

**FUNCTIONAL DISSECTION OF TRNA-CLEAVING TOXINS IN
MYCOBACTERIA**

by

Valdir Cristovao Barth Junior

A dissertation submitted to the
School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Graduate Program in Microbiology and Molecular Genetics

Written under the direction of

Nancy A. Woychik

And approved by

New Brunswick, New Jersey

October, 2020

ABSTRACT OF THE DISSERTATION

Functional dissection of tRNA-cleaving toxins in mycobacteria

By VALDIR CRISTOVAO BARTH JUNIOR

Dissertation Director:

Nancy A. Woychik

Tuberculosis (TB) is the leading cause of infectious disease-related deaths world-wide. The most common infectious agent of TB, *Mycobacterium tuberculosis*, is a highly resistant bacterium that evades the human immune response, presumably allowing it to persist in a dormant state in the lungs for decades. The factors that lead to its high persistence are not completely understood, but bacterial toxin-antitoxin (TA) systems have been implicated. Most TA systems do not have their mode of action elucidated, limiting our understanding of how they participate in bacterial persistence and latent TB. Therefore, the aim of this work was to identify the specific targets of VapC and MazF toxins and their physiological effects in *M. tuberculosis* and in the fast-growing model organism, *Mycobacterium smegmatis*. To do so, we applied 5' RNA-seq to accurately detect toxin-cleaved RNAs differentiating them from other cellular RNAs by their distinct 5' end left by the toxin. Knowing that VapCs and MazFs leave 5' monophosphate and 5' hydroxyl (OH) ends upon cleavage,

respectively, our method analyzes enrichment of those cleavage markers in mycobacterial cells expressing these TA toxins. We first found by 5' OH RNA-seq that MazF-mt9 is an isoacceptor-specific tRNAse which targets tRNA^{Lys-UUU} in *M. tuberculosis*. 5' OH RNA-seq also suggested that ribosomes were selectively stalling at lysine AAA codons due to the low levels of tRNA^{Lys-UUU}, which we confirmed using Ribo-seq. Expressing MazF-mt9 in the model organism *M. smegmatis* generates a shift in translational output favoring protein synthesis from transcripts with low levels of AAA codons and lowering the levels of AAA-rich proteins. Therefore, we documented a possible new mechanism of post-transcriptional regulation by tRNAs that dictates gene expression by codon usage. We predict this resulting cellular AAA-depleted proteome may induce persistence against antibiotics and protection against the host's immune system. In the second Chapter, we show that the only MazF toxin described in *M. smegmatis* genome (here named MazF-ms) is also a tRNA^{Lys}-specific tRNA-cleaving toxin. Expression of MazF-ms in *M. smegmatis* generates similar effects as observed for MazF-mt9 in *M. tuberculosis*. Newly synthesized protein production is heavily dictated by the transcript's Lys AAA codon content. The change in proteome favors genes involved in stress response and reduces expression of genes involved in cell division and DNA replication. In the third Chapter, we show that in vitro studies may be misleading. As we report for VapC-mt11, the toxin is highly specific when studied in vivo in its original host (*M. tuberculosis*) and loses specificity in in vitro assays. We propose the actual targets of VapC-mt11 are tRNA^{Gln} and tRNA^{Leu}, even though additional targets

can be observed in vitro or when expressing in *M. smegmatis*. In the fourth and last Chapter, we show that VapC-mt4 is another tRNA-cleaving toxin that behaves promiscuously in in vitro assays. We report that VapC-mt4 targets the only tRNA^{Cys} in vivo in *M. tuberculosis*. As observed with MazF-mt9, depletion of tRNA^{Cys} also triggered selective ribosome stalling in Cys codons and we were able to identify stalling sites by 5' OH RNA-seq. This selective ribosome stalling allowed the identification of dozens of putative small unannotated Cys-containing open reading frames (ORFs), some of which have been confirmed experimentally by quantitative mass spectrometry. Overall, our results indicate that tRNA cleavage is a common feature in mycobacterial TA toxins, which may be used to prompt responses to facilitate survival in stressful situations. Also, we demonstrate that our 5' OH RNA-seq method may be used to shed light on hidden ORFs in mycobacterial genomes.

Acknowledgements and Dedication

I dedicate this thesis first and foremost to my family whose physical company and support I painfully sacrificed to follow this dream.

Also, to my “family by choice”: “the originals” Karin, Luiz, Erik and Silvio, and the ones who arrived late to the Brazil/PhD-at-Rutgers party (yet somehow managed to finish earlier) Pedro, Lisia and Pupim. Your friendship made the hardest times bearable. To Vito and Theo, my sources of peace and unconditional love at the end of the day. To Julia and Eduardo for being the closest to a medicine I found against homesickness (if only all medicines were this much fun!). To friends who provided me with indispensable emotional support: Natasha, Bernardo, Dani Schabo, Bruna Hassan, Rosane, Mia, Laura, Maria Claudia, Anelise, Bruna Kern, Bruna “Bolota”, Sté, Belisa, Aline, Gior, Dani Cid, Amanda, Paula, .

To my current and former lab mates Unnati, Heather, Mike, Julia P., Julia G., Eduardo, Megan, Melvilí, John *et al.*, thank you for your help, the long discussions about science and other random subjects, the many fun times and for making the best environment for a productive work. A special thanks to Melvilí, “my favorite Puerto Rican”... Wait, is Ricky Martin Puerto Rican too? Oh, but she still my favorite. Mel was the one I looked up to the most and also who pushed me the most to achieve my goals during this rollercoaster of emotions we call graduate school.

I would also like to thank our collaborators Dr. Robert Husson and Dr. Ju-Mei Zeng without whom this work literally would not be the same. To the people at the Rutgers Quantitative Mass Spectrometry facility at the Proteomics building, especially Dr. Haiyan Zheng for guiding us through a steep learning curve during our proteomics experiments. To the members of the Patel lab and the Gartenberg lab who were always helpful and such good neighbors, especially in desperate times of not having specific reagents or in the need of sharing equipment.

To the most helpful staff I met at Rutgers: Diane Murano, Carolyn Ambrose, Caryn Carroll and Letice Smith.

To my committee members Dr. Robert Husson, Dr. Paul Copeland and Dr. Gary Brewer. Your invaluable input enriched my work and contributed tremendously to my formation as a scientist.

Last but certainly not least, to my “life and thesis” mentor Dr. Nancy Woychik. I have never met anyone who **genuinely cares** about their students’ personal and professional growth like you do. I consider myself very lucky to have had the opportunity to learn what a true mentor is, and I will take this lesson for life.

The work in this dissertation was supported in part by National Institutes of Health (NIH) grants R21 AI135461, RO1 154464 and R56 AI119055 to mentor Dr. Nancy A. Woychik. Also, I was fortunate enough to have received a full scholarship from the Brazilian government through the Science without Borders

program, which was the catalyzer for all the amazing opportunities I had to grow as a Scientist. The program was linked to the Coordination for the Improvement of Higher Education Personnel (CAPES), a part of the Ministry of Education, with the process number 99999.013433/2013-00. I have also received the Bevier Thesis Completion Fellowship and the Aaron Shatkin award from Rutgers University, both of which I am very grateful for.

Chapter 1 was fully published in Nature Communications (Barth et al., 2019), Chapter 2 was fully published in Frontiers in Genetics (Barth et al., 2020), Chapter 3 was fully published in Scientific Reports (Cintrón et al., 2019) in close collaboration with the first author Dr. Melvilí Cintrón.

Table of Contents

ABSTRACT OF THE DISSERTATION.....	ii
Acknowledgements and Dedication.....	v
Table of Contents	viii
List of illustrations	xiii
Introduction.....	1
Pathology of TB	5
Toxin-antitoxin (TA) systems and bacterial persistence	7
VapBC and MazEF TA systems in Mycobacteria.....	12
VapBC in <i>M. tuberculosis</i>	14
MazEF in <i>M. tuberculosis</i>	15
Rationale	16
Chapter 1: Toxin-mediated ribosome stalling reprograms the <i>Mycobacterium tuberculosis</i> proteome	19
Introduction.....	19
Methods.....	21
Strains, plasmids and reagents	21
RNA isolation.....	23
Northern analysis of tRNA levels.....	24
Labeling of newly synthesized <i>M. tuberculosis</i> proteins	25
5' RNA-seq	25
RNA-seq	28

Ribo-seq	28
Construction of RNase J deletion strain	29
Proteomics.....	30
Data availability	32
Results	32
5' RNA-seq suggests ribosome stalling at AAA Lys codons	36
Ribo-seq confirms ribosome stalling	38
Ribosome stalling at AAAs upon RNase J expression in <i>E. coli</i>	41
Cleavage at stalled ribosomes favors AAA-deficient mRNAs	45
MazF-mt9 translation inhibition is codon specific.....	49
Discussion	50
Chapter 2: The sole <i>Mycobacterium smegmatis</i> MazF toxin targets tRNA ^{Lys} to impart highly selective, codon-dependent proteome reprogramming	54
Introduction.....	54
Materials and Methods	55
Strains, plasmids and reagents	55
RNA isolation.....	56
5' RNA-seq	57
Labeling of newly synthesized proteins and proteomics analysis.....	59
Results	60
MazF-ms expression arrests growth in <i>M. smegmatis</i>	60
MazF-ms exclusively targets both tRNA ^{Lys} isoacceptors for cleavage at a single site within their anticodons	62
5' RNA-seq of MazF-ms-expressing cells reveals ribosome stalling	67
MazF-ms promotes codon-specific translation.....	71

MazF-ms reduces translation of critical components of the DNA replication machinery while concomitantly supporting synthesis of stress response proteins	74
Discussion	78
 Chapter 3 - Accurate target identification for <i>Mycobacterium tuberculosis</i>	
endoribonuclease toxins requires expression in their native host.....	83
Introduction.....	83
Methods.....	85
Strains, Plasmids and Reagents	85
Growth Assays in Mycobacteria	86
Purification of Recombinant VapC-mt11	87
In vitro Transcription of <i>M. tuberculosis</i> tRNAs	87
In vitro tRNA Cleavage Assay	88
In vitro <i>M. tuberculosis</i> tRNA Primer Extension	88
tRNA Northern Analysis	90
In vitro Translation Inhibition	91
Metabolic Labeling in <i>M. smegmatis</i>	91
5' RNA-seq	92
Results	94
VapC-mt11 expression causes growth arrest in both <i>M. smegmatis</i> and <i>M. tuberculosis</i>	94
VapC-mt11 inhibits translation	96
VapC-mt11 is a tRNase that targets multiple synthetic <i>M. tuberculosis</i> tRNAs.....	97
VapC-mt11 tRNA recognition and cleavage is not modification dependent	100
VapC-mt11 generates tRNA halves upon cleavage after a GG sequence in the anticodon loop.....	103
VapC-mt11 requires a GG sequence for tRNA recognition and accurate cleavage	106
VapC-mt11 cleavage is both sequence- and structure-dependent.....	109

VapC-mt11 cleaves only two GG consensus-containing tRNA targets when expressed in its natural <i>M. tuberculosis</i> host	112
Discussion	114
Chapter 4: Toxin-mediated, cysteine-specific ribosome stalling reveals hidden sORFs in the <i>Mycobacterium tuberculosis</i> genome and triggers targeted stress survival pathways	
Abstract	120
Introduction.....	121
Methods.....	123
Strains, plasmids and reagents	123
Construction of vapBC4 deletion strain	123
RNA isolation.....	124
In vitro cleavage assays with purified VapC4.....	125
5' RNA-seq	125
Total RNA-seq.....	126
Labeling of newly synthesized <i>M. tuberculosis</i> proteins	127
Proteomics.....	127
Small protein identification	128
Results	130
VapC4 cleaves many <i>M. tuberculosis</i> tRNAs within their anticodon stem loop in vitro.	130
VapC4 does not inhibit protein synthesis.	131
VapC4 only targets tRNA ^{Cys} in <i>M. tuberculosis</i> cells.....	133
tRNA ^{CysGCA} cleavage leads to ribosome stalling at Cys codons.....	136
Ribosome stalling at Cys codons uncovered an abundance of unannotated small proteins.	140
One sORF is the ortholog of Cys-responsive attenuator in <i>M. smegmatis</i>	142

VapC4 causes ribosome stalling at more leaderless sORFs within other Cys attenuation-responsive regulons.	145
Other unannotated ORFs encode stable Cys-containing proteins that can also map immediately before, within or opposite annotated genes.	146
Four unannotated sORFs encode novel EsxB-like proteins.	148
VapC4 selectively upregulates sulfur metabolism and assimilation, oxidative stress, copper detoxification and antiviral defense pathways.	151
Discussion	157
Conclusions	160
References	162

List of illustrations

Figure 1. Top 10 global causes of deaths worldwide in 2017.	1
Figure 2. Estimated TB incidence rates.	2
Figure 3. Effect of COVID-19 pandemics in the notification of new cases in India in 2020.....	3
Figure 4. Latent (a) and active (b) stages of <i>Mycobacterium tuberculosis</i> infection.	5
Figure 5. Distinguishing antimicrobial resistance, tolerance and persistence.	11
Figure 6. MazF-mt9 expression leads to slow growth in <i>M. tuberculosis</i>	23
Figure 7. Only tRNA ^{Lys43-UUU} is cleaved by MazF-mt9 in <i>M. tuberculosis</i>	34
Figure 8. MazF-mt9 targets only tRNA ^{Lys43-UUU} in vivo.	35
Figure 9. 5' RNA-seq captures ribosome stalling events at AAA Lys codons.....	37
Figure 10. Ribo-seq confirms transcriptome-wide AAA Lys-specific ribosome stalling.	40
Figure 11. Ribo-seq shows ribosome stalling preference for early AAA codons.	41
Figure 12. <i>M. tuberculosis</i> RNase J cleaves 5' of stalled ribosomes favoring expression of AAA-deficient transcripts.	43
Figure 13. Ribosome stalling occurs upon expression of MazF-mt9 in a RNase J-deletion background.....	44
Figure 14. Induction of MazF-mt9 leads to global proteomic shifts based on the AAA codon content of transcripts.	47
Figure 15. mRNA levels correlate with protein expression.....	48
Figure 16. MazF-mt9 expression does not inhibit translation in <i>M. tuberculosis</i>	49
Figure 17. MazF-mt9 translation inhibition is codon specific.....	50
Figure 18. MazF-mt9 mechanism of action.....	51

Figure 19. Expression levels of pMC1s vector in <i>M. smegmatis</i>	61
Figure 20. MazF-ms expression in <i>M. smegmatis</i> leads to growth inhibition.	62
Figure 21. MazF-ms selectively targets both tRNA ^{Lys} isoacceptors.....	64
Figure 22. MazF-ms targets show high secondary structure and sequence similarity.....	66
Figure 23. 5' RNA-seq serendipitously reveals ribosome stalling at lysine codons.	69
Figure 24. Translation of proteins rich in lysine AAA codons is reduced upon MazF-ms expression.	73
Figure 25. Expression of MazF-ms leads to global proteomic shifts based on the Lys AAA, but not Lys AAG, codon content.	74
Figure 26. Functional classification of differentially translated proteins during MazF-ms expression.	76
Figure 27. Summary of the proposed MazF-ms mechanism of action.	80
Figure 28. Growth inhibition by VapC-mt11.	95
Figure 29. VapC-mt11 inhibits translation in vitro and in <i>M. smegmatis</i>	97
Figure 30. VapC-mt11 cleaves one third of <i>M. tuberculosis</i> tRNAs in vitro.	99
Figure 31. tRNA modifications are not required for VapC-mt11 target recognition in vitro.	101
Figure 32. Weak or uncleaved in vivo synthesized tRNAs mirror the subset of weakly cleaved in vitro synthesized tRNAs.....	102
Figure 33. VapC-mt11 cleaves tRNAs within the ASL.....	104
Figure 34. VapC-mt11 requires a GG sequence within the proper context.....	108
Figure 35. Sequence and structure influence the VapC-mt11 cleavage site.....	110

Figure 36. VapC-mt11 cleaves only two tRNAs containing the GG cleavage consensus sequence in vivo.....	113
Figure 37. VapC4 promiscuously cleave several tRNAs in vitro.	131
Figure 38. Cells expressing VapC4 sustained substantial translation despite growth inhibition.	132
Figure 39. tRNA ^{CysGCA} is the only VapC4 target in vivo.	135
Figure 40. VapC4 cleavage generates 5' monophosphate ends and lacks specificity in in vitro assays.....	136
Figure 41. 5' OH RNA-seq reveals ribosome stalling at Cys codons with a single nucleotide resolution.	138
Figure 42. <i>M. tuberculosis</i> expressing VapC4 preferentially downregulates translation of Cys-rich proteins.	140
Figure 43. Downstream genes of identified Cys-containing sORFs are generally upregulated.....	143
Figure 44. Unannotated sORFs can occur in a diverse gene organization context.....	147
Figure 45. Putative sORFs identified overlapping transposases are homologues of EsxB/CFP-10.	149
Figure 46. Genes involved in sulfate assimilation, namely part of the cysteine biosynthesis pathway, are upregulated in <i>M. tuberculosis</i> cells expressing VapC4.	152
Figure 47. A putative regulon formed by Rv0846c, Rv0847 (lpqS), Rv0848 (cysK2), Rv0849, Rv0850 is active upon expression of VapC4.	156

Introduction

According to the World Health Organization (WHO), in 2016, microorganisms were involved in at least three of the top ten causes of deaths worldwide, having an especially pronounced impact in lower income countries (Figure 1). Tuberculosis (TB) is among them, being the fifth most common cause of death in low income countries (World Health Organization, 2018b). TB's etiological agent is *Mycobacterium tuberculosis*, and it ranks as the deadliest pathogen in the world since 2007. Even though the most common form of the disease occurs when *M. tuberculosis* infects exclusively the lungs (pulmonary TB), it can also propagate to other organs causing extrapulmonary TB.

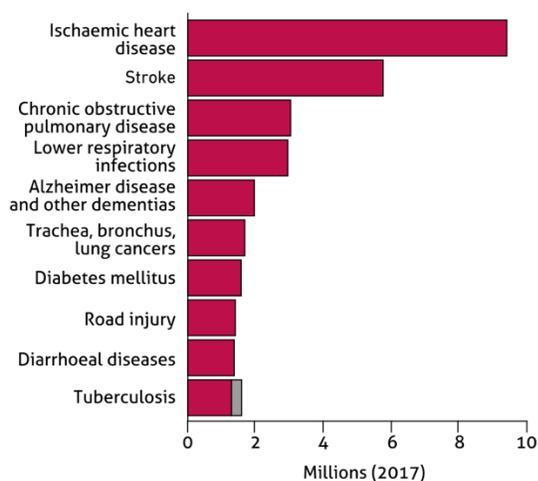


Figure 1. Top 10 global causes of deaths worldwide in 2017. Deaths from TB among HIV-positive people are shown in grey (as they are officially classified as death by HIV/AIDS). Source: World Health Organization, 2019.

Ten million new cases were reported in 2018 and about 1.45 million people died due to the infection (World Health Organization, 2019). However, the

disease incidence is not homogeneously distributed worldwide (Figure 2). India, China, Indonesia, the Philippines, Pakistan, Nigeria, Bangladesh and South Africa account for two thirds of the world's new cases in 2018 (World Health Organization, 2019). It is predicted that the COVID-19 pandemic scenario of 2020 will aggravate even more the reporting and spread of TB in these countries. Only in India, there was a 75% drop in new cases reported in the weeks following the mandated lockdown to avoid SARS-CoV-2 spread (Figure 3). Similar trends were observed in China (-20% reported cases) and Indonesia (-68% reported cases) (Glaziou, 2020). These numbers indicate a drastic deficiency in diagnosing infected patients, signaling an almost inevitable rise in active cases that will remain untreated until the health system in those countries returns to normality. In fact, it is projected that if global TB cases detection decreases by 25% in 3 months, there will be a 13% increase in the number of deaths, bringing the total death toll to 1.66 million in 2020 (Glaziou, 2020).

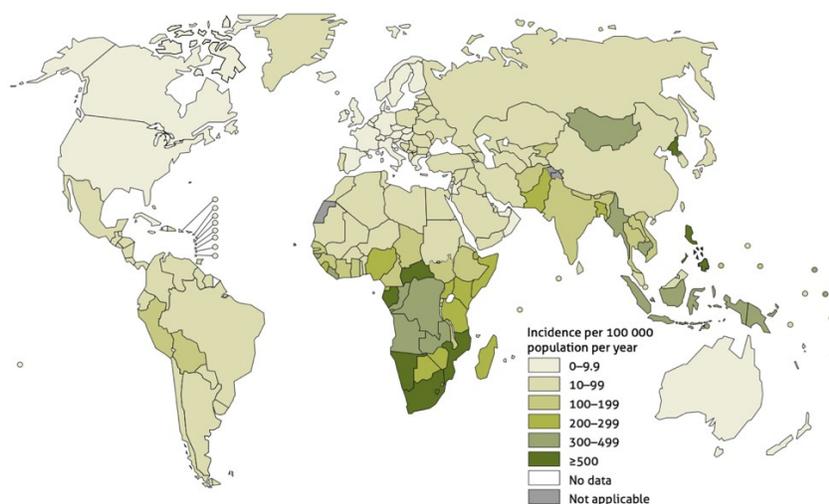


Figure 2. Estimated TB incidence rates. Source: World Health Organization, 2019.

The fact that the majority of infections occur in lower-middle-income countries is not coincidental. Many factors that contribute to TB infection and spread are tightly tied to socio-economical aspects of the population, such as: undernutrition, access to health coverage, treatment availability, housing density, and poverty.

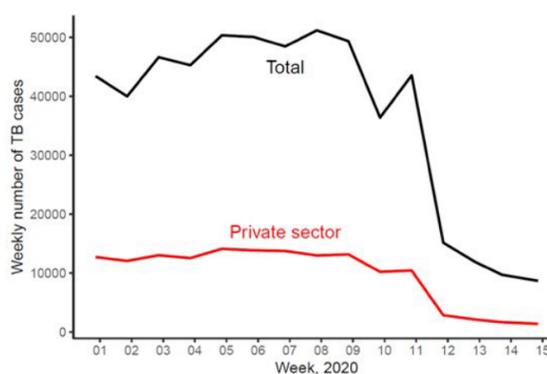


Figure 3. Effect of COVID-19 pandemics in the notification of new cases in India in 2020. Source: Glaziou, 2020.

Other risk factors include diabetes, smoking and alcohol use, but the strongest risk factor associated to TB is HIV infection. Approximately 21% of deaths by TB occur in people co-infected with HIV, most of them in South Africa. Moreover, TB is the leading cause of hospitalizations (18% of HIV-related illnesses) among HIV-positive people (Ford et al., 2015). These hospitalizations were associated with low CD4⁺ lymphocyte counts, a consequence of viral replication (Ford et al., 2015).

In high burden countries, the WHO recommends immunization with the BCG vaccine at or close to birth. This is the only licensed vaccine to prevent TB cases and contains an attenuated version of *M. tuberculosis*'s sister species,

Mycobacterium bovis (strain bacillus Calmette-Guérin, BCG). It is recommended especially to prevent leprosy, tuberculous meningitis and some extrapulmonary TB cases during childhood but provide limited protection later in life (World Health Organization, 2018a). Conversely, some recent studies have suggested extended protection spanning into adulthood, with lower rates of infection associated with early vaccination (Katelaris et al., 2020; Nguipdop-Djomo, Heldal, Rodrigues, Abubakar, & Mangtani, 2016), contrasting previous reports that protection is mostly lost in adulthood (Herzmann et al., 2014). Vaccination with BCG however may create false positive results in the tuberculin skin test (TST), the most widely used method for TB screening. This one-hundred year old test evaluates an immune response generated against a complex mix of *M. tuberculosis* antigens (Old Tuberculin or purified protein derivative, PPD) applied intradermally (Huebner, Schein, & Bass, 1993). TST is inexpensive and easy to perform, hence its widespread distribution even in low income countries. However, its interpretation can be dubious and possible false positive results may arise not only from previous BCG immunization but also from exposure to other mycobacteria (Diel, 2017). Other diagnostic tests are available, and as they can detect the disease at different stages and/or give contrasting results, multiple confirmatory tests are often necessary. Interferon Gamma Release Assays (IGRAs), such as QuantiFERON-TB® and T-SPOT TB®, are diagnostic alternatives that can detect latent TB without being affected by previous BCG vaccination, since they use antigens that are not present in *M. bovis* BCG genome. Nonetheless, due to its high cost, these tests are not popular in low-

income countries. Direct detection of *M. tuberculosis* bacteria by microbiology techniques (sputum smears or microbiological cultures) are also inexpensive but only detect the active state of the disease (Pai et al., 2016). Still, false negatives results may occur even in active disease cases, due to low bacterial load expelled in the sputum (Pai et al., 2016).

Pathology of TB

TB is classically classified into latent TB infection (LTBI) and active TB (Figure 4), although recently these terminologies have been argued against (Behr, Edelstein, & Ramakrishnan, 2018).

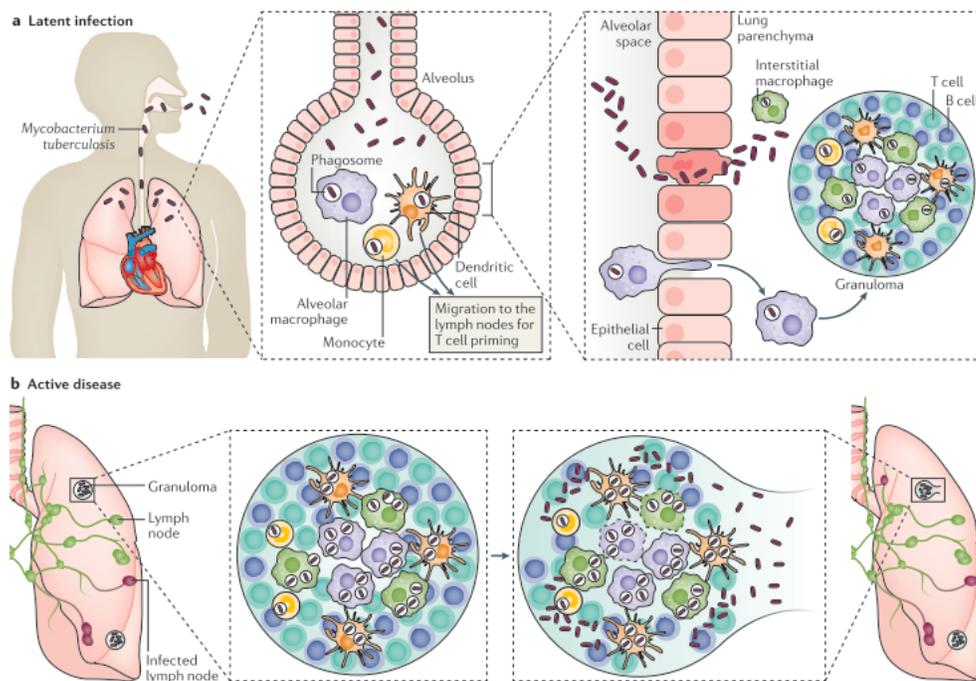


Figure 4. Latent (a) and active (b) stages of *Mycobacterium tuberculosis* infection. Source: Pai et al., 2016.

Dissemination of *M. tuberculosis* occurs during the active form of the disease through the inhalation of infected droplets expelled by a patient with active TB. The bacilli travel to lower respiratory tract and are detected by immune cells, particularly alveolar macrophages (Pai et al., 2016). If *M. tuberculosis* survives this first line of attack by the innate immune response, it migrates to the parenchymal space of the lungs. The exact mechanism by which *M. tuberculosis* is transported from the alveolus lumen to the interstitium is unknown, but it is thought to occur by either one of four mechanisms: traveling inside infected macrophages, by directly infecting epithelial cells, by infecting the antigen-presenting M cells present in the epithelium or by dendritic cell trafficking (Moule & Cirillo, 2020). Once in the interstitium, it can disseminate to other organs and cause extrapulmonary TB or it can be contained in the lungs by a barrier-like structure formed by immune cells, called “granuloma” (Figure 4). The granuloma morphology in humans is immensely understudied because it is in an inaccessible location in patients, therefore models for granuloma formation are generally built from data obtained from experimental animals (such as guinea pigs, mice, zebra fish and non-human primates) (Orme & Basaraba, 2014). In terms of structure, granulomas consist of several layers of immune cells, including macrophages, neutrophils, monocytes, TH1 T lymphocytes and B lymphocytes, with a necrotic environment in its core, where *M. tuberculosis* is entrapped (Orme & Basaraba, 2014). This structure may control the infection for decades in the form of a LTBI. However, when these structures can no longer contain the bacteria (thus initiating the active TB), *M. tuberculosis* can re-enter

the airways to be expelled via aerosols to start a new cycle of dissemination (Figure 4). Recently, it has been debated which of these two forms (LTBI or active TB infection) is the main factor responsible for the ~10 million new annual cases. Three independent studies have shown that in most cases of *M. tuberculosis* infection – verified by TST conversion – patients developed the active form of TB within 2 years rather than the latent form (as reviewed in Behr et al., 2018), indicating that the active disease may be the main source of new infections in countries with high number of cases (and not late reactivation of LBTI). Also, contrary to the widely accepted model of latent TB reactivation in the field, the authors argue that the reactivation of latent TB is less common than thought and only has a heavier weight in countries with low TB burden (Behr et al., 2018).

Still, the mechanisms by which *M. tuberculosis* may cause reactivation of the disease or survive the first attacks by the innate immune system are generally unknown. Some genes have been proposed to play an important role in these processes, including toxin-antitoxin systems.

Toxin-antitoxin (TA) systems and bacterial persistence

Toxin-antitoxin (TA) systems have been studied for decades, and it is still a rapidly developing field. They can be found in virtually all main phylogenetic branches of the Bacteria and Archaea domains (Xie et al., 2018) and are known to be involved in regulating cell growth. TA toxins have been implicated in important bacterial physiological processes, including plasmid maintenance,

antiviral defense and bacterial persistence (Harms, Brodersen, Mitarai, & Gerdes, 2018). There are at least four well-described types (I-IV) of TA systems (Harms et al., 2018), however, type II (in which both toxin and antitoxin are proteins) are the most studied. Type II TA systems typically comprise adjacent genes encoding two small (~10 kDa) proteins, a toxin and its cognate antitoxin that inhibits toxin activity through formation of a stable TA protein-protein complex (Fraikin, Goormaghtigh, & Van Melderen, 2020). The toxin has inhibitory activity on essential physiological processes, such as translation and DNA replication. Therefore, when the toxin is in its free, active form (that is, in the absence of antitoxin), the cells present a significant growth defect, the hallmark of TA toxins (Harms et al., 2018). In fact, this growth regulation prompted by TA toxins and other factors has been proposed as the basis of bacterial persistence (an antibiotic tolerance phenotype) (Pontes & Groisman, 2020).

Although several species encode TA system homologs in their genome, our knowledge on the subject is fundamentally obtained from experiments with the model microorganism *Escherichia coli*. For instance, according to the Toxin-antitoxin Database (Xie et al., 2018), *E. coli* K-12 strain harbors 31 predicted TA toxins in its chromosome, 26 of which were empirically tested. In contrast, *Salmonella enterica* harbors 18 predicted TA system, only 4 of which have been experimentally studied (Xie et al., 2018). A common theme among many TA toxins is their ribonuclease (RNase) activity, particularly in the MazEF and VapBC families of TA systems. Both MazF toxins encoded by *Bacillus subtilis* and *Staphylococcus aureus*, for instance, cleave RNAs at 5'-UACAU-3' (Park,

Yamaguchi, & Inouye, 2011; Zhu et al., 2009). Accordingly, several MazFs described so far cut their target at single stranded regions after recognizing a specific sequence, normally 3-, 5- or 7-nucleotide long (Masuda & Inouye, 2017). The *E. coli* MazF toxin (also referred to as MazF-ec) recognizes the 5'-ACA-3' triplet in mRNAs and rRNAs, although flanking bases may play a secondary role in recognition (Culviner & Laub, 2018). A complex mechanism of action has been proposed in which MazF-ec preferentially cleaves ACA-containing mRNAs upstream of start codons generating leaderless transcripts (Vesper et al., 2011). Concomitantly, it generates specialized ribosomes by removing their anti-Shine Dalgarno (SD) sequence (Vesper et al., 2011). This double cleavage would generate compatibility between cleaved transcripts and specialized ribosomes, stimulating the translation of proteins involved in stress response (Nigam, Ziv, Oron-Gottesman, & Engelberg-Kulka, 2019; Vesper et al., 2011). However, this idea has been criticized due to flaws in the methodology (Wade & Laub, 2019). Moreover, recent studies failed to reproduce their main findings (Culviner & Laub, 2018; Mets et al., 2017). Transcriptomic analysis of cells expressing MazF-ec did not show primary cleavage of mature rRNA nor preferential ACA cleavage in leader regions of mRNAs (Culviner & Laub, 2018). Nonetheless, it is widely accepted that MazFs cleave mRNAs and/or rRNA (Schifano & Woychik, 2017). The only unusual case is the tRNA-cleaving toxin MazF-mt9 present in *M. tuberculosis* genome (Schifano et al., 2016; Schifano & Woychik, 2017).

tRNA cleavage is, in fact, a more common trait among the VapBC family (named after virulence associated protein). Toxins from the VapBC family are

characterized by the RNase-associated PiIT N-terminal (PIN) domain, which is composed of conserved aspartic acid and glutamic acid residues that form the enzymatic active site (Matelska, Steczkiewicz, & Ginalski, 2017). One of the earliest biochemical characterizations of VapBC family members was both VapCs present in a plasmid from *Shigella flexnerii* and in the chromosome of *Salmonella enterica* serovar Typhimurium LT2 (K. S. Winther & Gerdes, 2011). These toxins were site-specific endoribonucleases that targeted the anticodon stem loop (ASL) of the initiator tRNA, tRNA^{fMet}, shutting down translation and transiently arresting growth (K. S. Winther & Gerdes, 2011). Yet, the ability to digest tRNA^{fMet} is not exclusive of enterobacterial VapCs. Other examples can be found in *Haemophilus influenzae* (VapC1^{NTHi} and VapC2^{NTHi}, which similarly target tRNA^{fMet} at the ASL), and *Leptospira interrogans* (VapC_{Lin}) (Lopes et al., 2014; Walling & Butler, 2017). VapC toxins from *M. tuberculosis*, on the other hand, present unusually high diversity in targets, including other elongating tRNAs and 16S/23S rRNA (Schifano et al., 2014; K. Winther, Tree, Tollervey, & Gerdes, 2016). The preference for tRNAs and rRNAs suggest an additional layer of precision in target recognition that extends beyond only recognizing sequence-specific regions like most MazFs. VapCs also require a secondary structure (eg. stem loop) for cleavage, as will be discussed further.

In a few occasions, TA systems were directly proposed to be underlying antimicrobial persistence, including MazF toxins (Cheverton et al., 2016; Harrison et al., 2009; Kaspary et al., 2013; Tripathi, Dewan, Siddique, & Varadarajan, 2014). Antimicrobial persistence is multi-drug unresponsiveness

phenotype with obscure molecular mechanisms. It is characterized by a biphasic killing curve presented by a bacterial population challenged with high concentrations of bactericidal antibiotics (Balaban et al., 2019). In other words, it is the ability of a small fraction of cells in the population (called persisters) to survive a bactericidal dose of antimicrobial drugs despite presenting similar Minimum Inhibitory Concentrations (MIC) to susceptible cells (Figure 5 a-b). It is distinguished from drug resistance and tolerance due to the higher MIC values and kinetics of antimicrobial killing, respectively (Figure 5).

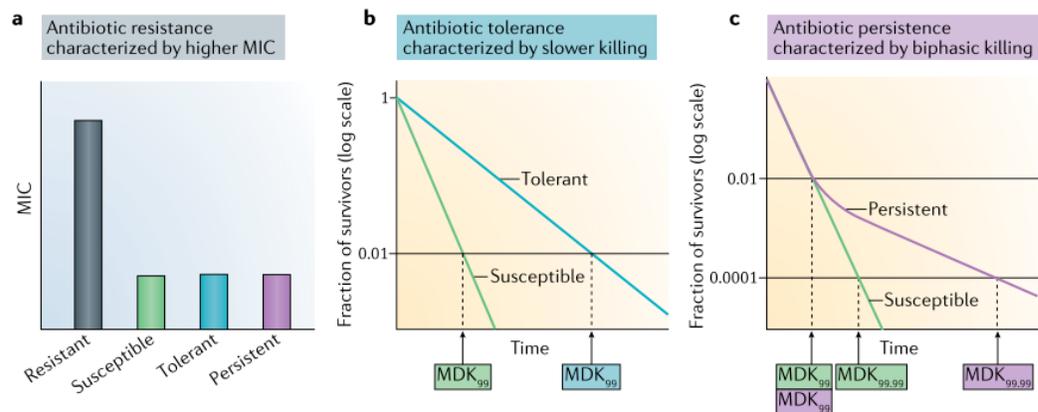


Figure 5. Distinguishing antimicrobial resistance, tolerance and persistence. Drug resistant cells present a high Minimum Inhibitory Concentration (MIC), while tolerant and persisters cells present similar MIC values to susceptible cells (a). However, tolerant cells are eliminated homogeneously at population level upon drug exposure (b). Instead, persister cell subpopulation behave differently from most of the population by remaining unaffected by the antimicrobial drug. Source: Balaban et al., 2019.

The observations that not all cells are equally affected by antimicrobials suggest a phenotypic heterogeneity in the isogenic bacterial population. It is thought that persister cells can arise spontaneously (attributed to the stochastic nature of transcription initiation) or be triggered by certain environmental causes (Balaban et al., 2019). Nevertheless, persisters have been studied in laboratory settings, therefore their involvement in disease is, although expected, underdocumented. Still, the phenotypes classically associated with TA systems and persister cells (activation of stress response, growth arrest, unresponsiveness to antimicrobial drug treatment) strikingly resemble the non-replicating, stressful state *M. tuberculosis* cells withstand during latent infection (Wayne & Sohaskey, 2001). Moreover, the failure of sterilizing the site of infection is witnessed in the traditional Cornell murine model and its variants (Scanga et al., 1999). This model consists of treatment with two antimicrobial drugs (isoniazid and pyrazinamide) resulting in no detectable bacilli at the infection site. However, the disease still reactivates weeks to months after the treatment is ceased (Scanga et al., 1999), indicating the existence of a subpopulation of cells that did not respond to treatment, similar to the antimicrobial persistence observed in liquid cultures. Also, in line with the idea that TA system may play a role in this persistence-like condition, *M. tuberculosis* ranks number one among pathogens with the highest number of TA systems in their genome (Xie et al., 2018).

VapBC and MazEF TA systems in Mycobacteria

Among the species belonging to the *Mycobacterium tuberculosis* complex (MTBC), a phylogenetic group of ~10 mycobacterial species that can cause tuberculosis in humans or animals, TA systems are extremely abundant. There are at least >80 type II TA systems in *M. tuberculosis* H37Rv (Sala, Bordes, & Genevaux, 2014) in remarkable contrast to only ~3 in the non-MTBC model species *M. smegmatis* (Frampton, Aggio, Villas-Bôas, Arcus, & Cook, 2012). *M. smegmatis* possesses only a single member of each of the three families: MazEF, VapBC and PhD/doc. All their targets are unknown. On the other hand, *M. tuberculosis* has members of the families: VapBC, MazEF, RelBE and YefM/YoeB, HigBA, ParDE, and others in non-canonical and/or uncharacterized families (Sala et al., 2014). Despite a greater diversity in families, most of TA systems belong to either the VapBC (comprising ~50 modules) or MazEF (comprising ~11 modules) families (Sala et al., 2014). In addition to being abundant in mycobacterial species that can cause tuberculosis (MTBC), the deletion of members from both of these two families greatly decreases the bacillary loads in the guinea pig infection model, providing further evidence for their involvement in pathogenesis (Agarwal et al., 2018; Tiwari et al., 2015). Accordingly, quantitative mass spectrometry showed that VapC toxins and MazF-*mt3* are upregulated in a nutrient starvation model commonly used to mimic nutritional condition of phagosomes (Albrethsen et al., 2013). These findings highlight a probable role in establishing tuberculosis and emphasize the importance of elucidating their molecular mechanisms and targets.

VapBC in *M. tuberculosis*

Despite early reports mischaracterizing them as mRNAses (Ramage, Connolly, & Cox, 2009), *M. tuberculosis* VapBC modules, just like their homologues in enterobacteria, seem to primarily target tRNAs, with a few exceptions that are predicted to cleave rRNA (K. Winther et al., 2016). In vitro characterizations were customarily performed using MS2 phage RNA or common mRNAs from model microorganism as substrates to detect ribonucleolytic activity (Ramage et al., 2009; Sharp et al., 2012). By using this approach, it was possible to determine some sequence specificity and RNase activity, but this method does not consider secondary structures that were later found to be essential for VapC proper recognition of its actual target (Cintrón et al., 2019; Jonathan W. Cruz et al., 2015). Cruz and collaborators extensively explored the determinants necessary for in vitro cleavage of tRNAs by VapC-mt4, showing that both secondary structure and a single-stranded region at the cleavage site are required (Jonathan W. Cruz et al., 2015). More recently, one study attempted to investigate the targets of numerous VapC toxins from *M. tuberculosis* (K. Winther et al., 2016). Surprisingly, the initiator tRNA^{fMet} was not among them. Instead, elongating tRNAs (such as tRNA^{Cys}, tRNA^{Leu}, tRNA^{Gln}, tRNA^{Ser} and tRNA^{Trp}) were redundantly cleaved by distinct VapCs (K. Winther et al., 2016). For instance, specific cleavage products of tRNA^{Trp} were generated when total *M. tuberculosis* RNA was incubated with VapC-mt25, VapC-mt29, VapC-mt33, VapC-mt37 and VapC-mt39. Moreover, the 23S rRNA cleavage at the sarcin-ricin loop reported to be generated by VapC-mt26 had already been observed in

a study from the same group when VapC-mt20 was ectopically expressed in *E. coli* (K. S. Winther, Brodersen, Brown, & Gerdes, 2013). These examples strongly suggest that, for reasons yet to be determined, *M. tuberculosis* suffered a selective pressure along its co-evolution with humans to maintain multiple functionally redundant TA genes. It is intriguing, however, that not all these toxins produce a growth arrest phenotype when expressed, despite targeting the same tRNA isoacceptor (K. Winther et al., 2016). Nevertheless, all studies that aimed to identify *M. tuberculosis* VapC targets used one or a combination of the following approaches: 1) in vitro cleavage assays with purified toxins, 2) expression of the toxin in *E. coli* or 3) expression of the toxin in *M. smegmatis* (Jonathan W. Cruz et al., 2015; K. S. Winther et al., 2013; K. Winther et al., 2016). The results presented in the upcoming Chapters are the first to analyze VapC toxin activity by expressing them in their original host cells, therefore being closer to physiological conditions and accounting for variables that may prevent the correct target recognition (eg. unexpected toxin interactions with cellular components or tRNA modifications).

MazEF in *M. tuberculosis*

Analogous to VapCs, the little we know about MazFs in *M. tuberculosis* suggest they might have convergently evolved to target secondary structure-rich transcripts. The most well characterized toxin in this group is MazF-mt9 (aka MaF7). In contrast to the well-studied Maz-ec, which is primarily a mRNA-

cleaving enzyme, MazF-mt9 target recognition depends on two features: a specific sequence (5' UUU 3') in a single-stranded region and a stem loop structure (Schifano et al., 2016). tRNA^{Pro-GGG} and tRNA^{Lys-UUU} both match these criteria and were shown to be cleaved by MazF-mt9 in vitro (Schifano et al., 2016). The cleavage sites were in the anticodon stem loop of tRNA^{Lys-UUU} and in the D-loop of tRNA^{Pro-GGG} (Schifano et al., 2016). MazF-mt3 (aka MazF6) and MazF-mt6 (aka MazF3), on the other hand, are specialized in endoribonucleolytic cleavage of rRNA, despite having a slightly distinct sequence recognition (Schifano et al., 2013, 2014). MazF-mt3 mutually cleaves 16S and 23S rRNA at conserved regions (5' UCCUU 3' as recognition sequence) where the anti-SD and the helix/loop 70 are located, respectively (Schifano et al., 2014). MazF-mt6 only cleaves 23S at 5' UUCCU 3' occurring at the helix/loop 70 of dissociated 50S subunits, impeding the interaction between the small and large subunits of the ribosome (Schifano et al., 2013). Cleavage at helix/loop 70 is predicted to harshly inhibit translation in *E. coli* due to its structural participation an essential region of the ribosome (A site). An atypical molecular mechanism was proposed for MazF-mt7 (aka MazF4). MazF-mt7 was shown to interact with topoisomerase I, inhibiting its activity and promoting growth arrest (F. Huang & He, 2010). This curious mode of action contradicts all other examples of MazFs known to date but it could suggest additional roles for MazFs in addition to the canonical RNA cleavage.

Rationale

With all the data linking TA systems to antimicrobial persistence, nutrient starvation, hypoxia and dormancy, it becomes tempting to propose their role in TB. During active and latent TB, *M. tuberculosis* must cope with several stresses imposed by the host's immune system. *M. tuberculosis* does not have any intermediate hosts, meaning its whole life cycle revolves around humans. Therefore, it is likely that this pathogen genome had to be evolutionarily shaped to thrive in the host-pathogen battle. That might include duplication of genes that are critical for successfully maintain its life cycle (ie. infection). This hypothesis is reinforced by the essentiality of TA systems in establishing infections in the guinea pig animal model (Agarwal et al., 2018; Tiwari et al., 2015) and by the disproportionate number of duplicated TA genes exclusively seen in mycobacterial species that cause TB (Ramage et al., 2009).

Also, the trend observed in the in vitro assays using *M. tuberculosis* toxins is the preference for highly structured RNAs required for translation (Jonathan W. Cruz et al., 2015; Schifano et al., 2016, 2013, 2014; K. Winther et al., 2016). In contrast to other VapCs previously characterized in other bacterial species, *M. tuberculosis* toxins target isoacceptor-specific elongating tRNAs rather than initiator tRNAs, pointing that translation initiation and elongation with untargeted tRNAs might still occur once these toxins are expressed. If that is the case, activation of these toxins by specific environmental cues could generate personalized proteomic responses according to the targeted tRNA isoacceptor. Moreover, tRNA biology has been historically studied under physiological conditions. Little is known about their involvement in diseases, particularly

infectious diseases. Due to their role in such an essential process (ie. protein synthesis), they could be used as a molecular switch exploited by bacteria to regulate gene expression and adapt to survive stresses. Therefore, determining the physiological impact of tRNA-targeting toxins in mycobacteria will help to decipher the molecular basis of persistence in TB infection and possibly result in new therapeutic strategies to eradicate this persistent pathogen.

Chapter 1: Toxin-mediated ribosome stalling reprograms the *Mycobacterium tuberculosis* proteome

Introduction

The molecular switches that enable *M. tuberculosis* to slow or stop replication, become dormant and establish latent tuberculosis infection are poorly characterized. Toxin-antitoxin (TA) systems/modules are thought to be involved in *M. tuberculosis* stress survival and the establishment of latent tuberculosis infection because they typically impart reversible growth inhibition in their host in response to stresses relevant to this state (Masuda & Inouye, 2017). TA systems are operons comprising adjacent genes encoding two small (~10 kDa) proteins, a toxin and its cognate antitoxin that inhibits toxin activity through formation of a stable TA protein-protein complex. Stress conditions lead to lower levels of the antitoxin and thus, a preponderance of free toxin which exerts its growth-regulating and/or other functions from within the bacterial cells (Masuda & Inouye, 2017). In fact, *M. tuberculosis* cells subjected to stresses relevant to latent tuberculosis infection—nutrient limitation (Albrethsen et al., 2013; Betts, Lukey, Robb, McAdam, & Duncan, 2002; Cortes et al., 2013; Tiwari et al., 2015), hypoxia (Ramage et al., 2009; Rustad, Harrell, Liao, & Sherman, 2008; Tiwari et al., 2015), macrophage infection (Cappelli et al., 2006; Fontan et al., 2008; Korch, Contreras, & Clark-Curtiss, 2009; Ramage et al., 2009) or antibiotic treatment (Denkin, Byrne, Jie, & Zhang, 2005; Provvedi, Boldrin, Falciani, Palù, & Manganello, 2009; R. Singh, Barry, & Boshoff, 2010; Tiwari et al., 2015)—exhibit

enhanced expression of TA toxins. Thus, the phenotypes associated with toxin expression in *M. tuberculosis* are consistent with a role for TA systems in the establishment and maintenance of latent tuberculosis and persistence of this pathogen.

Among the estimated 88 TA systems in *M. tuberculosis*, 11 belong to the MazEF family (Ramage et al., 2009; Sala et al., 2014). All MazF toxins are unified by their hallmark sequence-specific endoribonuclease activity. With a single exception, all MazF toxins in bacteria cleave mRNA and/or rRNA at specific three-, five-, or seven-base recognition sequences (Masuda & Inouye, 2017). The exception, the *M. tuberculosis* MazF-mt9 toxin (aka MazF7, Rv2063A), specifically recognizes and cleaves tRNA based on both sequence and structure determinants (Schifano et al., 2016; Schifano & Woychik, 2017). However, the series of downstream events that lead to growth arrest following MazF toxin-mediated cleavage of the *M. tuberculosis* target RNA are not well understood, especially for tRNA-cleaving toxins.

It has been widely assumed that since these toxins cleave one or more RNAs involved in protein synthesis—mRNA, rRNA and/or tRNA—they arrest growth by global translation inhibition (Harms et al., 2018; Masuda & Inouye, 2017; Walling & Butler, 2019; Wen, Behiels, & Devreese, 2014). However, deployment of 88 toxins to reach the same endpoint represents a redundancy that is at odds with the relatively compact *M. tuberculosis* genome adapted for survival within host granulomas during latent tuberculosis. Here we report the molecular mechanism of toxin MazF-mt9 which demonstrates that RNA-cleaving

M. tuberculosis toxins do not necessarily act by simply inhibiting translation, it illuminates a sophisticated mode of transcriptome recalibration and proteome reprogramming through highly selective ribosome stalling.

Methods

Strains, plasmids and reagents

All experiments were performed using either *M. tuberculosis* strain H37Rv (ATCC 25618), *M. tuberculosis* mc² 6206 (Δ panCD Δ leuCD, generously provided by William Jacobs laboratory, Albert Einstein College of Medicine), *M. smegmatis* mc² 155 (ATCC 700084), or *Escherichia coli* BW25113 Δ 6. *E. coli* BW25113 Δ 6 (a gift from the Masayori Inouye laboratory, Rutgers RWJMS) is a derivative of BW25113 (Guzman, Belin, Carson, & Beckwith, 1995) that lacks loci for six *E. coli* chromosomal TA modules (mazEF, chpBIK, relBE, hipBA, yefM-yoeB and dinJ-yafQ). For experiments using mycobacteria, MazF-mt9 (Rv2063A locus) was amplified by PCR from genomic DNA extracted from *M. tuberculosis* H37Rv, using oligonucleotides containing ClaI and EcoRV restriction sites at the 5' and 3' ends respectively (NWO1646 – 5' CCCATCGATGGCTGAGCCACGGCGAG 3' and NWO1647 – 5' CGATATCTCACGGGTCCCGGCCACC 3'). The resulting fragment was cloned adjacent to an anhydrotetracycline (ATC)-inducible promoter in the pMC1s plasmid. MazF-mt9 expression was induced by adding ATC (Clontech) to the media to a final concentration of 200 ng/ml.

M. tuberculosis H37Rv and *M. smegmatis* mc² 155 cells were grown in 7H9 Middlebrook medium with a final concentration of 0.05% Tween 80, 0.5%

bovine albumin, 0.2% dextrose, 0.085% NaCl (7H9-TW80-ADN) and 20 µg/ml kanamycin or zeomycin a 25 µg/ml. The attenuated strain mc² 6206 was grown in 7H9 Middlebrook media containing 1x OADC supplement (Sigma), 0.05% of tyloxapol, 50 µg/ml of pantothenic acid, 100 µg/ml of leucine, and kanamycin at 20 µg/ml.

For co-expression experiments in *E. coli*, MazF-mt9 was cloned into the arabinose-inducible vector pBAD33 (Guzman et al., 1995) and *M. tuberculosis* RNase J (Rv2752c) was cloned into the IPTG-inducible pinIII plasmid. *E. coli* harboring mazF-mt9-pBAD33 and/or RNase J-pinIII was grown in M9 minimal media using glycerol as carbon source and chloramphenicol at 25 µg/ml or ampicillin at 100 µg/ml as selection markers. The inducers (arabinose and/or IPTG) were added to the media at final concentration of 0.2% and 1 mM, respectively. All growth experiments were conducted at 37°C and under constant agitation at 230 rpm.

For co-expression experiments in *M. smegmatis*, codon-adjusted mCherry was cloned into the nitrile-inducible plasmid pNIT-zeo. *M. smegmatis* containing both pMC1s-mazF-mt9 and pNIT-zeo-mCherry were grown in triplicate to OD_{600nm} 0.2 and co-induced for 5 h by adding ATC and isovaleronitrile to a final concentration of 200 ng/ml and 1 µM, respectively. mCherry fluorescence (Ex: 585 nm/Em: 610 nm) and ODs were measured on a Synergy HT spectrophotometer (Biotek) and normalized against a -ATC (no MazF-mt9 induction) control group. Background fluorescence signal was inferred by measuring mCherry uninduced cultures and subtracted from all other groups.

RNA isolation

To isolate RNA from cells expressing MazF-mt9, bacteria were grown in the absence or presence of inducer for 2 or 7 days (corresponding growth curve in Figure 6) for *M. tuberculosis*, 7.5 hours for *M. smegmatis* or 60 min for *E. coli*. ATC was replenished into the media every 48 hours, as necessary.

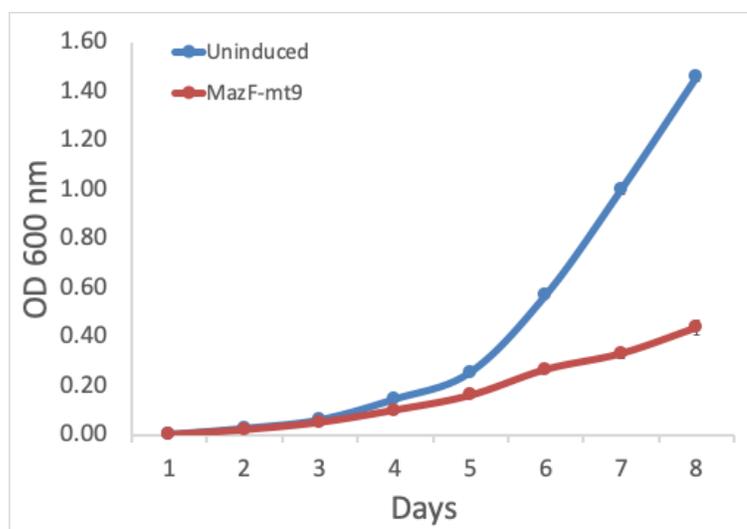


Figure 6. MazF-mt9 expression leads to slow growth in *M. tuberculosis*. *M. tuberculosis* cultures harboring pMC1s-MazF-mt9 were grown in 7H9 media in the presence (red) or absence (blue) of inducer (ATC) for 8 days. To estimate growth, absorbance at 600 nm was measured every 24 h. Error bars represent the standard error of the mean (S.E.M.) obtained from three biological replicates.

Cells were centrifuged at $2000 \times g$ for 10 min at 4°C , and supernatants were removed. Mycobacterial cell pellets were washed three times with PBS resuspended in 1 ml of Trizol and approximately 200 μl of 0.1 mm zirconia or

glass beads were added to the tube. Cells were lysed in 4 cycles of 30 s at 9000 rpm using Precellys Evolution homogenizer (Bertin Corp.) or 4 times at 7000 rpm for 45 s using a MagNA Lyser (Roche) with 1 min intervals on ice in between each cycle. The lysate was centrifuged at 12000 rpm at 4°C and total RNA was extracted from the supernatant using Direct-zol™ RNA MiniPrep Plus (Zymo Research). The flow-through from the first centrifugation step was saved for protein isolation and proteomics analysis. RNA was eluted in nuclease-free water, treated with TURBO DNase (Thermo Fisher) for 30 min at 37°C, and re-purified using the same RNA extraction kit. The extracted RNA was quantified by spectrophotometry using a µCuvette in a BioSpectrometer (Eppendorf) or Nanodrop (Thermo Fisher).

Northern analysis of tRNA levels

To detect mycobacterial tRNA, specific DNA probes complementary to the 3' end of *M. tuberculosis* tRNA^{Lys43-UUU}, tRNA^{Lys19-CUU} and tRNA^{Pro14-GGG} genes (NWO2517 5'-TGCCCCCACCAGGACTCGAACCTGGGAC-3', NWO2570 5'-GCCGTCAGGGTTTCGAACCCC-3' and NWO2518 5'-GTCAAGTGGTCGCAGGTTCAAATCCTGTCAGC-3', respectively) were radiolabeled at the 5' end with T4 polynucleotide kinase (New England Biolabs) and [γ -32P]ATP (PerkinElmer) for 1 h at 37°C. Then, total RNA (2 µg) from *M. tuberculosis* was resolved on a 9% polyacrylamide, 7 M urea gel and stained with SYBR Gold (Invitrogen) to assess and ensure overall quality. The RNA was

transferred to a nylon Hybond-N+ membrane (GE Healthcare) and hybridized with the ³²P-labeled oligonucleotides at 45°C overnight. To further remove non-specific signals, the membrane was washed twice with 0.01% SDS and 0.1X SCC at 50°C for 10 min. The membranes were exposed to phosphorimager screens and scanned using the Typhoon FLA 9500 (GE Healthcare) image system.

Labeling of newly synthesized *M. tuberculosis* proteins

To assess the levels of newly synthesized proteins in *M. tuberculosis* after toxin induction, MazF-mt9 was induced for 72 h and cultures (uninduced and induced) were incubated with AHA for 5 h, in duplicate. Cells were then pelleted and lysed in 4 cycles of 30 s at 9000 rpm using Precellys Evolution homogenizer (Bertin). The lysate was centrifuged, and the supernatant used as input in the Click-IT Protein Reaction Buffer kit (ThermoFisher), linking the AHA-containing proteins with an alkyne-containing version of the fluorophore TAMRA (ThermoFisher). 20 ug of protein for each sample were run on a SDS-PAGE gel and subsequently scanned using the Typhoon FLA 9500 (GE Healthcare) image system.

5' RNA-seq

Preparation of 5'-dependent libraries was performed as described in Schifano et al. (Schifano et al., 2014). Briefly, for sequencing of RNAs containing 5' hydroxyl ends (5'-OH), total RNA from pMC1s-mazF-mt9 (+ATC and -ATC) samples was digested using 1 U of Terminator 5'-Phosphate-Dependent Exonuclease (Epicentre) to remove RNAs containing 5'-monophosphate (5'-P). Subsequently, the RNA was purified using RNA Clean & Concentrator™-5 (Zymo Research) and 5'-OH ends were phosphorylated using 3 U of T4 PNK (New England Biolabs) to allow subsequent adapter ligation. The Illumina small RNA 5' adapter (5'- GUUCAGAGUUCUACAGUCCGACGAUCNNNNNN - 3') was ligated using T4 RNA ligase 1 at 16°C overnight (New England Biolabs), according to the manufacturer's instructions. Excessive unbound adapters were removed by gel excision and purification. The purified RNA was used in a Superscript IV reverse transcription reaction (ThermoFisher) using the primer 5'- GCCTTGGCACCCGAGAATTCCANNNNNNNNN-3' and the resulting cDNA was gel extracted, selecting fragments from 80 to 500 nts. The cDNA libraries were amplified in 12 cycles of PCR using Phusion HF DNA polymerase (ThermoFisher). The oligonucleotides used for PCR amplification were RP1 (5'- AATGATACGGCGACCACCGAGATCTACACGTTTCAGAGTTCTACAGTCCGA - 3') and RPIX (5'- CAAGCAGAAGACGGCATACTGAGATNNNNNNGTGACTGGAGTTCCTTGGCA CCGAGAATTCCA-3'), where underlined N's represent the individual Illumina barcodes for each library. After electrophoresis on a 10% (wt/vol) polyacrylamide

gel, amplified DNA between the sizes 150 bp and 450 bp was isolated by gel excision and sequenced in an Illumina NextSeq 500 or HiSeq 2500 platform.

The FASTQ files generated by the 5' RNA-seq protocol had the adapter and the first 6 nucleotides of the high quality end trimmed using Trimmomatic (Bolger, Lohse, & Usadel, 2014). Then, the sequences were trimmed to 20 nts (discarding shorter sequences) and aligned either to *M. tuberculosis* genome (Genbank accession: AL123456), *M. smegmatis* (Genbank accession: CP000480) or to *E. coli* (Genbank accession: NC_000913) using Bowtie 1.2.0 applying the parameters `-n 0 -l 20` (Langmead, Trapnell, Pop, & Salzberg, 2009). For each nucleotide in the genome, we calculated the number of reads that started at that position (i.e. the amount of RNA molecules that had their 5' end starting at that nucleotide in the genome). Read counts were normalized to sequencing depth and expressed as "reads per million of mapped reads" (rpm). Next, we calculated the ratio for each position (counts in induced sample / counts in uninduced sample) to generate a fold change. Positions that had 0 counts in the uninduced library were adjusted to a pseudo-count of 1. We only considered reads that had at least 5 rpm for mRNAs and 50 rpm for tRNAs in the induced sample and a ratio of fold change of at least 10. For nucleotide frequency visualization, we ran weblogo (Crooks, 2004) on the top hits using the default parameters. kpLogo (Wu & Bartel, 2017) was used adjusting to 3-mers to preferentially find codons in the experiments using RNase J and MazF-mt9 co-expressing *E. coli*; the Bonferroni corrected p-value was used to find significant k-mer positions. The significant hits from the 5' RNA-seq from a culture

expressing RNase J only (compared to an uninduced culture) was used as background hits in the KpLogo program in order to remove RNase J cleavage products in the absence of MazF-mt9.

RNA-seq

In order to deplete 16S/23S ribosomal RNA from the extracted RNA, the samples were treated using the RiboZero (Illumina) kit for bacteria. Approximately 100 ng of rRNA-depleted RNA were used to generate the libraries using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs) and sequenced on an Illumina HiSeq 2500 or similar. The resulting sequences were mapped to the *M. smegmatis* reference genome (Genbank accession: CP000480) using the default parameters of Bowtie 1.2. 37, Stringtie 1.3.4 37 and DESeq2 2.11.40.2 38 programs were used for transcript assembly and differential expression analysis, respectively.

Ribo-seq

In order to confirm ribosome stalling, we analyzed the position of translating ribosomes on the mRNA through Ribo-Seq (Becker, Oh, Weissman, Kramer, & Bukau, 2013; Ingolia, Brar, Rouskin, McGeachy, & Weissman, 2012). *M. smegmatis* or *M. tuberculosis* cells were grown to OD_{600nm} = 0.2, separated into induced (+ATC) and uninduced (-ATC) cultures and incubated at 37°C for

7.5 h (*M. smegmatis*) or 72 h (*M. tuberculosis*). The cells were harvested by quick filtration using a 0.22 µm filter apparatus. Immediately after, cells were quickly scraped off the filter surface using a scoopula and flash frozen in liquid nitrogen. Cells were ruptured while frozen through bead beating with the Precellys Evolution (Bertin Corp). The procedure was performed at 4 degrees Celsius using 7-ml metal tubes and metal beads for 8 cycles of 30 s at 6000 rpm. Cells were kept in liquid nitrogen for 30 s between each cycle. After lysis, the resulting cell powder was resuspended in lysis buffer, and the Becker et al. no-crosslinking protocol was followed (Becker et al., 2013). The A-site positioning was predicted using RiboTools (Michel et al., 2016) and the coverage around specific codons was analyzed using the bamCoverage of the DeepTools (Ramírez et al., 2016) package.

Construction of RNase J deletion strain

An RNase J mutant strain (MSMEG_2685) was constructed using the ORBIT recombineering method (Woolstenhulme, Guydosh, Green, & Buskirk, 2015). We designed the Bxb1 attP-containing oligonucleotide with the first and last 70 base pairs of the MSMEG_2685 gene (NWO3007, 5'-

TCAGATCTCTATGACGGTCTGGGACGATCATCGGCTGCCTGCGGTAGGTCTC
 GCCCACCCACTTGCCGACCGGTTTGTACCGTACACCACTGAGACCGCGGT
 GGTTGACCAGACAAACCTACCGCCCAGGGCGGTGACACGCAGTCCGCCCG
 GAGCGAGCGGTGGTGGCGGCGCGAGTTCGGCGCTCAT -3'). To confirm

successful deletion, we performed PCR using oligonucleotides NWO3007 (5' - GGG GAT TCA TGA GCG CCG A – 3') and NWO3009 (5' - ACG CGT ATG TCG CGT TGG AG – 3'), which are complimentary to the flanking regions of the deleted gene, and performed whole genome sequencing.

Proteomics

Total protein samples in Trizol were incubated with an equal volume of isopropanol for 30 min at RT with constant mixing. Next, the mixture was centrifuged at 12,000 g for 5 min at 4°C and the supernatant was removed. The protein pellets were incubated with 2 mL of 0.3 M guanidine hydrochloride in 95% ethanol for 20 min on an orbital shaker at room temperature and centrifuged at 7,500g for 5 min. This was repeated for a total of three rounds. After the third round, 2 ml of ethanol was added to the pellet and incubated for 20 min. The supernatant was decanted, and pellets were dried for 5 min followed by resuspension in 9.5 M urea/2% CHAPS pH 9.1. Protein concentrations were assessed using the Pierce™ 660nm Protein Assay (ThermoFisher). Extracted proteins were run into SDS-PAGE as gel plug and in-gel digested with trypsin using standard procedure.

For assessing newly synthesized proteins, *M. smegmatis* cultures (induced or uninduced, in triplicate) were incubated for 2 h with azidohomoalanine (AHA, ThermoFisher). AHA is an azide-containing methionine analog that is incorporated into proteins, allowing the capture of the newly

synthesized, AHA-incorporated proteins by click chemistry. We selectively captured the newly synthesized proteins using an alkyne-containing column from the Click-iT™ Protein Enrichment Kit (ThermoFisher), following the manufacturer protocol, and proceeded with in-column trypsin digestion.

Digests were analyzed using a Q Exactive HF tandem mass spectrometer coupled to a Dionex Ultimate 3000 RLSCnano System (Thermo Scientific). Samples were solubilized in 5% acetonitrile/0.1% TFA and loaded onto a fused silica trap column Acclaim PepMap 100, 75 µm x 2 cm (ThermoFisher). After washing for 5 min at 5 µl/min with solvent A (0.2% formic acid), the trap column was brought in-line with an analytical column (Nanoease MZ peptide BEH C18, 130A, 1.7 µm, 75 µm x 250mm, Waters) for LC-MS/MS. Peptides were eluted using a segmented linear gradient from 4 to 90% B (B: 0.08% formic acid, 80% acetonitrile): 4% B for 5min, 4–15% B for 30 min, 15-25% B for 40 min, 25-50% B for 44 min and 50-90% B for 8 min. Mass spectrometry data was acquired using a data-dependent acquisition procedure with a cyclic series of a full scan with resolution of 120,000 followed by MS/MS (HCD, relative collision energy 27%) of the 20 most intense ions and a dynamic exclusion duration of 20 sec.

The raw LC-MS data was converted into MASCOT Generic Format (MGF) using Proteome Discover 2.1 (ThermoFisher) and searched against either the NCBI M. tuberculosis database (Accession: AL123456) together with a database of common laboratory contaminants (<http://www.thegpm.org/crap/>) using a local implementation of the global proteome machine (GPM Fury) (Beavis, 2006).

Differential expression was estimated using the QLSpline option of the QuasiSeq package, only using proteins having 10 or more spectral counts total (from the 12 LC-MS/MS runs on the six different samples (<https://cran.r-project.org/web/packages/QuasiSeq/index.html>) (Lund, Nettleton, McCarthy, & Smyth, 2012). Data is presented as estimated log₂ ratios of Induced/uninduced. Q-values are calculated using the fdrtool package of Strimmer (Strimmer, 2008) with a q-value cutoff of 0.05 to consider significant change.

Data availability

The sequencing datasets generated in this study were deposited in the NCBI Sequence Read Archive under BioProject accession number “PRJNA490371 [<https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA490371>]”. Mass spectrometry data were deposited in the MassIVE database under the accession number “MSV000083670 [<ftp://massive.ucsd.edu/MSV000083074>]”.

Results

MazF-mt9 inactivates a single tRNA in *M. tuberculosis*. In our earlier work, tRNA was identified as the primary target of MazF-mt9—the first MazF toxin to exhibit a preference for tRNA—but we had not expressed this toxin in *M. tuberculosis* cells to identify the true target in vivo. Here we enlisted our specialized RNA-seq method (Schifano et al., 2014), 5' RNA-seq, to specifically

identify the RNA(s) cleaved by the MazF-mt9 toxin in *M. tuberculosis* cells, as well as the precise site of cleavage within the RNA(s). 5' RNA-seq differentially detects one or more subpopulations of RNA depending on the modification present at the 5' end of the transcript. The 5' RNA-seq method used here selectively detected transcripts with a 5'-hydroxyl (OH) moiety generated by MazF-mt9 and other MazF family toxins (Zhang, Zhang, Hara, Kato, & Inouye, 2005).

Only one tRNA, the Lys tRNA^{Lys43-UUU} isoacceptor, was identified as the primary target of MazF-mt9 when 5' RNA-seq was performed on *M. tuberculosis* H37Rv cells expressing MazF-mt9 versus control cells (Figure 8A). None of the other 44 *M. tuberculosis* tRNA species were cleaved by MazF-mt9, not even the other isoacceptor Lys tRNA, tRNA^{Lys19-CUU} (Figure 7a,c).

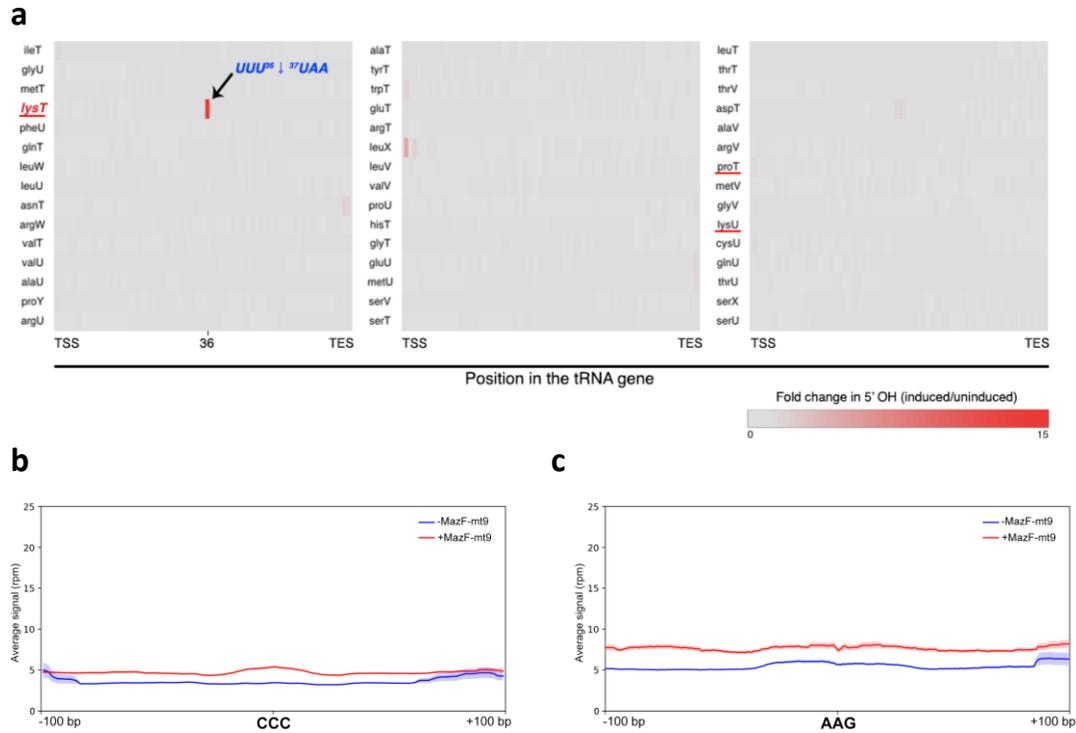


Figure 7. Only tRNA^{Lys43-UUU} is cleaved by MazF-mt9 in *M. tuberculosis*. a) Heat map representing fold changes in tRNA cleavage of all 45 tRNA species when MazF-mt9 is expressed for 2 days in H37Rv *M. tuberculosis* cells relative to the uninduced control. The previously identified in vitro targets *proT* and *lysU*, which were not confirmed in vivo, are underlined in red. TSS, transcription start site; TES, transcription end site. Cleavage of tRNA^{Lys43-UUU} is shown at position 36 within the anticodon. b-c, Average read counts (reads per million, rpm) of mapped ribosome footprints surrounding the cognate codons of the in vitro mistargets *proT* tRNA (b) or the other Lys isoacceptor tRNA *lysU* (c) in *M. tuberculosis*, showing no significant stalling. Blue or red solid line represent uninduced cells or cells expressing MazF-mt9, respectively, with the light-colored shading denoting standard deviation.

Cleavage occurred before the third U, i.e. $^{35}\text{UU}/\text{U}^{37}$, within the anticodon sequence (Figure 8B). This contrasted with our earlier 5' RNA-seq of *M. tuberculosis* RNA incubated with recombinant MazF-mt9 (Schifano et al., 2016). In that in vitro experiment, both $\text{tRNA}^{\text{Lys43-UUU}}$ and $\text{tRNA}^{\text{Pro14-GGG}}$ were targeted by MazF-mt9. Both tRNAs contain a UUU consensus sequence, $\text{tRNA}^{\text{Lys43-UUU}}$ within its single-stranded anticodon-loop or $\text{tRNA}^{\text{Pro14-GGG}}$ in its D-loop. Many UUU-containing transcripts were also cleaved in vitro when the secondary structure of the UU/U cleavage site and flanking sequences happen to mimic that of a tRNA anticodon stem loop (Schifano et al., 2016).

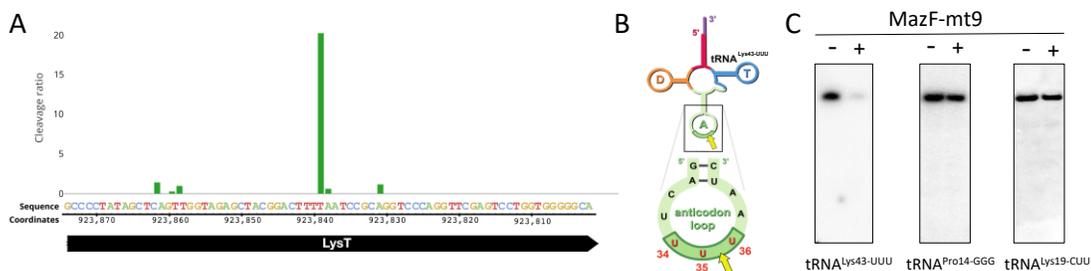


Figure 8. MazF-mt9 targets only tRNA^{Lys43-UUU} in vivo. A, Histogram representing the ratio of cleavage by MazF-mt9 identified using 5' RNA-seq at each nucleotide within the *lysT* gene (tRNA^{Lys43-UUU}) in *M. tuberculosis* after 7 days of toxin induction. Genomic positions and the negative strand sequence are shown. B, Representation of the only MazF-mt9 target, tRNA^{Lys43-UUU} (adapted from Schifano et al.15). Anticodon (UUU), in red; cleavage site, yellow arrow. C,

Northern analysis of RNA from \pm MazF-mt9 *M. tuberculosis* cells using the tRNA isoacceptor-specific oligonucleotides indicated.

Since three tRNA targets were previously identified in vitro, we used northern analysis to validate our in vivo 5' RNA-seq result, which identified tRNA^{Lys43-UUU} as the sole tRNA target (Figure 8C). Consistent with our in vivo 5' RNA-seq data, only tRNA^{Lys43-UUU} was cleaved upon expression of MazF-mt9 in *M. tuberculosis* cells. Therefore, the MazF-mt9 toxin needs to be expressed in living cells, not simply incubated with *M. tuberculosis* RNA in vitro, to accurately pinpoint the sole tRNA target. This requirement for in vivo expression in the natural host is likely broadly relevant for accurate target identification for the other *M. tuberculosis* RNA-cleaving toxins.

5' RNA-seq suggests ribosome stalling at AAA Lys codons

In addition to identification of tRNA^{Lys43-UUU} as the primary target of MazF-mt9, three unique features were uncovered in the 5' RNA-seq dataset that suggested ribosome stalling was intrinsic to the mechanism of action of MazF-mt9. First, in addition to lysine tRNA^{Lys43-UUU}, there were nearly two hundred mRNAs that were also significantly cleaved (fold change > 10) when MazF-mt9 was expressed. Second, there was a conspicuous fixed distance, ~15 nt, between the cleavage site in these mRNAs and the presence of an in-frame AAA

Lys codon Figure 9A,B). Third, there was no consensus sequence at the cut site (Figure 9A,B), indicating that these mRNAs were not cleaved by MazF-mt9 (which only cleaves UU/U sequences in the proper structural context (Schifano et al., 2016)). Instead, these mRNAs were cut by another RNase, whose activity was detectable because it also generated an RNA cleavage product with a 5'-OH.

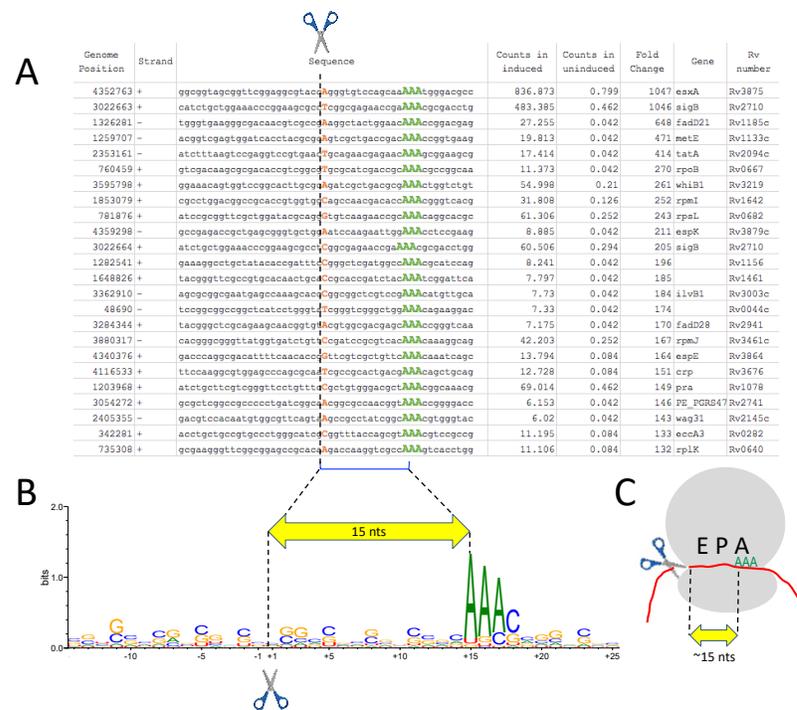


Figure 9. 5' RNA-seq captures ribosome stalling events at AAA Lys codons.

A, Top mRNA hits detected by 5' RNA seq, ± 25 nts surrounding the cleavage site (scissor). The first nucleotide of each mapped read, orange; AAA Lys codons, green. Counts are represented as reads per million (rpm). B, Nucleotide frequency logo (weblogo) at each position constructed from the top 50 of the 181

mRNA hits identified. Positions numbered relative to the cleavage site. **C**, Schematic of a ribosome (grey) and the ~15 nt length (yellow arrow) from the 5' end of transcript to the A-site.

We first investigated the significance of the 15 nt periodicity in the 5' RNA-seq dataset. The precise distance between the 5'-OH cleavage sites and the AAA Lys codons (~15 nts) suggested that ribosomes were stalling at the AAA Lys codon because its cognate tRNA^{Lys43-UUU} was depleted. It is known that the bacterial ribosome footprint is ~28 nt and the distance from the P-site to the 3' end of this footprint is 15 nts (Woolstenhulme et al., 2015). Therefore, we posited that the ~15 nt gap between the 5' RNA-seq cleavage site and the in-frame AAA Lys codon represented the ribosome footprint from the 5' end to the tRNA^{Lys43-UUU} codon at the A-site (illustrated in Figure 9C).

Ribo-seq confirms ribosome stalling

To confirm that the ~15 nt gap between the cleavage site and the in-frame AAA Lys represented ribosome stalling at AAA codons, we performed Ribo-seq of MazF-mt-expressing *M. smegmatis* and *M. tuberculosis* cells (Figure 10). MazF-mt9 cleaves tRNA^{Lys23-UUU} and exhibits the same telltale markers of stalling as *M. tuberculosis* when analyzed by 5' RNA-seq. Four representative transcripts, two from *M. smegmatis* (hisC and cobN, Figure 10A,B) and two from *M. tuberculosis* (nrdB and Rv0178, Figure 10E,F) revealed exclusive stalling at

in-frame Lys AAA codons. There was no ribosome stalling at out-of-frame AAA codons (blue arrows, Figure 10A,B and E,F). In fact, this striking preference for ribosome stalling at AAA Lys codons was consistent throughout the transcriptomes of each mycobacteria species when stalling was analyzed at AAA codons only (Figure 10 C,G) or when codon occupancy for mapped ribosomal footprints of all codons was compared (Figure 10D,H). Not only was there a preference for AAA Lys codons, stalling occurred predominantly at the first AAA Lys codon in any given mRNA, as the number of stalled ribosomes captured at the second, third or fourth AAA Lys codon by Ribo-seq successively diminished (Figure 11). Consistent with the northern analysis in Figure 8C, we did not observe ribosome stalling at tRNA^{Pro14-GGG} (that we previously reported was cleaved by MazF-mt9 in vitro (Schifano et al., 2016)) or at the other Lys isoacceptor tRNA^{Lys19-CUU} (Figure 7b,c).

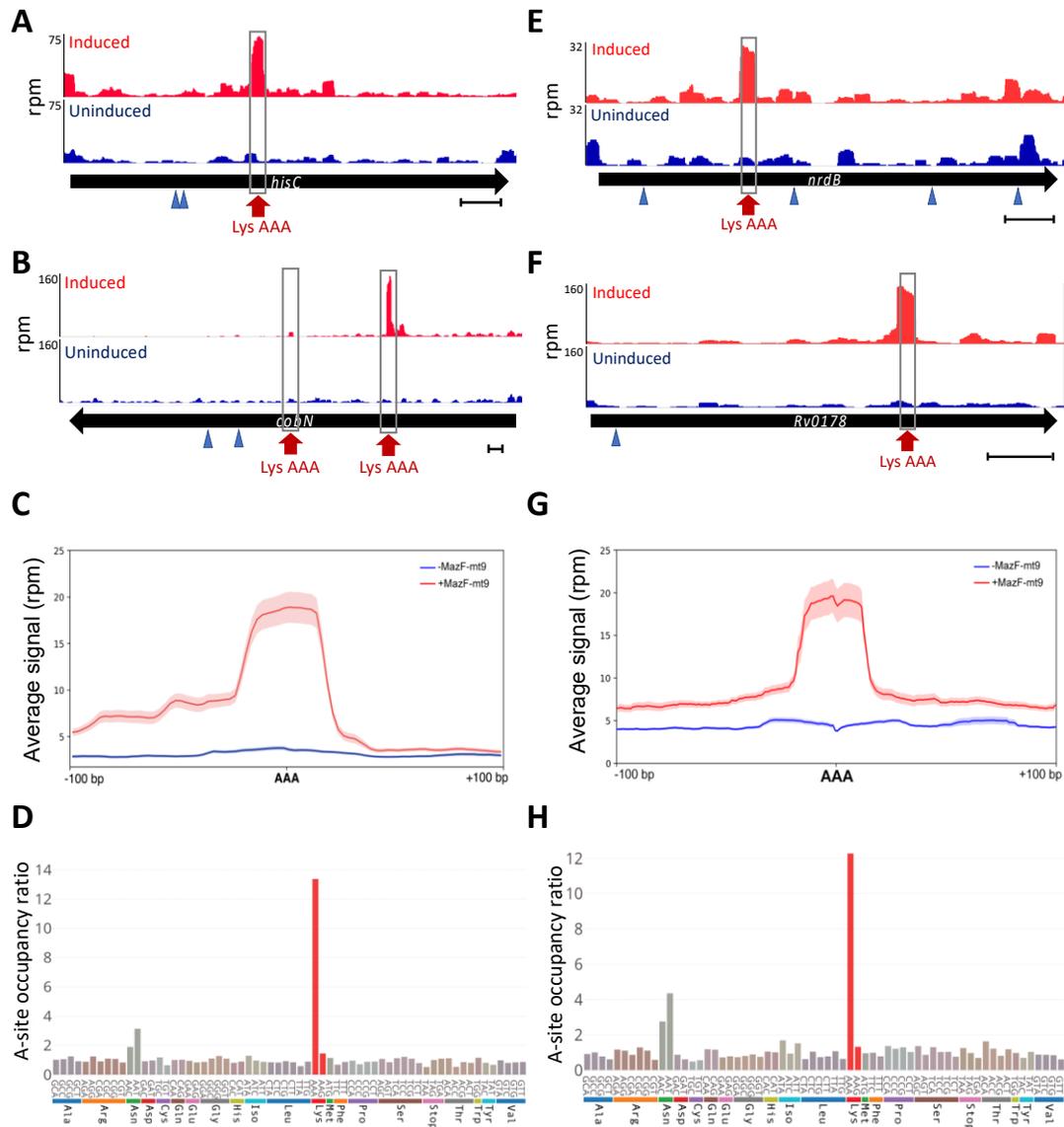


Figure 10. Ribo-seq confirms transcriptome-wide AAA Lys-specific ribosome stalling. A-B, E-F, Ribo-seq read coverage in reads per million (rpm) of AAA Lys-containing *M. smegmatis* (a-b) or *M. tuberculosis* (E-F) transcripts in \pm MazF-mt9 cells (arrows indicate direction of translation). The location of in-frame AAA Lys codons and out-of-frame AAAs are indicated (red arrow/grey shading and blue arrow head, respectively). Scale bars illustrate a 100 bp

genomic distance. C,G, Average read counts (rpm) of mapped ribosomes surrounding AAA Lys codons in +MazF-mt9 (red line) or -MazF-mt9 (blue) cells; light shading represents the standard deviation. D, H, Ratio of codon occupancy (induced vs. uninduced) at the predicted ribosomal A-site position identified in the mapped ribosomal footprints, grouped by amino acid or stop codon. Scale bars represent a 100 bp range.

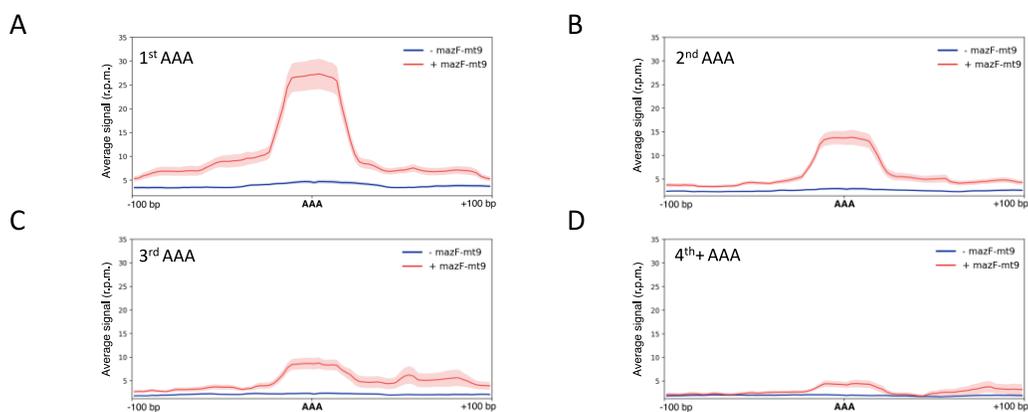


Figure 11. Ribo-seq shows ribosome stalling preference for early AAA

codons. (A-D) Average read counts (reads per million, rpm) of mapped ribosome footprints surrounding the first (A), second (B), third (C) or fourth and higher (D) AAA Lys codon of *M. smegmatis* genes. Blue or red solid line represent uninduced cells or cells expressing MazF-mt9, respectively, with the light-colored shading denoting standard deviation.

Ribosome stalling at AAAs upon RNase J expression in *E. coli*

Next we sought to identify the RNase that cut on the 5' side of the stalled ribosome and generated a 5'-OH detectable by 5' RNA-seq. Although typically involved in rRNA maturation and mRNA turnover (Mathy et al., 2007), we suspected RNase J because it has been reported to cleave on the 5' side of erythromycin-stalled ribosomes in *Bacillus subtilis* ~15 nt from the A-site (Yao, Blaustein, & Bechhofer, 2008). Therefore, *M. tuberculosis* RNase J (Rv2752c) was ectopically expressed in *E. coli* cells to determine if ribosome stalling was now detectable. We chose *E. coli* because it naturally lacks this enzyme and has no other known RNase J-like activity. Also, MazF-mt9 expression in *E. coli* has two essential features that enable us to assay for RNase J gain-of-function activity. First, as in *M. tuberculosis*, the only tRNA that MazF-mt9 cleaves in *E. coli* is tRNA^{LysUUU} (Schifano et al., 2016). Second, even though tRNA^{LysUUU} was exclusively cleaved, there was no evidence of ribosome stalling in our 5' RNA-seq dataset (Schifano et al., 2016).

The kplogo in Figure 12 illustrates the key features of the experimental design and outcome. As is the case with incubation of recombinant MazF-mt9 with total RNA from *M. tuberculosis*, MazF-mt9 expression in *E. coli* does not exhibit exclusive specificity for tRNA^{LysUUU}, it also cleaves some mRNAs at UU/U that are predicted to form structures that recapitulate the tRNA^{LysUUU} anticodon stem loop (reflected as a stacked UUU at -2 using kplogo in Figure 12A). However, the presence of these cleaved mRNAs did not interfere with interpretation of this experiment because we looked for a phenotype independent of the direct cleavage of mRNA at UU/U sequences by MazF-mt9. Instead, we tested whether

the presence of RNase J now enabled 5' RNA-seq detection of stalled ribosomes at AAA Lys codons in mRNAs just as it had in *M. tuberculosis*. Comparison of 5' RNA-seq datasets of *E. coli* cells expressing MazF-mt9 +/- RNase J revealed that ribosome stalling only occurred when RNase J was present (as indicated by the stacked AAA codon appearing on the kplogo at a distance consistent with ribosome stalling; Figure 12 A vs. B). This result suggested that RNase J collaborates with MazF-mt9 in *M. tuberculosis* cells, i.e. the toxin first inactivates a single tRNA (tRNA^{Lys43}-UUU) and RNase J subsequently recognizes and cleaves mRNA on the 5' side of ribosomes stalled at AAA Lys codons.

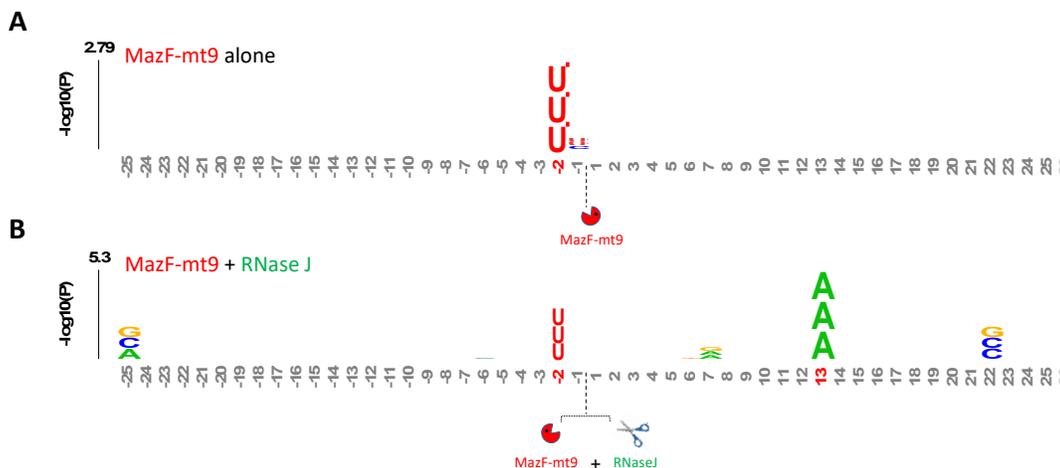


Figure 12. *M. tuberculosis* RNase J cleaves 5' of stalled ribosomes favoring expression of AAA-deficient transcripts. KpLogo graphs derived from *E. coli* 5' RNA-seq +MazF-mt9 (A) or +MazF-mt9 +RNase J (B) datasets compared to -MazF-mt9 samples. The enriched k-mers in the datasets are stacked vertically at the position of their initial nucleotide. The positions are numbered relative to the

cleavage site (dotted line), with those achieving statistical significance (Bonferroni corrected p-value < 0.01) shown in red.

We next deleted the RNase J gene in *M. smegmatis* (MSMEG_2685) to more rigorously test the role of RNase J in mycobacteria. In contrast to the gain-of-function phenotype we observed upon coexpression of Rv2752 with MazF-mt9 in *E. coli*, we were surprised to find that there was no loss-of-function without MSMEG_2685 because we could still detect ribosome stalling by 5' RNA-seq (Figure 13).

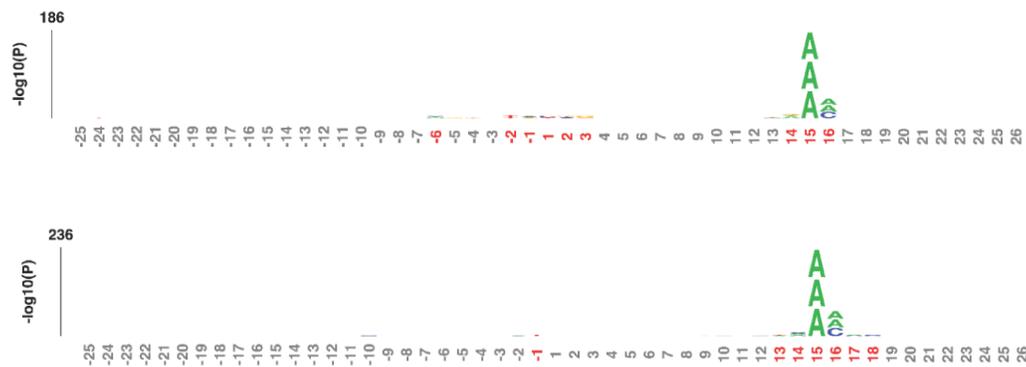


Figure 13. Ribosome stalling occurs upon expression of MazF-mt9 in a RNase J-deletion background. An *M. smegmatis* mutant lacking the entire RNase J gene (MSMEG_2685) was analyzed with 5' RNA-seq after expressing MazF-mt9 for 5 h and compared to an uninduced mutant sample. Positions are numbered related to the cleavage site. Enrichment of Lys AAA codon is still seen at position +15, suggesting the existence of a functionally compensatory RNase in mycobacteria. The wild-type (no MSMEG_2685 RNase J deletion) k-plot shows

essentially the same pattern of stalling with the stacked AAA sequence ~15 nts from the cleavage site.

Given the clear result in *E. coli*, both *M. smegmatis* and *M. tuberculosis* genomes might encode an additional functionally related enzyme with RNase J-like activity that complements the enzymatic activity we deleted. There is precedent for this type of functional redundancy as *Bacillus subtilis* has two related RNase J genes (Condon, 2010; Jamalli, Hébert, Zig, & Putzer, 2014; Mathy et al., 2010). However, in contrast to *B. subtilis*, the other *M. tuberculosis* enzyme(s) that compensates for loss of RNase J might elude identification because it (they) may lack significant sequence similarity to RNase J. MazF-mt9 itself serves as precedent for functional similarity without sequence similarity. It is the only known tRNA-cleaving MazF family member in bacteria (Schifano et al., 2016; Schifano & Woychik, 2017). Its substrate specificity and cleavage activity mimics many tRNA-cleaving VapC toxins even though MazF-mt9 completely lacks sequence or structural similarity to any VapC family member (Schifano et al., 2016; Schifano & Woychik, 2017).

Cleavage at stalled ribosomes favors AAA-deficient mRNAs

The single cleavage site before one or more stalled ribosomes at AAA Lys codons is expected to result in a truncated, nonfunctional mRNA. This, in turn, should result in a decrease in the abundance of the proteins encoded by these

transcripts. To test this directly, we metabolically labeled *M. smegmatis* cells with the Met mimetic azidohomoalanine (AHA) whose azido group can be coupled to an alkyne-containing reagent using click chemistry (Dieterich, Link, Graumann, Tirrell, & Schuman, 2006). AHA is efficiently incorporated into nascent proteins without cytotoxicity with a sensitivity comparable to [³⁵S]Met labeling. We selectively captured the newly synthesized proteins after MazF-mt9 was induced using an alkyne-containing column and performed an in-column trypsin digestion. The peptides released from the column were subjected to quantitative mass spectrometry. The results are graphed as a volcano plot where all individual proteins are represented as colored circles whose diameter and saturation is scaled to reflect AAA Lys codon content (Figure 14). The effect of AAA Lys content on the abundance of newly synthesized protein is striking, with a clear downward shift in synthesis of AAA Lys codon-containing proteins relative to the uninduced control. These data also clearly illustrate that MazF-mt9 does not inhibit protein synthesis globally as there are hundreds of proteins that are synthesized in the presence of MazF-mt9.

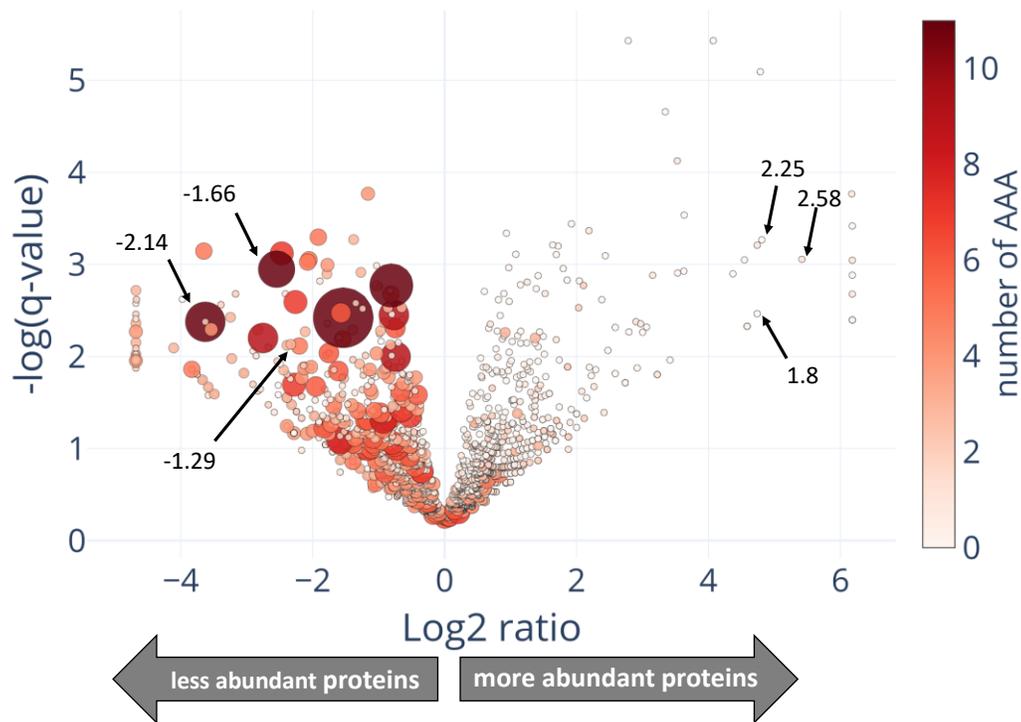


Figure 14. Induction of MazF-mt9 leads to global proteomic shifts based on the AAA codon content of transcripts. Volcano plot showing the proteomic changes in newly synthesized proteins between induced (+MazF-mt9) and uninduced cells. The diameter and color intensity of each represented protein (circle) is proportional to the number of AAA codons in its coding sequence. Fold changes in mRNA levels obtained by RNA-seq are shown for three examples of up- and down-regulated proteins (arrows).

We also pointed out three up- and down-regulated proteins on each side of the volcano plot and included the respective fold-changes of their mRNAs obtained by conventional RNA-seq (see arrows and corresponding +/- values in Figure 14). We extended the comparison of the three relevant parameters

uncovered in this study—AAA content, fold-changes in proteins, and fold-changes in mRNAs—to a genome scale (Figure 15). Comparison of these three analyses demonstrated that the anticipated trends were maintained genome-wide.

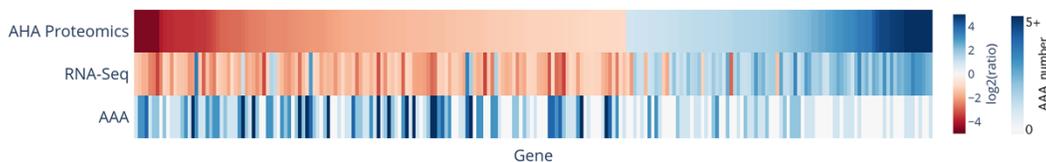


Figure 15. mRNA levels correlate with protein expression. Heatmaps comparing fold changes in newly synthesized proteins (AHA-labeling Proteomics, top row) and mRNA levels (RNA-Seq, middle row) in \pm MazF-mt9 cultures. Genes with at least a log₂ratio of ± 1 were considered. The number of AAA codons (bottom row) of each gene is shown.

Finally, we demonstrated sustained protein synthesis after MazF-mt9 induction in *M. tuberculosis*. After metabolically labeling with AHA, newly synthesized AHA-containing *M. tuberculosis* proteins were coupled to the fluorophore TAMRA by click chemistry and visualized after PAGE (Figure 16). Comparison of \pm MazF-mt9 samples revealed similar protein levels and band patterns (Figure 16).

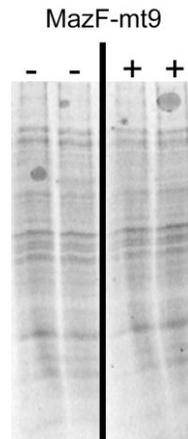


Figure 16. MazF-mt9 expression does not inhibit translation in *M.*

tuberculosis. Newly synthesized proteins from two biological replicates of *M. tuberculosis* cultures with (+) or without (-) MazF-mt9 expression were visualized by fluorescent imaging after metabolic labeling with AHA, click-chemistry coupling to the fluorophore TAMRA, followed by SDS-PAGE.

MazF-mt9 translation inhibition is codon specific

To further solidify that MazF-mt9 expression leads to AAA Lys codon-directed inhibition of protein synthesis, we compared the level of protein synthesis in two transcripts engineered to contain only AAA Lys codons or AAG Lys codon. We used the mCherry fluorescence reporter gene as a convenient vehicle for translation read-out. Consistent with the single tRNA^{Lys43-UUU} target and the series of downstream events documented above, we measured a 70% reduction in mCherry fluorescence in the AAA-containing gene relative to the AAG only counterpart (Figure 17).

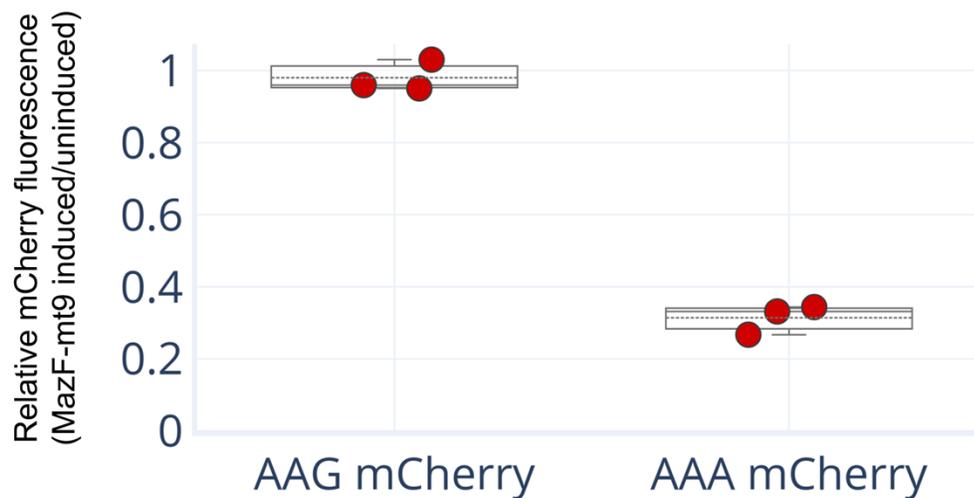


Figure 17. MazF-mt9 translation inhibition is codon specific. MazF-mt9 was co-expressed with two different versions of mCherry: one containing its lysine codons mutated to AAA (Lys-AAA mCherry) and the other to AAG (Lys-AAG mCherry). mCherry fluorescence was measured 5 h after induction of MazF-mt9 and normalized against an uninduced (-MazF-mt9) control. Each individual data point derives from a biological replicate (n = 3). Box plot center line, median; box limits, upper and lower quartiles; whiskers, max and min values.

Discussion

In summary, 5' RNA-seq captured the unexpected, highly precise “one-two punch” dealt by the MazF-mt9 toxin that would typically require parallel Ribo-seq and RNA-seq analyses. MazF-mt9 inactivated just one tRNA, tRNA^{Lys43-UUU} in *M. tuberculosis* cells through UU/U cleavage at its anticodon. Concordant with

the depletion of tRNA^{Lys43-UUU}, the dataset revealed compelling evidence of ribosome stalling exclusively at cognate AAA Lys codons (rare in the 65.6% GC-rich *M. tuberculosis* genome). These stalled ribosome signatures were detected in the 5' RNA-seq dataset because another RNase—possibly RNase J working in concert with one or more additional RNases—cleaved the transcript 5' of the stalled ribosomes. We confirmed that this RNase cleavage of AAA Lys codon-containing mRNA transcripts is associated with their removal from the transcriptome since the abundance of proteins encoded by AAA-containing mRNAs decreased in cells expressing MazF-mt9. Thus, we have tracked a succession of unexpected events set in motion by inactivation of only one tRNA in *M. tuberculosis* cells, which revealed an intricate mechanism for precise transcriptome recalibration that in turn reprograms the *M. tuberculosis* proteome (summarized in Figure 18).

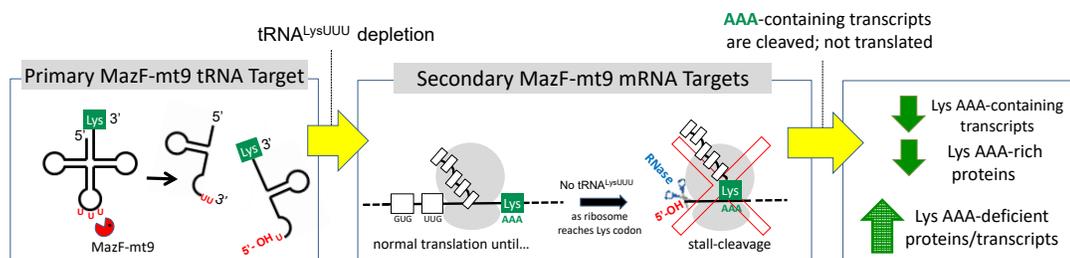


Figure 18. MazF-mt9 mechanism of action. The toxin controls gene expression through the cleavage of tRNA^{Lys43-UUU} (left panel), causing ribosome stalling and subsequent degradation of the mRNAs by one or more RNases (middle panel).

This cooperative action reduces the steady levels of AAA-rich proteins and favors an increase in AAA-deficient proteins such as virulence factors (right panel).

This cascade of molecular events is initiated by a single, highly specific, cleavage event at the 35UU/U37 site within the anticodon of tRNA^{Lys43-UUU}. MazF-mt9 specifically cleaves the tRNA^{Lys43-UUU}, no other tRNA, not even the other Lys tRNA isoacceptor tRNA^{Lys19-CUU}. Wobble rules dictate that tRNA^{Lys43-UUU} can recognize the less common Lys AAA codon (5.3/1000 codons), as well as the more frequently represented Lys codon in *M. tuberculosis*, AAG (15/1000 codons). However, MazF-mt9 cuts and inactivates only tRNA^{Lys43-UUU}, resulting in a short supply of the tRNA that services both AAA and AAG Lys codons. For tRNA^{Lys19-CUU}, wobble rules only allow recognition to Lys AAG codons. Since we did not detect stalling at AAG codons upon expression of MazF-mt9, tRNA^{Lys19-CUU} must be able to adequately supply Lys amino acids to AAG codons in growing peptide chains.

Paradoxically, although TA toxins are unusually abundant in *M. tuberculosis*, they have been proposed to universally act by simply inhibiting protein translation (K. Winther et al., 2016). However, our data does not support this model. MazF-mt9 did not inhibit translation in *M. tuberculosis* (Figure 16). In fact, there are hundreds of newly synthesized proteins in our quantitative mass spectrometry dataset (Figure 14). Instead, MazF-mt9 appears to reprogram *M. tuberculosis* physiology and growth by enlisting codon-bias as a means to

surgically remove distinct subsets of mRNA transcripts. This targeted transcript ablation leads to the global reduction in the levels of the proteins they encode, and likely enables facile adaptation to stress. It appears that the growth arrest phenotype associated with MazF-mt9-mediated proteome remodeling might be a consequence of the calibrated increase or reduction of key regulatory proteins or enzymatic pathways that influence growth rate based on their codon content.

Finally, our findings are consistent with codon-biased translation reported by the Dedon laboratory upon hypoxic persistence in *M. bovis* (Chionh et al., 2016). However, in their case the codon-biased translation was attributed to alterations in tRNA modification tracked by mass spectrometry, not tRNA abundance (Chionh et al., 2016). The MazF-mt9 toxin-mediated, codon-driven shift in the transcriptome is predicted to alter the spectrum of proteins in *M. tuberculosis* cells and sabotage the efficacy of the immunological and therapeutic assaults intended to eradicate this deadly pathogen. Since there are many predicted tRNA-cleaving toxins in *M. tuberculosis*, the activity of MazF-mt9 might be the first of several examples that enlist this approach to selectively reprogram cell physiology. However, the mechanics of the putative orchestrated enlistment of multiple tRNA-cleaving toxins are unclear since physiological triggers have not yet been identified for any of these toxins in *M. tuberculosis*.

Chapter 2: The sole *Mycobacterium smegmatis* MazF toxin targets tRNA^{Lys} to impart highly selective, codon-dependent proteome reprogramming

Introduction

Stress is a constant threat to the survival of free-living organisms. Toxin-antitoxin (TA) systems are believed to act as one line of defense enlisted by free-living bacteria to survive of the constant barrage of environmental assaults in their native habitats (reviewed in (Harms et al., 2018; Yoshihiro Yamaguchi, Park, & Inouye, 2011)). TA systems are operons comprising adjacent genes encoding two small (~10 kDa) proteins, an antitoxin and its cognate toxin. Based on current models developed through studies of *Escherichia coli* TA systems (reviewed in (Harms et al., 2018; Yoshihiro Yamaguchi et al., 2011)), in the absence of stress the intrinsic activity of the toxin is sequestered by the formation of a stable TA protein-protein complex. However, in response to one or more specific stress triggers, the antitoxin is degraded by proteases. The resulting paucity of antitoxin results in an excess of free toxin which then acts on its specific intracellular target. Toxin action on its target(s) generally results in growth arrest which is characteristically reversible when the stress is released, enabling replenishment of antitoxin. Thus, TA systems appear to work well for pulses of stress instead of prolonged stress where a major physiological and structural transformation is warranted, i.e. conversion from vegetative state to a nearly dormant spore.

Mycobacterium smegmatis is most recognized as a laboratory surrogate for the study of features it shares with its pathogenic relative, *M. tuberculosis*. However, in nature *M. smegmatis* is a saprophyte that lives in ever changing environments within soil, water and on plants. Thus, it should be genetically hardwired for stress survival even though its genome harbors only a small fraction of TA systems compared to *M. tuberculosis* (three in *M. smegmatis* (Frampton et al., 2012) vs. ~90 in *M. tuberculosis* (Ramage et al., 2009; Sala et al., 2014). In this study, we determine the intracellular target and study the function of one of the three distinct TA system family toxins in *M. smegmatis*, the sole MazF family member that we designate as MazF-ms (MSMEG_4448). We found that the primary RNA targeted for cleavage by this endoribonuclease toxin was tRNA^{LysUUU}, identical to one of the 11 MazF toxins in *M. tuberculosis*, MazF-mt9 (Barth et al., 2019; Schifano et al., 2016). We then documented a spectrum of downstream events that lead to surgical remodeling of the proteome, which closely paralleled that which occur in *M. tuberculosis* (Barth et al., 2019). This striking conservation of tRNA targets underscores the importance of the relatively rare Lys AAA mRNA codon as an efficient conduit for modulating the physiology in both mycobacterial species through activation of MazF toxins.

Materials and Methods

Strains, plasmids and reagents

All experiments were performed using *Mycobacterium smegmatis* strain mc² 155 (ATCC 700084). *M. smegmatis* cells were grown at 37 °C in Difco Middlebrook 7H9 media (BD) supplemented with 5 g/L albumin, 2 g/L dextrose, 0.085 g/L NaCl, 0.05 % Tween 80 and 25 µg/mL kanamycin (for plasmid selection), under constant shaking at 200 rpm.

The gene MSMEG_4448, here referred as MazF-ms, was amplified by PCR from *M. smegmatis* genomic DNA using the oligos NWO2791 (5'-AGA TAC ATA TGC GGC GCG GCG ATA TCT ACA CCG CGG-3') and NWO2792 (5'-AGA TAA AGC TTC ACC CGG CGA TTC CCA GAA AAA CC-3'). The amplified DNA was cloned into the anhydrotetracycline (ATC)-inducible plasmid pMC1s (Ehrt et al., 2005) modified to substitute unique NdeI-HindIII sites in place of ClaI-EcoRI to enable insertion of a gene with 5'NdeI-3'HindIII sites. MazF-ms expression was induced by adding ATC to the media at a final concentration of 200 ng/mL when cells reached an OD (600 nm) between 0.1 and 0.2 and compared to uninduced (-ATC) samples.

RNA isolation

In order to extract total RNA, ~50 mL of *M. smegmatis* cells were collected by centrifugation at 2000 g at 4 °C for 5 min. Cell pellets were resuspended in Tri reagent (Zymo Research) and transferred to 2 ml lysing kit tubes (Bertin Corp.) containing 0.1 mm glass beads. Cells lysis was performed on a Precellys Evolution homogenizer (Bertin Corp.) by three consecutive 30-second pulses at 9000 rpm, with 1 min cooling periods on ice in between each cycle. The samples

were centrifuged for 5 minutes at 14000 rpm at 4 °C, and RNA was isolated from the supernatant using Direct-zol RNA Miniprep Plus extraction kit (Zymo Research). After isolation, the samples were treated with 1 U of Turbo DNase as an extra genomic DNA removal step, purified using RNA Clean and Concentration kit (Zymo Research) and eluted in 40 µL of RNase-free water. RNA concentration was measured in a BioSpectrometer (Eppendorf) with a µCuvette.

5' RNA-seq

5' OH libraries were constructed as previously described (Schifano et al., 2014). Briefly, in order to remove 5' monophosphate RNA species, three µg of the purified RNA from induced and uninduced cultures were digested with 1 U Terminator at 30 °C for 1 h. After purification using RNA Clean and Concentration kit (Zymo Research), the samples were phosphorylated using 3 U of T4 PNK at 37 °C for 1 h and re-purified with the same kit. 5' adapter (5'-GUUCAGAGUUCUACAGUCCGACGAUCNNNNNN-3') was ligated using 1 U of T4 RNA ligase 1 (New England Biolabs) at 16 °C for approximately 18 h. In order to remove the remaining free adapters, the adapter-ligated RNAs were resolved on a 6% TBE-Urea PAGE gel, excised and precipitated in isopropanol at -20 °C. The purified RNAs were used in a reverse transcription reaction using Superscript IV (Thermo Fisher) and the degenerate primer (5'-GCCTTGGCACCCGAGAATTCCANNNNNNNNN-3'). The resulting cDNA was

loaded into a 10% TBE-Urea gel and fragments between 80 and 500 nts were excised and precipitated. The cDNA libraries were amplified in a PCR reaction with Phusion HF DNA Polymerase (Thermo Fisher). The primers used were RP1 (5'-AATGATACGGCGACCACCGAGATCTACACGTTTCAGAGTTCTACAGTCCGA - 3') and RPIX (5'-CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCCTTGGCA CCGAGAATTCCA-3'), where the N's represent the individual Illumina barcodes for each library. The amplified libraries between 150-450 bp were gel purified and subjected to single-end sequencing in an Illumina HiSeq 2500 or HiSeq 4000 sequencer.

Note that our 5' RNA-seq method specifically selects for RNA molecules with 5' hydroxyl ends created upon cleavage by MazF-ms. Based on the detailed schematic of our method in Schifano et al. (Schifano et al., 2016), if reverse transcriptase should pause at tRNA modifications, the truncated cDNA would not contain the complement to the 5' adapter sequence that was exclusively ligated to RNAs containing a 5'-OH. Without the adapter sequence, truncated cDNAs would not be amplified by PCR nor could they be sequenced by the Illumina primer which is also complementary to the 5' adapter.

The resulting FASTQ files had the adapter sequences and the first 6 nucleotides of the 5' end trimmed using Trimmomatic (Bolger et al., 2014). Reads were trimmed to 20 nts and the ones containing fewer than 20 nucleotides were excluded. The remaining reads were mapped to *M. smegmatis* genome (NCBI

accession: CP000480.1) using bowtie 1.2 applying the parameters $-n\ 0\ -l\ 20$ (Langmead et al., 2009). Next, we calculated the number of reads that started at a given genome position for each nucleotide in the genome. Genomic positions with 0 counts received a pseudo count of 1 in the uninduced sample. The counts were normalized by sequencing depth, in reads per million (rpm) of mapped reads and the counts of the induced sample were divided by the uninduced control to generate a fold change. Unless otherwise stated, we only considered positions with at least 50 rpm and 5 rpm in the induced sample for tRNA and mRNA genes, respectively, and a fold change >10 . Sequence and frequency logos were generated by kpLogo (Wu & Bartel, 2017). The fastq files were submitted to NCBI's Sequence Read Archive (SRA), under BioProject number PRJNA564437.

Labeling of newly synthesized proteins and proteomics analysis

To identify and quantify which proteins are translated after induction of MazF-ms (MSMEG_4448), three biological replicates were grown to an OD600 between 0.1-0.2 and divided into induced (+ATC) and uninduced samples. In order to label newly synthesized proteins, the methionine analog azidohomoalanine (AHA, AnaSpec) was added to the media at 50 μM after 4.5 hours of MazF-ms induction. After 1.5 h of incubation with AHA, the cells were pelleted by centrifugation at 2000 g at 4 $^{\circ}\text{C}$ for 10 min and washed with PBS. The cells were resuspended in a 2% CHAPS, 8M Urea buffer and lysed as described

in the “RNA isolation” section, using Precellys Evolution homogenizer (Bertin Corp). The cell lysate was centrifuged and the AHA-containing proteins in the supernatant were captured using Click-iT™ Protein Enrichment Kit (ThermoFisher) followed by in-column trypsin digestion.

Digests were analyzed in two separate runs and combined. The data was analyzed as described previously (Barth et al., 2019), considering proteins with ≥ 15 detected spectral counts. Q-values are calculated using the fdrtool package of Strimmer (Strimmer, 2008) with significant changes at or below a q-value of 0.05. The raw files were deposited in Mass Spectrometry Interactive Virtual Environment (MassIVE) repository (accession number: MSV000084300).

Results

MazF-ms expression arrests growth in *M. smegmatis*

According to the Toxin-Antitoxin Database (TADB 2.0) (Xie et al., 2018), *M. smegmatis* reference genome holds only one gene from the MazF toxin family, annotated as MSMEG_4448 (here referred to as MazF-ms). To establish whether or not MazF-ms is toxic (i.e. leads to cell growth arrest) when expressed in *M. smegmatis*, we cloned the gene under the control of an ATC-inducible promoter in the pMC1s plasmid and transformed *M. smegmatis* mc² 155 cells. The expression level from this plasmid is modest, a 4.5-fold induction, based on measurement of mCherry fluorescence (Figure 19). MazF-ms expression led to a pronounced growth arrest that started between 4.5 to 6 h (Figure 20) and was sustained for at least 18 h.

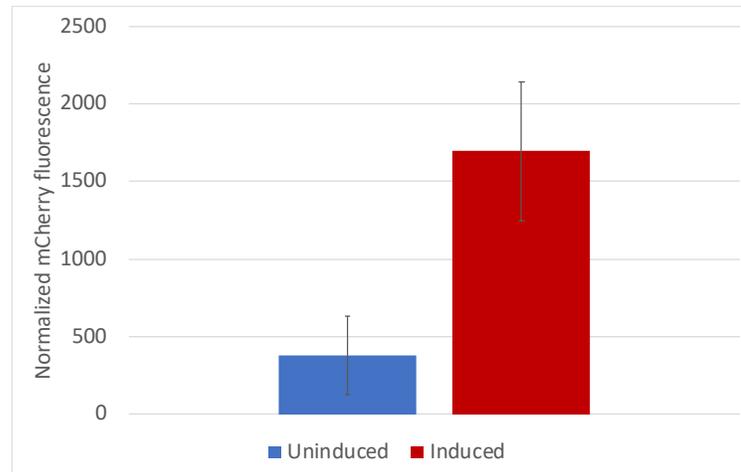


Figure 19. Expression levels of pMC1s vector in *M. smegmatis*. pMC1s vector expression was estimated by cloning a fluorescent reporter gene (mCherry) and inducing its expression with 200 ng/ml of anhydrotetracycline for approximately 6h. mCherry fluorescence was measured (Excitation: 585 nm, Emission: 610 nm) in a Synergy HT 96-well plate spectrophotometer. Raw fluorescence reads were normalized by OD and by auto-fluorescence of a culture that did not contain mCherry.

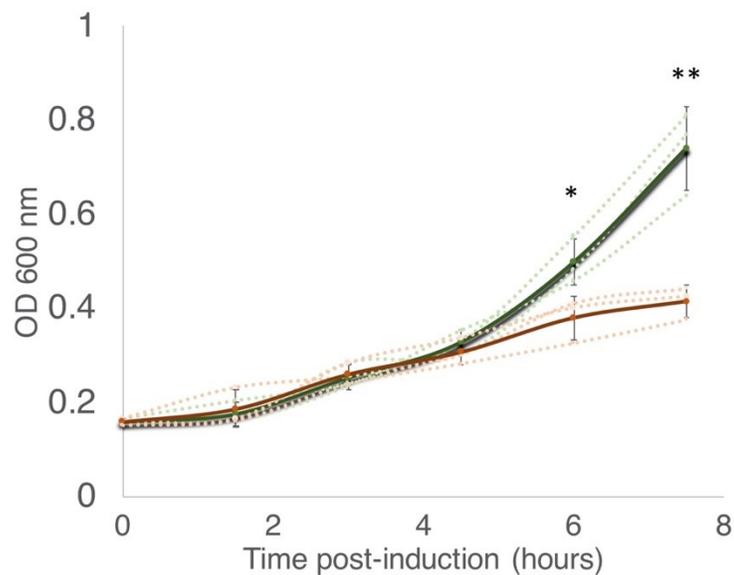


Figure 20. MazF-ms expression in *M. smegmatis* leads to growth

inhibition. *M. smegmatis* cells harboring pMC1s-MazF-ms were grown in triplicate in supplemented 7H9 media until $OD_{600nm} \approx 0.16$. The cultures were split into induced (+ATC, orange lines) and uninduced (-ATC, green lines) and absorbance at 600 nm was determined every 1.5 hours. Error bars represent the standard deviation from the average (solid line) of the three biological replicates (dotted lines). Asterisks represent statistical significance between induced and uninduced in a Student's t test comparison (*, p-value = 0.011; **, p-value = 0.0046).

MazF-ms exclusively targets both tRNA^{Lys} isoacceptors for cleavage at a single site within their anticodons

All members of the MazF family reported to date are single-strand, sequence-specific endoribonucleases. To help elucidate the molecular mechanism by which the ribonuclease activity of MazF-ms is able to regulate growth, we applied our specialized RNA seq method, 5' RNA seq, to find its RNA target(s) (Schifano et al., 2014). This technique was originally developed to selectively sequence transcripts based on their 5' ends. Here, we apply 5' RNA seq to identify RNA fragments containing a 5' hydroxyl (5' OH) end, which are products of MazF toxin activity (Schifano et al., 2014). This approach also allows us to precisely map the cleavage position at a single nucleotide resolution.

Given that the difference in growth between induced and uninduced cultures is more dramatically observed after 6 h post induction, we selected two time points for RNA isolation: one immediately before we observed growth separation (at 4.5 h) and one where the separation is significant (at 6 h). In both time points, 5' RNA seq identified internal cleavage of the only two tRNA^{Lys} isoacceptors (tRNA^{Lys23-UUU} and tRNA^{Lys18-CUU} (Lowe & Chan, 2016) annotated in the *M. smegmatis* genome (Figure 21A-D). When compared to controls, MazF-ms-induced datasets showed an enrichment of 41 to 163-fold in intragenic 5' OH ends at position 36 of both tRNA^{Lys} genes. There are 46 standard tRNAs *M. smegmatis* and one selenocysteine tRNA, none of the other 45 tRNAs were cleaved by MazF-ms (Figure 21E).

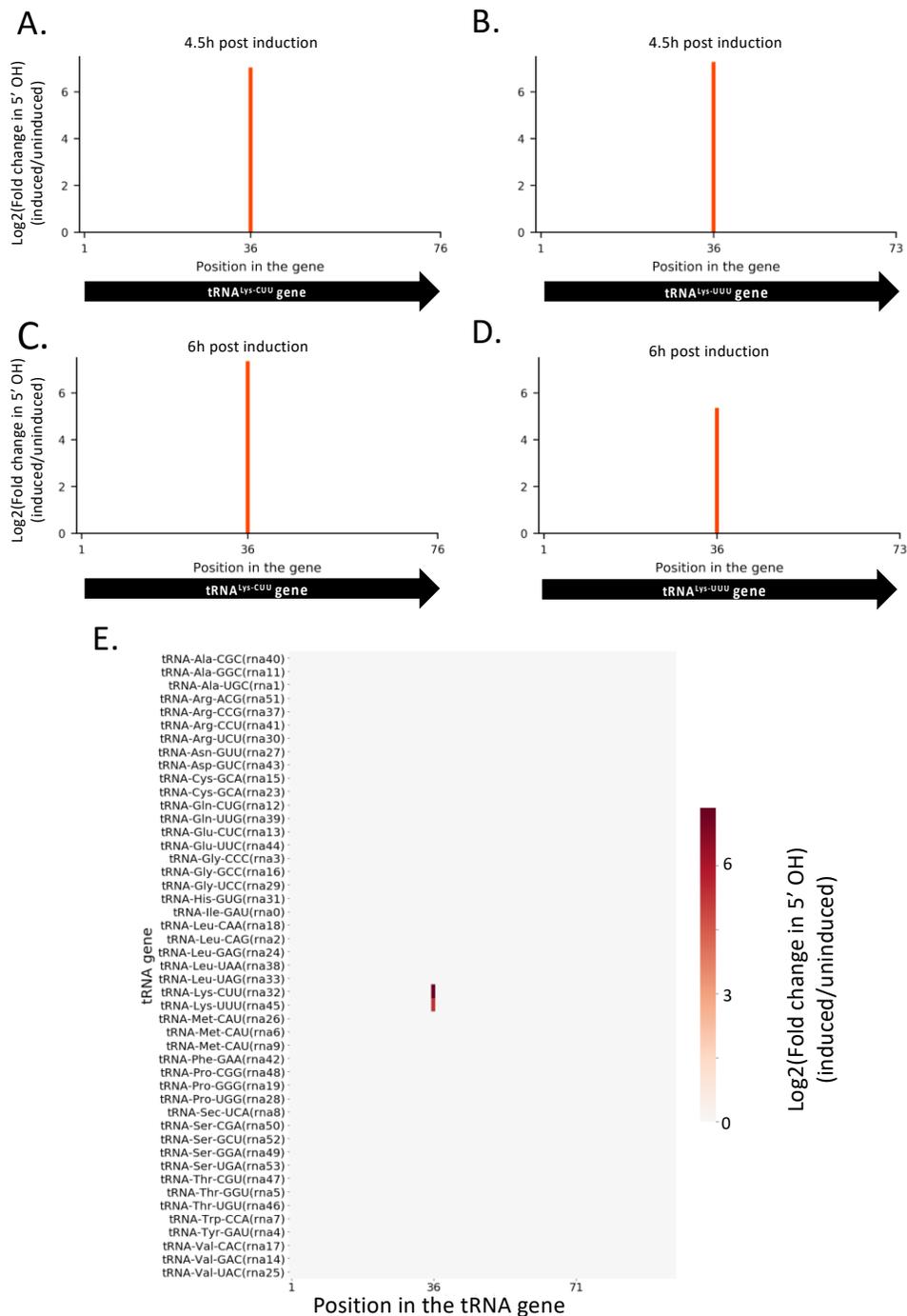


Figure 21. MazF-ms selectively targets both tRNA^{Lys} isoacceptors. (A-D)

Fold changes of 5'OH (indicating endonucleolytic cleavage) within both tRNA^{Lys}-

^{CUU} and tRNA^{Lys-UUU} genes detected by 5' RNA-seq. *M. smegmatis* cells expressing MazF-ms for 4.5 (A, B) or 6 hours (C, D) were compared to uninduced controls. E) Heatmap showing the fold change in 5' OH levels in each position of all 47 *M. smegmatis* tRNA genes after 6 h of MazF-ms induction. The annotated tRNA gene ID (from genome CP000480.1) is show in parenthesis.

Other tRNA-cleaving toxins, such as MazF-mt9 and VapC-mt11 (Cintrón et al., 2019; Schifano et al., 2016), rely on both sequence and secondary structure to accurately recognize their targets. Accordingly, the two tRNA^{Lys} identified here as MazF-ms targets only differ by one nucleotide in the ± 5 nt region surrounding the cleavage site (Figure 22A). This site is located in the anticodon stem loop, where the predicted secondary structure is highly conserved. More specifically, cleavage occurred between the second and third bases of the anticodons (UU/U and CU/U, Figure 22B), presumably inactivating these tRNAs.

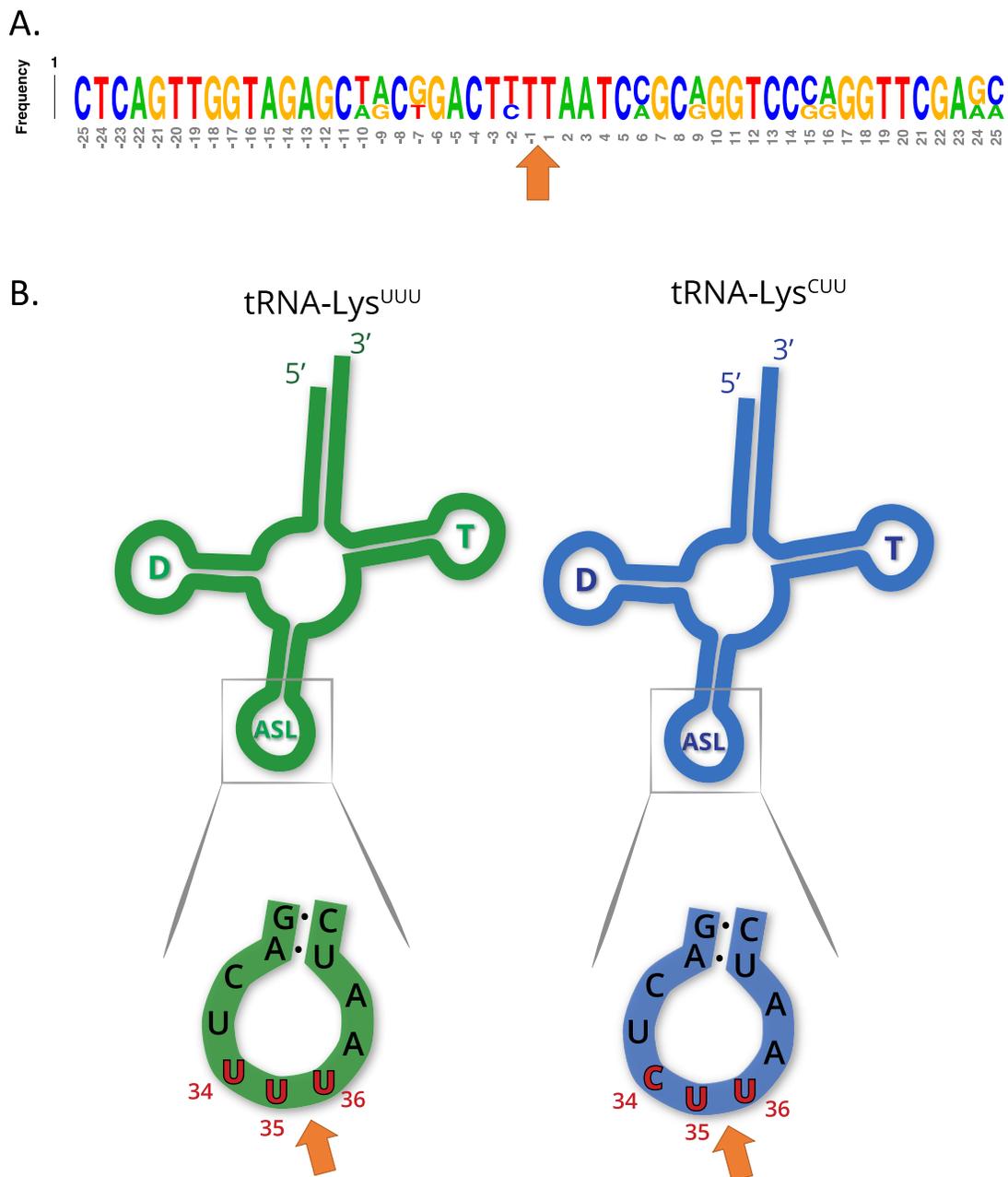


Figure 22. MazF-ms targets show high secondary structure and sequence similarity. A) Frequency logo showing the DNA sequence similarity between tRNA^{Lys-UUU} and tRNA^{Lys-CUU} genes in the 50 nucleotides surrounding MazF-ms cleavage position (orange arrow). B) Schematic representation of tRNA^{Lys-UUU}

and tRNA^{Lys-CUU}, illustrating the D-arm (D), T-arm (T) and anticodon stem loop (ASL) portions of the tRNA. ASL is partially shown in greater detail to emphasize the sequence and secondary structure near the cleavage site (orange arrow). Numbering in red indicates the nucleotide position of the anticodon in the mature tRNA molecule.

5' RNA-seq of MazF-ms-expressing cells reveals ribosome stalling

Having established that the primary targets were tRNA^{Lys-UUU} and tRNA^{Lys-CUU}, we questioned whether the depletion of these tRNAs would lead to ribosome stalling in *M. smegmatis* at the AAA and AAG mRNA codons requiring these tRNAs as we had previously observed for *M. tuberculosis* (Barth et al., 2019). Indeed, we identified 130 cleaved mRNAs in our 5' RNA seq dataset that were not similar in sequence to the tRNA targets at the cleavage site and are not expected to have the secondary structure requirements for MazF-ms recognition demonstrated by Schifano et al. (Schifano et al., 2016). When aligned by their 5' ends, these transcripts showed a clear AAA or AAG consensus sequence approximately 15 nt downstream of the 5' OH end (Figure 23A,B), the cognate Lys codons for tRNA^{Lys-UUU} and tRNA^{Lys-CUU}, respectively. As we had recently proven by Ribo-seq, the 15-nt spacing from the codon to the 5' OH end indicates that a stalled ribosome was bound at this position of the mRNA in vivo (Barth et al., 2019). The 15 nts represents the approximate distance from the 5' side of the translating ribosome to the A-site (Figure 23D) (Barth et al., 2019). Thus,

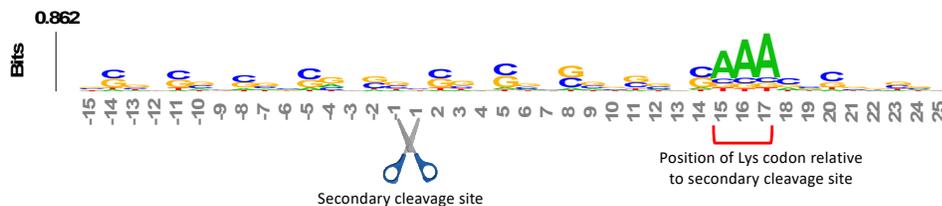
evidence of the stalled ribosome is fortuitously revealed by 5' RNA-seq because upon stalling, the mRNA is then cleaved on its 5' side of the ribosome by one or more cellular RNase(s)—not MazF-ms—that generates a 5' OH upon cleavage.

A.

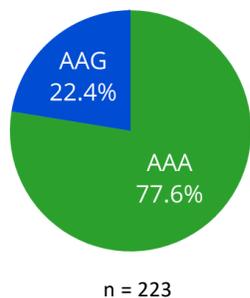
Genome Position	Strand	Counts in induced (rpm)	Counts in uninduced (rpm)	Fold Change	Sequence
417556	+	32.795	0.081	404	atcgtggcgcaactggccaagcagcTcgcggaagaagaagAAAtccaccggc
2106779	-	90.932	0.242	375	ctggaccgtccgatcgtgctgttcaCgacgaccggcgccAAAaccggcaag
3000443	+	19.379	0.081	239	ctcggcgccgaggggtcttcgctcgGgtcgggcacatctcAAAaccggcaac
1024664	+	36.77	0.161	228	catcgcaagctgcttcggcggaactGcgtgcagcgcagaAAAaccggtaag
4386732	+	17.267	0.081	213	tatggtgctgcaaacacctcccagaGtgttgtggacaccAAAaccggccag
772224	-	32.174	0.161	199	acagccgcgaaacctcgagcagatCgtcgacctggcgcgAAAaccatcagc
3071873	+	15.901	0.081	196	atcacgccgagcagcagcaaggaatCcgggatcgagcgcAAAaccaggga
6533313	-	31.18	0.161	193	aagcagctcagcacctacaacgatTgcatctgacgctgAAAaccgtccac
4730773	-	14.783	0.081	182	ccccacgacaccgcccgtctgcaatCcttgcctgctgggtAAAaccacttgg
1498767	+	43.975	0.242	181	gtcaccagctcgtgaacaaggctcTgctcgagggcaagAAAaccgtggc
2014040	+	13.168	0.081	162	tccgcgcttcgacgaggtcaaggccTcggtcgacgagtgcAAAaccaggga
1461954	+	12.795	0.081	157	gcatcgcaagctgcttcggcgaactCgcgtgcagcgcagAAAaccgtgtaa
1024663	+	71.304	0.483	147	gacggatccggcgcccagccggaggAcggctctggggatAAAaccgacgcc
799223	+	22.733	0.161	141	gacgggttcacctcaacgatctggTctcgtacaacgagAAAaccaacgag
3264531	-	54.41	0.403	135	gacaccgctcgtggaactcggcaccgAgaccttcccggtcAAAaccgctgtg



B.



C.



D.

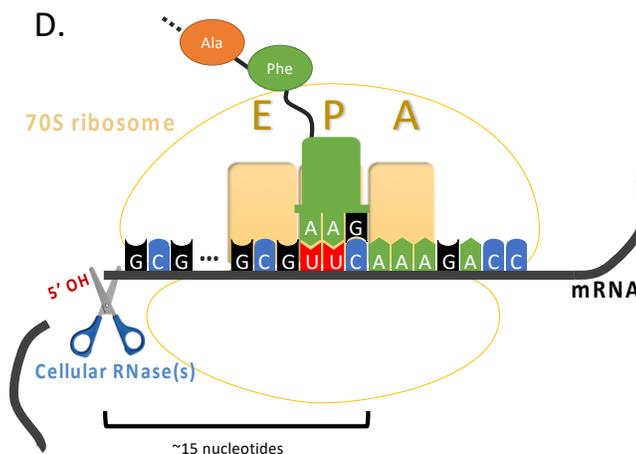


Figure 23. 5' RNA-seq serendipitously reveals ribosome stalling at lysine codons. A) Top mRNA hits found in the 5'RNA-seq dataset. The 50 nucleotides

surrounding the secondary cleavage site (dotted line) generated by cellular RNases (scissor) are shown. Lysine AAA codons are indicated in green, approximately 15 nucleotides downstream of the cleavage site. The first nucleotide (adjacent to the 5' OH) of the read is highlighted in orange. Counts are normalized to reads per million (rpm). B) Sequence logo summarizing the 130 detected mRNA hits. The height of each nucleotide is proportional to its frequency at that given position. Positions are numbered relative to the secondary cut site by cellular RNases (scissor). Red bracket indicates the position of lysine AAA codons. C) Proportion of stalling in AAA codons vs. AAG codons found by 5' RNA-seq in 223 annotated transcripts with at least 1 rpm in the induced sample and containing a lysine codon at position +15. D) Schematic representation demonstrating the events following the depletion of the cellular levels of tRNA^{Lys} by MazF-ms. Due to the lack of available tRNA^{Lys}, translating ribosomes stall mainly at lysine AAA codons at the A site. Ribosome stalling events likely trigger mRNA cleavage at 5' side of the stalled ribosome by cellular RNases (scissor).

We have described the same cascade of events for just one of the 11 MazF family members in *M. tuberculosis*, MazF-mt9, in which tRNA^{Lys43-UUU} depletion leads to ribosome stalling and cleavage on the 5' end of the ribosome. The observation of an analogous trend in the MazF-ms 5' RNA-seq datasets indicating ribosome stalling and subsequent cleavage strongly suggests mechanistic conservation between MazF-mt9 and MazF-ms toxins regarding

initial toxin-mediated tRNA cleavage followed by a secondary ribosome stalling/mRNA cleavage event. In *M. smegmatis*, however, although tRNA cleavage was significant for both isoacceptors, the vast majority of the observed stalled ribosomes (75%) paused at the rarer Lys AAA codon rather than the more frequent Lys AAG (Figure 23C). Therefore, as in *M. tuberculosis*, our data support model in which MazF-ms acts by depleting the cellular pool of tRNA^{Lys} causing ribosome stalling at Lys codons (predominantly Lys AAA), followed by recruitment of another RNase that cleaves 5' of the stalled ribosome (Figure 23D).

MazF-ms promotes codon-specific translation

Next, we sought to characterize the proteomic changes promoted by the Lys AAA/AAG-specific ribosome stalling events mediated by MazF-ms. In order to distinguish proteins that were only synthesized after MazF-ms induction from preexisting “old” proteins, we adopted a method that utilize a methionine mimetic amino acid, called azidohomoalanine (AHA). AHA is incorporated into nascent peptides during translation, therefore only marking proteins that were synthesized after its addition to the media. Due to its azide moiety, proteins containing AHA residues can be captured using an alkyne-containing resin through a Cu(I)-catalyzed click reaction and then analyzed by quantitative mass spectrometry. We added AHA to the cultures after 4.5 h of MazF-ms induction, tagging only proteins that were newly translated.

In contrast with a translation shut-off model proposed for other tRNA-cleaving toxins (K. Winther et al., 2016), global translation was not halted. One hundred twenty six proteins were significantly more abundant in cells expressing MazF-ms compared to the controls (q-value < 0.05). Interestingly, a striking difference in AAA codon content was detected in these proteins compared to the less abundant ones (Figure 24A), i.e. upregulated proteins were generally Lys AAA codon deficient while downregulated proteins were generally Lys AAA codon rich. This trend was not as striking when the proteomics data was instead graphed based on AAG codon content (Figure 24B). These trends observed in newly synthesized proteins were concordant with the relative abundance of ribosome stalling events detected in our 5' RNA seq datasets at Lys AAA codons (>75% of transcripts with evidence of ribosome stalling). Finally, the ability of MazF-ms to preferentially influence the overall cellular pool of Lys AAA codon-containing proteins over those containing Lys AAG codons is also graphed in Figure 25.

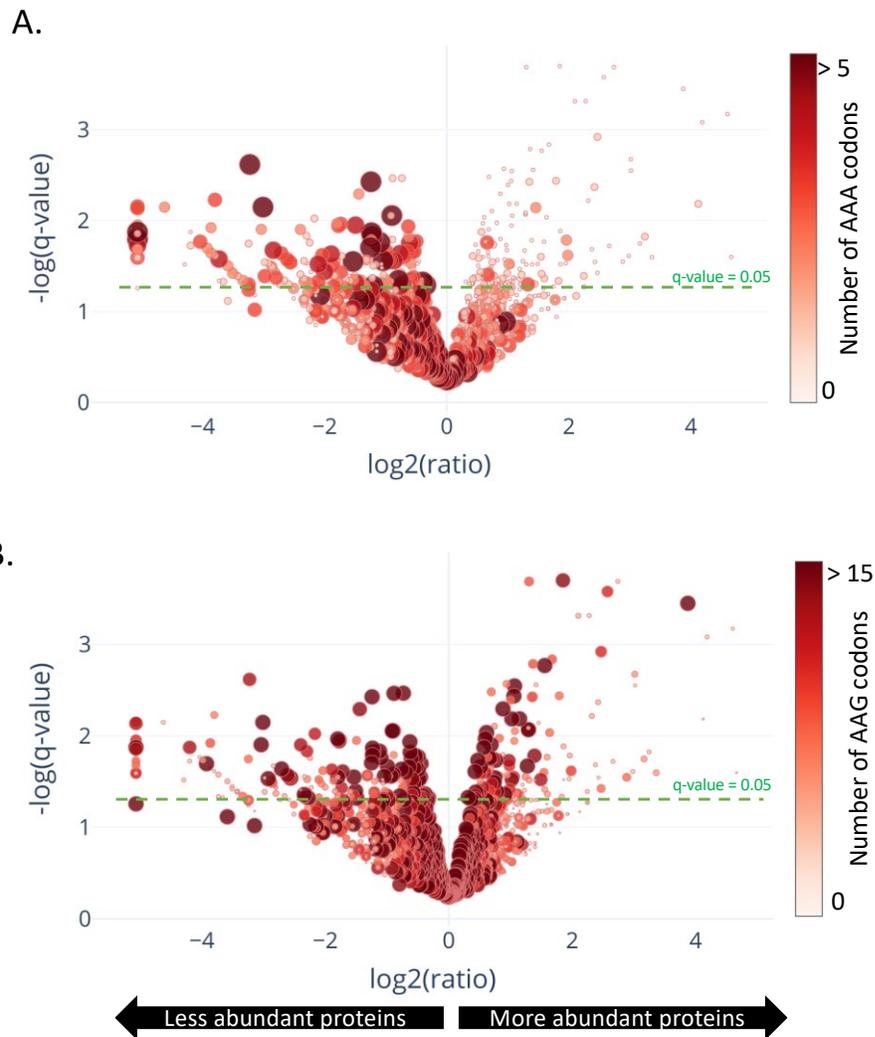


Figure 24. Translation of proteins rich in lysine AAA codons is reduced upon MazF-ms expression. A-B) Volcano plot showing differentially translated proteins (circles) detected by quantitative mass spectrometry and their AAA (A) or AAG (B) codon content. *M. smegmatis* cultures expressing MazF-ms for 4.5 h were incubated with azidohomoalanine (AHA) to label only newly synthesized proteins after toxin expression. The color saturation and circle size are proportional to the number of AAA (A) or AAG (B) codons in the gene.

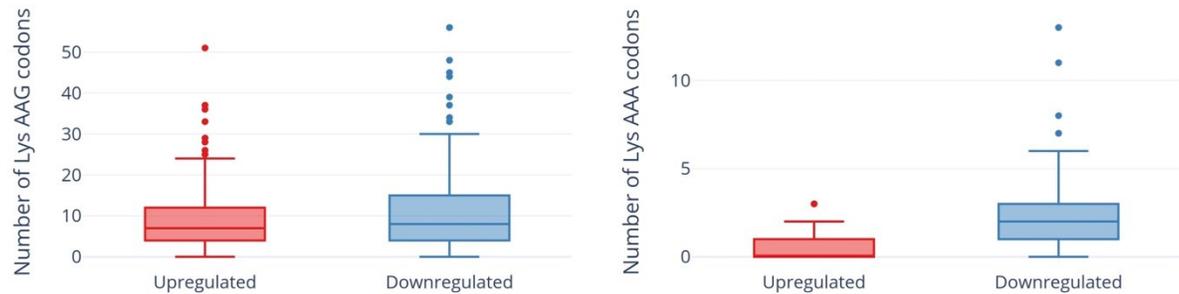


Figure 25. Expression of MazF-ms leads to global proteomic shifts based on the Lys AAA, but not Lys AAG, codon content. Distribution of AAG (left) or AAA (right) codons in significantly up- or downregulated proteins. Outliers are shown as individual dots.

MazF-ms reduces translation of critical components of the DNA replication machinery while concomitantly supporting synthesis of stress response proteins

In TA systems, specific stresses trigger degradation of the cognate antitoxin by a protease, freeing toxin to act within the cell (Harms et al., 2018; Song & Wood, 2018; Yoshihiro Yamaguchi et al., 2011). Therefore, overexpression of MazF-ms (used in this analysis and throughout this work) is intended to mimic natural toxin activation that occurs when cells are exposed to the relevant stress. After observing hundreds of differentially translated proteins in Figure 24, we subjected the two datasets comprising the more abundant or

less abundant proteins following MazF-ms toxin expression to the Function Annotation Tool in the Database for Annotation, Visualization and Integrated Discovery (DAVID) platform (D. W. Huang et al., 2009; Huang da et al., 2009). This tool within DAVID was used to identify was used to identify functionally similar proteins which were enriched in the two datasets (Figure 26). Since toxin overexpression simulates toxin activation by exposure to stress, the proteins identified using DAVID are expected to be physiologically relevant.

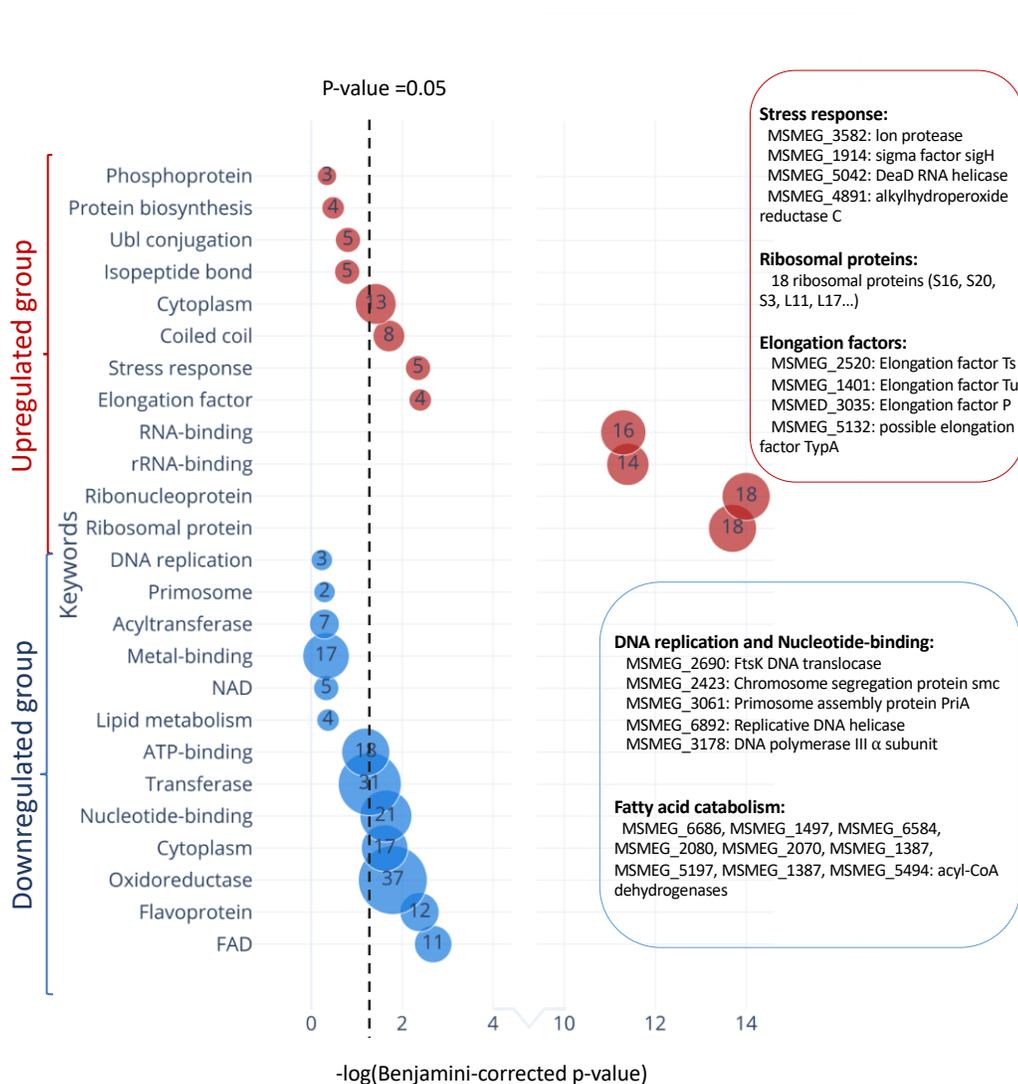


Figure 26. Functional classification of differentially translated proteins during MazF-ms expression. Enriched DAVID UP_KEYWORDS (UniProt keywords) terms were generated by analyzing the up- (red) or down-regulated (blue) proteins identified in AHA-proteomics. The diameter of each circle and its numbering correspond to the number of genes associated with the given keyword. Examples of genes in the main groups are described by relevant categories in the corresponding color-matched boxes.

Among the 171 statistically significant downregulated proteins from Figure 24A (in blue in Figure 26), there were some notable trends consistent with cells in a state of growth arrest. First, there were 11 downregulated enzymes involved in “fatty acid catabolism” (FAD heading; blue box Figure 26). Fatty acid catabolism occurs predominantly through successive rounds of β -oxidation, a process whereby even-chain fatty acids are degraded to acetyl-CoA and odd-chain fatty acids are degraded to acetyl-CoA and propionyl-CoA. Second, there were five functionally related proteins within the “DNA replication” and “nucleotide binding” categories (blue box Figure 26). In the DNA replication category, one critical protein was downregulated, the only catalytic subunit of the three subunit core DNA polymerase III enzyme (α subunit, MSMEG_3178). This core enzyme is a component of the DNA polymerase III holoenzyme that mediates DNA replication in bacteria. Within the nucleotide binding group were a spectrum of proteins whose downregulation is also logical for cells in a state of growth arrest: RecA (MSMEG_2723) is involved in DNA repair, the DNA translocase FtsK (MSMEG_2690) is localized at the septum where cell division occurs, chromosome segregation protein SMC (MSMEG_2423) and the priA (MSMEG_3061) are components of the primosome protein complex that activates DNA replication forks.

There were 126 upregulated proteins upon MazF-ms expression. When this dataset was subjected to the Functional Annotation Tool in DAVID, there were several proteins whose upregulation were also consistent with the growth

arrested state of MazF-ms expressing cells. Four proteins were in the “stress response” category: Lon protease (MSMEG_3582), an ATP-dependent RNA helicase DEAD/DEAH box family protein (MSMEG_5042), sigma factor SigH (MSMEG_1914) and alkylhydroperoxide reductase (MSMEG_4891). Lon is a stress-responsive protease. Since it is known to degrade all TA system antitoxins in *Escherichia coli* (Gerdes and Maisonneuve, 2012), it may also have an analogous role and cleave the MazE-ms antitoxin in *M. smegmatis*. Uniprot places the MSMEG_5042 RNA helicase as functioning in ribosome biogenesis, mRNA degradation and translation initiation. The alternate sigma factor SigH is activated by oxidative, heat and nitric oxide stress (Sharp et al., 2016) while Uniprot places alkylhydroperoxide reductase MSMEG_4891 in protection from oxidative damage by detoxifying peroxides.

Consistent with the sustained protein synthesis while MazF-ms was being expressed, we observed new synthesis of a subset of elongation factors and ribosomal proteins. Four elongation factors were upregulated: EF-Ts (MSMEG_2520), EF-Tu (MSMEG_1401), EF-P (MSMEG_3035) and BipA/TypA ribosome-binding GTPase (MSMEG_5132). Many of these elongation factors are associated with the bacterial stress response. Finally, 18 of the 51 ribosomal proteins were also upregulated. The other 33 ribosomal protein levels were stable.

Discussion

Bacterial genomes are under constant pressure to remain compact while also retaining genes that provide a competitive edge for survival in their natural environments. Acquisition of TA systems in bacterial genomes is thought to represent one potent vehicle for stress protection. In contrast to the ~ 90 TA systems in its pathogenic relative *M. tuberculosis*, the *M. smegmatis* genome harbors just three TA systems (MazEF, PhD-Doc, and VapBC) as one facet of its stress survival armamentarium. A thorough understanding on how a TA system acts to protect its host from stress requires determination of the function of the toxin, and the function of the toxin is informed by determining its intracellular target. Since all MazF toxins are generally single-strand and sequence-specific endoribonucleases, in this work we identified the RNA target of the MazF-ms toxin using 5' RNA-seq which revealed its detailed mechanism of action. To our surprise, MazF-ms did not behave like the vast majority of MazF toxins that appear to predominantly cleave mRNAs (reviewed in Masuda & Inouye, 2017). Instead, MazF-ms behaves almost exactly like the only other known exception, MazF-mt9, one of the 11 MazF family members in *M. tuberculosis* (Barth et al., 2019).

MazF-mt9 is an outlier because it requires both structure and sequence for its highly specific recognition of a single tRNA isoacceptor. This requirement for structure is much like VapC toxins, even though MazF-mt9 and VapC toxins lack sequence or structural similarity (Cintrón et al., 2019; Jonathan W. Cruz et al., 2015; Schifano et al., 2016; Schifano & Woychik, 2017; K. Winther et al., 2016). MazF-ms now represents the second example of a MazF toxin that

targets tRNA for cleavage, thus reducing the levels of only this tRNA species in vivo. Since we were able to unequivocally detect MazF-ms target RNAs with the required 5'-OH using 5' RNA-seq, there was no apparent masking of the precise cleavage site due to the presence of an RNA modification. This surgical depletion of just one tRNA results in ribosome stalling at codons requiring this depleted tRNA and proteome remodeling to favor sustained synthesis of only AAA-deficient proteins (Figure 27).

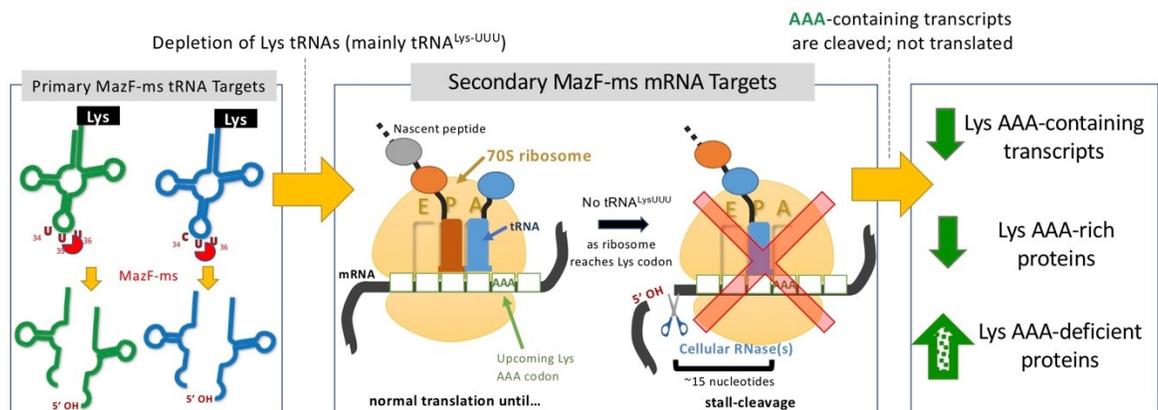


Figure 27. Summary of the proposed MazF-ms mechanism of action. MazF-ms primarily targets tRNA^{Lys-UUU} and tRNA^{Lys-CUU}, depleting their intracellular levels. Inactivation of these tRNAs lead to selective ribosome stalling predominantly in Lys-AAA codons, culminating in cleavage of the transcript. This results in codon-biased proteomic changes, favoring Lys-AAA-depleted transcripts.

Why is Lys tRNA such an important conduit for proteome remodeling, ostensibly during stress, in mycobacteria? It appears that many proteins critical for stress survival in these GC rich genomes (67.4% GC in *M. smegmatis*; 65.6% in *M. tuberculosis*) are deficient in AAA Lys codons, allowing for their sustained translation while the cell can save energy by not synthesizing proteins that do not contribute to this endpoint. Indeed, as discussed above, many of the upregulated proteins annotated in the stress response category have direct roles in one or more stress responses: cold shock, heat shock, oxidative stress and nitric oxide stress. The stress-specific sigma factor SigH was also upregulated. Therefore, it is implicated as the primary RNA polymerase sigma factor enlisted for the sustained transcription of *M. smegmatis* genes whose proteins were upregulated after toxin expression. However, other upregulated proteins in our dataset that are not commonly associated with the stress response, i.e. at least some elongation factors, appear to have indirect roles (reviewed in Starosta et al. (Starosta, Lassak, Jung, & Wilson, 2014)). EF-Tu is maximally expressed during stress in *E. coli* (Muela et al., 2008). Since the guanine nucleotide exchange factor EF-Ts assembles with EF-Tu in a 2:2 stoichiometry (Kawashima, Berthet-Colominas, Wulff, Cusack, & Leberman, 1996), it is expected to be expressed at levels equivalent to EF-Tu. EF-P rescues ribosomes stalled at poly-proline stretches. However, EF-P is only active when lysinylated or hydroxylysinylated (Doerfel et al., 2013; Ude et al., 2013). Thus, the lysine from cleaved Lys tRNAs might be recycled and used to activate EF-P. Finally, BipA/TypA, while not

essential, appears to confer a growth advantage by regulating the synthesis of a subclass of proteins in cells under cold shock, low pH, oxidative stress, antimicrobial peptide stress and detergent stress.

Chapter 3 - Accurate target identification for *Mycobacterium tuberculosis* endoribonuclease toxins requires expression in their native host

Introduction

The bacterial pathogen that causes tuberculosis, *M. tuberculosis*, is able to evade the immune system and persist within its host for extended periods of time as a latent infection. Elucidation of the molecular mechanisms that underlie the latent state of *M. tuberculosis* is essential for developing more effective therapeutics for latent tuberculosis infection.

The *M. tuberculosis* genome harbors ~90 TA systems of which 50 belong to the VapBC family (Ahidjo et al., 2011; Arcus, Rainey, & Turner, 2005; Harms et al., 2018; Ramage et al., 2009; Sala et al., 2014). All VapC toxins contain a conserved catalytic PiT N-terminus (PIN) domain (Arcus, McKenzie, Robson, & Cook, 2011; Arcus et al., 2005); an intact PIN domain is essential for VapC toxicity and nuclease activity (Sharp et al., 2012). However, the stresses that trigger VapC toxin activities during infection are not known. In fact, the precise physiological roles of these 50 VapC toxins in *M. tuberculosis* infection and virulence are not well understood, although they are proposed to reduce protein synthesis (K. Winther et al., 2016). It is also a conundrum why this pathogen harbors so many VapBC TA systems if collectively their only role is to simply reduce translation. To begin to tackle these broad and complex questions, we and others have been systematically studying their enzymatic activity as a first

step toward understanding how they may influence the course of tuberculosis infections and reduce the efficacy of antituberculars.

RNA cleavage by members of the *M. tuberculosis* VapC-family of toxins has been reported using one of two general approaches. The first in vitro approach involves incubation of recombinant toxin with synthetic tRNAs (Jonathan W. Cruz et al., 2015; Deep et al., 2018), synthetic RNAs (Ahidjo et al., 2011; McKenzie, Duyvestyn, Smith, Bendak, MacKay, et al., 2012; McKenzie, Robson, Berney, Smith, Ruthe, et al., 2012; Sharrock, Ruthe, Andrews, Arcus, & Hicks, 2018), or total RNA isolated from *M. tuberculosis* (Jonathan W. Cruz et al., 2015). The second approach involves ectopic expression of the toxin in alternate rapidly growing hosts that do not require biosafety containment (*Escherichia coli* and *M. smegmatis*) (Jonathan W. Cruz et al., 2015; McKenzie, Duyvestyn, Smith, Bendak, MacKay, et al., 2012; McKenzie, Robson, Berney, Smith, Ruthe, et al., 2012; Ramage et al., 2009; Sharrock et al., 2018; K. S. Winther et al., 2013; K. Winther et al., 2016). Both of these approaches appeared to be reliable because only one or a few RNA targets were identified within the total RNA pool. Therefore, they are routinely employed to identify and characterize *M. tuberculosis* VapC toxin targets.

The enzymatic activities of 13 of the 50 *M. tuberculosis* VapC toxins have been studied to varying degrees (Jonathan W. Cruz et al., 2015; Sharp et al., 2012; K. S. Winther et al., 2013; K. Winther et al., 2016). Two of these 13 VapC toxins specifically target 23S rRNA at the sarcin-ricin loop (K. S. Winther et al., 2013; K. Winther et al., 2016). The proposed RNA targets for the other 11 *M.*

tuberculosis VapC toxins studied previously by Winther et al. (K. Winther et al., 2016)—including VapC-mt11 (aka VapC11, Rv1561) that is the focus of our study presented here—are derived from an RNA-VapC interaction screen using an *M. smegmatis* host.

We performed a thorough characterization of the enzymatic properties of the *M. tuberculosis* VapC-mt11 TA toxin in vitro and in vivo as well as its effect on protein synthesis and mycobacterial growth. We demonstrate that the *M. tuberculosis* VapC-mt11 toxin exhibits highly precise target specificity only when expressed in its natural *M. tuberculosis* host, in striking contrast to its much broader spectrum tRNase activity in vitro. These results also differ from those of Winther et al. (K. Winther et al., 2016) and serve as a useful guide for the benefits and limitations of the array of approaches implemented by those studying endoribonuclease TA toxins.

Methods

Strains, Plasmids and Reagents

The *E. coli* strain BL21(DE3) (F- ompT hsdS β (r β -m β) dcm gal (DE3) tonA) (Novagen) was used for all protein expression. *E. coli* K-12 Mach1 T1 cells (Δ recA1398 endA1 tonA Φ 80 Δ lacM15 Δ lacX74 hsdR(rk+mk+); Invitrogen) were used for all cloning experiments. *M. smegmatis* mc² 155 and *M. tuberculosis* H37Rv strains were both used for growth profiles. Experiments involving the virulent H37Rv strains of *M. tuberculosis* were conducted in a BSL-3 laboratory

following institutionally approved protocols. *M. tuberculosis* H37Rv was also used for the plating efficiency assay to determine recovery of viable bacteria upon VapC-mt11 expression. *M. smegmatis* mc² 155 was used for metabolic labeling. The vapC-mt11 (Rv1561; “mt” refers to *M. tuberculosis*) gene was cloned using *M. tuberculosis* H37Rv genomic DNA. The DNA sequences of PCR fragments used for cloning were confirmed by automated DNA sequence analysis. The VapC-mt11 coding region was cloned into the pET28a expression vector (EMD Millipore) and the arabinose inducible pBAD33 plasmid (ATCC) after the addition of 5' NdeI and 3' BamHI restriction enzyme sites by PCR. For expression in Mycobacteria, the VapC-mt11 coding region with PCR generated 5' ClaI and 3'Sall restriction sites was cloned into the corresponding sites in the anhydrotetracycline (ATc) inducible vector pMC1s (Ehrt et al., 2005).

Growth Assays in Mycobacteria

M. smegmatis mc² 155 was transformed by electroporation with 0.5 – 1 µg of pMC1s-vapC-mt11 DNA. Cultures were inoculated with single colonies, grown at 37°C in 7H9-TW80-AND medium containing 25 µg/ml kanamycin and induced at an OD600 of 0.4 by addition of 200 ng/ml ATc. Cell pellets were collected from uninduced and induced samples at intervals up to 6 h. *M. tuberculosis* H37Rv was transformed by electroporation with 0.5 µg pMC1s-vapC-mt11 DNA. Cultures were inoculated with single colonies, grown at 37°C in 7H9-TW80-AND medium containing 30% spent medium and 25 µg/ml kanamycin. Upon inoculation, cultures were induced by 200 ng/ml ATc; additional ATc was added every other day to maintain the initial ATc concentration. Three independent

experiments were performed, and error bars used to represent the standard deviation (S.D.). For plating efficiency assays, *M. tuberculosis* H37Rv VapC-mt11 clones were plated on 7H9-TW80-ADN agar plates containing 25 µg/ml kanamycin and with or without 500 ng/ml ATc. Colonies were counted 3-weeks after incubation at 37°C.

Purification of Recombinant VapC-mt11

E. coli BL21(DE3) cells transformed with the pET28a-vapC-mt11 were grown in M9 minimal medium supplemented with 0.1% glycerol and 50 µg/ml kanamycin at 37°C to exponential phase. Induction of the protein was achieved by adding 1 mM isopropyl 1-thio-D-galactopyranoside (IPTG). After 4 h, cells were harvested by centrifugation and resuspended in lysis buffer (50 mM NaH₂PO₄ (pH 8.0), 500 mM NaCl, 20 mM imidazole, 10 mM β-mercaptoethanol, 1 mM PMSF, 1 mg/ml lysozyme). Cell pellets were then lysed by sonication and lysates applied to a nickel-nitrilotriacetic acid (Ni-NTA) resin (Qiagen) to purify the protein as described previously (Sharp et al., 2012).

In vitro Transcription of *M. tuberculosis* tRNAs

M. tuberculosis wild-type and mutant tRNAs were synthesized in vitro using synthetic DNA oligonucleotides containing the T7 RNA polymerase promoter and the 5' end sequence of the tRNA of interest along with a second

oligonucleotide with the 3' end sequence of the tRNA. Oligonucleotides were designed with a region of overlap to serve as a template for PCR with Taq DNA polymerase to create a tDNA. The tDNA was gel extracted from a 2% agarose gel using the QIAquick Gel Extraction kit (Qiagen). 200 ng of the tDNA was used to transcribe the tRNAs of interest using the RiboMAX Large Scale RNA Production System (Promega) as recommended by the manufacturer. The tRNA transcription reactions were separated on a 9% polyacrylamide, 7 M urea gel and visualized by ethidium bromide staining. The product corresponding to the correct size was 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 elutions were ethanol precipitated and resuspended in nuclease-free water.

In vitro tRNA Cleavage Assay

The 45 in vitro transcribed *M. tuberculosis* tRNAs (2 pmol) were initially incubated with or without VapC-mt11 (10 pmol) for 3 hr at 37 °C in 10 mM HEPES pH 7.5, 15 mM KCl, 3 mM MgCl₂, 10% glycerol. Mutants were also incubated with increasing amounts of the toxin (0, 10, 20 and 30 pmol) for 3 h at 37 °C. Samples were analyzed in 7 M Urea, 9% polyacrylamide gel and visualized by SYBR Gold (Invitrogen) staining.

In vitro *M. tuberculosis* tRNA Primer Extension

Two pmol of the in vitro transcribed *M. tuberculosis* wild-type or mutant tRNAs were incubated with 10 pmol VapC-mt11 toxin or without toxin for 1 h at 37 °C. To detect cleavage products, the following oligonucleotides were 5' labeled with [γ -³²P] ATP (Perkin Elmer Life Sciences) using T4 polynucleotide kinase (NEB): tRNA^{Ala31} (5'GGA GCT AAG GGG ACT CGA ACC C 3'), tRNA^{Arg27} (5'GCG CCC GAA GAG ATT CGA ACT C3'), tRNA^{Arg40} (5' CCG GCA GGA TTC GAA CCT GCG 3'), tRNA^{Gln32} (5' CTG GGG TAC CAG GAC TCG A 3'), tRNA^{Gln41} (5' TCC GTC GCC AGG ACT CGA ACC 3'), tRNA^{Leu13} (5' GCG GGC GGA GGG ACT CGA ACC 3'), tRNA^{Leu15} (5'GGG ACT TGA ACC CCC ACG C 3'), tRNA^{Pro14} (5' CGG GCT GAC AGG ATT TGA ACC TGC G 3'), tRNA^{Pro23} (5' CGG GGT GGC GGG ATT TGA AC 3'), tRNA^{Pro35} (5' CGG GGT GAC AGG ATT TGA ACC TG 3'), tRNA^{Ser24} (5' GGA GGA TGC GGG ATT TGA ACC C 3'), tRNA^{Ser26} (5' GAG GCG AGA GGA TTT GAA CCT CC 3'), tRNA^{Ser28} (5' GGT GGC GGA GGG ATT TGA ACC CTC 3'), tRNA^{Thr5} (5' GCC CCC TAA CGG AAT CGA ACC 3'). For tRNA^{Pro14} mutants, the tRNA^{Pro14} (5' CGG GCT GAC AGG ATT TGA ACC TGC G 3') was used. Labeled oligonucleotides were added to the reactions and incubated for 3 min at 95 °C and left to cool down to room temperature. Primer extension with Superscript III or IV (Invitrogen) was performed at >60 °C for 80 min. The DNA sequencing ladder was produced using the Sequenase version 2.0 DNA sequencing kit (USB) according to the manufacturer's instructions. Loading buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol) was added to each sample followed by 15% Urea PAGE and autoradiography.

tRNA Northern Analysis

In brief, total RNA from *M. tuberculosis* H37Rv was obtained from cells grown to exponential phase using TRIzol Reagent (Invitrogen). RNA was treated with TURBO DNase (Invitrogen). 1.5 µg of total RNA was incubated with 98 pmol of recombinant VapC-mt11 toxin for 3 h at 37°C. Reactions were subjected to Urea-PAGE separation on a 9% polyacrylamide, 7 M urea gel and visualized by ethidium bromide staining and transferred to nitrocellulose. The following oligonucleotides were 5' labeled with [γ -³²P] ATP (Perkin Elmer Life Sciences) using T4 polynucleotide kinase (NEB): tRNA^{Ala31} (5' CCA CAC TGC CAG TGT GGT GCG C 3'), tRNA^{Arg27} (5' CCT TCT GAT CCG TAG TCA GAT GC 3'), tRNA^{Arg40} (5' CTT CTG CTC CGG AGG CAG ACG 3'), tRNA^{Gln32} (5' ATG GCT GAA CCA GAA TCA GCT GT 3'), tRNA^{Gln41} (5' CTA TCT GAA CCA AAA TCA GAG GTG C 3'), tRNA^{Leu3} (5'GGA CAC TGG CAC CTG AAG CCA 3'), tRNA^{Leu13} (5' GCA CCG GCA CCT AAA ACC GGC 3'), tRNA^{Leu15} (5' TAG GGC ACT AGC ACC TCA AGC TAG CG 3'), tRNA^{Pro14} (5'GAC CAC TTG ACC CCC AGT CAA G'), tRNA^{Pro23} (5'GGC CTC TTC GTC CCG AAC GAA GC3'), tRNA^{Pro35} (5' GGC CTT CCG CTC CCA AAG CGG AT 3'), tRNA^{Ser24} (5'AGG GCT GTT AAC CCA ACC CGC G 3'), tRNA^{Ser26} (5' CCC TTG AAG GGG GAC AAC TCA TTA 3'), tRNA^{Ser28} (5' CAC ACG CTT TCG AGG CGT GCT CC 3'), tRNA^{Thr5} (5'CCT TTT

CCT TAC CAT GGA AAC G 3'). Hybridization temperatures used precluded cross-reactivity with other tRNAs.

In vitro Translation Inhibition

To test whether the VapC-mt11 strong toxic phenotype was due to translation inhibition, we added recombinant VapC-mt11 to the PURExpress kit (New England Biolabs) at a concentration of 200 pmol. Production of the DHFR protein (~20 kDa) was assessed via [³⁵S] Express Protein Labeling Mix (Perkin Elmer Life Sciences) incorporation. An equal volume of 2X Laemmli buffer (125 mM Tris (pH 6.8), 20% glycerol, 4% SDS, 0.01% bromophenol blue) was added to terminate the translation reaction. Samples were heated to 95°C for 5 min prior to separation by 17.5% SDS-PAGE followed by autoradiography.

Metabolic Labeling in *M. smegmatis*

Transformants were obtained as previously described. Individual colonies were grown at 37°C in 7H9-TW80-AND medium containing 20 µg/ml kanamycin until OD₆₀₀ 0.3-0.5. The culture was then split into equal portions, and 200 ng/ml anhydrotetracycline (ATc) was added to one portion. 1 ml aliquots were removed at 0, 0.5, 1, 2, 4 and 6 h post-induction and incubated with 37.5 µCi of [³⁵S] Express Protein Labeling Mix (Perkin Elmer Life Sciences) at 37°C for 20 min. Cell pellets were collected by centrifugation (13,200 rpm for 5 min) and

resuspended in Laemmli buffer. To ensure normalization, the volume of Laemmli buffer added to each pellet was determined by multiplying the OD600 by 500. To lyse the *M. smegmatis* cells resuspended in Laemmli buffer, a 40% volume of acid washed glass beads (<160 µm; Sigma) were added and the mixture vortexed for 5 min. Samples were heated to 95°C for 5 min prior to 17.5% SDS-PAGE followed by autoradiography.

5' RNA-seq

Total RNA was isolated from *M. tuberculosis* H37Rv harboring the plasmid pMC1s containing vapC-mt11 were grown to an OD600 of 0.4 and split into uninduced and induced cultures (+ 200 ng/ml anhydrotetracycline). Samples were collected 8 h and 24 h post induction and pelleted by centrifugation. Preparation of 5'-dependent libraries and data analysis was performed as described in Schifano et al. (Schifano et al., 2014). Briefly, for sequencing of RNAs containing 5' hydroxyl ends (5'-OH), total RNA samples were digested using 1 U of Terminator 5'-Phosphate-Dependent Exonuclease (Epicentre) to remove RNAs containing 5'-monophosphate (5'-P). After purification using RNA Clean & Concentrator™-5 (Zymo Research), 5'-OH ends were phosphorylated using 3 U of T4 PNK (New England Biolabs). The small RNA 5' adapter (5'-GUUCAGAGUUCUACAGUCCGACGAUCNNNNNN - 3') was ligated to the phosphorylated 5' ends using T4 RNA ligase 1 (New England Biolabs) at 16°C overnight. Ligated RNAs were separated from free adapters on a 6% Urea-PAGE gel and isolated by gel excision. Subsequently, cDNA was generated using Superscript IV (ThermoFisher) with the primer 5'-

GCCTTGGCACCCGAGAATTCCANNNNNNNNN-3', and gel extracted selecting fragments from 80 to 500 nts. The cDNA libraries were subjected to 12 cycles of PCR amplification with Phusion HF DNA polymerase (ThermoFisher), using the oligonucleotides RP1 (5'-
AATGATACGGCGACCACCGAGATCTACACGTTTCAGAGTTCTACAGTCCGA -
3') and RPIX (5'-
CAAGCAGAAGACGGCATACTGAGATNNNNNNGTGACTGGAGTTCCTTGGCA
CCCGAGAATTCCA-3'), where underlined N's represent the library-specific Illumina barcodes. After 10% PAGE, amplified DNA between the sizes 150 bp and 450 bp was isolated by gel excision and sequenced in an Illumina NextSeq 500 platform.

After trimming the adapter sequences from the resulting FASTQ files, the sequences were trimmed to 20 nts (discarding shorter sequences) and aligned to the *M. tuberculosis* H37Rv genome (Genbank accession: AL123456) using Bowtie 1.2.0, not allowing mismatches. For each nucleotide in the genome, we calculated the number of reads that started at that position (i.e. the number of RNA molecules that had their 5' OH end starting at each nucleotide). Read counts were normalized to sequencing depth and expressed as “reads per million of mapped reads” (rpm). The ratio of counts between induced and uninduced was calculated. Positions that had 0 counts in the uninduced library were adjusted to a pseudo-count of 0.5. We only considered reads that had at least 5 rpm for mRNAs and 50 rpm for tRNAs in the induced sample and a ratio of at least 10. Among the three tRNAs that met these criteria—tRNA^{Gln32-CUG},

tRNA^{Leu3-CAG} and tRNA^{Glu8-UUC}—only two were counted as legitimate primary VapC-mt11 targets (tRNA^{Gln32-CUG} and tRNA^{Leu3-CAG}) because they had a GG cleavage consensus sequence immediately before the cleavage site. tRNA numbering based on the Lowe lab genomic tRNA database <http://gtrnadb.ucsc.edu> (Lowe & Chan, 2016). The sequencing datasets generated in this study were deposited in the NCBI Sequence Read Archive (Submission ID: PRJNA509278).

Results

VapC-mt11 expression causes growth arrest in both *M. smegmatis* and *M. tuberculosis*

M. tuberculosis H37Rv and *M. smegmatis* mc² 155 cells were transformed with a VapC-mt11-containing anhydrotetracycline (ATc) inducible vector (pMC1s, (Ehrt et al., 2005)). For *M. smegmatis*, VapC-mt11 expression resulted in growth arrest 1 h post-induction which was sustained through the length of the growth profile (Figure 28A). Likewise, VapC-mt11 expression arrested growth in *M. tuberculosis*. In contrast to the uninduced control, the OD600 for cells expressing VapC-mt11 remained near zero for the duration (10 days) of the growth profile (Figure 28B). In concordance with the growth profile, the average CFU/ml recovered when *M. tuberculosis* cells were plated in the presence of ATc inducer was 50 CFU/ml compared to 12,000 CFU/ml from the uninduced control (Figure

28C). These results were consistent with the strong growth inhibition we observed when VapC-mt11 was expressed in *E. coli* cells (Sharp et al., 2012).

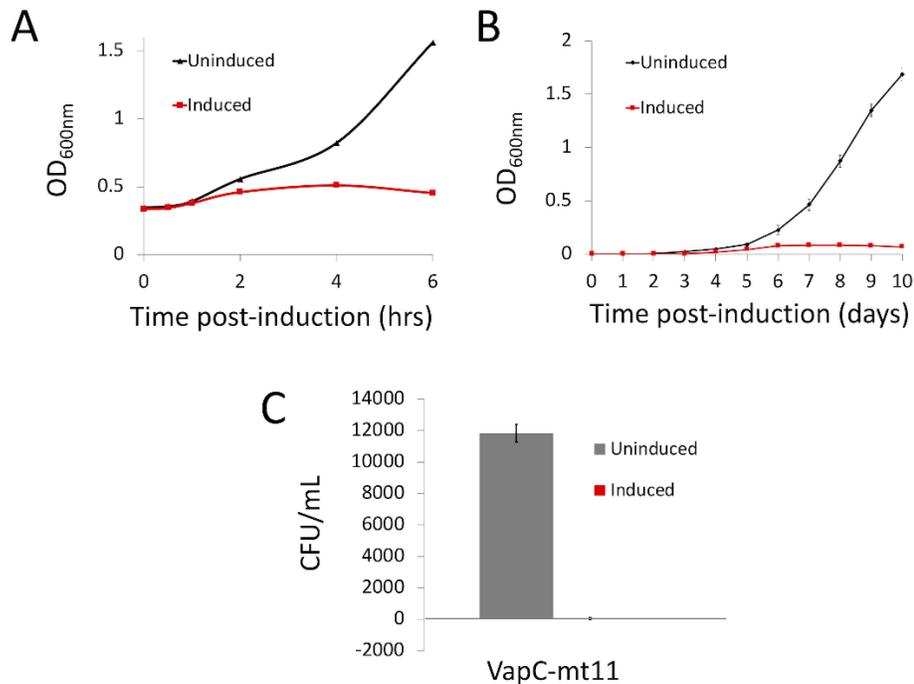


Figure 28. Growth inhibition by VapC-mt11. (A) *M. smegmatis* mc² 155 harboring the anhydrotetracycline (ATc) inducible vector pMC1s with VapC-mt11 was grown at 37 °C in 7H9 + TW80 + ADN. Cultures were split and one culture was induced by the addition of ATc. Time points for uninduced (black) and induced (red) samples were collected up to 6 h. Growth curve shown is representative of the trend documented in three independent experiments. (B) *M. tuberculosis* H37Rv transformed with pMC1s-VapCmt11 were grown at 37 °C in 7H9-TW80-ADN medium containing 30% spent medium and supplemented with 25 ug/mL kanamycin. Cultures were separated into uninduced (black) and induced (red). Induced cultures were supplemented with 200 ng/ml ATc and

additional ATc was added every 48 h to maintain the ATc concentration between 100 and 250 ng/ml. Data points represent the average of three independent experiments; error bars represent the S.D. (C) Plate efficiency assay for *M. tuberculosis* H37Rv cells harboring pMC1s-VapC-mt11 between uninduced (gray) and induced (red). Transformants were plated on 7H9 + TW80 + ADN agar with 25 ug/mL kanamycin with or without 500 ng/mL of ATc and incubated for 3 weeks at 37 °C. Error bars represent the S.D. for biological and technical replicates.

VapC-mt11 inhibits translation

An in vitro translation system was used to determine if the strong growth arrest phenotype characteristic of VapC-mt11 was a consequence of defective protein synthesis. Preincubation of the in vitro translation mix with VapC-mt11 before DHFR DNA template addition resulted in complete inhibition of synthesis relative to the control lane which was not treated with recombinant VapC-mt11 (Figure 29A). This translation defect was confirmed upon VapC-mt11 expression in *M. smegmatis* (Figure 29B). Cells were radioactively labeled at intervals after VapC-mt11 induction to monitor new protein synthesis. Nearly complete shutdown of protein synthesis was observed 2 h after induction (the point where growth inhibition commences Figure 28A) followed by sustained and complete inhibition of protein synthesis after 4 h of toxin induction. Therefore, the strong growth arrest phenotype characteristic of VapC-mt11 expression appears to be a consequence of complete translation inhibition in mycobacteria.

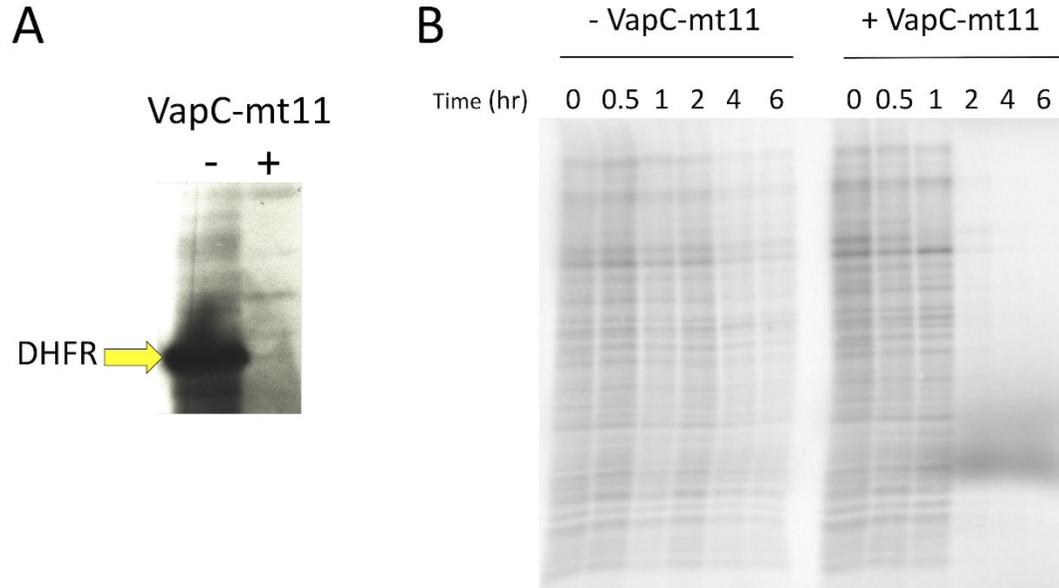


Figure 29. VapC-mt11 inhibits translation in vitro and in *M. smegmatis*. (A)

The PURExpress translation reaction was incubated with (+) or without (-) VapC-mt11. Production of the control DHFR template was assayed (yellow arrow).

(B) [³⁵S]-Methionine incorporation in *M. smegmatis* mc² 155 cells grown to exponential phase and split into uninduced (-VapC-mt11) and induced (+VapC-mt11) cultures. Cell aliquots were collected for up to 6 h post induction.

Equivalent amounts of cell lysate (resuspended in appropriate Laemmli buffer volumes to normalize for differences in OD600) were subjected to SDS-PAGE and visualized on a phosphorimager.

VapC-mt11 is a tRNase that targets multiple synthetic *M. tuberculosis* tRNAs

Since there is precedent for VapC recognition of both tRNA and rRNA targets that is highly structure-dependent (Jonathan W. Cruz et al., 2015; K. S.

Winther et al., 2013), we assayed VapC-mt11 cleavage activity with each of these two intrinsically folded classes of RNAs.

We first analyzed 23S and 16S rRNA from *E. coli* cells expressing VapC-mt11 and did not observe a decrease in the overall abundance of these rRNAs or the presence of any degradation products (data not shown). This suggested that VapC-mt11 likely targets tRNAs. Next, to determine which tRNAs were preferred substrates of VapC-mt11, each of the 45 tRNAs present in *M. tuberculosis* were synthesized and incubated with recombinant VapC-mt11 (Figure 30). To our surprise, VapC-mt11 cleaved 15 tRNAs, one third of the 45 *M. tuberculosis* tRNAs. However, among these 15, ten were cleaved efficiently (eight to completion, two ~90% cleaved; shown in red, Figure 30) while five exhibited only weak cleavage (blue, Figure 30) The extent of cleavage did not change when assay times were increased (data not shown).

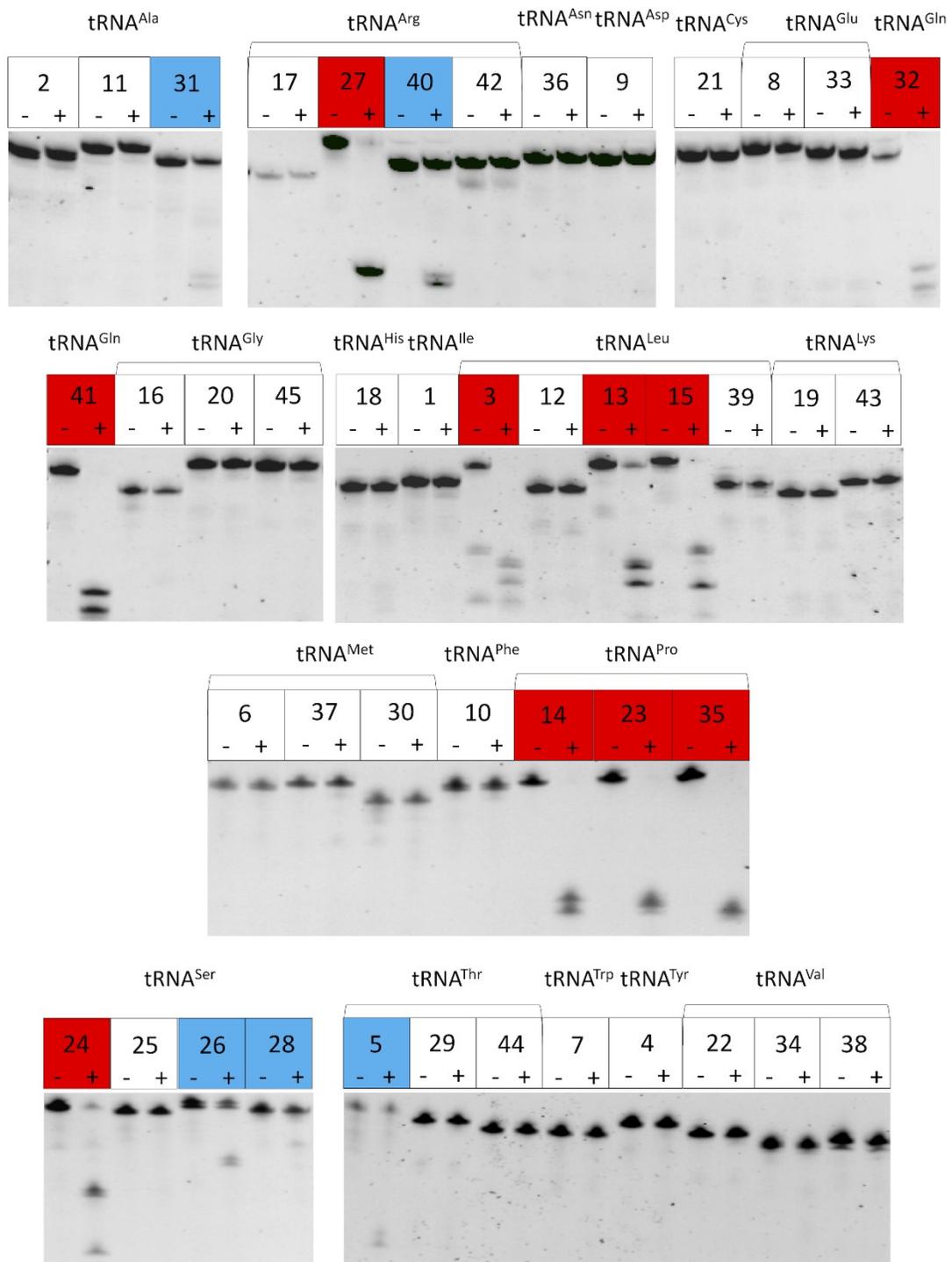


Figure 30. VapC-mt11 cleaves one third of *M. tuberculosis* tRNAs in vitro.

tRNAs and their cleavage products visualized by SYBR Gold staining. 2 pmol of

each of the 45 in vitro synthesized *M. tuberculosis* tRNAs were incubated with (+) or without (-) 10 pmol of recombinant VapC-mt11 for 3 h at 37 °C to ensure detection of both strong and weak tRNA cleavage reactions. Full length tRNAs that were cleaved to completion or near completion by VapC-mt11 shown in red; these tRNAs were also cleaved to the same extent shown when incubated for only 15 min at 37 °C. Weak VapC-mt11 tRNA targets shown in blue. tRNA numbering and anticodon sequences of each numbered tRNA from the Lowe lab genomic tRNA database.

VapC-mt11 tRNA recognition and cleavage is not modification dependent

To determine if posttranscriptional modifications present on tRNAs influenced cleavage, we used northern analysis to assess the cleavage efficiency of VapC-mt11 for all 15 tRNAs synthesized in vitro (Figure 30) versus in vivo in *M. tuberculosis* cells (Figure 31 and Figure 32). Generally, tRNA cleavage efficiency trends were in agreement. The in vivo counterparts for eight of the ten in vitro tRNAs that were cleaved to ~90-100% completion (Figure 30, red) were also completely cleaved (Figure 31). The two exceptions, tRNA^{Leu15} and tRNA^{Ser24}, were ~50% cleaved by VapC-mt11 when purified from *M. tuberculosis* cells (Figure 32C and E, respectively). Four of the five very weakly cleaved in vitro synthesized tRNAs (blue tRNAs in Figure 30) exhibited marginal or no detectable cleavage by northern analysis (Figure 32). The exception was in vivo synthesized tRNA^{Arg40}, which was fully cleaved by VapC-mt11 (Figure 32B). Since cleavage efficiency was generally comparable between synthetic tRNA

and its counterpart isolated from *M. tuberculosis* cells, VapC-mt11 target recognition and cleavage did not appear to require tRNA modifications.

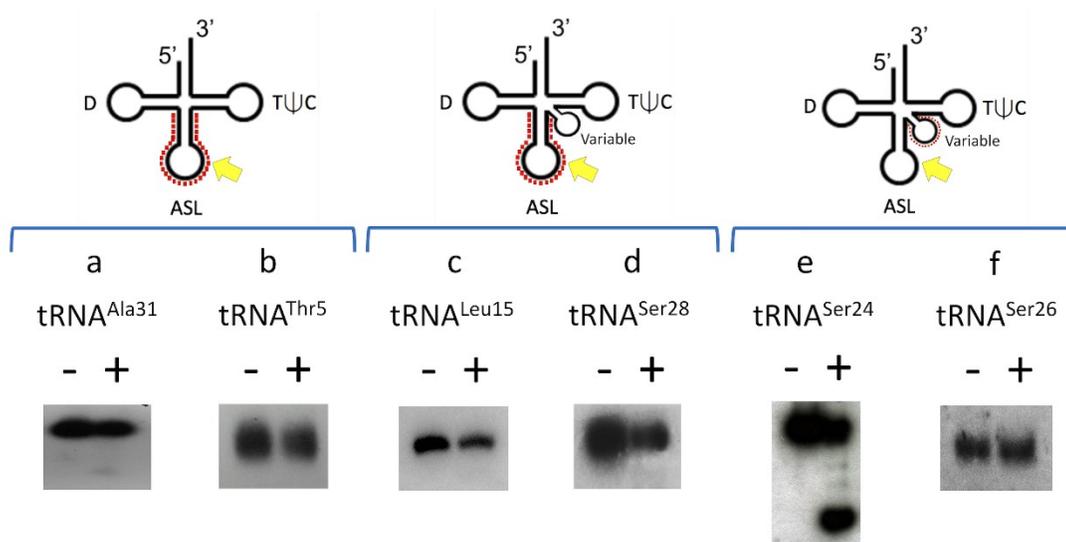


Figure 31. tRNA modifications are not required for VapC-mt11 target recognition in vitro. Total *M. tuberculosis* RNA was incubated with (+) or without (-) recombinant VapC-mt11. Northern analysis of (a) tRNA^{Arg27} (b) tRNA^{Arg40} (c) tRNA^{Gln32} (d) tRNA^{Gln41} (e) tRNA^{Pro14} (f) tRNA^{Pro23} (g) tRNA^{Pro35} (h) tRNA^{Leu3} (i) tRNA^{Leu13}. The light band in the + lane of panel b is background, not uncleaved tRNA. Position of the oligonucleotides used are shown by the red dots on the tRNA diagrams above each bracketed group. Cleavage positions, yellow arrow. Oligonucleotides were designed to hybridize to the ASL to optimally differentiate between tRNA species. Therefore, cleavage products were generally not visible because the oligonucleotides could no longer hybridize to the cleaved ASL. Hybridization temperatures were optimized to preclude cross-hybridization

to other tRNAs. tRNA numbering and anticodon sequences of each numbered tRNA from the Lowe lab genomic tRNA database (<http://gtrnadb.ucsc.edu>).

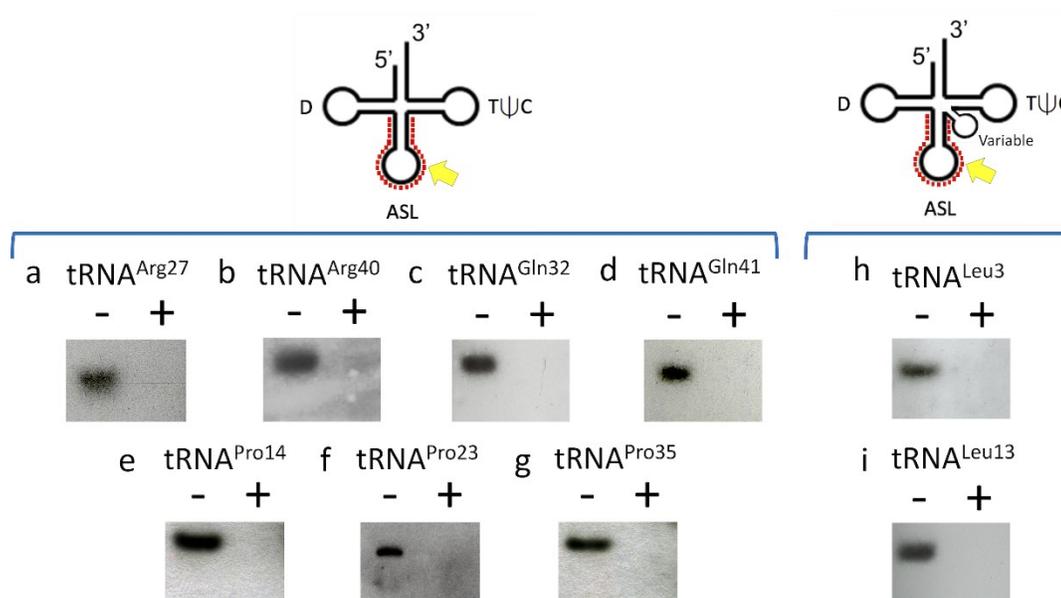
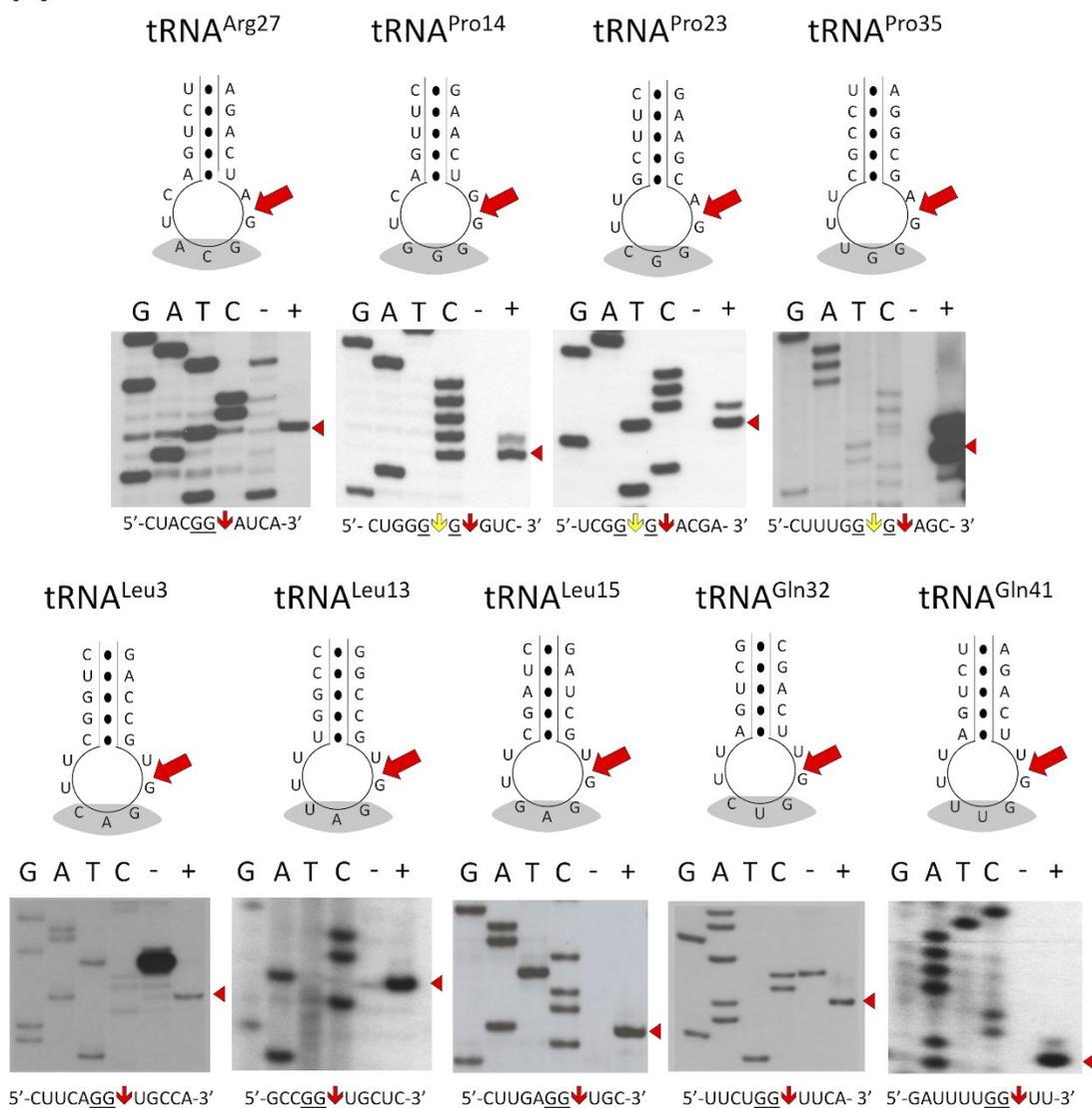


Figure 32. Weak or uncleaved in vivo synthesized tRNAs mirror the subset of weakly cleaved in vitro synthesized tRNAs. Total *M. tuberculosis* RNA was incubated with (+) or without (-) recombinant VapC-mt11. Northern analysis of (a) tRNA^{Ala31} (b) tRNA^{Thr5} (c) tRNA^{Leu15} (d) tRNA^{Ser28} (e) tRNA^{Ser24} (f) tRNA^{Ser26}. Position of the oligonucleotides used are shown by the red dots on the tRNA diagrams above each bracketed group. Cleavage positions, yellow arrow. Oligonucleotides were designed to hybridize to the ASL to optimally differentiate between tRNA species. Therefore, cleavage products are not visible. Hybridization temperatures were optimized to preclude cross-hybridization to other tRNAs. tRNA numbering and anticodon sequences of each numbered tRNA from the Lowe lab genomic tRNA database (<http://gtrnadb.ucsc.edu>).

VapC-mt11 generates tRNA halves upon cleavage after a GG sequence in the anticodon loop

We used primer extension analysis to map the VapC-mt11 cleavage sites for nine tRNAs that were efficiently cleaved in vitro (Figure 33A), all but one of these (tRNA^{Leu15}) were also 100% cleaved in vivo. In each case, three trends were observed.

A



B

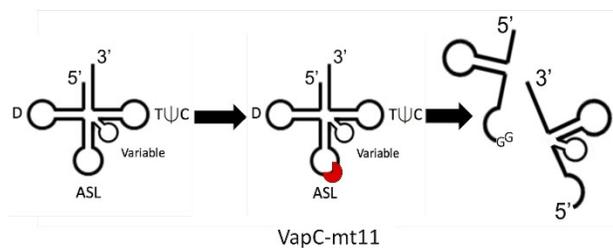


Figure 33. VapC-mt11 cleaves tRNAs within the ASL. (A) Primer extension analysis of in vitro synthesized *M. tuberculosis* tRNA^{Arg27}, tRNA^{Pro14}, tRNA^{Pro23},

tRNA^{Pro35}, tRNA^{Leu3}, tRNA^{Leu13}, tRNA^{Leu15}, tRNA^{Gln32} and tRNA^{Gln41} treated with (+) or without (-) recombinant VapC-mt11. G, A, T, and C correspond to DNA-sequencing ladders using the same oligonucleotide as in the primer extension reactions for each tRNA. RNA sequence shown below the gels and major cleavage sites are indicated by the red arrow; alternate weak cleavage sites are indicated by yellow arrow for the three proline tRNAs. Major cleavage products in the gels are depicted by the red arrow head (right). The ASL sequence is illustrated above each tRNA sequencing gel, the anticodon (grey shading) and major cleavage position indicated by red arrow. Bands visible in (-) lanes of tRNA^{Arg27}, tRNA^{Leu3} and tRNA^{Gln32} correspond to secondary structure. (B) Diagram of tRNA halves produced by VapC-mt11. tRNA numbering and anticodon sequences of each numbered tRNA from the Lowe lab genomic tRNA database (<http://gtrnadb.ucsc.edu>).

First, the primary cleavage site always occurred after a GG sequence (GG↓). Since we observed cleavage at GG↓A, GG↓U and GG↓G sequences among these nine tRNAs, the base following cleavage site does not appear to contribute to target specificity. Yet, several GG-containing mRNAs were not cleaved by VapC-mt11. For example, *E. coli* OmpF and OmpC mRNAs contain 67 and 65 GG motifs, respectively, but were not cleaved by VapC-mt11 (data not shown). Likewise, as mentioned earlier, *E. coli* 23S and 16S rRNAs were not also cleaved, and they contain hundreds of GG motifs. Therefore, the presence of a GG RNA consensus sequence alone cannot serve as the sole determinant for tRNA target recognition. Second, all nine cleavage sites occurred at the same

position in the anticodon loop, 3' of the ribonucleotide that followed the anticodon. Third, VapC-mt11 always targeted the anticodon loop. This is consistent with our earlier observation for another VapC family member and a subsequent paper demonstrating that several *M. tuberculosis* VapC toxins target tRNAs for cleavage at their anticodon loop (Jonathan W. Cruz et al., 2015; K. Winther et al., 2016). Therefore, VapC-mt11, and all *M. tuberculosis* VapC toxins known to cleave tRNA, result in the generation of tRNA halves (Figure 33B).

VapC-mt11 requires a GG sequence for tRNA recognition and accurate cleavage

To pinpoint the determinants that contribute to toxin recognition and cleavage by VapC-mt11, we created a variety of mutants using one of the efficiently cleaved tRNAs, tRNA^{Pro14}, as the test tRNA template (Figure 34A). First, we interrogated the importance of the conserved GG sequence 5' of the cleavage site. Mutation of the second G of the GG consensus to an A (GG→GA) resulted in the shifting of the cleavage site one nt 5' from the major cleavage site (which also corresponded to the position of the minor cleavage site in wild-type tRNA^{Pro14}, Figure 34B). Next, we mutated the first G of the GG consensus to an A (GG→AG). In this case, we observed very inefficient cleavage, with only a

small percentage of the tRNA^{Pro14} template cut by VapC-mt11 (Figure 34C).

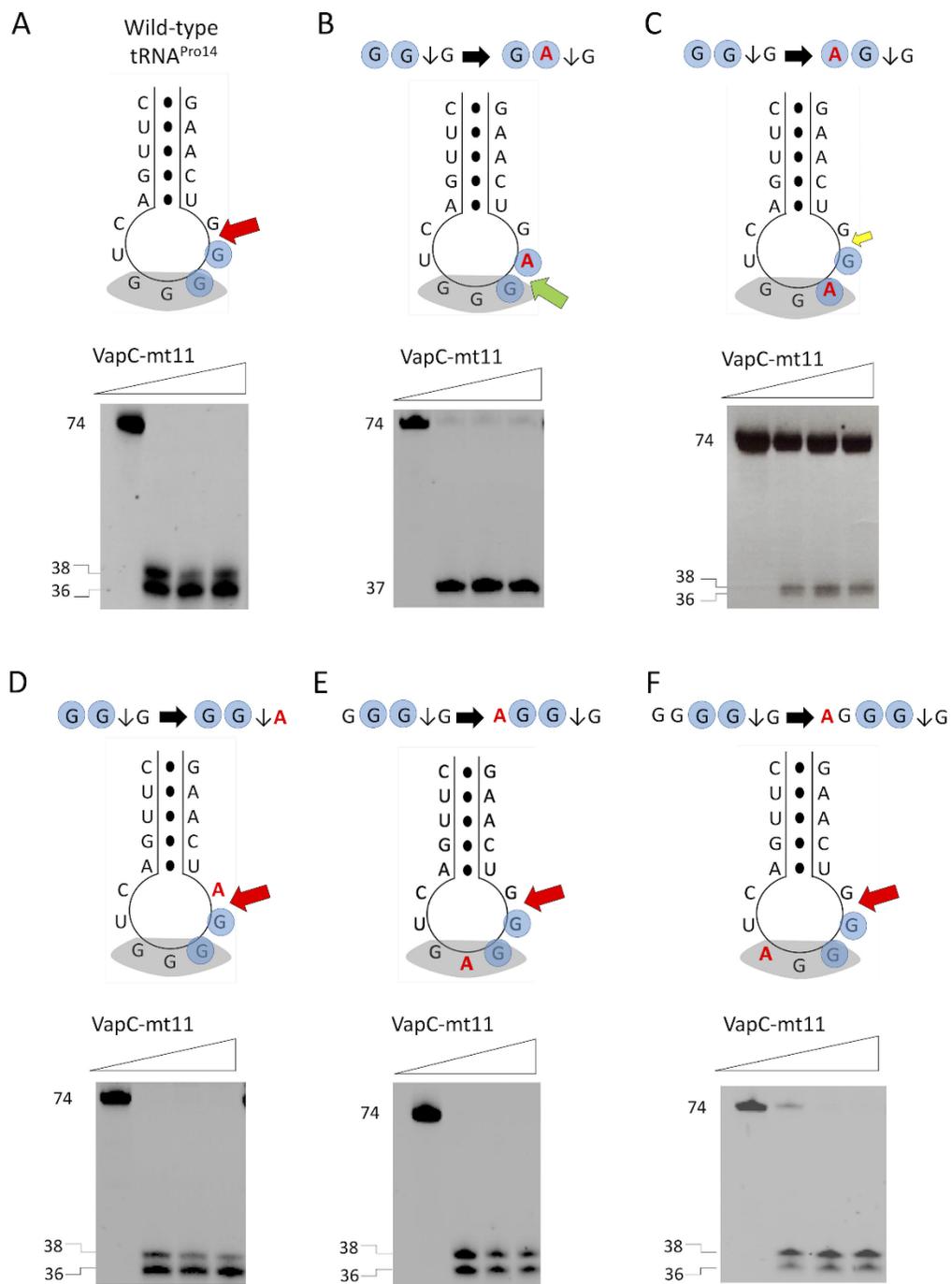


Figure 34. VapC-mt11 requires a GG sequence within the proper context. In vitro synthesized *M. tuberculosis* tRNA^{Pro14} mutants were incubated with increasing amounts of recombinant VapC-mt11 (ratios of toxin:tRNA were 0:1, 5:1, 10:1, 15:1) for 3 h at 37 °C. Sizes for full-length and cleavage products (left). Cleavage assays with in vitro synthesized wild-type tRNA^{Pro14} (A), tRNA^{Pro14} point mutation (G → A) of each G residue of the consensus GG sequence (blue shaded circles) mutants (B,C), tRNA^{Pro14} mutant with point mutation 3' of cut site (D) and tRNA^{Pro14} mutants with point mutations within the anticodon sequence (E,F). Secondary structure of *M. tuberculosis* tRNA^{Pro14} wild-type or mutants shown above gels. Anticodon sequence (shaded grey), base pairing represented as black dots (●), mutated bases (red), consensus sequence (shaded blue circles), wild-type cleavage site (red arrow), alternate cleavage site (green arrow), weak cleavage at wild-type cleavage site (small yellow arrow).

Since we observed cleavage at GG↓A, GG↓U and GG↓G sequences among the 10 preferred tRNA targets (Figure 33A), the base following the GG consensus sequence did not appear to contribute to target specificity. In agreement with this prediction, cleavage was not affected when we mutated the base 3' of the GG consensus in tRNA^{Pro14} (Figure 34D).

Next, we individually mutated each of the two bases 5' of the GG cleavage consensus sequence, which in each case alters the anticodon sequence (GGG↓G → AGG↓G, Figure 34E; GGGG↓G → AGGG↓G, Figure 34F). Neither mutation altered the cleavage of tRNA^{Pro14} by VapC-mt11. Therefore, the only

sequence determinants for cleavage of tRNA^{Pro14} by VapC-mt11 were the two GG residues directly 5' of the cleavage site.

VapC-mt11 cleavage is both sequence- and structure-dependent

The importance of a GG sequence for VapC-mt11 cleavage is underscored when both consensus GG residues were simultaneously mutated to AA (GG→AA, Figure 35A). In this case, we observed a two base 5' shift in the cleavage site to the only other GG sequence remaining in the anticodon loop,

suggesting that in vitro VapC-mt11 will seek out a GG in proximity when the native sequence is no longer in place.

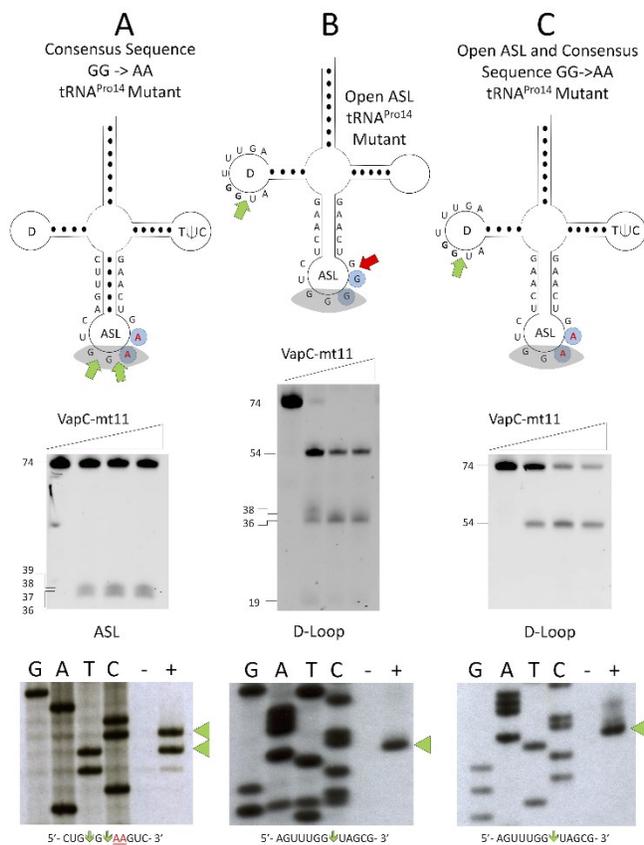


Figure 35. Sequence and structure influence the VapC-mt11 cleavage site.

Cleavage assays of (A) tRNA^{Pro14} mutant with a GG consensus sequence replacement by AA, (B) an open stem loop mutant, and a (C) GG → AA and open stem loop mutant. Full length and cleavage product sizes shown on the left. The small 19 nucleotide cleavage product ran off the gel in (C). Secondary structures of the in vitro synthesized mutants of tRNA^{Pro14} are depicted above cleavage reaction gels. Anticodon stem loop (ASL), anticodon sequence (shaded grey), base pairing represented as black dots (●), mutated bases (red),

consensus sequence (blue shaded circles), weak cleavage at wild-type cleavage site (small yellow arrow), and alternate cleavage site (green arrow). Primer extension of these mutants shown below. G, A, T, and C correspond to DNA-sequencing ladders using the same oligonucleotide as in the primer extension reactions for each tRNA. Alternate cleavage sites are indicated by the green arrow head to the right of the primer extension gel images. RNA sequence corresponding to the alternate cleavage sites (green arrow) within the (A) ASL or (B,C) D-loop are shown below each primer extension gel image, mutations indicated by underlined red text.

Since we established that the presence of the GG sequence alone could not account for the specificity of VapC-mt11 for tRNAs, we first altered the structure of the anticodon stem loop (ASL) by abolishing stem formation (Figure 35B). This relatively severe structural change resulted in an equally dramatic shift in the cleavage site. This open ASL mutant primarily shifted the cleavage site from the GG in the anticodon loop to a GG sequence in the D-loop of tRNA^{Pro14} (Figure 35B, green arrow); only marginal cleavage at the native GG site was detected (Figure 35B, small yellow arrow).

Finally, when we combined the open ASL mutation (Figure 35B) with the consensus GG→AA mutation (Figure 35A), we observed a complete shift of the cleavage site to GG↓ in the D-loop (Figure 35C). Apparently, a GG sequence in the loop of an intact stem-loop (much like the original site harbored in the loop of the ASL) was a preferred alternate substrate over other spatially distinct GG motifs present in tRNA^{Pro14}. More specifically, a second GG located in a single-

stranded region between the ASL and T_ψC loop of tRNA^{Pro14} was not cut in this mutant. Alternatively, the selection of a new cleavage site in this mutant may simply be dictated by toxin accessibility as the D-loop would be predicted to be more surface exposed than the other GG site tucked between two stem loops.

In summary, VapC-mt11 requires the GG consensus sequence to be in the proper structural context within the anticodon loop for tRNA cleavage in vitro.

VapC-mt11 cleaves only two GG consensus-containing tRNA targets when expressed in its natural *M. tuberculosis* host

We sought to understand why our in vitro cleavage experiments using either synthetic tRNAs or those isolated from *M. tuberculosis* (with modifications) identified numerous VapC-mt11 cleaved tRNAs compared to three identified by Gerdes and colleagues in a screen for *M. tuberculosis* VapC RNA targets expressed in another mycobacteria, *M. smegmatis* (K. Winther et al., 2016).

We expressed VapC-mt11 in *M. tuberculosis* and performed a specialized RNA-seq method, 5' RNA-seq, developed in our laboratory to identify RNA targets of endoribonuclease toxins (Schifano et al., 2014). 5' RNA-seq methodology enables global analysis of specific populations of RNA transcripts based on the modification at their 5' end (Schifano et al., 2014). VapC family members as well as other RNases have a hydroxyl group (5'-OH) at their 5' ends (Jonathan W. Cruz et al., 2015; K. S. Winther & Gerdes, 2011). Therefore, we analyzed only those transcripts carrying a 5'-OH in VapC-mt11 expressing *M.*

tuberculosis cells compared to an uninduced control. The resulting dataset revealed both the RNA target and the position of toxin cleavage within these RNAs. Only two tRNAs were identified as VapC-mt11 targets, tRNA^{Gln32-CUG} and tRNA^{Leu3-CAG} (Figure 36A). Both of these tRNAs were among the ten that were cleaved to completion in vitro (Figure 30). In concordance with the requirements for substrate recognition and specificity we documented in vitro, both tRNA^{Gln32-CUG} and tRNA^{Leu3-CAG}, were cleaved within their ASL and immediately after the GG consensus sequence to generate stable tRNA halves (Figure 36B). Finally, 5' RNA-seq did not identify any other class of RNA—mRNA, rRNA or other small stable noncoding RNA—that was directly cleaved by VapC-mt11.

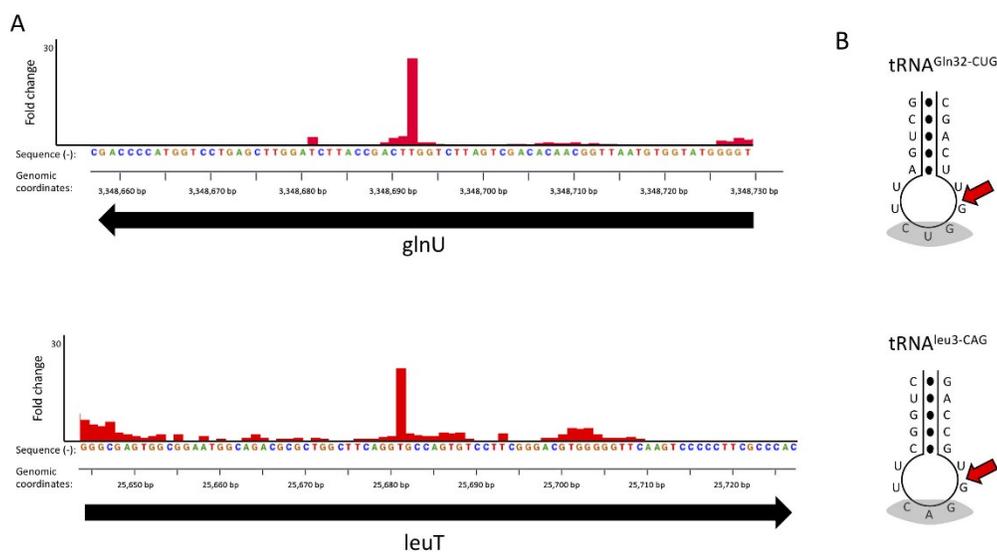


Figure 36. VapC-mt11 cleaves only two tRNAs containing the GG cleavage consensus sequence in vivo. (A) Histogram representing the ratio of cleavage by VapC-mt11 identified using 5' RNA-seq at each nucleotide within the glnU

gene (encoding tRNA^{Gln32-CUG}) and LeuT gene (encoding tRNA^{Leu3-CAG}) in *M. tuberculosis* after 24 hours of toxin induction. Genomic positions and the negative strand sequence for glnU and positive strand sequence for leuT are shown. (B) ASL sequence of the two in vivo VapC-mt11 tRNA targets, with cleavage site following the conserved GG consensus sequence indicated (red arrow).

Discussion

Here we showed that a single toxin is capable of recognizing and cleaving nine *M. tuberculosis* tRNAs, one-fifth of the 45 distinct tRNA species present in this pathogen in vitro (Figure 30, Figure 31). The broad scope of targets identified in vitro seemed consistent with the growth and translation phenotypes observed. Expression of VapC-mt11 dramatically impaired *M. tuberculosis* growth (Figure 28B) and precluded recovery of viable cells (Figure 28C). Growth in *M. smegmatis* was also arrested ~2 h postinduction and beyond (Figure 28A), concomitant with a virtually complete shutdown of new protein synthesis (Figure 29B).

A common feature among these tRNA targets was sequence- (GG↓) and context-specific (anticodon loop of an intact ASL) cleavage. In fact, the requirement for proper structural context is emerging as a general feature of tRNA-cleaving *M. tuberculosis* toxins (Jonathan W. Cruz et al., 2015; Schifano et

al., 2016). However, RNA target selection by VapC-mt11 in *M. tuberculosis* is dictated by more than the enzymatic properties of this toxin.

In its natural host VapC-mt11 was far more discriminating, cleaving only tRNA^{Gln32-CUG} and tRNA^{Leu3-CAG}. This much higher level of target discrimination in vivo may be due to modifications on the tRNAs or the binding of an accessory factor that alters enzyme specificity. However, evidence only exists to support the former explanation. Although we did not detect any difference in cleavage for synthetic tRNA^{Gln32-CUG} and tRNA^{Leu3-CAG} versus those prepared from unstressed *M. tuberculosis* cells, newly added or newly removed post-transcriptional modifications to the tRNA that prevent cleavage of the other seven tRNAs in vivo might occur only upon expression of the toxin. Although modifications on *M. tuberculosis* tRNAs have not been characterized in any detail, the ASL of tRNAs is generally a modification hotspot that influences tRNA structure and thermostability (Lorenz, Lünse, & Mörl, 2017). If the changes we observed in specificity are indeed attributable to modifications, the VapC-mt11 target recognition seen in vivo cannot be accurately recapitulated in vitro until the exact location and nature of the chemical changes on the entire *M. tuberculosis* tRNA population is elucidated.

The toxin components of Type II TA systems are characteristically activated in response to stress (Harms et al., 2018). In *M. tuberculosis*, certain TA toxin transcripts are upregulated when cells are exposed to stresses relevant to latent tuberculosis infection—nutrient limitation (Albrethsen et al., 2013; Betts et al., 2002; Cortes et al., 2013; Gupta, Venkataraman, Vasudevan, & Gopinath

Bankar, 2017; Tiwari et al., 2015), hypoxia (Ramage et al., 2009; Rustad et al., 2008; Tiwari et al., 2015), macrophage infection (Cappelli et al., 2006; Fontan et al., 2008; Korch et al., 2009; Ramage et al., 2009) or antibiotic treatment (Denkin et al., 2005; Gupta et al., 2017; Provvedi et al., 2009; R. Singh et al., 2010; Tiwari et al., 2015). However, there is no definitively confirmed stress trigger for any *M. tuberculosis* TA toxin linking upregulation of toxin mRNA to increased enzymatic activity of the toxin protein. By analogy, TA toxin expression may underlie the observation that hypoxia regulates tRNA modifications in the *M. tuberculosis* relative *M. bovis* BCG (Chionh et al., 2016).

There are five Leu tRNAs in *M. tuberculosis* to service the six Leu codons. tRNA^{Leu3-CAG} (depleted by VapC-mt11) services the most abundant of the six Leu codons, CUG. The Leu CUG codon is also one of the three most frequently represented codons in the *M. tuberculosis* transcriptome (50.4/1000 codons; equal to Gly GGC but less than Ala GCC at 59.8/1000 codons). The other four Leu tRNAs, are spared but may not be able to fully compensate for loss of tRNA^{Leu3-CAG}.

There are two Gln tRNAs in *M. tuberculosis* to service the two Gln codons, CAG and CAA. tRNA^{Gln32-CUG} (depleted by VapC-mt11) services the CAG codon (22.8/1000 codons). If the 5' anticodon C for tRNA^{Gln32-CUG} is modified, it may also service the CAA Gln codon (8.1/1000 codons). The second Gln tRNA (spared by VapC-mt11) services both Gln CAG and CAA codons.

Gerdes and colleagues identified RNA targets for several *M. tuberculosis* VapC toxins using a genome-scale toxin-RNA interaction screen in the rapidly growing, nonpathogenic *M. smegmatis* (K. Winther et al., 2016). This screen identified three VapC-mt11-interacting *M. smegmatis* tRNAs—two isoacceptors of leucine, and one of glutamine. These three *M. smegmatis* tRNAs are designated tRNA^{Leu3-CAG}, tRNA^{Leu13-GAG}, and tRNA^{Gln10-CUG} and are orthologs of *M. tuberculosis* tRNA^{Leu3-CAG}, tRNA^{Leu15-GAG} and tRNA^{Gln32-CUG}, respectively. Of the three *M. smegmatis* tRNAs, only tRNA^{Leu3-CAG} was significantly cleaved in vivo upon ectopic expression of VapC-mt11 in this surrogate host. Thus, only tRNA^{Leu3-CAG} in both *M. smegmatis* and *M. tuberculosis* was subjected to more detailed study. More specifically, 1) the sites of VapC-mt11 cleavage of *M. smegmatis* tRNA^{Leu3-CAG} in vitro were identified by primer extension and 2) relatively efficient in vitro cleavage of *M. tuberculosis* tRNA^{Leu3-CAG} was demonstrated (K. Winther et al., 2016). Therefore, their data focused on tRNA^{Leu3-CAG} as the likely preferred RNA target for *M. tuberculosis* VapC-mt11. Our 5' RNA-seq approach has three advantages over this interaction screen as well as in vitro cleavage assays similar to those presented in Figure 30 and Figure 31. First, by focusing on only 5'-OH transcripts we are identifying transcripts that have been explicitly altered by the enzymatic activity of the toxin, not simply a cross-linkable RNA-protein interaction. Second, this method enables both the detection of the RNA target(s) as well as the exact position of cleavage without needing to perform laborious primer extension experiments. Third,

enzyme activity occurs under physiological conditions (i.e., an *M. tuberculosis* host) to best approximate toxin activity during infection.

Recently, Deep et al. reported a high resolution x-ray crystal structure of VapBC-mt11 along with physiological studies of VapC-mt11 (Deep et al., 2018). VapC-mt11 overexpression in *M. tuberculosis* followed by RNA-seq led to an altered transcriptome that mirrored those of *M. tuberculosis* cells undergoing the enduring hypoxic response as well as cells in a nonreplicating persistent state (Deep et al., 2018). Yet paradoxically, deletion of the vapBC11 locus resulted in decreased recovery of viable cells only upon exposure to oxidative stress. Other stress conditions tested—including nonreplicating persistence, hypoxia using the Wayne model, nitrosative stress, nutritional stress, macrophage infection, and antibiotic treatment—did not alter cell recovery (Deep et al., 2018). Finally, although not essential for cell survival, the vapBC11 locus is required for establishment of an *M. tuberculosis* infection in the guinea pig model (Deep et al., 2018).

In summary, our results reveal that analysis of toxin cleavage targets is most accurate in the in vivo setting in which the toxin is active, i.e. within the *M. tuberculosis* cell. Although in vitro cleavage methods are useful for identification of consensus sequences and structural features required for toxin cleavage, they have inherent limitations that may preclude accurate identification of the true toxin targets in vivo. Likewise, while ectopic toxin expression in alternate hosts followed by 5' RNA-seq can provide useful clues to RNA targets or general RNA class favored by the toxin, the results will likely differ from that performed in the

true host because of the inherent variability between transcriptomes and the presence and position of RNA modifications. Finally, expression of toxins in their natural host not only identifies the precise RNA target(s), it provides more physiologically relevant clues about the impact of toxin activity on discrete pathways and biological processes.

Chapter 4: Toxin-mediated, cysteine-specific ribosome stalling reveals hidden sORFs in the *Mycobacterium tuberculosis* genome and triggers targeted stress survival pathways

Abstract

The vast majority of those infected with *Mycobacterium tuberculosis* do not develop active tuberculosis (TB). Instead, this bacterial pathogen reprograms its physiology to establish a latent TB infection. The ~90 toxin-antitoxin (TA) systems harbored by this pathogen have been implicated in this process. To better understand the molecular underpinnings of this transition, we studied the physiological role of Vap4 toxin using a multi-pronged, systems-level approach to first identify the initial toxin target, then track the downstream alterations to the *M. tuberculosis* transcriptome and proteome initiated by toxin activity. Our findings reveal surprisingly nuanced roles for a representative TA system in modulating *M. tuberculosis* physiology. VapC4 exclusively inactivated the sole tRNA^{Cys} through cleavage at a single site within its anticodon sequence. Depletion of the pool of tRNA^{Cys} led to ribosome stalling at Cys codons within actively translating mRNAs. Genome mapping of the hundreds of transcripts carrying these Cys-stalled ribosomes unexpectedly uncovered 96 unannotated, apparently legitimate, ORFs undergoing active translation. Among these, ~50% were small ORFs (sORFs) encoding Cys-rich proteins <50 amino acids that appear to function as novel Cys-responsive attenuators. Tracking downstream VapC4-mediated molecular events by RNA-seq and quantitative mass spectrometry

revealed that cells deplete tRNA^{Cys} on one hand, so that they can enlist Cys-attenuation to redirect cell physiology toward the synthesis of L-cysteine. While seemingly counterintuitive, this new pool of L-cysteine serves as a precursor for synthesis of mycothiol, the *M. tuberculosis* glutathione counterpart, to act as *the* major cellular redox buffer. L-cysteine also serves as a sulfur donor for the synthesis of critical redox-sensing iron sulfur transcription regulators. In parallel, VapC4 activates other pathways that ameliorate the killing properties of copper as well as reactive oxygen and nitrogen species imparted by macrophages.

Introduction

The *M. tuberculosis* genome harbors ~90 Type II toxin-antitoxin (TA) systems (Ramage et al., 2009; Sala et al., 2014), the majority belonging to the 50 member virulence associated protein VapBC family (VapB - antitoxin, VapC -toxin) unique to pathogenic bacteria. In general, Type II TA systems comprise adjacent genes encoding two small (~10 kDa) proteins, a toxin and its cognate antitoxin that inhibits toxin activity through formation of a stable TA protein-protein complex. In the prevailing model, stress conditions lead to lower levels of the antitoxin and thus, a preponderance of free toxin, which exerts its growth-regulating and/or other functions from within the bacterial cells (Harms et al., 2018; Masuda & Inouye, 2017). In *M. tuberculosis*, several of these toxins are upregulated when cells are exposed to stresses relevant to latent tuberculosis infection: nutrient limitation (Albrethsen et al., 2013; Betts et al., 2002; Cortes et al., 2013; Gupta et al., 2017; Tiwari et al., 2015), hypoxia (Ramage et al., 2009; Rustad et al., 2008;

Tiwari et al., 2015), macrophage infection (Cappelli et al., 2006; Fontan et al., 2008; Korch et al., 2009; Ramage et al., 2009) or antibiotic treatment (Denkin et al., 2005; Gupta et al., 2017; Provvedi et al., 2009; R. Singh et al., 2010; Tiwari et al., 2015). Therefore, since the expression of individual *M. tuberculosis* toxins also typically leads to growth inhibition, they have been implicated in the establishment of latent TB infection (Arcus et al., 2011, 2005; Harms et al., 2018; Ramage et al., 2009; Sala et al., 2014).

M. tuberculosis VapC toxins are structure and sequence-specific endoribonucleases (J W Cruz & Woychik, 2016). Only a few of the 50 *M. tuberculosis* VapC toxins have been studied in detail. In vitro cleavage studies demonstrated that VapC4 (aka VapC-mt4, Rv0595c) and VapC11 (aka VapCmt-11, Rv1561) require a cleavage consensus sequence and specifically target tRNA (Cintrón et al., 2019; Jonathan W. Cruz et al., 2015; Sharp et al., 2012). Other *M. tuberculosis* VapC toxins with apparent tRNase activity were identified using an RNA-VapC interaction screen in an *M. smegmatis* host (K. Winther et al., 2016). Only recently did we learn that accurate tRNA target identification requires in vivo cleavage in the matched mycobacterial host cell.

Here we use physiologically relevant in vivo studies in tandem with powerful genome-scale approaches to accurately identify the single tRNA^{Cys} target of VapC4 in *M. tuberculosis* then track and model how this toxin controls growth and facilitates pathogen survival during latent TB infection by counterattacking deadly assaults from the host innate immune response.

Methods

Strains, plasmids and reagents

All experiments were performed using either *M. tuberculosis* strain H37Rv (ATCC 25618), *M. tuberculosis* mc² 6206 ($\Delta panCD \Delta leuCD$, generously provided by William Jacobs laboratory, Albert Einstein College of Medicine) or *M. tuberculosis* mc² 6206 $\Delta VapBC4$. VapC4 (also known as VapC-mt4, Rv0595c locus) was amplified by PCR from *M. tuberculosis* H37Rv genomic DNA. The amplified gene was cloned under the control of an anhydrotetracycline (ATC)-inducible promoter in the pMC1s plasmid. Induction of VapC4 was obtained by adding ATC to the media to a final concentration of 200 ng/ml, and replenishing it every 48 hours as necessary.

M. tuberculosis cells were grown under constant shaking at 200 rpm at 37° C in 7H9 Middlebrook media containing 1x OADC supplement (Sigma), 0.05% of tyloxapol, and kanamycin at 25 µg/ml for plasmid selection. The media was supplemented with 50 µg/ml of pantothenic acid and 100 µg/ml of leucine for growth of the attenuated strain *M. tuberculosis* mc² 6206.

Construction of vapBC4 deletion strain

A strain with the VapBC4 ($\Delta VapBC4$ or $\Delta Rv0596c-Rv0595c$) module deleted was constructed from the *M. tuberculosis* mc² 6206 parental strain using ORBIT recombineering method (Murphy et al., 2018). The Bxb1 attP-containing

oligonucleotide was designed to contain the first and last 60 base pairs of the Rv0596c-Rv0595c toxin-antitoxin module (NWO3093, 5'-GAT CGT GAC CGC AGA CAT CGA GAC GAG GCG GCG CGA CAG CAG CTC CGC TAC ACT CGT AGC GGT TTG TCT GGT CAA CCA CCG CGG TCT CAG TGG TGT ACG GTA CAA ACC TGC GAT CCT GAC CCA AGA CAA CGA CTA CGC CGC CAT GCC CGA CGT CGA GGT CAT AAC GAT CTG ACG GTT G-3'). To confirm successful deletion, we performed PCR using oligonucleotides targeting the module flanking regions (NWO3126, 5' - ACA AAT CAC GGC ACT TCG GC – 3' and NWO3127, 5' - GAG CGG CAA AGT CGT AGC AC – 3') in combination with internal pKM464 plasmid oligonucleotides (NWO3012, 5' - CAG GTA TCC GGT AAG CGG CA - 3' and NWO3013, 5' - CAC CGA TCC GGA GGA ACT GG - 3').

RNA isolation

Mycobacterial cells were grown in the absence or presence of inducer for 24 or 72 hours. Cells were centrifuged at 2000 × *g* for 10 min at 4 °C, and supernatants were removed. The cell pellets were resuspended in 1 ml of Trizol and transferred to lysing tubes (Bertin Corp.) containing 0.1 mm glass beads. Cells were lysed in 4 cycles of 30 s at 9000 rpm using Precellys Evolution homogenizer (Bertin Corp.) with 1 min cooldown periods on ice in between each cycle. The lysate was centrifuged at 14000 rpm at 4°C and the supernatant was used to extract total RNA using the Direct-zol™ RNA MiniPrep Plus kit (Zymo Research). RNA was treated with an additional step of genomic DNA removal in a TURBO

DNase (Thermo Fisher) digestion reaction for 30 min at 37°C, and re-purified using the Zymo RNA Clean and Concentrator kit. The extracted RNA was quantified by spectrophotometry using a μ Cuvette in a BioSpectrometer (Eppendorf).

In vitro cleavage assays with purified VapC4

VapC4 purification and cleavage reactions were performed as described in Sharp et al. and Cruz et al., respectively (Jonathan W. Cruz et al., 2015; Sharp et al., 2012). Briefly, ten picomoles of purified VapC4 was individually incubated with 2 pmol of all 45 in vitro synthesized tRNAs from the *M. tuberculosis* H37Rv genome or 2 μ g of total RNA extracted from *M. tuberculosis* mc² 6206 for 3 hours at 37° C. Cleavage reactions with in vitro synthesized tRNAs were run on a 15% Urea-PAGE gel, stained with SYBR Gold and visualized in a UV transilluminator. Cleavage reactions using total RNA were purified using the Zymo RNA Clean and Concentrator kit following the manufacturer's protocol and were used for constructing 5' RNA seq libraries.

5' RNA-seq

Preparation and analysis of 5'OH libraries was performed as described in Barth et al. (Barth & Woychik, 2020). For 5' monophosphate libraries, the same procedure was followed except for the steps of digestion with Terminator 5'-Phosphate-Dependent Exonuclease (Epicentre) and phosphorylation with T4 PNK

(New England Biolabs). The libraries were sequenced in an Illumina HiSeq 2500/4000 platform at Genewiz Corp or at New York University's Genome Technology Center.

For data analysis, we only considered reads that had at least 1 read per million of mapped reads (rpm) for mRNAs and 5 rpm for tRNAs in the induced sample and a fold change of at least 20. Frequency logos were generated with kpLogo (Wu & Bartel, 2017) or weblogo (Crooks, 2004).

Total RNA-seq

In order to remove 16S/23S ribosomal RNA from the total RNA, the samples were treated using the NEBNext bacterial rRNA depletion kit (New England Biolabs). Approximately 100 ng of rRNA-depleted RNA were used to generate the libraries using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs) and sequenced on an Illumina HiSeq 2500 or similar. Mapping of reads were done against *M. tuberculosis* H37Rv reference genome (Genbank accession: AL123456.3) in a local Galaxy instance (Afgan et al., 2018) using the default parameters of Bowtie 2.3.4.3 (Langmead & Salzberg, 2012), featureCounts 1.6.4 (Liao, Smyth, & Shi, 2014) and limma 3.38.3 (Ritchie et al., 2015). Significantly up- or down-regulated genes (with an adjusted p-value < 0.05) were analyzed using the Functional Annotation tool in the DAVID platform using an Benjamini p-value cut off of 0.1 (D. W. Huang et al., 2009).

Labeling of newly synthesized *M. tuberculosis* proteins

To assess the levels of newly synthesized proteins in *M. tuberculosis* after toxin induction, VapC4 was induced for 12, 24 and 48 hours and azidohomoalanine (AHA, Anaspec Inc.) was added to the media at a final concentration of 50 μ M for 6 h, in triplicate. AHA is an azide-containing methionine mimetic that is incorporated into proteins, allowing the capture or visualization of the newly synthesized proteins by a copper-catalyzed azide-alkyne cycloaddition reaction. To extract the AHA-labeled proteins, cells were pelleted, resuspended in lysis buffer (2% CHAPS, 8M Urea) and lysed in Precellys Evolution homogenizer as described in RNA isolation section. The lysate was centrifuged, and the proteins from the supernatant were linked to an alkyne-containing fluorophore (TAMRA) using the Click-IT Protein Reaction Buffer kit (ThermoFisher). Ten micrograms of protein from each sample were run on an SDS-PAGE gel and scanned using the Typhoon FLA 9500 (GE Healthcare) image system. Fluorescence intensities from each lane were compared using the ImageJ software.

Proteomics

For assessing newly synthesized proteins by quantitative mass spectrometry, *M. tuberculosis* Δ VapBC4 containing vapC4-pMC1s plasmid were grown to Optical Density at 600nm of 0.1 and induced for 48 hours along with uninduced cultures, in triplicate. Fifty milliliters of cultures were centrifuged at 2000 g at 4 °C for 5 minutes and washed with 1X PBS two times to remove traces of the albumin-

containing 7H9 media. The cell pellets were resuspended in lysis buffer (2% CHAPS, 8M Urea) and lysed using Precellys Evolution homogenizer as described in the RNA Extraction section. The lysates were pelleted at 12000 g at 4 °C for 10 minutes and the AHA-labeled proteins contained in the supernatant were selectively captured using alkyne-containing agarose beads from the Click-iT™ Protein Enrichment Kit (ThermoFisher), following the manufacturer protocol.

Tryptic digests were analyzed using a Q Exactive HF tandem mass spectrometer coupled to a Dionex Ultimate 3000 RLSCnano System (Thermo Scientific), as described in (Barth & Woychik, 2020).

The raw LC-MS data was converted into MASCOT Generic Format (MGF) using Proteome Discover 2.1 (ThermoFisher) and searched against either the NCBI *M. tuberculosis* database (Accession: AL123456) together with a database of common laboratory contaminants (<http://www.thegpm.org/crap/>) using a local implementation of the global proteome machine (GPM Fury) (Beavis, 2006).

Spectral counts were analyzed using the QuasiSeq package (<https://cran.r-project.org/web/packages/QuasiSeq/index.html>), with proteins containing 15 or more spectral counts total (Lund et al., 2012). Q-values are calculated using the fdrtool package of Strimmer (Strimmer, 2008) and considered significant if below 0.05.

Small protein identification

In order to validate the small proteins suggested by 5' OH RNA-seq, we searched for the estimated masses of their trypsin digestion products in publicly available *M. tuberculosis* proteomic datasets (retrieved from the PRIDE Archive, www.ebi.ac.uk/pride) and also in the datasets generated in this study. The public dataset accession numbers used for this purpose were: PXD003842, PXD004165, PXD005290, PXD006039, PXD006389, PXD008555, PXD009239, PXD010929, PXD011466, PXD012584 and PXD013677. Peak lists obtained from MS/MS spectra were identified using both X!Tandem version X! Tandem Vengeance (2015.12.15.2) (Craig & Beavis, 2004) and MS-GF+ (Kim & Pevzner, 2014). The search was conducted using SearchGUI version 3.3.19 (Vaudel, Barsnes, Berven, Sickmann, & Martens, 2011).

Protein identification was conducted against a concatenated target/decoy version of the Unknown complement of the *M. tuberculosis* H37Rv database (NCBI Genbank: AL123456.3) complemented with all Cys-containing peptides found in this study and list of common contaminants (<http://www.thegpm.org/crap/>). The decoy sequences were created by reversing the target sequences in SearchGUI. The identification settings were as follows: Trypsin, Specific, with a maximum of 2 missed cleavages, 10.0 ppm as MS1 and 20.0 ppm as MS2 tolerances; fixed modifications: Carbamidomethylation of C (+57.021464 Da), variable modifications: Acetylation of protein N-term (+42.010565 Da), Deamidation of N (+0.984016 Da), Deamidation of Q (+0.984016 Da), Dioxidation of M (+31.989829 Da), Oxidation of M (+15.994915 Da), Phosphorylation of S (+79.966331 Da), Phosphorylation of T (+79.966331

Da), Phosphorylation of Y (+79.966331 Da), variable modifications during refinement procedure: Pyrolydione from E (--18.010565 Da), Pyrolydione from Q (--17.026549 Da), Pyrolydione from carbamidomethylated C (--17.026549 Da).

Peptides and proteins were inferred from the spectrum identification results using PeptideShaker version 1.16.45 (Vaudel et al., 2015). Peptide Spectrum Matches (PSMs), peptides and proteins were validated at a 1.0% False Discovery Rate (FDR) estimated using the decoy hit distribution.

Results

VapC4 cleaves many *M. tuberculosis* tRNAs within their anticodon stem loop in vitro.

VapC4 expression in *M. tuberculosis* leads to growth arrest (Jonathan W. Cruz et al., 2015). To understand how this toxin downregulates cell growth, we first determined how many *M. tuberculosis* tRNAs were cleaved by VapC4 in vitro. Eight of the complete set of 45 tRNAs (Chan & Lowe, 2009, 2016) were cleaved to completion: tRNA^{Ala2}-UGC, tRNA^{Ala11}-CGC, tRNA^{Ala31}-GGC, tRNA^{Cys21}-GCA, tRNA^{Gly20}-GCC, tRNA^{Phe10}-GAA, tRNA^{Ser24}-GGA and tRNA^{Ser26}-GCU (Figure 37a,b *shaded black*). Surprisingly, five of these eight *M. tuberculosis* in vitro tRNA targets did not harbor an ACGC, ACUGC or ACAGC cleavage consensus sequence identified in earlier reports. In addition, 19 of the 45 tRNAs were partially cleaved (Figure 37a,b *shaded gray*). Therefore, 27 of the 45 tRNAs were fully or partially cleaved by VapC4. This data suggested that the growth inhibition

associated with VapC4 toxin expression might be due to depletion of multiple tRNAs.

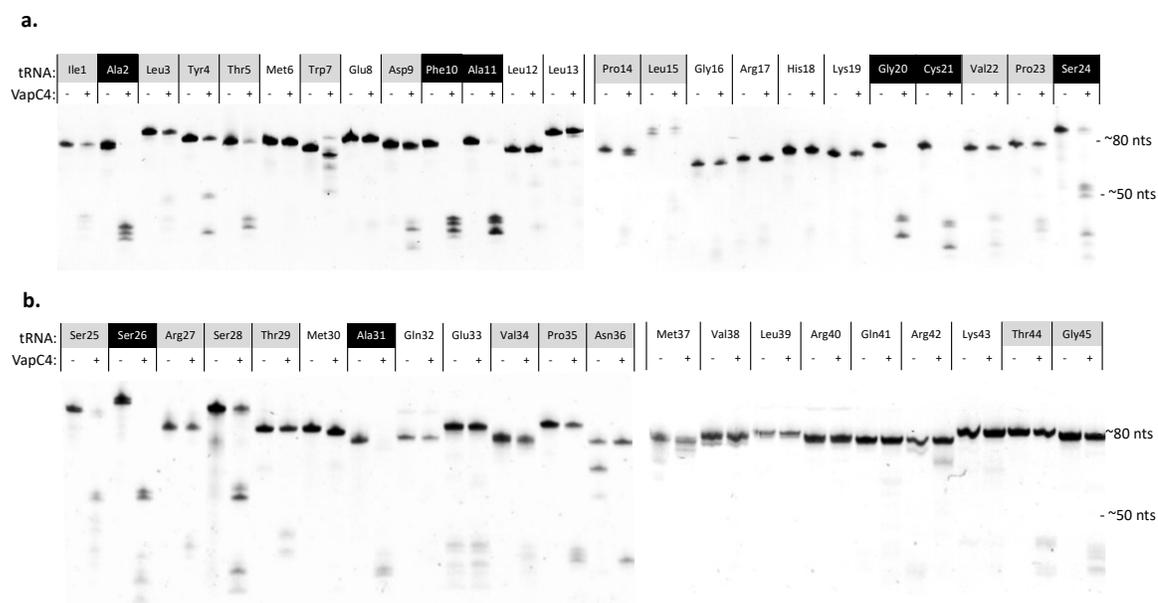


Figure 37. VapC4 promiscuously cleave several tRNAs in vitro. (a-b)

Cleavage assays with all 45 in vitro synthesized tRNAs (a, 1-24 and b, 25-45) from *M. tuberculosis* H37Rv. tRNAs were individually cleaved by VapC4 at 37 °C for 3 hours (the ratio of toxin to tRNA was 5:1). Controls (no VapC4) are run in parallel. Cleavage levels were classified in strong cleavage (black), weak cleavage (grey) and no cleavage (white).

VapC4 does not inhibit protein synthesis.

Since VapC4 appeared to disable so many tRNAs, we next tested if its role is to comprehensively shut down translation. In fact, others have concluded that tRNA-cleaving toxins generally act by globally inhibiting translation because their expression characteristically leads to bacterial cell growth arrest (K. S.

Winther & Gerdes, 2011; K. Winther et al., 2016). We used a click chemistry approach for metabolic labeling to monitor steady state protein synthesis in *M. tuberculosis* cells with and without toxin expression. Incorporation of the azide-containing Met mimetic azidohomoalanine (AHA) enabled fluorescent visualization of AHA-containing proteins upon coupling to the alkyne tetramethylrhodamine azide (TAMRA). Total proteins were resolved by SDS-PAGE and the newly synthesized population was visualized by fluorescent imaging (Figure 38a,b). Surprisingly, VapC4 did not completely inhibit protein synthesis even though its expression leads to growth arrest (Figure 38c).

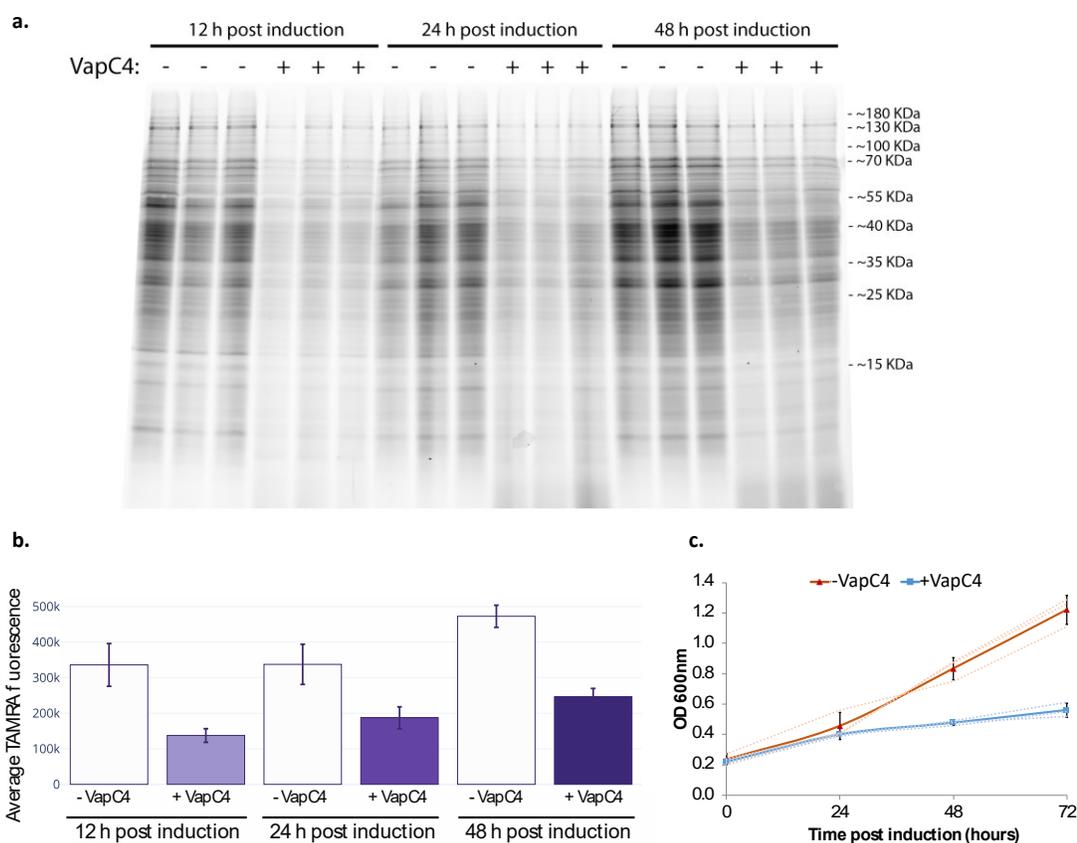


Figure 38. Cells expressing VapC4 sustained substantial translation despite growth inhibition. a, Newly synthesized (AHA-labeled) proteins from

induced and uninduced samples at 12, 24 and 48 hours post-induction were visualized by reacting with an alkyne-TAMRA conjugate. Each lane of a 9% SDS PAGE gel was normalized to contain 10 micrograms of total protein and visualized under the Typhoon fluorescence scanner. Approximate protein masses are indicated obtained from the Pre-stained PAGE-Ruler marker lane run on the same gel. **b**, Whole lane fluorescence was measured by ImageJ software selecting equal areas. The average fluorescence signals from triplicates are shown. Error bars correspond to standard error of mean. **c**, Growth curve obtained from *M. tuberculosis* H37Rv cultures with (blue) or without (orange) VapC4 inducer after 0, 24, 48 and 72 hours the induction time. Dotted lines are OD_{600nm} readings for each replicate and solid lines represent the average of the triplicates. Error bars represent the standard error of the mean.

VapC4 only targets tRNA^{Cys} in *M. tuberculosis* cells.

Because the effects on translation were not consistent with the large number of tRNAs cleaved by VapC4 in vitro, we next expressed this toxin in *M. tuberculosis* to determine if the tRNAs cleaved by VapC4 in vitro were also targeted in vivo. We enlisted a specialized RNA-seq approach developed in our laboratory, 5' RNA-seq, that enabled genome-wide VapC4 target identification as well as high resolution, single nt cleavage site mapping (Schifano et al., 2014). 5' RNA-seq captures and amplifies RNAs carrying the specific chemical moieties, a 5'-phosphate (5'-P) or 5'-hydroxyl (5'-OH), that are generated upon cleavage by VapC4 and other endoribonuclease toxins.

Total RNA was isolated from VapC4 induced and uninduced cells. We created 5'-P and 5'-OH 5' RNA-seq libraries and identified a single tRNA, tRNA^{CysGCA}, as the sole target of VapC4 (Figure 39a,b). This is the only Cys tRNA in *M. tuberculosis*. It services both the UGC and the UGU Cys codons. The site of toxin cleavage was within the Cys anticodon (Figure 39c), thus functionally disabling this tRNA for protein synthesis. This tRNA^{CysGCA} target is also consistent with that identified upon ectopic expression of VapC4 in *M. smegmatis* (K. Winther et al., 2016). Interestingly, in contrast to the MazF family of toxins, which generate a 5'-OH upon RNA cleavage, VapC4 instead appeared to generate a 5'-P upon RNA cleavage since there were no tRNA targets detected in the 5'-OH RNA-seq libraries. To prove that VapC4 directly generated a 5'-P upon RNA cleavage, instead of downstream phosphorylation of a 5' end that was initially hydroxylated, we incubated VapC4 with total *M. tuberculosis* RNA and again created 5'-P and 5'-OH 5' RNA-seq libraries. tRNA^{CysGCA} was cleaved in the 5'-P library (Figure 40a) and not the 5'-OH (Figure 40b). Based on the in vitro cleavage targets, we predict the true in vivo target has an absolute requirement for the sequence U G³⁴ C³⁵ A³⁶ A in the proper structural context to recognize and cleave only tRNA^{CysGCA} (Figure 40c). Therefore, VapC4, and likely all VapC family toxins, are distinct from other ribonuclease toxins in that they generate 5'-P ends upon RNA cleavage.

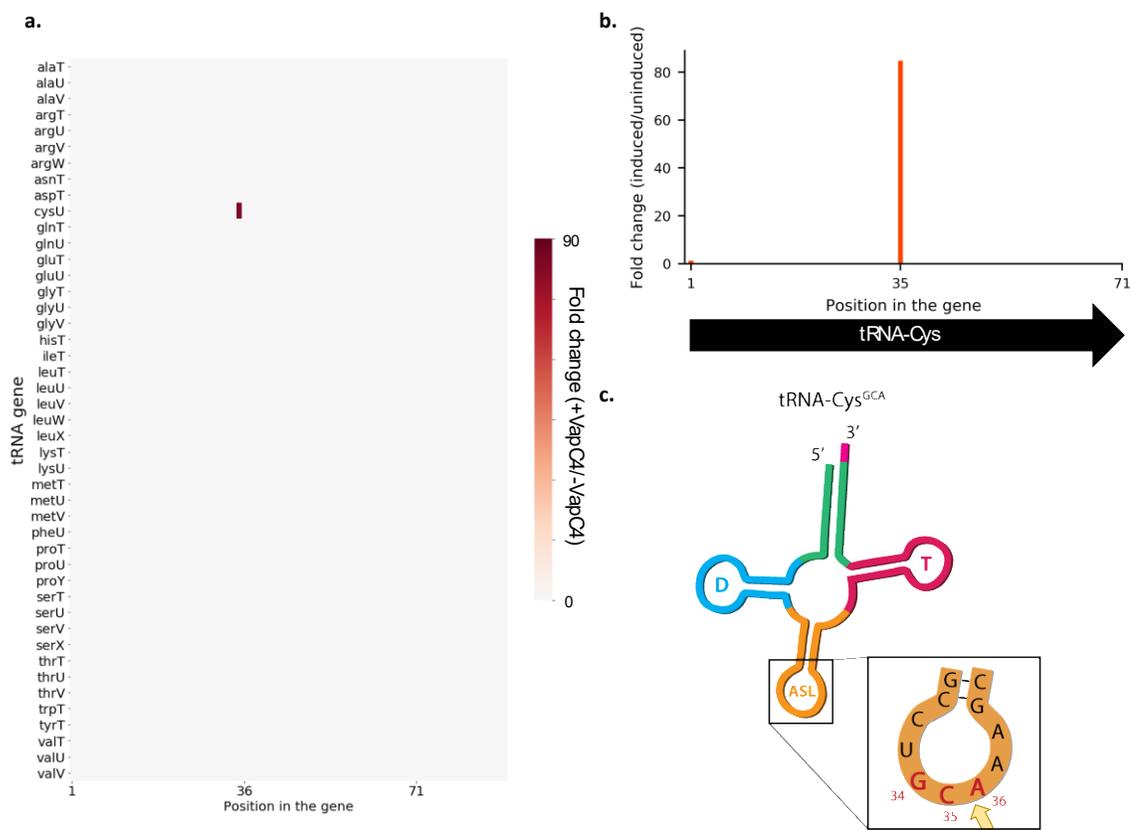


Figure 39. tRNA^{CysGCA} is the only VapC4 target in vivo. **a**, Heatmap obtained from 5'P RNA-seq libraries indicating internal cleavage (represented in fold change of internal 5' monophosphate ends in induced vs. uninduced) at each position for all 45 tRNA genes in *M. tuberculosis* mc² 6206, 24 hours post VapC4 induction. **b**, Bar graph showing the fold change of 5' monophosphate ends (induced vs. uninduced) in the only Cys tRNA gene after 24 hour of VapC4 induction. **c**, Representation of tRNA^{CysGCA} showing its main regions (D, D-loop; T, T Ψ C loop; ASL, anticodon stem loop). Cleavage site is indicated by a yellow arrow. Anticodon positions are numbered according to tRNA position standard guidelines.

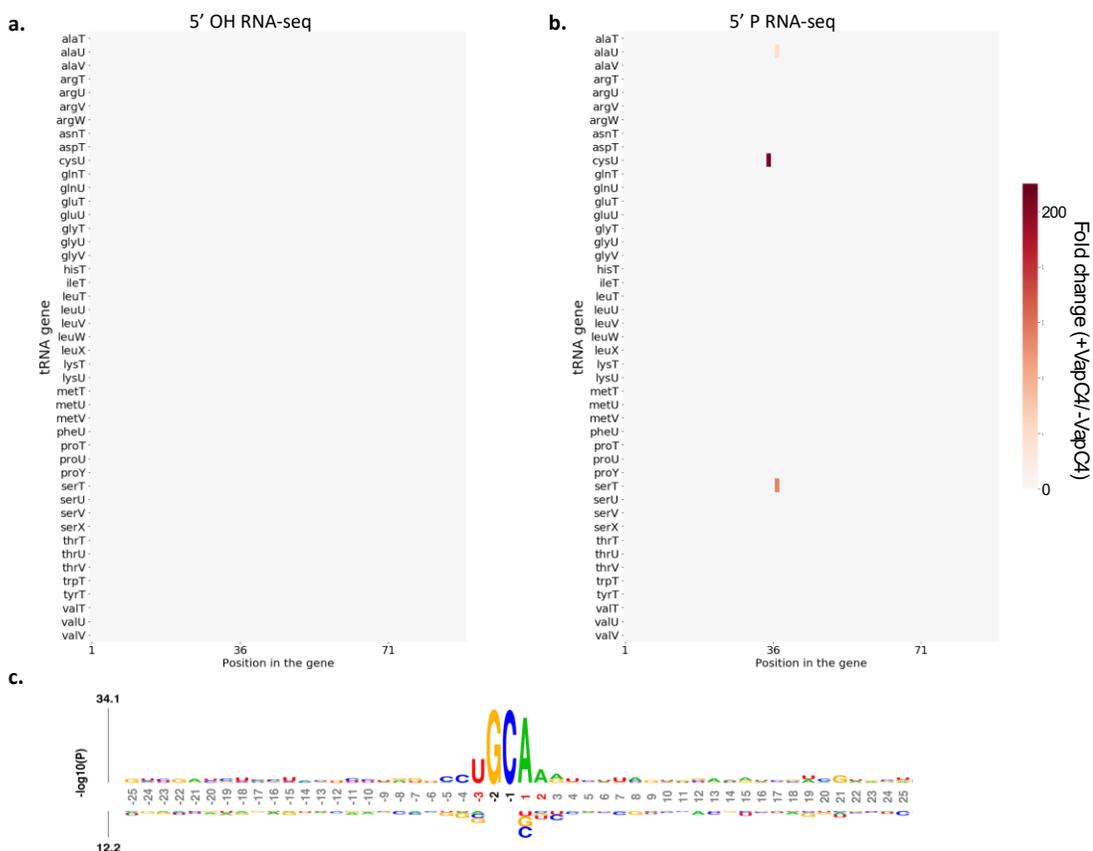


Figure 40. VapC4 cleavage generates 5' monophosphate ends and lacks specificity in in vitro assays. Heatmap showing fold changes obtained in 5' RNA-seq for internal 5'OH (a) or 5'P (b) ends in all 45 tRNA genes after incubating total RNA extracted from *M. tuberculosis* mc² 6206 with purified VapC4. **c**, Weblogo showing the consensus sequence observed in 100 RNA hits with highest fold change (induced vs. uninduced) in the in vitro cleavage 5'P libraries. The flanking region of the cleavage site (25 nucleotides up- and downstream) is shown. Positions are numbered relative to the cleavage site and colored in red if the nucleotide is statistically enriched at the position or in black if the nucleotide frequency is above 75%.

tRNA^{CysGCA} cleavage leads to ribosome stalling at Cys codons.

We recently reported that MazF-mt9 toxin-mediated cleavage of a single species of tRNA, tRNA^{LysUUU}, leads to selective ribosome stalling at the Lys codon (AAA) requiring that depleted tRNA (Barth et al., 2019). This highly specific, genome-wide ribosome stalling leads to preferential depletion of proteins containing the Lys AAA codon. Serendipitously, these ribosome stalling events were first suggested within the MazF-mt9 5'-OH RNA-seq dataset and then rigorously proven using Ribo-seq (Barth et al., 2019). In perfect agreement, we documented the same two trends for VapC4 following cleavage of tRNA^{CysGCA}.

As the first confirmed trend, within our VapC4 5'-OH RNA-seq datasets, we saw the hallmarks of ribosome stalling at several hundred *M. tuberculosis* transcripts. Each stalling event is readily detected because there is a conspicuous ~15 nt distance in this 5'-OH RNA-seq dataset to an in-frame “hungry” Cys codon requiring the depleted tRNA (Figure 41a,b). In Barth et al., we demonstrated that this 15 nt distance corresponds to the footprint of the stalled ribosome spanning its 5' edge to the position of the hungry codon at the A-site using Ribo-seq (Figure 41c) (Barth et al., 2019).

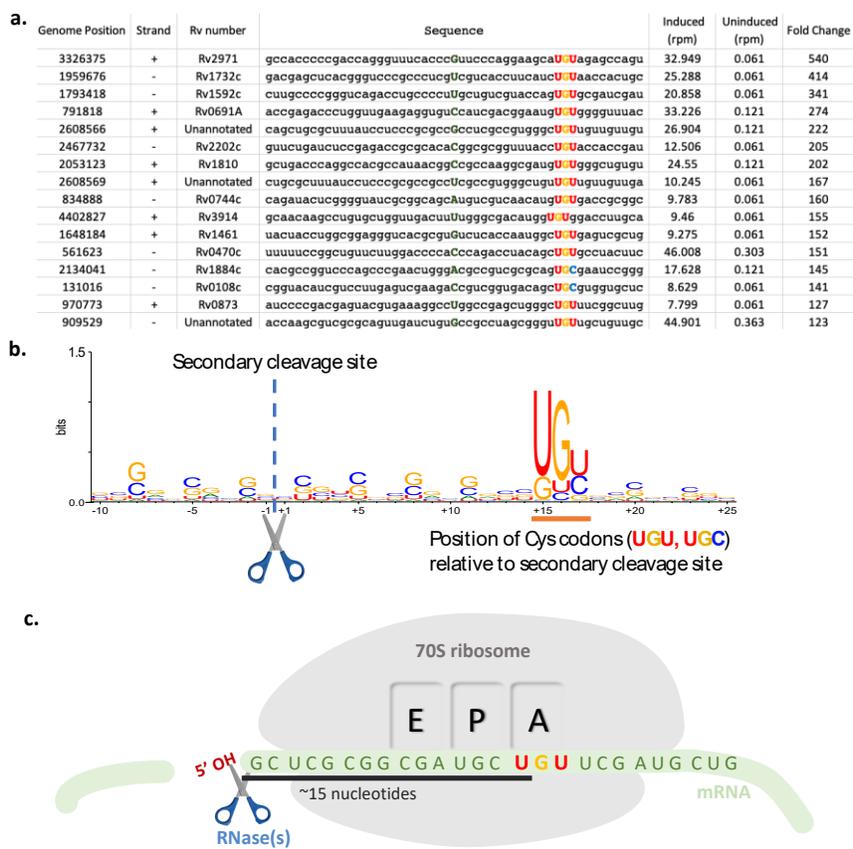


Figure 41. 5' OH RNA-seq reveals ribosome stalling at Cys codons with a single nucleotide resolution. a, mRNA top hits in 5' OH RNA-seq libraries constructed from *M. tuberculosis* mc² 6206 RNA extracted after 24 hours of VapC4 induction. Cysteine codons are highlighted, approximately 15 nucleotides upstream of the detected 5'OH. The genome position and strand where the secondary cleavage occurs is shown, as well as the Rv number of the gene containing the Cys codon. **b,** Weblogo showing the consensus sequence from the top 100 mRNA hits found by 5' OH RNA-seq. Positions are numbered relative to cleavage site (scissor). Cysteine codons are predominantly observed from +15 to +17 (orange underline). **c,** Schematic representation illustrating the ribosome stalling events following tRNA^{CysCGA} depletion by VapC4. Due to lack of

tRNA^{CysCGA}, ribosomes stall with Cys codons (UGU or UGC) at the A site. A secondary cleavage event on the unprotected 5' portion of the transcript is done by an unknown RNase (scissor), which leaves a 5'OH approximately 15 nucleotides downstream of the stalled codon.

We fortuitously detect these stalled ribosomes because some transcripts that harbor them are recycled by an RNase distinct from VapC4 that happens to leave a 5'OH upon cleavage. As the second trend consistent with our earlier study, we found that the proteome in VapC4-expressing cells contained fewer Cys-containing proteins relative to control cells (Figure 42). Therefore, VapC4 expression leads to inactivation of the primary target of the toxin, tRNA^{CysGCA}, followed by ribosome stalling at Cys codons because the only Cys tRNA in *M. tuberculosis* is now in deficit. These events extend to the proteome, resulting in an overall reduction in the steady state level of Cys-containing proteins.

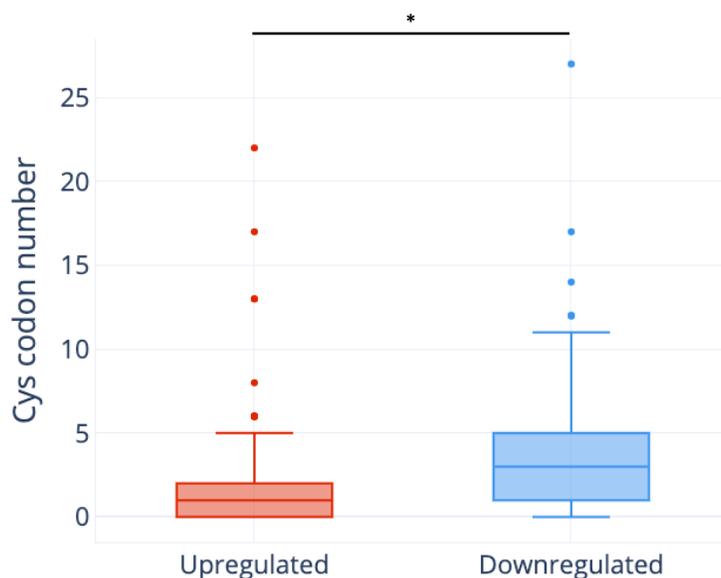


Figure 42. *M. tuberculosis* expressing VapC4 preferentially downregulates translation of Cys-rich proteins. Box plot summarizing Cys codon number in up- or down-regulated proteins identified by quantitative mass spectrometry after 48h of VapC4 induction. To exclude proteins synthesized before VapC4 induction, newly synthesized proteins were enriched by labeling with azidohomoalanine (AHA) for approximately 6 hours. Upregulated (n = 104, shown in red) or downregulated (n = 214, shown in blue) proteins were defined by presenting a fold change of at least +/- 1.5 and a q-value ≤ 0.05 . Predicted outliers are represented by individual dots and asterisk represents p-value < 0.0001 in a statistical comparison using Student's t test.

Ribosome stalling at Cys codons uncovered an abundance of unannotated small proteins.

Our ability to definitively map the presence, and precise position, of stalled

ribosomes on mRNAs within the VapC4 5'-OH RNA-seq dataset was exploited as a powerful tool to identify transcripts actively undergoing translation, and thus, whose expression is impacted by these stalling events. Consequently, we detected hundreds of stalled ribosomes on mRNAs within the VapC4 5'-OH RNA-seq dataset. The ribosomes were stalled at either of the two Cys codons, UGC and the UGU, serviced by the single Cys tRNA (tRNA^{CysGCA}) in *M. tuberculosis* (Figure 41b). We mapped the position of the stalled ribosomes relative to annotated *M. tuberculosis* H37Rv genes in NCBI and Mycobrowser and were surprised to discover that 96 transcripts harboring a stalled ribosome were from Cys codon-containing transcripts whose genes were not annotated. Many (75%) of these 96 transcripts encoded proteins less than 100 amino acids. Of note, 50% of the unannotated transcripts (48 of 96) were derived from short ORFs (sORFs) that encode transcripts ≤ 150 nts and proteins ≤ 50 amino acids (following convention for small proteins (Hemm, Weaver, & Storz, 2020) and similar to the <150 nt cutoff used for mycobacteria sORFs (Canestrari et al., 2020).

We mined 11 publicly available *M. tuberculosis* quantitative mass spectrometry datasets containing 102 raw data files for evidence that the unannotated Cys codon-containing proteins we uncovered are translated and stable. We identified one or more tryptic peptides mapping within 17% of the unannotated proteins (17 of the 96), suggesting that these proteins are translated and stable. Detection of tryptic peptides generated from 17% of the unannotated proteins is actually very encouraging given that half of these are small proteins of

≤50 amino acids. These small proteins are notoriously difficult to detect by mass spectrometry because they contain few, if any, trypsin cleavage sites and in general, they may not be abundant enough for detection (Hemm et al., 2020). Also, some of the very small Cys-rich proteins with regulatory roles discussed below might be intrinsically unstable. Below we highlight several exciting themes that came to light using conventional RNA-seq and/or proteomic approaches on RNA or protein derived from *M. tuberculosis* cells expressing the VapC4 toxin.

One sORF is the ortholog of Cys-responsive attenuator in *M. smegmatis*.

Although the majority of sORFs and unannotated proteins we identified have no known function, we were fortunate that some were accompanied with compelling functional clues. One sORF with an instructive recent history was leaderless (lacking a 5' UTR) and encoded a small protein of only 29 amino acids with an extremely Cys-rich sequence near its carboxy-terminus:

VSARIEPMLTKRRAVDLCRLAGCCCCCSC. Coincidentally, this small protein is the ortholog of the *M. smegmatis* sORF (Ms5788A) first described as one of the many unannotated small leaderless transcripts that are inexplicably common in mycobacteria (Figure 43a) (Shell et al., 2015). Moreover, recent follow-up studies on the function of Ms5788A support its role in a novel mode of Cys attenuation (Canestrari et al., 2020). Ms5788A is so named because it exerts translational control of the downstream three-gene operon beginning with the Ms5788 (unknown function) followed by the *M. smegmatis* CysA2 sulfotransferase and SseC putative sulfotransferase both of whose upregulation is logical under Cys

limiting conditions. When Cys availability is low, ribosomes are proposed to stall at the Cys codons of Ms5788A and preclude formation of a secondary structure that would otherwise block the availability of the Shine-Dalgarno (SD) ribosome binding site needed to recruit the ribosome and translate the downstream protein (illustrated in Figure 43b) (Canestrari et al., 2020). Distinct from the classic tryptophan attenuation in *Escherichia coli*, where the absence or presence of ribosome stalling at tryptophan codons in a small leader peptide dictates the formation of alternate secondary structures that either permit or prevent transcription termination, this *M. smegmatis* Cys-rich small protein acts as a sensor for Cys concentration and dictates whether transcripts that follow are translated.

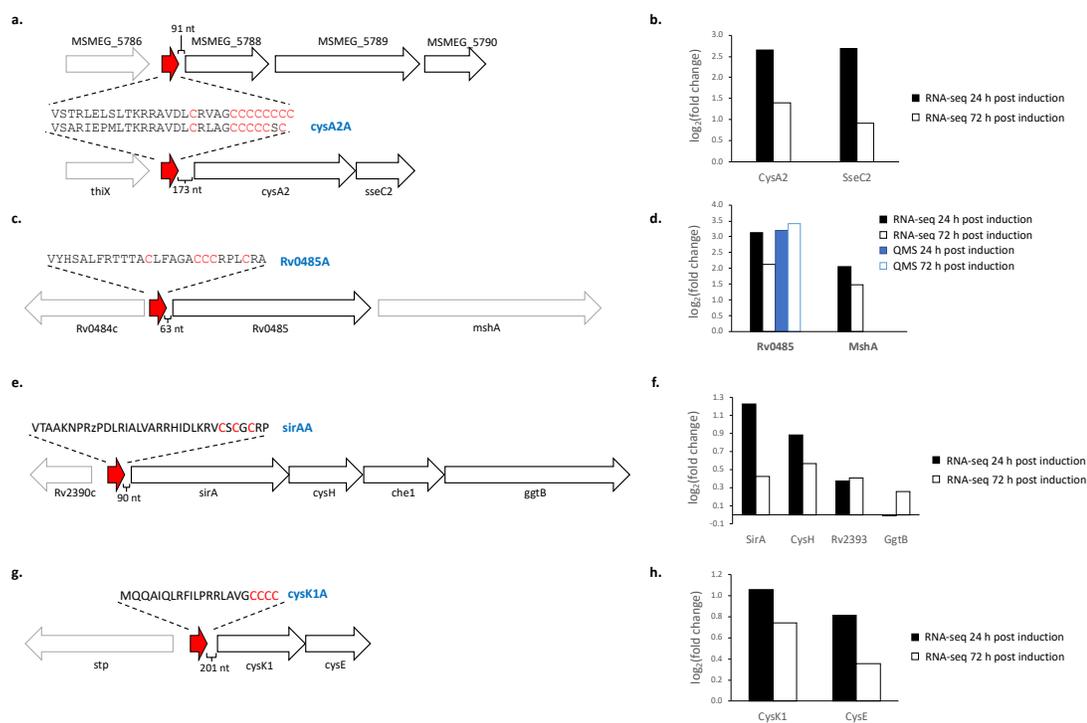


Figure 43. Downstream genes of identified Cys-containing sORFs are generally upregulated. Genomic organization of the region surrounding the

Cys-containing sORFs that occur upstream of annotated genes (**a, c, e, g**) and their respective transcriptional and/or translational changes, when detected by Quantitative Mass Spectrometry (QMS), in *M. tuberculosis* cells expressing VapC4 (**b, d, f, h**). Unannotated putative sORFs (red arrow) and their amino acid sequence (Cys residues highlighted in red) are shown. At least one of the identified Cys-containing sORFs has a known homolog in *M. smegmatis* (**a**).

Since the function of the Ms5788A sORF in *M. smegmatis* is now well described, its leaderless *M. tuberculosis* counterpart residing in front of an operon with similar architecture is predicted to also serve as a Cys-responsive attenuator. We designated this *M. tuberculosis* sORF as *cysA2A* following the convention of Canestrari et al. to add an A after the name of the gene it precedes since it resides 173 nt in front of the *cysA2* (Rv0815c) and *sseC* genes (Figure 43a). We also already knew that ribosomes stall within *cysA2A* because that was the basis of its detection within our 5' RNA-seq dataset. Therefore, stalling at its Cys-rich carboxy terminus would be expected to alter the secondary structure and thus influence expression of *cysA2* and *sseC*. We performed conventional RNA-seq coupled with quantitative mass spectrometry to determine how expression of VapC4 influenced the *M. tuberculosis* transcriptome and proteome.

We expected upregulation of the *cysA2* and *sseC*, consistent with the luciferase reporter experiments for *M. smegmatis* Ms5788A (Canestrari et al., 2020). Indeed, RNA-seq performed on RNA harvested 24 and 72 hrs after VapC4 induction was in alignment with this model, showing significant log₂-fold increases in both *cysA2* and *sseC* transcripts relative to the uninduced controls

(Figure 6b). Although quantitative mass spectrometry did detect increases in spectral counts for both proteins upon VapC4 expression, the sensitivity of this method is much lower than RNA-seq. Thus, the relatively low number of spectral counts for each protein were not supported by statistically significant q values. Although the detection limit of quantitative mass spectrometry is better than ever, identification of individual proteins in a total cell extract is dependent on their relative concentration. Since *cysA2* and *sseC* are both specialized enzymes, they likely function at relatively low concentrations. However, the significantly high levels of both *cysA2* and *sseC* transcripts in RNA-seq datasets support the model predicted for *M. smegmatis* Ms5788A because it is well established that active translation leads to increased mRNA stability because ribosome occupancy prevents RNA degradation/recycling by ribonucleases (Radhakrishnan & Green, 2016).

VapC4 causes ribosome stalling at more leaderless sORFs within other Cys attenuation-responsive regulons.

We documented VapC4-dependent ribosome stalling on three more unannotated *M. tuberculosis* leaderless sORF mRNAs that are also implicated as Cys-responsive attenuators. Two were previously recognized as orthologs of *M. smegmatis* leaderless sORFs encoding small proteins with the telltale Cys-rich tract (Shell et al., 2015) (Figure 43c,e). In agreement with a role in attenuation, the abundance of both of these *M. smegmatis* small proteins increased under low Cys conditions (Shell et al., 2015). The third sORF was identified in *M.*

tuberculosis (Shell et al., 2015) based on its distinctive Cys-rich carboxy-terminus MQQAIQLRFILPRRLAVGCCCC and characteristic genome position (sORF starts 270 bps upstream of the CysK1 cysteine synthase and CysE serine acetyltransferase genes) (Figure 43g). In support of their role as *M. tuberculosis* Cys-responsive attenuators, our H37Rv RNA-seq data 24 and 72 hrs after VapC4 induction showed significant increases in the abundance of all transcripts immediately downstream of each of the three leaderless mRNAs, consistent with active translation of these transcripts (Figure 43d,f,h). We observed a polar effect on the polycistronic sirA-cysH-che1-ggtB operon (Figure 43h,i), i.e. a decrease in translation as the ribosomes traverse the polycistronic mRNA, in alignment with the findings reported by Canestrari et al. (Canestrari et al., 2020) (Figure 43e,f).

One of the four examples of Cys-responsive attenuator regulation in Figure 43 was validated by both RNA-seq and mass spectrometry, revealing striking increases in the levels of the Rv0485 transcript and newly synthesized protein (Figure 43c,d). In fact, the leaderless mRNA regulating Rv0485 appears to be naturally abundant as well; it consistently appeared at or near the top of the list (by fold change VapC4 induced/uninduced) of mRNAs harboring stalled ribosomes at Cys codons in our 5'-OH RNA-seq datasets.

Other unannotated ORFs encode stable Cys-containing proteins that can also map immediately before, within or opposite annotated genes.

Finally, when VapC4 was expressed, we also identified stalled ribosomes on leaderless mRNAs (Shell et al., 2015) that do not exhibit the same genome

architecture as Cys attenuation-responsive regulons shown in Figure 43. Instead, some leaderless unannotated Cys-containing ORFs reside immediately upstream of another ORF, with no intervening region, as illustrated in Figure 44a. Another novel leaderless ORF overlaps with two other annotated genes on the opposite strand (Figure 44b).

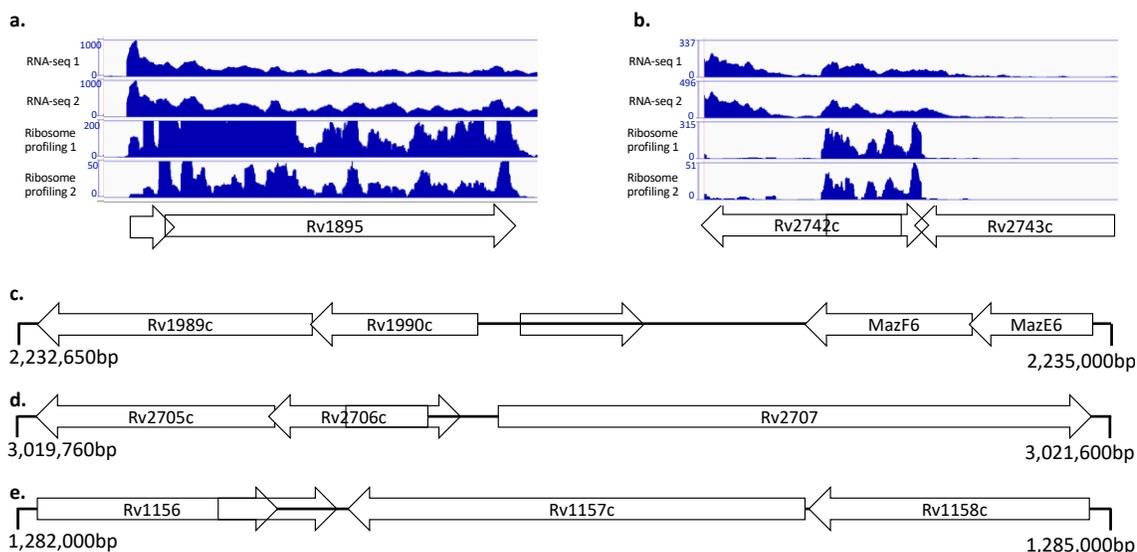


Figure 44. Unannotated sORFs can occur in a diverse gene organization

context. RNA-seq and Ribosome profiling coverage data (retrieved from

<http://mtb.wadsworth.org>) support the transcription and translation of

unannotated sORFs (red arrows) identified by 5' OH RNA-seq occurring

upstream (**a**) or in the opposite direction (**b**) of annotated genes. Other examples

of sORFs were found without any overlap (**c**) or overlapping annotated genes in

the reverse (**d**) or the same (**e**) direction. Genomic coordinates of the illustrated

regions in the *M. tuberculosis* H37Rv genome are shown based on the GenBank

accession AL123456.

We could not clearly establish if many of the other unannotated ORFs were leadered or leaderless based on the data from Shell et al. (Shell et al., 2015). Nevertheless, Figure 44 illustrates three examples of typical mapped locations of unannotated ORFs: upstream of a gene in the opposite orientation (Figure 44c), overlapping a gene in the opposite orientation (Figure 44d) and overlapping a gene in the same orientation (Figure 44e).

The Cys attenuators in Figure 43 appear to have regulatory roles similar to the 14 amino acid TrpL leader peptide that is unstable and thought to be nonfunctional in the cell beyond its role in facilitating Trp operon attenuation (Merino, Jensen, & Yanofsky, 2008). Therefore, while the Cys attenuator proteins may be intrinsically unstable as well, the stability (and legitimacy) of each of the five representative unannotated proteins in Figure 44 is confirmed by high confidence identification of multiple tryptic peptides from published *M. tuberculosis* quantitative mass spectrometry datasets.

Four unannotated sORFs encode novel EsxB-like proteins.

We NCBI-BLASTP searched all unannotated proteins that harbored stalled ribosomes at Cys codons to determine if their sequences revealed functional clues. We identified four proteins with significant similarity to a 156 amino acid “ESAT-6-like” protein sequenced from an *M. tuberculosis* strain originating from Peru. Alignments to investigate sequence similarity of these four proteins to EsxA (formerly ESAT-6) vs. EsxB (formerly CFP-10) established that these proteins are actually orthologs of EsxB (Figure 45a). However, each of the four

unannotated proteins have distinctly different amino terminal sequences whose legitimacy was supported by the presence of a stalled ribosome at a Cys codon within these variable amino terminal sequences. We also identified a tryptic peptide confirmed by published mass spectrometry datasets within the variable region (Figure 45a). All four proteins contain the hallmark Y-X₃-D/E motif and an appropriately aligned possible counterpart (i.e. W-G) of the W-X-G motif of EsxB (Figure 45a). The Y-X₃-D/E sequence is required for secretion of the EsxA/EsxB complex by type VII secretion systems; the W-X-G motif is thought to also be part of the secretion signal (Ates, Houben, & Bitter, 2016; Daleke et al., 2012).

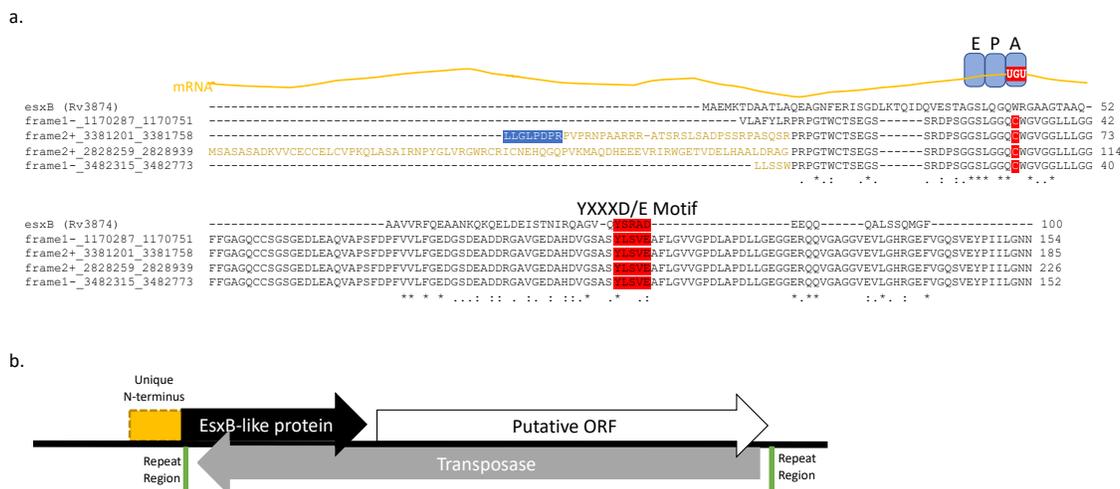


Figure 45. Putative sORFs identified overlapping transposases are homologues of EsxB/CFP-10. a) A multi-sequence alignment of four Cys-containing unannotated small ORFs against EsxB (Rv3874) showing identities (*), strong similarities (:), and weak similarities (.) was obtained using Clustal Omega. The identified ORFs present a possible variable N-terminus (yellow) and overlap the transposases annotated as Rv2512c, Rv1047, Rv3023c and Rv3114.

The Cys codons where ribosome stalling was observed by 5' OH RNA-seq are highlighted in red. At least one confident peptide was found by quantitative mass spectrometry for ORF “frame2+_3381201_3381758” (sequence highlighted in blue). **b)** Representative scheme of the region surrounding the identified putative small EsxB-like ORFs overlapping the transposase genes on the opposite strand. A putative ORF can be found downstream of the EsxB-like protein.

Curiously, all four EsxB-like genes overlap with a portion of distinct, but nearly identical, transposase genes (Figure 45b). The EsxA-EsxB heterodimer is a major virulence factor that resides within the 15-gene ESX-1 type VII secretion system locus. ESX-1 is essential for *M. tuberculosis* evasion of the host immune response (Gröschel et al., 2016). Since EsxB is followed by EsxA in the ESX-1 locus, we examined each EsxB-like gene for a downstream ORF. Although we identified a putative ORF with some similarity to EsxA, these ORFs were more than twice the size of EsxA and were missing the W-X-G motif. Therefore, they do not appear to be clear functional orthologs of EsxA (Figure 45b). Because these four EsxB-like proteins do not have an apparent binding partner, they may function on their own. Although EsxA and the EsxA-EsxB complex are most commonly the focus of functional studies and EsxB is often referred to as simply a chaperone, one published report uncovered a novel biological activity for EsxB alone: recruitment and activation of human neutrophils (Welin et al., 2015). If these EsxB-like proteins function in a similar manner, then ribosome stalling upon VapC4 expression is expected to block their translation and thus, dampen neutrophil activation through this pathway.

VapC4 selectively upregulates sulfur metabolism and assimilation, oxidative stress, copper detoxification and antiviral defense pathways.

In addition to the functional insights obtained through identification of nearly 100 unannotated ORFs described above, we used the annotation tool “Database for Annotation, Visualization and Integrated Discovery” (DAVID) (D. W. Huang et al., 2009; Huang da et al., 2009) to obtain a comprehensive summary of biological themes from RNA-seq datasets derived from total RNA harvested 24 hours after VapC4 induction in *M. tuberculosis* H37Rv. These biological themes were consistent when DAVID was performed on RNA-seq datasets from different induction times (24 and 72 hrs) in H27Rv. Notably, DAVID analysis also integrates the global effects imparted by the 156 annotated mRNAs for which we identified stalled ribosomes at one or more Cys codons.

Among downregulated gene categories, some were consistent with the growth arrested state of VapC4-expressing cells, i.e. “growth of symbiont in host” (Figure 46a). Gene groups involved in ATP generation—“quinone/quinone binding/ubiquinone”, “oxidative phosphorylation/electron transport”—were well represented. The “cell wall/cell membrane” gene groups are downregulated as well as “two component systems” and “ABC transporters”. Finally, the keyword/GO term “phosphopantetheine” was represented. Phosphopantetheine is an intermediate in the synthesis of the cofactor coenzyme A (CoA) from pantothenate (vitamin B₅). Since CoA is indispensable for several essential biosynthetic pathways, limiting its synthesis certainly affects cell growth. In fact,

deletion of two genes that participate in CoA synthesis, $\Delta panCD$, inhibits *M. tuberculosis* growth and represents an attenuated vaccine candidate (Sambandamurthy et al., 2002). Downregulation of “phosphate transport” is more provocative as phosphate depletion triggers *M. tuberculosis* persistence (Rifat, Bishai, & Karakousis, 2009).

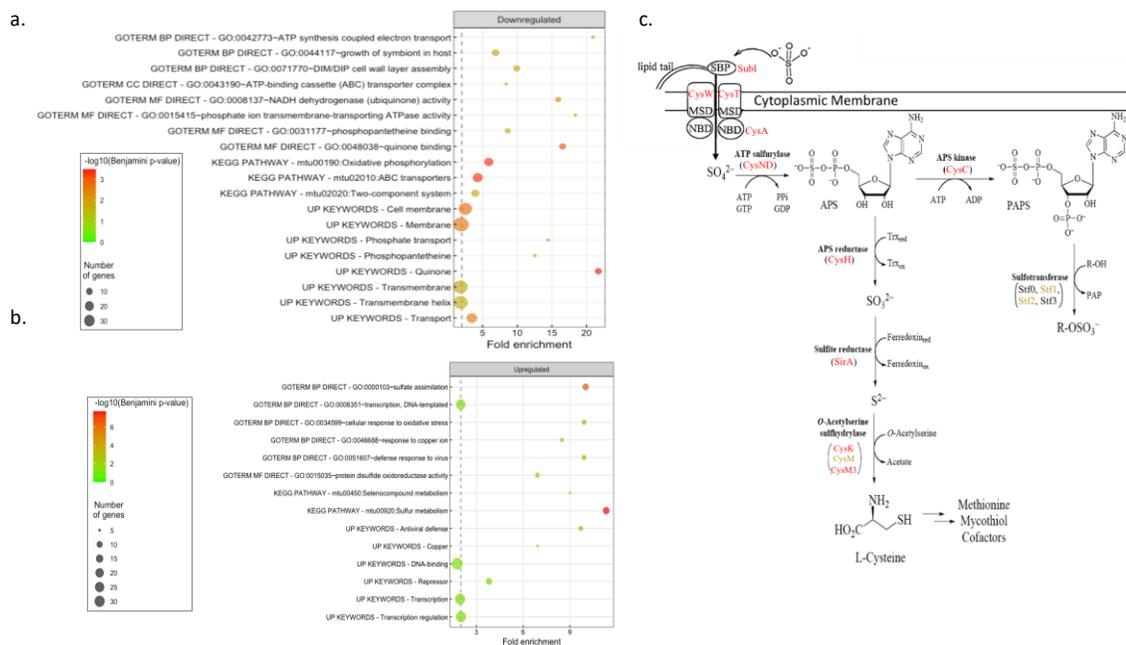


Figure 46. Genes involved in sulfate assimilation, namely part of the cysteine biosynthesis pathway, are upregulated in *M. tuberculosis* cells expressing VapC4. Significant DAVID Functional Analysis Tool terms (<http://david.ncifcrf.gov>) associated to the genes that were significantly upregulated (a) or downregulated (b) after 24 hours of VapC4 induction. Cysteine biosynthesis pathway (c) highlighting significant upregulated transcripts at 24 h (red) and at 72 h (yellow). Adapted from Devayani P. Bhavne, Wilson B. Muse III, & Kate S. Carroll, 2008; Zeng, Shi, Zhao, & Xie, 2013.

Biological themes among upregulated transcripts (Figure 46b) were especially

informative. The “sulfate assimilation/sulfur metabolism” categories had the highest statistical significance and are consistent with a toxin whose primary target is tRNA^{Cys}. In fact, all of the enzymes in the *M. tuberculosis* sulfur assimilation pathway were upregulated (Figure 46c). Overall, 16 of the 19 the “Cys”-assigned genes in Mycobrowser were upregulated. Other themes include groups in related categories: “oxidative stress/disulfide oxidoreductase”, “antiviral defense/defense response to virus” and “copper/response to copper ion”.

Upregulation of the oxidative stress/disulfide oxidoreductase pathways is also consistent with an increase in sulfur metabolism shown in Figure 46c. The major end product, Cys, also supports the production of mycothiol, a small thiol molecule with a cysteine-ligated disaccharide that serves as the *M. tuberculosis* glutathione counterpart responsible maintaining an intracellular reducing environment and minimizing the damaging effects of hydrogen peroxin and NO (Jothivasan & Hamilton, 2008). Independent of DAVID analysis, we also found significant upregulation 24 and/or 72 hrs post VapC4 induction of all 11 well vetted genes confirmed by multiple experimental approaches in the alternate sigma factor SigH regulon (Rv0350, dnaK; Rv1471, trxB1; Rv1528c, papA4; Rv1875; RV2466c; Rv2707; Rv2710, sigB; Rv3206c, moeB1; Rv3223c, sigH; Rv3913, trxB2) (Sharp et al., 2016). SigH regulates a network of genes that control the oxidative and heat stress responses (Raman et al., 2001; Sharp et al., 2016).

The “transcription/transcription regulation/DNA-binding/repressor” themes are very broad and likely represent the spectrum of regulatory proteins that work in

concert to favor activation of the predominant DAVID biological pathways over the many other pathways in the cell. Of note, some proteins within this transcription regulation-oriented theme bridge with the oxidative theme. Four of the seven redox-sensing WhiB transcription factors (WhiB1,3,4,6) were upregulated following VapC4 induction, most markedly WhiB6 (3.2 log₂ fold change) and WhiB1 (1.9 log₂ fold change). The seven *M. tuberculosis* WhiB family members are redox-sensing, Fe-coordinating, transcription factors whose ability to bind DNA is dependent on oxidation state (Saini, Farhana, Glasgow, & Steyn, 2012). Their synthesis is dependent on the availability of Cys through the sulfur assimilation pathway because they contain four conserved Cys residues responsible for coordinating Fe, [4Fe-4S]. Although WhiB1 is relatively insensitive to O₂, it is extremely reactive to NO, resulting in cluster nitrosylation and its conversion from a non-DNA-binding form to a DNA-binding transcription repressor (Smith, Stapleton, Buxton, & Green, 2012; Smith et al., 2010). WhiB3 (A. Singh et al., 2007) and WhiB4 (Chawla et al., 2012; Larsson et al., 2012) are both sensitive to both O₂ and NO; WhiB4 increases resistance to oxidative stress in vitro and enhances survival in macrophages (Chawla et al., 2012). Unfortunately, the function of *M. tuberculosis* WhiB6, which showed the highest upregulation in response to VapC4, has not been well characterized. When exposed to NO, *M. marinum* WhiB6 differentially regulates expression of the ESX-1 regulon and DevR/DosR dormancy regulon in zebrafish (Chen et al., 2016). There are also three more [4Fe-4S] proteins upregulated that are not in the WhiB family: Rv2733c (unknown function), Rv2392 (CysH) and Rv1594

(quinolinate synthetase NadA) (Saini et al., 2012).

Finally, within the theme of copper/response to copper ion, the copper oxidase MmcO/Rv0846c exemplifies the importance of copper detoxification for *M. tuberculosis* survival (Rowland & Niederweis, 2013). We documented high MmcO transcript and protein levels after VapC induction (Figure 47a, *far left bar graph*). Copper detoxification is critical for evasion of macrophage-mediated killing. Macrophages increase their concentration of copper upon *M. tuberculosis* infection. As a counterattack, the MmcO copper oxidase inactivates the copper, enabling *M. tuberculosis* survival and effectively imparting copper resistance. In fact, copper resistance is required for full *M. tuberculosis* virulence (Wolschendorf et al., 2011). More recent studies suggest that MmcO also scavenges reactive oxygen species in THP-1 cells (Kinkar, Kinkar, & Saleh, 2019).

Curiously, the Tuberculosis Database TBDB revealed that the mmcO/Rv0846c gene happens to reside within an apparent divergently transcribed regulon composed of genes encoding other highly upregulated transcripts, with some also present in companion mass spectrometry datasets (Figure 47a,b). This configuration is feasible as coregulation of divergent transcripts/operons has recently been documented in bacteria (Warman, Forrest, Wade, & Grainger, 2020). mmcO/Rv0846c opposes and shares its intergenic region with the four gene LpqS-CysK2-Rv0849-Rv0850 operon. CysK2 catalyzes the synthesis of S-sulfocysteine and also uses sulfide as donor to produce L-cysteine as in Figure 9c. The other genes in the four-gene operon are not well

characterized. Interestingly, three of the transcripts in this divergent regulon—*mmcO/Rv0846c*, *lpqS* and *CysK2*—were consistently upregulated during the Wayne model of nonreplicating persistence in independent microarray experiments (Muttucumaru, Roberts, Hinds, Stabler, & Parish, 2004; Voskuil, 2004; Voskuil, Visconti, & Schoolnik, 2004). In summary, *VapC4* expression triggers a multi-pronged counterattack to evade the host innate immune response and establish a latent infection to insure its long-term survival.

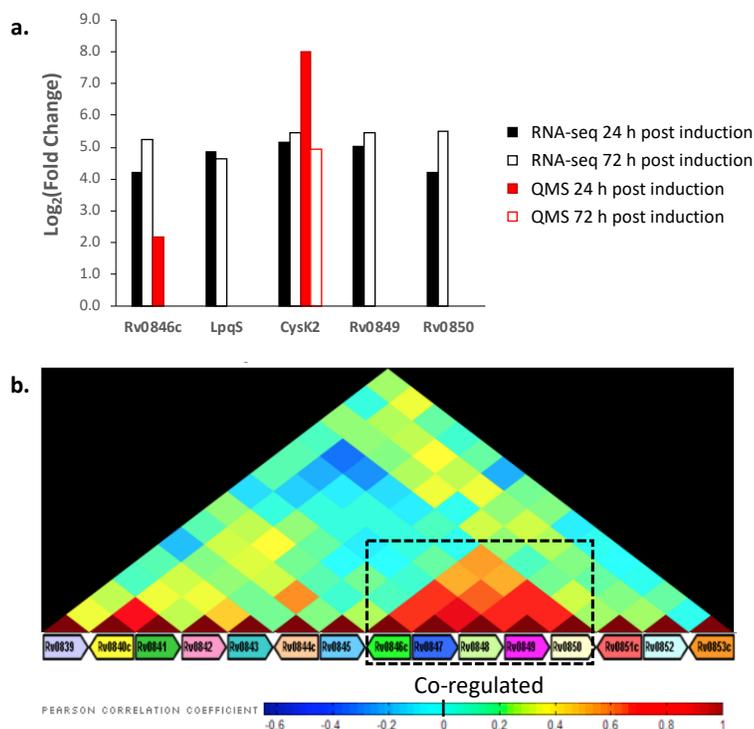


Figure 47. A putative regulon formed by *Rv0846c*, *Rv0847* (*lpqS*), *Rv0848* (*cysK2*), *Rv0849*, *Rv0850* is active upon expression of *VapC4*. Bar graph showing differential expression of genes within the putative regulon as seen in RNA-seq and proteomics data after 24 and 72 hours post *VapC4* induction. Only showing proteomics data for proteins detected by quantitative mass spectrometry (QMS) with a minimum of 15 spectral counts. Tuberculosis Database (TBDB

found at http://genome.tdb.org/tbdb_sysbio/OperonBrowser.html) shows strong correlation in the expression of these genes **(b)**.

Discussion

The abundant *M. tuberculosis* TA systems have long been implicated in the establishment of latent tuberculosis infection because expression of their toxin components in *E. coli* typically leads to a reversible state of growth arrest and a “quasidormant” metabolic state (Harms et al., 2018; Masuda & Inouye, 2017). Early studies of TA toxins in *E. coli*—most of which are endoribonucleases—almost exclusively identified mRNA as the toxin target (Masuda & Inouye, 2017; Y Yamaguchi & Inouye, 2011; Yoshihiro Yamaguchi et al., 2011). Therefore, for many years the prevailing view was that all endoribonuclease TA toxins in bacteria were sequence-specific “mRNA interferases” that imparted reversible growth arrest through widespread mRNA degradation that then leads to inhibition of protein synthesis. However, once our laboratory and others began to characterize the targets of TA toxins in *M. tuberculosis*, it was clear that this pathogen was quite different from *E. coli* and the gram-positive *Bacillus subtilis*. Not only does *M. tuberculosis* employ endoribonuclease toxins for inactivation of 23S rRNA at helix loop 70 (Schifano et al., 2013, 2014; Schifano & Woychik, 2014) or the highly conserved sarin-ricin loop (K. S. Winther et al., 2013), it engages part of its artillery of over 90 TA systems for highly selective isoacceptor-specific tRNA cleavage (Barth et al., 2019; Cintrón et al., 2019; Jonathan W. Cruz et al., 2015; Schifano et al., 2016).

In our recent study, we demonstrated that targeted inactivation of a single tRNA^{LysUUU} species surgically manipulates the transcriptome based on codon-usage, representing a novel mode of action for tRNA-cleaving toxins that is thought to enable reprogramming of *M. tuberculosis* physiology (Barth et al., 2019). With over 90 *M. tuberculosis* TA systems thought to be initiated by stress, and a finite number of infection-associated stresses (e.g. hypoxia, oxidative stress, nutrient limitation, nitrosative stress, low pH), it is likely that more than one toxin is induced in response to infection. In this study, we implemented low level expression of VapC4 to recapitulate the not yet identified natural in vivo trigger(s) of toxin activation, thus enabling us to isolate and distinguish the effect of this toxin on *M. tuberculosis* physiology from others.

In this study we used a spectrum of powerful systems-level approaches to pinpoint the molecular events following expression of a single toxin, VapC4. We discovered a highly targeted, surprisingly ingenious *M. tuberculosis* response based on the selection of tRNA^{Cys} as the primary toxin target. This toxin engaged tRNA^{Cys} inactivation to mimic a state of Cys starvation so that the cell can retool its proteome for the defense against attacks from macrophages with toxic copper and reactive oxygen and nitrogen species using a novel series of paths that include sORFs acting as Cys-attenuators. By systematically deconstructing the proteome based on Cys codon ribosome stalling and refashioning it towards the synthesis of L-Cys—the essential building block of multiple key protectors against reactive oxygen and nitrogen species such as mycothiol and Fe-sulfur enzymes—this pathogen has adeptly adapted to survive in the harsh

macrophage environment and long term in the lung as a granuloma.

The effects we see with VapC4 are in sync with what is known about the importance of sulfur in this pathogen. Bacteria differ from mammalian cells in that they assimilate inorganic sulfur into cysteine, which then serves as a precursor to a variety of key metabolites to maintain protein synthesis and a redox balance (reviewed in (Hatzios & Bertozzi, 2011; Schelle & Bertozzi, 2006)). Cysteine biosynthesis is critical for *M. tuberculosis* virulence and survival. Inhibition of cysteine biosynthesis impairs the ability of this pathogen to infect the host, fight oxidative stress and to establish latent infections.

Finally, because we demonstrate that this response appears to be imparted by codon-dependent proteome remodeling, it is likely that a full blown, comprehensive counterattack will enlist the coordinated activation of multiple tRNA-cleaving toxins that each target a single tRNA isoacceptor to each surgically remodel the proteome and collectively attain protection. We speculate that each toxin engages unique arms of the multifaceted *M. tuberculosis* stress response endured upon infection. These arms of defense differ because each is restricted to the codon-dependence dictated by the toxin tRNA target. If so, a coordinated defense system engaging an armamentarium of tRNA-cleaving toxins would exemplify a sophisticated, unrivaled multitier evolutionary adaptation for survival upon infection and within the harsh environment of the granuloma.

Conclusions

Until recently, none of the *M. tuberculosis* ribonucleolytic toxins had been functionally studied in vivo in its original host. Most studies performed their experiments using the model organism *M. smegmatis* or in vitro assays. We showed that this practice, although customary in the field, introduce artifacts that generate results that may not be transposable to the physiological conditions of the original host. Here, we studied for the first time the targets of one MazF (MazF-mt9) and two VapCs (VapC-mt4 and VapC-mt11) when the toxins were expressed in *M. tuberculosis*. All of them showed tRNAse activity in vivo (particularly against tRNA^{Lys}, tRNA^{Cys}, tRNA^{Gln} and tRNA^{Leu}), indicating a universal non-canonical role of tRNAs in controlling growth. Surprisingly, analysis of 5' OH RNA seq libraries revealed codon-specific ribosome stalling events caused by toxin-mediated tRNA depletion, which we used not only to validate our target identification, but also to uncover novel unannotated ORFs in *M. tuberculosis* genome.

Moreover, this was the first time that a toxin found in *M. smegmatis* had its target identified, despite the relevance of this species in mycobacterial physiology studies. The observation that MazF-mt9 and MazF-ms cleave the same tRNA isotype (tRNA^{Lys}) was intriguing and suggests a conserved relevance of tRNA^{Lys} in mycobacterial biology that needs to be elucidated. The expression of these toxins individually was enough to create a codon content-dependent proteome in *M. smegmatis* that seemed to be specialized to respond to stress.

Whether this altered proteome is also observable in *M. tuberculosis* infection or not still needs to be determined.

Finally, much still needs to be discovered in TA systems from mycobacteria, especially their effect in conditions that they are naturally expressed and the triggers that activate each one of them. Finding the molecular mechanisms by which these toxins act might be the basis for discovering new drugs that aim to eradicate *M. tuberculosis*, even in non-replicating conditions.

References

- Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Ech, M., ... Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537–W544. <https://doi.org/10.1093/nar/gky379>
- Agarwal, S., Tiwari, P., Deep, A., Kidwai, S., Gupta, S., Thakur, K. G., & Singh, R. (2018). System-Wide Analysis Unravels the Differential Regulation and In Vivo Essentiality of Virulence-Associated Proteins B and C Toxin-Antitoxin Systems of *Mycobacterium tuberculosis*. *The Journal of Infectious Diseases*, *217*(11), 1809–1820. <https://doi.org/10.1093/infdis/jiy109>
- Ahidjo, B. A., Kuhnert, D., McKenzie, J. L., Machowski, E. E., Gordhan, B. G., Arcus, V., ... Mizrahi, V. (2011). VapC toxins from *Mycobacterium tuberculosis* are ribonucleases that differentially inhibit growth and are neutralized by cognate VapB antitoxins. *PloS One*, *6*(6), e21738. <https://doi.org/10.1371/journal.pone.0021738>
- Albrethsen, J., Agner, J., Piersma, S. R., Højrup, P., Pham, T. V., Weldingh, K., ... Rosenkrands, I. (2013). Proteomic profiling of *Mycobacterium tuberculosis* identifies nutrient-starvation-responsive toxin-antitoxin systems. *Molecular and Cellular Proteomics*, *12*(5), 1180–1191. <https://doi.org/10.1074/mcp.M112.018846>
- Arcus, V. L., McKenzie, J. L., Robson, J., & Cook, G. M. (2011). The PIN-domain ribonucleases and the prokaryotic VapBC toxin-antitoxin array. *Protein Engineering, Design & Selection : PEDS*, *24*(1–2), 33–40. <https://doi.org/10.1093/protein/gzq081>
- Arcus, V. L., Rainey, P. B., & Turner, S. J. (2005). The PIN-domain toxin-antitoxin array in mycobacteria. *Trends in Microbiology*, *13*(8), 360–365. <https://doi.org/10.1016/j.tim.2005.06.008>
- Ates, L. S., Houben, E. N. G., & Bitter, W. (2016). Type VII Secretion: A Highly Versatile Secretion System. *Microbiol Spectr*, *4*(1). <https://doi.org/10.1128/microbiolspec.VMBF-0011-2015>
- Balaban, N. Q., Helaine, S., Lewis, K., Ackermann, M., Aldridge, B., Andersson, D. I., ... Zinkernagel, A. (2019). Definitions and guidelines for research on antibiotic persistence. *Nature Reviews Microbiology*, *17*(7), 441–448. <https://doi.org/10.1038/s41579-019-0196-3>
- Barth, V. C., & Woychik, N. A. (2020). The Sole *Mycobacterium smegmatis* MazF Toxin Targets tRNALys to Impart Highly Selective, Codon-Dependent Proteome Reprogramming. *Frontiers in Genetics*, *10*(February), 1–11. <https://doi.org/10.3389/fgene.2019.01356>
- Barth, V. C., Zeng, J.-M., Vvedenskaya, I. O., Ouyang, M., Husson, R. N., & Woychik, N. A. (2019). Toxin-mediated ribosome stalling reprograms the *Mycobacterium tuberculosis* proteome. *Nature Communications*, *10*(1), 3035. <https://doi.org/10.1038/s41467-019-10869-8>
- Beavis, R. C. (2006). Using the global proteome machine for protein identification. *Methods in Molecular Biology (Clifton, N.J.)*, *328*(4), 217–228.

- <https://doi.org/10.1385/1-59745-026-x:217>
- Becker, A. H., Oh, E., Weissman, J. S., Kramer, G., & Bukau, B. (2013). Selective ribosome profiling as a tool for studying the interaction of chaperones and targeting factors with nascent polypeptide chains and ribosomes. *Nature Protocols*, 8(11), 2212–2239. <https://doi.org/10.1038/nprot.2013.133>
- Behr, M. A., Edelstein, P. H., & Ramakrishnan, L. (2018). Revisiting the timetable of tuberculosis. *The BMJ*, 362(August), 1–10. <https://doi.org/10.1136/bmj.k2738>
- Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A., & Duncan, K. (2002). Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Molecular Microbiology*, 43(3), 717–731. <https://doi.org/10.1046/j.1365-2958.2002.02779.x>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Canestrari, J. G., Lasek-Nesselquist, E., Upadhyay, A., Rofaeil, M., Champion, M. M., Wade, J. T., ... Gray, T. A. (2020). Polycysteine-encoding leaderless short ORFs function as cysteine-responsive attenuators of operonic gene expression in mycobacteria. *Molecular Microbiology*, (March), 1–16. <https://doi.org/10.1111/mmi.14498>
- Cappelli, G., Volpe, E., Grassi, M., Liseo, B., Colizzi, V., & Mariani, F. (2006). 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. *Research in Microbiology*, 157(5), 445–455. <https://doi.org/10.1016/j.resmic.2005.10.007>
- Chan, P. P., & Lowe, T. M. (2009). GtRNADB: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res*, 37(Database issue), D93–7. <https://doi.org/10.1093/nar/gkn787>
- Chan, P. P., & Lowe, T. M. (2016). GtRNADB 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res*, 44(D1), D184–9. <https://doi.org/10.1093/nar/gkv1309>
- Chawla, M., Parikh, P., Saxena, A., Munshi, M., Mehta, M., Mai, D., ... Singh, A. (2012). *Mycobacterium tuberculosis* WhiB4 regulates oxidative stress response to modulate survival and dissemination in vivo. *Mol Microbiol*, 85(6), 1148–1165. <https://doi.org/10.1111/j.1365-2958.2012.08165.x>
- Chen, Z., Hu, Y., Cumming, B. M., Lu, P., Feng, L., Deng, J., ... Chen, S. (2016). Mycobacterial WhiB6 Differentially Regulates ESX-1 and the Dos Regulon to Modulate Granuloma Formation and Virulence in Zebrafish. *Cell Rep*, 16(9), 2512–2524. <https://doi.org/10.1016/j.celrep.2016.07.080>
- Cheverton, A. M., Gollan, B., Przydacz, M., Wong, C. T., Mylona, A., Hare, S. A., & Helaine, S. (2016). A *Salmonella* Toxin Promotes Persister Formation through Acetylation of tRNA. *Molecular Cell*, 63(1), 86–96. <https://doi.org/10.1016/j.molcel.2016.05.002>

- Chionh, Y. H., McBee, M., Babu, I. R., Hia, F., Lin, W., Zhao, W., ... Dedon, P. C. (2016). tRNA-mediated codon-biased translation in mycobacterial hypoxic persistence. *Nature Communications*, 7, 1–12. <https://doi.org/10.1038/ncomms13302>
- Cintrón, M., Zeng, J.-M., Barth, V. C., Cruz, J. W., Husson, R. N., & Woychik, N. A. (2019). Accurate target identification for *Mycobacterium tuberculosis* endoribonuclease toxins requires expression in their native host. *Scientific Reports*, 9(1), 5949. <https://doi.org/10.1038/s41598-019-41548-9>
- Condon, C. (2010). What is the role of RNase J in mRNA turnover? *RNA Biology*, 7(3), 316–321. <https://doi.org/10.4161/rna.7.3.11913>
- Cortes, T., Schubert, O. T., Rose, G., Arnvig, K. B., Comas, I., Aebersold, R., & Young, D. B. (2013). Genome-wide Mapping of Transcriptional Start Sites Defines an Extensive Leaderless Transcriptome in *Mycobacterium tuberculosis*. *Cell Reports*, 5(4), 1121–1131. <https://doi.org/10.1016/j.celrep.2013.10.031>
- Craig, R., & Beavis, R. C. (2004). TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics*, 20(9), 1466–1467. <https://doi.org/10.1093/bioinformatics/bth092>
- Crooks, G. E. (2004). WebLogo: A Sequence Logo Generator. *Genome Research*, 14(6), 1188–1190. <https://doi.org/10.1101/gr.849004>
- Cruz, J W, & Woychik, N. A. (2016). tRNAs taking charge. *Pathog Dis*, 74(2), ftv117. <https://doi.org/10.1093/femspd/ftv117>
- Cruz, Jonathan W., Sharp, J. D., Hoffer, E. D., Maehigashi, T., Vvedenskaya, I. O., Konkimalla, A., ... Woychik, N. A. (2015). Growth-regulating *Mycobacterium tuberculosis* VapC-mt4 toxin is an isoacceptor-specific tRNase. *Nature Communications*, 6(May), 1–12. <https://doi.org/10.1038/ncomms8480>
- Culviner, P. H., & Laub, M. T. (2018). Global Analysis of the *E. coli* Toxin MazF Reveals Widespread Cleavage of mRNA and the Inhibition of rRNA Maturation and Ribosome Biogenesis. *Molecular Cell*, 70(5), 868-880.e10. <https://doi.org/10.1016/j.molcel.2018.04.026>
- Daleke, M. H., Ummels, R., Bawono, P., Heringa, J., Vandenbroucke-Grauls, C. M., Luijck, J., & Bitter, W. (2012). General secretion signal for the mycobacterial type VII secretion pathway. *Proc Natl Acad Sci U S A*, 109(28), 11342–11347. <https://doi.org/10.1073/pnas.1119453109>
- Deep, A., Tiwari, P., Agarwal, S., Kaundal, S., Kidwai, S., Singh, R., & Thakur, K. G. (2018). Structural, functional and biological insights into the role of *Mycobacterium tuberculosis* VapBC11 toxin–antitoxin system: targeting a tRNase to tackle mycobacterial adaptation. *Nucleic Acids Research*, 46(21), 11639–11655. <https://doi.org/10.1093/nar/gky924>
- Denkin, S., Byrne, S., Jie, C., & Zhang, Y. (2005). Gene expression profiling analysis of *Mycobacterium tuberculosis* genes in response to salicylate. *Archives of Microbiology*, 184(3), 152–157. <https://doi.org/10.1007/s00203-005-0037-9>
- Devayani P. Bhave, Wilson B. Muse III, & Kate S. Carroll. (2008). Drug Targets in Mycobacterial Sulfur Metabolism. *Infectious Disorders - Drug Targets*,

- 7(2), 140–158. <https://doi.org/10.2174/187152607781001772>
- Diel, R. (2017). Long-term Effect of Bacille Calmette-Guérin Vaccination in Tuberculin Skin Testing. *Chest*, *152*(2), 235–236. <https://doi.org/10.1016/j.chest.2017.03.011>
- Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A., & Schuman, E. M. (2006). Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proceedings of the National Academy of Sciences of the United States of America*, *103*(25), 9482–9487. <https://doi.org/10.1073/pnas.0601637103>
- Ehrt, S., Guo, X. V., Hickey, C. M., Ryou, M., Monteleone, M., Riley, L. W., & Schnappinger, D. (2005). Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Research*, *33*(2), 1–11. <https://doi.org/10.1093/nar/gni013>
- Fontan, P., Aris, V., Ghanny, S., Soteropoulos, P., Smith, I., Fontán, P., ... Smith, I. (2008). Global transcriptional profile of *Mycobacterium tuberculosis* during THP-1 human macrophage infection. *Infect Immun*, *76*(2), 717–725. <https://doi.org/10.1128/IAI.00974-07>
- Ford, N., Vitoria, M., Penazzato, M., Doherty, M., Shubber, Z., Meintjes, G., ... Calmy, A. (2015). Causes of hospital admission among people living with HIV worldwide: A systematic review and meta-analysis. *The Lancet HIV*, *2*(10), e438–e444. [https://doi.org/10.1016/S2352-3018\(15\)00137-X](https://doi.org/10.1016/S2352-3018(15)00137-X)
- Fraikin, N., Goormaghtigh, F., & Van Melderen, L. (2020). Type II Toxin-Antitoxin Systems: Evolution and Revolutions. *Journal of Bacteriology*, *202*(7), 1–14. <https://doi.org/10.1128/JB.00763-19>
- Frampton, R., Aggio, R. B. M. M., Villas-Bôas, S. G., Arcus, V. L., & Cook, G. M. (2012). Toxin-antitoxin systems of *Mycobacterium smegmatis* are essential for cell survival. *The Journal of Biological Chemistry*, *287*(8), 5340–5356. <https://doi.org/10.1074/jbc.M111.286856>
- Glaziou, P. (2020). Predicted impact of the COVID-19 pandemic on global tuberculosis deaths in 2020. *MedRxiv*, *000*, 2020.04.28.20079582. <https://doi.org/10.1101/2020.04.28.20079582>
- Gröschel, M. I., Sayes, F., Simeone, R., Majlessi, L., Brosch, R., Groschel, M. I., ... Brosch, R. (2016). ESX secretion systems: mycobacterial evolution to counter host immunity. *Nat Rev Microbiol*, *14*(11), 677–691. <https://doi.org/10.1038/nrmicro.2016.131>
- Gupta, A., Venkataraman, B., Vasudevan, M., & Gopinath Bankar, K. (2017). Co-expression network analysis of toxin-antitoxin loci in *Mycobacterium tuberculosis* reveals key modulators of cellular stress. *Scientific Reports*, *7*(1), 5868. <https://doi.org/10.1038/s41598-017-06003-7>
- Guzman, L. M., Belin, D., Carson, M. J., & Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of Bacteriology*, *177*(14), 4121–4130. <https://doi.org/10.1128/jb.177.14.4121-4130.1995>
- Harms, A., Brodersen, D. E., Mitarai, N., & Gerdes, K. (2018). Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Mol Cell*, *70*(5), 768–784. <https://doi.org/10.1016/j.molcel.2018.01.003>

- Harrison, J. J., Wade, W. D., Akierman, S., Vacchi-Suzzi, C., Stremick, C. A., Turner, R. J., & Ceri, H. (2009). The chromosomal toxin gene *yafQ* is a determinant of multidrug tolerance for *Escherichia coli* growing in a biofilm. *Antimicrob Agents Chemother*, *53*(6), 2253–2258. <https://doi.org/10.1128/AAC.00043-09>
- Hatzios, S. K., & Bertozzi, C. R. (2011). The Regulation of Sulfur Metabolism in *Mycobacterium tuberculosis*. *PLoS Pathogens*, *7*(7), e1002036. <https://doi.org/10.1371/journal.ppat.1002036>
- Hemm, M. R., Weaver, J., & Storz, G. (2020). *Escherichia coli* Small Proteome. *EcoSal Plus*, *9*(1). <https://doi.org/10.1128/ecosalplus.ESP-0031-2019>
- Herzmann, C., Sotgiu, G., Schaberg, T., Ernst, M., Stenger, S., & Lange, C. (2014). Early BCG vaccination is unrelated to pulmonary immunity against *Mycobacterium tuberculosis* in adults. *European Respiratory Journal*, *44*(4), 1087–1090. <https://doi.org/10.1183/09031936.00086514>
- Huang, D. W., Sherman, B. T., Lempicki, R. A., Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, *37*(1), 1–13. <https://doi.org/10.1093/nar/gkn923>
- Huang da, W., Sherman, B. T., Lempicki, R. A., Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, *4*(1), 44–57. <https://doi.org/10.1038/nprot.2008.211>
- Huang, F., & He, Z. G. (2010). Characterization of an interplay between a *Mycobacterium tuberculosis* MazF homolog, Rv1495 and its sole DNA topoisomerase i. *Nucleic Acids Research*, *38*(22), 8219–8230. <https://doi.org/10.1093/nar/gkq737>
- Huebner, R. E., Schein, M. F., & Bass, J. B. (1993). The Tuberculin Skin Test. *Clinical Infectious Diseases*, *17*(6), 968–975. <https://doi.org/10.1093/clinids/17.6.968>
- Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M., & Weissman, J. S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols*, *7*(8), 1534–1550. <https://doi.org/10.1038/nprot.2012.086>
- Jamalli, A., Hébert, A., Zig, L., & Putzer, H. (2014). Control of expression of the RNases J1 and J2 in *Bacillus subtilis*. *Journal of Bacteriology*, *196*(2), 318–324. <https://doi.org/10.1128/JB.01053-13>
- Jothivasan, V. K., & Hamilton, C. J. (2008). Mycothiol: synthesis, biosynthesis and biological functions of the major low molecular weight thiol in actinomycetes. *Nat Prod Rep*, *25*(6), 1091–1117. <https://doi.org/10.1039/b616489g>
- Kaspy, I., Rotem, E., Weiss, N., Ronin, I., Balaban, N. Q., & Glaser, G. (2013). HipA-mediated antibiotic persistence via phosphorylation of the glutamyl-tRNA-synthetase. *Nature Communications*, *4*, 1–7. <https://doi.org/10.1038/ncomms4001>
- Katellaris, A. L., Jackson, C., Southern, J., Gupta, R. K., Drobniowski, F., Lalvani, A., ... Abubakar, I. (2020). Effectiveness of BCG Vaccination Against

- Mycobacterium tuberculosis* Infection in Adults: A Cross-sectional Analysis of a UK-Based Cohort. *The Journal of Infectious Diseases*, 221(1), 146–155. <https://doi.org/10.1093/infdis/jiz430>
- Kawashima, T., Berthet-Colominas, C., Wulff, M., Cusack, S., & Leberman, R. (1996). The structure of the *Escherichia coli* EF-Tu.EF-Ts complex at 2.5 Å resolution. *Nature*, 379(6565), 511–518. <https://doi.org/10.1038/379511a0>
- Kim, S., & Pevzner, P. A. (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. *Nature Communications*, 5. <https://doi.org/10.1038/ncomms6277>
- Kinkar, E., Kinkar, A., & Saleh, M. (2019). The multicopper oxidase of *Mycobacterium tuberculosis* (MmcO) exhibits ferroxidase activity and scavenges reactive oxygen species in activated THP-1 cells. *Int J Med Microbiol*, 309(7), 151324. <https://doi.org/10.1016/j.ijmm.2019.06.004>
- Korch, S. B., Contreras, H., & Clark-Curtiss, J. E. (2009). Three *Mycobacterium tuberculosis* Rel toxin-antitoxin modules inhibit mycobacterial growth and are expressed in infected human macrophages. *Journal of Bacteriology*, 191(5), 1618–1630. <https://doi.org/10.1128/JB.01318-08>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
- Larsson, C., Luna, B., Ammerman, N. C., Maiga, M., Agarwal, N., & Bishai, W. R. (2012). Gene expression of *Mycobacterium tuberculosis* putative transcription factors whiB1-7 in redox environments. *PLoS One*, 7(7), e37516. <https://doi.org/10.1371/journal.pone.0037516>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>
- Lopes, A. P. Y., Lopes, L. M., Fraga, T. R., Chura-Chambi, R. M., Sanson, A. L., Cheng, E., ... Martins, E. A. L. (2014). VapC from the leptospiral VapBC toxin-antitoxin module displays ribonuclease activity on the initiator tRNA. *PLoS One*, 9(7), e101678. <https://doi.org/10.1371/journal.pone.0101678>
- Lorenz, C., Lünse, C. E., & Mörl, M. (2017). tRNA Modifications: Impact on Structure and Thermal Adaptation. *Biomolecules*, 7(2). <https://doi.org/10.3390/biom7020035>
- Lowe, T. M., & Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research*, 44(W1), W54-7. <https://doi.org/10.1093/nar/gkw413>
- Lund, S. P., Nettleton, D., McCarthy, D. J., & Smyth, G. K. (2012). Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. In *Statistical Applications in Genetics and Molecular Biology* (Vol. 11). <https://doi.org/10.1515/1544-6115.1826>
- Masuda, H., & Inouye, M. (2017). Toxins of Prokaryotic Toxin-Antitoxin Systems with Sequence-Specific Endoribonuclease Activity. *Toxins*, 9(4), 140. <https://doi.org/10.3390/toxins9040140>

- Matelska, D., Steczkiewicz, K., & Ginalski, K. (2017). Comprehensive classification of the PIN domain-like superfamily. *Nucleic Acids Research*, *45*(12), 6995–7020. <https://doi.org/10.1093/nar/gkx494>
- Mathy, N., Bénard, L., Pellegrini, O., Daou, R., Wen, T., & Condon, C. (2007). 5'-to-3' Exoribonuclease Activity in Bacteria: Role of RNase J1 in rRNA Maturation and 5' Stability of mRNA. *Cell*, *129*(4), 681–692. <https://doi.org/10.1016/j.cell.2007.02.051>
- Mathy, N., Hébert, A., Mervelet, P., Bénard, L., Dorléans, A., Li de la Sierra-Gallay, I., ... Condon, C. (2010). *Bacillus subtilis* ribonucleases J1 and J2 form a complex with altered enzyme behaviour. *Molecular Microbiology*, *75*(2), 489–498. <https://doi.org/10.1111/j.1365-2958.2009.07004.x>
- McKenzie, J. L., Duyvestyn, J. M., Smith, T., Bendak, K., MacKay, J., Cursons, R., ... Arcus, V. L. (2012). Determination of ribonuclease sequence-specificity using Pentaproboscids and mass spectrometry. *RNA*, *18*(6), 1267–1278. <https://doi.org/10.1261/rna.031229.111>
- McKenzie, J. L., Robson, J., Berney, M., Smith, T. C., Ruthe, A., Gardner, P. P., ... Cook, G. M. (2012). A VapBC toxin-antitoxin module is a posttranscriptional regulator of metabolic flux in mycobacteria. *Journal of Bacteriology*, *194*(9), 2189–2204. <https://doi.org/10.1128/JB.06790-11>
- Merino, E., Jensen, R. A., & Yanofsky, C. (2008). Evolution of bacterial trp operons and their regulation. *Curr Opin Microbiol*, *11*(2), 78–86. <https://doi.org/10.1016/j.mib.2008.02.005>
- Mets, T., Lippus, M., Schryer, D., Liiv, A., Kasari, V., Paier, A., ... Kaldalu, N. (2017). Toxins MazF and MqsR cleave *Escherichia coli* rRNA precursors at multiple sites. *RNA Biology*, *14*(1), 124–135. <https://doi.org/10.1080/15476286.2016.1259784>
- Michel, A. M., Mullan, J. P. A., Velayudhan, V., O'Connor, P. B. F., Donohue, C. A., & Baranov, P. V. (2016). RiboGalaxy: A browser based platform for the alignment, analysis and visualization of ribosome profiling data. *RNA Biology*, *13*(3), 316–319. <https://doi.org/10.1080/15476286.2016.1141862>
- Moule, M. G., & Cirillo, J. D. (2020). *Mycobacterium tuberculosis* Dissemination Plays a Critical Role in Pathogenesis. *Frontiers in Cellular and Infection Microbiology*, *10*(February), 1–12. <https://doi.org/10.3389/fcimb.2020.00065>
- Muela, A., Seco, C., Camafeita, E., Arana, I., Orruño, M., López, J. A., & Barcina, I. (2008). Changes in *Escherichia coli* outer membrane subproteome under environmental conditions inducing the viable but nonculturable state. *FEMS Microbiology Ecology*, *64*(1), 28–36. <https://doi.org/10.1111/j.1574-6941.2008.00453.x>
- Murphy, K. C., Nelson, S. J., Nambi, S., Papavinasasundaram, K., Baer, C. E., & Sasseti, C. M. (2018). ORBIT: a new paradigm for genetic engineering of mycobacterial chromosomes. *BioRxiv*, *9*(6), 1–20. <https://doi.org/10.1101/249292>
- Muttucumar, D. G., Roberts, G., Hinds, J., Stabler, R. A., & Parish, T. (2004). Gene expression profile of *Mycobacterium tuberculosis* in a non-replicating state. *Tuberculosis (Edinb)*, *84*(3–4), 239–246. <https://doi.org/10.1016/j.tube.2003.12.006>

- Nguipdop-Djomo, P., Haldal, E., Rodrigues, L. C., Abubakar, I., & Mangtani, P. (2016). Duration of BCG protection against tuberculosis and change in effectiveness with time since vaccination in Norway: A retrospective population-based cohort study. *The Lancet Infectious Diseases*, *16*(2), 219–226. [https://doi.org/10.1016/S1473-3099\(15\)00400-4](https://doi.org/10.1016/S1473-3099(15)00400-4)
- Nigam, A., Ziv, T., Oron-Gottesman, A., & Engelberg-Kulka, H. (2019). Stress-Induced MazF-Mediated Proteins in *Escherichia coli*. *MBio*, *10*(2), 1–9. <https://doi.org/10.1128/mBio.00340-19>
- Orme, I. M., & Basaraba, R. J. (2014). The formation of the granuloma in tuberculosis infection. *Seminars in Immunology*, *26*(6), 601–609. <https://doi.org/10.1016/j.smim.2014.09.009>
- Pai, M., Behr, M. A., Dowdy, D., Dheda, K., Divangahi, M., Boehme, C. C., ... Raviglione, M. (2016). Tuberculosis. *Nature Reviews Disease Primers*, *2*. <https://doi.org/10.1038/nrdp.2016.76>
- Park, J. H., Yamaguchi, Y., & Inouye, M. (2011). *Bacillus subtilis* MazF-bs (EndoA) is a UACAU-specific mRNA interferase. *FEBS Lett*, *585*(15), 2526–2532. <https://doi.org/10.1016/j.febslet.2011.07.008>
- Pontes, M. H., & Groisman, A. (2020). A Physiological Basis for Nonheritable Antibiotic Resistance. *MBio*, *11*(3), 1–13.
- Provvedi, R., Boldrin, F., Falciani, F., Palù, G., & Manganeli, R. (2009). Global transcriptional response to vancomycin in *Mycobacterium tuberculosis*. *Microbiology (Reading, England)*, *155*(Pt 4), 1093–1102. <https://doi.org/10.1099/mic.0.024802-0>
- Radhakrishnan, A., & Green, R. (2016). Connections Underlying Translation and mRNA Stability. *J Mol Biol*, *428*(18), 3558–3564. <https://doi.org/10.1016/j.jmb.2016.05.025>
- Ramage, H. R., Connolly, L. E., & Cox, J. S. (2009). Comprehensive functional analysis of *Mycobacterium tuberculosis* toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution. *PLoS Genet*, *5*(12), e1000767. <https://doi.org/10.1371/journal.pgen.1000767>
- Raman, S., Song, T., Puyang, X., Bardarov, S., Jacobs Jr., W. R., & Husson, R. N. (2001). The alternative sigma factor SigH regulates major components of oxidative and heat stress responses in *Mycobacterium tuberculosis*. *J Bacteriol*, *183*(20), 6119–6125. <https://doi.org/10.1128/JB.183.20.6119-6125.2001>
- Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., ... Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*, *44*(W1), W160-5. <https://doi.org/10.1093/nar/gkw257>
- Rifat, D., Bishai, W. R., & Karakousis, P. C. (2009). Phosphate depletion: a novel trigger for *Mycobacterium tuberculosis* persistence. *J Infect Dis*, *200*(7), 1126–1135. <https://doi.org/10.1086/605700>
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47. <https://doi.org/10.1093/nar/gkv007>

- Rowland, J. L., & Niederweis, M. (2013). A multicopper oxidase is required for copper resistance in *Mycobacterium tuberculosis*. *J Bacteriol*, *195*(16), 3724–3733. <https://doi.org/10.1128/JB.00546-13>
- Rustad, T. R., Harrell, M. I., Liao, R., & Sherman, D. R. (2008). The enduring hypoxic response of *Mycobacterium tuberculosis*. *PLoS One*, *3*(1), e1502. <https://doi.org/10.1371/journal.pone.0001502>
- Saini, V., Farhana, A., Glasgow, J. N., & Steyn, A. J. (2012). Iron sulfur cluster proteins and microbial regulation: implications for understanding tuberculosis. *Curr Opin Chem Biol*, *16*(1–2), 45–53. <https://doi.org/10.1016/j.cbpa.2012.03.004>
- Sala, A., Bordes, P., & Genevaux, P. (2014). Multiple Toxin-Antitoxin Systems in *Mycobacterium tuberculosis*. *Toxins*, *6*(3), 1002–1020. <https://doi.org/10.3390/toxins6031002>
- Sambandamurthy, V. K., Wang, X., Chen, B., Russell, R. G., Derrick, S., Collins, F. M., ... Jacobs Jr., W. R. (2002). A pantothenate auxotroph of *Mycobacterium tuberculosis* is highly attenuated and protects mice against tuberculosis. *Nat Med*, *8*(10), 1171–1174. <https://doi.org/10.1038/nm765>
- Scanga, C. A., Mohan, V. P., Joseph, H., Yu, K., Chan, J., & Flynn, J. L. (1999). Reactivation of Latent Tuberculosis: Variations on the Cornell Murine Model. *Infection and Immunity*, *67*(9), 4531–4538. <https://doi.org/10.1128/IAI.67.9.4531-4538.1999>
- Schelle, M. W., & Bertozzi, C. R. (2006). Sulfate Metabolism in Mycobacteria. *ChemBioChem*, *7*(10), 1516–1524. <https://doi.org/10.1002/cbic.200600224>
- Schifano, J. M., Cruz, J. W., Vvedenskaya, I. O., Edifor, R., Ouyang, M., Husson, R. N., ... Woychik, N. A. (2016). tRNA is a new target for cleavage by a MazF toxin. *Nucleic Acids Research*, *44*(3), 1256–1270. <https://doi.org/10.1093/nar/gkv1370>
- Schifano, J. M., Edifor, R., Sharp, J. D., Ouyang, M., Konkimalla, A., Husson, R. N., & Woychik, N. A. (2013). Mycobacterial toxin MazF-mt6 inhibits translation through cleavage of 23S rRNA at the ribosomal A site. *Proc Natl Acad Sci U S A*, *110*(21), 8501–8506. <https://doi.org/10.1073/pnas.1222031110>
- Schifano, J. M., Vvedenskaya, I. O., Knoblauch, J. G., Ouyang, M., Nickels, B. E., & Woychik, N. A. (2014). An RNA-seq method for defining endoribonuclease cleavage specificity identifies dual rRNA substrates for toxin MazF-mt3. *Nature Communications*, *5*, 3538. <https://doi.org/10.1038/ncomms4538>
- Schifano, J. M., & Woychik, N. A. (2014). 23S rRNA as an a-Maz-ing new bacterial toxin target. *RNA Biology*, *11*(2), 101–105. <https://doi.org/10.4161/rna.27949>
- Schifano, J. M., & Woychik, N. A. (2017). Cloaked dagger: tRNA slicing by an unlikely culprit. *RNA Biology*, *14*(1), 15–19. <https://doi.org/10.1080/15476286.2016.1255396>
- Sharp, J. D., Cruz, J. W., Raman, S., Inouye, M., Husson, R. N., & Woychik, N. A. (2012). Growth and translation inhibition through sequence-specific RNA binding by *Mycobacterium tuberculosis* VapC toxin. *J Biol Chem*, *287*(16),

- 12835–12847. <https://doi.org/10.1074/jbc.M112.340109>
- Sharp, J. D., Singh, A. K., Park, S. T., Lyubetskaya, A., Peterson, M. W., Gomes, A. L. C., ... Husson, R. N. (2016). Comprehensive Definition of the SigH Regulon of *Mycobacterium tuberculosis* Reveals Transcriptional Control of Diverse Stress Responses. *PLoS One*, *11*(3), e0152145. <https://doi.org/10.1371/journal.pone.0152145>
- Sharrock, A., Ruthe, A., Andrews, E. S. V., Arcus, V. A., & Hicks, J. L. (2018). VapC proteins from *Mycobacterium tuberculosis* share ribonuclease sequence specificity but differ in regulation and toxicity. *PLoS One*, *13*(8), e0203412. <https://doi.org/10.1371/journal.pone.0203412>
- Shell, S. S., Wang, J., Lapierre, P., Mir, M., Chase, M. R., Pyle, M. M., ... Gray, T. A. (2015). Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape. *PLoS Genetics*, *11*(11), 1–31. <https://doi.org/10.1371/journal.pgen.1005641>
- Singh, A., Guidry, L., Narasimhulu, K. V., Mai, D., Trombley, J., Redding, K. E., ... Steyn, A. J. (2007). *Mycobacterium tuberculosis* WhiB3 responds to O₂ and nitric oxide via its [4Fe-4S] cluster and is essential for nutrient starvation survival. *Proc Natl Acad Sci U S A*, *104*(28), 11562–11567. <https://doi.org/10.1073/pnas.0700490104>
- Singh, R., Barry, C. E. 3rd, & Boshoff, H. I. M. (2010). The three RelE homologs of *Mycobacterium tuberculosis* have individual, drug-specific effects on bacterial antibiotic tolerance. *Journal of Bacteriology*, *192*(5), 1279–1291. <https://doi.org/10.1128/JB.01285-09>
- Smith, L. J., Stapleton, M. R., Buxton, R. S., & Green, J. (2012). Structure-function relationships of the *Mycobacterium tuberculosis* transcription factor WhiB1. *PLoS One*, *7*(7), e40407. <https://doi.org/10.1371/journal.pone.0040407>
- Smith, L. J., Stapleton, M. R., Fullstone, G. J., Crack, J. C., Thomson, A. J., Le Brun, N. E., ... Green, J. (2010). *Mycobacterium tuberculosis* WhiB1 is an essential DNA-binding protein with a nitric oxide-sensitive iron-sulfur cluster. *Biochem J*, *432*(3), 417–427. <https://doi.org/10.1042/BJ20101440>
- Song, S., & Wood, T. K. (2018). Post-segregational Killing and Phage Inhibition Are Not Mediated by Cell Death Through Toxin/Antitoxin Systems. *Frontiers in Microbiology*, *9*(APR), 1–6. <https://doi.org/10.3389/fmicb.2018.00814>
- Starosta, A. L., Lassak, J., Jung, K., & Wilson, D. N. (2014). The bacterial translation stress response. *FEMS Microbiology Reviews*, *38*(6), 1172–1201. <https://doi.org/10.1111/1574-6976.12083>
- Strimmer, K. (2008). fdrtool: A versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics*, *24*(12), 1461–1462. <https://doi.org/10.1093/bioinformatics/btn209>
- Tiwari, P., Arora, G., Singh, M., Kidwai, S., Narayan, O. P., & Singh, R. (2015). MazF ribonucleases promote *Mycobacterium tuberculosis* drug tolerance and virulence in guinea pigs. *Nature Communications*, *6*, 1–12. <https://doi.org/10.1038/ncomms7059>
- Tripathi, A., Dewan, P. C., Siddique, S. A., & Varadarajan, R. (2014). MazF-induced Growth Inhibition and Persister Generation in *Escherichia coli*.

- Journal of Biological Chemistry*, 289(7), 4191–4205.
<https://doi.org/10.1074/jbc.M113.510511>
- Vaudel, M., Barsnes, H., Berven, F. S., Sickmann, A., & Martens, L. (2011). SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics*, 11(5), 996–999.
<https://doi.org/10.1002/pmic.201000595>
- Vaudel, M., Burkhart, J. M., Zahedi, R. P., Oveland, E., Berven, F. S., Sickmann, A., ... Barsnes, H. (2015). PeptideShaker enables reanalysis of MS-derived proteomics data sets: To the editor. *Nature Biotechnology*, 33(1), 22–24.
<https://doi.org/10.1038/nbt.3109>
- Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A. C., Engelberg-Kulka, H., & Moll, I. (2011). Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in *Escherichia coli*. *Cell*, 147(1), 147–157. <https://doi.org/10.1016/j.cell.2011.07.047>
- Voskuil, M. I. (2004). *Mycobacterium tuberculosis* gene expression during environmental conditions associated with latency. *Tuberculosis (Edinb)*, 84(3–4), 138–143. <https://doi.org/10.1016/j.tube.2003.12.008>
- Voskuil, M. I., Visconti, K. C., & Schoolnik, G. K. (2004). *Mycobacterium tuberculosis* gene expression during adaptation to stationary phase and low-oxygen dormancy. *Tuberculosis (Edinb)*, 84(3–4), 218–227. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&opt=Citation&list_uids=15207491
- Wade, J. T., & Laub, M. T. (2019). Concerns about “Stress-Induced MazF-Mediated Proteins in *Escherichia coli*” *MBio*, 10(3).
<https://doi.org/10.1128/mBio.00825-19>
- Walling, L. R., & Butler, J. S. (2017). Homologous VapC Toxins Inhibit Translation and Cell Growth by Sequence-Specific Cleavage of tRNA-fMet. *Journal of Bacteriology*, 200(3), 1–15. <https://doi.org/10.1128/JB.00582-17>
- Walling, L. R., & Butler, J. S. (2019). Toxins targeting transfer RNAs: Translation inhibition by bacterial toxin-antitoxin systems. *Wiley Interdisciplinary Reviews. RNA*, 10(1), e1506. <https://doi.org/10.1002/wrna.1506>
- Warman, E., Forrest, D., Wade, J. T., & Grainger, D. C. (2020). Widespread divergent transcription from prokaryotic promoters. *BioRxiv*, 44(0).
<https://doi.org/10.1101/2020.01.31.928960>
- Wayne, L. G., & Sohaskey, C. D. (2001). Nonreplicating persistence of *Mycobacterium tuberculosis*. *Annu Rev Microbiol*, 55(1), 139–163.
<https://doi.org/10.1146/annurev.micro.55.1.139>
- Welin, A., Bjornsdottir, H., Winther, M., Christenson, K., Oprea, T., Karlsson, A., ... Bylund, J. (2015). CFP-10 from *Mycobacterium tuberculosis* selectively activates human neutrophils through a pertussis toxin-sensitive chemotactic receptor. *Infect Immun*, 83(1), 205–213. <https://doi.org/10.1128/IAI.02493-14>
- Wen, Y., Behiels, E., & Devreese, B. (2014). Toxin-Antitoxin systems: their role in persistence, biofilm formation, and pathogenicity. *Pathogens and Disease*, 70(3), 240–249. <https://doi.org/10.1111/2049-632X.12145>
- Winther, K. S., Brodersen, D. E., Brown, A. K., & Gerdes, K. (2013). VapC20 of *Mycobacterium tuberculosis* cleaves the Sarcin-Ricin loop of 23S rRNA.

- Nature Communications*, 4(1), 2796. <https://doi.org/10.1038/ncomms3796>
- Winther, K. S., & Gerdes, K. (2011). Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. *Proc Natl Acad Sci U S A*, 108(18), 7403–7407. <https://doi.org/10.1073/pnas.1019587108>
- Winther, K., Tree, J. J., Tollervey, D., & Gerdes, K. (2016). VapCs of *Mycobacterium tuberculosis* cleave RNAs essential for translation. *Nucleic Acids Research*, 44(20), 9860–9871. <https://doi.org/10.1093/nar/gkw781>
- Wolschendorf, F., Ackart, D., Shrestha, T. B., Hascall-Dove, L., Nolan, S., Lamichhane, G., ... Niederweis, M. (2011). Copper resistance is essential for virulence of *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A*, 108(4), 1621–1626. <https://doi.org/10.1073/pnas.1009261108>
- Woolstenhulme, C. J., Guydosh, N. R., Green, R., & Buskirk, A. R. (2015). High-Precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. *Cell Reports*, 11(1), 13–21. <https://doi.org/10.1016/j.celrep.2015.03.014>
- World Health Organization. (2018a). BCG vaccine: WHO position paper, February 2018 – Recommendations. *Vaccine*, 36(24), 3408–3410. <https://doi.org/10.1016/j.vaccine.2018.03.009>
- World Health Organization. (2018b). The top 10 causes of death. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- World Health Organization. (2019). *Global Tuberculosis Report 2019*. Geneva.
- Wu, X., & Bartel, D. P. (2017). KpLogo: Positional k-mer analysis reveals hidden specificity in biological sequences. *Nucleic Acids Research*, 45(W1), W534–W538. <https://doi.org/10.1093/nar/gkx323>
- Xie, Y., Wei, Y., Shen, Y., Li, X., Zhou, H., Tai, C., ... Ou, H.-Y. (2018). TADB 2.0: an updated database of bacterial type II toxin-antitoxin loci. *Nucleic Acids Research*, 46(D1), D749–D753. <https://doi.org/10.1093/nar/gkx1033>
- Yamaguchi, Y., & Inouye, M. (2011). Regulation of growth and death in *Escherichia coli* by toxin-antitoxin systems. *Nat Rev Microbiol*, 9(11), 779–790. <https://doi.org/10.1038/nrmicro2651>
- Yamaguchi, Yoshihiro, Park, J.-H., & Inouye, M. (2011). Toxin-antitoxin systems in bacteria and archaea. *Annual Review of Genetics*, 45, 61–79. <https://doi.org/10.1146/annurev-genet-110410-132412>
- Yao, S., Blaustein, J. B., & Bechhofer, D. H. (2008). Erythromycin-induced ribosome stalling and RNase J1-mediated mRNA processing in *Bacillus subtilis*. *Molecular Microbiology*, 69(6), 1439–1449. <https://doi.org/10.1111/j.1365-2958.2008.06370.x>
- Zeng, L., Shi, T., Zhao, Q., & Xie, J. (2013). *Mycobacterium* Sulfur Metabolism and Implications for Novel Drug Targets. *Cell Biochemistry and Biophysics*, 65(2), 77–83. <https://doi.org/10.1007/s12013-012-9410-x>
- Zhang, Y., Zhang, J., Hara, H., Kato, I., & Inouye, M. (2005). Insights into the mRNA cleavage mechanism by MazF, an mRNA interferase. *The Journal of Biological Chemistry*, 280(5), 3143–3150. <https://doi.org/10.1074/jbc.M411811200>
- Zhu, L., Inoue, K., Yoshizumi, S., Kobayashi, H., Zhang, Y., Ouyang, M., ...

Inouye, M. (2009). *Staphylococcus aureus* MazF specifically cleaves a pentad sequence, UACAU, which is unusually abundant in the mRNA for pathogenic adhesive factor SraP. *Journal of Bacteriology*, 191(10), 3248–3255. <https://doi.org/10.1128/JB.01815-08>