA^Thr59
tRNA^Trp35
tRNA^Tyr44
tRNA^Tyr77/78
tRNA^Val10/12/23/24/25
tRNA^Val16
tRNA^Val17

Figure 2.9 VapC-mt4 cleaves tRNA^{Ala2} in *E. coli*. Histograms representing the fold change in reads observed in tRNAs from the analysis of RNAs carrying a 5'-OH ends isolated from cells induced to express VapC-mt4 versus uninduced cells. The sequence surrounding the site of cleavage in tRNA is shown in the corresponding histogram with the position of cleavage indicated by a red arrow.

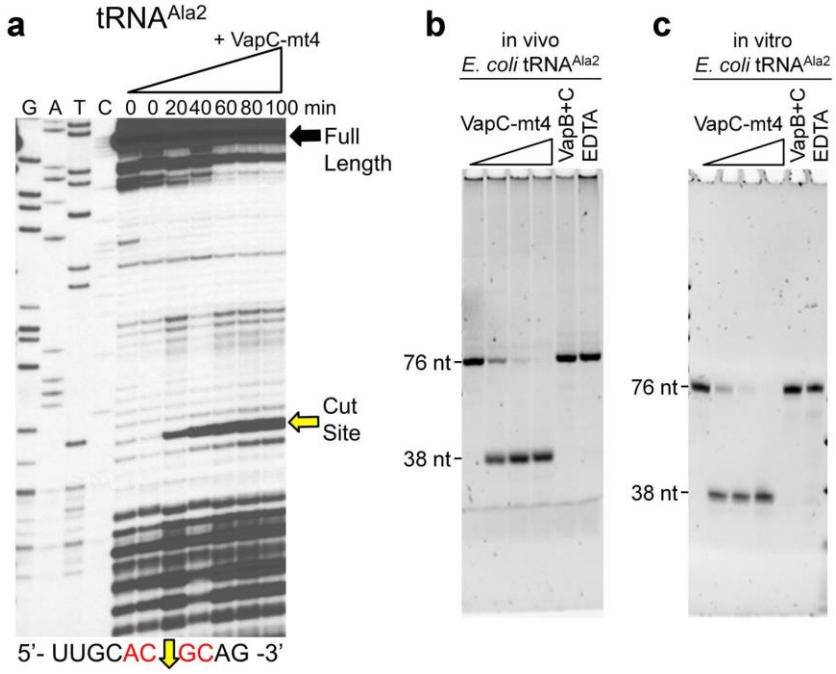

Figure 2.10 VapC-mt4 Cleaves tRNA^{Ala2} in *E. coli* (complete gel images of Fig. 1b, d, e). (**a**) Primer extension analysis with total *E. coli* RNA following induction of VapC-mt4 for the times indicated. The 0 minute time point was loaded twice. The RNA sequence (ACGC consensus in red) and positions of cleavage (yellow arrow) shown below the gel image. G, A, T, and C lanes correspond to DNA sequencing ladders using the same primer and a tRNA^{Ala2} DNA template. The major primer extension band migrates between the G and T residues in the sequencing ladder instead of aligning exactly to the G residue. We attribute this to the repeatable aberrant migration of the sequencing ladder below this tRNA sequence. (**b,c**) Cleavage assay with in vivo purified tRNA^{Ala2} (**b**) or in vitro synthesized tRNA^{Ala2} (**c**) and increasing amounts of VapC-mt4 (ratios of toxin to RNA were 0:1, 1.25:1, 2.5:1, and 5:1). Control reactions on the right contained the highest concentration of VapC-mt4 preincubated with VapB antitoxin or EDTA before addition of the respective tRNAs. Reactions were incubated at 37^{o}C for 3 h. Sizes of full length and cleaved tRNA products on the left.

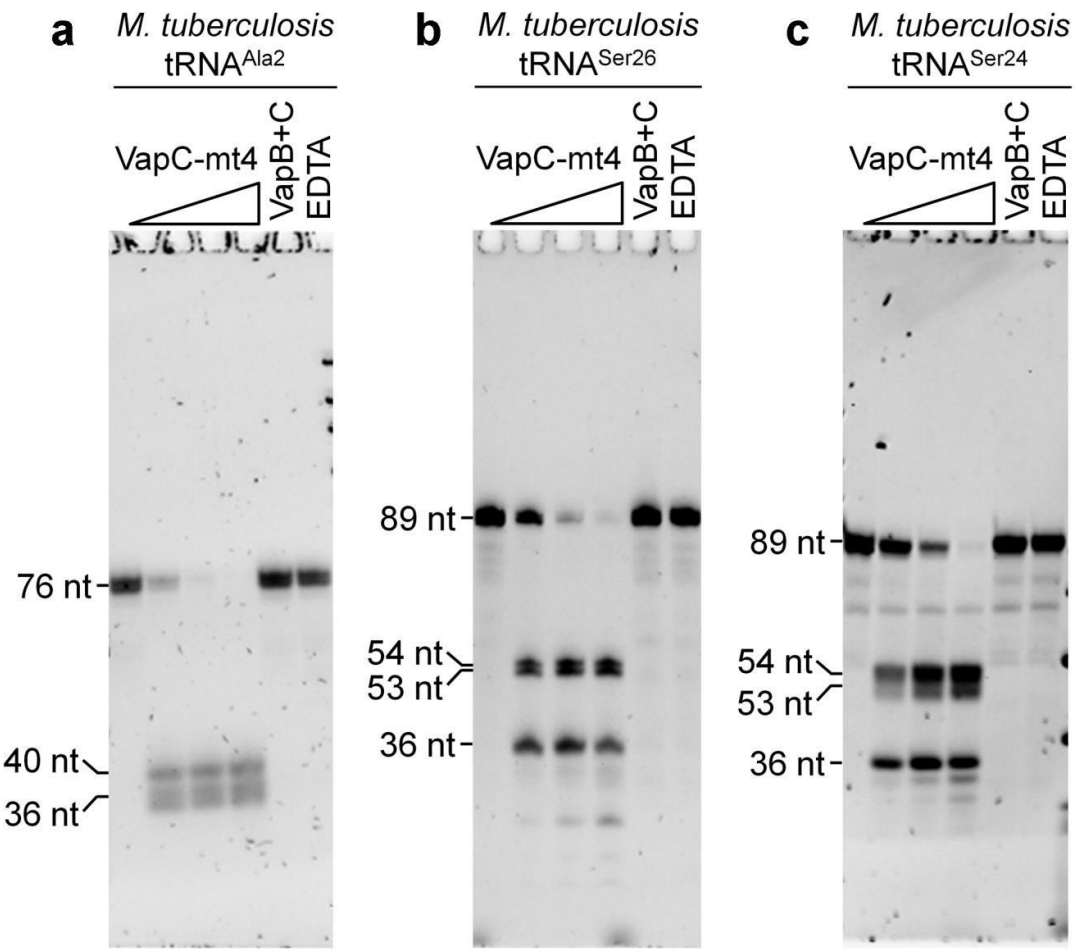

Figure 2.11 VapC-mt4 targets a specific subset of tRNAs in *M. tuberculosis* (complete gel images of Fig. 2 a, b and c). (**a-c**) In vitro VapC-mt4 cleavage assays showing the three tRNAs from among the 13 consensus-containing *M. tuberculosis* tRNAs tested that are cleaved to completion: $tRNA^{Ala2}$ (**a**), $tRNA^{Ser26}$ (**b**), and $tRNA^{Ser24}$ (**c**). In vitro synthesized tRNAs were incubated with increasing amounts of VapC-mt4 (ratios of toxin to RNA were 0:1, 1.25:1, 2.5:1, and 5:1). Control reactions on the right contained the highest concentration of VapC-mt4 preincubated with VapB antitoxin or EDTA before addition of the respective tRNAs. Reactions were incubated at 37°C for 3 h. Sizes of full length and cleaved tRNA products on the left.

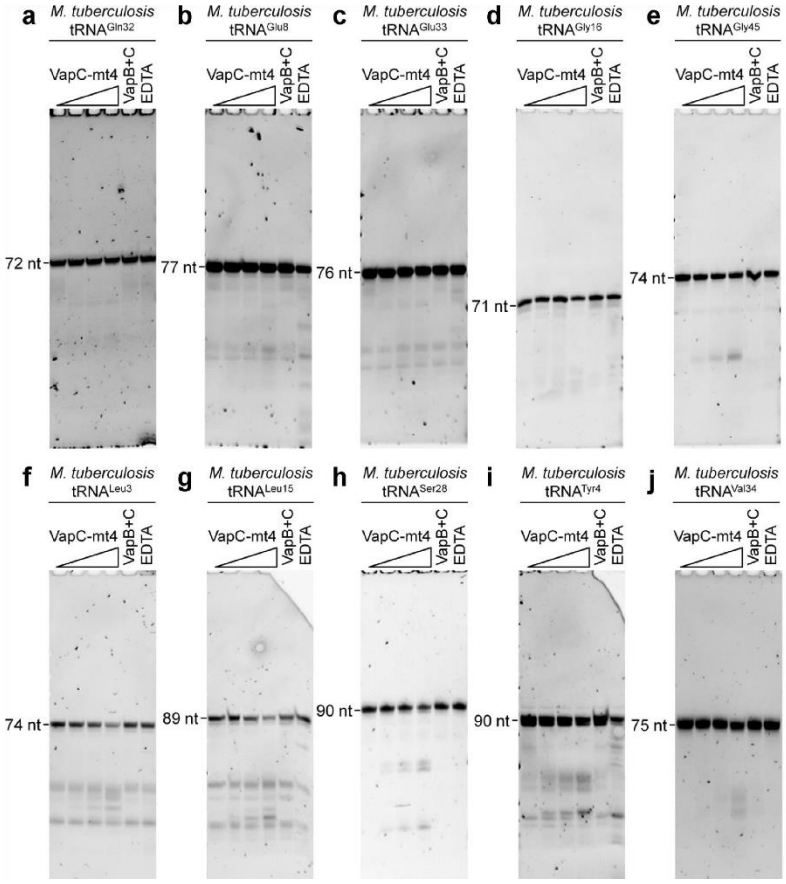

Figure 2.12 VapC-mt4 does not cleave all consensus-containing tRNAs in *M. tuberculosis*. In vitro VapC-mt4 cleavage assays showing the 10 tRNAs from among the 13 consensus-containing *M. tuberculosis* tRNAs tested that are not cleaved: $tRNA^{Gln32}$ (**a**), $tRNA^{Glu8}$ (**b**), $tRNA^{Glu33}$ (**c**), $tRNA^{Gly16}$ (**d**), $tRNA^{Gly45}$ (**e**), $tRNA^{Leu3}$ (**f**), $tRNA^{Leu15}$ (**g**), $tRNA^{Ser28}$ (**h**), $tRNA^{Tyr4}$ (**i**), and $tRNA^{Val34}$ (**j**). In vitro synthesized tRNAs were incubated with increasing amounts of VapC-mt4 (ratios of toxin to RNA were 0:1, 1.25:1, 2.5:1, and 5:1). Control reactions on the right contained the highest concentration of VapC-mt4 preincubated with VapB antitoxin or EDTA before addition of the respective tRNAs. Reactions were incubated at 37°C for 3 h. Sizes of full length and cleaved tRNA products on the left.

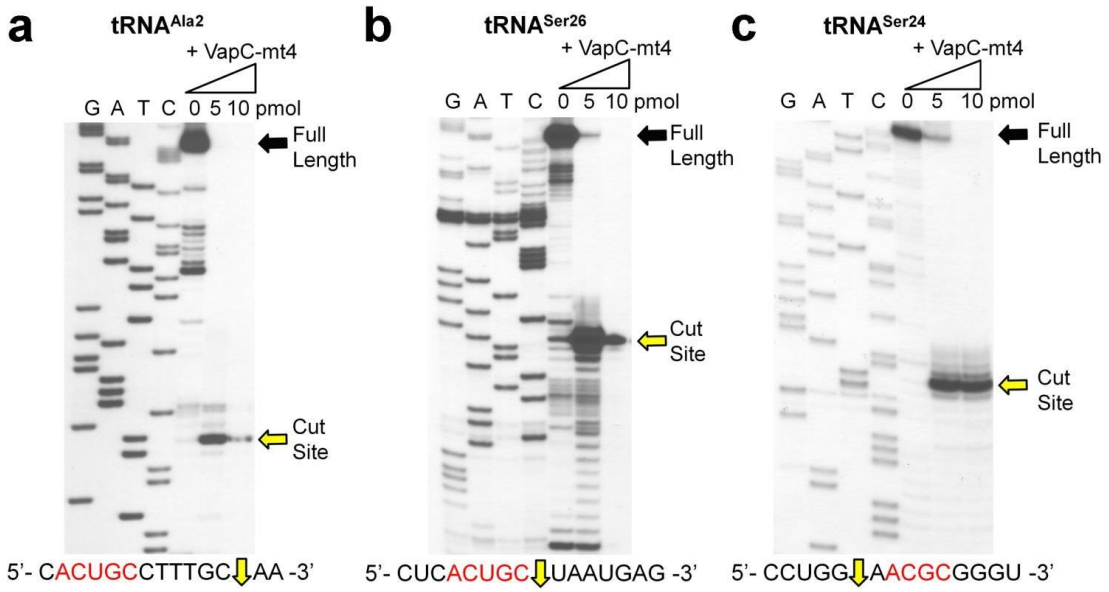

Figure 2.13 VapC-mt4 cleaves three *M. tuberculosis* tRNAs near consensus sequences (complete gel images of Fig. 2.3a, c, e). (**a-c**) Primer extension analysis with in vitro synthesized *M. tuberculosis* $tRNA^{Ala2}$ (**a**), $tRNA^{Ser26}$ (**b**), and $tRNA^{Ser24}$ (**c**). Full gels are shown for each tRNA. The RNA sequence (ACGC or ACUGC consensus in red) and positions of cleavage (yellow arrow) shown below gel images in left panels. G, A, T, and C lanes correspond to DNA sequencing ladders using the same primer and matched tRNA DNA template as the corresponding primer extension. tRNAs were incubated with increasing amounts of VapC-mt4 (ratios of toxin to RNA were 0:1, 2.5:1, and 5:1) for 3h at 37°C.

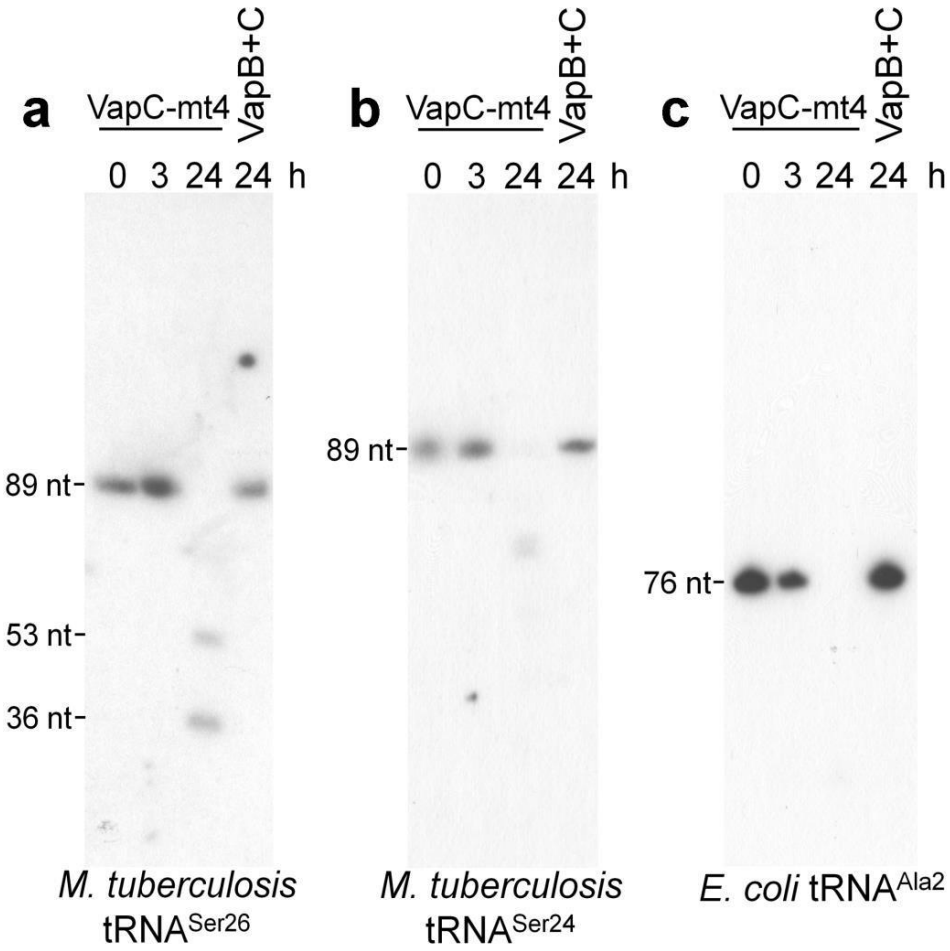

Figure 2.14 VapC-mt4 also cleaves modified *M. tuberculosis* tRNAs (complete blot images of Fig. 2.4). (a, b) *M. tuberculosis* total RNA (2 µg) was incubated with VapC-mt4 (20 pmol) for the times indicated and northern analysis was performed using isoacceptor-specific oligonucleotides complementary to each ASL. VapB+C samples denote preincubation of the VapC-mt4 toxin (20 pmol) with VapB-mt4 antitoxin (120 pmol) before addition of total RNA. (**c**) Analogous experiment for *E. coli* $tRNA^{Ala2}$ using *E. coli* total RNA and an isoacceptor-specific oligonucleotide complementary to the ASL.

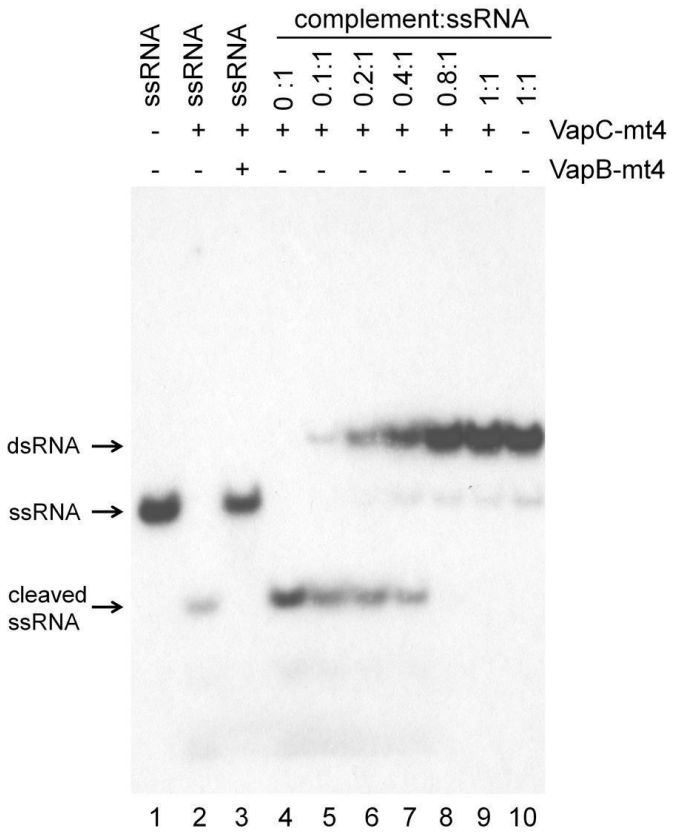

Figure 2.15 VapC-mt4 cleavage requires a single-stranded RNA template (complete gel image of Fig. 2.5). A 20 nt 5'-end labeled RNA containing an ACGC consensus sequence alone (lane 1), after incubation with VapC-mt4 (lane 2), or after incubation with VapB-mt4 and VapC-mt4 (lane 3). This 20 nt RNA was also preincubated with increasing amounts of an RNA complement lacking an ACGC consensus (lanes 410) followed by incubation with VapC-mt4. The positions of the dsRNA, ssRNA and cleaved ssRNA are shown on the left. The ratio of toxin to RNA was 64.2:1 and the assay was incubated at 37^{o}C for 3 h.

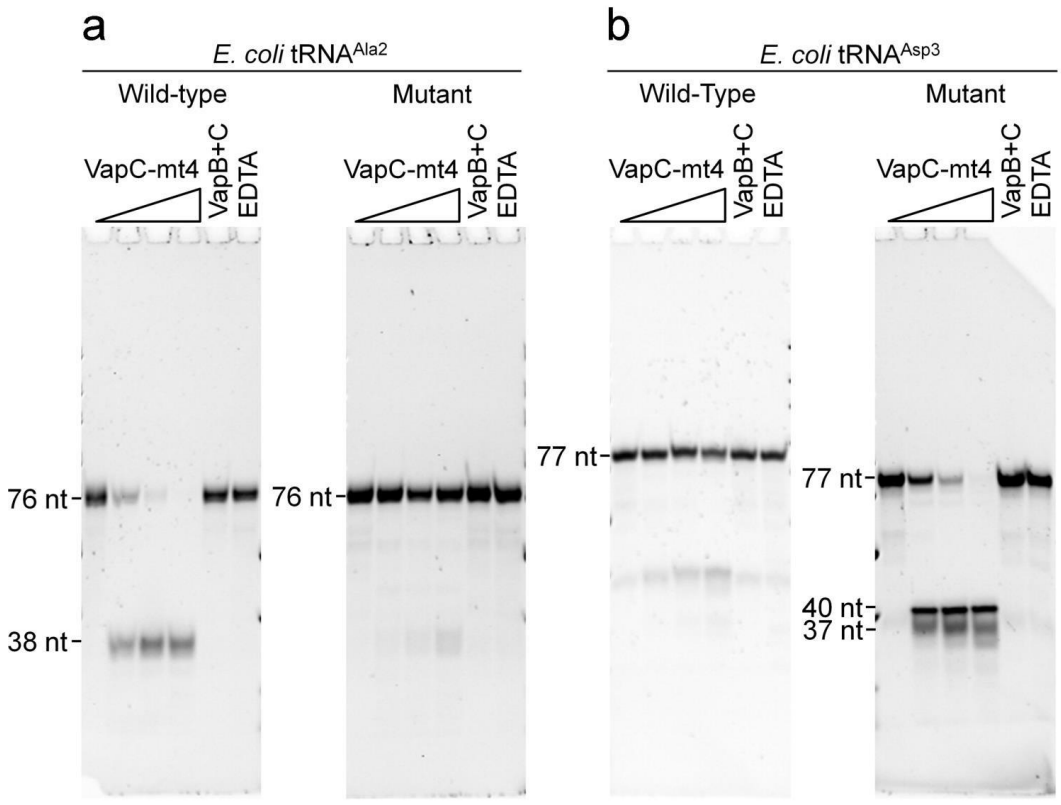

Figure 2.16 VapC-mt4 cleavage requires an ACGC in the proper context (complete gel images of Fig. 2.6a, b). (**a**) *left panel*, complete cleavage of wild-type $tRNA^{Ala2}$ by VapC-mt4; *right panel*, no cleavage after mutation of two bases in the $tRNA^{Ala2}$ anticodon to match those in $tRNA^{Asp3}$ (**b**) *left panel*, no cleavage of wild-type $tRNA^{Asp3}$ by VapC-mt4; *right panel*, complete cleavage after mutation of two bases in the $tRNA^{Asp3}$ anticodon to match those in $tRNA^{Ala2}$. In vitro synthesized tRNAs were incubated with increasing amounts of VapC-mt4 (ratios of toxin to RNA were 0:1, 1.25:1, 2.5:1, and 5:1). Control reactions on the right contained the highest concentration of VapC-mt4 preincubated with VapB antitoxin or EDTA before addition of the respective tRNAs. Reactions were incubated at 37°C for 3 h. Sizes of full length and cleaved tRNA products on the left. The sizes of the $tRNA^{Asp3}$ mutant cleavage products are predicted based on the site of cleavage in $tRNA^{Ala2}$.

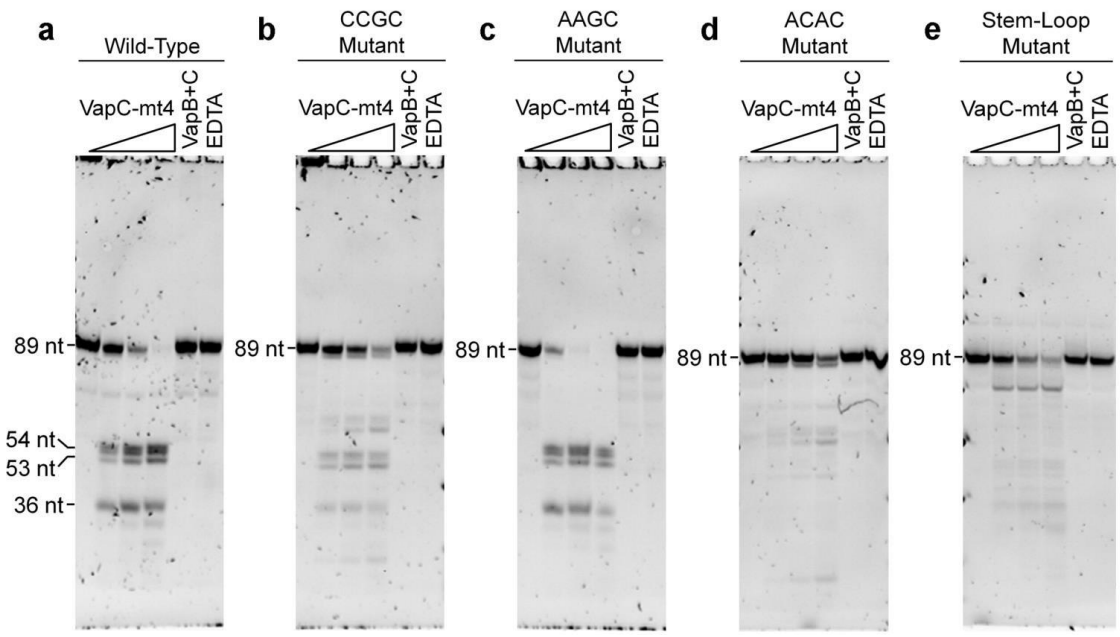

Figure 2.17 tRNA cleavage by VapC-mt4 requires both structure and sequence determinants (complete gel images of Fig. 2.7). (**a-e**) Cleavage assays comparing wild-type VapC-mt4 cleavage (**a**) to mutants that alter the consensus sequence but not ASL base pairing (**b,c**), a mutant that alters the consensus and disrupts one base pair in the stem (**d**), a mutant that retains the consensus sequence but removes all base pairing in the stem (**e**). In vitro synthesized tRNAs were incubated with increasing amounts of VapC-mt4 (ratios of toxin to RNA were 0:1, 1.25:1, 2.5:1, and 5:1). Control reactions on the right contained the highest concentration of VapC-mt4 preincubated with VapB antitoxin or EDTA before addition of the respective tRNAs. Reactions were incubated at 37°C for 3 h. Sizes of full length and cleaved tRNA products on the left.

REFERENCES

1. E. Aizenman, H. Engelberg-Kulka, G. Glaser, An Escherichia coli chromosomal "addiction module" regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death. *Proc Natl Acad Sci U S A* **93**, 6059-6063 (1996).
2. R. Hazan, B. Sat, H. Engelberg-Kulka, Escherichia coli mazEF-mediated cell death is triggered by various stressful conditions. *J Bacteriol* **186**, 3663-3669 (2004).
3. Y. Yamaguchi, M. Inouye, Regulation of growth and death in Escherichia coli by toxin-antitoxin systems. *Nat Rev Microbiol* **9**, 779-790 (2011).
4. G. R. Stewart *et al.*, Dissection of the heat-shock response in Mycobacterium tuberculosis using mutants and microarrays a. *Microbiology* **148**, 3129-3138 (2002).
5. H. R. Ramage, L. E. Connolly, J. S. Cox, Comprehensive functional analysis of Mycobacterium tuberculosis toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution. *PLoS Genet* **5**, e1000767 (2009).
6. T. R. Rustad, M. I. Harrell, R. Liao, D. R. Sherman, The enduring hypoxic response of Mycobacterium tuberculosis. *PLoS One* **3**, e1502 (2008).
7. L. Rand *et al.*, The majority of inducible DNA repair genes in Mycobacterium tuberculosis are induced independently of RecA. *Mol Microbiol* **50**, 1031-1042 (2003).
8. J. C. Betts, P. T. Lukey, L. C. Robb, R. A. McAdam, K. Duncan, Evaluation of a nutrient starvation model ofMycobacterium tuberculosispersistence by gene and protein expression profiling. *Molecular Microbiology* **43**, 717-731 (2002).
9. G. Cappelli *et al.*, Profiling of Mycobacterium tuberculosis gene expression during human macrophage infection: upregulation of the alternative sigma factor G, a group of transcriptional regulators, and proteins with unknown function. *Res Microbiol* **157**, 445-455 (2006).
10. S. B. Korch, H. Contreras, J. E. Clark-Curtiss, Three Mycobacterium tuberculosis Rel toxin-antitoxin modules inhibit mycobacterial growth and are expressed in infected human macrophages. *J Bacteriol* **191**, 1618-1630 (2009).
11. R. Provvedi, F. Boldrin, F. Falciani, G. Palu, R. Manganelli, Global transcriptional response to vancomycin in Mycobacterium tuberculosis. *Microbiology* **155**, 1093-1102 (2009).
12. R. Singh, C. E. Barry, H. I. M. Boshoff, The Three RelE Homologs of Mycobacterium tuberculosis Have Individual, Drug-Specific Effects on Bacterial Antibiotic Tolerance. *Journal of Bacteriology* **192**, 1279-1291 (2010).
13. T. Cortes *et al.*, Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in Mycobacterium tuberculosis. *Cell Rep* **5**, 1121-1131 (2013).
14. B. A. Ahidjo *et al.*, VapC toxins from Mycobacterium tuberculosis are ribonucleases that differentially inhibit growth and are neutralized by cognate VapB antitoxins. *PLoS One* **6**, e21738 (2011).
15. V. L. Arcus, P. B. Rainey, S. J. Turner, The PIN-domain toxin-antitoxin array in mycobacteria. *Trends Microbiol* **13**, 360-365 (2005).
16. K. S. Winther, K. Gerdes, Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. *Proc Natl Acad Sci U S A* **108**, 7403-7407 (2011).
17. A. P. Lopes *et al.*, VapC from the leptospiral VapBC toxin-antitoxin module displays ribonuclease activity on the initiator tRNA. *PLoS One* **9**, e101678 (2014).

18. J. L. McKenzie *et al.*, A VapBC Toxin-Antitoxin Module Is a Posttranscriptional Regulator of Metabolic Flux in Mycobacteria. *Journal of Bacteriology* **194**, 2189-2204 (2012).
19. J. D. Sharp *et al.*, Growth and Translation Inhibition through Sequence-specific RNA Binding byMycobacterium tuberculosisVapC Toxin. *Journal of Biological Chemistry* **287**, 12835-12847 (2012).
20. K. S. Winther, D. E. Brodersen, A. K. Brown, K. Gerdes, VapC20 of Mycobacterium tuberculosis cleaves the Sarcin-Ricin loop of 23S rRNA. *Nature communications* **4**, 2796 (2013).
21. J. M. Schifano *et al.*, An RNA-seq method for defining endoribonuclease cleavage specificity identifies dual rRNA substrates for toxin MazF-mt3. *Nature communications* **5**, 3538 (2014).
22. Y. Zhang, J. Zhang, H. Hara, I. Kato, M. Inouye, Insights into the mRNA Cleavage Mechanism by MazF, an mRNA Interferase. *Journal of Biological Chemistry* **280**, 3143-3150 (2004).
23. J. L. McKenzie *et al.*, Determination of ribonuclease sequence-specificity using Pentaprobes and mass spectrometry. *Rna-a Publication of the Rna Society* **18**, 1267-1278 (2012).
24. P. P. Chan, T. M. Lowe, GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* **37**, D93-97 (2009).
25. Y. L. Zhang *et al.*, MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli. *Molecular Cell* **12**, 913-923 (2003).
26. Y. Huang, S. Liu, D. Guo, L. Li, Y. Xiao, A novel protocol for three-dimensional structure prediction of RNA-protein complexes. *Scientific reports* **3**, 1887 (2013).
27. A. Sali, L. Potterton, F. Yuan, H. van Vlijmen, M. Karplus, Evaluation of comparative protein modeling by MODELLER. *Proteins* **23**, 318-326 (1995).
28. J. Soding, A. Biegert, A. N. Lupas, The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**, W244-248 (2005).
29. L. Miallau *et al.*, Structure and proposed activity of a member of the VapBC family of toxin-antitoxin systems. VapBC-5 from Mycobacterium tuberculosis. *J Biol Chem* **284**, 276-283 (2009).
30. M. J. Mate *et al.*, Crystal structure of the DNA-bound VapBC2 antitoxin/toxin pair from Rickettsia felis. *Nucleic Acids Res* **40**, 3245-3258 (2012).
31. K. Mattison, J. S. Wilbur, M. So, R. G. Brennan, Structure of FitAB from Neisseria gonorrhoeae bound to DNA reveals a tetramer of toxin-antitoxin heterodimers containing pin domains and ribbon-helix-helix motifs. *J Biol Chem* **281**, 37942-37951 (2006).
32. R. D. Bunker, J. L. McKenzie, E. N. Baker, V. L. Arcus, Crystal structure of PAE0151 from Pyrobaculum aerophilum, a PIN-domain (VapC) protein from a toxin-antitoxin operon. *Proteins* **72**, 510-518 (2008).
33. K. Xu *et al.*, Protein expression, crystallization and preliminary X-ray crystallographic analysis of the isolatedShigella flexneriVapC toxin. *Acta Crystallographica Section F Structural Biology and Crystallization Communications* **69**, 762-765 (2013).
34. A. Tovchigrechko, I. A. Vakser, GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res* **34**, W310-314 (2006).
35. S. Lyskov, J. J. Gray, The RosettaDock server for local protein-protein docking. *Nucleic Acids Res* **36**, W233-238 (2008).

36. N. Demeshkina, L. Jenner, E. Westhof, M. Yusupov, G. Yusupova, A new understanding of the decoding principle on the ribosome. *Nature* **484**, 256-259 (2012).
37. B. Hingerty, R. S. Brown, A. Jack, Further refinement of the structure of yeast tRNAPhe. *Journal of Molecular Biology* **124**, 523-534 (1978).
38. H. Hasegawa, L. Holm, Advances and pitfalls of protein structural alignment. *Curr Opin Struct Biol* **19**, 341-348 (2009).
39. D. P. Pandey, in *Nucleic Acids Research*. (Oxford University Press (OUP), 2005), vol. 33, pp. 966-976.
40. J. I. Rood, C. L. Wright, V. Haring, M. E. Katz, Molecular analysis of virulence associated gene regions from the ovine footrot pathogen, Dichelobacter nodosus. *Developments in Plant Pathology*, 615-624 (1994).
41. W. H. McClain, J. Schneider, S. Bhattacharya, K. Gabriel, The importance of tRNA backbone-mediated interactions with synthetase for aminoacylation. *Proc Natl Acad Sci U S A* **95**, 460-465 (1998).
42. K. Tomita, T. Ogawa, T. Uozumi, K. Watanabe, H. Masaki, A cytotoxic ribonuclease which specifically cleaves four isoaccepting arginine tRNAs at their anticodon loops. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 8278-8283 (2000).
43. T. Ogawa *et al.*, A cytotoxic ribonuclease targeting specific transfer RNA anticodons. *Science* **283**, 2097-2100 (1999).
44. M. David, G. D. Borasio, G. Kaufmann, T4 bacteriophage-coded polynucleotide kinase and RNA ligase are involved in host tRNA alteration or repair. *Virology* **123**, 480-483 (1982).
45. G. Kaufmann, Anticodon nucleases. *Trends Biochem Sci* **25**, 70-74 (2000).
46. B. Meineke, S. Shuman, Structure-function relations in the NTPase domain of the antiviral tRNA ribotoxin Escherichia coli PrrC. *Virology* **427**, 144-150 (2012).
47. E. Cascales *et al.*, Colicin biology. *Microbiol Mol Biol Rev* **71**, 158-229 (2007).
48. T. Ogawa, S. Inoue, S. Yajima, M. Hidaka, H. Masaki, Sequence-specific recognition of colicin E5, a tRNA-targeting ribonuclease. *Nucleic Acids Res* **34**, 6065-6073 (2006).
49. J. Albrethsen *et al.*, Proteomic profiling of Mycobacterium tuberculosis identifies nutrient-starvation-responsive toxin-antitoxin systems. *Mol Cell Proteomics* **12**, 1180-1191 (2013).
50. J. Gebetsberger, N. Polacek, Slicing tRNAs to boost functional ncRNA diversity. *RNA Biol* **10**, 1798-1806 (2013).
51. E. M. Phizicky, A. K. Hopper, tRNA biology charges to the front. *Genes Dev* **24**, 1832-1860 (2010).
52. D. M. Thompson, R. Parker, Stressing out over tRNA cleavage. *Cell* **138**, 215-219 (2009).
53. M. Sisido, K. Ninomiya, T. Ohtsuki, T. Hohsaka, Four-base codon/anticodon strategy and non-enzymatic aminoacylation for protein engineering with non-natural amino acids. *Methods* **36**, 270-278 (2005).
54. T. Yokogawa, Y. Kitamura, D. Nakamura, S. Ohno, K. Nishikawa, in *Nucleic Acids Research*. (Oxford University Press (OUP), 2009), vol. 38, pp. e89-e89.
55. I. O. Vvedenskaya, S. R. Goldman, B. E. Nickels, Preparation of cDNA libraries for high-throughput RNA sequencing analysis of RNA 5' ends. *Methods Mol Biol* **1276**, 211-228 (2015).
56. R. E. Bruccoleri, G. Heinrich, An improved algorithm for nucleic acid secondary structure display. *Comput Appl Biosci* **4**, 167-173 (1988).

57. Y. Zhao *et al.*, Automated and fast building of three-dimensional RNA structures. *Sci Rep* **2**, 734 (2012).
58. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta crystallographica* **66**, 486-501 (2010).

Chapter 3: The structure and function of *Mycobacterium tuberculosis* MazF-mt6 provides insights into conserved features of MazF endonucleases

Eric D. Hoffer, Stacey J. Miles, and Christine M. Dunham

E. coli MazF cleaves free mRNA after stress is induced while nine *Mycobacterium tuberculosis* MazF family members cleave mRNA, rRNA and tRNAs. *Mtb* MazF-mt6 cleaves 23S rRNA Helix 70 (H70) to inhibit protein synthesis. Here I report the X-ray crystal structure of MazF-mt6 and perform experiments both *in vivo* and *in vitro* to demonstrate that MazF-mt6 residues Arg-13, Asp-10 and Thr-36 are critical for RNase activity. I also compare the structure of MazF-mt6 to previously solved MazF structure to provide further insight into how MazF toxins recognize different RNAs.

This research was originally published in the Journal of Biological Chemistry. Hoffer ED, Miles SJ, Dunham CM. The structure and function of *Mycobacterium tuberculosis* MazF-mt6 provides insights into conserved features of MazF endonucleases. *JBC*. 2017.

Author Contributions: EDH and CMD designed research; SJM purified wild-type MazF-mt6. E.D.H performed all other experiments; EDH and CMD analyzed data; EDH and CMD wrote the paper

ABSTRACT

Toxin-antitoxin systems are ubiquitous in prokaryotic and archaea genomes and regulate growth in response to stress. *E. coli* contains at least 36 putative toxin-antitoxin gene pairs, and some pathogens such as *Mycobacterium tuberculosis* (*Mtb*) have over 90 toxin-antitoxin operons. *E. coli* MazF cleaves free mRNA after encountering stress and nine *Mtb* MazF family members cleave mRNA, tRNA, or rRNA. Moreover, *Mtb* MazF-mt6 cleaves 23S rRNA Helix 70 (H70) to inhibit protein synthesis. The overall tertiary folds of these MazFs are predicted to be similar, therefore it is unclear how they recognize structurally distinct RNAs. Here we report the 2.7 Å X-ray crystal structure of MazF-mt6. MazF-mt6 adopts a PemK-like fold, but lacks an elongated β1-β2 linker, a region that typically acts as a gate to direct RNA or antitoxin binding. In the absence of an elongated β1-β2 linker, MazF-mt6 is unable to transition between open and closed states, suggesting that the regulation of RNA or antitoxin selection may be distinct from other canonical MazFs. Additionally, a shortened β1-β2 linker allows for the formation of a deep, solvent accessible, active-site pocket which may allow recognition of specific, structured RNAs like H70. Structure-based mutagenesis and bacterial growth assays demonstrate that MazF-mt6 residues Asp-10, Arg-13, and Thr-36 are critical for RNase activity and likely catalyze the proton-relay mechanism for RNA cleavage. These results provide further critical insights into how MazF secondary structural elements adapt to recognize diverse RNA substrates.

INTRODUCTION

Toxin-antitoxin gene pairs are two component operons distributed widely among prokaryotic and archaea genomes where they have diverse roles in bacterial physiology. These small gene pairs are critical for post-segregational killing, adaptive responses to environmental stress, antibiotic tolerance and host-pathogen virulence responses (1, 2). Toxin-antitoxins are organized into six different classes based on how the antitoxin component regulates toxin activity (3). In all classes, the toxin is proteinaceous and inhibits an essential cellular function such as replication, translation or an assembly pathway including ribosome and membrane biogenesis, in response to stress (2, 4). The antitoxin is either RNA or protein that inhibits toxin activity by direct binding (Type II, III, VI), preventing toxin protein expression (Type I, V) or competing with the toxin for substrate binding (Type IV) (3).

Type II toxin-antitoxin complexes are the most abundant and best characterized toxin-antitoxin systems (5). In this type, the antitoxin protein neutralizes the toxin protein by direct binding. Most Type II toxins are endoribonucleases and are either ribosome-dependent or -independent. Ribosome-dependent toxins such as *E. coli* RelE, YoeB, YafQ, and *P. vulgaris* HigB cleave mRNA but only when bound to the ribosome (6-8). In contrast, *E. coli* and *B. subtilis* MazFs are endoribonucleases that degrade free mRNAs at sequences containing 5'-N↓ACA-3' and 5'-U↓ACAU-3', respectively (9, 10). The *Mycobacterium tuberculosis* (*Mtb*) genome contains more than 90 putative toxin-antitoxin operons with nine MazEF paralogs (annotated as MazF-mt1 - MazF-mt9) (5). Three of the nine MazF toxins (MazF-mt3, MazF-mt6 and MazF-mt9) cause growth arrest when overexpressed in *E. coli* (11). Surprisingly, in addition to cleaving mRNAs, *Mtb* MazF toxins cleave different RNAs that are structurally distinct. For example, MazF-mt9 cleaves tRNAs, MazF-mt6 cleaves 23S rRNA and MazF-mt3 cleaves both 16S and 23S rRNA, all resulting in the inhibition of protein synthesis (12-14). Although the overall tertiary fold of

these MazFs are predicted to be similar (13), it is unclear how they then recognize very structurally different RNAs.

MazF-mt6 toxin cleaves 23S rRNA at Helix 70 (H70) of the large 50S subunit at a consensus sequence of 5-^{1939}UU↓CCU^{1943}-3'(13). H70 is at the interface between the 30S and 50S subunits and adopts a 10-nucleotide (nt) stem-loop with the consensus sequence located in an internal bulged region. Here, we solved the 2.7 Å X-ray crystal structure of *Mtb* MazF-mt6 and show MazF-mt6 adopts a PemK-like fold, an architecture commonly adopted by several bacterial toxins including PemK and CcdB (15). MazF-mt6 contains a significantly shorter β1-β2 linker, the region predicted to interact with RNA (13). We find that MazF-mt6 cleaves a 35-mer H70 RNA strongly suggesting that the entire 23S rRNA is not required for MazF-mt6 activity. We further determine that MazF-mt6 residues Asp-10, Arg-13 and Thr-36 are critical for activity and likely perform mRNA cleavage via a proton relay mechanism. These data provide a structural and functional framework for how MazF endoribonucleases recognize diverse RNA substrates by the remodeling of loops to mediate RNA recognition.

RESULTS

MazF-mt6 cleaves both 23S rRNA and Helix 70 (H70) - MazF-mt6 inhibits protein synthesis by cleaving the 23S rRNA at H70 (^{1939}UU↓CCU^{1943}), an RNA sequence conserved between *Mtb* and *E. coli* 70S (13). H70 forms interactions with the A-site tRNA and ribosome recycling factor during elongation and interacts directly with 16S rRNA decoding helix 44 (16, 17). To test for activity, we incubated MazF-mt6 with purified *E. coli* 50S and monitored 23S rRNA cleavage product formation using reverse transcription (RT) from a complementary DNA oligo annealed to 23S rRNA nts 1959-1945 (**Fig. 3.1*B***). We used a poisoned RT reaction that yields full length and cleaved products of 19 and 25 nts, respectively (18). The extension is halted by incorporating dideoxyCTP at position 1935,

six nts after the cleavage site. Using this experimental setup, we confirmed MazF-mt6 primarily cleaves between residues 1940 and 1941 (**Fig. 3.1*C*; Fig. 3.8)**. We next tested whether MazF-mt6 required an intact 50S by incubating MazF-mt6 with a 35 nt RNA containing the H70 sequence (nts 1933-1967) (**Fig. 3.1*B***). The major RT product is eleven nts corresponding to cleavage between 1940-1941 (64% of the product formed) as previously demonstrated (13). However, there are additional minor RT products corresponding to cleavage after C1942 and U1943 in the same bulged loop (8% and 23%; **Fig. 3.1*C***). While MazF-mt6 cleaves H70 almost to completion, MazF-mt6 only cleaves 23S rRNA in the 50S to ~30% completion.

Structure determination of MazF-mt6- MazF-mt6 was overexpressed and purified fused to an N-terminal hexahistidine $(His)_6$ Trigger Factor (TF) as previously described (13). In addition to immobilized metal affinity chromatography (IMAC) purification, the $(His)_6$TF-MazF-mt6 complex was incubated with thrombin protease to release TF. MazF-mt6 was further purified by IMAC and size exclusion chromatography (Superdex S200, GE Healthcare) and concentrated to 10 mg/ml. Crystals formed in the $P6_422$ space group with two identical, symmetry related MazF monomers in the asymmetric unit. The X-ray crystal structure was solved to 2.7 Å resolution by molecular replacement using *B. subtilis* MazF (BsMazF) as a search model in the Phaser program of the PHENIX suite (**Table 3.1, Fig. 3.2,** PDB code 4MDX) (19, 20). The MazF-mt6 model was manually built in the Crystallographic Object-Oriented Toolkit program (COOT) and geometry and B-factor refinements in PHENIX and rebuilding in COOT were iteratively performed (20, 21). The final model contains residues 2-99 (out of 103 total residues).

MazF-mt6 adopts a PemK-like, ribonuclease fold- MazF-mt6 is a small globular protein (10.8 kDa) consisting of two β-sheets and three α-helices. One β-sheet adopts a five-stranded, antiparallel SH3-like barrel fold (β7 ↑ β1 ↓ β2 ↓ β3 ↑ β6 ↓) linked by four small α-helices (α1, α2, α3, α4) while the second β-sheet is three-stranded and antiparallel

(β4↓ β5 ↑), connected to a large C-terminal α-helix (α4) (22) (**Fig. 3.2*A***). While MazF-mt6 shares low amino acid sequence identity with orthologous family members (15-28%) (**Fig. 3.9**), MazF-mt6 adopts a plasmid emergency maintenance K(PemK)-like fold as seen in other bacterial toxins including *E. coli* CcdB, MazF and PemK (23-27). A DALI search (28) reveals that MazF-mt6 shares high structural homology with *S. aureus* MazF (PDB code 4MZP; root mean square deviation (rmsd)=1.4 Å, Z=15.7), BsMazF toxin (PDB code 4MDX; rmsd=1.4 Å, Z=15.5), *E. coli* MazF toxin (EcMazF; PDB code 5CR2; rmsd=1.9 Å, Z=12.6), *B. anthracis* MoxT toxin (PDB code 4HKE; rmsd=1.4 Å, Z=15.3), *E. coli* CcdB toxin (PDB code 2VUB; rmsd=2.3 Å, Z=12.9) and *E. coli* Kid toxin (PDB code 1M1F; rmsd=2.0 Å, Z=11.8). Dimerization of MazF-mt6 is mediated by C-terminal α4-α4 interactions similar to BsMazF (19, 29)(**Fig. 3.2*A***). MazF-mt6 dimerization buries ~2,000 $Å^2$ of hydrophobic and electrostatic interactions between β6 and α4 in comparison to other MazFs that have an additional 500 $Å^2$ buried.

Sequence alignments of MazF-mt6 with BsMazF and EcMazF (referred to as MazF hereafter) reveal RNA-binding regions that differ in length (19). Comparison of the structure of MazF-mt6 with MazFs that interact with mRNA reveals that MazF-mt6 contains a significantly shorter β1-β2 linker, the region that interacts with mRNA (19, 29) (**Fig. 3.2*B***). The MazF β1-β2 linker is thirteen residues in length, while in MazF-mt6, β1-β2 linker is only five residues. The MazF β1-β2 linker extends ~10 Å from one monomer to the other monomer forming part of the dimer interface that packs against and covers a net positive charge where mRNA binds (19). Since there is a shortened MazF-mt6 β1-β2 linker, this region does not extend towards the dimer interface leaving the positively charged cleft exposed. An additional positive region is exposed upon MazE antitoxin binding (19, 29). This second positively charged region is missing in MazF-mt6 and instead, a negative solvent exposed patch is present.

MazF-mt6 residues Asp-10, Arg-13 and Thr-36 are important for activity and inhibition of bacterial growth- The MazF-mt6 structure reveals that residues Asp-10, Lys-11, Arg-13, Thr-36, Thr-37 and Thr-38 are adjacent to the putative active site, which also contains a bound sulfate ion (**Fig. 3.3*A***; **Fig. 3.10**). The negatively charged sulfate ion can mimic the phosphate backbone of RNA and provides further support that this location is likely the active site. To identify which MazF-mt6 residues are important for toxin activity, we used an established cell-based assay whereby overexpression of wild-type (WT) MazF-mt6 inhibits growth (11, 30). MazF-mt6 variants that restore growth are interpreted as disrupting canonical MazF-mt6 function. As demonstrated previously, overexpression of WT MazF-mt6 inhibits cell growth in *E. coli* while uninduced or vector control only exhibit normal growth (**Fig. 3.3*A*; Fig. 3.11**) (19). MazF residues previously identified as important for activity include an arginine that functions as both an acid and a base and either a serine or threonine that stabilizes the transition state during mRNA cleavage (19, 29, 31). MazF-mt6 residue Arg-13 is located on the shortened β1-β2 linker, in a similar location as catalytic residues EcMazF Arg-29 and BsMazF Arg-25. Expression of MazF-mt6 R13A variant allows for normal growth demonstrating this residue is important for activity (**Fig. 3.3*B***). Proximal residues to the MazF-mt6 active site include Asp-10 and Lys-11. Expression of the MazF-mt6 D10A and K11A variants allow for growth suggesting these residues have a critical role in activity (**Fig. 3.3*B***). MazF-mt6 contains three threonine residues adjacent to its active site suggesting each could potentially stabilize the transition state. We find that the T38A variant has no effect on MazF-mt6 inhibition of growth while the T37A variant has a modest growth effect (**Fig. 3.3*C***). In contrast, the T36A variant causes a reversal of growth inhibition suggesting Thr-36 plays a critical role in MazF-mt6 activity.

MazF-mt6 residues Asp-10, Arg-13 and Thr-36 have the largest impact on rRNA degradation – Bacterial growth assays suggested MazF-mt6 residues Asp-10, Arg-13,

Thr-36, and possibly Lys-11 are important for endoribonuclease activity (**Fig. 3.3**). To assess the quantitative impact each MazF-mt6 residue has on mRNA cleavage, we performed single turnover assays where 5'-^{32}P labeled H70 RNA was incubated with a 10-fold molar excess of MazF-mt6 over a time course of 45 mins (**Fig. 3.4; Fig. 3.12**). The cleavage of H70 by WT MazF-mt6 was monitored by the accumulation of RNA products corresponding to 8-11 nts. The major product for WT MazF-mt6 cleavage was an RNA product corresponding to cleavage between U1940 and C1941 (3.7×10^{-3} s^{-1}) (**Table 3.2**). MazF-mt6 variants R13A and T36A had the most significant impact on activity with minimal RNA cleavage products detected (**Fig. 3.4*A***). In contrast, MazF-mt6 variants K11A, T37A and T38A had minor effects on RNA cleavage as compared to WT suggesting that although these residues are proximal to the active site, they do not contribute substantially to catalysis. Although the K11A variant appeared to have a large effect on activity in the bacterial growth assays (**Fig. 3.3*B***), the variant has a minimal effect on activity in single turnover experiments. These data suggest that the K11A variant either has impaired mRNA binding or lower protein solubility which could reduce its expression in the cell-based assays. MazF-mt6 residue Asp-10 is adjacent to critical residue Arg-13. The MazF-mt6 D10A variant decreased the rate of RNA cleavage by 10-fold (4×10^{-4} s^{-1}), suggesting this residue is important for catalysis (**Fig. 3.4*A***).

In addition to the canonical eight nt product produced by all MazF-mt6 variants, RNA products of 9-11 nts were also seen but at varying extents depending on the variant (**Fig. 3.4*A***). For example, MazF-mt6 variants D10A, R13A and T36A that had large impacts on RNA degradation produced additional RNA products in similar amounts as compared to the canonical eight nt product. In contrast, MazF-mt6 variants T37A and T38A that had a modest impact on overall activity, produced products at the canonical position to the largest extent (40 and 60% respectively of overall product production), and 9-11 nt RNAs were only seen at low levels (2-13%). The MazF-mt6 K11A variant impacts cleavage to a

minimal extent (**Table 3.2**; 3-fold), however, the major RNA product was eleven nts corresponding to cleavage between U1943 and U1944. Thus, in addition to these MazF variants altering activity at the major cleavage site, they also seem to impact specificity. Another possibility is that in the absence of the 50S subunit, the MazF-mt6 dimer may be more flexible in how it engages the bulged loop containing the consensus sequence, thus resulting in nt promiscuity.

DISCUSSION

Bacterial toxins regulate growth in response to environmental stress including antibiotic treatment (32). In *E. coli*, there is a single MazF toxin family member that cleaves free mRNA to inhibit translation (31). The MazF family is expanded to nine members in *Mtb* concurrent with the expansion of different target RNAs including tRNAs and rRNAs (12-14). This expansion suggests that *Mtb* MazFs may contain different structural elements to recognize diverse RNA substrates. In support of this, *Mtb* MazF sequence alignments reveal truncations in specific loop regions known to interact with RNA (11). The structure of MazF-mt6 reveals this MazF family member adopts a nearly identical tertiary fold as BsMazF and EcMazF (19, 29) (rmsd = 1.4 and 1.9 Å, respectively). Despite this similarity, MazF-mt6 contains a significantly shorter β1-β2 linker, the region known to contact its RNA substrate (19). MazF β1-β2 linkers typically contain thirteen residues which surround the single-stranded mRNA substrate. The MazF-mt6 β1-β2 linker is truncated to five residues which we predict would alter how it interacts with H70. H70 adopts a stem-loop structure with an internal bulged loop region where the MazF-mt6 cleavage site is located. Therefore, the length of the β1-β2 linker may be a consequence of the rigidity of the RNA it encounters. For example, MazFs that recognize flexible RNAs such as single-stranded mRNAs may require the loop to confine and orient the substrate. The H70 RNA likely require less remodeling by the β1-β2 linker and therefore alteration

of this MazF-mt6 region may allow for different MazF homologues to distinguish among diverse RNAs.

In addition to the β1-β2 linker interacting directly with RNA substrates, this region also modulates an open to a closed MazF state effectively functioning as a gate to allow either RNA or its cognate antitoxin MazE to bind (29). In the context of a MazF dimer which forms the closed state, the β1-β2 linker extends across the face of the monomer to interact with the adjacent MazF monomer (19). These interactions are mediated by symmetry-related electrostatics between the distal end of the β1-β2 linker (Glu-24) and β6/α3 of the other monomer (Lys-79 and Arg-86) (**Fig. 3.5*A***). This β1-β2 linker extension additionally establishes a positively charged RNA binding pocket and obscures an adjacent and slightly overlapping positively charged region where the antitoxin MazE binds (29). Therefore, this closed state appears to direct mRNA binding while preventing MazE binding. Antitoxin MazE binding induces an open state by displacing the β1-β2 linker and thus disrupting interactions between the MazF monomers, causing a reorganization of MazF active site residues (29). β1-β2 linker residue Glu-24 is critical for MazF activity (29) suggesting that maintenance of interactions between the monomers is important for MazF homologues that recognize mRNA.

The structure of MazF-mt6 reveals a truncated β1-β2 linker suggesting that it likely does not undergo conformational changes to convert from a closed to an open state as seen in other MazF homologues. The shortening of β1-β2 allows for a permanently accessible, positively-charged channel where mRNA binds, while at the dimer interface, a negatively charged rather than a positively charged surface forms (**Fig. 3.5*B***). A C-terminal MazE α-helix typically binds at the positively charged, MazF dimer interface (19), however, MazF-mt6 lacks this positively charged surface (**Fig. 3.5*B***). This suggests that MazE-mt6 antitoxin binding required to neutralize MazF-mt6 may be distinct from other MazEs solved to date. Homology modeling of MazE-mt6 predicts a C-terminal α-helix

similar to other structures of MazE and ParD (19, 29, 33, 34). Two positively charged residues align on one side of the α-helix, which could bind the negatively charged dimer interface **(Fig. 3.5*C*).** As mentioned, MazE binding disrupts the MazF active site and in particular the catalytic arginine (29). Modeling of the MazE-mt6 C-terminal α-helix bound to MazF-mt6 indicates MazE-mt6 may be able to interact with catalytic Arg-13 suggesting active site disruption could occur.

MazF endoribonucleases cleave RNA substrates via a proton charge relay mechanism whereby an arginine functions as both a general acid and base and a threonine or serine stabilizes the transition state (29, 35) (**Fig. 3.6*A***). In the absence of MazE, and therefore containing a longer β1-β2 linker in a closed state, the catalytic arginine of MazF forms a hydrogen bond with the backbone (Gly-22) which is located at the distal end of the β1-β2 linker (**Fig. 3.6*A***), a region missing in MazF-mt6. Instead, the shorter MazF-mt6 β1-β2 linker forms an electrostatic interaction with Arg-13 using the negatively charged carboxyl group of Asp-10 (**Fig. 3.6*A***). In MazF, a threonine is located adjacent to the scissile phosphate and is predicted to stabilize the transition state and in MazF-mt6, Thr-36 is in an equivalent location. We show that both Asp-10 and Thr-36 are critical for MazF-mt6 activity by either stabilizing the catalytic Arg-13 or the RNA, respectively. Other nearby MazF-mt6 residues including Lys-11, Thr-37 and, to a lesser extent, Thr-38 potentially contribute to RNA substrate stabilization by the formation of electrostatic interactions with the phosphate backbone or ribose as these residues are close to Arg-13 and Thr-36 (**Fig. 3.2*B***). Although Thr-38 is adjacent to the active site, our data suggests this residue has no significant role in RNA catalysis. Therefore, despite the significantly shorter β1-β2 linker, MazF-mt6 active site residues form functionally equivalent interactions seen in other MazF orthologs.

MazFs cleave single-stranded mRNA at a consensus sequence of 5'-ACA-3', with BsMazF requiring at least one uridine flanking the 5' and 3' ends of the consensus (27,

31). MazF-mt6 differs by requiring an additional uridine upstream of the consensus 5'-UU↓CCU-3' sequence of the cleavage site (13). Structurally aligning the MazF-RNA complex to MazF-mt6 suggests several potential residues that could contribute to H70 RNA specificity. For upstream specificity, MazF-mt6 Arg-75 (located on β7) is located in a similar position as BsMazF Phe-10 and Asp-91, which recognize the -2 uridine position **(Fig. 3.6*B*)**. A conserved BsMazF serine (Ser-73) is located in a pocket formed by the β4-β5 linker and interacts with the Watson-Crick face of the -1 uridine. MazF-mt6 Ser-58 likely fulfils an analogous role **(Fig. 3.6*B*)**. In previously solved structures of MazF, the β1-β2, β3-β4 and β5-β6 linker residues recognize the downstream nts ACAU (19). Since MazF-mt6 has a significantly shorter β1-β2 linker, the β5-β6 and β3-β4 regions are likely the only linkers responsible for base-specific recognition. The +1 position is recognized by the elongated β1-β2 linker, but, in the case of MazF-mt6, the +1 cytidine may be recognized by β5-β6 linker residue Gln-66 because it is located in an equivalent position **(Fig. 3.6*B*)**. MazF-mt6 residues Asp-63 and Asn-64 align with the MazF residues Glu-78 and Gln-79 suggesting these residues play similar roles in base specific nt interactions of the +2 and +3 regions. Lastly, by containing a truncated β1-β2 linker, the RNA binding path appears deeper, potentially allowing a structured RNA like the bulged loop region containing the H70 consensus sequence, to better fit on the surface of MazF-mt6.

MazF homologues are part of the PemK-like endonuclease family that cleave single-stranded mRNA transcripts (27). The one exception is the *E. coli* F-plasmid CcdB toxin that inhibits bacterial growth by binding to DNA gyrase to prevent replication (36). Despite having different functions and targeting different types of molecules, CcdB toxins are almost identical to MazFs (37) (**Fig. 3.7**). However, one key difference is the surface each uses to engage their different substrates. For example, β1-β2, β3-β4 and β5-β6 linkers of MazF interact with RNA while the α4 dimer interface of CcdB binds GyrA, a region on the opposite side of MazF (**Fig. 3.7**) (38). Additionally, in order to bind GyrA, CcdB toxin bends

12º at the α4-α4 dimer interface, a movement which disorders the β1-β2 linker (**Fig. 3.7*C***) (37). CcdA antitoxin binding orders the CcdB β1-β2 linker but prevents the α4-α4 conformational rearrangement required to bind GyrA. Therefore, the ordering/disordering transition is opposite between MazFs and CcdB (19, 29, 38), as has been suggested previously (37). The lack of apparent β1-β2 linker regulation in MazF-mt6 indicates that how it is regulated has diverged from other MazFs.

EXPERIMENTAL PROCEDURES

Mycobacterium tuberculosis MazF-mt6 and MazF-mt6 variant purifications- Overnight cultures *of E. coli* C41(DE3) cells (Lucigen) transformed with pCold-$(His)_6$-Trigger Factor (TF)-MazF-mt6 (kind gift of the Woychik lab) (13) were diluted 1:100 in Lysogeny Broth (LB) medium supplemented with 100 µg/mL ampicillin. Cultures were grown to an OD_{600} of 1.0 at 37˚C, transferred to 15˚C for 30 min, and protein expression was induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and grown for 24 hrs at 15°C. Cells were harvested and resuspended in lysis buffer (40 mM Tris-Cl pH 8.0, 500 mM NaCl, 10% (w/v) glycerol, 0.1% (w/w) Triton X-100, 5 mM β-mercaptoethanol (β-Me), 20 mM imidazole, 0.1 mM benzamidine and 0.1 mM phenylmethanesulfonylfluoride), and lysed using an Emulsiflex C5 (Avastin). Cell debris was clarified by centrifugation at 35,000 *x g* for 50 min and applied to a 5 mL Ni^{2+}-NTA affinity column (GE Healthcare) at 10°C. The $(His)_6$-TF-MazF-mt6 fusion protein was eluted with a linear gradient of lysis buffer without Triton X-100 or protease inhibitors but containing 500 mM imidazole. The sample was dialyzed overnight into thrombin cleavage buffer (20 mM Tris-Cl pH 8.0, 150 mM NaCl, 2.5 mM $CaCl_2$), 1 unit of thrombin per 200 µg protein was added at room temp for 3 hrs, and further incubated overnight at 10˚C. The cleaved $(His)_6$-TF-MazF was reapplied to the Ni^{2+}-NTA affinity column and MazF-mt6 was collected from the flow through fraction. MazF-mt6 was further purified and buffer exchanged on a size exclusion Superdex 200

column (GE Healthcare) into 20 mM Tris-Cl pH 7.5, 250 mM KCl, 10% (w/v) glycerol and 5 mM β-Me. MazF-mt6 was determined to be 95% pure by 15% SDS-PAGE, flash frozen in liquid nitrogen and stored at -80˚C.

MazF-mt6 variants were constructed using site-directed mutagenesis and sequences were verified by DNA sequencing (oligos listed in **Table 3.3)**. MazF-mt6 variant purifications were performed similar to wild-type. Each variant eluted as a single, symmetrical peak at the expected volume during the gel filtration step as wild-type suggesting the variants are properly folded.

MazF-mt6 crystallization, data collection and structure determination- MazF-mt6 (10 mg/mL) crystallized by hanging drop vapor diffusion in 50 mM sodium cacodylate pH 6.0, 10 mM $MgCl_2$, and 1.05 M $LiSO_4$ at 20˚C in three weeks to dimensions of ~100 μm^3. The crystals were cryoprotected in 50% w/v lithium sulfate, 30% glycerol and 150 mM ammonium tartrate for approximately 60 s and flash frozen in liquid nitrogen. Native X-ray diffraction datasets were collected at the Northeastern Collaborative Access Team (NE-CAT) 24-IDC beamline at the Advanced Photon Source. The datasets were collected under cryogenic conditions (100 K) using 0.979 Å radiation. The data were integrated and scaled to 2.9 Å using the X-ray Detector Software (XDS) software package (39). The *B. subtilis* MazF-mRNA complex (PDB code 4MDX) was used as a search model (with the mRNA removed) and all MazF residues were changed to alanine during molecular replacement using AutoMR from the PHENIX software suite (20). Multiple rounds of positional, B-factor and energy minimization refinement in PHENIX were followed by manual model rebuilding in Coot was performed with R_{work} and R_{free} converging to 23.9/27.1% (21). The final MazF-mt6 model consisted of residues 1-99 out of 103 total. The model quality was determined using Molprobity of the PHENIX validation suite and figures were generated using PyMOL (40).

23S rRNA cleavage and reverse transcription- E. coli MRE600 70S were purified as previously described (41). The 50S subunit was purified by dialyzing the 70S ribosome in low magnesium (2.5 mM) and separated by sucrose gradient. The 23S rRNA Helix 70 (H70) residues 1933-1967 was chemically synthesized (Integrated DNA Technologies (IDT); 5'-^{1933}GCGAAAU UCCUUGUCGGGUAAGUUCCGACCUG C^{1967}-3'). *E. coli* 50S or H70 RNA was incubated at 37°C for 45 min with a ten-fold molar excess of wild-type MazF-mt6 in 5 mM HEPES pH 7.5, 50 mM KCl and 10 mM NH_4Cl. Reactions were halted by the addition of either phenol:chloroform (5:1) or phenol:chloroform:isoamyl alcohol (25:24:1) for H70 RNA or 50S, respectively. Both RNAs were next chloroform extracted once, precipitated overnight at -20°C by the addition of 0.3 M sodium acetate pH 5.2 and three volumes of ice cold 100% ethanol, and resuspended in milliQ water.

The reverse transcription primer (IDT; 5'-GTCGGAACTTACCCGAC-3') was 5'-^{32}P-labeled using T4 polynucleotide kinase (New England Biolabs) and 4 pmols of the primer was annealed to 4 or 7 pmols of RNA in 50 mM Tris-Cl, pH 8.0 and 50 mM NaCl to 90°C or 65°C. These reactions were slow cooled to 42°C by removing the heat block and placing on the benchtop at room temperature. The reverse transcription reactions were initiated with the addition of 1 μM dATP, 1 μM dGTP and 1 μM dTTP, limiting amounts of 0.1 μM ddCTP and Protoscript II reverse transcriptase (NEB) as previously described (18). Reactions were incubated at 42°C for 1 hour, then halted with the addition of 1 volume of formamide dye (100% v/v deionized formamide, 0.01% w/v bromophenol blue and 0.01% w/v xylene cyanol) and heated to 90°C for 2 min. Dideoxy sequencing reactions were performed as previously described (18). Samples were run on a preheated 50% urea/20% polyacrylamide gel at 37 mA limiting for 90 mins. The gel was fixed in 20% (v/v) methanol, 20% (v/v) acetic acid and 3% glycerol (v/v) for 60 min, dried for 1 hr, exposed to a phosphor screen for 1 hr, and visualized using a Typhoon FLA 7000 gel imager (GE Healthcare).

Exposure was visualized using a Typhoon FLA 7000 gel imager (GE Healthcare), with the band density determined by ImageQuant (GE Healthcare).

mRNA cleavage assays- H70 was 5'-^{32}P labeled as described above and incubated (0.2 μM RNA) with 2 μM wild-type MazF-mt6 or MazF-mt6 variants for 45 min at 37°C. Aliquots of the reactions were halted at 1, 3, 5, 10, 25 and 45 min time points with the addition of 1 volume of formamide gel loading buffer and heated to 70°C for 2 min. Samples were run on a 50% urea/20% polyacrylamide gel at 37mA limiting for 2 hrs. The gel was fixed in 40% ethanol, 20% acetic acid and 3% glycerol for 1 hr, dried for 90 min, then exposed to a GE phosphor screen for 30 min and visualized using a Typhoon FLA 7000 gel imager (GE Healthcare). Exposure was visualized using a Typhoon FLA 7000 gel imager (GE Healthcare), with the band density determined by ImageQuant (GE Healthcare). Background density was subtracted using default settings. The amount of cleavage product was determined by determining the amount of full length RNA in each lane. The amount of full length RNA was divided by the total lane density for each sample and the reciprocal of the resulting reaction was recorded as the amount of RNA cleaved. The amount of cleaved product against time were fit by GraphPad Prism 5 using the single exponential equation:

$$\text{Product} = P_{max}(1 - e^{-kt})$$

where P_{max} is the mRNA cleavage product plateau (pmol mRNA cleaved), k is the observed rate constant (min^{-1}), and t is the time the reaction progressed (min). The rates of both MazF-mt6 R13A and T36A were so low that we were unable to effectively fit the points to the single turnover equation with an R-squared > 80%.

Homology modeling of MazE-mt6- The homology model of MazE-mt6 was generated by the HHpred homology model workflow and Modeller using the optimal

multiple template function (42, 43). The structural templates used in Modeller to generate the MazE-mt6 homology model were *E. coli* ParD (from the ParD-ParE complex; PDB code 3KXE, chain C (33), 20% sequence identity to MazE-mt6), *M. opportunistum* ParD3 (from the ParD3-ParE3 complex; PDB code 5CEG, chain A (34), 12% sequence identity), *H. sapiens* Calmodulin (PDB code 2K61, chain A (44), 7.8% sequence identity), *E. coli* NikR (from NikR-DNA operator complex; PDB code 2HZA, chain A (45); 9.3% sequence identity) and *S. agaiactiae* CopG (PDB code 2CPG, chain A (46); 16.3% sequence identity).

ACKNOWLEDGMENTS:

We thank Dunham lab members Drs. T. Maehigashi, J. Meisner, and M.A. Schureck for their helpful discussions throughout the project, Dr. M. Witek for providing purified *E. coli* 50S and Dr. G.L. Conn for critical reading of the manuscript. We thank Dr. Woychik for reagents and Woychik lab member Dr. Schifano for technical guidance. This work was supported in part by an NSF CAREER award number MCB 0953714 (CMD) and Emory School of Medicine Bridge Funds. This work is based upon research conducted at the NE-CAT beamline, which are funded by the National Institute of General Medical Sciences from the NIH [P41 GM103403] and at the SER-CAT beamline. The Pilatus 6M detector on 24-ID- C beam line is funded by a NIH-ORIP HEI grant [S10 RR029205]. Use of the Advanced Photon Source, an Office of Science User Facility operated for the US Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the US DOE under Contract DE-AC02-06CH11357.

FIGURES AND TABLES

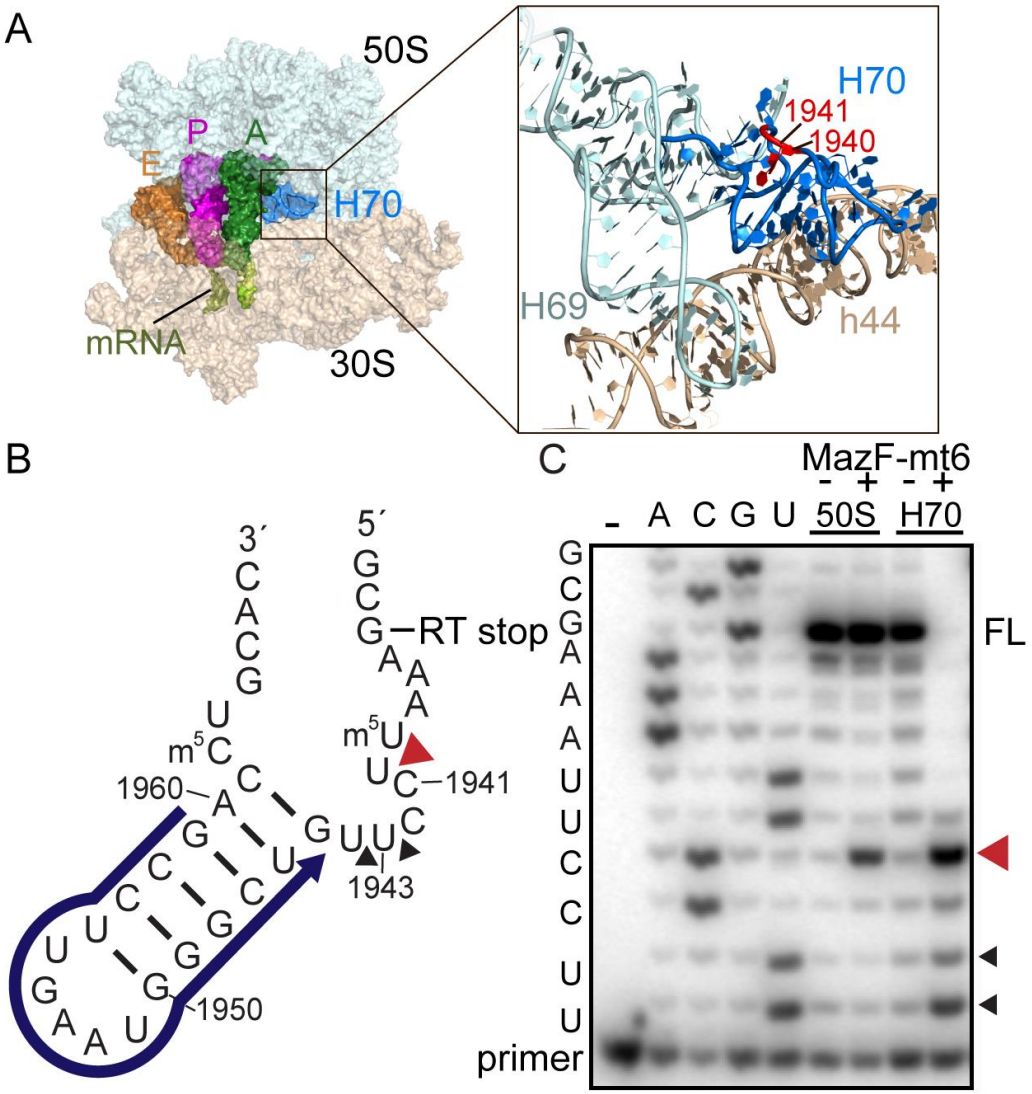

Figure 3.1 Location of 23S rRNA Helix 70 in the context of the ribosome. *A*) Overview of the 70S containing A-site, P-site and E-site tRNAs (PDB code 4Y4R)(left panel). An expanded view of 23S rRNA Helix 69 (H69) and Helix 70 (H70) and 16S rRNA helix 44 (h44) is shown in the right panel. H70 residues that MazF-mt6 targets are shown in red. *B*) H70 (nucleotide position 1933–1967) indicating the location of the reverse transcription (RT) primer (blue), the MazF-mt6 cleavage site (red arrowhead) and the RT stop at nucleotide 1937. *C*) RT assays demonstrating MazF-mt6 induced 23S rRNA or H70 cleavage. The cleavage products were monitored by RT and the reactions were run on an 8M Urea/PAGE sequencing gel. The MazF-mt6 full length RT stop (FL), the major cleavage site (red arrowhead) and a minor cleavage site (black arrowhead) are indicated.

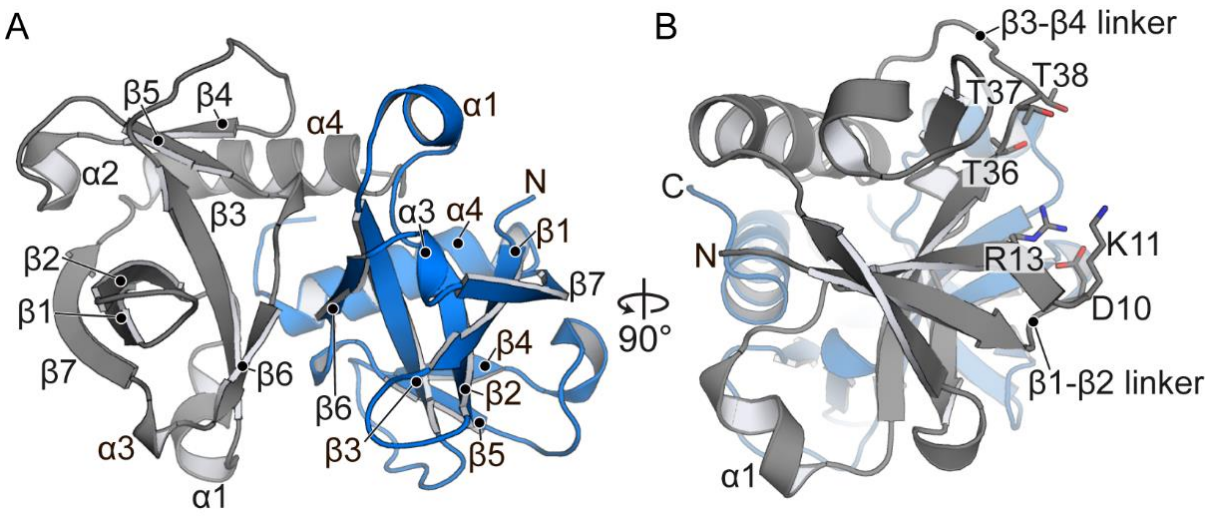

Figure 3.2 Structure of *Mycobacterium tuberculosis* MazF-mt6. *A*) The MazF-mt6 is a homodimer consisting of seven β-strands and four α-helices. *B*) A 90° rotated view to show the SH3-like antiparallel β-sheet and the proximity of active site residues location on the β1-β2 and β3-β4 linkers.

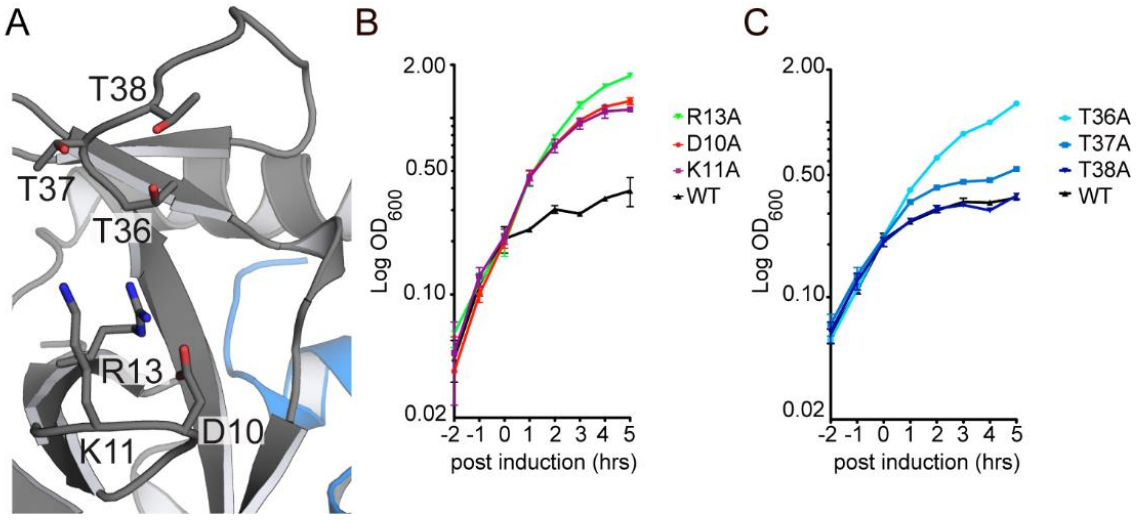

Figure 3.3 MazF-mt6 residues important for activity. *A*) MazF-mt6 putative active site highlighting the residues that were targeted for single variant analysis. *B & C) E. coli* BW25113 growth monitored over 5 hrs post induction of wild-type (WT) MazF-mt6 and active site variants. Cells were grown in M9 media supplemented with 0.2% glycerol and 0.2% casamino acids and induction was initiated by the addition of 0.2% arabinose after cells reached an optical density of 0.2 at 600 nm. Error bars represent standard error of means from three independent experiments.

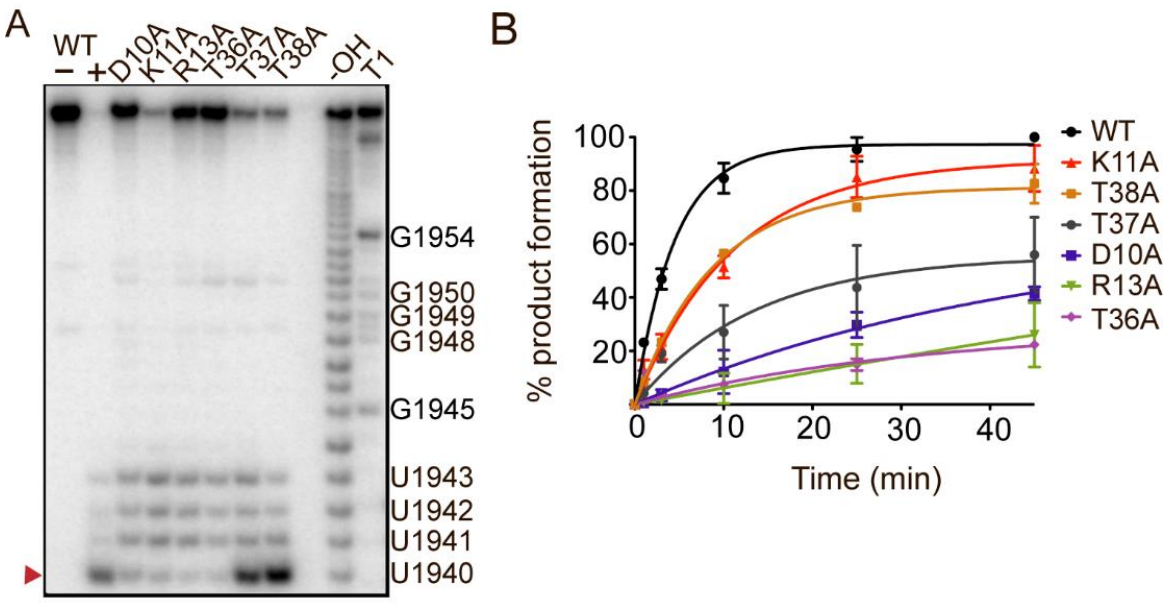

Figure 3.4 MazF-mt6 residues Asp-10, Arg-13 and Thr-36 are critical for endonuclease activity. *A*) Single turnover experiments monitoring H70 cleavage by wild-type (WT) MazF-mt6 and MazF-mt6 variants. The products were analyzed on an 8M Urea/PAGE sequencing gel. The major cleavage product is denoted with a red arrowhead at U1940 and other minor cleavage products are indicated at positions 1941-43. *B*) The product progression plot of WT MazF-mt6 and MazF-mt6 variants as monitored over a time course of 45 mins.

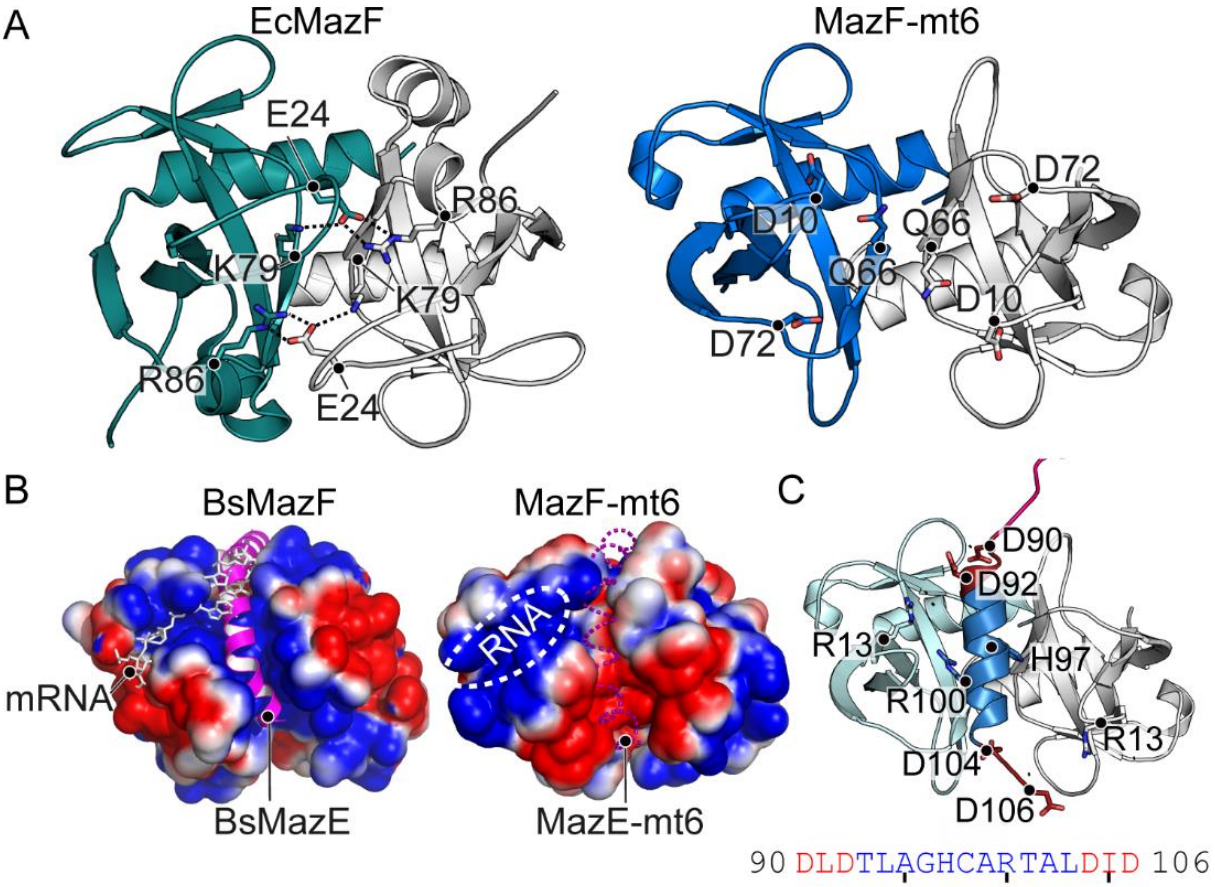
A
EcMazF
E24
R86
K79
K79
R86
E24
MazF-mt6
D72
D10
Q66
Q66
D10
D72
B
BsMazF
MazF-mt6
RNA
mRNA
BsMazE
MazE-mt6
C
D90
D92
R13
H97
R100
R13
D104
D106
90 DLDTLAGHCARTALDID 106

Figure 3.5 Comparison of MazFs. *A*) Comparison of electrostatic interactions between the EcMazF β1-β2 and β5-β6 linkers with an adjacent EcMazF dimer (PDB code 5CR2) (left) with the absence of electrostatic interactions between MazF-mt6 dimers (right). Residues responsible for stabilizing the β1-β2 linker in EcMazF are shown as sticks with dotted lines highlighting ionic bonds. The equivalent residues to EcMazF are shown for MazF-mt6, but there are no electrostatics between these residues. *B*) Electrostatic surface potential of BsMazF with both the MazE C-terminal α helix (magenta) and the mRNA (white) shown to emphasize regions where their binding overlaps (PDB code 4ME7) (left). The electrostatic surface potential of MazF-mt6 with both the putative antitoxin binding path (dashed magenta α-helix) and the RNA binding pocket (white circle). The putative antitoxin binding path is negatively charged in contrast to BsMazF. *C*) Homology model of MazE-mt6 C-terminal α helix (blue) superimposed onto MazF-mt6. MazF-mt6 catalytic residue Arg-13 is predicted to be close to MazE-mt6 residues Asp-90, Asp-92, Asp 104 and Asp-106 which could potentially disorder the MazF-mt6 active site. The MazE-mt6 peptide sequence is shown below.

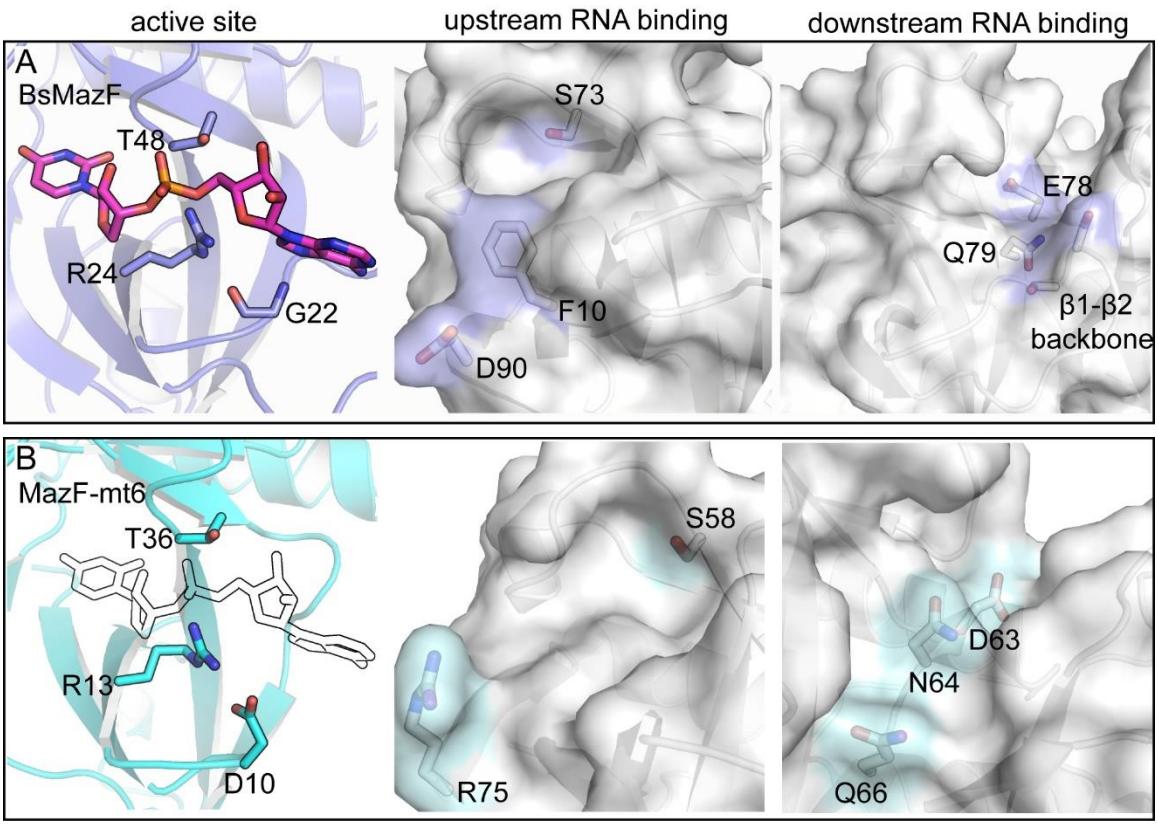

Figure 3.6 Comparison of MazF nucleotide specificities. *A*) BsMazF active site residues Gly-22, Arg-24 and Thr-48 interact with its mRNA substrate (PDB code 4MDX) (left panel). The BsMazF upstream and downstream RNA binding regions with proposed key residues are indicated. *B*) The MazF-mt6 active site residues Asp-10, Arg-13 and Thr-36 with a modeled RNA bound (outline). The equivalent MazF-mt6 upstream and downstream RNA binding regions with potentially key residues are highlighted.

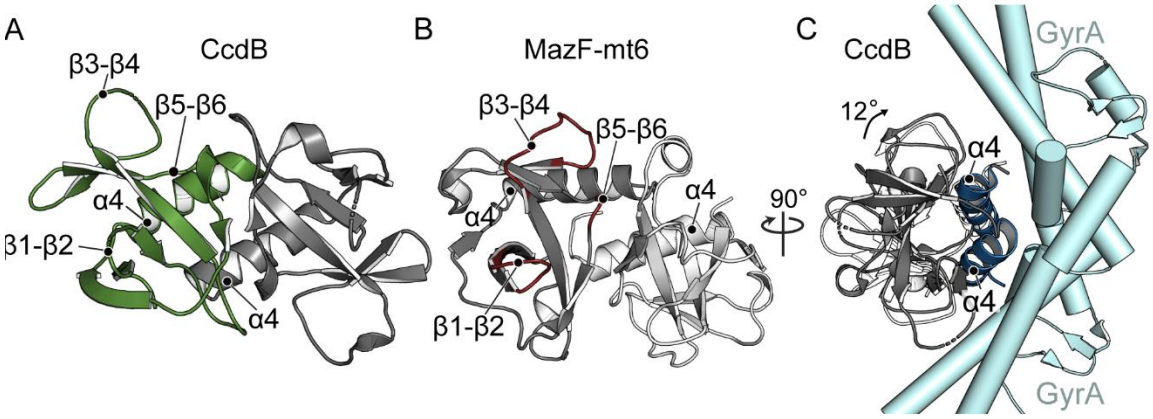

Figure 3.7 Structural comparison of MazF-mt6 and CcdB toxin. *A*) *E. coli* CcdB toxin similarly forms a homodimer toxin (one monomer shown in green with the other shown in gray) with important secondary structural regions labeled (PDB code 1X75). *B*) Two homodimers of MazF-mt6 are shown (one monomer shown in light gray with the other shown in dark gray) with the putative RNA binding regions colored in red. *C*) A 90° rotated view of panel A of CcdB bound to GyrA (cyan) illustrating how CcdB α4 from both monomers engages its substrate, which is located on the opposite side of the RNA binding domain of MazF-mt6. To bind GyrA, CcdB bends 12° (shown in dark gray). CcdB toxin in the absence of GyrA is shown for comparison (white).

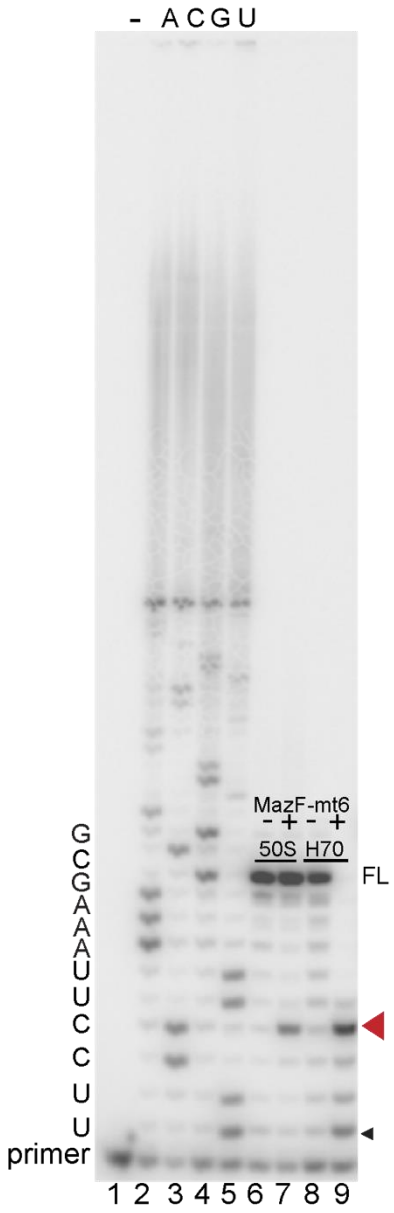

Figure 3.8 rRNA cleavage by MazF-mt6. The full length 8M Urea/PAGE gel as shown in Figure 1*C*. Sequencing reactions are located in lanes 2-5 and assays performed with 50S subunits and H70 are shown in lanes 6-7 and 8-9, respectively. The cleavage products were indicated by a reverse transcription stop. The MazF-mt6 full length RT stop (FL), the major cleavage site (red arrowhead) and a minor cleavage site (black arrowhead) are indicated.

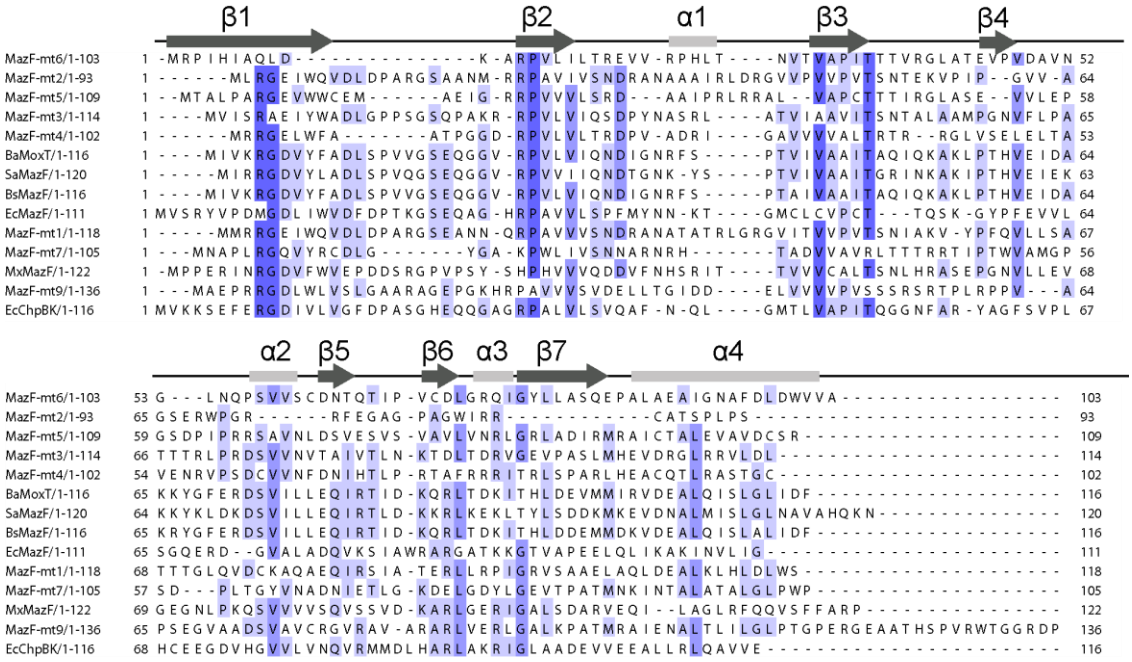

Figure 3.9 Sequence alignment of MazF-mt6 and other MazF homologs. Amino acid sequence alignments of *M. tuberculosis* MazFs (MazF-mt1, -mt2, -mt3, -mt4, -mt5, -mt7 and -mt9), *E. coli* MazF (EcMazF), *B. subtilis* MazF (BsMazF), *E. coli* ChpBK (EcChpBK), *S. aureus* MazF (SaMazF), *M. xanthus* MazF (MxMazF), and *B. anthraces* MoxT (BaMoxT) aligned to MazF-mt6 using the program T-coffee (47). Sequences were sorted based on their pairwise identity to MazF-mt6 and visualized in Jalview (48). Conserved residues are color coded based on the following identities: 41-60% (light purple), 61-80% (purple), and >80% (dark purple). Secondary structural elements from the MazF-mt6 structure are also shown.

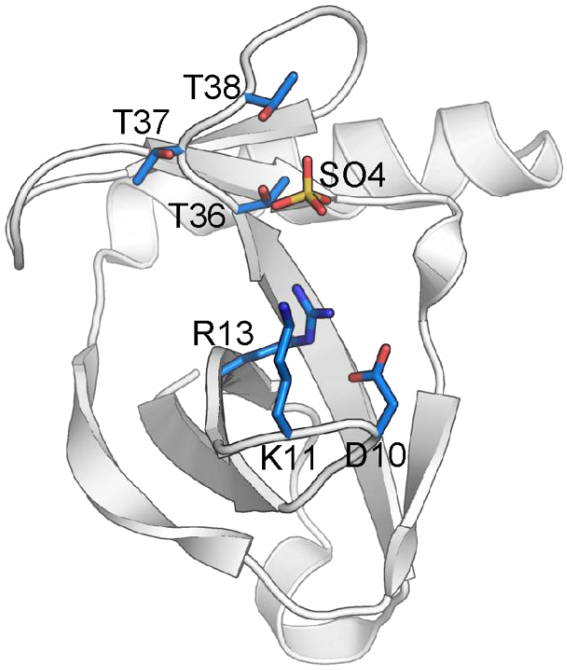

Figure 3.10 Location of a sulfate ion in the putative active site of MazF-mt6. The sulfate ion is surrounded by putative active site residues Asp-10, Lys-11, Arg-13, Thr-36, Thr-37, and Thr-38.

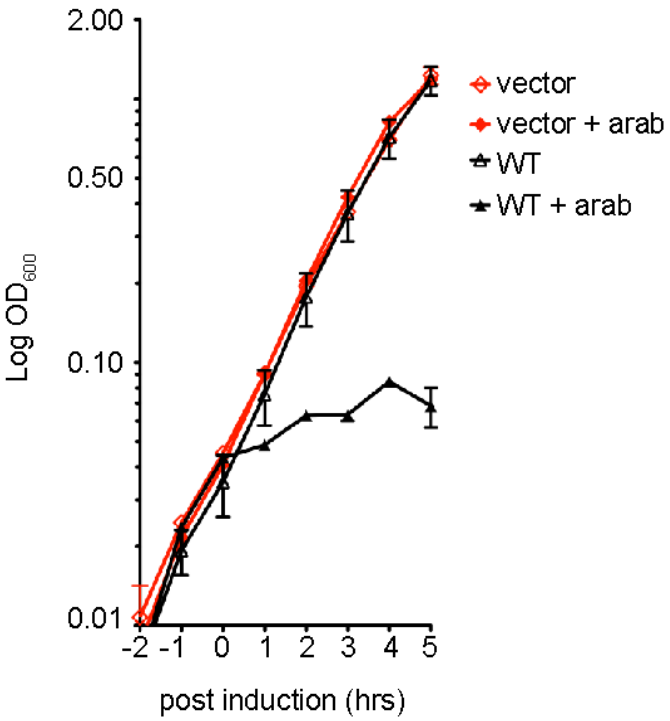

Figure 3.11 Bacterial growth assays. *E. coli* BW25113 growth monitored over 5 hrs post induction of either pBAD33 vector or pBAD33 WT MazF-mt6. Cells were grown in M9 media supplemented with 0.2% glycerol and 0.2% casamino acids and induction was initiated by the addition of 0.2% arabinose after cells reached an optical density of 0.2 at 600 nm. Error bars represent standard error of means from three independent experiments.

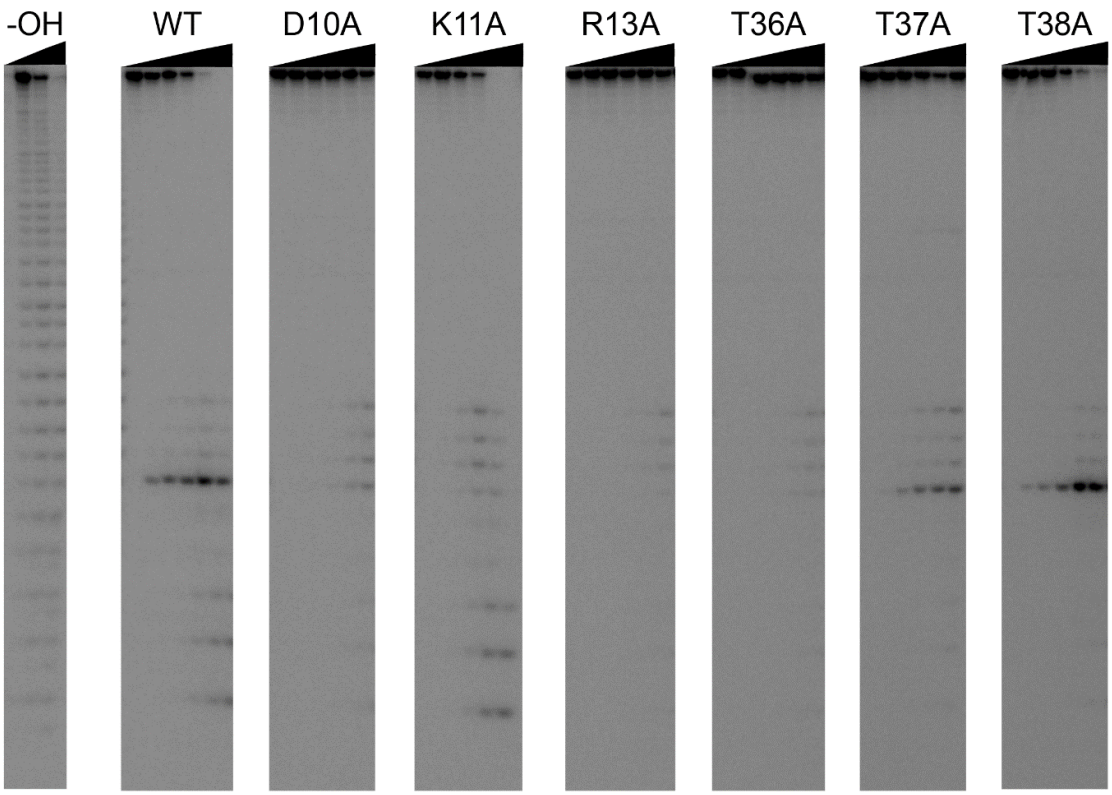

Figure 3.12 *In vitro* H70 cleavage by MazF-mt6 variants. Single turnover experiments monitoring 23S rRNA H70 cleavage by wild-type (WT) MazF-mt6 and MazF-mt6 variants. Samples were removed at time points of 0, 1, 3, 10, 25 and 45 min, the reactions were stopped by the addition of loading dye, and products were analyzed on an 8M Urea/PAGE sequencing gel.

Table 3.1 Data collection and refinement statistics.

	MazF-mt6
Data collection	
Space group	P $6_4$22
Wavelength (Å)	0.97920
Cell dimensions	
a, b, c (Å)	105.2, 105.2, 144.4
α, β, γ (°)	90, 90, 120
Resolution (Å)	77.0-2.70
$R_{p.i.m}$ (%)	3.37 (53.1)*
I/σI	12 (1.4)
Completeness (%)	95 (99)
Redundancy	6.3 (6.5)
CC1/2 (%)	99.8 (66.2)
Refinement	
Reflections	13309
R_{work}/R_{free} (%)	23.3/25.2
No. atoms	
Macromolecule	1540
Ligand/ion	20
Water	5
B-factors (Å^2)	
Overall	84.9
Macromolecule	84.3
Ligand/ion	132
R.m.s deviations	
Bond lengths (Å)	0.005
Bond angles (°)	1.12

*Highest resolution shell is shown in parenthesis

Table 3.2 Effects MazF-mt6 variants have on mRNA cleavage.

MazF-mt6 variant	K_{obs}
Wild-type	$3.5 \times 10^{-3} \pm 2 \times 10^{-4}\ s^{-1}$
D10A	$2.53 \times 10^{-4} \pm 2 \times 10^{-5}\ s^{-1}$
K11A	$8.3 \times 10^{-4} \pm 1.0 \times 10^{-4}\ s^{-1}$
R13A	NA
T36A	NA
T37A	$2.9 \times 10^{-4} \pm 2 \times 10^{-5}\ s^{-1}$
T38A	$1.03 \times 10^{-3} \pm 1.0 \times 10^{-4}\ s^{-1}$

Table 3.3 Oligos for the construction of MazF-mt6 variants.

Primer	Sequence (5'-3')
MazF-mt6 E10A fw	CAG CTT GCC AAG GCT AGA CCC GTC CTG ATC CTT ACC C
MazF-mt6 E10A rv	GCC TTG GCA AGC TGT GCG ATG TGG ATA GGT CGC ATA TG
MazF-mt6 K11A fw	CTT GAC GCT AGA CCC GTC CTG ATC CTT ACC CG
MazF-mt6 K11A rv	CTA GCC GCG TCA AGC TGT GCG ATG TGG ATA GGT CGC
MazF-mt6 R13A fw	CAA GGC TGC ACC CGT CCT GAT CCT TAC CCG CGA GGT CG
MazF-mt6 R13A rv	GGA CGG GTG CAG CCT TGT CAA GCT GTG CGA TGT GGA TAG G
MazF-mt6 T36A fw	GTC GCC CCG ATC GCG ACA ACC GTG C
MazF-mt6 T36A rv	GCA CGG TTG TCG CGA TCG GGG CGA C
MazF-mt6 T37A fw	GCC CCG ATC ACG GCA ACC GTG CGT G
MazF-mt6 T37A rv	CAC GCA CGG TTG CCG TGA TCG GGG C
MazF-mt6 T38A fw	TCG CCC CGA TCA CGA CAG CCG TGC GTG
MazF-mt6 T38A rv	CAC GCA CGG CTG TCG TGA TCG GGG CGA
pCOLD-R	GGC AGG GAT CTT AGA TTC TG
pBAD-Reverse	GAT TTA ATC TGT ATC AGG

REFERENCES

1. E. Maisonneuve, M. Castro-Camargo, K. Gerdes, (p)ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity. *Cell* **154**, 1140-1150 (2013).
2. Y. Yamaguchi, J. H. Park, M. Inouye, Toxin-antitoxin systems in bacteria and archaea. *Annu Rev Genet* **45**, 61-79 (2011).
3. D. Lobato-Marquez, R. Diaz-Orejas, F. Garcia-Del Portillo, Toxin-antitoxins and bacterial virulence. *FEMS Microbiol Rev*, (2016).
4. K. Gerdes, E. Maisonneuve, Bacterial persistence and toxin-antitoxin loci. *Annu Rev Microbiol* **66**, 103-123 (2012).
5. D. P. Pandey, K. Gerdes, Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res* **33**, 966-976 (2005).
6. K. Pedersen *et al.*, The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. *Cell* **112**, 131-140 (2003).
7. J. M. Hurley, N. A. Woychik, Bacterial toxin HigB associates with ribosomes and mediates translation-dependent mRNA cleavage at A-rich sites. *J Biol Chem* **284**, 18605-18613 (2009).
8. M. H. Prysak *et al.*, Bacterial toxin YafQ is an endoribonuclease that associates with the ribosome and blocks translation elongation through sequence-specific and frame-dependent mRNA cleavage. *Mol Microbiol* **71**, 1071-1087 (2009).
9. E. Aizenman, H. Engelberg-Kulka, G. Glaser, An Escherichia coli chromosomal "addiction module" regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death. *Proc Natl Acad Sci U S A* **93**, 6059-6063 (1996).
10. R. Hazan, B. Sat, H. Engelberg-Kulka, Escherichia coli mazEF-mediated cell death is triggered by various stressful conditions. *J Bacteriol* **186**, 3663-3669 (2004).
11. L. Zhu *et al.*, Characterization of mRNA interferases from Mycobacterium tuberculosis. *J Biol Chem* **281**, 18638-18643 (2006).
12. J. M. Schifano *et al.*, An RNA-seq method for defining endoribonuclease cleavage specificity identifies dual rRNA substrates for toxin MazF-mt3. *Nature communications* **5**, 3538 (2014).
13. J. M. Schifano *et al.*, Mycobacterial toxin MazF-mt6 inhibits translation through cleavage of 23S rRNA at the ribosomal A site. *Proc Natl Acad Sci U S A* **110**, 8501-8506 (2013).
14. J. M. Schifano *et al.*, tRNA is a new target for cleavage by a MazF toxin. *Nucleic Acids Res* **44**, 1256-1270 (2016).
15. K. S. Makarova, Y. I. Wolf, E. V. Koonin, Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol Direct* **4**, 19 (2009).
16. M. M. Yusupov *et al.*, Crystal structure of the ribosome at 5.5 A resolution. *Science* **292**, 883-896 (2001).
17. A. Weixlbaumer *et al.*, Crystal structure of the ribosome recycling factor bound to the ribosome. *Nat Struct Mol Biol* **14**, 733-737 (2007).
18. J. E. Jackman, E. M. Phizicky, tRNAHis guanylyltransferase catalyzes a 3'-5' polymerization reaction that is distinct from G-1 addition. *Proc Natl Acad Sci U S A* **103**, 8640-8645 (2006).
19. D. K. Simanshu, Y. Yamaguchi, J. H. Park, M. Inouye, D. J. Patel, Structural basis of mRNA recognition and cleavage by toxin MazF and its regulation by antitoxin MazE in Bacillus subtilis. *Mol Cell* **52**, 447-458 (2013).
20. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica* **66**, 213-221 (2010).

21. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta crystallographica* **66**, 486-501 (2010).
22. G. M. Cook *et al.*, Ribonucleases in bacterial toxin-antitoxin systems. *Biochim Biophys Acta* **1829**, 523-531 (2013).
23. R. Loris *et al.*, Crystal structure of CcdB, a topoisomerase poison from E. coli. *Journal of molecular biology* **285**, 1667-1677 (1999).
24. D. Hargreaves *et al.*, Structural and functional analysis of the kid toxin protein from E. coli plasmid R1. *Structure* **10**, 1425-1433 (2002).
25. J. Zhang, Y. Zhang, L. Zhu, M. Suzuki, M. Inouye, Interference of mRNA function by sequence-specific endoribonuclease PemK. *J Biol Chem* **279**, 20678-20684 (2004).
26. K. Kamada, F. Hanaoka, S. K. Burley, Crystal structure of the MazE/MazF complex: molecular bases of antidote-toxin recognition. *Mol Cell* **11**, 875-884 (2003).
27. J. H. Park, Y. Yamaguchi, M. Inouye, Bacillus subtilis MazF-bs (EndoA) is a UACAU-specific mRNA interferase. *FEBS letters* **585**, 2526-2532 (2011).
28. L. Holm, P. Rosenstrom, Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-549 (2010).
29. V. Zorzini *et al.*, Substrate Recognition and Activity Regulation of the Escherichia coli mRNA Endonuclease MazF. *J Biol Chem* **291**, 10950-10960 (2016).
30. J. S. Han *et al.*, Characterization of a chromosomal toxin-antitoxin, Rv1102c-Rv1103c system in Mycobacterium tuberculosis. *Biochem Biophys Res Commun* **400**, 293-298 (2010).
31. Y. Zhang *et al.*, MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli. *Mol Cell* **12**, 913-923 (2003).
32. V. Hauryliuk, G. C. Atkinson, K. S. Murakami, T. Tenson, K. Gerdes, Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat Rev Microbiol* **13**, 298-309 (2015).
33. K. M. Dalton, S. Crosson, A conserved mode of protein recognition and binding in a ParD-ParE toxin-antitoxin complex. *Biochemistry* **49**, 2205-2215 (2010).
34. C. D. Aakre *et al.*, Evolving new protein-protein interaction specificity through promiscuous intermediates. *Cell* **163**, 594-606 (2015).
35. Y. V. Guillen Schlippe, L. Hedstrom, A twisted base? The role of arginine in enzyme-catalyzed proton abstractions. *Arch Biochem Biophys* **433**, 266-278 (2005).
36. M. H. Dao-Thi *et al.*, Molecular basis of gyrase poisoning by the addiction toxin CcdB. *Journal of molecular biology* **348**, 1091-1102 (2005).
37. V. Zorzini *et al.*, Structural and biophysical characterization of Staphylococcus aureus SaMazF shows conservation of functional dynamics. *Nucleic Acids Res* **42**, 6709-6725 (2014).
38. M. H. Dao-Thi *et al.*, Crystallization of CcdB in complex with a GyrA fragment. *Acta crystallographica* **60**, 1132-1134 (2004).
39. W. Kabsch, Xds. *Acta crystallographica* **66**, 125-132 (2010).
40. Schrodinger, LLC. (2010).
41. M. A. Witek, G. L. Conn, Functional dichotomy in the 16S rRNA (m1A1408) methyltransferase family and control of catalytic activity via a novel tryptophan mediated loop reorganization. *Nucleic Acids Res* **44**, 342-353 (2016).
42. A. Sali, L. Potterton, F. Yuan, H. van Vlijmen, M. Karplus, Evaluation of comparative protein modeling by MODELLER. *Proteins* **23**, 318-326 (1995).

43. J. Soding, A. Biegert, A. N. Lupas, The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**, W244-248 (2005).
44. I. Bertini *et al.*, Accurate solution structures of proteins from X-ray data and a minimal set of NMR data: calmodulin-peptide complexes as examples. *J Am Chem Soc* **131**, 5134-5144 (2009).
45. E. R. Schreiter, S. C. Wang, D. B. Zamble, C. L. Drennan, NikR-operator complex structure and the mechanism of repressor activation by metal ions. *Proc Natl Acad Sci U S A* **103**, 13676-13681 (2006).
46. F. X. Gomis-Ruth *et al.*, The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator. *EMBO J* **17**, 7404-7415 (1998).
47. C. Notredame, D. G. Higgins, J. Heringa, T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology* **302**, 205-217 (2000).
48. A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton, Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191 (2009).

Chapter 4: Purification and crystal screening of two tRNA cleaving toxins.

Eric D. Hoffer, Regan Esposito, Christine M. Dunham

Toxin-antitoxin (TA) complexes are two component regulatory systems found in bacteria that inhibit growth and have important roles in several cellular pathways including plasmid maintenance and antibiotic tolerance. Toxin proteins promote a non-growth phenotype in numerous ways, but the most common is degradation of cellular RNAs. These toxins are classified as either ribosome dependent or independent. However, it is unclear how these endonucleases recognize and cleave specific types of RNA. type II toxin family members VapC and MazF-mt9 both cleave transfer RNAs (tRNA) despite adopting different structural folds. Here I show preliminary results of the purification and activity of an inactive variant of *Shigella flexneri* VapC ($VapC^{D7A}$), *Mycobacterium tuberculosis* MazF-mt9 and their tRNA substrates.

Author contributions: E.D.H. and C.M.D. designed the experiments; E.D.H. supervised rotation student R.E. who purified $tRNA^{Lys}$ and $tRNA^{Pro}$. E.D.H performed the remaining experiments.

INTRODUCTION

Toxin-antitoxin (TA) complexes are two component regulatory systems that encode a protein antitoxin that prevents its toxin partner from inhibiting cell growth until specific environmental conditions are encountered. TA systems have been associated with several critical roles in bacteria including antibiotic tolerance, host-pathogen virulence responses, post-segregational killing and adaptive responses to environmental stress (1, 2). TA systems consist of six different subfamilies classified by how the antitoxin component regulates toxin activity (3). The toxin gene product is always a protein that inhibits growth by perturbing essential cellular functions like DNA replication or protein synthesis (2, 4). The antitoxin is either a non-coding RNA or a protein that inhibits toxin activity by directly binding the toxin protein (type II, III, VI), preventing toxin protein expression (type I, V) or competing with the toxin for substrate binding (type IV) (3). The best studied class of TA system is the type II, which both toxin and antitoxin are proteinaceous.

The type II TA system toxins are mostly endoribonucleases that inhibit protein translation to retard cell growth and can be classified as ribosome-dependent or independent. Ribosome-dependent toxins RelE and HigB only cleave mRNA when bound to the A site of the ribosome. In contrast, ribosome-independent toxins like MazF and VapC are fully functional in the absence of the ribosome. The ribosome-dependent toxins have been extensively studied and our lab has provided in-depth mechanistic insights into how they interact with A site of the ribosome and how they cleave mRNA (5-7). Ribosome-independent toxins like MazF have been well studied (8-11). In contrast, little is known about the most abundant type II TA VapBC or MazF toxins from *Mycobacterium tuberculosis* (*Mtb*) that also cleave tRNAs for example, MazF-mt9.

VapC toxins are endoribonucleases are members of the PilT N-terminus (PIN) family (12). The PIN domain is ~140 amino acids in length containing a conserved motif

of four metal chelating acidic residues essential for ribonuclease activity (13-15). In the gram-negative pathogens *Shigella flexneri* and *Salmonella enteric* serovar Typhimurim LT2 VapC cleaves the anticodon stem-loop of initiator $tRNA^{fMet}$ (14). VapC toxins from *Mtb* have been shown to cleave elongator tRNAs and the sarcin-ricin loop of the bacterial ribosome (15-17). Additionally, recent studies on MazF-mt9 have identified tRNA as its primary target (18). Since MazF and VapC toxins share no sequence or structural homology, it remains elusive how MazFs have evolved to recognize more complex RNA targets.

Here, my goals were to solve the X-ray crystal structures of the *S. flexneri* $VapC^{D7A}$-$tRNA^{fMet}$ complex or the *Mtb* MazF-mt9- tRNA complex. I purified $VapC^{D7A}$ and MazF-mt9 to >95% homogeneity and *in vitro* transcribed $tRNA^{fMet}$, $tRNA^{Lys}$ and $tRNA^{Pro}$. I demonstrated that $VapC^{D7A}$ binds $tRNA^{fMet}$ using an electrophoretic mobility shift assay (EMSA) confirming $VapC^{D7A}$ binds to $tRNA^{fMet}$. I set up high-throughput crystallization trials but did not observe a crystal hit with the conditions tested. Future experiments will be aimed at expanding the crystallization screen conditions and concentrations of the $VapC^{D7A}$-$tRNA^{fMet}$ and *Mtb* MazF-mt9- tRNA complexes.

MATERIALS AND METHODS

The *Shigella flexneri vapC* gene was designed and purchased from GeneArt. A D7A mutation was introduced into VapC using primers EH074_sflex_vc (5'-CATATGCTG AAATTTATG CTG GCTACCAACATCTGCATCTTTACC-3') and EH075_sflex_vc_anti (5'-GGTAAAGATGCAGATGTTGGTAGCCAGCATAAATTTCAGCATATG-3'). The $vapC^{D7A}$ gene was digested with BamHI and HindIII and ligated into pET-HT, a modified version of pET41 that contains an N-terminal 6xHis tag. The pUC18-$tRNA^{fMet}$ was generated by a previous lab member (unpublished). A new pUC18-$tRNA^{fMet}$ plasmid containing a 3' BsaI site was generated using primers EH088_BsaI_ pUC18fMet_fwd (5'-

GCCCCCGCCACCATGAGACCGGCGCGG-3') and EH089_BsaI_pUC18fMet_rev (5'-CCGCGCCGGTCTCATGGTGGCGGGGGC-3'). The pUC18-tRNA^{Pro} and pUC18-tRNA^{Lys} vectors were generated by subcloning synthesized genes (IDT) into EcoRI and BamHI sites.

Expression and purification of VapC^{D7A}

Expression and purification of *S. flexneri* VapC^{D7A} was performed as described previously with a few modifications (19). *Escherichia coli* C41 (DE3) cells (Lucigen) were transformed with pET-HT-VapC^{D7A} and overnight cultures were grown in Lysogeny Broth (LB) medium containing 100 mg mL^{-1} ampicillin at 37°C and 250 rpm. The overnights were diluted 1:100 in fresh 1 L cultures of LB containing 100 mg mL^{-1} ampicillin at 37°C and 250 rpm. Once the OD_{600} for the culture reached 0.5, the VapC^{D7A} expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and incubated overnight at 25°C and 120 rpm. Cells were harvested at 3,500 xg for 30 min at 4°C and the cell pellet was resuspended in lysis buffer (50 mM Tris-Cl pH 8.0, 500 mM KCl, 5 mM $MgCl_2$, 10 mM imidazole, 5 mM β-mercaptoethanol (β-Me), 0.1 mM phenylmethylsulfonyl fluoride (PMSF) and 0.1 mM benzamidine (BZA)). Cells were lysed using a high-pressure continuous flow C5 Emulsiflex homogenizer by applying three times at a chamber pressure ≥ 10,000 psi. Lysed cells were clarified by spinning at 30,000 xg for 45 min at 4°C in a JA-20 rotor. The supernatant was filtered and applied to a 5 mL immobilized metal affinity column (IMAC) equilibrated in buffer A (50 mM Tris-Cl pH 8.0, 500 mM KCl, 5 mM $MgCl_2$, 35 mM imidazole, 5 mM β-Me). The protein was eluted with a linear gradient of buffer B (50 mM Tris-Cl pH 8.0, 500 mM KCl, 5 mM $MgCl_2$, 500 mM imidazole, 5 mM β-Me) over the course of 20 column volumes (cv). Fractions containing VapC^{D7A} were combined and concentrated before running on a Superdex S200 16/60 size exclusion column (GE Healthcare) equilibrated in buffer C (25 mM MES-KOH pH 6.0, 50 mM KCl, 5 mM $MgCl_2$,

5 mM β-Me). The protein was greater than 90% homogeneous as assessed by SDS-PAGE. The protein was concentrated to 5 mg/mL and flash frozen before storing at -80°C.

Expression and purification of MazF-mt9

Overnight cultures *of E. coli* C41(DE3) cells (Lucigen) transformed with pCold-$(His)_6$-Trigger Factor (TF)-MazF-mt9 (kind gift of the Woychik Lab, Rutgers University) were diluted 1:100 in LB medium supplemented with 100 µg/mL ampicillin. Cultures were grown to an OD_{600} of 1.0 at 37˚C, transferred to 15˚C for 30 min, and protein expression was induced with 0.5 mM IPTG and grown for 24 hrs at 15°C. Cells were harvested and resuspended in lysis buffer (40 mM Tris-Cl pH 8.0, 500 mM NaCl, 10% (w/v) glycerol, 0.1% (w/w) Triton X-100, 5 mM β-Me, 20 mM imidazole, 0.1 mM BZA and 0.1 mM PMSF). Cells were lysed using a high-pressure continuous flow C5 Emulsiflex homogenizer (Avastin) by applying three times at a chamber pressure ≥ 10,000 psi. Lysed cells were clarified by spinning at 30,000 xg for 45 min at 4°C in a JA-20 rotor. Cell debris was clarified by centrifugation at 35,000 xg for 50 min and applied to a 5 mL IMAC. The $(His)_6$-TF-MazF-mt9 fusion protein was eluted with a linear gradient of lysis buffer without Triton X-100 or protease inhibitors but containing 500 mM imidazole.

Affinity purified $(His)_6$-TF-MazF-mt9 was dialyzed overnight into protease cleavage buffer (20 mM Tris-Cl pH 8.0, 150 mM NaCl, 2.5 mM $CaCl_2$) and either 1 unit of Factor Xa (Sigma-Aldrich) or Bovine Thrombin (Calbiochem) per 200 µg protein was added at 10°C and incubated overnight. The cleaved $(His)_6$-TF-MazF was reapplied to the IMAC and MazF-mt9 was collected from the flow through fraction. The MazF-mt9 was further purified and buffer exchanged on a size exclusion Superdex 200 column (GE Healthcare) into 20 mM Tris-Cl pH 7.5, 250 mM KCl, 10% (w/v) glycerol and 5 mM β-Me. MazF-mt9 was determined to be ~95% pure by 15% SDS-PAGE and the fractions that contained only MazF-mt9 were flash frozen in liquid nitrogen then stored at -80˚C.

In vitro transcription and purification of tRNA toxin substrates

E. coli DH5α cells were transformed with pUC18-tRNA^{fMet} , pUC18-tRNA^{Lys} or pUC18-tRNA^{Pro}, grown to an OD_{600} of 2.0 and large-scale plasmid preps were performed as previously described (20). The purified DNA template was digested overnight using *Bsa*I restriction endonuclease (NEB). The digested template was purified on a 1% agarose gel using a gel extraction kit (Qiagen). IVT reactions were performed at 37°C for 4 hrs with a final concentration of 200 mM HEPES-KOH pH 7.5, 28 mM $MgCl_2$, 2 mM spermidine, 40 mM DTT, 6 mM rNTPs, 50 ng/µL linearized template DNA and T7 RNA polymerase (Dunham lab purified). Large scale reactions volumes were 500 µL each. After 4 hrs, the reactions were spun at 3,000 *x g* for 10 min to pellet the magnesium pyrophosphate precipitate and all supernatants were combined in a fresh tube. The reaction was halted by adding ~50 mM EDTA pH 8.0 to each tube at room temperature for 1 min.

The *in vitro* transcription reaction was phenol:chloroform:isoamyl alcohol (25:24:1) extracted and back extracted with 100% chloroform twice. The RNA was desalted using a 10DG desalting column (Biorad). The RNA was concentrated by ethanol precipitation by adding a final concentration of 0.3M sodium acetate then applying 2 volumes of ice cold 100% ethanol. 2x formamide dye was added to RNA before purification on a large 10% polyacrylamide- 50% urea sequencing gel. The gel was run at 200 V constant for 9 hrs, the tRNA was visualized by UV shadowing and cut out of the gel and extracted overnight at 4°C in RNA elution buffer (300 mM sodium acetate pH 5.2, 0.1% SDS, 1 mM EDTA). The eluted RNA was ethanol precipitated and the dried pellet was resuspended in annealing buffer (20 mM HEPES-KOH pH 7.5, 1 mM EDTA pH 8.0). The tRNA was heated for 3 min at 80°C then cooled at room temp and buffer (20 mM HEPES-KOH pH 7.5, 20 mM $MgCl_2$) was added before incubated at 37°C for 15 min. The final concentration of tRNA was determined using absorbance at 260 nm and an estimated molecular weight of 25 kDa.

Electrophoretic mobility shift assay (EMSA)

Assays were performed as described previously (21) but with slight modifications. Purified *E. coli* $tRNA^{fMet}$ (Chemblock, Russia), was labeled with $ATP_{\gamma}P^{32}$ using T4 RNA ligase (NEB). Protein at final concentrations ranging from 0 to 20 µM was incubated with 30 nM $^{32}P\text{-}tRNA^{fMet}$ for 20 mins at room temp in 20 mM HEPES-KOH pH 7.5, 50 mM KCl, 10% glycerol, 10 mM $MgCl_2$, 1 mM DTT and 0.01 mg/mL BSA. Free and protein-bound $tRNA^{fMet}$ were resolved on a native 8% PAGE gel prepared with 0.5x Tris-Borate pH 8-EDTA (TE) buffer including 10% glycerol (v/v). The gel was run at 10°C for 2 hrs before the gel was dried and exposed to a phosphor screen (GE Healthcare) and imaged on a Typhoon Trio (GE Healthcare).

Co-crystallization trials

Initial screening for crystallization conditions was carried out using a Phoenix 96-well screen dispenser (Art Robbins Instruments) by mixing 200 nl of 5 mg/mL protein/tRNA complex solution with 200 nl reservoir solution (using Crystal Screen 1, Crystal Screen 2, PEG/Ion, Index and Natrix commercial screens; Hampton Research) in sitting drop, vapor-diffusion experiments in 96-well plates (Hampton Research) at 4°C, 10°C and 20°C.

RESULTS AND DISCUSSION

Successful complex formation of $VapC^{D7A}$-$tRNA^{fMet}$

$VapC^{D7A}$ was purified according to an established protocol (19) and I was able to purify 5.0 mg/L of culture. **(Fig. 4.1A)**. One difference as that in the original protocol, $VapC^{D7A}$ was ammonium sulfate precipitated after the affinity purification step, but I was unable to replicate this. Next I *in vitro* transcribed and purified $tRNA^{fMet}$ by denaturing urea PAGE gel extraction **(Fig. 4.3C)**. I tested if purified $VapC^{D7A}$ binds $tRNA^{fMet}$ using native

EMSAs. $tRNA^{fMet}$ ran as three distinct species **(Fig. 4.3F)**. The highest MW tRNA band appears to bind $VapC^{D7A}$ **(Fig. 4.4)**. Notably, I did not observe any cleavage of tRNA by $VapC^{D7A}$ in this assay and I moved forward with co-crystallization trials.

I used two strategies when setting up the VapC ^{D7A}-tRNA crystallization screens. The first was to sequentially add the protein and tRNA into the wells, and the second was forming the $VapC^{D7A}$-tRNA complex before setting up the trays. I tested 5 mg/mL complex or 2.5 mg/mL complex using a variety of screens at three temperatures of 4°C, 10°C and 20°C. All screens were set up in a 3-well vapor diffusion screen with the first well containing the VapC ^{D7A}-tRNA complex and the other two were either protein alone or RNA alone to control for single component crystals. A majority of the drops precipitated when the components were added to the wells sequentially, but the pre-complexed drops had about a 50% precipitation rate at all temperatures at 5 mg/mL, which means that pre-complexing is likely the optimal way to setup trays at that concentration. I identified a few conditions that produced crystals, but the single component control wells had the same crystals for either VapC or tRNA in all cases.

Purification of MazF-mt9 toxin and $tRNA^{Pro}$ and $tRNA^{Lys}$

$(His)_6$TF-MazF-mt6 was purified as previously described (22). Next I needed to remove the fused TF from MazF-mt9. Previous work with the pCOLD-TF-MazF-mt9 used the endopeptidase Factor Xa {Schifano, 2013, 8501-6}, so I performed a test proteolysis over the course of 22 hrs **(Fig 2.2A)**. We noticed that the proteolysis by Factor Xa produced a two protein bands on the gel. To make sure the different sized protein bands were not caused by Factor Xa, I also performed the assay with thrombin **(Fig 2.2B)**. The thrombin proteolysis also produced two bands, which could either mean that the MazF-mt9 dimer is formed from two different length monomers or both thrombin and Factor Xa cleaves two sites on the protein. Since the thrombin cleavage products were stable after

22 hours when compared to factor Xa **(Fig. 4.2A, B)**, I used thrombin to remove the tag. I separated the $(His)_6$TF from MazF-mt9 by reapplying the reaction products over a nickel column and analyzed the fractions. Any remaining TF was removed by size exclusion chromatography **(Fig. 4.2C, D)**. We only collected fractions that did not contain a visible band form trigger factor and concentrated to 5 mg/mL **(Fig. 4.2C, D)**.

In order to generate enough tRNA for our crystallization trails, I *in vitro* transcription (IVT) the tRNA. I initially used a pUC18-$tRNA^{fMet}$ plasmid, but the restriction enzyme located at the 3' end of the gene used to linearize the template also cut in the middle of the gene. To get around this issue, I inserted a new BsaI restriction site at the 3' of the tRNA gene called the plasmid pUC18-$tRNA^{fMet}$-BsaI. Unlike type II restriction enzymes that cleave DNA within their recognition sequence, BsaI is a type II restriction enzyme that recognizes a unique six nucleotide sequence, but cleaves downstream. I determined that a 4 hour IVT reaction produced the same amount of full length product and had less runoff products than overnight incubation **(Fig. 4.3B)**. I incubated the largescale IVT reaction for 4 hours and then purified by gel extraction on a 50% urea-PAGE gel **(Fig. 4.3B)**. The tRNA was annealed and flash froze in liquid nitrogen and stored at -80°C. Based on my success with $tRNA^{fMet}$, I designed the substrate tRNA genes for MazF-mt9 using the same BsaI site downstream for $tRNA^{Pro}$ and $tRNA^{Lys}$ **(Fig. 4.3D-F)**.

I performed vapor diffusion crystallization trials of MazF-mt9 protein (5 mg/ml) at two temperatures with one well containing buffer only and another well containing MazF-mt9. I did not generate an inactive variant of MazF-mt9 so I did not test tRNA binding or set up crystallization trials. However, I have optimized the tRNA purification protocol for both $tRNA^{Pro}$ and $tRNA^{Lys}$.

CONCLUSIONS/FUTURE DIRECTIONS

TA systems are ubiquitous in bacteria and have been shown to increase virulence (1). Here, I attempted to determine the structural basis of action of two systems with their known substrates identified as tRNA. There has yet to be structure of tRNA endoribonuclease toxin from the VapC family nor the novel MazF-mt9 toxin, which hinders our ability to understand how toxins recognize such complex RNA.

Among the type II bacterial toxins, VapC is the least understood. VapC is a metal dependent, PIN containing ribonucleases that cleaves the anticodon stem-loop of tRNAs (13, 14). VapC is homologous to RNase H, which cleaves RNA found in DNA-RNA hybrids (23). VapC and RNase H share some similarities such as both requiring metal ions for activity and recognizing RNA helices, but in the absence of structural information, the mechanism of RNA recognition or cleavage is unclear (23). Although MazF and VapC are different toxin family members, they both recognize tRNAs and I hypothesize do so in different ways.

Here, I was able to replicate *S. flexneri* VapC^{D7A} protein expression and purification and substrate RNA and show they form a complex. The determination of crystallization conditions is a future goal. Another interesting question is whether VapC cleaves an RNA containing only the stem-loop of tRNA^{fMet} which may help in crystallization.

FIGURES

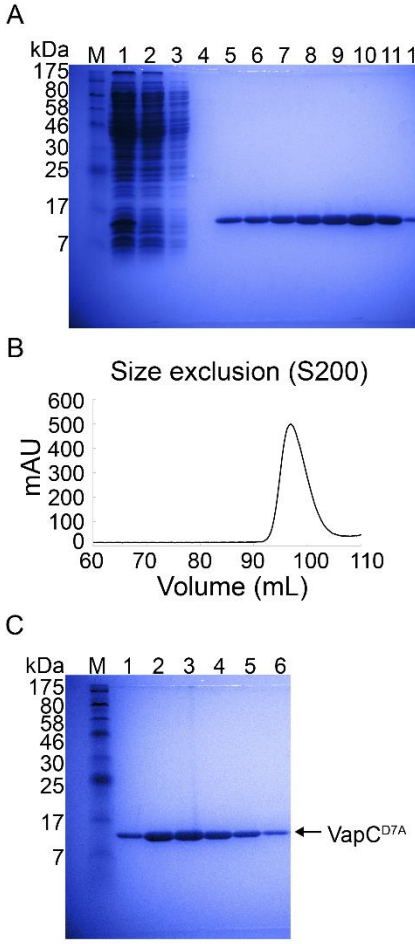

Figure 4.1 Purification of VapC^{D7A}. A) SDS-PAGE of VapC^{D7A} soluble lysate (lane 1) and flow through of the affinity column (lanes 2-4) and imidazole elution (lanes 5-12). B) Superdex S200 SEC chromatograph of VapC^{D7A}. C) 15% SDS-PAGE gel of VapC^{D7A} after Superdex S200.

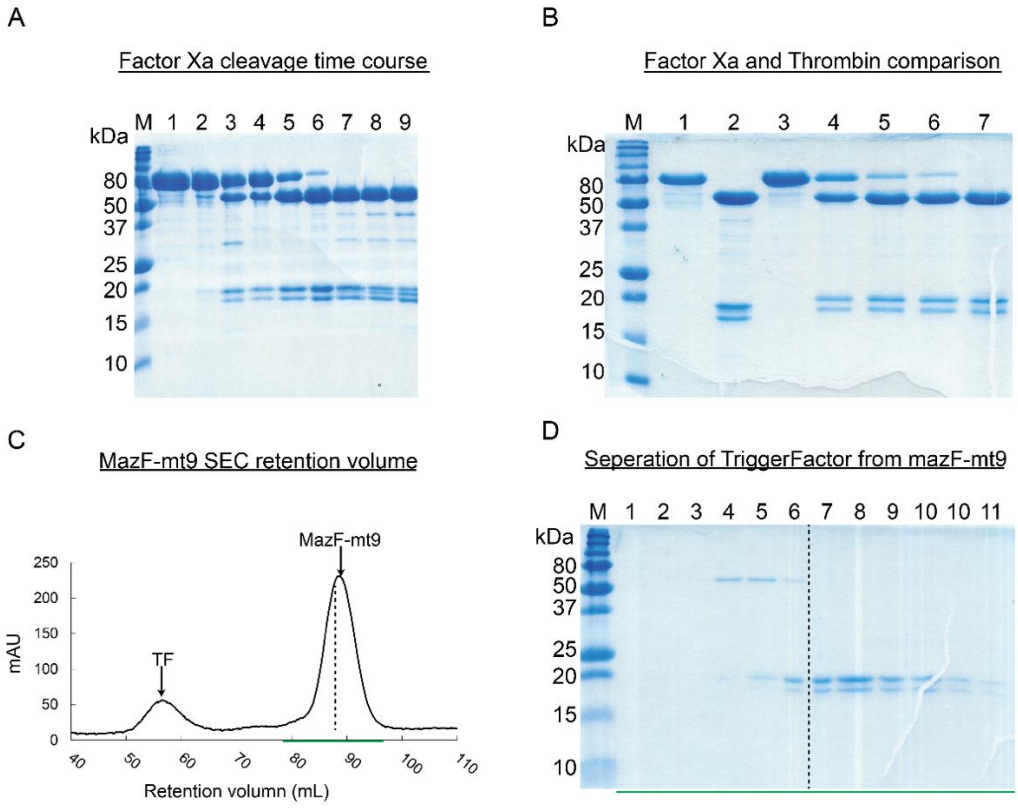

Fig 4.2 Purification of wild-type MazF-mt9. A) Purified $(His)_6$-TF-MazF-mt9 from IMAC (lane 1) was subjected to Factor Xa cleavage over a 22 hour time course. Lanes 2-6 range are 1-5 hours of incubation, respectively, while lanes 7-9 correspond to 16, 18 and 22 hours, respectively. B) Purified $(His)_6$-TF-MazF-mt9 from Nickel IMAC before Factor Xa incubation (lane 1) and after (lane 2) was compared to Thrombin cleave over the course of 22 hours. Timepoints were taken at 0, 1, 2, 4 and 22 hours (lanes 3-7) after addition of thrombin. C) Superdex S200 SEC chromatograph of cleaved $(His)_6$-TF-MazF-mt9 with TF and MazF-mt9 peaks labeled. The green line indicates which fractions were collected and the dotted line indicates where TF contamination in the prep ended. D) 15% SDS-PAGE of fractions collected in panel C. All fractions to the right of the dotted line were combined and concentrated for crystallization trials.

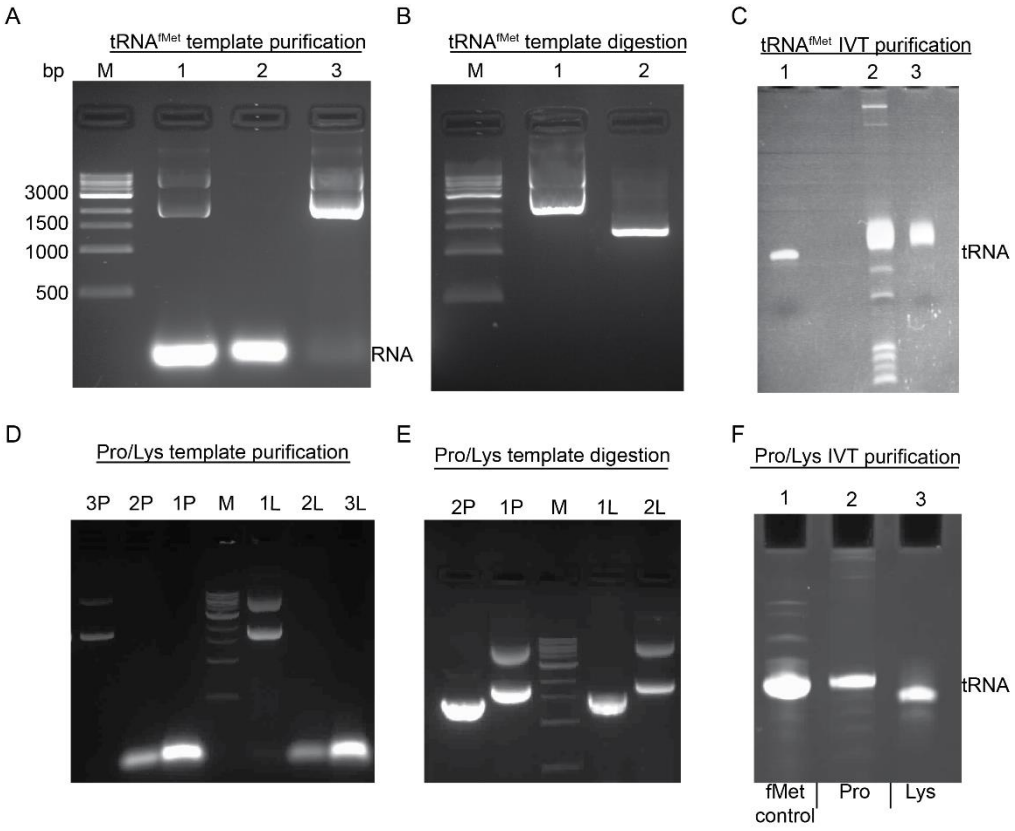

Figure 4.3 DNA template purification and IVT of tRNAs. A) 1% agarose gel of purified pUC18-tRNA^{fMet}-BsaI plasmid containing RNA contaminants (lane 1) were removed by a PEG 8000 precipitation step (lane 2) and resuspended in TE buffer (lane 3). B) 1% agarose gel of pUC18-tRNA^{fMet}-BsaI (lane 1) and overnight BsaI digested pUC18-tRNA^{fMet}. C) 50% urea-10% polyacrylamide gel of IVT reaction of tRNA^{fMet}. tRNA standard (lane 1) before gel purification (lane 2) and after gel purification (lane 3). D) Similar to panel A, the purified pUC18-tRNA^{Pro} and pUC18-tRNA^{Lys} plasmid containing RNA contaminants were removed by a PEG 8000 precipitation step (lane 1 and 2) and resuspended in TE buffer (lane 3). E) 1% agarose gel of pUC18-tRNA^{Pro} and pUC18-tRNA^{Lys} (lane 1) and overnight BsaI digested plasmid (lane 2). F) 50% urea-10% polyacrylamide gel of IVT reaction of tRNA^{Pro} (lane 2) and tRNA^{Lys} (lane 3). tRNA^{Pro} was not gel purified while the tRNA^{Lys} was.

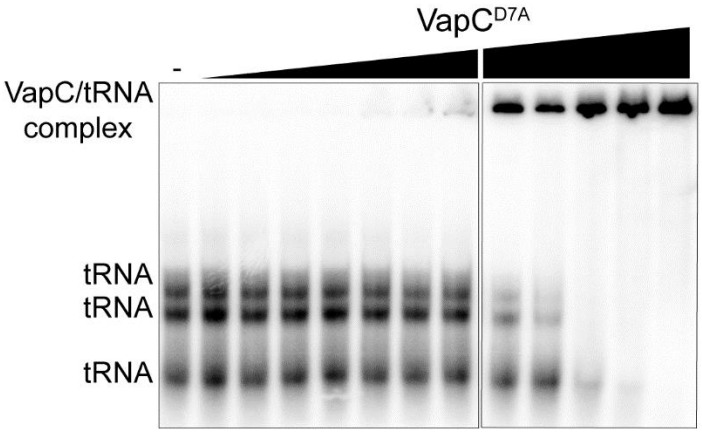

Figure 4.4 Vapc^{D7A} forms a complex with tRNA^{fMet}. 32P-tRNA^{fMet} was incubated with increasing amounts of VapC^{D7A} at room temp before running on a native 8% PAGE gel. tRNA^{fMet} ran as three distinct species on the gel (labeled on the left) and the VapC^{D7A}-tRNA^{fMet} complex is labeled at the top of the gel.

References:

1. E. Maisonneuve, M. Castro-Camargo, K. Gerdes, (p)ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity. *Cell* **154**, 1140-1150 (2013).
2. Y. Yamaguchi, J. H. Park, M. Inouye, Toxin-antitoxin systems in bacteria and archaea. *Annu Rev Genet* **45**, 61-79 (2011).
3. D. Lobato-Marquez, R. Diaz-Orejas, F. Garcia-Del Portillo, Toxin-antitoxins and bacterial virulence. *FEMS Microbiol Rev*, (2016).
4. K. Gerdes, E. Maisonneuve, Bacterial persistence and toxin-antitoxin loci. *Annu Rev Microbiol* **66**, 103-123 (2012).
5. M. A. Schureck, T. Maehigashi, S. J. Miles, J. Marquez, C. M. Dunham, mRNA bound to the 30S subunit is a HigB toxin substrate. *RNA* **22**, 1261-1270 (2016).
6. M. A. Schureck, A. Repack, S. J. Miles, J. Marquez, C. M. Dunham, Mechanism of endonuclease cleavage by the HigB toxin. *Nucleic Acids Res*, (2016).
7. C. Neubauer *et al.*, The structural basis for mRNA recognition and cleavage by the ribosome-dependent endonuclease RelE. *Cell* **139**, 1084-1095 (2009).
8. V. Zorzini *et al.*, Substrate Recognition and Activity Regulation of the Escherichia coli mRNA Endonuclease MazF. *J Biol Chem* **291**, 10950-10960 (2016).
9. V. Zorzini *et al.*, Structural and biophysical characterization of Staphylococcus aureus SaMazF shows conservation of functional dynamics. *Nucleic Acids Res* **42**, 6709-6725 (2014).
10. D. K. Simanshu, Y. Yamaguchi, J. H. Park, M. Inouye, D. J. Patel, Structural basis of mRNA recognition and cleavage by toxin MazF and its regulation by antitoxin MazE in Bacillus subtilis. *Mol Cell* **52**, 447-458 (2013).
11. K. Kamada, F. Hanaoka, S. K. Burley, Crystal structure of the MazE/MazF complex: molecular bases of antidote-toxin recognition. *Mol Cell* **11**, 875-884 (2003).
12. R. D. Bunker, J. L. McKenzie, E. N. Baker, V. L. Arcus, Crystal structure of PAE0151 from Pyrobaculum aerophilum, a PIN-domain (VapC) protein from a toxin-antitoxin operon. *Proteins* **72**, 510-518 (2008).
13. V. L. Arcus, K. Backbro, A. Roos, E. L. Daniel, E. N. Baker, Distant structural homology leads to the functional characterization of an archaeal PIN domain as an exonuclease. *J Biol Chem* **279**, 16471-16478 (2004).
14. K. S. Winther, K. Gerdes, Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. *Proc Natl Acad Sci U S A* **108**, 7403-7407 (2011).
15. J. W. Cruz *et al.*, Growth-regulating Mycobacterium tuberculosis VapC-mt4 toxin is an isoacceptor-specific tRNase. *Nature communications* **6**, 7480 (2015).
16. J. D. Sharp *et al.*, Growth and translation inhibition through sequence-specific RNA binding by Mycobacterium tuberculosis VapC toxin. *J Biol Chem* **287**, 12835-12847 (2012).
17. K. S. Winther, D. E. Brodersen, A. K. Brown, K. Gerdes, VapC20 of Mycobacterium tuberculosis cleaves the Sarcin-Ricin loop of 23S rRNA. *Nature communications* **4**, 2796 (2013).
18. J. M. Schifano *et al.*, tRNA is a new target for cleavage by a MazF toxin. *Nucleic Acids Res* **44**, 1256-1270 (2016).
19. K. Xu *et al.*, Protein expression, crystallization and preliminary X-ray crystallographic analysis of the isolated Shigella flexneri VapC toxin. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **69**, 762-765 (2013).

20. G. L. Conn, *Recombinant and in vitro RNA synthesis : methods and protocols*. Methods in molecular biology, (Humana Press, New York, 2012), pp. xiv, 274 p.
21. M. A. Schureck *et al.*, Structure of the Proteus vulgaris HigB-(HigA)2-HigB toxin-antitoxin complex. *J Biol Chem* **289**, 1060-1070 (2014).
22. J. M. Schifano *et al.*, Mycobacterial toxin MazF-mt6 inhibits translation through cleavage of 23S rRNA at the ribosomal A site. *Proc Natl Acad Sci U S A* **110**, 8501-8506 (2013).
23. K. Xu, E. Dedic, D. E. Brodersen, Structural analysis of the active site architecture of the VapC toxin from Shigella flexneri. *Proteins* **84**, 892-899 (2016).

Chapter 5: Discussion

Abstract

Bacterial toxin-antitoxins systems play important roles in plasmid maintenance, antibiotic tolerance, and a variety of other stress survival pathways (1, 2). Here, I focus on the distinguishing structural features of the MazF/CcdB toxin family. The majority of MazF/CcdB toxins belong to the type II class of toxins and are endoribonucleases that cleave several different types of RNA- from single-stranded mRNA to highly structured rRNA and tRNA. My structure of MazF-mt6 has revealed that not all MazF toxins are regulated through an open/closed state of their β1-β2 linkers as previously hypothesized. Here I discuss the implications of MazF-mt6 lacking a regulated closed state and how it might impact substrate recognition and antitoxin binding. Furthermore, based upon a closer inspection of the MazF family, one branch of the MazF/CcdB family appears to belong to the type III class of toxins, but much less is known about these systems. I argue that the type III toxin class should be the focus of more studies in the future due to their ability to recognize structured RNA and active site conservation with the CcdB/MazF toxin family. Finally, I summarize the importance of further structural studies of toxins aimed at the understanding for how these unique gene pairs function. Additionally, further insights will have a potential impact in combating antibiotic tolerance or persistence.

Understanding toxin superfamilies through structure

Bacterial toxins are small, single domain proteins that have poor sequence conservation, yet adopt very similar structural folds (3). Therefore, they represent excellent model systems to study divergent functions of structurally-related proteins. However, due to their poor sequence conservation, it is typically unclear what toxin family they will resemble, therefore it is critical to solve their molecular structures. For example, the CcdB/MazF toxin family has two very distinct activities and divergent sequences, yet they

share a conserved tertiary fold (4). MazF toxins are endoribonucleases and CcdB toxins are inhibitors of DNA gyrase. In Chapter 3, I solved the X-ray crystal structure of one of the six *Mycobacterium tuberculosis* (*Mtb*) MazFs, MazF-mt6, and performed additional functional assays to identify important residues. One unique feature of *Mtb* MazF-mt6 is its truncated β1-β2 linker, as compared to MazFs from *E. coli* and *B. subtilis* MazF (5, 6). The MazFs β1-β2 linker is part of the active site and is responsible for base-specific substrate recognition (5-7). In the MazF-mt6 structure, the shorter linker allows for the partial formation of its active site. As a consequence, MazF-mt6 has a much deeper RNA binding pocket and therefore may use other regions for base-specific RNA recognition (7). The deeper RNA binding pocket could also explain how MazF-mt6 is able to recognize structured RNA. Therefore, my structure has contributed to the field by expanding our understanding for how diverse MazFs recognize different RNAs.

Comparison of MazFs with CcdB toxins reveal that they share a SH3-like, β-sheet core fold with an antitoxin binding region at their homodimer interfaces **(Fig 5.1)** (6, 8-11). MazF and CcdB additionally contain the same number of β-strands, so I will refer to the regions or loops of the proteins based on their location between the conserved strands (*e.g.* the β1-β2 linker is a loop region between β1 and β2) **(Fig 5.1)**. Although MazF and CcdB toxins are structurally similar, these two toxins use mutually exclusive regions for catalysis and are regulated in opposite manners (12, 13). For example, the β1-β2 linker is important for MazF RNase activity, however, the β1-β2 linker has no direct role in binding DNA gyrase subunit GyrA by CcdB (5, 9, 10, 14). The MazF active site is located between the β1-β2 and β3-β4 linkers and contains a highly-conserved arginine and either a serine or threonine residue (5, 6, 12). Additional RNA sequence recognition is achieved through the β5-β6 linker and in some cases, with elongated RNA recognition sequences (> 5 nts), the β3-β4 linker and β7(6). The MazF β1-β2 linker is import for both RNA recognition and RNase activity regulation by MazE antitoxin binding (6, 15). This is in contrast to CcdB,

which uses the β1-β2 linker for antitoxin regulation, but primarily uses its C-terminal α-helix to bind and inhibit the GyrA subunits **(Fig. 5.1D)** (9, 16). The α-helix of CcdB presents four amino acids, three are typically asparagine or glutamine, and the C-terminal residue is either a tryptophan (*E. coli*) or threonine (*V. fischeri*) that form hydrogen bonds at the GyrA dimer interface and cause a conformational change in the gyrase complex (9, 13, 16). This means that MazF and CcdB share a common fold and regions that are important for toxin regulation are conserved, but the regions responsible for activity have diverged significantly.

Both MazF and CcdB form homodimers, which is important for antitoxin binding (5, 6, 8, 9, 12). In addition, the structures of CcdB bound to GyrA strongly suggest that dimerization is required for gyrase inhibition (9, 16, 17), yet it is unclear if MazFs require dimerization for their RNase activity. Unlike CcdB that uses the entire dimer to inhibit GyrA, MazF toxins have two independent active sites, each contained within one monomer **(Fig. 5.1D)** (5, 6). Additionally, the same monomer that cleaves RNA also performs base-specific recognition (5, 6). One possible reason for MazF dimerization is that each MazF monomer is not active on its own and may contribute to non-specific interactions with the substrate RNA sugar-phosphate backbone **(Fig 5.1D)**. Another possibility is that the dimerization is critical for the antitoxin MazE to bind at the dimer interface. In this context, MazE could inhibit both active sites of MazF simultaneously, which may be easier for cellular regulation of MazF by requiring only one MazE antitoxin for every two MazF toxins. Therefore, dimerization of MazF may have more to do with the regulation of antitoxin binding than inhibiting RNase activity (5).

Protein dynamics, that is specifically how the antitoxin regulates the flexibility of the MazF and CcdB homodimers, also appears to be different for each toxin. CcdB undergoes a hinging motion at its α-helical dimer interface to inhibit GyrA and this hinging motion is inhibited by the antitoxin CcdA (10). Regulation of the α-helical hinging is

achieved through the β1-β2 linkers, which are normally disordered when CcdB is bound to GyrA and transition to an ordered state when CcdA binds (9, 10, 16). Interestingly, the opposite is true for MazF toxins where the β1-β2 linker is ordered when bound to substrate RNA and disordered when bound to the antitoxin MazE (5, 6, 12, 15). When both structures of *S. aureus* and *E. coli* MazF were solved, the authors proposed that both MazF and CcdB are capable of the same dynamic states and are regulated in an opposite manner as compared to one another (**Fig 5.2A**) (5, 12). However, my structure of MazF-mt6 seemingly lacks this regulation due to the shorter β1-β2 linkers (Chapter 3). Using my structure of MazF-mt6 as an example, I argue that the two common dynamic states that MazF and CcdB adopt, termed open and closed, are not only important for toxin regulation by its antitoxin, but also the respective activities of MazF and CcdB independent from their antitoxins.

An interesting observation to arise from my research on MazF-mt6 is that because the MazF and CcdB active sites are located on opposite regions, in theory, a single toxin could contain both activities. However, at this time, there is no evidence that a MazF or CcdB family member inhibits both functions. If the main purpose of MazF/CcdB toxins is to inhibit growth during stress, is it impossible to inhibit both protein synthesis and DNA replication with one toxin? One potential reason why a MazF/CcdB dual toxin may not exist is that binding RNA could prevent binding to gyrase and vice versa because of the different requirements for an open versus a closed state. Typically, MazF toxins never enter the open state because the β1-β2 linker forms interactions across the homodimer interface and keeps the closed state stable **(Fig 5.2A)** (5). If mutations are introduced in *E. coli* MazF that disrupt the β1-β2 linker bridges, all RNase activity is lost (5, 18). However, I predict that analogous mutations in the CcdB β1-β2 linker would have the opposite effect on gyrase inhibition by hindering the closed state with bound antitoxin CcdA. Therefore, for every member of the CcdB/MazF family, one of the states is preferred

for activity and the other is inhibitory. My structures of MazF-mt6 reveal that it has a significantly shorter β1-β2 linker and lacks intersubunit bridges **(Fig 5.2B)** (Chapter 3). Without the intersubunit bridge, MazF-mt6 could theoretically interchange between open and closed states, which is opposite to other CcdB/MazF family members. However, I propose that regulation of MazF-mt6 is not linked to the β1-β2 linker, but instead, possibly linked to the highly-structured C-terminus. MazF-mt6 may lock itself in the closed state by preventing the C-terminal α-helix from moving, and therefore, the open state may never occur. This would prevent MazF-mt6 binding GyrA even without the β1-β2 linker regulation of the open/closed state. Another possibility is that MazF-mt6 is active in an open state, and the C-terminal residues obscure what would be a GyrA binding site. Future studies could test if the C-terminus of MazF-mt6 is important for maintaining the closed state by generating truncation mutants and testing for activity.

Insights into RNA recognition by type III toxins

In the toxin-antitoxin field, the type II toxins were the first to be characterized and consequently, the other two main classes of toxin-antitoxins (type I and type III) are not as well understood (19). Type III toxins play roles in plasmid maintenance and aborting phage infections, but our understanding of these systems is limited to only a couple of examples (20-23). Although they may not be as well studied as the type II systems, I argue that we can learn a lot about how toxins recognize complex RNA structures through more detailed studies of the type III systems.

Type III toxins belong to the MazF toxin family despite being a different toxin class (24). This is because toxin classes are based on the antitoxin's mode of inhibiting the toxin, not the homology of the toxins themselves (24, 25). Type III antitoxins are non-coding RNAs that form pseudoknots (21). The best-studied type III system is ToxIN from *Pectobacterium atrosepticum* and *Bacillus thuringiensis*, where ToxI is the RNA antitoxin

and ToxN is the protein toxin (21, 23). One key difference though is that *mazEF* and *toxIN* genes are organized and regulated in a completely different way **(Fig 5.3)**. MazEF, like all other type II systems, contains a single transcript for the antitoxin and toxin and the protein antitoxin is a transcriptional repressor of its own expression (26-28). There is no evidence to suggest ToxI RNA antitoxin regulates the transcription of the ToxIN operon. Instead, the operon contains several copies of the ToxI non-coding RNA and a Rho-independent terminator upstream of the ToxN gene, ensuring that ToxN levels do not exceed ToxI levels under non-stress conditions **(Fig. 5.3)** (20-23). MazF and ToxN are both sequence-specific endoribonucleases and cause inhibition of growth when overexpressed (11, 20). However, MazF toxins have been associated with the stringent response and bacterial virulence, while ToxN is involved in abortive phage infections and plasmid maintenance (2, 20, 21, 29, 30). The differences in the cellular roles of MazF and ToxN might be attributed to their sequence specificity, whereby MazF toxins cleave a variety of different sequences, while ToxN cleaves sequences that contain three to five adenosines, which might be found in phage transcription products (4, 11, 31, 32).

Although ToxN and MazF differ substantially in regulation and function, they are structural homologues. The structure of *Pectobacterium atrosepticum* ToxN bound to the non-coding RNA (ncRNA) antitoxin ToxI revealed that ToxN shares the same core β-barrel fold as MazF (23). The antitoxin ToxI ncRNA inhibits ToxN by binding the same positive groove that MazF toxins use to bind and cleave RNA (23). One of the striking differences between MazF and ToxN is the quaternary structure upon comparison of the toxin alone and toxin-antitoxin complexes. MazF toxins form homodimers and the MazEF complex is a heterohexamer of two MazE antitoxins and four MazF toxins **(Fig. 5.3 C)**. In contrast, ToxN is the only member of the MazF/CcdB family to exist as a monomer in solution and it forms a triangular ring structure composed of three ToxN and three ToxI molecules **(Fig 5.3 D)** (22, 23, 25). ToxN proteins have a significantly extended C-terminal α-helix that

wraps around and packs against the usual homodimer interface of other MazF/CcdB toxins (23). Despite the differences in quaternary structure, ToxN recognizes RNA using the same surface as MazF that is, the β1-β2, β3-β4 and β5-β6 linkers and in addition, β3-β4 linker and its unique C-terminal region **(Fig. 5.4)** (20, 22, 23). Recognition of structural features in RNA is a novel feature of the MazF/CcdB toxin family, with only MazF-mt3, MazF-mt6 (Chapter 3) and MazF-mt9 (Chapter 4) exhibiting specificity for structured RNAs like tRNA and rRNA (31-33). There are currently three examples of type III systems bound to their cognate RNA antitoxin and they provide a wealth of information on how a MazF toxin could potentially recognize structured RNA substrates (21, 23).

One reason that the similarities between MazF and ToxN have not been more thoroughly explored is because it was incorrectly assumed there is a weak relationship between ToxN and the MazF/CcdB toxin family (20, 23, 34). When the first structure of ToxIN was solved, it was found that ToxN shares the same core fold with the CcdB/MazF family confirmed by structural searches identifying that ToxN was most similar to *kid* (PemK), MazF and CcdB toxins (22). However, because the proposed active site of ToxN did not align well with the MazF/CcdB family member PemK, these proteins were thought to be functionally different and therefore should be classified into different toxin families (22). However, when the *B. subtilis* MazF-RNA structure was solved, it revealed that the MazF active site was not where it was originally proposed using the PemK structure as a guide **(Fig. 5.5A)** (6). The new active site was further supported by the structure of *E. coli* MazF bound to DNA and biochemical characterization of MazF-mt6 (Chapter 3) (5). The originally proposed active site is responsible for RNA binding in both *B. subtilis* MazF and *E. coli* MazF structures, while the actual active site aligns significantly better with the proposed active site of ToxN **(Fig 5.5B)** (5, 6, 20, 22, 23). Unfortunately, the old active site has continued to be supported in numerous studies that followed the structural work (34). In summary, I suggest that MazF and ToxN *do* share the same active site location and

recognize and cleave RNA in a conserved mechanism found in the type II MazF/CcdB endoribonuclease toxins.

MazF and ToxN are an excellent example as to why both toxins and the antitoxin should be used in classifications. Subdividing each class based on the toxin would allow further clarification of toxin-antitoxin systems that share members of the same structural family. Currently, the type II toxin-antitoxin class is enormous and has several toxins with unique activities. If a subdividing classification was adopted then type II-a and type II-b would be the two component protein systems, type II-c would be the three component type II toxin-antitoxin-chaperon (35), and type II-d would be the current type IIIs **(Fig. 5.6)**. I argue that names like type II-a and type II-d imply a closer relationship and would aid in clarifying the relationship between similar toxins along with the cognate antitoxin. To prevent confusion in the literature, I also propose to maintain the type I, IV, V and VI as is and either remove type III or make it interchangeable with type II-d. Not only would my classification help to clarify the type II class, it also allows for easier additions of novel toxin inhibitory mechanisms to a given class in the future.

MazF-mt9 is a potential link between ToxN and MazF toxins

An interesting development in the MazF toxin field is the expansion of the MazF family in *Mtb* where now instead of cleaving free single-stranded mRNAs, MazFs degrade more complex RNAs including tRNAs and rRNAs (31-33). One interesting MazF homologue is MazF-mt9 which cleaves $tRNA^{Lys}$ in its anticodon loop and in the D-loop of $tRNA^{Pro}$ at sequences containing 5’-UUU-3’ (32). Additionally, MazF-mt9 has a slight preference for 5’-GGA-3’ sequences upstream of the UUU consensus site with a gap of two nucleotides that have no sequence requirement (5’-GGANNUUU-3’) (32). I attempted to solve the crystal structure of MazF-mt9 (Chapter 4) to identify the structural basis for

recognition of tRNAs. I was unable to grow crystals of MazF-mt9 but I did optimize the purification of MazF-mt9 as well as its substrate tRNAs for future studies.

One interesting thing about MazF-mt9 is it contains an extension of 20 amino acids at its C-terminus upon comparison with other MazF toxins (Figure 3.9) (31, 32, 36). Other regions of MazF-mt9 align well with other MazFs including the β1-β2 and β3-β4 linkers and the catalytic arginine and threonine. I hypothesize that the C-terminal extension may play a role in the ability of MazF-mt9 to recognize and cleave tRNAs. I was unable to model or predict any secondary structure for the last 20 C-terminal residues (37, 38), however I have some predictions for how MazF-mt9 might recognize tRNA using the solved structures of the ToxIN complex (21, 23).

The MazF-mt9 structure is not known but because it appears its active site is conserved based on sequence alignments, I predict that β1-β2, β3-β4 and β5-β6 likely recognizes UUU and the opposite side of the β3-β4 linker likely recognizes GGA (consensus sequence of 5'-GGANNUUU-3') similar to how ToxN binds to ToxI RNA **(Fig 5.4)** (23). Comparing the antitoxin ToxI to the anticodon stem loop of tRNAs suggests that the anticodon packs in the active site with the RNA wrapped around the β3-β4 linker **(Fig 5.7)**. Also, the β3-β4 linker region may be in a similar location as the same ToxN linker that is found positioned in a groove formed by the ToxI pseudoknot. Based on my size exclusion chromatography data, MazF-mt9 exists as a dimer, but always forms two distinct bands on SDS-PAGE (Chapter 4). One idea is that perhaps only one C-terminus is stable in the MazF-mt9 homodimer and the other is degraded because it is intrinsically disordered. The C-terminus of MazF-mt9 is not nearly as long as ToxN, but it may pack against structured RNA in a similar manner to how ToxN packs against ToxI. MazF-mt9 presents a unique opportunity to expand our understanding of the CcdB/MazF toxin family similar to how the structure of MazF-mt6 provided insights into how it is able to interact with rRNA.

Concluding remarks

Bacterial infections have been the leading cause of death throughout human history and is still a significant problem in some parts of the world. Antibiotics may have been the go-to solution for infections for over 70 years (31), but bacteria have had millions of years to evolve and establish mechanisms to evade their effectiveness. Our current regimen of antibiotics is becoming less effective and we must find new ways to combat bacterial infections in the coming years. Bacterial stress response mechanisms are a critical component for pathogen survival and lead to the activation of toxin-antitoxin systems. By studying toxins and their substrates, we will expand our understanding of bacterial stress responses and growth regulation thus possibly finding new ways to combat antibiotic-resistant infections.

Structural studies of toxin-antitoxin systems have greatly expanded our understanding of both the mechanism and substrate specificities of these important bacterial factors. However, there is still much that we do not understand, especially about the type II ribosome-independent and type III toxins. Here, I explored some of the gaps in our knowledge about the CcdB/MazF structural family and how a common protein fold can have multiple activities depending upon how the antitoxin regulates it. I have expanded our understanding of MazF substrate with the structure of MazF-mt6 **(Fig. 5.8)**. I also provided a hypothesis for how MazF-mt6, which lacks the usual regulatory motif which might prevent a closed to open transitions that regulate RNA or antitoxin binding **(Fig. 5.2)**. The lack of regulation in MazF-mt6 appears to be regulated by the formation of a new negatively charged surface which would repel RNA but would allow positively charged residues as found in MazE-mt6. Without the MazF-mt6 structure, it would have been impossible to predict these features, again emphasizing the importance of structural studies of toxin-antitoxin complexes.

Again, the importance of solving high-resolution structures toxin-antitoxin complexes cannot be understated. For example, if the crystal structure of MazF-RNA was not solved, the field would still be focused on a misidentified active site. Another example is that upon solving the structure of the first type III toxin (ToxN), it was revealed that ToxN belonged to the type II MazF/CcdB structural family. The ToxN structure has now provided the field with a MazF/CcdB toxin example that recognizes structured RNA. Based upon these insights, it has the potential to aid in prediction of substrate specificity of other family members. I propose that MazF-mt9, which cleaves tRNAs, might recognize RNA in a similar manner to ToxN which may be dependent upon shared, unique features like an extended C-terminus **(Fig. 5.7)**. However, until a structure of MazF-mt9 is solved, we will not have a clear understanding of how the CcdB/MazF toxin family recognizes tRNA **(Fig. 5.8)**. In conclusion, structural studies of toxins are paramount to our understanding their biology and help to inform on their unique abilities to target important molecular targets.

Figures

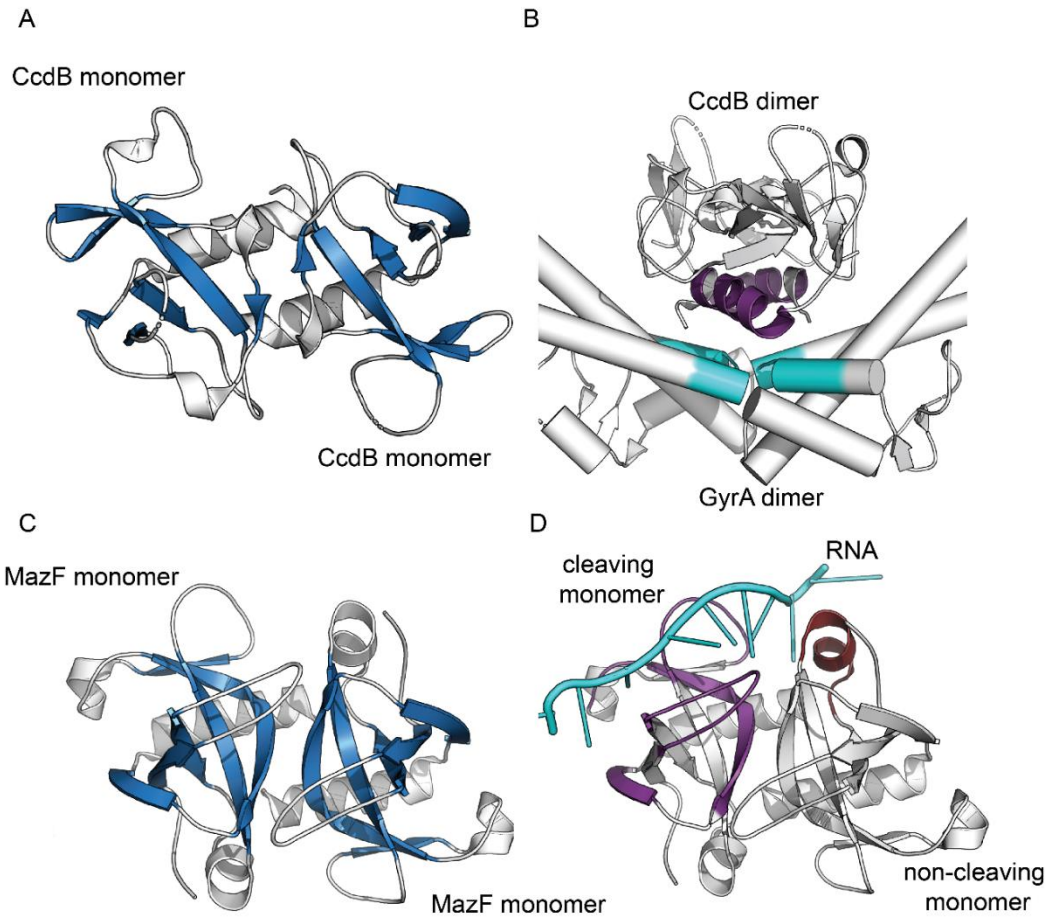

Figure 5.1 Secondary structure and ligand binding comparisons of CcdB and MazF toxins. A) Structure of *E. coli* CcdB homodimer with the β-barrel core highlighted in blue (PDB ID: 1X75). B) A 90° rotation of CcdB with *E. coli* GyrA homodimer bound (PDB ID: 1X75). The C-terminal α-helix is highlighted in purple and the regions of GyrA that CcdB interacts with are highlighted in cyan. C) Structure of *B. subtilis* MazF homodimer with the β-barrel core highlighted in blue similar to CcdB in panel A (PDB ID: 4MDX). D) *B. subtilis* MazF (white) bound to an RNA analog (cyan) (PDB ID: 4MDX). The MazF regions that bind RNA are highlighted in purple for the monomer and that cleave RNA are in red for the non-cleaving monomer.

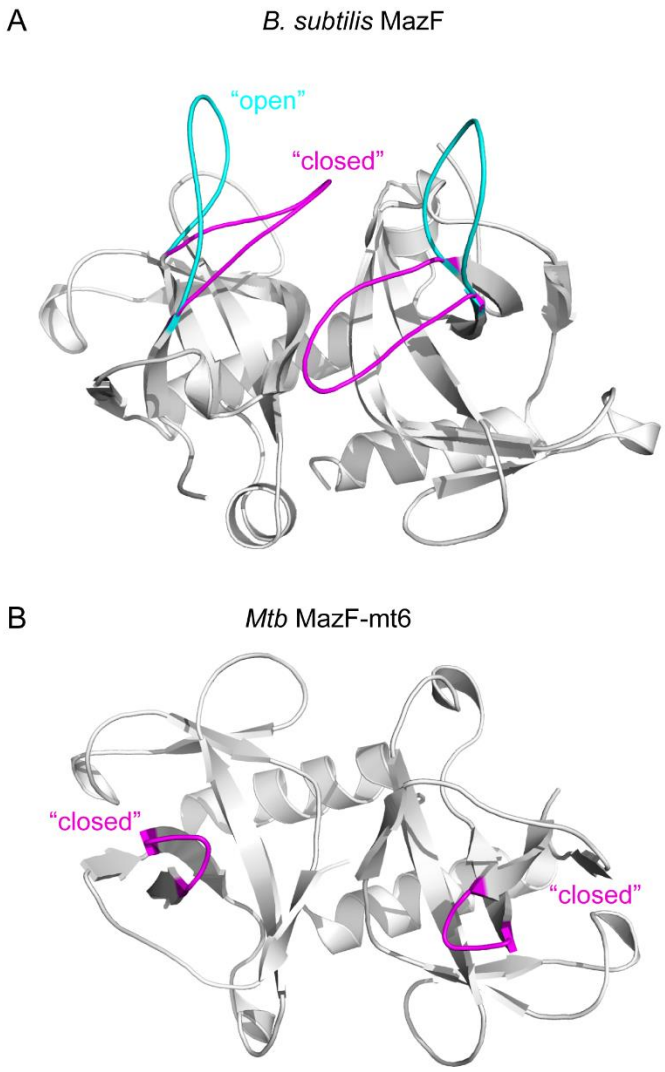

Figure 5.2 Conformations of the CcdB/MazF β1-β2 linker. A) *B. subtilis* MazF (PDB ID: 4MDX) showing the active "closed" state and a modeled "open" state. The closed state is seen in the toxin alone and RNA-bound structures and the open state is seen when antitoxin MazE is bound. For CcdB, the open state is observed when alone and bound to GyrA and the closed state is seen when antitoxin CcdA is bound. B) Structure of *Mtb* MazF-mt6 with the β1-β2 linker highlighted in pink. Unlike other CcdB/MazF structures, the β1-β2 linker does not seem capable of an open state.

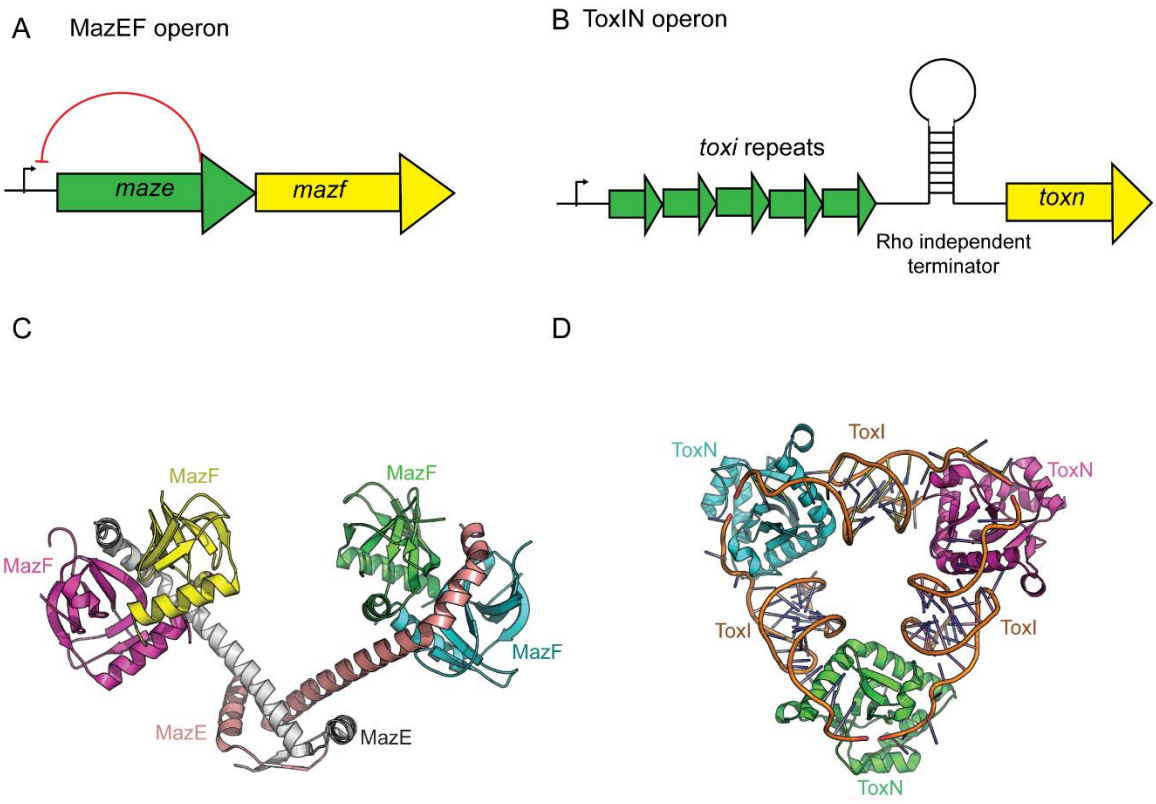

Figure 5.3 Genetic organization and quaternary structure of MazEF and ToxIN.

A) Genetic organization of a typical MazEF operon. Once the genes are transcribed, the antitoxin MazE inhibits further transcription and prevents toxin activity. B) Genetic organization of a typical ToxIN operon. An average of about 5 *toxi* ncRNA genes are upstream of the *toxn* gene. A rho independent terminator sits between the *toxi* and *toxn* gene, which keeps the levels of ToxI RNA higher than ToxN. C) Quaternary structure of *B. subtilis* MazEF (PDB ID: 4MDx) with each MazF and MazE monomer highlighted in a different color. There are two MazF homodimers bound to one MazE homodimer, which makes up the heterohexametric structure. D) Quaternary structure of *Bacillus thuringiensis* ToxIN (PDB ID:4ATO) showing the three ToxN monomers bound by three ToxI ncRNA antitoxins making up the ToxIN heterohexamer.

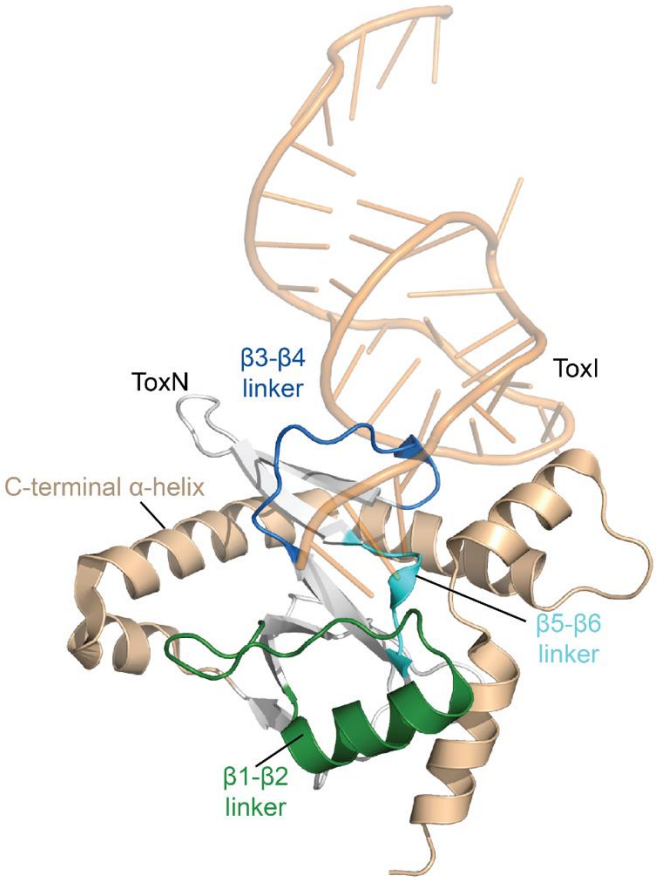

Figure 5.4 RNA binding regions of ToxN. A single ToxN monomer bound to ToxI ncRNA antitoxin with important regions of ToxN highlighted. The β1-β2, β3-β4 and β5-β6 linkers bind RNA similar to MazF with additional interactions from the C-terminal α-helix.

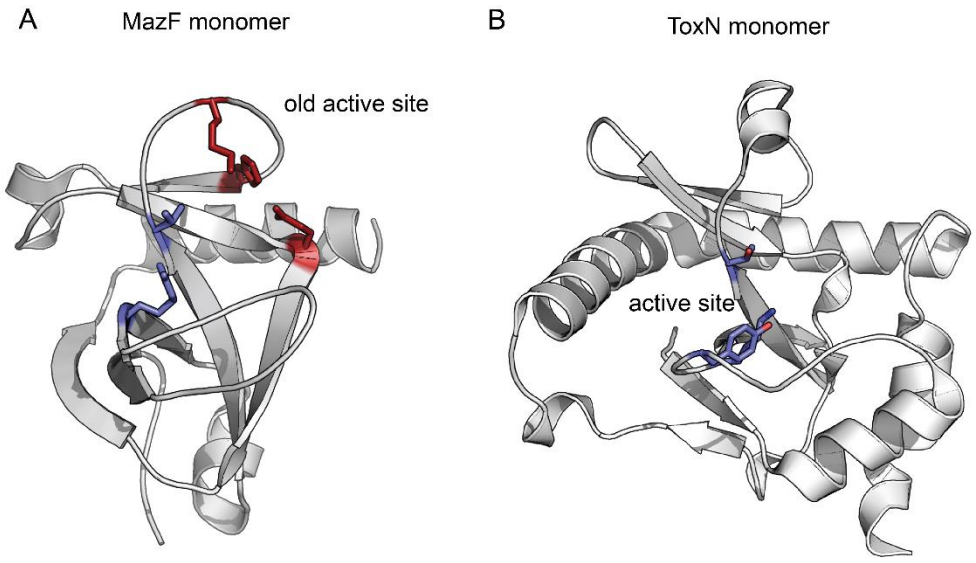

Figure 5.5 Comparison of the incorrect and correct MazF active sites as compared to ToxN. A) *B. subtilis* MazF monomer (PDB ID: 4MDX) with the old active site identified by NMR highlighted in red and the actual active site highlighted in purple. B) ToxN monomer (PDB ID: 4ATO) with the proposed active site highlighted in purple.

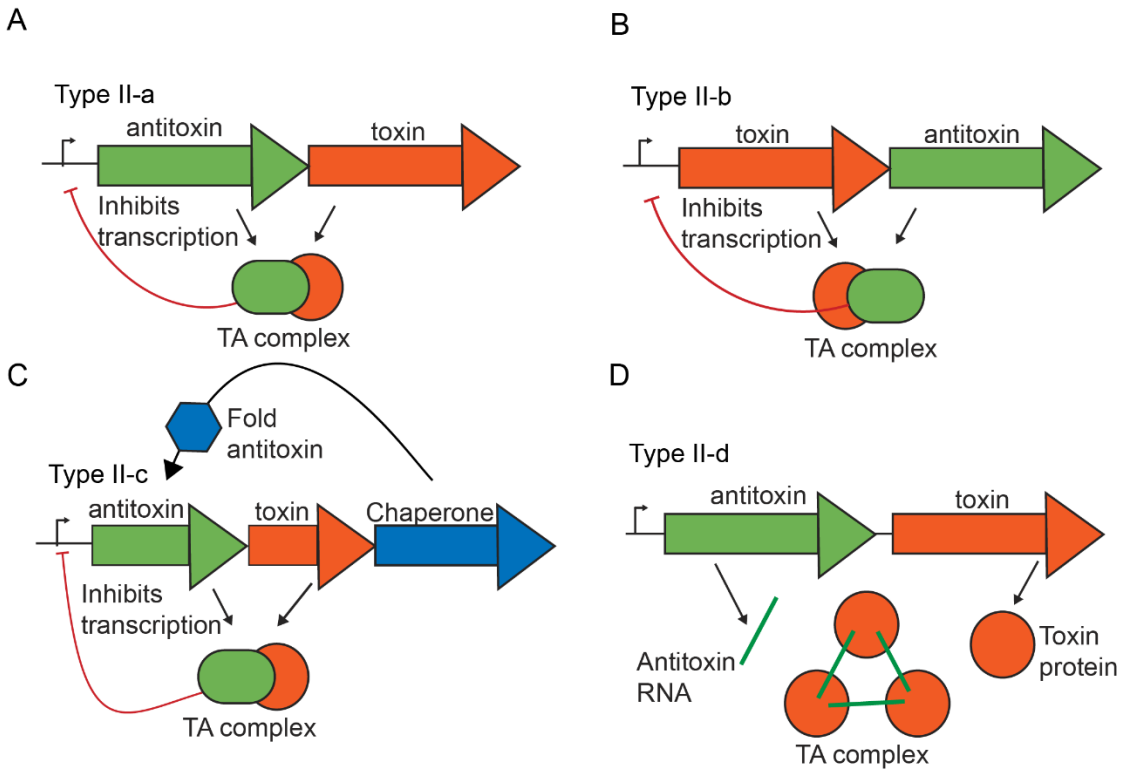

Figure 5.6 Proposed alternative nomenclature for toxin-antitoxin classification. A) The canonical type II toxin-antitoxin class could be classified as type II-a as it was the first type II toxin-antitoxin system to be defined. B) Type II-b toxin-antitoxins are nearly identical to type II-a, except the gene order in the operon is reversed with toxin first and antitoxin second. C) Some type II toxin-antitoxins systems have a third component called a chaperone. The chaperone is needed to properly fold the antitoxin protein before it can inhibit the toxin. The tripartite type II systems could be called type II-c as they contain three components and C is the third letter in the alphabet. D) The current type III toxin-antitoxin systems are now type II-d, which implies they are related to the other type II systems, but have a unique mode of regulation.

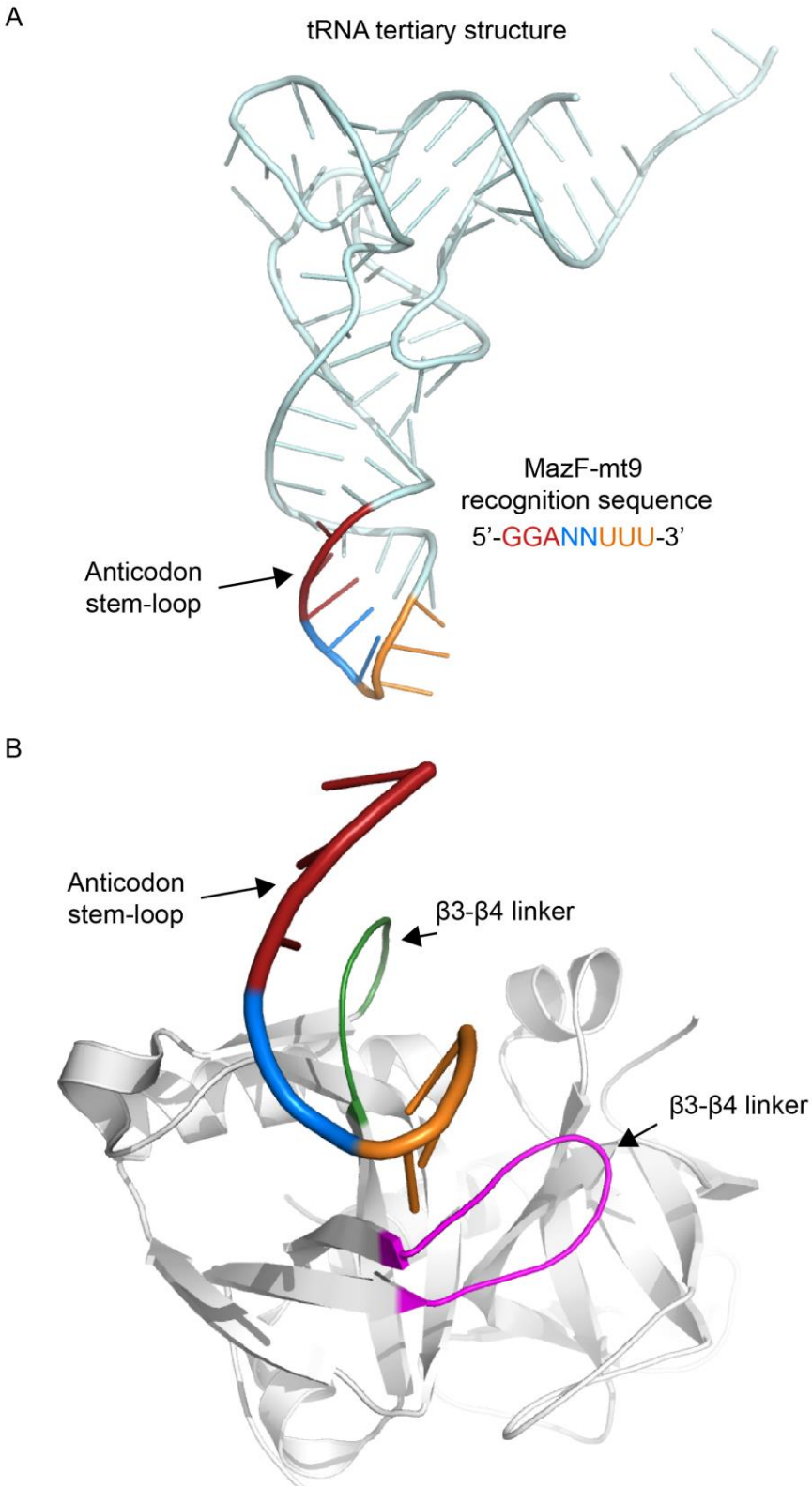
A
tRNA tertiary structure
MazF-mt9
recognition sequence
5'-GGANNUUU-3'
Anticodon
stem-loop
B
Anticodon
stem-loop
β3-β4 linker
β3-β4 linker

Figure 5.7 Model for how MazF-mt9 recognizes tRNAs. A) Tertiary structure of a tRNA with the location of MazF-mt9 RNA recognition on the anticodon stem-loop highlighted. MazF-mt9 RNA recognition is split into three parts with the upstream recognition (red), unrecognized nucleotide spacing notated as N (blue) and the sequence that is cleaved (orange) highlighted. B) Anticodon stem-loop RNA from panel A is superimposed onto the structure of *B. subtilis* MazF (PDB ID: 4MDX). MazF-mt9 likely uses the β1-β2 (magenta) and β3-β4 (green) linkers to recognize RNA. The β3-β4 linker might also recognize the upstream portion of the recognition sequence using the opposite side of the β3-β4 linker.

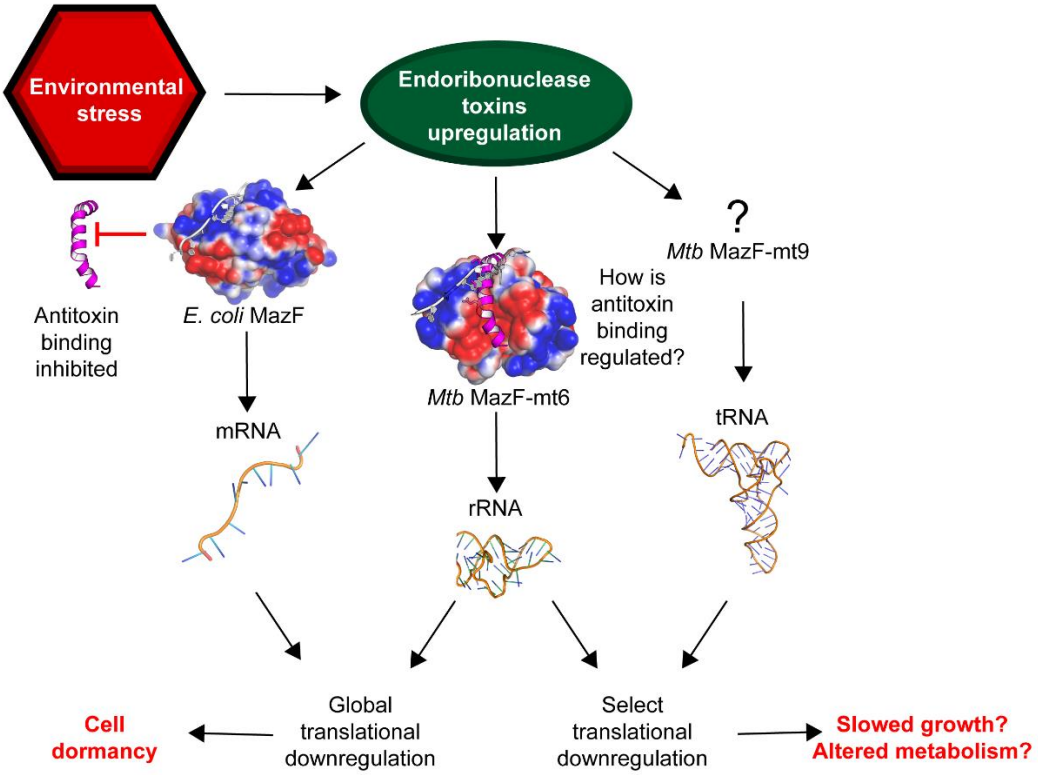

Figure 5.8 MazF RNA substrates and consequences of their cleavage. When a cell encounters stress, endoribonucleases like MazF are upregulated and act on their respective RNA targets. The *E. coli* MazF has been well studied and it is clear that mRNA cleavage leads to global translational inhibition and that RNA binding excludes antitoxin binding. Mtb MazF-mt6 cleaves 23S rRNA and my studies have provided critical insights into how the MazF family can recognize structured RNA. Additionally, my studies suggest that the binding of antitoxin or RNA may be regulated in different ways. Unlike mRNA, the cleavage of rRNA could lead to selective translation along with MazF-mt9 cleavage of elongator tRNAs. How MazF-mt9 cleaves tRNA is unknown and its structure will provide great insights into the regulation of translation by this toxin family. The implications of selective translation could lead to slower growth instead of latency or even alter the proteome during stress.

References

1. S. Maki, S. Takiguchi, T. Miki, T. Horiuchi, Modulation of DNA supercoiling activity of Escherichia coli DNA gyrase by F plasmid proteins. Antagonistic actions of LetA (CcdA) and LetD (CcdB) proteins. *J Biol Chem* **267**, 12244-12251 (1992).
2. X. Wang, T. K. Wood, Toxin-antitoxin systems influence biofilm and persister cell formation and the general stress response. *Appl Environ Microbiol* **77**, 5577-5583 (2011).
3. Y. Yamaguchi, J. H. Park, M. Inouye, Toxin-antitoxin systems in bacteria and archaea. *Annu Rev Genet* **45**, 61-79 (2011).
4. L. Zhu *et al.*, Staphylococcus aureus MazF specifically cleaves a pentad sequence, UACAU, which is unusually abundant in the mRNA for pathogenic adhesive factor SraP. *J Bacteriol* **191**, 3248-3255 (2009).
5. V. Zorzini *et al.*, Substrate Recognition and Activity Regulation of the Escherichia coli mRNA Endonuclease MazF. *J Biol Chem* **291**, 10950-10960 (2016).
6. D. K. Simanshu, Y. Yamaguchi, J. H. Park, M. Inouye, D. J. Patel, Structural basis of mRNA recognition and cleavage by toxin MazF and its regulation by antitoxin MazE in Bacillus subtilis. *Mol Cell* **52**, 447-458 (2013).
7. E. D. Hoffer, S. J. Miles, C. M. Dunham, The structure and function of Mycobacterium tuberculosis MazF-mt6 provides insights into conserved features of MazF endonucleases. *J Biol Chem*, (2017).
8. R. Loris *et al.*, Crystal structure of CcdB, a topoisomerase poison from E. coli. *Journal of molecular biology* **285**, 1667-1677 (1999).
9. N. De Jonge *et al.*, Structural and thermodynamic characterization of Vibrio fischeri CcdB. *J Biol Chem* **285**, 5606-5613 (2010).
10. N. De Jonge *et al.*, Rejuvenation of CcdB-poisoned gyrase by an intrinsically disordered protein domain. *Mol Cell* **35**, 154-163 (2009).
11. Y. Zhang *et al.*, MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli. *Mol Cell* **12**, 913-923 (2003).
12. V. Zorzini *et al.*, Structural and biophysical characterization of Staphylococcus aureus SaMazF shows conservation of functional dynamics. *Nucleic Acids Res* **42**, 6709-6725 (2014).
13. N. De Jonge *et al.*, Alternative interactions define gyrase specificity in the CcdB family. *Mol Microbiol* **84**, 965-978 (2012).
14. J. H. Park, Y. Yamaguchi, M. Inouye, Bacillus subtilis MazF-bs (EndoA) is a UACAU-specific mRNA interferase. *FEBS letters* **585**, 2526-2532 (2011).
15. K. Kamada, F. Hanaoka, S. K. Burley, Crystal structure of the MazE/MazF complex: molecular bases of antidote-toxin recognition. *Mol Cell* **11**, 875-884 (2003).
16. M. H. Dao-Thi *et al.*, Crystallization of CcdB in complex with a GyrA fragment. *Acta crystallographica* **60**, 1132-1134 (2004).
17. P. Bernard, M. Couturier, Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *Journal of molecular biology* **226**, 735-745 (1992).
18. A. Tripathi, P. C. Dewan, S. A. Siddique, R. Varadarajan, MazF-induced growth inhibition and persister generation in Escherichia coli. *J Biol Chem* **289**, 4191-4205 (2014).
19. T. Ogura, S. Hiraga, Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proc Natl Acad Sci U S A* **80**, 4784-4788 (1983).

20. T. R. Blower *et al.*, Mutagenesis and functional characterization of the RNA and protein components of the toxIN abortive infection and toxin-antitoxin locus of Erwinia. *J Bacteriol* **191**, 6029-6039 (2009).
21. T. R. Blower *et al.*, A processed noncoding RNA regulates an altruistic bacterial antiviral system. *Nat Struct Mol Biol* **18**, 185-190 (2011).
22. P. C. Fineran *et al.*, The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc Natl Acad Sci U S A* **106**, 894-899 (2009).
23. F. L. Short *et al.*, Selectivity and self-assembly in the control of a bacterial toxin by an antitoxic noncoding RNA pseudoknot. *Proc Natl Acad Sci U S A* **110**, E241-249 (2013).
24. D. Lobato-Marquez, R. Diaz-Orejas, F. Garcia-Del Portillo, Toxin-antitoxins and bacterial virulence. *FEMS Microbiol Rev* **40**, 592-609 (2016).
25. G. M. Cook *et al.*, Ribonucleases in bacterial toxin-antitoxin systems. *Biochim Biophys Acta* **1829**, 523-531 (2013).
26. I. Marianovsky, E. Aizenman, H. Engelberg-Kulka, G. Glaser, The regulation of the Escherichia coli mazEF promoter involves an unusual alternating palindrome. *J Biol Chem* **276**, 5975-5984 (2001).
27. C. F. Schuster *et al.*, Characterization of a mazEF toxin-antitoxin homologue from Staphylococcus equorum. *J Bacteriol* **195**, 115-125 (2013).
28. E. Aizenman, H. Engelberg-Kulka, G. Glaser, An Escherichia coli chromosomal "addiction module" regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death. *Proc Natl Acad Sci U S A* **93**, 6059-6063 (1996).
29. K. Gerdes, S. K. Christensen, A. Lobner-Olesen, Prokaryotic toxin-antitoxin stress response loci. *Nat Rev Microbiol* **3**, 371-382 (2005).
30. P. Tiwari *et al.*, Corrigendum: MazF ribonucleases promote Mycobacterium tuberculosis drug tolerance and virulence in guinea pigs. *Nature communications* **6**, 7273 (2015).
31. J. M. Schifano *et al.*, Mycobacterial toxin MazF-mt6 inhibits translation through cleavage of 23S rRNA at the ribosomal A site. *Proc Natl Acad Sci U S A* **110**, 8501-8506 (2013).
32. J. M. Schifano *et al.*, tRNA is a new target for cleavage by a MazF toxin. *Nucleic Acids Res* **44**, 1256-1270 (2016).
33. J. M. Schifano *et al.*, An RNA-seq method for defining endoribonuclease cleavage specificity identifies dual rRNA substrates for toxin MazF-mt3. *Nature communications* **5**, 3538 (2014).
34. N. Goeders, R. Chai, B. Chen, A. Day, G. P. Salmond, Structure, Evolution, and Functions of Bacterial Type III Toxin-Antitoxin Systems. *Toxins* **8**, (2016).
35. P. Bordes *et al.*, Chaperone addiction of toxin-antitoxin systems. *Nature communications* **7**, 13339 (2016).
36. L. Zhu *et al.*, Characterization of mRNA interferases from Mycobacterium tuberculosis. *J Biol Chem* **281**, 18638-18643 (2006).
37. J. Soding, A. Biegert, A. N. Lupas, The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**, W244-248 (2005).
38. V. Alva, S. Z. Nam, J. Soding, A. N. Lupas, The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res* **44**, W410-415 (2016).

Appendix I: Uniformity of Peptide Release Is Maintained by Methylation of Release Factors

William E. Pierson*, Eric D. Hoffer*, Hannah E. Keedy, Carrie L. Simms, Christine M. Dunham, and Hani S. Zaher

*Co-first authors

This research was originally published in Cell Reports. William E. Pierson, Eric D. Hoffer, Hannah E. Keedy, Carrie L. Simms, Christine M. Dunham, and Hani S. Zaher. Uniformity of Peptide Release Is Maintained by Methylation of Release Factors. *Cell Reports*. 2016 17(1):11-8.

AUTHOR CONTRIBUTIONS

W.E.P. designed and performed the peptide release assays. E.D.H. designed and carried out the X-ray crystallography assays. H.E.K. performed the nonenzymatic release assays. C.L.S. carried out the in vivo readthrough assays. All authors interpreted the results. C.M.D. and H.S.Z. designed the experiments and wrote the manuscript.

SUMMARY

Termination of protein synthesis on the ribosome is catalyzed by release factors (RFs), which share a conserved glycine-glycine-glutamine (GGQ) motif. The glutamine residue is methylated in vivo, but a mechanistic understanding of its contribution to hydrolysis is lacking. Here, we show that the modification, apart from increasing the overall rate of termination on all dipeptides, substantially increases the rate of peptide release on a subset of amino acids. In the presence of unmethylated RFs, we measure rates of hydrolysis that are exceptionally slow on proline and glycine residues and approximately two orders of magnitude faster in the presence of the methylated factors. Structures of 70S ribosomes bound to methylated RF1 and RF2 reveal that the glutamine side-chain methylation packs against 23S rRNA nucleotide 2451, stabilizing the GGQ motif and placing the side-chain amide of the glutamine toward tRNA. These data provide a framework for understanding how release factor modifications impact termination.

INTRODUCTION

Protein synthesis is an intricate process that involves many coordinated events on the ribosome, ensuring that the genetic code is decoded accurately and efficiently. Termination of protein synthesis occurs when one of three nearly universally conserved stop codons (UAA, UGA, and UAG) enters the A site of the ribosome. Stop codons are recognized by protein factors called release factors (RFs), which mediate the hydrolytic reaction in the peptidyl transferase center (PTC) of the ribosome. Unlike eukaryotes and archaea, which have one factor (eRF1 and aRF1, respectively) that recognizes all stop codons (1), bacteria have two very similar factors with overlapping specificities (2). RF1 recognizes UAG and UAA, whereas RF2 recognizes UGA and UAA.

Despite the conserved nature of the translation machinery in all domains of life, bacterial and eukaryotic RFs are unrelated and instead appear to have evolved

independently (3). The factors, however, share a universally conserved sequence motif of Gly-Gly-Gln (GGQ) (4). This motif is critical for the hydrolytic reaction (4), and early low-resolution structures of termination complexes revealed that the motif engages the PTC of the ribosome (5, 6). Mutation of either glycine to any other amino acid (aa) completely inhibits activity, whereas certain substitutions at the glutamine residue are only partially defective for promoting catalysis (4, 7). The ribosome also plays a critical role during the reaction. This is best highlighted by early observations showing that binding of a cognate deacylated tRNA to the A site accelerates the hydrolysis of the peptidyl-tRNA (8). Interestingly, in the presence of a deacylated tRNA, the specificity for water as a nucleophile is compromised akin to what has been observed for the RF glutamine mutants (7). These findings suggest that RFs contribute largely to catalysis through inducing conformational rearrangements in the PTC that expose the ester carbon for the nucleophilic attack. Consistent with this proposal, mutations of the so-called inner shell nucleotides of the PTC as well as the 2'-OH of the terminal adenosine of the peptidyl tRNA are detrimental for RF-mediated release, but not for peptidyl transfer (9-12).

Crystal structures of post-termination complexes revealed important clues about the mechanism of the reaction. The glycine residues in the GGQ motif adopt a conformation in the PTC that is incompatible with any other amino acid rationalizing earlier mutational studies (13-18). In contrast, the role of the glutamine in catalysis has been a source of considerable debate. One group argued that the main-chain amide of the glutamine is involved in coordinating a nucleophilic water molecule (16). Another group argued for both the main-chain amide and the side-chain carbonyl oxygen playing a critical role in coordinating the hydrolytic water (17). The conserved nature of the residue adds support to the side chain of glutamine contributing in some way to the mechanism.

It is worth noting that, although the 3.0–3.5 Å 70S-RF structures are at resolutions typically sufficient to observe interpretable electron density of protein side chains, there is

little to no electron density surrounding the unmethylated glutamine residue in all structures but the pre-termination 70S-RF2 complex and the recent post-termination 70S-*E. coli* RF1 (**Figure A1.5**; (13-15, 17, 18)). Therefore, it is still ambiguous whether the glutamine side-chain carbonyl or amide is directed toward the 3^0-OH of the substrate tRNA A76, where it could potentially coordinate the catalytic water molecule (7). Recent kinetic analyses examining the pK_a of the termination reaction suggested instead that a hydroxide ion could participate in the reaction, with the glutamine side-chain amide group donating a proton to the hydroxide to stabilize the reaction (19).

The glutamine residue of the GGQ motif in bacterial and eukaryotic RFs is N^5 methylated. In E. coli, RFs are methylated by prmC (20). The methylation is required for optimal growth in certain minimal media (21), and in vitro, the methylation contributes modestly to the rate of release by RF2 (22). Therefore, the coordination of water in the active site of the ribosome could be further stabilized by the methylation. Indeed, molecular dynamics studies suggest that the methylation removes the side-chain amide of the glutamine from the active site and helps in orienting the carbonyl oxygen toward the active site to coordinate the catalytic water (23). However, beyond these models, a mechanistic understanding for the role of the modification during peptide release is missing. Here, we explore the effects of methylation on the efficiency of peptide release using a well-defined bacterial reconstituted translation system. We measured the rates of peptide release on all 20 common amino acids with unmethylated and methylated RFs. Whereas the modification was found to increase the rate of release on all amino acids, this enhancement varies significantly, depending on the identity of the amino acid. We report four structures of unmethylated and methylated *E. coli* RF1 or RF2 in complex with the *Thermus thermophilus* (*Tth*) ribosome. In all structures, the placement of the glutamine backbone amide is within hydrogen bonding distance to the 3^0-OH of the terminal A76 of the P-site tRNA. However, in the 70S-$RF1^{Me}$ post-termination structure, the side-chain

amide packs against A2451 of 23S rRNA and is proximal to the tRNA, both suggesting an unexpected role of the side-chain amide in stabilizing the GGQ motif.

RESULTS

Experimental Approach

To explore the full contribution of the methylation of RFs on peptide release, we took advantage of our in vitro bacterial reconstituted system to generate dipeptidyl-tRNA ribosomal nascent chains (DRNCs) programmed with a UAA stop codon in the A site **(Figure A1.6A)**. Twenty initiation complexes (24) were generated by incubating 70S ribosomes with 20 mRNAs that had the following coding sequence (AUG NNN UAA) in the presence of [^{35}S]-Met-tRNA^{fMet}, initiation factors 1, 2, and 3, and GTP. We note that the NNN sequence coded for the most-abundant *E. coli* codon for each amino acid. DRNCs were generated by incubating the initiation complexes with their corresponding ternary complexes composed of the appropriate aa-tRNA, elongation factor Tu (EF-Tu), and GTP in the presence of elongation factor G (EF-G). Following peptide-bond formation and EF-G catalyzed translocation, the complexes were purified over sucrose cushions to remove unincorporated factors. We assessed the yield of the peptidyl-transfer reaction using an electrophoretic TLC system to follow the release reaction ((11); **Figure A1.6B**).

Uncatalyzed Hydrolysis of Dipeptidyl tRNAs On and Off the Ribosome

Previous experiments showed that the efficiency of nonsense suppression in *E. coli* is dependent on the amino acid prior to the stop codon (ultimate amino acid; (25)), suggesting that RF-mediated release is influenced by the identity of the C-terminal amino acid. Before exploring this hypothesis in our in vitro system, we sought to determine the rates of nonenzymatic release on all 20 dipeptidyl-tRNAs to assess whether any differences observed are due to inherent chemical differences among amino acids.

Nonenzymatic hydrolysis and aminolysis off the ribosome were initiated by diluting the complexes in the presence of EDTA, which facilitates splitting of the ribosomal subunits and the dissociation of the dipeptidyl-tRNA. To increase the rate of release, the pH of the dilution buffer was increased to 8.0 using Tris. Fortuitously, in these assays, Tris serves as an analog for peptide bond formation, because similar to the a-amine group of aa-tRNA, its amine group attacks the peptidyl-tRNA bond to form an amide bond **(Figures A1.1A and A1.1B)**. The rates we observed for hydrolysis and aminolysis **(Figures A1.1C and A1.1D)** were generally similar to one another for the same dipeptides, with one exception. Aminolysis rates were higher than hydrolysis rates for the fMet-Trp dipeptide. Apart from fMet-Ile, fMet-Lys, fMet-Pro, fMet-Arg, fMet-Thr, and fMet-Val, which displayed overall reduced rates, the rates of hydrolysis/aminolysis were overall similar among the 20 dipeptides. The reduced rate for fMet-Pro can be easily rationalized given the unique structure of proline. Similarly, it appears that having a positive side chain stabilizes the ester bond of the dipeptidyl tRNA for lysine and arginine. The remaining fMet-Ile, fMet-Thr, and fMet-Val all share the common feature of having a branched b-carbon, but beyond that, no discernible similarities are obvious.

The environment in the PTC of the ribosome is very different from that in solution and is likely to affect the accessibility/orientation of the attacking water. We measured the nonenzymatic hydrolysis rates on the ribosome in the absence of release factor. Hydrolysis rates on the ribosome were significantly slower—2to 30-fold but on average 10-fold—than those determined off the ribosome **(Figure A1.1E)**. Similar to the solution rates of hydrolysis, fMet-Pro exhibited the slowest rates on the ribosome. Apart from this similarity, we observe no correlation between the nonenzymatic solution rates to rates on the ribosome **(Figure A1.7A)**, suggesting that the environment in the PTC has an effect on the rate of hydrolysis. However, we note that the nonenzymatic rates on the ribosome, except for proline, do not vary significantly relative to what we observe for those measured

off the ribosome (<4-fold; **Figures A1.1D and A1.1E**). Thus, it appears that the environment in the PTC provides some uniformity to the termination process.

GGQ N^5-Methylation Significantly Increases the Rate of Peptide Release on Proline- and Glycine-Terminating Peptides

Having established that the nonenzymatic rates of release do not vary significantly among the 20 amino acids, we set out to measure the RF-mediated rates of peptide release. Our experiments were motivated by previous in vivo studies, which suggested that peptide release is particularly inefficient on peptides terminating with proline and glycine residues (25). In agreement with these studies, we measured greatly reduced rates of RF1-mediated peptide release for fMet-Pro and fMet-Gly of 0.0050 s^1 and 0.013 s^1 **(Figure A1.2A)**, respectively. Similarly for the RF2 reaction, we measured slow rates of 0.020 s^{-1} and 0.035 s^{-1} **(Figure A1.2B)**. These rates are almost two orders of magnitude slower than those previously reported rates for fMet-Ala **(**(26)**; Figures A1.2C and A1.2D)**. Furthermore, similar to what has been observed for other amino acids (26), the addition of RF3 has no effect on RF1- and RF2-mediated peptide release on proline and glycine residues (**Figure A1.8**; data not shown). As a result, our measurements suggest that peptide release is likely to be a major bottleneck for gene expression on proline- and glycine terminating peptides. In direct contrast to this notion, an *E. coli* proteome-wide analysis shows that these residues are not underrepresented at the C terminus **(Figure A1.8C)**. Therefore, we wondered whether the above rates we measured in vitro do not truly represent the in vivo efficiency of peptide release and that some other factor might be contributing to peptide release.

Recently, it has been shown that, during peptide-bond formation, the elongation factor P (EF-P) increases the rate of peptide-bond formation on ribosomes stalled on dipropyl motifs (27, 28). As a result, we sought to determine whether the factor could also

increase the rate of release of peptides having C-terminal proline and glycine residues. Lys^{34} of *E. coli* EF-P is modified post-translationally through the addition of a hydroxylated lysine residue by three enzymes (29). We generated modified EF-P by coexpressing the modifying enzymes **((27); Figure A1.9A)**, which resulted in high levels of modified protein as judged by liquid chromatography-tandem mass spectrometry (LC-MS/MS) **(Figure A1.9B)**. We further examined the function of the EF-P variants—unmodified, lysylated, and hydroxylysylated—by assessing their effect on the yield of polyproline peptides **(Figure A1.9C)**. Having established that our modified EF-P is active, we next compared the release rates of fMet-Pro and fMet-Gly dipeptides from DRNC in the presence and absence of EF-P. Incubation with EF-P had little to no effect on the release rate of fMet-Pro and fMet-Gly, regardless of the RF added **(Figures A1.3 and A1.9D)**. These observations suggest that, in contrast to peptide-bond formation, EF-P does not significantly promote peptide release on proline and glycine residues.

As discussed earlier, in E. coli, the glutamine residue of the GGQ motif is N^5 methylated by prmC (20). Our experiments, so far, have been conducted with overexpressed release factors **(Figure A1.9E)**, which exhaust prmC and as a result are unmethylated **(Figure A1.9F)**. To address the effect of methylation on peptide release of fMet-Pro and fMet-Gly, we generated methylated RFs by coexpressing the methyltransferase prmC alongside the His-tagged RF. Using this strategy, we obtained fully methylated RF1 (99.8%) and 75% methylated RF2 as assessed by LC-MS/MS **(Figures A1.9F and A1.9G)**.

Although we expected the methylation to stimulate peptide release (22), we were surprised by its extent of stimulation on particular dipeptides. Methylation of RF1 increased the rate of peptide release by 70-fold and 30-fold for glycine and proline, respectively **(Figure A1.2)**. Likewise, for RF2, methylation increased the rate of peptide release by 30and 10-fold for glycine and proline, respectively **(Figure A1.2)**. Similar to the

unmethylated factors, the addition of RF3 had minimal effects on $RF2^{Me}$-mediated release **(Figure A1.8B)**. In contrast, the addition of EF-P to the methylated factors further increased the rate of release by 2- to 5-fold, bringing the rates of peptide release to 1 s^1 **(Figures A1.2 and A1.9D)**. Collectively, these findings suggest that methylation is critical for peptide release on a subset of amino acids.

Effects of GGQ N^5-Methylation on Peptide Release on All Amino Acids

Our data on peptide release with glycine- and proline-terminating peptides suggest that methylation of release factor may play a larger role in catalysis than previously appreciated (22) and is likely to be specific to certain amino acids. To test this, we measured the k_{hyd} values for the remaining 18 amino acids by incubating our DRNCs with saturating concentrations of RF1 and RF2. In agreement with previous in vivo studies (25), release rates were significantly influenced by the identity of the terminal amino acid. The rates vary by more than 500-fold for RF1-mediated and by more than 800-fold for RF2-mediated release **(Figures A1.3A and A1.3B)**. We note that, although rates of peptide release by both factors on glycine and proline were the obvious outliers, rates on other amino acids were also relatively slow. This variation in the rates of peptide release is in direct contrast to what we observe for the nonenzymatic reactions, for which we measure a maximum of 10-fold difference in the rate of hydrolysis **(Figure A1.1E)**. These data suggest that the release factor significantly reshapes the environment in the PTC and perhaps alters the reaction pathway. In agreement with this proposal, rates for RF1-mediated release do not correlate well to those of nonenzymatic release (R = 0.332; **Figure A1.7B**). RF2-mediated release correlates a little better to nonenzymatic release (R = 0.562; **Figure A1.7C**). This is consistent with recent data from our laboratory showing that, whereas RF1 and RF2 share many similarities, under certain conditions, the factors interact differently with the ribosome (30). Nevertheless, statistically significant correlation

(R = 0.544; p = 0.01) was observed between RF1 and RF2 k_{hyd} values **(Figure A1.7D)**. In summary, our observations suggest that, in the absence of methylation, peptide release varies significantly and could potentially affect gene expression.

This disparity in the rates of peptide release appears to be alleviated by the methylation of the factors. GGQ N^5 methylation increased release rates for both factors **(Figures A1.3A and A1.3B)**. Similar to our earlier analysis, we observe little to no correlation between the nonenzymatic rates and those of RF^{Me}-mediated release **(Figures A1.7E and A1.7F)**. We do observe, however, significant correlation between rates of release for $RF1^{Me}$ and $RF2^{Me}$ (R = 0.651; p = 0.0014; **Figure A1.7G**). Perhaps more interesting was the observation of a correlation between the effect of RF methylation on dipeptide release and the release rate observed for the unmethylated RF **(Figure A1.3D)**. DRNCs that exhibited Negative correlation between release rate for unmethylated RF and increase in release rate due to GGQ N^5 methylation, consistent with the notion that methylation has a more significant effect on slow release reactions.

Slow dipeptide release rates for RF1 and RF2 generally exhibited disproportionately high release rates for $RF1^{Me}$ and $RF2^{Me}$, whereas those that exhibited high release rates for RF1 and RF2 exhibited more-modest increases for $RF1^{Me}$ and $RF2^{Me}$. For instance, methylation of RF2 increased the rate of fMet-Ser release by nearly 40-fold (0.07 s^1 to 2.5 s^1) but had little effect on the rate of fMet-Tyr release (10 s^1 to 14 s^1). In addition, dipeptide release rates varied less between DRNCs for $RF1^{Me}$and $RF2^{Me}$-mediated release than for RF1 and RF2 **(Figure A1.3C)**. The coefficient of variation (31) for dipeptide hydrolysis rates was 1.215 for RF1, 1.164 for $RF1^{Me}$, 1.599 for RF2, and 0.613 for $RF2^{Me}$. This is also visible in the condensed vertical spread of data points for $RF1^{Me}$ and $RF2^{Me}$. These observations suggest that the GGQ N^5 methylation plays an important role in tuning the environment of the PTC during peptide release to ensure that different peptidyl-tRNA substrates are released with somewhat similar efficiency.

Structural Insights into Catalysis by RF1^{Me} and RF2^{Me}

To understand the structural basis of the methylated glutamine, we next determined four structures of *E. coli* unmethylated and methylated RF1 and RF2 bound to the *Tth* 70S ribosome. In all previous structures of *Tth* RFs bound to the 70S, domain three closely packs against 23S rRNA nucleotides of the PTC **(Figure A1.4A;** (13, 14, 32, 33)**)**. The universally conserved GGQ motif is located at the tip of domain three, and the glutamine (Q235 in RF1; Q252 in RF2) backbone amide is located in an equivalent position as the 2^0 end of the A-site tRNA as previously reported **(Figures A1.4B and A1.5;** (13, 14, 32, 33)). The structures of the 70S-RF1^{Me} with and without the Q235 methylation reveal that the overall location of the backbone of the GGQ motif is unaffected by the modification. Surprisingly, in the 70SRF1^{Me} complex, the side-chain amide of glutamine in RF1^{Me} is positioned facing A76 of the tRNA and packs against 23S rRNA A2451 and C2452 **(Figure A1.4B)**. This positioning is clear based on our electron density maps and leaves positive difference density if rotated 180 with the side-chain carbonyl facing the tRNA instead **(Figure A1.5)**. In the RF2 structure, the methylated Q252 positioning is more ambiguous, likely due to the lower resolution (3.4 Å; **Figure A1.5**). In contrast to the side-chain carbonyl, the side-chain amide has historically been proposed to play little to no role in catalysis (17). Recent studies have suggested that the packing of the methyl group against A2451 of the 23S provides support for the glutamine side-chain amide group to participate in hydrogen bonding with a hydroxide ion (19). However, future studies assessing the role of the methylated glutamine in pretermination states and with different substrates will be required to reconcile the mechanism of termination.

DISCUSSION

During termination of protein synthesis, release factors encounter a diverse set of peptidyl-tRNA substrates. As expected, for the nonenzymatic reaction, this diversity of the leaving group was found to impart a significant effect on the rate of hydrolysis **(Figure A1.1)** that is likely the result of chemical differences among the 20 amino acids. However, on the ribosome, this variation in the rates of hydrolysis is not simply the result of intrinsic chemical differences among the leaving groups, as we observe no correlation between ribosome-independent and dependent reactions **(Figure A1.7A)**. Instead, it appears that the side chain of the C-terminal amino acid, through interactions with the active site of the ribosome, affects the environment of the PTC and hence possibly the reaction pathway.

Release factors play an active role during the course of the hydrolysis reaction and accelerate the rate of the reaction by a factor of 10^5 **(Figures A1.2 and A1.3)**. It is therefore reasonable to assume that variations in the efficiency of peptide release that we observe for the uncatalyzed reactions are irrelevant during cellular protein synthesis. Release factors, in principle, could bring about uniformity to the reaction by reshaping the environment in the PTC, which they are known to do (13-18), and/or by altering the rate-limiting step of the reaction. As a result, we were surprised by the observations that, in the presence of RF1 and RF2, the rates are much more dramatically different relative to the uncatalyzed reaction. For fMet-Pro and fMetGly dipeptidyl tRNAs, the rates of release were so greatly reduced (0.005–0.01 s^1) relative to other dipeptidyl tRNAs that, if the rates we measure reflect the in vivo values, ribosomes terminating on these amino acids are likely to stall. Interestingly, Hayes and colleagues have previously documented that transcripts having proline codons before stop codons induce ssrA tagging in E. coli, indicating that termination following proline residues is inefficient and causes ribosomal stalling (34). In E. coli, 113 and 279 genes terminate with proline and glycine, respectively, and they are in fact not underrepresented, taking into account the overall amino acid

frequencies **(Figure A1.8C)**; unless there are mechanisms in vivo that improve termination efficiency on these amino acids, cellular fitness is likely to be impeded given the large fraction of ribosomes that would be unavailable for translation. This termination inefficiency appears to be partly alleviated by adjusting the sequence context around the C-terminal amino acid. In particular, the penultimate amino acid appears to play a critical role during the hydrolysis reaction as it affects the efficiency of stop-codon readthrough (35). In agreement with these findings, proteome-wide analysis of *E. coli* shows that the same subset of amino acids are either forbidden or significantly underrepresented at the penultimate position for genes terminating with proline **(Figure A1.8D)**. We note, however, all of our in vitro analysis conducted here was done with Met as the penultimate amino acid, which is not significantly underrepresented at this position for genes terminating with proline. Hence, it is highly unlikely that the termination inefficiencies that we document for certain peptidyl-tRNAs are the result of methylation status of release factor plays a critical role in ensuring termination proceeds uniformly, irrespective of the identity of the C-terminal amino acid.

What is clear from our analysis and a previous in vivo survey of nonsense suppression (25) is that the identity of the C-terminal amino acid is important for efficiency of the hydrolysis reaction. Our data suggest that the side chain of the ultimate amino acid is likely to influence the orientation of the catalytic water. Our structures of $RF1^{Me}$ and $RF2^{Me}$ bound to 70S reveals that the backbone placement of the GGQ motif is unaffected by the methylation and preserves the previously noted key interaction of the main-chain amide interaction of the glutamine with the 3'-OH of A76 **(Figures A1.4B and A1.4C)**. Previous models for how the methylation contributes to release argued that only the main-chain amide is responsible for coordinating the water whereas a second model argued that the side-chain carbonyl is also involved in coordinating the water (14-16, 32). A recent alternative hypothesis is that methylation could position the side-chain amide toward the

3^0-OH of A76, where it forms a hydrogen bond with a hydroxide ion (19). Our structures show that the methyl-glutamine of the GGQ motif maintains the main-chain amide interaction with the P-site tRNA. Unexpectedly, the methylation causes the side-chain carbonyl to face away from the active site and, instead, positions the side-chain amide toward the P-site tRNA and A2451 of the 23S **(Figure A1.4B)**. Clearly presenting the sidechain amide adjacent to the P-site tRNA must be beneficial to the hydrolysis reaction, and two possibilities exist. One is that the methylation rigidifies the important glutamine to provide a tight pocket for the catalytic water, regardless of the ultimate amino acid. The other possibility is that the methylation stabilizes the glutamine in the PTC to allow for the side-chain amide to interact with a hydroxide ion in the reaction. Future studies aimed at the determination of pre-termination state structures with different substrates will be required to reconcile this important mechanistic question.

EXPERIMENTAL PROCEDURES

70S tight-couple ribosomes were purified from *Escherichia coli* MRE600 (ATCC29417) as described (36). His-tagged IF1 and IF3 were cloned into pPROEXHtb1 (37). The proteins were purified over Ni-NTA resin. The His-tag was removed by tobacco etch virus (TEV) protease. His-tagged IF2 was overexpressed and purified as described elsewhere (38). The 20 aminoacyl-tRNA synthetases were expressed and purified on Ni-NTA resin as previously described (38). His-tagged EF-Tu and EFG were purified as previously described (39). For both factors the His tag was removed by TEV. RF1 and RF2 (Ala at position 246) were cloned into pETDuet, overexpressed, and purified on Ni-NTA resin. $RF1^{Me}$ and $RF2^{Me}$ were prepared by cloning PrmC into the same pETDuet vectors. *E. coli* $tRNA^{fMet}$ was purchased from ChemBlock; *E. coli* total tRNA was purchased from Roche. mRNAs encoding each of the dipeptide combinations were generated from DNA templates using run-off transcription (40). The mRNAs were gel purified from 10%

PAGE. The final mRNAs had the following sequence: GGGTTAACTTTAGAAGG AGGTAAAAAAA *ATG* X **TAA** CTC GCA TGC CCA CTT GTC GAT CAC CGC CCT TGA TTT GCC CTT CTG T (underlined is the Shine Dalgarno sequence, the initiation codon is italicized, X is the codon encoding the C terminal amino acid and occupies the P site and the stop codon is in bold letters).

tRNA aminoacylation

[^{35}S]-fMet-tRNA^{fMet} was prepared as described (41). Each of the 20 charged tRNA isoacceptors was charged by carrying out aminoacylation of a tRNA mix in the presence of the appropriate tRNA synthetase and amino acid. The complete procedure is described in detail elsewhere (37).

Ribosome Complexes

Initiation complexes were prepared by incubating 3 µM *E. coli* IF1, IF2 and IF3, 3 µM [^{35}S]-fMet-tRNA^{fMet}, 6 µM mRNA, and 2 µM 70s ribosomes at 37°C for 10 minutes in Polymix buffer (95 mM KCl, 5 mM NH_4Cl, 5 mM $Mg(OAc)_2$, 0.5 mM $CaCl_2$, 8 mM putrescine, 1 mM spermidine, 5 mM potassium phosphate pH 7.5, 1 mM DTT) and 2 mM GTP. Ternary complexes were prepared by first incubating 30 µM EF-Tu and 2 mM GTP in Polymix buffer for 5 minutes at 37°C to allow for the exchange of bound GDP to GTP by EF-Tu; 4 µM aa-tRNA and 6 µM EF-G were then added to the EF-Tu·GTP solution. The ternary complex and EF-G mixture was then added to the initiation complex at 37°C for 2 minutes, facilitating the process of elongation. The solution was placed on ice and $MgCl_2$ was added to a final concentration of 10 mM, after which the complex was layered on top of a sucrose cushion solution (1.1 M sucrose, 20 mM Tris pH 7.5, 500 mM NH_4Cl, 10 mM $MgCl_2$, 0.5 mM EDTA) and centrifuged for 2 hours at 69,000 RPM and 4°C. The purified DNCs were resuspended in Polymix buffer. The fraction of radioactivity that

pelleted with the complex was used to calculate the concentration of the complex. We also tested the efficacy of dipeptide formation and premature peptide dissociation from the ribosome by resolving a sample of the complex and a sample treated with 300 mM KOH via TLC electrophoresis (1200 V for 30 minutes), which resolved dipeptides from unreacted fMet.

Preparation of EFP

His-tagged EF-P was prepared by cloning it into a pET-28a vector. yjeA, yjeK and yfcM, which encode proteins that perform modifications on EFP, were sequentially cloned into the same vector downstream of the previous stop codon. The plasmids were transformed into BL21(DE3)pLys cells. EF-P was purified using Ni-NTA chromatography.

RF1 and EF-P LC MS/MS

Mass spec analysis was conducted at the Danforth Plant Science Center proteomics and mass spectrometry facility at Washington University. Protein samples were first reduced with TCEP (10 mM), then alkylated with Iodoacetamide (IAA; 20 mM). 2.5 µg of trypsin was added to each sample and proteins were digested at 37°C overnight. The digests were resuspended in 150 µL of 5% ACN/0.1% FA. 5 µL were injected and run on an LTQOrbitrap Velos Pro using a 1 hour gradient on a Dionex RSLCnano HPLC. All MS/MS samples were analyzed using Mascot (Matrix Science, London, UK; version 2.5.1).

Mascot was set up to first search the cRAP database and the custom database including the sequences provided assuming the digestion enzyme trypsin. Then Mascot was set up to search the cRAP database and the *E. coli* K12 database (from NCBI, 20150201, 4409 entries) assuming the digestion enzyme trypsin. Mascot was searched with a fragment ion mass tolerance of 0.80 Da and a parent ion tolerance of 15 PPM.

Carbamidomethyl of cysteine was specified in Mascot as a fixed modification. Methyl of glutamine and oxidation of methionine were specified in Mascot as variable modifications for the RF1 and RF2 samples Oxidation of methionine, lysinilation (TMAB) of lysine and lysinilation + hydroxylation (iTRAQ4plex115) of lysine were specified in Mascot as variable modifications for the EFP samples. Scaffold (version Scaffold_4.4.3, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Peptide Prophet algorithm (42) with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (43). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters.

RF2 LC MS/MS

E. coli RF2 was added to IP beads, the beads were spun down and residual wash solution was removed. Digestion buffer (200 µl of 50 mM NH_4HCO_3) was added and the bead solution was then treated with 1 mM dithiothreitol (DTT) at 25°C for 30 minutes, followed by 5 mM IAA at 25°C for 30 minutes in the dark. Proteins were digested with 1 µg of lysyl endopeptidase (Wako) at room temperature for 2 hours and further digested overnight with 1:50 (w/w) trypsin (Promega) at room temperature. Resulting peptides were desalted with a Sep-Pak C18 column (Waters) and dried under vacuum. The dried peptides were resuspended in 10 µL of loading buffer (0.1% formic acid, 0.03% trifluoroacetic acid, 1% acetonitrile). Peptide mixtures (2 µL) were separated on a selfpacked C18 (1.9 µm Dr. Maisch, Germany) fused silica column (25 cm x 75 µM internal

diameter (ID); New Objective, Woburn, MA) by a Dionex Ultimate 3000 RSLCNano and monitored on a Fusion mass spectrometer (ThermoFisher Scientific, San Jose, CA). Elution was performed over a 120 minute gradient at a rate of 350 nl/min with buffer B ranging from 3% to 80% (buffer A: 0.1% formic acid in water, buffer B: 0.1 % formic in acetonitrile). The mass spectrometer cycle was programmed to collect at the top speed for 3 second cycles. The MS scans (400-1600 m/z range, 200,000 AGC, 50 ms maximum ion time) were collected at a resolution of 120,000 at m/z 200 in profile mode and the HCD MS/MS spectra (0.7 m/z isolation width, 30% collision energy, 10,000 AGC target, 35 ms maximum ion time) were detected in the ion trap. Dynamic exclusion was set to exclude previous sequenced precursor ions for 20 seconds within a 10 ppm window. Precursor ions with +1, and +8 or higher charge states were excluded from sequencing. MS/MS spectra were searched against a Uniprot curated *E. coli* database (downloaded on 08/14/2015 with 4218 target sequences) with Proteome Discoverer 2.0 (ThermoFisher Scientific, San Jose, CA). Methionine oxidation (+15.9949 Da), asparagine and glutamine deamidation (+0.9840 Da), glutamine methylation (14.01565) and protein Nterminal acetylation (+42.0106 Da) were variable modifications (up to 3 allowed per peptide); cysteine was assigned a fixed carbamidomethyl modification (+57.0215 Da). Only fully tryptic peptides were considered with up to 2 miscleavages in the database search. A precursor mass tolerance of ±20 ppm and a fragment mass tolerance of 0.6 Da were applied. Spectra matches were filtered by Percolator to a psm fdr of less than 1 percent. Peptide level data was compiled with in-house written software.

70S complex formation and crystallization

Tth 70S ribosomes were purified as described (44). 70S ribosomes (4.4 µM) were incubated with twofold excess of mRNA (IDT, 8.8 µM; 5'- GGCAAGGAGGUAAAA AUG UAA CAGA -3'; Shine Dalgarno underlined) for 5 minutes at 55°C. A 2.5 molar excess of

E. coli $tRNA^{fMet}$ to fill the P site (Chemical Block; 11 µM) was incubated for 30 minutes at 55°C followed by 10 minutes at 37°C.

Either 2.2 molar excess of *E. coli* $RF1^{Me}/RF2^{Me}$ (9.7 µM) or a 3.2 molar excess of *E. coli* RF1/RF2 (14.1 µM) was added to the A site at 37°C for 30 minutes. Crystallization trials were performed as previously described in 100 mM Tris-Acetate pH 7.0, 200 mM KSCN, 10 mM MgCl2, 3.5-6% polyethylene glycol (PEG) 20K, 6-12% PEG 200 and 2.8 µM deoxy BigCHAP (Hampton Research) was added to the complex immediately before crystallization (16, 44). The crystals were cryoprotected by increasing the PEG concentrations to 7% PEG 20K with stepwise increases in PEG 200 (10% 12% 14%) and 2-methyl-2-4-pentanediol (MPD) (10%, 18%, 25%). The final cryoprotectant was introduced to the crystal wells twice before freezing in liquid nitrogen.

X-ray data collection and structure determination

Crystal were screened for diffraction at the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline and X-ray data were collected at the Northeastern Collaborative Access Team (NE-CAT) 24-IDC beamline, both beamlines at the Advanced Photon Source, Argonne National Laboratory. The data were integrated and scaled using the XDS software package (45). A search model composed of the *Tth* 70S ribosome (PDB code 4Y4O) with YfiA and waters removed was used for crystallographic refinement with the Phenix software suite (46). The initial model was place by rigid body refinement using the 30S and 50S as rigid groups. RF1 was placed using the unbound *E. coli* RF1 (PDB code 2B3T) structure and a *Tth* RF1 70S (PDB code 4V63) as a guide to place the domains. RF2 was placed similarly using unbound *E. coli* RF2 (PDB code 1GQE) and *Tth* RF2 70S (PDB code 4V5J). Modeling of mRNA and $tRNA^{fMet}$ along with conformational changes in rRNA and RFs were performed using Coot (47). Iterative rounds of model building were followed by positional and restrained individual

atomic displacement parameter (48) refinement along with Translation/Liberation/Screw refinement in Phenix, yielding final models with the statistics reported in Table S1. Figures were generated using PyMOL (49).

ACCESSION NUMBERS

The accession numbers for the coordinates and structure factors reported in this paper are PDB: 5CZP, PDB: 5DFE, PDB: 5J30, and PDB: 5J3C.

ACKNOWLEDGMENTS

The authors wish to thank Joe Jez and Doug Chalker for comments on an earlier version of the manuscript and members of the H.S.Z. and C.M.D. laboratories for useful discussions. This work was supported by the NIH (NIH R01GM093278 to C.M.D. and R01GM112641 to H.S.Z.), a Searle Scholars award (to H.S.Z.), and a Pew Scholar in the Biomedical Sciences award (to C.M.D.). The Donald Danforth Plant Science Center Proteomics and Mass Spectrometry Facility is supported the National Science Foundation grant no. DBI-0922879 for acquisition of the LTQ-Velos Pro Orbitrap LC-MS/MS. The Emory Integrated Proteomics Core is supported by the Neuroscience NINDS Core Facilities (P30NS055077), the Emory University School of Medicine, and is one of the Emory Integrated Core Facilities. The X-ray crystallography datasets were collected at the NE-CAT beamlines, which are funded by the National Institute of General Medical Sciences from the NIH (P41 GM103403), and at the SER-CAT beamlines. The Pilatus 6M detector on 24ID-C beam line is funded by a NIH-ORIP HEI grant (S10 RR029205). This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC0206CH11357

FIGURES AND TABLES

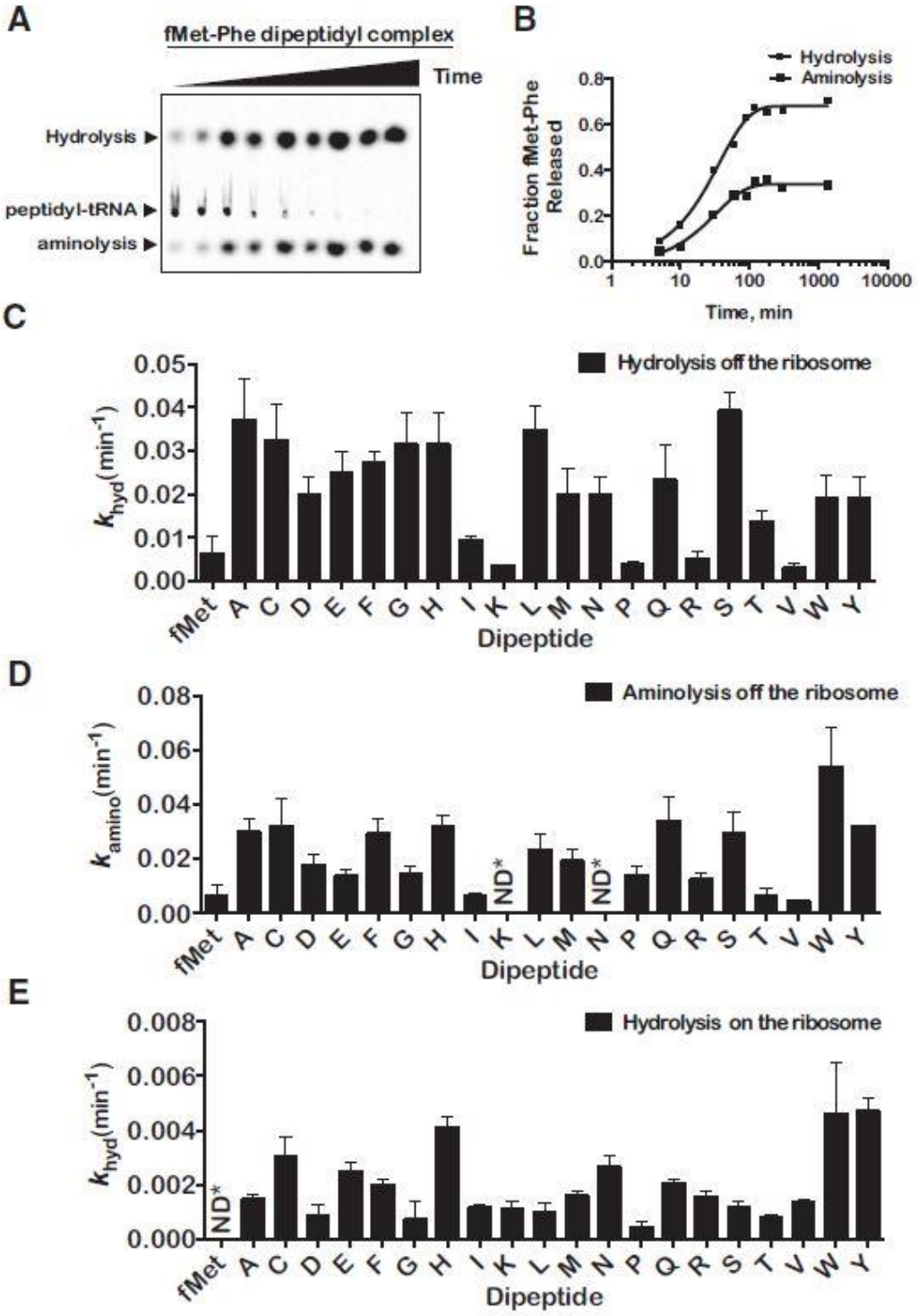

Figure A1.1 Nonenzymatic Dipeptide Hydrolysis and Aminolysis

(A) Representative electrophoretic TLC of a time course of nonenzymatic hydrolysis and aminolysis from free fMet-Phe peptidyl tRNA. (B) Quantification of data in (A). Fractional radioactivity corresponding to the released product was plotted against time and fit to one-phase exponential association equation. (C) Dipeptide release rates calculated for hydrolysis of free peptidyl-tRNA. (D) Dipeptide release rates calculated for aminolysis of free peptidyl-tRNA. Note the general similarity between hydrolysis and aminolysis rates. (E) Dipeptide release rates calculated for nonenzymatic hydrolysis of ribosomal peptidyl-tRNA. The error bars represent the SE obtained from the nonlinear regression. Most of the reactions were done in duplicates, and the deviation was less than 20% between the observed rates.

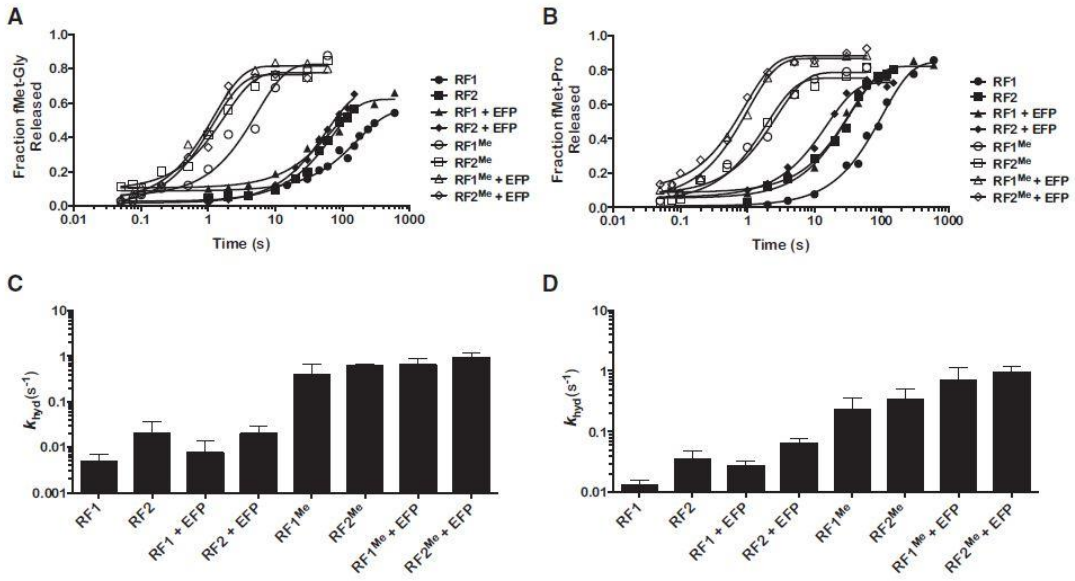

Figure A1.2 RF-Mediated Peptide Release of fMet-Glycyl and fMet-Prolyl Dipeptidyl tRNAs

(A and B) Representative time courses of (A) fMet-Gly and (B) fMet-Pro release in the presence of the indicated factors. For both dipeptides, only the addition of methylated RF1 or RF2 results in a significant increase in the rate of peptide release. (C and D) fMet-Gly (C) and fMet-Pro (D) dipeptide release rates under the indicated conditions. Error bars represent SE of measurement calculated from at least three replicates

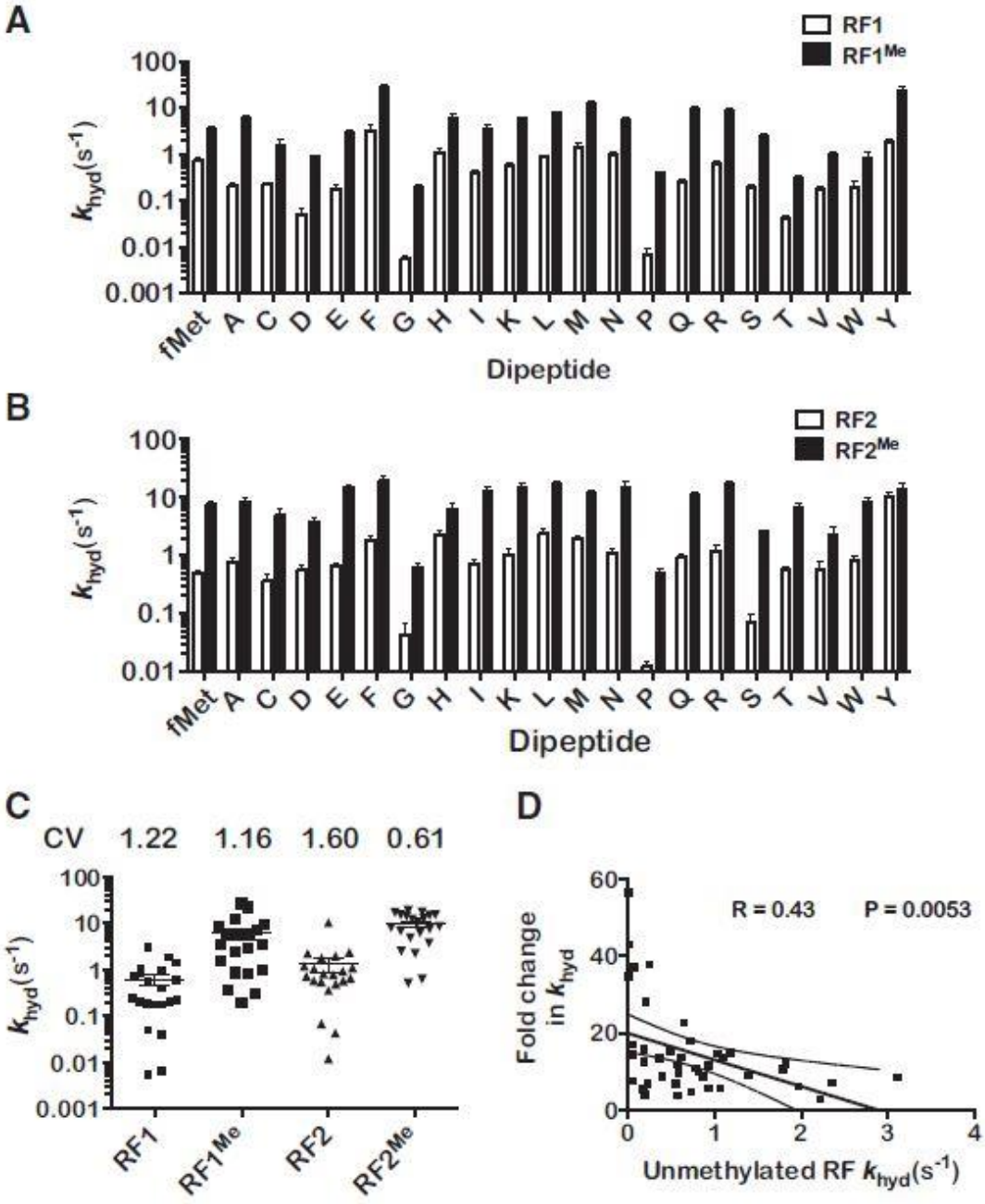
A
RF1
RF1Me
100
10
1
0.1
0.01
0.001
khyd(s-1)
fMet A C D E F G H I K L M N P Q R S T V W Y
Dipeptide
B
RF2
RF2Me
100
10
1
0.1
0.01
khyd(s-1)
fMet A C D E F G H I K L M N P Q R S T V W Y
Dipeptide
C
CV 1.22 1.16 1.60 0.61
100
10
1
0.1
0.01
0.001
khyd(s-1)
RF1
RF1Me
RF2
RF2Me
D
R = 0.43 P = 0.0053
Fold change in khyd
60
40
20
0
0 1 2 3 4
Unmethylated RF khyd(s-1)

Figure A1.3 Effect of RF Methylation on Dipeptide Release Rate

(A and B) Comparison of dipeptide release rates for (A) RF1 and $RF1^{Me}$ and (B) RF2 and $RF2^{Me}$. Note the near universal rate increase for methylated RFs. The error bars represent the SE obtained from the nonlinear regression. For most dipeptides, reactions were carried out at least in duplicates and the observed variation was less than 20% between the rates. (C) Variation in dipeptide release rates for each release factor. Each data point corresponds to the release rate of a different dipeptidyl ribosome complex. Numbers above plot indicate the coefficient of variation (CV) for the release rates for each RF. (D) Negative correlation between release rate for unmethylated RF and increase in release rate due to GGQ N5 methylation, consistent with the notion that methylation has a more significant effect on slow release reactions.

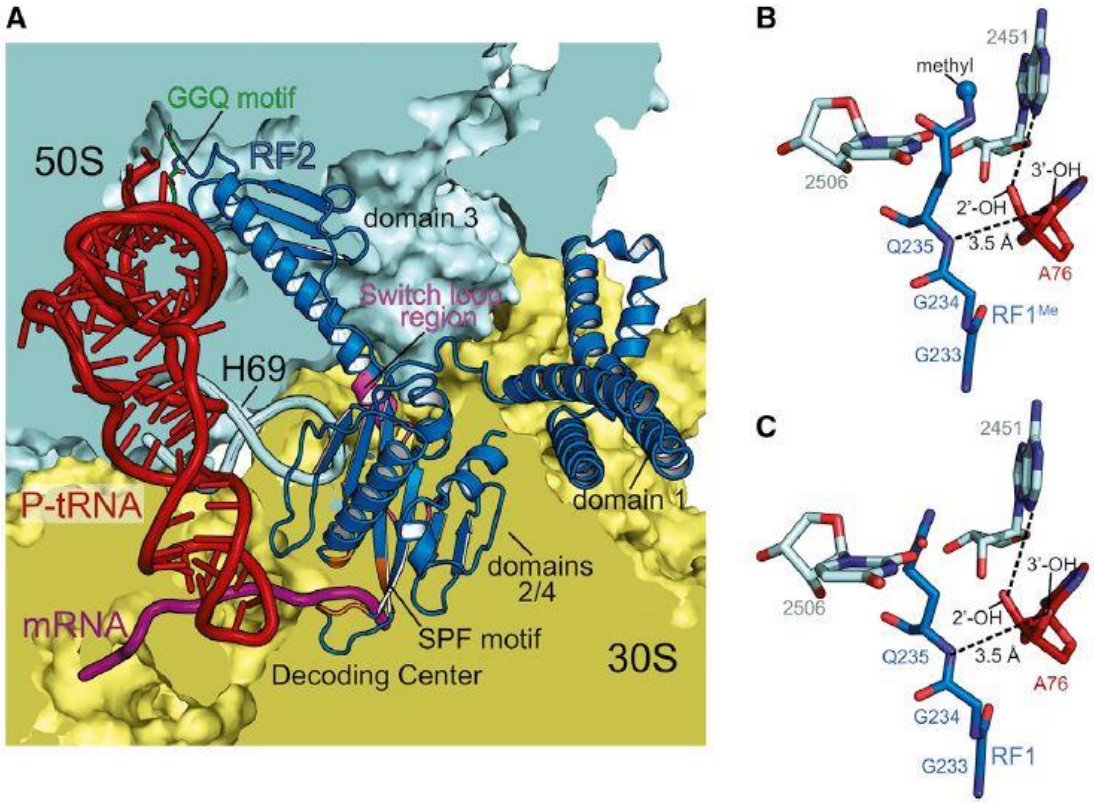

Figure A1.4 70S-RFMe Interactions

(A) Overview of *E. coli* RF2^{Me} (blue) interactions with the *Thermus thermophilus* 30S (yellow) and 50S (blue) subunits programmed with P-site tRNAfMet (red) and mRNA (magenta). (B) Interactions between RF1^{Me} GGQ motif (blue), the peptidyl transferase center 23S rRNA residues 2506 and 2452 (cyan), and the terminal A76 of P-site tRNAfMet (red). The backbone amide of Q235 is within hydrogen bond distance to the 30-OH of A76. The side-chain methyl group (blue sphere) packs against the nucleobase of 2451, placing its side chain positioned toward A76. (C) Interactions between RF1 GGQ motif. The backbone amide is still within hydrogen bond distance to the 30-OH of A76, but the glutamine side chain no longer interacts with 2451. The color scheme is the same as (B).

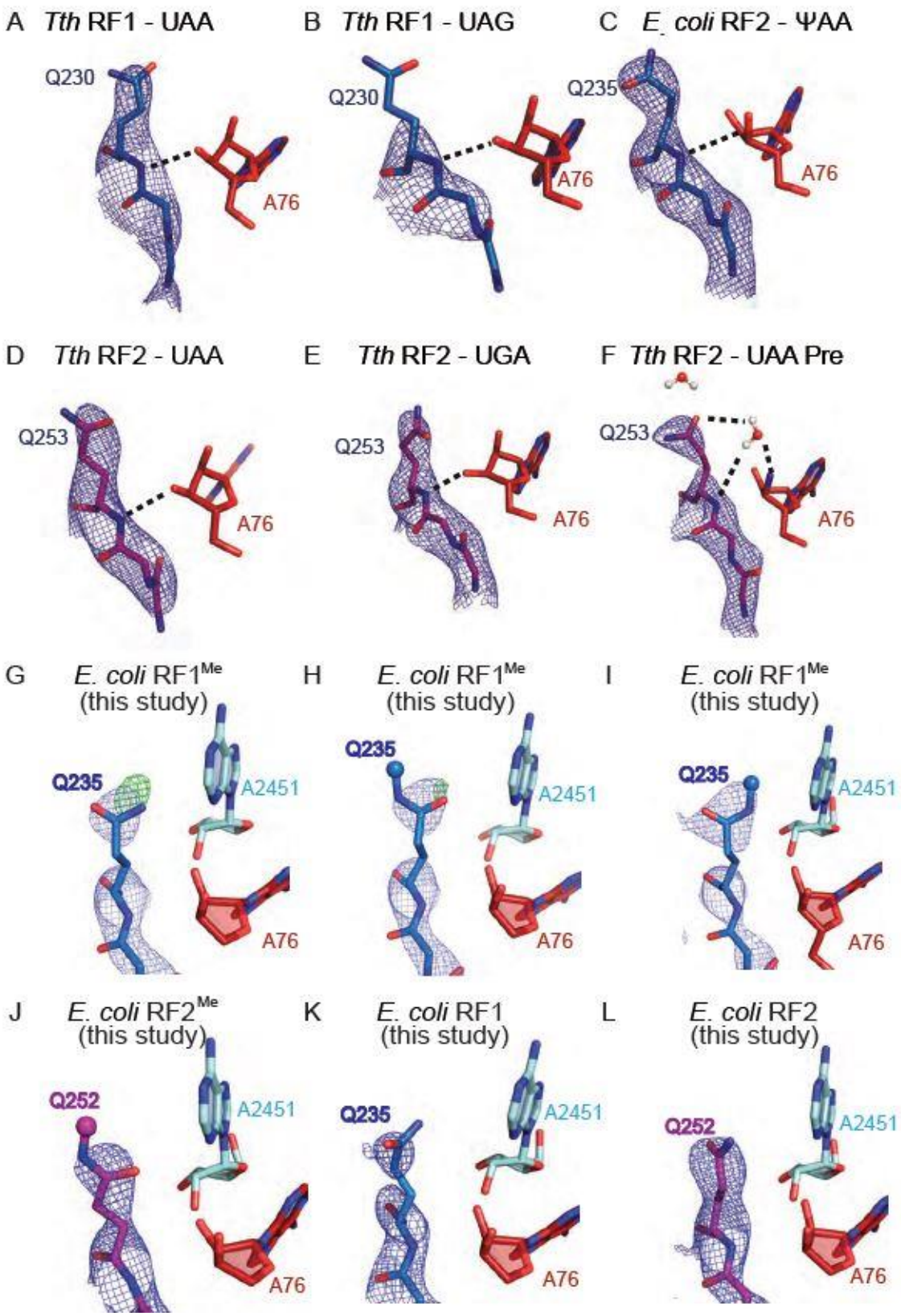
A *Tth* RF1 - UAA
Q230
A76
B *Tth* RF1 - UAG
Q230
A76
C *E. coli* RF2 - ΨAA
Q235
A76
D *Tth* RF2 - UAA
Q253
A76
E *Tth* RF2 - UGA
Q253
A76
F *Tth* RF2 - UAA Pre
Q253
A76
G *E. coli* RF1^{Me} (this study)
Q235
A2451
A76
H *E. coli* RF1^{Me} (this study)
Q235
A2451
A76
I *E. coli* RF1^{Me} (this study)
Q235
A2451
A76
J *E. coli* RF2^{Me} (this study)
Q252
A2451
A76
K *E. coli* RF1 (this study)
Q235
A2451
A76
L *E. coli* RF2 (this study)
Q252
A2451
A76

Figure A1.5 Interactions of Q252 with A76 of tRNA in different 70S-RF1 and RF2 structures.

The glutamine of the conserved GGQ motif as shown in the: (A) 70S - *Tth* RF1 structure (PDB code 4V63); (B) 70S - *Tth* RF1 structure (PDB code 4V7P); (C) 70S – *E. coli* RF1 structure (PDB code 5J4D); (D) 70S - *Tth* RF2 structure (PDB code 4V67); (E) 70S - *Tth* RF2 structure (PDB code 4V5E); (F) 70S - *Tth* RF2 structure (PDB code 4V5J). (G) Zoomed in view of the $RF1^{Me}$ Q235 side chain (blue). Crystallographic refinement was initially performed without the methyl side chain modification and resulting $2F_oF_c$ electron density (blue; I/σ = 1.0) and F_o-F_c difference electron density (green; I/σ = 2.0) maps are shown. Nucleotides 2451 from 23S rRNA (cyan) and A76 from tRNA (red) are shown as sticks. (H) The methyl amine side chain in the 70S- $RF1^{Me}$ structure was modeled facing away from A76 of tRNA with the carbonyl facing towards the 3'-OH. Crystallographic refinement with this model was performed and both electron density maps ($2F_o$-F_c maps; blue; I/σ = 1.0; and F_o-F_c difference maps; green; I/σ = 2.0) are shown. (I) Final model of $RF1^{Me}$ Q235 with the methyl amine side chain modeled facing A76 of the tRNA. $2F_o$-F_c electron density maps are shown (blue; I/σ = 1.0). No F_o-F_c electron difference density was observed at I/σ = 2.0. (J) Final model of the $RF2^{Me}$ Q252 side chain (purple) shown with $2F_o$-F_c electron density (blue; I/σ = 1.0). (K) Final model of the unmethylated RF1 Q252 side chain (blue) shown with $2F_o$-F_c electron density (blue; I/σ = 1.0). (L) Final model of the unmethylated RF2 Q252 side chain (purple) shown with $2F_o$-F_c electron density (blue; I/σ = 1.0). No F_o-F_c electron difference density was observed around the glutamine sidechain for panels C-F (I/σ = 2.0). All structures are in a post hydrolysis state except for (F).

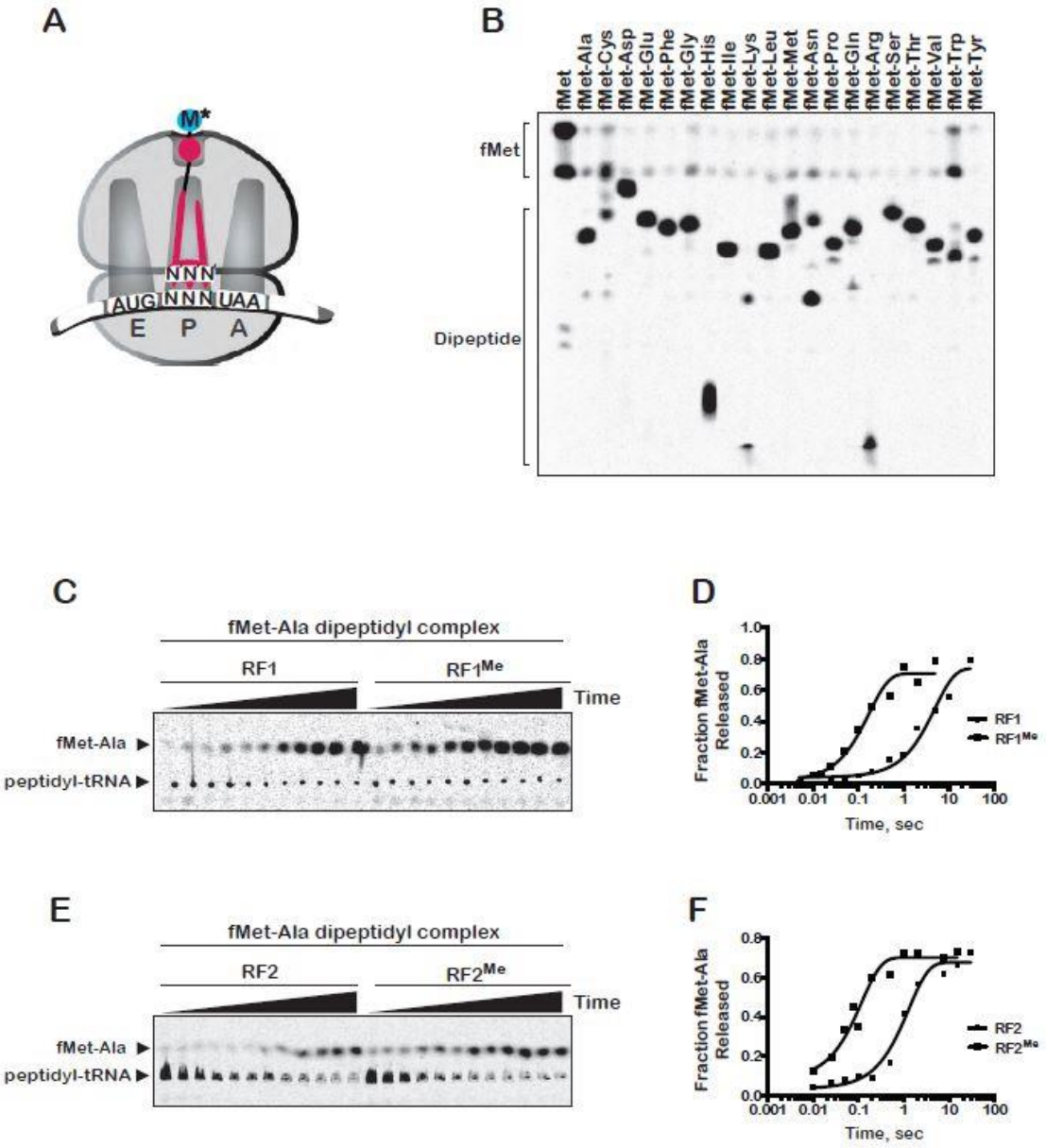
A
M*
NNN
AUGNNNUAA
E P A
B
fMet
fMet-Ala
fMet-Cys
fMet-Asp
fMet-Glu
fMet-Phe
fMet-Gly
fMet-His
fMet-Ile
fMet-Lys
fMet-Leu
fMet-Met
fMet-Asn
fMet-Pro
fMet-Gln
fMet-Arg
fMet-Ser
fMet-Thr
fMet-Val
fMet-Trp
fMet-Tyr
Dipeptide
C
fMet-Ala dipeptidyl complex
RF1
RF1Me
Time
fMet-Ala
peptidyl-tRNA
D
Fraction fMet-Ala Released
RF1
RF1Me
Time, sec
E
fMet-Ala dipeptidyl complex
RF2
RF2Me
Time
fMet-Ala
peptidyl-tRNA
F
Fraction fMet-Ala Released
RF2
RF2Me
Time, sec

Figure A1.6 Complex formation is nearly quantitative in the reconstituted system

(A) Graphical illustration of dipeptidyl (fMet-X-Stop) ribosome complexes. (B) Electrophoretic TLC resolution of all 20 dipeptidyl-tRNA complexes and fMet initiation complexes. Complexes were treated with KOH and electrophoresed on cellulose TLC for 30 minutes at 1200 V. (C) Phosphorimager scan of an electrophoretic TLC for release time courses of fMet-Ala RNC in the presence of unmethylated and methylated RF1. (D) A plot of fractional radioactivity corresponding to released dipeptides as a function of time for the times courses shown in (C). The data was fit to a first-order rate function. (E) Similar to (C), but in the presence of unmethylated and methylated RF2. (F) Similar to (D), but the fit corresponds to the data presented in E.

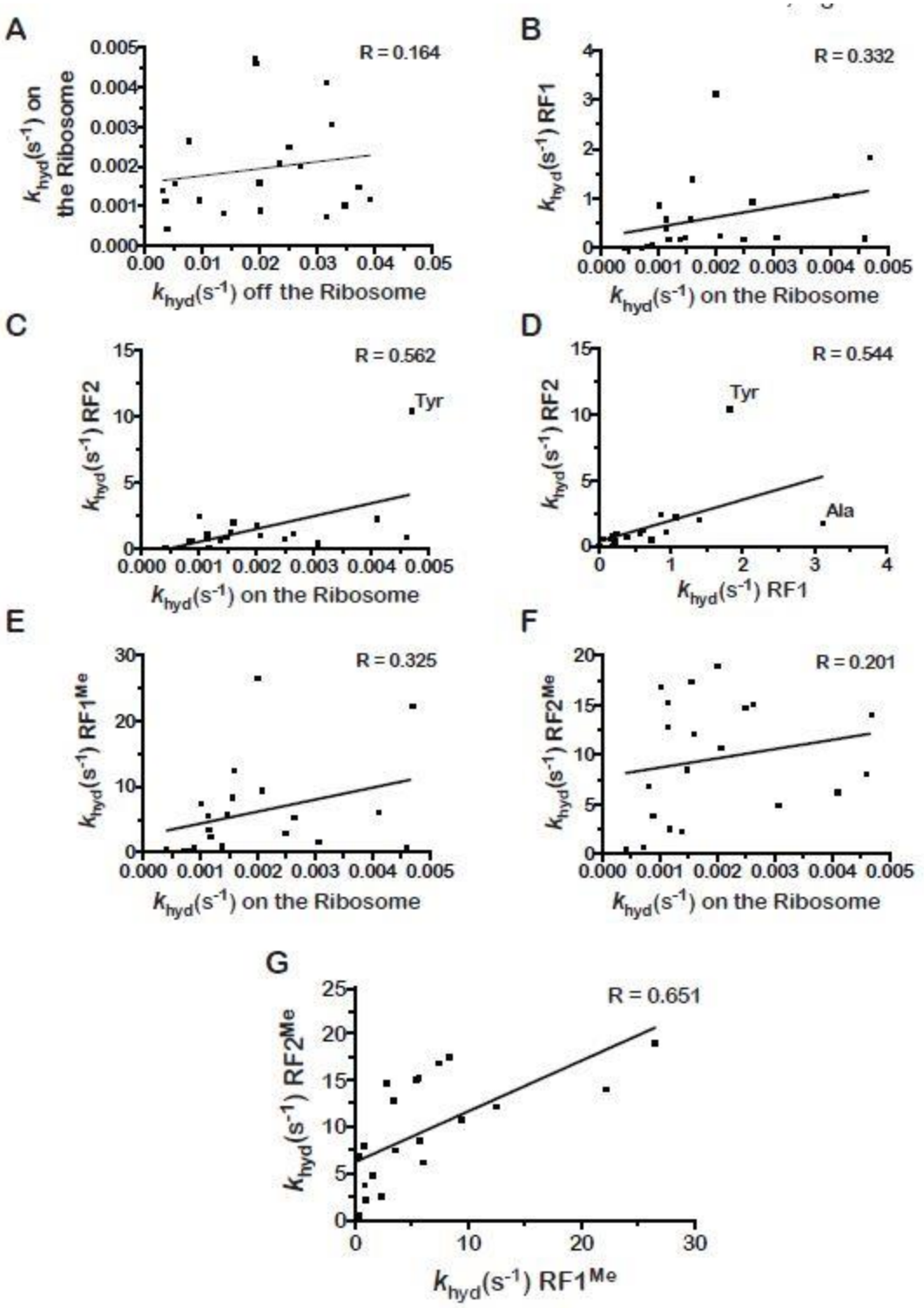
A
R = 0.164
$k_{hyd}(s^{-1})$ on the Ribosome
$k_{hyd}(s^{-1})$ off the Ribosome
B
R = 0.332
$k_{hyd}(s^{-1})$ RF1
$k_{hyd}(s^{-1})$ on the Ribosome
C
R = 0.562
Tyr
$k_{hyd}(s^{-1})$ RF2
$k_{hyd}(s^{-1})$ on the Ribosome
D
R = 0.544
Tyr
Ala
$k_{hyd}(s^{-1})$ RF2
$k_{hyd}(s^{-1})$ RF1
E
R = 0.325
$k_{hyd}(s^{-1})$ $RF1^{Me}$
$k_{hyd}(s^{-1})$ on the Ribosome
F
R = 0.201
$k_{hyd}(s^{-1})$ $RF2^{Me}$
$k_{hyd}(s^{-1})$ on the Ribosome
G
R = 0.651
$k_{hyd}(s^{-1})$ $RF2^{Me}$
$k_{hyd}(s^{-1})$ $RF1^{Me}$

Figure A1.7 Correlations of release rates.

(A) Plot of RF-independent rates of hydrolysis on the ribosome for the 20 dipeptides and fMet versus those measured off the ribosome. (B) Similar to (A) but for unmethylated RF2 versus unmethylated RF1. (C) and (D) Plot of RF1- and RF2-mediated release rates (respectively) versus RF independent rates. (E) and (F) Similar to (C) and (D)but for methylated RF1 and RF2, respectively. (G) Plot of $RF2^{Me}$-mediated release rates versus $RF1^{Me}$-mediated ones.

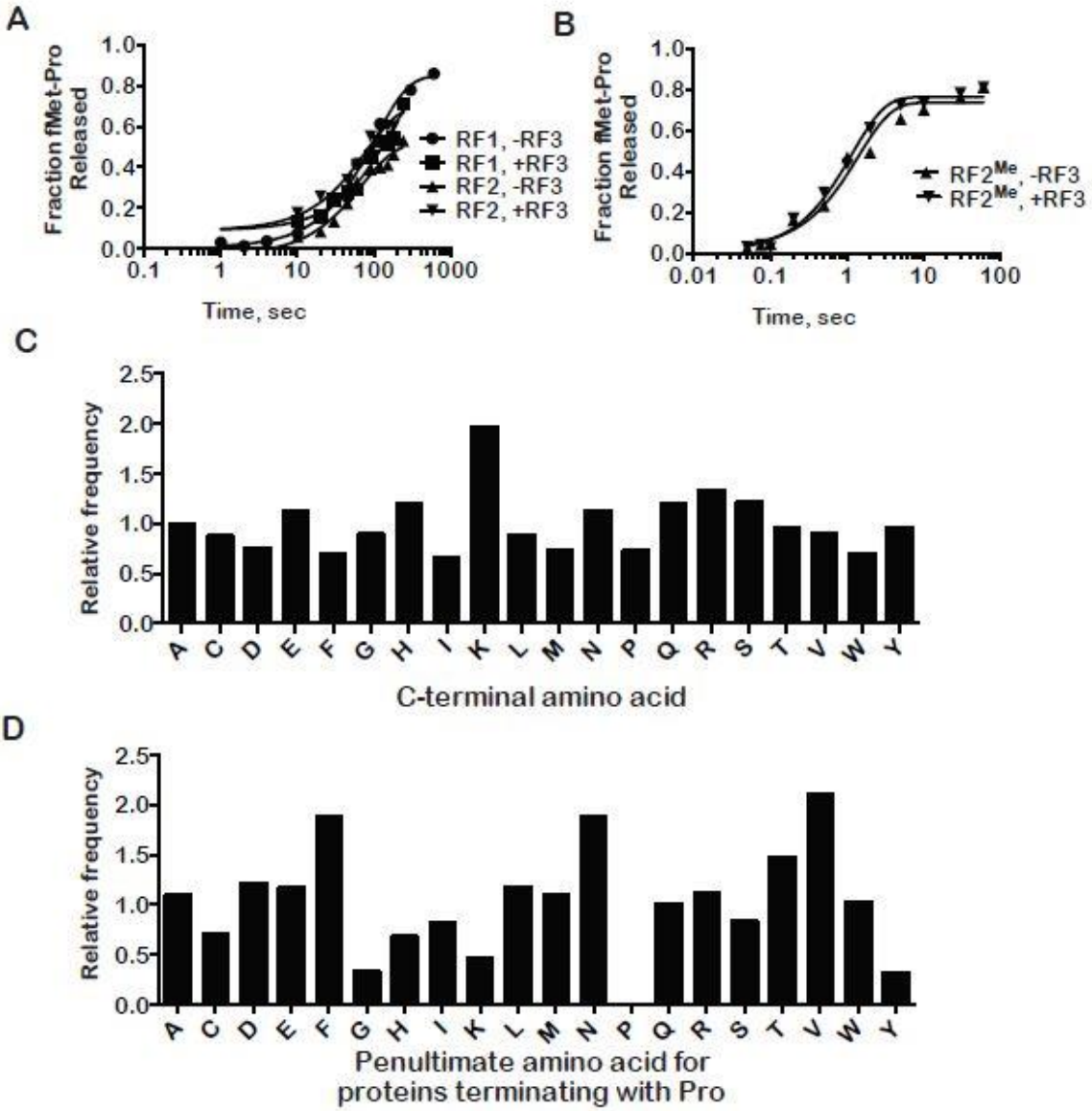
A
Fraction fMet-Pro Released
1.0
0.8
0.6
0.4
0.2
0.0
0.1
1
10
100
1000
Time, sec
RF1, -RF3
RF1, +RF3
RF2, -RF3
RF2, +RF3
B
Fraction fMet-Pro Released
1.0
0.8
0.6
0.4
0.2
0.0
0.01
0.1
1
10
100
Time, sec
RF2Me, -RF3
RF2Me, +RF3
C
Relative frequency
2.5
2.0
1.5
1.0
0.5
0.0
A C D E F G H I K L M N P Q R S T V W Y
C-terminal amino acid
D
Relative frequency
2.5
2.0
1.5
1.0
0.5
0.0
A C D E F G H I K L M N P Q R S T V W Y
Penultimate amino acid for
proteins terminating with Pro

Figure A1.8 RF3 has no effect on RF1- and RF2-mediated release of fMet-Pro and relative occurrence of different amino acids as the ultimate and penultimate residue in *E. coli*

(A) Time course of fMet-Pro dipeptide release in the presence of the indicated factors. The rates for RF1 are 0.009 s^{-1} and 0.006 s^{-1} in the absence and presence of RF3, respectively. The rates for RF2 are 0.014 s^{-1} and 0.013 s^{-1} in the absence and presence of RF3, respectively. (B) Time-course of $RF2^{Me}$-mediated fMet-Pro release in the absence and presence of RF3. In the absence of RF3, k_{hyd} = 0.7 s^{-1}; in the presence of RF3, k_{hyd} = 0.9 s^{-1}. (C) Bar graph showing the relative frequency of proteins ending with the depicted amino acid. Relative frequencies were determined by taking the ratio of the relative occurrence of that amino acid as the terminal reissue to that of its occurrence in the whole proteome. (D) Bar graph showing the penultimate amino acid relative frequencies for proteins terminating with proline. The values were determined by taking the ratio of the frequency a particular amino acid occurred at that position for proline-terminating proteins relative to its occurrence as penultimate amino acid in the whole proteome.

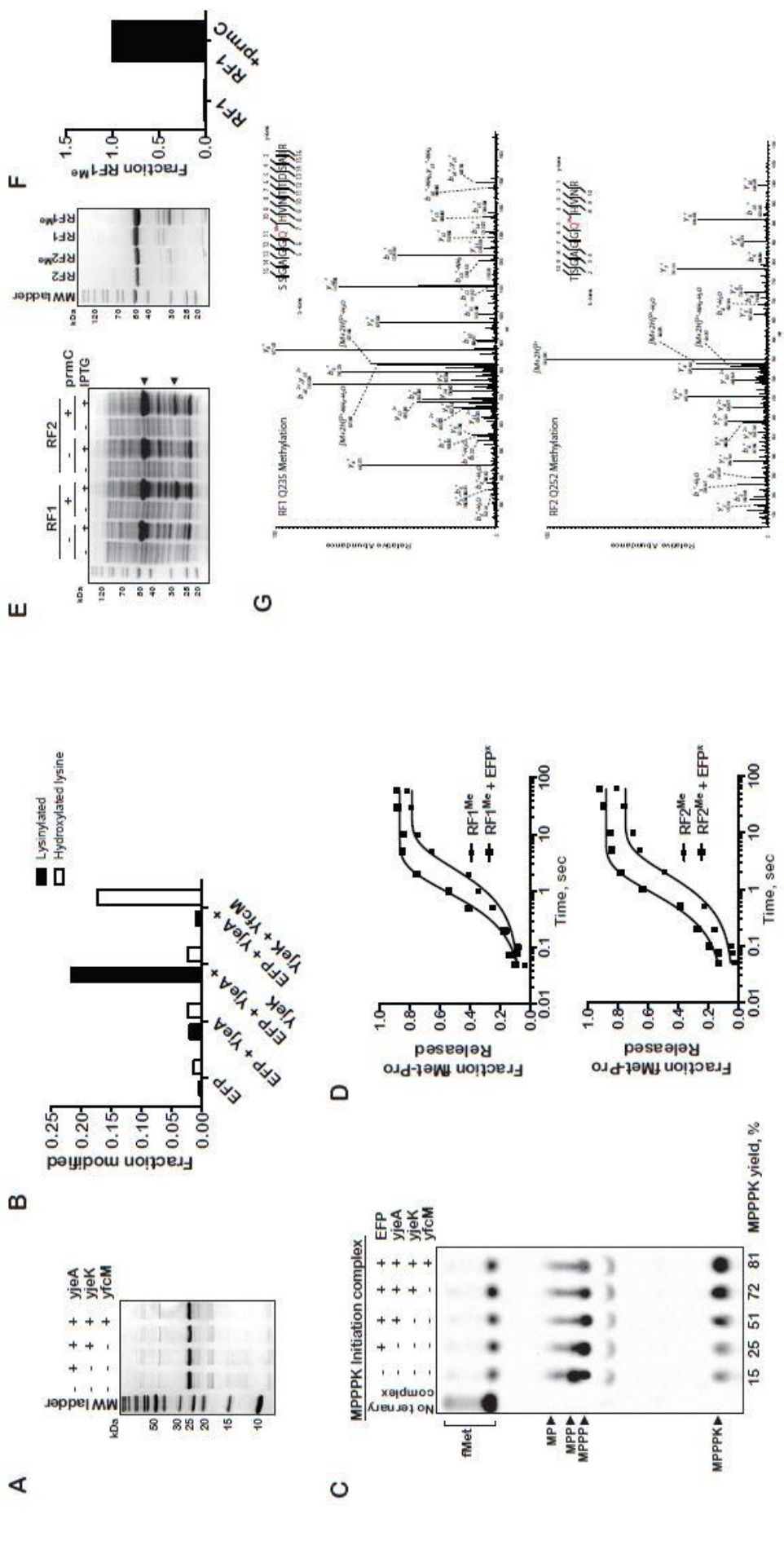
A
MW ladder
yjeA
yjeK
yfcM
kDa
50
30
25
20
15
10
B
Lysinylated
Hydroxylated lysine
Fraction modified
0.25
0.20
0.15
0.10
0.05
0.00
EFP
EFP + YjeA
EFP + YjeA + YjeK
EFP + YjeA + YfcM
YjeK + YfcM
C
MPPPK Initiation complex
No ternary complex
EFP
yjeA
yjeK
yfcM
fMet
MP
MPP
MPPP
MPPPK
15 25 51 72 81
MPPPK yield, %
D
Fraction fMet-Pro Released
RF1Me
RF1Me + EFPx
RF2Me
RF2Me + EFPx
Time, sec
E
RF1
RF2
prmC
IPTG
MW ladder
RF2
RF2Me
RF1
RF1Me
F
Fraction RF1Me
RF1
RF1 +prmC
G
RF1 Q235 Methylation
RF2 Q252 Methylation
Relative Abundance

Figure A1.9 MS/MS analysis of EFP and RFs and their effect on peptide-bond formation and peptide release.

(A) SDS-PAGE analysis of purified EFP proteins isolated in the presence of the indicated modification enzymes. (B) Bar graph showing the extent of lysinylation plus hydroxylation for the different purified proteins. (C) Electrophoretic TLC of *in vitro* translation of poly-Pro motif in the presence of successively modified EF-P protein. Note the increase in full-length peptide yield with successive EF-P modifications. (D) Time courses of dipeptide release from fMet-Pro dipeptidyl ribosome complex for the indicated RF in the absence or presence of modified EF-P. (E) SDS-PAGE analysis of RF1 and RF2 overexpression in the absence and presence of prmC (left panel). SDS-PAGE analysis of purified factors (right panel). (F) Bar graph showing the extent of N5-methylation of the GGQ motif of RF1. (G) Representative MS/MS spectra are shown for the RF1 doubly charged (M+2H) Q235 methylated and RF2 doubly charged (M+2H) Q252 methylated peptides. In both spectra, most fragment ions were matched including the ions representing the methylated glutamines.

Table A1.1 Data collection and refinement statistics for the 70S-RF complexes

	RF1	$RF1^{Me}$	RF2	$RF2^{Me}$
Data collection				
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Wavelength (Å)	0.97910	0.97910	0.97920	0.97890
Cell dimensions				
a, b, c (Å)	209.5, 450.9, 622.1	210.0, 449.7, 620.7	210.5, 450.4, 620.7	210.9, 450.1, 620.8
α, β, γ (°)	90, 90 ,90	90, 90 ,90	90, 90 ,90	90, 90 ,90
Resolution (Å)	49.9-3.04	49.8-3.20	49.9-3.10	49.9-3.30
$R_{p.i.m}$ (%)	9.7 (60.5)*	8.7 (87.8)*	7.3 (40.5)*	12.8 (94.4)*
I/σI	5.7 (1.3)	8.2 (1.1)	8.9 (1.8)	5.3 (0.84)
Completeness (%)	99.3 (98.5)	99.9 (99.9)	99.9 (99.9)	98.9 (99.2)
Redundancy	13 (12)	13 (10)	14 (6.6)	5.4 (5.2)
CC1/2 (%)	98.9 (20.4)	99.9 (28.9)	99.3 (57.7)	99.4 (22.3)
Refinement				
Reflections (#)	1109007	954884	1057109	866669
R_{work}/R_{free} (%)	24.7/27.9	21.4/24.4	18.7/22.4	19.8/23.6
No. atoms				
Macromolecule	292,465	292,280	294,133	293,949
Ligand/ion	2464	2459	2541	2548
Water	0	0	0	0
B-factors (Å^2)				
Overall	82.1	117	114	78.7
Macromolecule	83.0	118	114	78.8
Ligand/ion	60.5	93.0	94.6	69.3
R.m.s deviations				
Bond lengths (Å)	0.895	0.737	0.968	1.017
Bond angles (°)	0.005	0.003	0.006	0.008

*Highest resolution shell is shown in parenthesis

REFERENCES

1. D. S. Konecki, K. C. Aune, W. Tate, C. T. Caskey, Characterization of reticulocyte release factor. *J Biol Chem* **252**, 4514-4520 (1977).
2. E. Scolnick, R. Tompkins, T. Caskey, M. Nirenberg, Release factors differing in specificity for terminator codons. *Proc Natl Acad Sci U S A* **61**, 768-774 (1968).
3. L. Frolova *et al.*, A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. *Nature* **372**, 701-703 (1994).
4. L. Y. Frolova *et al.*, Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *Rna* **5**, 1014-1020 (1999).
5. S. Petry *et al.*, Crystal structures of the ribosome in complex with release factors RF1 and RF2 bound to a cognate stop codon. *Cell* **123**, 1255-1266 (2005).
6. U. B. Rawat *et al.*, A cryo-electron microscopic study of ribosome-bound termination factor RF2. *Nature* **421**, 87-90 (2003).
7. J. J. Shaw, R. Green, Two distinct components of release factor function uncovered by nucleophile partitioning analysis. *Mol Cell* **28**, 458-467 (2007).
8. Z. Vogel, A. Zamir, D. Elson, The possible involvement of peptidyl transferase in the termination step of protein biosynthesis. *Biochemistry* **8**, 5161-5168 (1969).
9. M. Y. Pavlov *et al.*, Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proc Natl Acad Sci U S A* **106**, 50-54 (2009).
10. N. Polacek *et al.*, The critical role of the universally conserved A2602 of 23S ribosomal RNA in the release of the nascent peptide during translation termination. *Mol Cell* **11**, 103-112 (2003).
11. E. M. Youngman, J. L. Brunelle, A. B. Kochaniak, R. Green, The active site of the ribosome is composed of two layers of conserved nucleotides with distinct roles in peptide bond formation and peptide release. *Cell* **117**, 589-599 (2004).
12. H. S. Zaher, J. J. Shaw, S. A. Strobel, R. Green, The 2'-OH group of the peptidyl-tRNA stabilizes an active conformation of the ribosomal PTC. *Embo J* **30**, 2445-2453 (2011).
13. H. Jin, A. C. Kelley, D. Loakes, V. Ramakrishnan, Structure of the 70S ribosome bound to release factor 2 and a substrate analog provides insights into catalysis of peptide release. *Proc Natl Acad Sci U S A* **107**, 8593-8598 (2010).
14. A. Korostelev *et al.*, Crystal structure of a translation termination complex formed with release factor RF2. *Proc Natl Acad Sci U S A* **105**, 19684-19689 (2008).
15. A. Korostelev, J. Zhu, H. Asahara, H. F. Noller, Recognition of the amber UAG stop codon by release factor RF1. *Embo j* **29**, 2577-2585 (2010).
16. M. Laurberg *et al.*, Structural basis for translation termination on the 70S ribosome. *Nature* **454**, 852-857 (2008).
17. A. Weixlbaumer *et al.*, Insights into translational termination from the structure of RF2 bound to the ribosome. *Science* **322**, 953-956 (2008).
18. J. Zhou, A. Korostelev, L. Lancaster, H. F. Noller, Crystal structures of 70S ribosomes bound to release factors RF1, RF2 and RF3. *Curr Opin Struct Biol* **22**, 733-742 (2012).
19. G. Indrisiunaite, M. Y. Pavlov, V. Heurgue-Hamard, M. Ehrenberg, On the pH dependence of class-1 RF-dependent termination of mRNA translation. *Journal of molecular biology* **427**, 1848-1860 (2015).
20. V. Heurgue-Hamard, S. Champ, A. Engstrom, M. Ehrenberg, R. H. Buckingham, The hemK gene in Escherichia coli encodes the N(5)-glutamine methyltransferase that modifies peptide release factors. *Embo j* **21**, 769-778 (2002).

21. L. Mora, V. Heurgue-Hamard, M. de Zamaroczy, S. Kervestin, R. H. Buckingham, Methylation of bacterial release factors RF1 and RF2 is required for normal translation termination in vivo. *J Biol Chem* **282**, 35638-35645 (2007).
22. V. Dincbas-Renqvist *et al.*, A post-translational modification in the GGQ motif of RF2 from Escherichia coli stimulates termination of translation. *EMBO J* **19**, 6900-6907 (2000).
23. S. Trobro, J. Aqvist, A model for how ribosomal release factors induce peptidyl-tRNA cleavage in termination of protein synthesis. *Mol Cell* **27**, 758-766 (2007).
24. G. R. Bjork *et al.*, Transfer RNA modification. *Annu Rev Biochem* **56**, 263-287 (1987).
25. A. Bjornsson, S. Mottagui-Tabar, L. A. Isaksson, Structure of the C-terminal end of the nascent peptide influences translation termination. *Embo j* **15**, 1696-1704 (1996).
26. D. V. Freistroffer, M. Kwiatkowski, R. H. Buckingham, M. Ehrenberg, The accuracy of codon recognition by polypeptide release factors. *Proc Natl Acad Sci U S A* **97**, 2046-2051 (2000).
27. L. K. Doerfel *et al.*, EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. *Science* **339**, 85-88 (2013).
28. S. Ude *et al.*, Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science* **339**, 82-85 (2013).
29. L. Peil *et al.*, Lys34 of translation elongation factor EF-P is hydroxylated by YfcM. *Nat Chem Biol* **8**, 695-697 (2012).
30. A. D. Petropoulos, M. E. McDonald, R. Green, H. S. Zaher, Distinct roles for release factor 1 and release factor 2 in translational quality control. *J Biol Chem* **289**, 17589-17596 (2014).
31. A. Batra *et al.*, Quantifying through-space charge transfer dynamics in pi-coupled molecular systems. *Nature communications* **3**, 1086 (2012).
32. N. Santos, J. Zhu, J. P. Donohue, A. A. Korostelev, H. F. Noller, Crystal structure of the 70S ribosome bound with the Q253P mutant form of release factor RF2. *Structure* **21**, 1258-1263 (2013).
33. A. Weixlbaumer *et al.*, Crystal structure of the ribosome recycling factor bound to the ribosome. *Nat Struct Mol Biol* **14**, 733-737 (2007).
34. C. S. Hayes, B. Bose, R. T. Sauer, Proline residues at the C terminus of nascent chains induce SsrA tagging during translation termination. *J Biol Chem* **277**, 33825-33832 (2002).
35. S. Mottagui-Tabar, A. Bjornsson, L. A. Isaksson, The second to last amino acid in the nascent peptide as a codon context determinant. *Embo j* **13**, 249-257 (1994).
36. D. Moazed, H. F. Noller, Transfer RNA shields specific nucleotides in 16S ribosomal RNA from attack by chemical probes. *Cell* **47**, 985-994 (1986).
37. C. L. Simms, B. H. Hudson, J. W. Mosior, A. S. Rangwala, H. S. Zaher, An active role for the ribosome in determining the fate of oxidized mRNA. *Cell reports* **9**, 1256-1264 (2014).
38. Y. Shimizu *et al.*, Cell-free translation reconstituted with purified components. *Nat Biotechnol* **19**, 751-755 (2001).
39. S. C. Blanchard, H. D. Kim, R. L. Gonzalez, Jr., J. D. Puglisi, S. Chu, tRNA dynamics on the ribosome during translation. *Proc Natl Acad Sci U S A* **101**, 12893-12898 (2004).
40. H. S. Zaher, P. J. Unrau, T7 RNA polymerase mediates fast promoter-independent extension of unstable nucleic acid complexes. *Biochemistry* **43**, 7873-7880 (2004).
41. S. E. Walker, K. Fredrick, Preparation and evaluation of acylated tRNAs. *Methods* **44**, 81-86 (2008).

42. A. Keller, A. I. Nesvizhskii, E. Kolker, R. Aebersold, Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74**, 5383-5392 (2002).
43. A. I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75**, 4646-4658 (2003).
44. M. Selmer *et al.*, Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935-1942 (2006).
45. W. Kabsch, Xds. *Acta crystallographica* **66**, 125-132 (2010).
46. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica* **66**, 213-221 (2010).
47. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta crystallographica* **66**, 486-501 (2010).
48. N. D. Heintzman *et al.*, Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-112 (2009).
49. Schrodinger, LLC. (2010).