GENOME SCALE TRACKING OF MYCOBACTERIUM TUBERCULOSIS MAZF TOXIN TARGETS

By

JULIA RENEE GREENDYK

A thesis submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Microbial Biology

Written under the direction of

Nancy Ann Woychik

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October 2017

ABSTRACT OF THE THESIS

Genome scale tracking of *Mycobacterium tuberculosis* MazF toxin targets

by JULIA RENEE GREENDYK

Thesis Director:
Dr. Nancy Ann Woychik

Mycobacterium tuberculosis, the bacteria that causes tuberculosis (TB), claims millions of lives every year. Due to coinfection of TB and HIV, increasing antibiotic resistance, and the unique ability of *M. tuberculosis* to persist in a non-replicating state within its host as a latent infection, TB eradication attempts have consistently failed. Latent TB infects one-third of the world's population and is refractory to many antibiotics, yet the molecular mechanisms by which *M. tuberculosis* enters and maintains latency are poorly understood. However, members of the MazF family of toxins, which act by cleaving specific regions of single-stranded RNA, are implicated in this process. To help elucidate the role that MazF toxins play in latent tuberculosis, we used a novel RNA-sequencing approach, 5'RNA-seq, to determine the RNA targets of 4 *M. tuberculosis* MazF toxins. While MazF-mt10 cleaved at a specific consensus sequence, CACCU, MazF-mt4, -mt7, and -mt11 did not appear to cleave at a specific sequence. MazF-mt7 cleaved at a number of positions within 23S rRNA; degradation of 23S rRNA could serve as a potent mechanism by which to inhibit translation and modulate growth. MazF-mt7 and -mt10 cleaved 16S rRNA slightly upstream of

the 3’ end, thus removing the anti-Shine-Dalgarno sequence and potentially aiding in the formation of stress ribosomes that can preferentially translate leaderless mRNAs in *M. tuberculosis*. MazF-mt4, -mt7 and –mt11 cleaved distinct tRNAs at their anticodon stem loop, providing another way through which to regulate translation. Results from this study could lead to the identification of MazF-related biomarkers for latent tuberculosis, which could in turn be used to develop new, more effective antimicrobials to fight latent tuberculosis.

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Nancy Woychik, for her invaluable contributions to my academic, professional and personal development. Without her help and guidance, this thesis would not have been possible. She challenged me to learn and grow as a scientist, while also showing concern for my life outside of academia. She exhibited patience, enthusiasm and creativity as she explained countless concepts, fostered my science communication skills, and exemplified how to search for knowledge with endless curiosity. I will forever respect and emulate the ethical way in which she conducted research and by which she treated others. I would like to thank my committee members, Dr. Debashish Bhattacharya, Dr. Jeffrey Boyd, and Dr. Gerben Zylstra, for their input and support. I would also like to thank Dr. Zylstra, the director of the Microbial Biology Program, for all the guidance he provided, as well as for all the details he takes care of every day to help Microbial Biology students during grad school. I would like to thank Drs. Zylstra and Boyd for teaching Microbial Genetics and Physiology with clarity and enthusiasm; I believe that these two classes prepared me to perform microbiological research and think like a scientist more than any other coursework I have taken at Rutgers. I would also like to thank Jennifer Krumins, my undergraduate advisor and microbiology professor, for sparking my interest in microbiology. I am also indebted to the Microbial Biology Program and H. and J. Woodruff for supporting me during my first year of graduate school through the H. Boyd and Jeanette I. Woodruff Microbiology Fellowship in Soil and Environmental Microbiology.

Throughout my time in Dr. Woychik's lab, I was privileged to have lab mates who were intelligent, helpful, and entertaining, among many other wonderful qualities. I would like to especially thank Valdir Barth for teaching me (almost) everything I now know how to do in lab, from molecular cloning to 5'RNA-sequencing. Thank you for patiently answering question after question and demonstrating technique after technique, and especially for helping me with the analysis of all the RNA-seq data. I would also like to thank my other lab mates, past and present, for all the ways that have influenced my time in grad school. Melvili Cintron befriended me from my first day in lab, and has been a shaft of light and energy even on dark days. Thank you (and Valdir and undergraduate Megan Chen) for getting really excited about food and traveling and music with me. Thanks to all three of you for providing great conversation that helped me endure many hours of waiting for bacteria to grow and reactions to incubate. Kimberly Cruz, a former postdoc in the lab taught me how to do protein extraction and labeling. Jillian Cortese, a former Master's student, made me feel comfortable in lab as we learned the ropes together, and always had a listening ear to lend. Our lab tech Mike was always willing to help with things in lab, as well as to answer my random questions about metabolism. Tatsuki Miyamoto, a visiting PhD student from Japan, was a great late-night lab buddy while we finished up experiments long after everyone else had gone home. Heather, a current Master's student, was a pleasure to train and is a kindred spirit who shares my love for teaching. Unnati, a new PhD student, is cheerful and curious, inspiring me to ask more question about how things work.

Finally, I would like to thank my family and friends for all their support through this process. My parents, Paul and Ellen, have instilled in me a love for learning, an appreciation for critical thinking, and a value for communication, all of which have aided this journey incredibly. They have taught me to do everything to the glory of God and to do whatever I can to help others. They have exemplified their own willingness to help others by sacrificing their time and energy so that I could focus on my education, and I am forever grateful. My brothers and their wives, Philip and Asriel and Eric and Ashley, as well as my sister, Aria, have always been encouraging of and interested in my research, consistently asking how things were coming along and providing much-needed breaks when necessary. My nephew and nieces, Vincent, Victoria, Genevieve and Lorraine, have reminded me of the beauty of life amid the stress, showering me with the light-hearted energy and effusive love that only children can. My grandparents, Don & Gerda Van Grouw and Augusta Greendyk, never attained college degrees, but they shared life-earned wisdom with me that no college degree can teach you. My Microbial Biology friends, specifically Hoa Vu, Rachel Dean, and Kevin Dillon have been a constant support, and I am so grateful that we can discuss both detailed scientific problems and mundane life events with each other. I am also thankful to the Carkle Crew and many other friends, particularly Abbey Cotton and Ashley Van Grouw, for providing encouragement, prayer, and comic relief, while also being willing to listen as I practiced explaining my research to non-scientists.

Table of Contents

List of figures

Page

Lists of tables

Page

Introduction

General Introduction

According to the 2016 Global Tuberculosis Report of the World Health Organization (WHO), tuberculosis (TB) ranks alongside HIV as a leading cause of death throughout the world, causing 1.8 million deaths in 2015 (WHO, 2016). *Mycobacterium tuberculosis* (Mtb), the causative agent of TB, is an airborne pathogen that has plagued humanity for thousands of years. While Mtb can infect various areas of the body, including the bones, intestines, and lymph nodes, Mtb most commonly infects the lungs, causing pulmonary TB. Pulmonary TB (referred to as TB from this point on) can exist in two forms: active TB and latent TB. Active TB is contagious and has visible symptoms, including fever, cough with bloody sputum, weight loss, and night sweats. Around 70% of patients with active TB will die if the infection is left untreated (Tiemersma et al., 2011). On the other hand, latent TB is noncontagious and asymptomatic, but has acted as a reservoir for new epidemics throughout history, making the disease difficult to eradicate even with modern antimicrobial drugs.

Current treatments for TB include a lengthy and side-effect laden 6-9 month treatment of 4 first-line drugs: isoniazid, rifampin, ethambutol and pyrazinamide (CDC, 2017). Over the past 50 years, antibiotic resistant TB has emerged, often necessitating the use of second-line (reserve) drugs, including kanamycin and fluoroquinolones (WHO, 2016). However, some strains of TB are even resistant to the second-line drugs, demonstrating the need for the

development of new and more effective TB drugs. Further highlighting the need for new techniques to tackle TB, a recent study found evidence of Mtb mRNA in the sputum of TB patients who had gone through the 6-9 month treatment and had been considered clinically cured for over a year (Malherbe et al., 2016). This indicates that even patients who are considered free from TB may still have viable Mtb in their lungs.

Tuberculosis: ancient disease, recurring issue

Evidence for historic Mtb infection is widespread, from bones and soft tissue of mummies in ancient Egypt, to records of “pthisis” and “consumption” in Asia, Europe, North America and South America. Despite much studying and surmising by ancient and medieval scientists, the cause of tuberculosis was not discovered until the 1800s, after Rene Laennec discovered that tubercles (a type of lesion) were found in the lungs of TB patients. In 1882, Robert Koch discovered microorganisms in these tubercles, named them the *Tubercle bacillus* and coined the abbreviation “TB.” The *Tubercle bacillus*, later renamed *Mycobacterium tuberculosis*, has be intensely studied since its discovery, yet it has become evident that our current knowledge is not sufficient for the eradication of this persistent pathogen.

Toward the end of the 1800s TB cases began to decline due to improved sanitation techniques. Then, in 1943, Selman Waksman and Albert Schatz discovered the soil microorganism *Streptomyces griseus* at Rutgers University; this led to the development of streptomycin, the first effective cure for

tuberculosis (Schatz et al., 1944). Numerous other antimicrobials effective against TB were developed shortly thereafter, leading to a significant decline in TB cases, particularly in Europe and North America (Frieden et al., 1993). However, resistance to streptomycin and other anti-tubercular drugs was quickly observed, in some cases within a few months of its introduction (e.g. streptomycin) (Lewis, 2013). In addition, the emergence of HIV/AIDS in the 1980s caused a resurgence of TB, especially among the poor. New York City was particularly affected by this resurgence, reporting a 132% increase in TB cases from 1980 to 1990 (Frieden et al., 1993). In 1993, the World Health Organization declared that TB was a global emergency (Nakajima, 1993).

In response to this global emergency, the WHO instituted the DOTS (directly observed therapy, short-course) program, which required that patients take their medication in the presence of a healthcare worker for the first two months to ensure that the medication was taken consistently. While this strategy was quite effective for classic cases of TB, it proved ineffective against the growing number of drug-resistant, multidrug-resistant (MDR-) and extensively drug-resistant (XDR-) TB cases (Raviglione and Uplekar, 2006). A revised DOTS program called DOTS Plus included strategies for fighting drug-resistant TB, with only limited success. Cure rates are currently 83% for TB, 52% for MDR -TB and 28% for XDR-TB (WHO, 2016). Although cases of TB are once again falling worldwide, drug-resistance is still on the rise – there were 580,000 drug resistant cases in 2015 alone (WHO, 2016). As the WHO Global Tuberculosis Report attests, TB is still a major global health problem; besides killing 1.8 million people

in 2015, TB also caused an active infection in approximately 10.4 million people (WHO, 2016). In addition, one-third of the world's population is infected with the latent form of TB (Dye et al., 1999).

Since latent TB makes up more than 90% of the current TB burden and acts as a potential reservoir for new epidemics, understanding Mtb latency is crucial for effective control of TB worldwide. During a latent TB infection, Mtb penetrates deep into the lungs, where it is engulfed by macrophages, forming granulomas where Mtb can persist for decades in a non-replicating state without causing active manifestations of the disease (Cruz and Woychik, 2016; Mwinga and Bernard Fourie, 2004; Ramage et al., 2009). Latent Mtb cells in granulomas are characterized by downregulated cellular processes and arrested growth. When the host is immunocompromised or otherwise stressed, Mtb can exit from this latent state, causing an active case of TB (Cruz and Woychik, 2016). Because latent Mtb cells attenuate or pause many of the cellular processes that are targeted by most antibiotics, it is difficult to treat latent TB with current antimicrobials (Ramage et al., 2009). Thus, the ability of Mtb to enter and exit latency is vital to its success as a pathogen.

Toxin-antitoxin systems

The molecular mechanisms through which Mtb establishes and maintains latency are largely unknown; however, several studies point to the involvement of toxin-antitoxin (TA) systems (Cruz et al., 2015; Korch et al., 2009; Provvedi et al., 2009; Ramage et al., 2009; Schifano et al., 2016; Schifano et al., 2013; Schifano

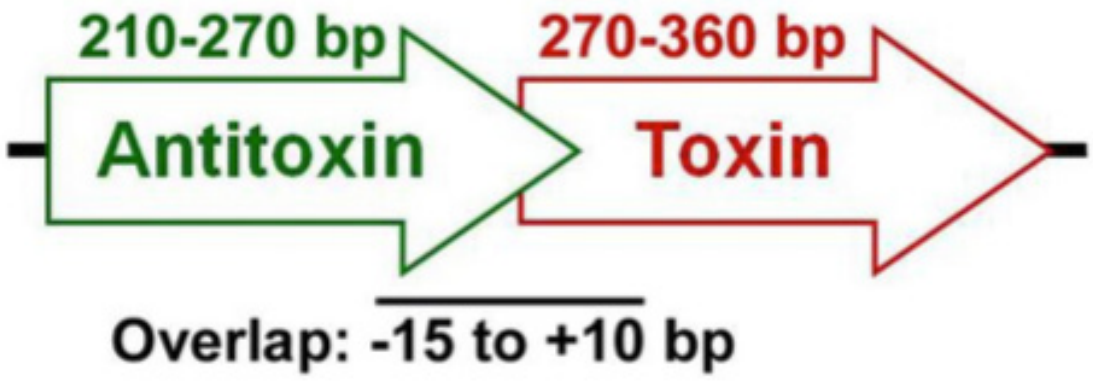
and Woychik, 2014; Sharp et al., 2012; Winther et al., 2013). TA systems are ubiquitously distributed amongst free-living prokaryotes. They comprise an autoregulated operon made up of an antitoxin and a toxin; the antitoxin serves to neutralize the toxin under normal conditions (Provvedi et al., 2009; Van Melderen and De Bast, 2009; Yamaguchi and Inouye, 2011; Yamaguchi et al., 2011).

There are five different functional categories of TA systems, based on the characteristics and actions of the antitoxin. Type I and III antitoxins have RNA antitoxins, while Types II, IV, and V have protein antitoxins. Toxins of all five types are proteins (Goeders and Van Melderen, 2014). Type I antitoxins are made up of antisense RNA and bind to the mRNA of their cognate toxins, thus inhibiting them (Fozo et al., 2010). Type II antitoxins are the most common and well-studied of the TA systems; they inhibit their toxins by physically interacting with them (Yamaguchi et al., 2011). Types III, IV and V each only have one example. ToxI, the Type III RNA antitoxin, has tandem repeats that inhibit the toxin by binding directly to the toxin protein (Short et al., 2013). The Type IV antitoxin, YeeU, binds to the targets of its cognate toxin, thereby keeping the toxin from acting (Masuda et al., 2012). GhoS, the Type V antitoxin, cleaves the mRNA of its cognate toxin, thus preventing its translation (Wang et al., 2012).

The TA systems found in Mtb are all Type II antitoxins, and typically have one promoter, with an upstream antitoxin and a downstream toxin that are cotranscribed and cotranslated. The open reading frames (ORFs) for the antitoxin and the toxin are typically very close to each other, either overlapping slightly or within 10 bases of each other. Both the toxin and the antitoxin are

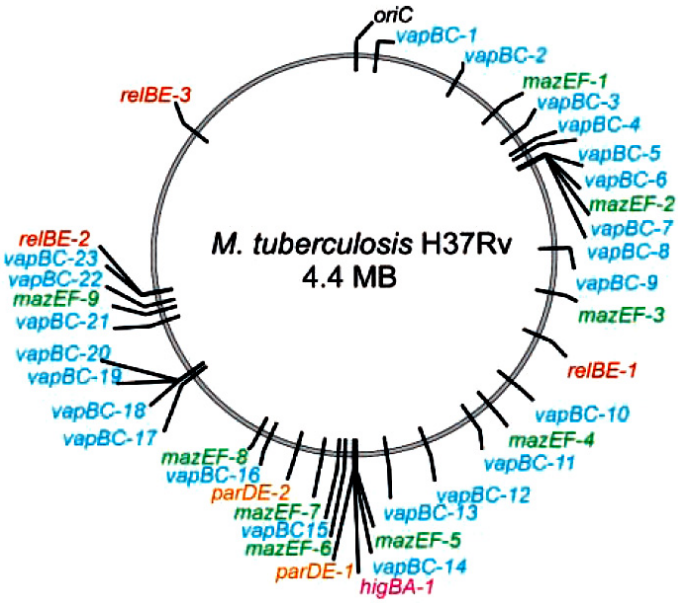
relatively small, about 80-120 amino acids (Figure 1). Under normal conditions, the antitoxin protein binds to the toxin protein, inactivating the toxin; however, during periods of stress, proteases degrade the antitoxin, leaving free toxin. Under these conditions, the toxin can interrupt many cellular processes, such as ATP production, RNA stabilization, DNA and protein synthesis, and various metabolic processes. As a result, it alters the physiology of the cell, often inhibiting growth (Ramage et al., 2009; Winther et al., 2013; Yamaguchi and Inouye, 2011; Yamaguchi et al., 2011). These features have a striking resemblance to characteristics seen in latent TB cells (Cruz et al., 2015; Schifano et al., 2016; Schifano et al., 2013; Schifano and Woychik, 2014). In addition, Mtb has over 88 putative TA systems (Figure 2) further implicating these systems in latent TB (Cruz and Woychik, 2016). There are six families of Type II TA systems found in Mtb, including VapBC, MazEF, RelBE, ParDE, HigBA, and YefM/YoeB (Sala et al., 2014). The MazEF family of TA toxins has 11 orthologs in Mtb, and they are the focus of this thesis work.

Figure 1. Genomic structure of a Type II toxin-antitoxin system operon.



Type II TA system operons are typically made up of an antitoxin and a toxin whose ORFs either overlap by up to 15 bases or are separated by less than 10 bases.

Figure 2. Map of TA systems in *M. tuberculosis*.



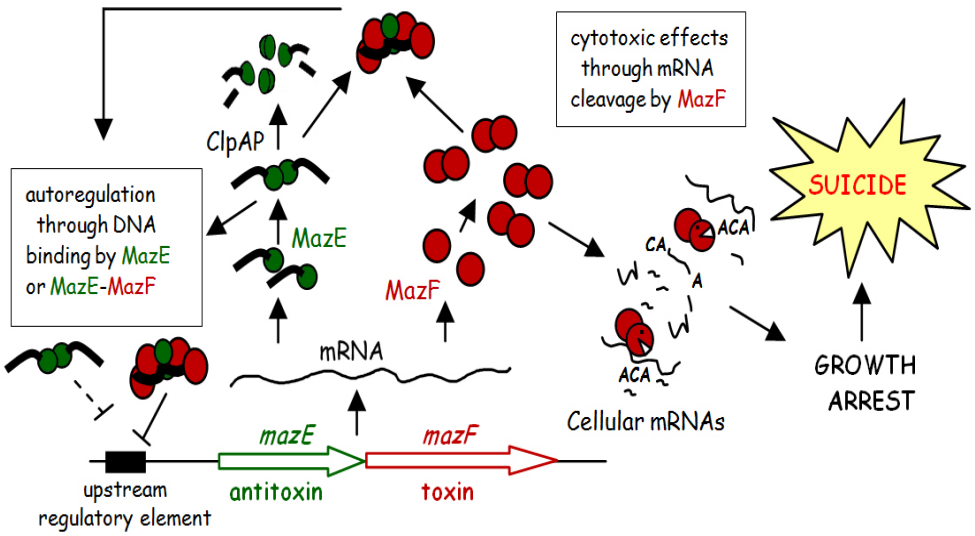
Mtb has an unusual number of TA systems, with over 88 putative TA systems, a number of which are in the MazEF family (highlighted in green).

MazEF toxin-antitoxin systems in *E. coli*

MazEF TA systems from *E. coli* have been characterized in detail. There are two TA loci in the *mazEF* family on its chromosome: *mazEF* (originally named *chpAI*/*chpAK* after the chromosomal homolog of *pem* A locus where it is found on the chromosome) and *chpBI*/*chpBK* (Amitai et al., 2004; Erental et al., 2012; Masuda et al., 1993; Yamaguchi and Inouye, 2011). The general features of *E. coli* MazEF TA systems are illustrated in Figure 3. During periods of low or no stress, MazE and MazF form heterohexamer complexes, with alternating toxin and antitoxin homodimers (2 MazF - 2 MazE - 2 MazF) (Kamada et al., 2003). This serves two purposes; not only does it neutralize the toxin, but it also binds to an upstream regulatory element, thus repressing the operon. In addition, free MazE dimers can also autoregulate the operon by binding to the same upstream regulatory element (Marianovsky et al., 2001; Tsuchimoto and Ohtsubo, 1993). During periods of stress – such as oxidative stress, nutrient limitation, high temperature, antibiotic exposure and DNA damage – the cellular proteases Lon and ClpAP can degrade the MazE antitoxin, activating the MazF toxin (Aizenman et al., 1996; Christensen et al., 2003; Hazan et al., 2004; Sat et al., 2001). Activated MazF toxins have endoribonuclease activity, cleaving single-stranded RNA and thus disrupting translation and inhibiting growth. This activity serves to regulate the growth of the bacteria under stressful conditions. When stressors are removed, the antitoxin is produced once again, and growth returns to normal. However, if the stress is not removed after a long period of time, the cell will reach a “point of no return,” after which it will no longer be able to survive in the

presence of toxin without antitoxin and will initiate cell death (Aizenman et al., 1996; Amitai et al., 2004; Hazan et al., 2004; Sat et al., 2001).

Figure 3. *E. coli* MazEF as a model TA system.



The *E. coli* MazEF operon is cotranscribed and cotranslated and the toxin and antitoxin bind to each other during normal conditions. During conditions of stress, proteases such as ClpAP degrade the antitoxin, allowing the toxin to cleave mRNA, thus causing growth arrest, and if the stress is sustained, cell suicide.

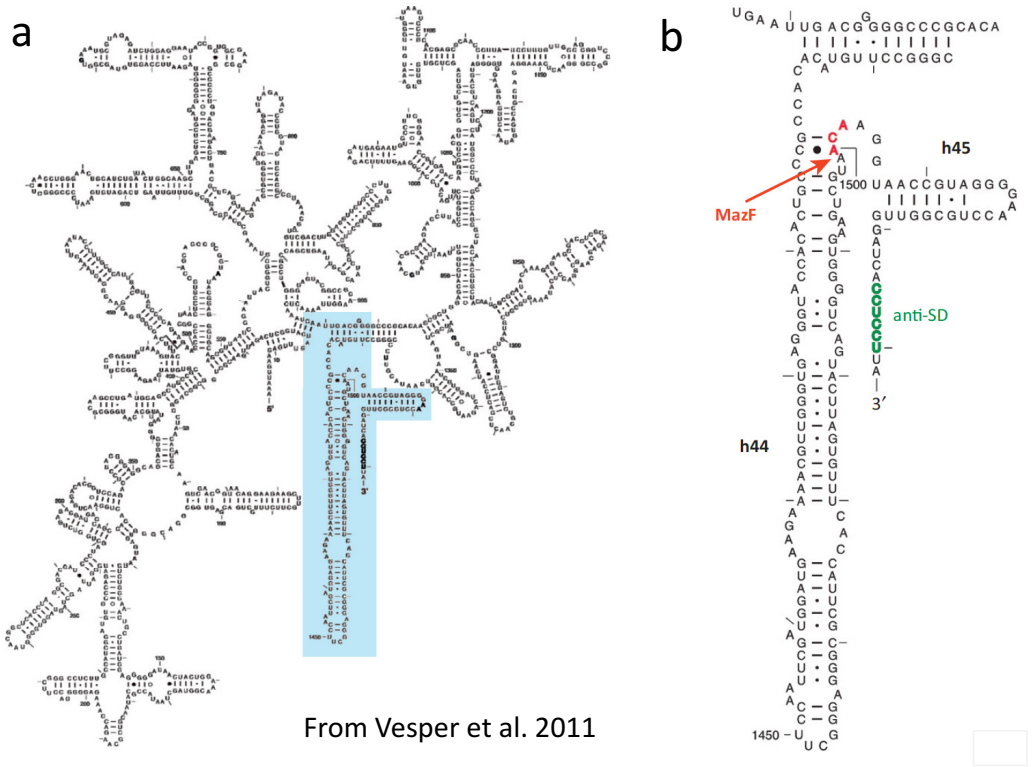
MazFs typically recognize and cleave RNA at a unique and specific sequence – called a consensus or recognition sequence – that can be 3-, 5-, or 7-base sequences long (Zhang et al., 2005; Zhang et al., 2003; Zhu et al., 2008; Zhu et al., 2006). *E. coli* MazF cleaves RNA at a 3-base sequence, ACA. Because this is a short sequence, MazF cleaves most mRNAs in *E. coli,* inhibiting protein synthesis; however, transcripts lacking an ACA sequence can still be translated (Zhang et al., 2003). In addition, several studies demonstrate that rRNA is stable even when MazF is active (Baik et al., 2009; Zhang et al., 2003). This led scientists to postulate that MazF toxins only cleave mRNA, not

tRNA or rRNA, and MazFs came to be known as “mRNA interferases” (Zhang et al., 2005; Zhang et al., 2003; Zhu et al., 2008; Zhu et al., 2006).

However, recent studies indicate that 16S rRNA, the component of the 30S small ribosomal subunit that contains the anti-Shine-Dalgarno (aSD) sequence, is cleaved 43 nucleotides from its 3’ end by *E. coli* MazF *in vivo*, removing the aSD sequence (Figure 4) (Mets et al., 2017; Sauert et al., 2016; Vesper et al., 2011). During typical translation initiation, the aSD region of the 30S subunit binds to the Shine-Dalgarno (SD) sequence, a ribosome binding site found slightly upstream of the start codon on most canonical mRNAs. mRNAs that have an upstream SD sequence are called leadered mRNAs. Studies also found that MazF cleaves at the 5’UTR of many mRNAs, thus removing the SD sequence as well and resulting in a subpopulation of leaderless mRNAs (Sauert et al., 2016; Vesper et al., 2011). Historically, it was believed that translation was unable to initiate without both the SD sequence on mRNA and the aSD on 16S rRNA. Contradicting this idea, Vesper et al. found that *E. coli* MazF cleaved both SD and aSD sequences, resulting in a subpopulation of ribosomes that are able to selectively translate leaderless mRNAs (2011). In addition, a study by Temmel et al. found that MazF-cleaved 16S rRNA is stable in *E. coli* during periods of stress. Remarkably, they also found that *E. coli* has a mechanism by which to repair 16S rRNA when the stress has been removed; 16S rRNA can be rejoined with its MazF-cleaved 43 nt 3’ end by the RNA ligase RtcB, restoring canonical function of ribosomes (Temmel et al., 2017). The opposing activities of

MazF and RtcB provide an elegant and paradigm-shifting method by which *E. coli* cells can regulate translation and thus respond to and recover from stress.

Figure 4. *E. coli* MazF removes the aSD sequence of 16S rRNA.



From Vesper et al. 2011

Contrary to prior belief that MazF cleaves only mRNAs, Vesper et al. found evidence that *E. coli* MazF cleaves at the ACA recognition sequence near the 3'-end of 16S rRNA, effectively removing the anti-Shine-Dalgarno (aSD) sequence. a) Schematic of 16srRNA with region containing the aSD highlighted in blue. b) Highlighted area of 16s rRNA with aSD in green and MazF consensus sequence in red (arrow to specific cleavage site).

Mycobacterium tuberculosis MazF toxins

Mtb MazF toxins have been specifically linked with the nonreplicating persistent state characteristic of latent TB and the general Mtb stress response for a number of reasons. First, MazF activity in other bacteria is known to play a role in stress response and persistence (Keren et al., 2004; Maisonneuve et al.,

2011; Pellegrini et al., 2005; Zhang et al., 2005; Zhang et al., 2003). Second, expression of a single Mtb MazF toxin can lead to translation inhibition and a growth arrested state in Mtb cells (Schifano et al., 2016; Schifano et al., 2013; Schifano et al., 2014; Schifano and Woychik, 2014; Winther et al., 2013). Third, MazF transcripts are upregulated after exposure to stresses encountered during TB infection such as hypoxia, DNA damage, nutrient starvation, or macrophage infection (Betts et al., 2002; Korch et al., 2009; Rand et al., 2003). While typical MazF-containing bacteria have only one or two MazF paralogs, the MazF family has 11 members in Mtb that exhibit significant protein sequence and structural similarity. Only 5 of these MazF paralogs have been studied to date: MazF-mt1 (Rv2801c), MazF-mt3 (Rv1991c), MazF-mt6 (Rv1102c), MazF-mt7 (Rv1495) and MazF-mt9 (Rv2063a) (Mtb gene names are given Rv numbers and are listed in the genome browser TubercuList[1], http://tuberculist.epfl.ch/) (Lew et al., 2011). These MazFs all cleave mRNA, modulate translation, and inhibit growth (Schifano et al., 2016; Schifano et al., 2013; Schifano et al., 2014; Zhu et al., 2008; Zhu et al., 2006) .

Interestingly, among the Mtb MazF toxins studied to date, some were found to cleave not only mRNA, but some also cleave rRNA (MazF-mt3, MazF-mt6) – as found in *E. coli* MazF – or tRNA (MazF-mt9), adding another level of complexity to the mechanism by which MazF toxins regulate translation and growth in Mtb. In 2008, Zhu et al. published a limited analysis revealing that

[1] Mtb MazF toxins were concurrently studied and numbered by two different labs, the Inouye and Gerdes labs. The numbering system used here is the one developed by the Inouye lab; this is not the numbering system used on TubercuList, so some of the MazF numbers differ.

MazF-mt3 cleaves at U↓CCUU sequences (↓ indicates position of cleavage) within single-stranded regions of mRNA. A few years later, the Woychik lab developed a novel genome-scale RNA-sequencing method, 5'RNA-seq, that allows for a more specific yet comprehensive search for MazF toxin targets. This technology was first applied to MazF-mt3, and revealed that while MazF-mt3 does cleave some mRNAs, it primarily cleaves within two regions of rRNA: the essential, conserved helix/loop 70 of 23S rRNA and the anti-Shine-Dalgarno (aSD) sequence of 16S rRNA (Schifano et al., 2014). Similarly, an initial study by Zhu et al. using primer extension of a limited number of RNA templates suggested that MazF-mt6 cleaves at UU↓CCU sequences of single-stranded mRNA. However, a follow-up study by Schifano et al. employed extensive primer extension analysis to show that MazF-mt6 also cleaves within helix/loop 70 of 23S rRNA (Schifano et al., 2013; Zhu et al., 2006). Finally, 5'RNA-seq analysis of MazF-mt9 indicated that MazF-mt9 cleaves both mRNA and tRNA at a UU↓U consensus sequence. MazF-mt9 cleaves two different tRNAs; it cleaves $tRNA^{Lys43}$ within the anticodon stem loop and $tRNA^{Pro14}$ within the D-loop. Further analysis revealed that MazF-mt9 also requires a specific structure within the stem of the tRNA that is cleaved in order for cleavage to occur (Schifano et al., 2016).

The other two MazF toxins that have been studied, MazF-mt1 and MazF-mt7, were shown to cleave single-stranded mRNA at U↓AC and U↓CGCU consensus sites, respectively (Zhu et al., 2008; Zhu et al., 2006). However, neither of these MazF toxins have undergone extensive primer extension or

5'RNA-seq analysis, so their full complement of RNA targets is not known. In addition, the RNA targets of the 6 uncharacterized MazFs also remains unknown. Two of the MazF toxins that are uncharacterized, MazF-mt10 and MazF-mt11, have not yet been recognized in the literature as MazF paralogs. However, an alignment of these genes with all the currently recognized MazF toxins was performed by Jason Schifano, a previous student in the Woychik lab and showed significant alignment (Figure 11). Jason also found that both MazF-mt10 and MazF-mt11 inhibit growth to varying degrees in *E. coli*, as do other MazF toxins. Thus, we decided to study these potential MazF toxins, along with several other MazF toxins. The goal of this study was to use 5' RNA-seq, to determine the RNA cleavage target(s) of MazF-mt4 (Rv0659c), MazF-mt5 (Rv1942c), MazF-mt7, MazF-mt8 (Rv2274c), MazF-mt10 (Rv3098a) and MazF-mt11 (Rv0299), and to identify the specific RNAs targeted by each toxin. Since all of the previously characterized MazF toxins require a precise consensus sequence for MazF toxin recognition and cleavage, another goal of this study was to determine whether these MazFs cleave in a sequence-specific manner, and if so, to determine what each MazF consensus sequence is.

Figure 5. Sequence alignment of 11 Mtb toxins in the MazF family

In order to determine the sequence similarity of MazF toxins found in M. tuberculosis, all 11 *mazF* genes found in the genome were aligned, using *E. coli* MazF as a reference point. Black and shaded backgrounds are used to show identical and homologous residues, respectively. Figure courtesy of Jason Schifano.

Materials and Methods

Strains, plasmids and reagents

The *E. coli* strain BW25113Δ6 [*lacI*q *rrnB*T14 Δ*lac-Z*WJ16 *hsdR514* Δ*araBAD*AH33 Δ*rhaBAD*LD78 Δ*chpBIK* Δ*dinJ-yafQ* Δ*hipBA* Δ*mazEF* Δ*relBE* Δ*yefM-yoeB*] (Prysak et al., 2009) was used for all *E. coli* growth profile and RNA cleavage studies. The attenuated *Mycobacterium tuberculosis* strain $mc^2$6030 [Δ*panCD* Δ*RD1]* (Hsu et al., 2003; Sambandamurthy et al., 2002) was used for *M. tuberculosis* growth profile studies. The *E. coli* strain Mach1-T1 [F– Δ*rec*A1398 *end*A1 *ton*A φ80(*lacZ)*ΔM15 Δ*lac*X74 *hsd*R(rκ–, mκ+); Invitrogen] was used for all cloning experiments. The plasmids used in this study were pBAD33 (Guzman et al., 1995) for expression in *E. coli* and pMC1s for expression in Mtb (Ehrt et al., 2005). Plasmids pBAD33-*mazF-mt7*, pBAD33-*mazF-mt8*, pBAD33-*mazF-mt10*, and pBAD33-*mazF-mt11* were obtained from Jason Schifano, along with plasmid pMC1s-*mazF-mt10*.

All *E. coli* liquid cultures were grown at 37°C in M9 minimal medium supplemented with either 0.2% glucose or 0.1% glycerol. Toxins were expressed via an arabinose-inducible promotor in pBAD33-*mazF-mtx* (x being the *mazF-mt* of interest, e.g. pBAD33-*mazF-mt4*). All Mtb liquid cultures were grown in 7H9 Middlebrook medium supplemented with 10% OADC Middlebrook enrichment, 0.05% Tyloxapol, 0.02% casamino acids, 0.024 mg pantothenate, and 0.5% glycerol. Toxins were expressed via a tetracycline-inducible promoter in pMC1s-*mazF-mtx*. The working concentrations of chloramphenicol in *E. coli* was 25 μg/mL, while the concentration of kanamycin in Mtb was 50 μg/mL.

RNA isolation for 5'RNA-seq

The following RNA isolation protocols were used to isolate total RNA for 5'RNA-Seq with mazF-mt7, mazF-mt10 and mazF-mt11 in *E. coli*. Total RNA was isolated from *E. coli* strain BW25113Δ6 with either pBAD33 or pBAD33-*mazF-mtx* grown to mid-logarithmic phase. Cultures were grown to an OD600nm of 0.3, after which arabinose was added to a final concentration of 0.2% and the cultures allowed to grow for an additional 60 (mazF-mt7), 90 (mazF-mt11), or 180 (mazF-mt10) min post-induction. The time of harvesting was determined based on the point at which each induced mazF started to show a growth separation from the cells containing empty pBAD33. Cells were pelleted by centrifugation at 3,000 × *g* at 4°C for 10 min, and supernatants were removed. Cell pellets were resuspended in TRIzol Reagent (Invitrogen) and lysed in a Precellys Evolution with 0.1mm silica beads at 10,000 rpm three times for 30 seconds each time. Lysates were precipitated with ethanol and added to a Zymo-Spin IIC column. After an initial wash with Direct-zol RNA Wash Buffer, gDNA was digested with 5U of DNAse I. The RNA was purified following the Direct-zol RNA Miniprep protocol (Zymo Research). To further purify the RNA, RNA was eluted from the column with nuclease-free water, treated with 2U of TURBO DNase (Invitrogen) for 30 min at 37°C, treated with TRIzol, and purified using ethanol, a Zymo-Spin IIC column, and the Direct-zol RNA PreWash and Wash Buffers as directed by the Direct-zol RNA Miniprep protocol. RNA was eluted from the column with nuclease-free water, and the quantity and quality of the RNA checked using an Eppendorf BioSpectrometer.

The same protocol was used for mazF-mt8 in *E. coli*, with the exception that rather than using empty pBAD33 as a control, pBAD33-mazF-mt8 that was not induced with arabinose was used as a control (total RNA from these samples was harvested at 180 hours). Total RNA from *E. coli* with mazF-mt4 and mazF-mt5 was obtained from Tatsuki Miyamoto. For MazF-mt4, growth in *E. coli* showed only slight growth inhibition, so cultures were diluted and grown to exponential phase again, as described in Zhu et al. 2006. MazF-mt5 total RNA also only showed limited growth inhibition, but was isolated after 130 minutes of induction. RNA was then harvested at 120 minutes post-dilution. To isolate total RNA from Mtb cells overexpressing mazF-mt10, the Mtb strain $mc^2$6030 containing pMC1s-mazF-mt10 was grown to an OD_{600} of 0.1. The culture was then split in half, and anhydrotetracycline (ATC) was added to one half to a final concentration of 100 ng/mL and the other half grown without ATC as a control. Cultures were grown for three days, after which cells were harvested and RNA isolated using the same protocol as above.

Preparation of RNA for high-throughput sequencing

RNA was prepared for high-throughput sequencing using the procedure described in Schifano *et al*. (2016), with some slight modifications. Since rRNA is a potential target of MazF toxins, we did not deplete rRNA before cDNA preparation. 3 ug of RNA were treated with 1 U of Terminator 5'-Phosphate-Dependent Exonuclease (Epicentre). This process removed RNAs with a 5'-monophosphate (5'-P), but left RNAs with a 5'monophosphate (5'-OH). Then, the

RNAs with a 5'-OH were phosphorylated with 50 U of T4 PNK to create 5'P ends that are suitable for ligation to a 5' adaptor. Next, these RNAs were ligated to the Illumina 5' RNA Adaptor, s1086rna, RA5+N6 (5'-GUUCAGAGUUCUACAGUCCGACGAUCNNNNNN-3'). Ligation reactions contained 5'-P RNA, 30 pmol Illumina 5'Adaptor and 10 U T4 RNA Ligase I (New England Biolabs), and were incubated at 16°C for 20 hours. Excess 5' Adaptor was separated from the 5' Adaptor-ligated RNAS by running the ligation reactions on a 1mm 6% TBE-urea gel and gel-excising RNAs that migrated above the free 5' Adaptor. Then, cDNA libraries were generated via reverse transcription using a primer with nine degenerate nucleotides at the 3'end, Illumina s1082, RTP (5'-GCCTTGGCACCCGAGAATTCCANNNNNNNNN-3'). To perform the annealing step, 500-600 ng of 5' Adaptor was mixed with 500 ng of RT primer and incubated at 85°C for 3 mins and then at 4°C for 1 min. Then reverse transcription was performed with 200 U SuperScript IV Reverse Transcriptase (Invitrogen), dNTPs, and the RNA with the annealed primer. The mixture was incubated at 25°C for 5 min and then 55°C for 30 min. After the reverse transcription, the RT enzyme was inactivated at 80°C for 10 min. Then, 10 U of RNase H (Ambion) was added and the mixture incubated at 37°C for 30 min in order to digest the RNA from the DNA-RNA duplexes. The samples were electrophoresed on a 1mm 6% TBE-urea gel, and cDNAs from 80-500 nts were excised from the gel. The cDNA libraries were PCR amplified using reagents from a TruSeq® Small RNA Sample Prep Kit. The PCR had an initial denaturation step of 30 sec at 98°C, 12 cycles of annealing as follows:

denaturation - 10 sec at 98°C; annealing - 20 sec at 62°C; extension - 10 sec at 72°C, and a final extension for 7 min at 72°C. The amplified DNA was run on a 1.0 mm 10% TBE PAGE gel, and DNA from 150 bp -450 bp were excised from the gel and sent for Illumina High-Throughput Sequencing.

5'RNA-seq data analysis

We aligned the 5'RNA-seq reads from all *E. coli* samples with the genome of the *E. coli* strain K12 substrain MC 4100 (accession number HG738867 in the genome browser GenBank, https://www.ncbi.nlm.nih.gov/genbank/) (Benson et al., 2013) and obtained approximately 30 million sequencing reads per sample. WE aligned the 5'RNA-seq reads from the Mtb samples with the genome of M. tuberculosis strain H37Rv (from the genome browser Tuberculist, http://tuberculist.epfl.ch/) (Lew et al., 2011) and obtained approximately 40 million sequencing reads that aligned to the *M. tuberculosis* strain H37Rv genome for each sample. We used Bowtie version 1.1.2 (Langmead et al., 2009) to identify sequencing reads for which the first 25 bases mapped to the *E. coli* or *M. tuberculosis* genomes with zero mismatches. We determined the fold change difference between MazF-induced and empty vector/MazF-uninduced control samples by dividing induced sequencing read counts by uninduced read counts. For MazF-mt10 in *E. coli*, we identified genomic positions for which the fold-change was ≥ 50. For all other MazFs we identified genomic positions for which the fold-change was ≥ 25. In addition, we required that the MazF-induced sample counts be above 50 to reduce skewing of the data due to low read counts. We

further sifted the data by only including reads that represented local maxima within a 50-base window spanning 25 bases up and downstream. If redundant sequences caused cleavage sites to map to more than one position in the genome, then we only counted the sequence surrounding each site of cleavage once. In the analysis of *E. coli* expressing MazF-mt7, MazF-mt10 and MazF-mt11 we identified 107, 627 and 13 positions, respectively, that met these criteria. In the analysis of *E. coli* expressing MazF-mt4, *E. coli* expressing MazF-mt5, *E. coli* expressing MazF-mt8 and *M. tuberculosis* expressing MazF-mt10, we did not identify any positions that met these criteria. Finally, we aligned the significant reads from each MazF using WebLogo to determine the consensus sequence for each MazF.

Results

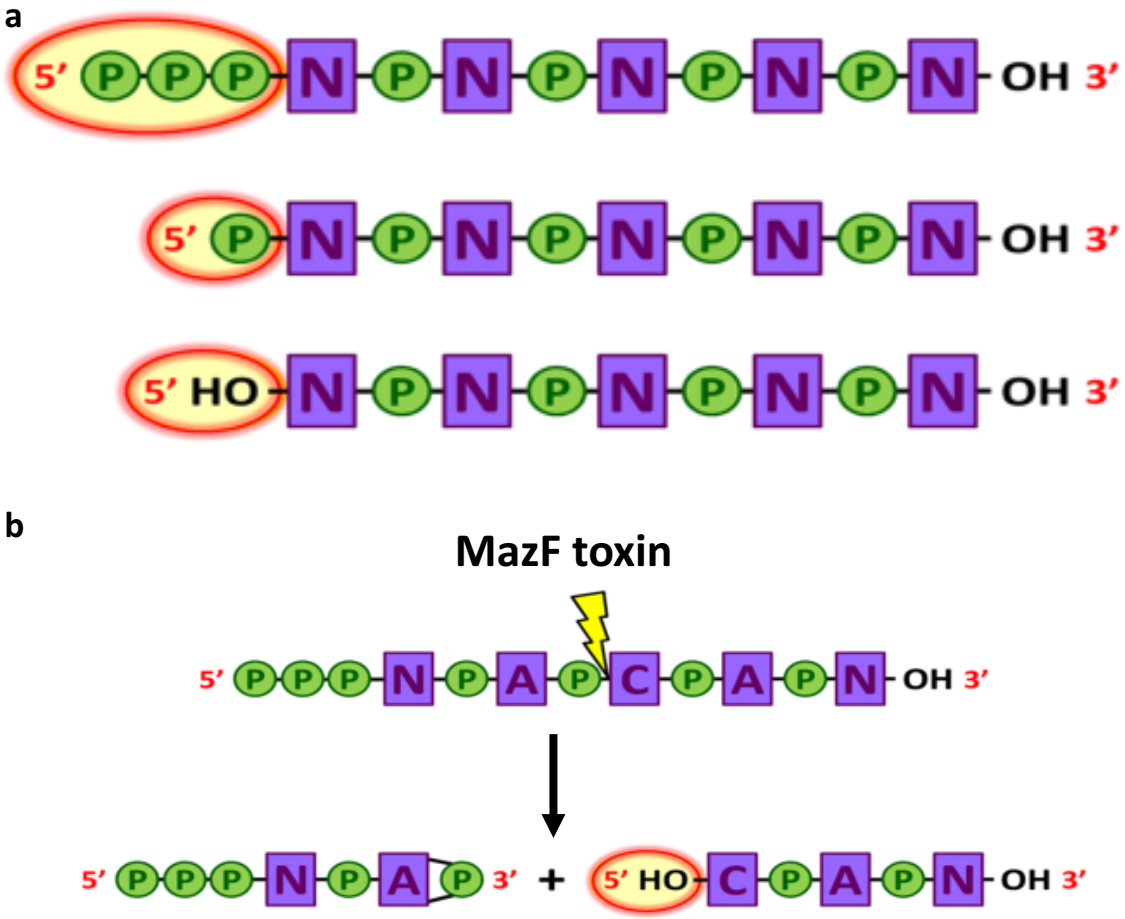
An RNA-seq approach to determine MazF RNA targets

Since MazF toxins are hypothesized to play a critical role in establishing and maintaining Mtb latency by cleaving RNA and thus modulating growth during periods of stress, we wanted to determine the RNA targets of some of the uncharacterized MazFs to inform us of how they might function in latent Mtb. To do this, we performed 5' RNA-seq, a specialized RNA-seq method developed by our laboratory, on 6 Mtb MazF toxins expressed *in vivo* in *E. coli* (MazF-mt4, MazF-mt5, MazF-mt7, MazF-mt8, MazF-mt10 and MazF-mt11) and one MazF toxin expressed *in vivo* in Mtb (MazF-mt10). 5' RNA-seq is a powerful tool because of its ability to differentiate between RNA products with distinctive 5' ends. Bacterial transcripts can possess 3 types of 5' ends: 5'-triphosphates (typical mRNAs), 5'-monophosphates (typical rRNAs and tRNAs) or 5'-hydroxyls (RNA cleaved by endonucleases such as MazF) (Figure 6) (Schifano et al., 2016; Zhang et al., 2005). 5' RNA-seq is able to select for RNA transcripts carrying a 5'-hydroxyl (5'OH), thus allowing for the detection and separation of transcripts that have been cleaved by MazFs.

In this study as well as previous studies, 5'RNA-seq was performed first in *E. coli* for several reasons. *E. coli* is a fast-growing bacterium that has a doubling time of ~30 minutes and only requires BSL1 containment (Chandler et al., 1975; Kimman et al., 2008). In contrast, Mtb is a fastidious, slow-growing bacterium that has a doubling time of 24 hours and requires BSL3 containment (BSL2 containment for the attenuated strain that we used) (Hsu et al., 2003; Kimman et

al., 2008; Sambandamurthy et al., 2002). RNA is highly conserved among bacteria, so cleavage of *E. coli* RNA is likely to be a good indicator of cleavage in Mtb. If cleavage targets are found in *E. coli*, results can be validated in Mtb using northern analysis, primer extension, and/or 5' RNA-seq.

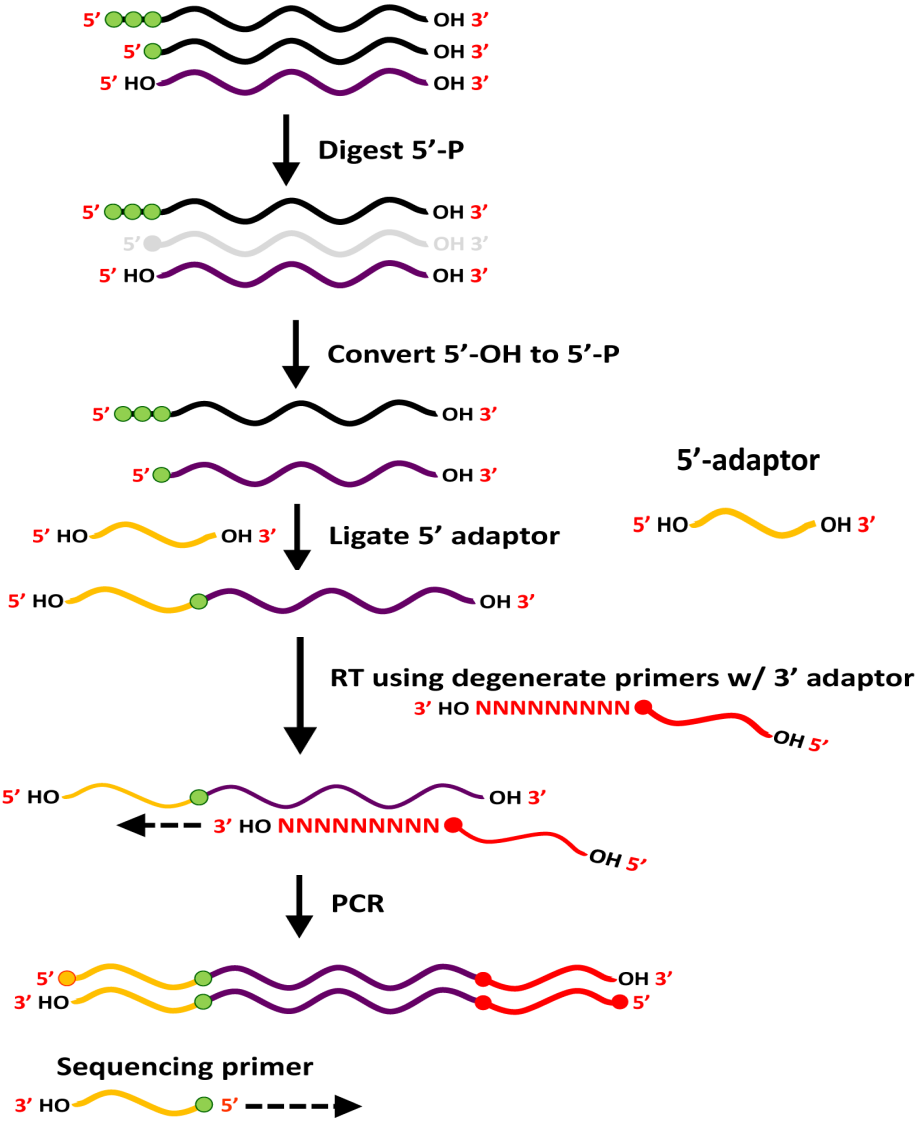
Figure 6. Bacterial transcripts have 3 distinct types of 5' ends.



a) The 5' ends of bacterial transcripts differ depending on the type of RNA: mRNAs have a 5'-triphosphate (top), tRNAs and rRNAs have a 5'-monophosphate (middle), and RNAs that have been cleaved by endonucleases have a 5'-hydroxyl (bottom). b) When MazF toxins cleave RNA, they leave a 5'-hydroxyl which can be detected by 5'RNA-seq. Adapted from Schifano et al. 2014.

To perform 5' RNA-seq on *E. coli* cells expressing individual MazF toxins *in vivo*, we transformed *E. coli* cells with an arabinose-inducible vector containing the MazF of interest, induced MazF expression with arabinose, and harvested total RNA from the cells when the cells began to show a decrease in growth as compared to controls (*E. coli* cultures containing the arabinose-inducible vector without MazF). RNA was harvested from experimental samples and controls at the same time. Next, we prepared the RNA for high-throughput sequencing (Figure 7). We digested RNA containing 5'-monophosphate (5'-P) ends, leaving only transcripts with 5'-OH and 5'-triphosphate (5'-3P) ends. Then, the 5'OH ends were converted to 5'-P ends because the 5'adaptor used in the next step can only anneal to 5'-P ends; this process can therefore isolate the cleaved RNA from what remains of the uncleaved RNA (5'-3P). After ligating the 5'adaptor, we performed reverse transcription to generate cDNA libraries, PCR amplified the cDNA libraries, and sent them for Illumina sequencing. We performed data analysis on the sequencing results to determine the RNA targets of each MazF. We required that there be a 25-fold change between experimental (induced MazF) and control samples, and that the induced count be 50 or above. We found potential RNA targets for MazF-mt4, MaxF-mt7, MazF-mt10 and MazF-mt1, but were unable to find RNA targets for MazF-mt5 and MazF-mt8 because none of read counts fell within the set limits. In addition, we aligned the reads for each MazF using WebLogo (Crooks et al., 2004) to determine whether each MazF cleaves at a unique consensus sequence. Surprisingly, we found that only MazF-mt10 showed cleavage at a clear consensus sequence.

Figure 7. Preparation of RNAs with a 5'-OH for Illumina sequencing.

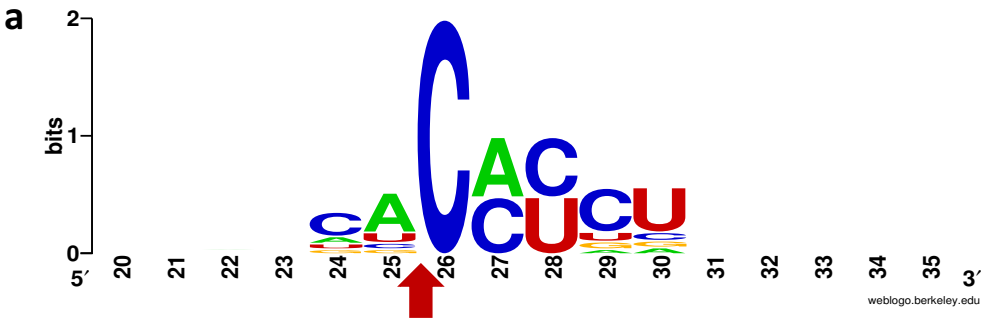


The procedure for preparing MazF-cleaved RNAs carrying a 5'-OH for Illumina sequencing are as follows: RNA with 5'-P ends are digested by a 5'-P-specific exonuclease; 5'-OH ends are then phosphorylated to 5'-Ps by a kinase and these ends are ligated to a 5'-adaptor; cDNAs are generated using reverse transcription (RT) and a primer with a 9-nt degenerate sequence at the 3' end and the sequence of the 3' adaptor at the 5' end; cDNAs are PCR amplified using primers that match the 5' and 3' adaptors; amplified cDNAs are sequenced using a primer that is complementary to the 5' adaptor. Thus, the first base of each sequencing read corresponds to the 5' end of an RNA transcript. The RNA that will be converted to cDNA and then sequenced is shown in purple, and phosphorylated ends are shown in green. Figure adapted from Schifano et al. 2014.

MazF-mt10 cleaves at consensus sequence CACCU

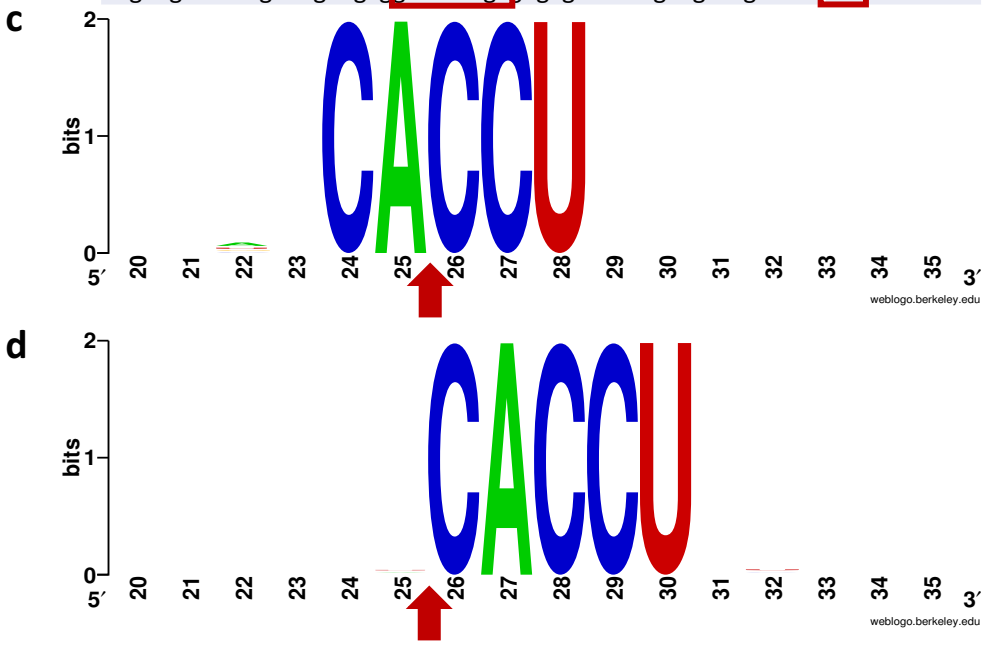
5’ RNA-seq of *E. coli* expressing MazF-mt10 revealed 627 hits with 50-fold or higher change and induced counts of 50 or above. Initially, alignment of all the RNA-seq reads for MazF-mt10 in *E. coli* did not reveal a clear consensus sequence (Figure 8a). However, further analysis revealed that the hits appeared to have a CACCU sequence at either the 24th or 26th position of the alignment (Figure 8b). When these two groups were divided and each group aligned separately, the WebLogo showed a clear consensus sequence of CACCU in each group (Figure 8c, 8d). This difference in the location of the consensus sequence indicated that while MazF-mt10 recognized one unique consensus sequence, it cleaved at two distinct locations, one right before the consensus sequence (↓CACCU) and one within the consensus sequence (CA↓CCU). This 2 nt cleavage site variability within the consensus sequence has not been observed for any other bacterial MazF toxin to date.

Figure 8. MazF-mt10 cleaves at two locations within the CACCU consensus.



b

Sequence	CACCU Position
gagggcguuguuguucguuaauaCACCUacuuugagccgguucacacuuuu	24
ucgccaucaaccagagaucgccacaCACCUcggugcugacguuaucuucac	26
uagaagcgaucgaccgucugcgugaCACCUcuucuucucaccagcguauuu	26
ccggcgcguuucguguuaacgacgcCACCUucuaugaaaaauauuacggcu	26
uggacguguaguaguuuaaguuuuCACCUgcacuguagaccggauaaggcau	24
cucaaggcggacgcaacggauaaCACCUcgcucgcugagcaucucagugaa	24
gcgugaauaucuggggugacgguagCACCUacaaaggaaacgauaucgaac	26
cugcaaaccagugcaaaagaaauuuCACCUgacuucaucaaugugaaaggg	26
gaugcugaugcggcggaucugauCACCUccgauugcgacccguaugacagc	24
uagaugaucaacguaaguagaggCACCUgagugagcaaucugucgcucgau	24

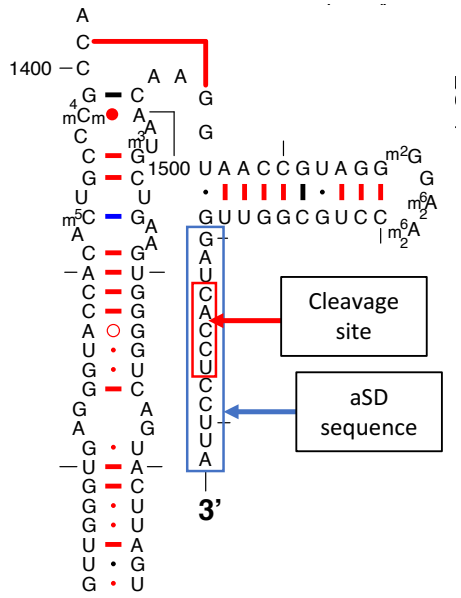


Cleavage occurs after the 25th position on the WebLogo and is marked with a red arrow. a) Alignment of all 627 5'RNA-seq hits revealed an unclear consensus sequence around the cleavage position. b) Further examination revealed a CACCU at either the 24th or 26th position. c, d) separation and alignment of these two groups showed a clear consensus of CACCU and cleavage at CA↓CCU and ↓CACCU, respectively.

MazF-mt10 cleaves 16S rRNA at the anti-Shine-Dalgarno sequence

Most of the 627 hits for MazF-mt10 in *E. coli* were mRNAs; however, there were 3 different 16S rRNA genes among the top hits: *rrnE*, *rrnG*, and *rrnD*, with 210-, 206-, and 205-fold-changes, respectively. Cleavage of these 16S rRNAs all occurred directly before (5' of) the anti-Shine-Dalgarno (aSD) sequence. Another member of the Woychik lab, Valdir Barth, also performed 5' RNA-seq of MazF-mt10 expressed in *Mycobacterium smegmatis*, a non-pathogenic environmental bacterium in the same genus as Mtb. This analysis revealed several potential RNA targets in *M. smegmatis*, but there was no clear consensus sequence when the *M. smegmatis* reads were aligned. We searched against the *M. smegmatis* reads for the consensus sequence found in *E. coli*, and found only one hit: 16S rRNA, cleaved before the aSD sequence. The cleavage of 16S rRNA in *M. smegmatis* occurred at the CA↓CCU (Figure 9).

Figure 9. MazF-mt10 cleaves within the aSD sequence of 16S rRNA.

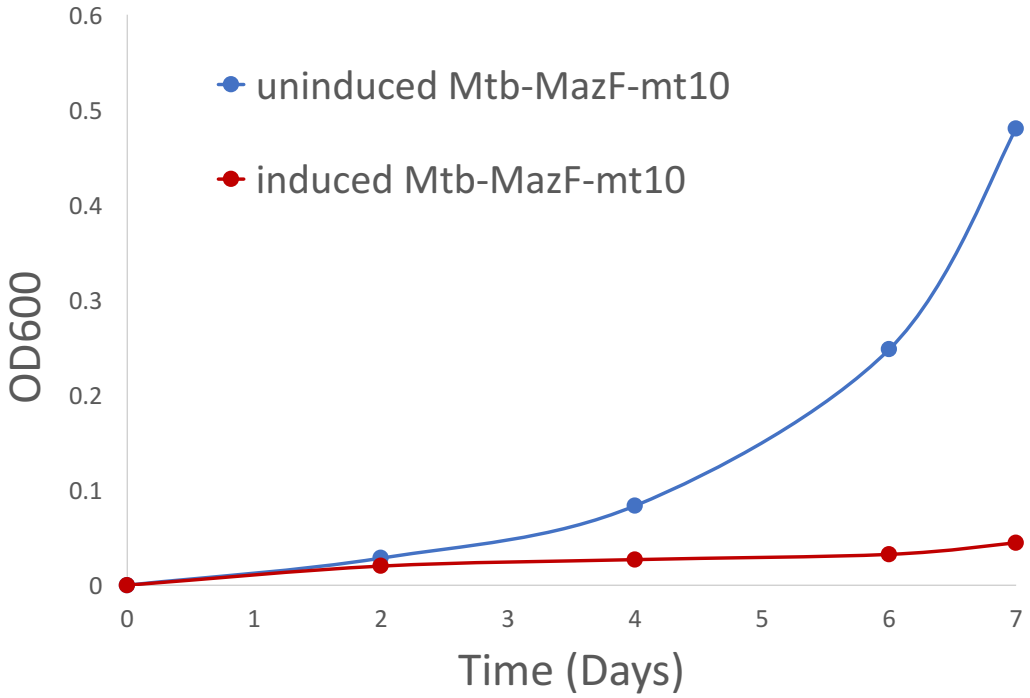


In *M. smegmatis* and *E. coli*, MazF-mt10 cleaves within the aSD sequence (boxed in blue) of 16S rRNA at the CA↓CCU consensus sequence (boxed in red). Figure adapted from The Comparative RNA Web (CRW) Site, http://www.rna.icmb.utexas.edu/ (Cannone et al., 2002).

The aSD sequence is a highly conserved, functionally important region within 16S rRNA. Therefore, cleavage of the aSD region in *E. coli* and *M. smegmatis* suggests that MazF-mt10 also cleaves 16S rRNA in Mtb. Therefore, we repeated the 5' RNA-seq with Mtb samples that were overexpressing MazF-mt10. To do this, we transformed Mtb with an anhydrotetracycline-inducible plasmid, grew the transformants in liquid media, split the culture, and induced one culture, while allowing the other culture to grow uninduced as a control. Total RNA was harvested from both cultures after 3 days, and the RNA prepared for sequencing the same way the *E. coli*. samples were. However, 5' RNA-seq did not reveal any differences in cleavage patterns between induced and uninduced samples for MazF-mt10 in Mtb.

When RNA was harvested from the induced and uninduced Mtb-MazF-mt10 cultures, a portion of the culture was grown for 4 more days (i.e. to day 7 post induction) to determine whether induction of the toxin inhibited growth. Growth profiles, measured by optical density (OD_{600}), did not show any difference in growth between induced and uninduced cultures (data not shown). This was not surprising, since MazF-mt10 only partially inhibits growth in *E. coli*, and MazF-mt3, another 16S rRNA cleaver, has only shown slight growth inhibition in Mtb (data not shown). However, the growth profile of MazF-mt10 in Mtb was repeated, and this time revealed clear growth arrest in the induced culture (Figure 10).

Figure 10. MazF-mt10 causes growth arrest in Mtb.



Mtb cells harboring an ATC-inducible plasmid containing MazF-mt10 were induced with ATC and OD readings taken to measure growth over a 7-day period. Mtb overexpressing MazF-mt10 inhibited growth (red line) as compared to the uninduced control (blue line).

MazF-mt11 and MazF-mt4 cleave tRNA

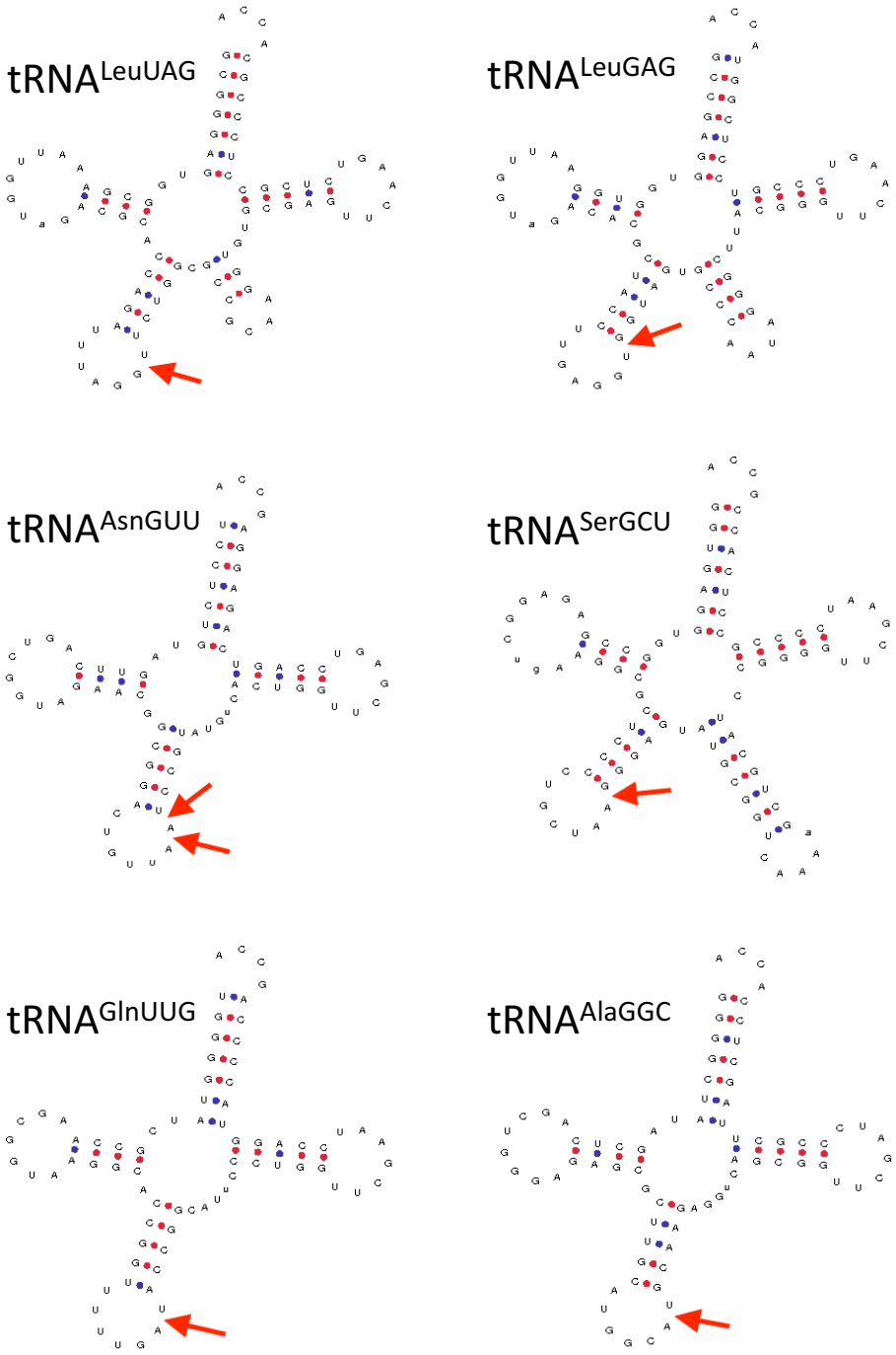
5' RNA-seq of *E. coli* expressing MazF-mt11 revealed 13 hits with 25-fold or higher change and induced counts of 50 or above. Out of these hits, 9 of them mapped to known genes in the *E. coli* genome, and 8 of them coded for tRNAs (Table 1). 5'-RNA-seq analysis further indicated that MazF-mt11 cleaved six tRNAs at their anticodon stem loop (ASL): $tRNA^{LeuUAG}$, $tRNA^{LeuGAG}$, $tRNA^{AsnGUU}$, $tRNA^{SerGCU}$, $tRNA^{GlnUUG}$, and $tRNA^{AlaGGC}$ (Figure 11). While 5 of the tRNAs are cleaved at only one location in a single-stranded region adjacent to their anticodon, $tRNA^{AsnGUU}$ is cleaved at two locations, both within the ASL. This suggests that MazF-mt11, like MazF-mt9, may also primarily target tRNA.

Table 1. MazF-mt11-mediated tRNA cleavage sites in *E. coli*.

	Gene Product	Gene	Fold Change	Sequence position
1	tRNA-Leu (UAG)	leuW	58	38
2	tRNA-Ile (GAU)	ileV	53	4
3	tRNA-Leu (GAG)	leuU	31	39
4	tRNA-Asn (GUU)	asnV*	26	37
5	tRNA-Ser (GCU)	serV	25	39
6	tRNA-Asn (GUU)	asnV*	25	38
7	tRNA-Gln (UUG)	glnW	25	36
8	tRNA-Ala (GGC)	alaX	25	37

***$tRNA^{AsnGUU}$ was cleaved at two positions, both within the ASL.**

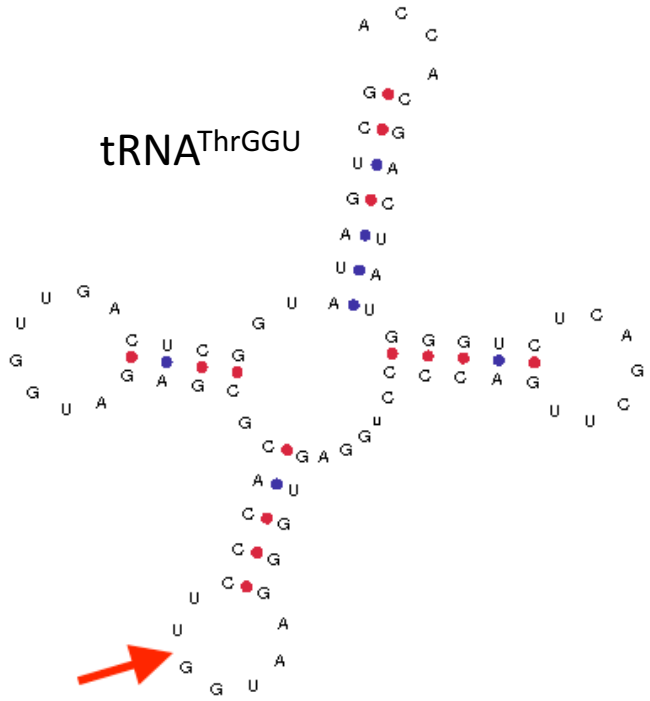
Figure 11. MazF-mt11 cleaves 6 *E. coli* tRNAs at their anticodon stem loop.



MazF-mt11 targets six tRNAs for cleavage within their anticodon stem loop. In each case, cleavage occurs in a single-stranded region of RNA adjacent to the anticodon. Cleavage of tRNA^{AsnGUU} occurs in two locations next to each other, while cleavage of the other tRNAs occurs at only one location each. Red arrows indicate the position of cleavage. Figures adapted from GtRNAdb, http://gtrnadb.ucsc.edu/ (Chan and Lowe, 2009).

5’ RNA-seq of *E. coli* expressing MazF-mt4 did not reveal any hits with both 25-fold or higher change and induced counts of 50 or above. However, there were only three hits in total, and two of these mapped to known genes in the *E. coli* genome and had fold-changes/induced counts close to the threshold. These genes code for two identical tRNA^{ThrGGU}s – *thrV*, with a fold-change of 32 and an induced count of 32 (vs. uninduced of 1), and *thrT*, with a fold-change of 24 and an induced count of 72 (vs. uninduced of 3) – with cleavage at the same location in the ASL for both (Figure 12). This suggests that MazF-mt4 may also be a tRNA cleaver.

Figure 12. MazF-mt4 cleaves tRNA^{ThrGGU}.



MazF-mt4 targets 2 identical E coli tRNAs for cleavage, *thrV* and *thrT*, within their anticodon stem loop. Figure adapted from GtRNAdb, http://gtrnadb.ucsc.edu/ (Chan and Lowe, 2009).

MazF-mt7 cleaves ribosomal proteins, tRNA, and 16S and 23S rRNA

5' RNA-seq of MazF-mt7 revealed 107 hits with 25-fold or higher change and induced counts of 50 or above. Out of these hits, 96 of them mapped to known genes in the *E. coli* genome (Table 2). 3 of the hits were identified as cleavage locations within tRNA, 14 of the hits were identified as cleavage locations within 23S rRNA, and 24 of the hits were identified as cleavage locations within 16S rRNA. All cleavage of RNA occurred within single-stranded regions. Interestingly, the top 6 hits, and 33 hits in total showed cleavage of ribosomal protein mRNA. Cleavage of all three tRNAs occurred within the anticodon stem loop. The tRNAs cleaved were $tRNA^{LeuUAG}$ and two identical versions of $tRNA^{AlaUGC}$ that were cleaved at the same position (Figure 13). 23S rRNA was cleaved at 5 positions, two of them within helix 101, and one each within helices 38, 55, and 79 (Figure 14). The two cleavage positions in helix 101 are directly adjacent to each other. 16S rRNA was cleaved at 9 locations, 2 in helix 23, 2 in helix 27, 1 at helix 43, 1 in helix 44, and 3 in helix 45, which contains the aSD sequence (Figure 15).

Table 2. MazF-mt7-mediated cleavage sites in *E. coli*

	Gene Product	Gene	Fold Change	Sequence position
1.	50S ribosomal protein L14	rplN	126	1
2.	30S ribosomal protein S5	rpsE	90	20
3.	30S ribosomal protein S5	rpsE	69	317
4.	30S ribosomal protein S12	rpsL	64	322
5.	50S ribosomal protein L10	rplJ	61	112
6.	30S ribosomal protein S12	rpsL	59	323
7.	primosomal replication protein N	priB	56	94
8.	preprotein translocase subunit SecY	secY	55	451

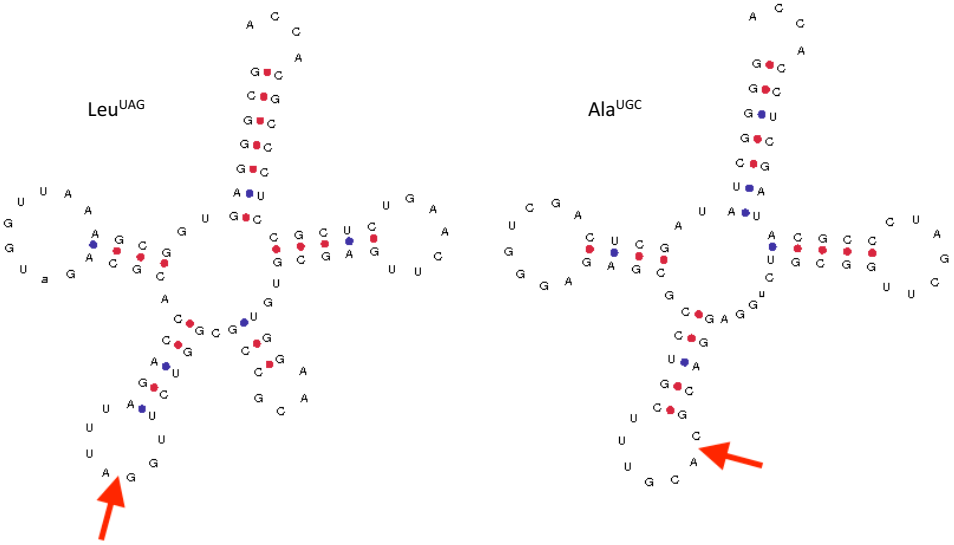
9.	elongation factor G	fusA	54	1849
10.	tRNA-Leu (UAG)	leuW	53	36
11.	50S ribosomal protein L15	rplO	53	50
12.	primosomal replication protein N	priB	52	92
13.	ATP-dependent RNA helicase DeaD	deaD	52	257
14.	elongation factor G	fusA	51	146
15.	DNA-directed RNA polymerase subunit alpha	rpoA	51	548
16.	30S ribosomal protein S7	rpsG	51	317
17.	30S ribosomal protein S1	rpsA	50	1665
18.	16S rRNA-processing protein RimM	rimM	50	271
19.	elongation factor G	fusA	49	147
20.	elongation factor G	fusA	48	148
21.	23S ribosomal RNA of rrnD operon	rrlD	47	2883
22.	23S ribosomal RNA of rrnH operon	rrlH	43	2213
23.	23S ribosomal RNA of rrnD operon	rrlD	43	2213
24.	23S ribosomal RNA of rrnE operon	rrlE	42	2213
25.	23S ribosomal RNA of rrnB operon	rrlB	42	2213
26.	tRNA	trmD	41	391
27.	elongation factor G	fusA	41	116
28.	16S ribosomal RNA of rrnG operon	rrsG	39	621
29.	DNA-directed RNA polymerase subunit alpha	rpoA	38	642
30.	23S ribosomal RNA of rrnE operon	rrlE	38	918
31.	16S ribosomal RNA of rrnH operon	rrsH	38	621
32.	16S ribosomal RNA of rrnE operon	rrsE	38	621
33.	lipoprotein NlpI	nlpI	37	559
34.	F0F1 ATP synthase subunit alpha	atpA	37	1278
35.	50S ribosomal protein L29	rpmC	37	26
36.	50S ribosomal protein L14	rplN	37	0
37.	50S ribosomal protein L1	rplA	37	542
38.	23S ribosomal RNA of rrnH operon	rrlH	36	918
39.	16S ribosomal RNA of rrnH operon	rrsH	36	919
40.	16S ribosomal RNA of rrnD operon	rrsD	36	621
41.	16S ribosomal RNA of rrnD operon	rrsD	35	1535
42.	50S ribosomal protein L15	rplO	34	11
43.	30S ribosomal protein S2	rpsB	34	542
44.	30S ribosomal protein S11	rpsK	34	44
45.	30S ribosomal protein S1	rpsA	34	1391
46.	16S ribosomal RNA of rrnD operon	rrsD	34	919
47.	tRNA-Ala (UGC)	alaU	33	37
48.	50S ribosomal protein L1	rplA	33	552

49.	16S ribosomal RNA of rrnH operon	rrsH	33	1433
50.	16S ribosomal RNA of rrnG operon	rrsG	33	919
51.	50S ribosomal protein L22	rplV	32	124
52.	30S ribosomal protein S8	rpsH	32	286
53.	23S ribosomal RNA of rrnB operon	rrlB	32	918
54.	16S ribosomal RNA of rrnG operon	rrsG	32	1433
55.	16S ribosomal RNA of rrnE operon	rrsE	32	919
56.	16S ribosomal RNA of rrnE operon	rrsE	32	1532
57.	30S ribosomal protein S9	rpsI	31	136
58.	16S rRNA-processing protein RimM	rimM	31	415
59.	50S ribosomal protein L2	rplB	30	450
60.	23S ribosomal RNA of rrnD operon	rrlD	30	918
61.	16S ribosomal RNA of rrnG operon	rrsG	30	620
62.	16S ribosomal RNA of rrnE operon	rrsE	30	620
63.	16S ribosomal RNA of rrnE operon	rrsE	30	1433
64.	16S ribosomal RNA of rrnD operon	rrsD	30	620
65.	elongation factor G	fusA	29	290
66.	50S ribosomal protein L14	rplN	29	199
67.	30S ribosomal protein S4	rpsD	29	135
68.	30S ribosomal protein S11	rpsK	29	108
69.	23S ribosomal RNA of rrnH operon	rrlH	29	1571
70.	23S ribosomal RNA of rrnD operon	rrlD	29	1571
71.	23S ribosomal RNA of rrnB operon	rrlB	29	1571
72.	16S ribosomal RNA of rrnH operon	rrsH	29	620
73.	16S ribosomal RNA of rrnG operon	rrsG	29	920
74.	translation initiation factor IF-3	infC	28	143
75.	50S ribosomal protein L22	rplV	28	122
76.	50S ribosomal protein L14	rplN	28	161
77.	16S ribosomal RNA of rrnD operon	rrsD	28	1433
78.	tRNA-Ala (UGC)	alaV	27	37
79.	30S ribosomal protein S5	rpsE	27	316
80.	30S ribosomal protein S1	rpsA	27	1664
81.	23S ribosomal RNA of rrnE operon	rrlE	27	1571
82.	16S ribosomal RNA of rrnG operon	rrsG	27	1505
83.	30S ribosomal protein S5	rpsE	26	19
84.	30S ribosomal protein S5	rpsE	26	331
85.	16S ribosomal RNA of rrnD operon	rrsD	26	1320
86.	preprotein translocase subunit SecY	secY	25	1192
87.	elongation factor G	fusA	25	289
88.	50S ribosomal protein L22	rplV	25	219
89.	50S ribosomal protein L19	rplS	25	149

90.	30S ribosomal protein S5	rpsE	25	78
91.	30S ribosomal protein S14	rpsN	25	117
92.	23S ribosomal RNA of rrnD operon	rrlD	25	2882
93.	16S rRNA-processing protein RimM	rimM	25	272
94.	16S ribosomal RNA of rrnH operon	rrsH	25	920
95.	16S ribosomal RNA of rrnE operon	rrsE	25	1320
96.	16S ribosomal RNA of rrnD operon	rrsD	25	920

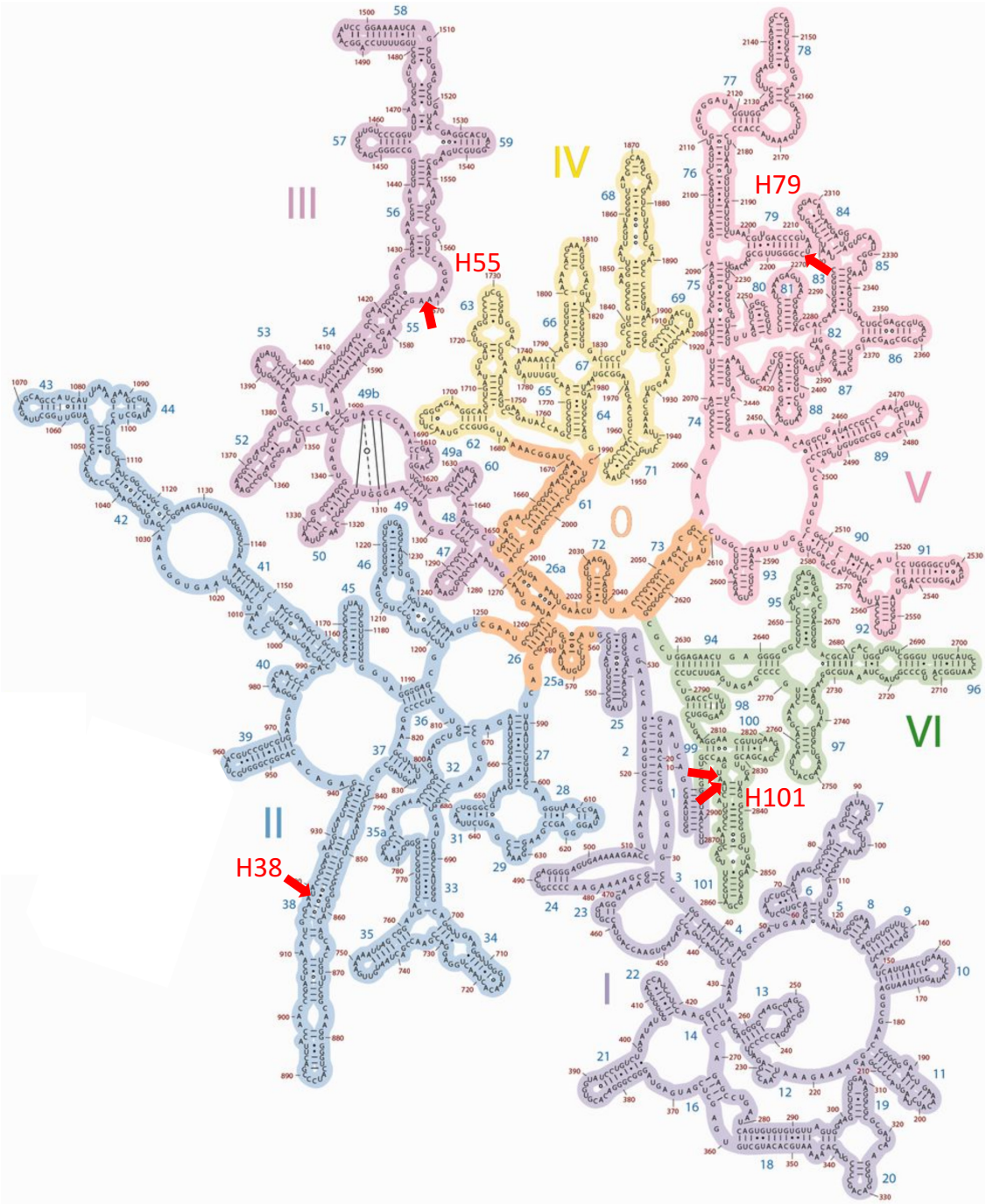
Ribosomal protein cleavage is in orange, tRNA cleavage is in purple, 23S rRNA cleavage is in blue, and 16S rRNA cleavage is in green.

Figure 13. MazF-mt7 in *E. coli* cleaves tRNA.



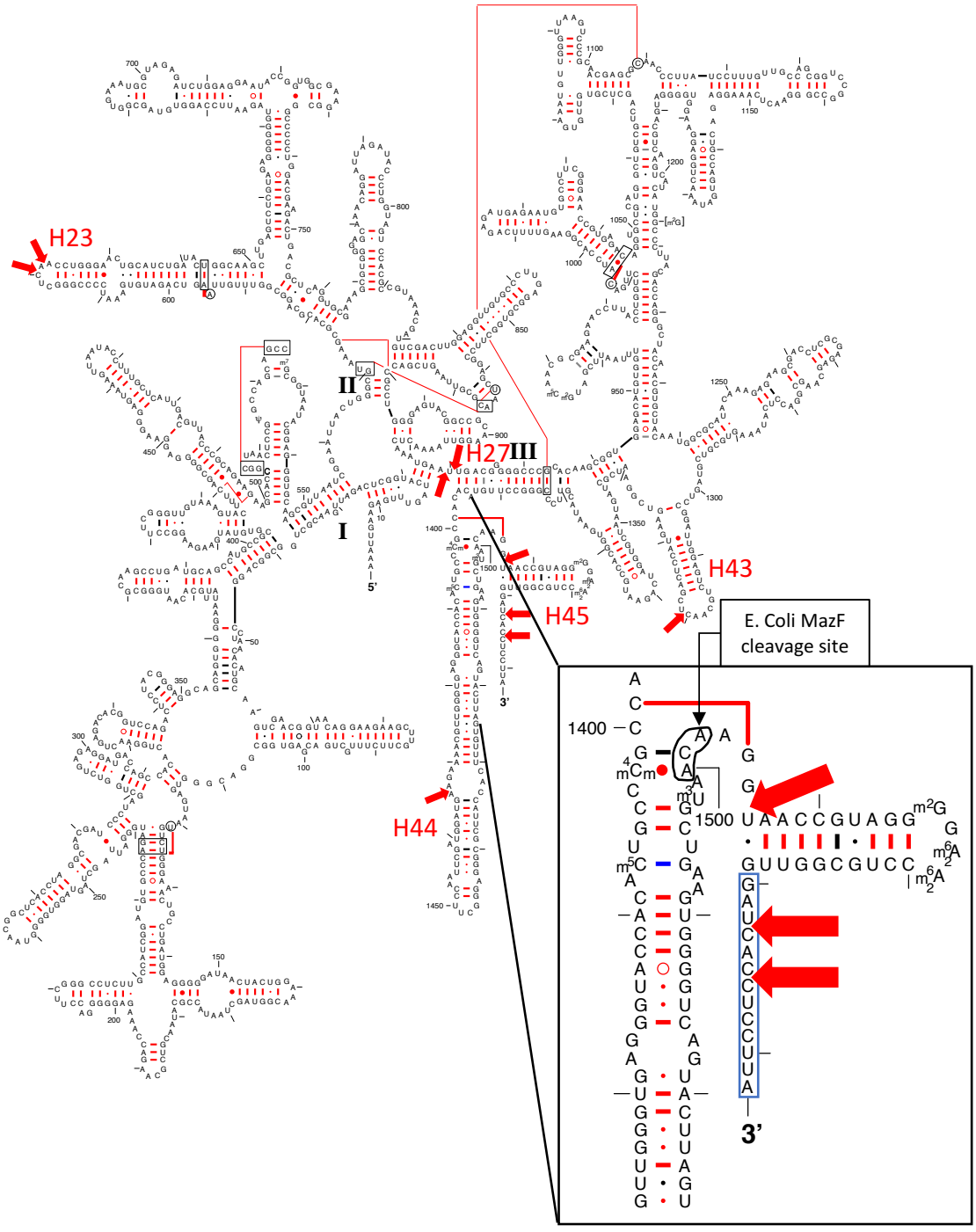
MazF-mt7 cleaves three tRNAs at the anticodon stem loop. It cleaves Leu^{UAG} and two identical versions of Ala^{UGC}.

Figure 14. MazF-mt7 cleaves 23S rRNA at single-stranded regions.



MazF-mt7 cleaves at one position within helices 38, 55 and 79, and two positions next to each other within helix 101. Cleavage positions are marked with red arrows. Figure adapted from Petrov et al. 2013.

Figure 15. MazF-mt7 cleaves single-stranded regions of 16S rRNA.



MazF-mt7 cleaves at 9 distinct locations in 16S rRNA, all of them within single-stranded regions. 3 of the cleavage sites are at or near the anti-Shine-Dalgarno sequence (aSD). The region that contains the aSD is enlarged, and the aSD is boxed in blue. Red arrows indicate positions of cleavage. Figure adapted from The Comparative RNA Web (CRW) Site, http://www.rna.icmb.utexas.edu/ (Cannone et al., 2002).

Discussion

For years, MazF toxins have been known as “mRNA interferases,” thought to cleave only mRNA (Zhang et al., 2005; Zhang et al., 2003; Zhu et al., 2008; Zhu et al., 2006). However, this work, in combination with previously published work, presents striking evidence otherwise. It appears that at least 7 of the 11 Mtb MazFs cleave other RNA species in addition to mRNA in Mtb (Schifano et al., 2016; Schifano et al., 2013; Schifano et al., 2014). There is evidence that four MazF toxins (MazF-mt4, MazF-mt7, MazF-mt9 and MazF-mt11) cleave tRNA, and four MazF toxins (MazF-mt3, MazF-mt6, MazF-mt7 and MazF-mt10) cleave rRNA. These findings indicate that MazF toxins may regulate growth during periods of stress by regulating translation through concerted cleavage of all three major forms of RNA that have a hand in this essential cellular process.

tRNA cleavage

5’ RNA-seq showed MazF-mt11 cleavage of 5 *E. coli* tRNAs at their single-stranded anticodon stem loop. tRNA cleavage by MazF toxins was recently discovered when Schifano and colleagues found that MazF-mt9 cleaves Mtb $tRNA^{Pro14}$ within the D-loop and $tRNA^{Lys43}$ within the anticodon stem loop (Schifano et al., 2016) and arrests translation and growth in *E. coli* and *Mtb*. In addition, cleavage of tRNA has been seen by other Mtb TA toxins. Another family of TA toxins with at least 48 members in Mtb, VapC, has been shown to contain at least 11 paralogs that cleave tRNA (Cruz et al., 2015; Winther et al., 2016). As a result, we thought of VapCs as the primary cleavers of tRNA; evidence from

this study and the Schifano et al. study of MazF-mt9 challenges this notion (Schifano et al., 2016). The ability of TA toxins, now including MazFs, to target tRNA for cleavage indicates that MazFs are able to regulate translation via mechanisms other than simple depletion of mRNA.

The tRNAs cleaved in the present study are *E. coli* tRNAs; however, precedent from previous studies of bona fide tRNA-cleaving Mtb TA toxins has shown that toxin cleavage of tRNAs found in *E. coli* strongly suggests TA toxin cleavage of Mtb tRNAs as well (Cruz et al., 2015; Cruz and Woychik, 2016; Schifano et al., 2016). For example, Schifano et al. found that MazF-mt9 cleaves 6 tRNAs in *E. coli* before discovering that MazF-mt9 cleaves two tRNAs in its native context of Mtb (Schifano et al., 2016). Cruz et al. also found that VapC-mt4 cleavage in *E. coli* was also an indicator of tRNA cleavage in Mtb (Cruz et al., 2015). Thus, although it is possible that MazF-mt11 may not cleave as many tRNA targets in Mtb as it did in *E. coli*, it is likely that it does indeed cleave Mtb tRNA. To confirm this and find the exact tRNA target(s) in Mtb, 5'RNA-seq or northern analysis should be done on Mtb tRNAs that have been pre-incubated with MazF-mt11 and/or Mtb cells overexpressing MazF-mt11.

Although results suggest that MazF-mt11 cleaves tRNA, they did not reveal a consensus sequence for MazF-mt11 cleavage. While all previously characterized MazF toxins are sequence-specific endoribonucleases, MazF-mt9 is the first MazF toxin that has been shown to also require a structural component for cleavage (Schifano et al., 2016). Thus, it is possible that some MazF toxins, including MazF-mt11, may not actually require a consensus

sequence but may cleave based on RNA secondary structure. Conversely, MazF-mt11 may have a consensus sequence that was missed due to the small 5' RNA-seq data set (only 11 hits). This small data set could have been a consequence of the time at which RNA was harvested for 5' RNA-seq. Because MazF-mt11 does not strongly inhibit growth in *E. coli*, it is difficult to determine exactly when the toxin is most active and therefore when to harvest RNA. Thus, 5' RNA-seq of MazF-mt11 in *E. coli* should be repeated with RNA isolated at a number of different time points within the MazF-mt11-induced growth profile to help decipher whether MazF-mt11 has a consensus sequence that was missed due to the limitations of our initial study.

Similarly, although 5' RNA-seq analysis of MazF-mt4 did not reveal results within our statistically significant threshold, it did identify cleavage of two identical $tRNA^{Trp}$s, indicating that MazF-mt4 may also cleave tRNA in Mtb. As with MazF-mt11, MazF-mt4 may not cleave the same tRNA in Mtb, and further studies will be performed to precisely determine the Mtb tRNA(s) cleaved by MazF-mt4.

One possible reason why MazF-mt4 did not produce higher read numbers is that the total RNA samples used for MazF-mt4 were harvested after the induced cells had been diluted and regrown to exponential phase, rather than right at or shortly after the point when cells started to show growth inhibition. As a result, the toxin activity may have started to wane, giving fewer cleavage products. This could also explain why MazF-mt5 and MazF-mt8 in *E. coli* did not reveal any significant differences in cleavage specificity between induced and uninduced samples. The 5' RNA-seq for MazF-mt4, -mt5, and –mt8 in *E. coli* will be repeated by an

incoming visiting scholar with total RNA from various time-points in the growth profile to determine whether cleavage patterns change over time. In addition, 5'RNA-seq should also be repeated in Mtb to validate the findings in the natural host.

23S rRNA cleavage

5'-RNA seq showed MazF-mt7 cleavage of five distinct single-stranded regions within 23S rRNA. This is the third Mtb MazF found to cleave 23S rRNA, after MazF-mt3, which cleaves at a single U↓CCUU site in helix/loop 70, and MazF-mt6, which also cleaves in helix/loop 70, but at a UU↓CCU recognition sequence (Schifano et al., 2013; Schifano et al., 2014). Helix/loop 70 of 23S rRNA, found in domain IV, is a critical part of the ribosome that facilitates the binding of tRNA in the ribosomal A site, stabilizes the ribosome recycling center, and aids in the 50S-30S subunit association (Ban et al., 2000; Leviev et al., 1995; Merryman et al., 1999). Cleavage of this critical site inhibits translation and growth (Schifano et al., 2013; Schifano et al., 2014).

In addition to MazF-mt3, MazF-mt6 and MazF-mt7, there are other TA toxins in Mtb that also cleave 23S rRNA. For example, VapC-mt20 cleaves the sarcin-ricin loop of 23S rRNA, another essential region of 23S rRNA, causing growth arrest (Winther et al., 2013). While MazF-mt7 does not cleave at helix/loop 70 or the sarcin-ricin loop, it does cleave at another crucial region, helix/loop 38. Helix/loop 38 is called the "A-site finger" because it directly interacts with tRNA in the A-site (Komoda et al., 2006). This adds another

potential mechanism by which MazF activity alters or inactivates ribosome activity, thereby downregulating translation and growth.

The 23S rRNA cleaving TA toxins mentioned, including MazF-mt7, share several commonalities. They all cleave in single-stranded regions, and they all cleave regions of 23S rRNA whose functions are critical for the activity of the ribosome (Schifano et al., 2013; Schifano et al., 2014; Winther et al., 2013). MazF-mt7 also cleaves within 4 other single-stranded regions of 23S rRNA whose functions are unknown; further study of the ribosome may lead to discovery of important functions within these regions. Conversely, like for tRNA, cleavage at 5 locations in *E. coli* 23S rRNA could have been a result of expressing the toxin outside its native host; the actual number of cleavage sites in Mtb may differ, and the results found in this study will be validated in follow on studies in Mtb.

16S rRNA cleavage

5’ RNA-seq indicated MazF-mt10 cleavage of 16S rRNA in both *E. coli* and *M. smegmatis*. Cleavage occurred at the MazF-mt10 CA↓CCU consensus sequence within the anti-Shine-Dalgarno (aSD) region of 16S rRNA, only a few base pairs from the 3’ end of the RNA. In addition to studies in the Moll lab that discovered MazF cleavage of 16S rRNA in *E. coli*, this is the second Mtb MazF with evidence of cleaving 16S rRNA within the aSD region (Schifano et al., 2016; Vesper et al., 2011). In 2014, Schifano et al. showed that MazF-mt3 cleaves within the aSD region of 16S rRNA, but at a cleavage position slightly different

than MazF-mt10 (U↓CCUU rather than CA↓CCU) (Schifano et al., 2014). In either case, cleavage results in the removal of the aSD region.

The aSD region is critical for canonical translation, as it binds to the Sine-Dalgarno (SD) sequence that is slightly upstream of the start codon of canonical, leadered mRNAs, thus aiding in translation initiation. If the aSD is removed, translation of canonical mRNAs cannot occur (Vesper et al., 2011). However, Vesper et al found that in *E. coli*, MazF cleavage of the aSD sequence results in translation of a subpopulation of non-canonical mRNAs that are leaderless, meaning that they lack part or all of the SD sequence (Vesper et al., 2011). Remarkably, two recent studies found that approximately 25% of the Mtb transcriptome is leaderless (Cortes et al., 2013; Shell et al., 2015). It is possible that these naturally leaderless mRNAs could be preferentially translated after cleavage of the aSD sequence, resulting in an altered population of proteins that better enable Mtb cells to cope with stress. Further studies of the functions of leaderless genes in Mtb and their interactions with MazF-cleaved ribosomes are needed in order to determine the extent to which MazF cleavage of 16S rRNA regulates translation in response to stress.

To confirm that MazF-mt10 also cleaves the aSD in Mtb, we performed 5' RNA-seq on total RNA from Mtb overexpressing MazF-mt10. However, we were unable to fund any significant results from this analysis. This is most likely due to suppressor mutations that occurred in the Mtb-MazF-mt10 culture, rendering MazF-mt10 no longer toxic. Originally, we thought that MazF-mt10 did not inhibit growth in Mtb when overexpressed; as a result, we isolated total RNA from an

Mtb-MazF-mt10 culture that was not showing growth inhibition, because it is possible for a toxin to be active without showing growth inhibition. We later discovered that MazF-mt10 does indeed show growth inhibition in Mtb, but can easily acquire mutations that suppress the activity of the toxin (data not shown). Thus, the 5'RNA-seq should be repeated with an Mtb-MazF-mt10 sample that is clearly toxic in order to get an accurate representation of MazF-mt10 RNase activity in Mtb cells.

5' RNA-seq also showed cleavage of 16S rRNA by MazF-mt7. While the cleavage positions found were scattered throughout 16S rRNA, the majority of cleavage positions were within or very close to the aSD region, indicating that MazF-mt7 may also help modulate translation by removing the aSD sequence. Distinct from the other MazFs studied to date, MazF-mt7 showed cleavage of all three major forms of RNA: tRNA, 23S rRNA, 16S rRNA. However, MazF-mt7 cleavage did not require a consensus sequence, despite having over 100 sequencing reads; these results differ from a limited primer extension study by Zhu et al. which found that MazF-mt7 cleaves at U↓CGCU. On the other hand, Zhu et al. found that cleavage specificity is less stringent for MazF-mt7 than for MazF-mt3, with some cleavage sites that contain a one- or two-base mismatch (Zhu et al., 2008). In addition, Zhu and colleagues' study of MazF-mt7 only included 23 mRNAs; this small sample size could have resulted in a consensus sequence that did not represent the entire population of RNA that was cleaved. A larger sample size, such as the one in our study, could provide a more accurate representation of the MazF-mt7 cleavage data. The findings that MazF-mt7

cleaved all major forms of RNA but did not cleave at a consensus sequence strongly suggest that MazF-mt7 cleavage is determined by structural determinants rather than consensus sequence. More specificity may also be imparted by two novel mechanisms. First, an additional factor may bind to MazF-mt7 in its native Mtb host to impart more target specificity. Alternatively, MazF-mt7 may bind to the translating ribosome and undergo an allosteric shift which facilitates target specificity.

Interestingly, the top hits for MazF-mt7 were not RNAs, but various ribosomal proteins. The structure of ribosomal protein mRNAs is not as well-defined as the structures of tRNA and rRNA, but at least one study found that some parts of ribosomal proteins have extensive secondary structure (Mackie, 1992). If structure is the determinant for cleavage by MazF-mt7, MazF-mt7 may be recognizing and cleaving at a similar structural motif within tRNA, 23S rRNA, 16S rRNA and ribosomal proteins. It is possible that ribosomal proteins are another MazF target as a means to regulate translation in response to stress. If so, it would represent a novel and potent multi-pronged attack on essential components of the translation machinery by a TA toxin. In order to further elucidate whether structure is a cleavage determinant for MazF-mt7, 5' RNA-seq will be repeated in Mtb by Woychik lab members, and further structural comparisons between the cleavage sites will be performed.

It is unclear why some of Mtbs MazF toxins appear to have redundant functions (e.g. cleavage of the aSD by MazF-mt3, MazF-mt7, and MazF-mt10). It is possible that these toxins are coactivated in order to rapidly and potently inhibit

growth of Mtb cells during periods of stress. On the other hand, each MazF toxin could be activated by different stressors. At this time, the triggers for Mtb MazF toxin activation are not known. Further studies are needed to illuminate the interaction of each of these MazF toxins, as well as other TA toxins in Mtb, and to elucidate their role in the establishment and maintenance of latent TB.

This project helped lay a foundation for development of biomarkers for latent TB and better treatment options for TB. First, latent TB infections can revert to active TB when the patient becomes immunocompromised. Therefore, it is important to identify and treat latently infected individuals. Yet, there are no reliable methods to identify, and then treat, this important cohort. The correlation of MazF signatures in the Mtb transcriptome to latent TB may serve as the basis for a rapid clinical test to identify individuals with latent TB so they can be treated to eradicate the infection. Second, existing therapies for TB require an excessively long course of treatment (6-9 months), and often still do not eliminate the pathogen from the host's body. Due to the length of current TB therapies, many patients do not consistently take their treatment; this is the perfect breeding-ground for antibiotic resistant Mtb. Shorter therapies could decrease the prevalence of these lapses in treatment, thereby decreasing antibiotic resistance. Current therapies also levy a significant financial burden on many countries – especially developing countries where TB incidence is highest – due to the need for long-term treatment for each patient; shorter therapies could remedy this problem. The data found in this study contribute to the critical knowledge base required for the development of shorter and more effective therapies for latent

TB. Specifically, if MazFs are involved in establishing Mtb latency, this could lead to the identification of latency-inducing signals as a new class of therapeutic targets in the fight against TB. Ultimately, new therapies could lead to more widespread control of TB worldwide, and could potentially lead to the eradication of TB.

References

Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (1996). An Escherichia coli chromosomal "addiction module" regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death. Proceedings of the National Academy of Sciences of the United States of America *93*, 6059-6063.

Amitai, S., Yassin, Y., and Engelberg-kulka, H. (2004). MazF-Mediated Cell Death in Escherichia coli : a Point of No Return MazF-Mediated Cell Death in Escherichia coli : a Point of No Return. Journal of Bacteriology *186*, 8295-8300.

Baik, S., Inoue, K., Ouyang, M., and Inouye, M. (2009). Significant Bias against the ACA Triplet in the tmRNA Sequence of Escherichia coli K-12. Journal of Bacteriology *191*, 6157-6166.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science (New York, NY) *289*, 905-920.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2013). GenBank. Nucleic Acids Res *41*, D36-42.

Betts, J.C., Lukey, P.T., Robb, L.C., McAdam, R.a., and Duncan, K. (2002). Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling. Molecular Microbiology *43*, 717-731.

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Müller, K.M.*, et al.* (2002). The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics *3*, 2.

CDC (2017). Treatment for TB Disease.

Chan, P.P., and Lowe, T.M. (2009). GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Research *37*, D93-D97.

Chandler, M., Bird, R.E., and Caro, L. (1975). The replication time of the Escherichia coli K12 chromosome as a function of cell doubling time. Journal of Molecular Biology *94*, 127-132.

Christensen, S.K., Pedersen, K., Hansen, F.G., and Gerdes, K. (2003). Toxin–antitoxin Loci as Stress-response-elements: ChpAK/MazF and ChpBK Cleave Translated RNAs and are Counteracted by tmRNA. Journal of Molecular Biology *332*, 809-819.

Cortes, T., Schubert, O.T., Rose, G., Arnvig, K.B., Comas, I., Aebersold, R., and Young, D.B. (2013). Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in Mycobacterium tuberculosis. Cell reports *5*, 1121-1131.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome research *14*, 1188-1190.

Cruz, J.W., Sharp, J.D., Hoffer, E.D., Maehigashi, T., Vvedenskaya, I.O., Konkimalla, A., Husson, R.N., Nickels, B.E., Dunham, C.M., and Woychik, N.A. (2015). Growth-regulating Mycobacterium tuberculosis VapC-mt4 toxin is an isoacceptor-specific tRNase. Nature Communications *6*, 1-12.

Cruz, J.W., and Woychik, N.A. (2016). tRNAs taking charge. FEMS Pathogens and Disease *74*, 1-8.

Dye, C., Scheele, S., Dolin, P., Raviglione, M.C., and Page, P. (1999). Global Burden of Tuberculosis: Estimated Incidence, Prevalence, and Mortality by Country. The Journal of the American Medical Association *282*, 677-686.

Ehrt, S., Guo, X.V., Hickey, C.M., Ryou, M., Monteleone, M., Riley, L.W., and Schnappinger, D. (2005). Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. Nucleic Acids Res *33*, e21.

Erental, A., Sharon, I., and Engelberg-Kulka, H. (2012). Two programmed cell death systems in escherichia coli: An apoptotic-like death is inhibited by the mazef-mediated death pathway. PLoS biology *10*.

Fozo, E.M., Makarova, K.S., Shabalina, S.A., Yutin, N., Koonin, E.V., and Storz, G. (2010). Abundance of type I toxin–antitoxin systems in bacteria: searches for new candidates and discovery of novel families. Nucleic Acids Research *38*, 3743-3759.

Frieden, T.R., Sterling, T., Pablos-Mendez, A., Kilburn, J.O., Cauthen, G.M., and Dooley, S.W. (1993). The emergence of drug-resistant tuberculosis in New York City. The New England journal of medicine *328*, 521-526.

Goeders, N., and Van Melderen, L. (2014). Toxin-Antitoxin Systems as Multilevel Interaction Systems. Toxins *6*, 304-324.

Hazan, R., Hazan, R., Sat, B., Sat, B., Engelberg-kulka, H., and Engelberg-kulka, H. (2004). Escherichia coli mazEF-Mediated. Microbiology (Reading, England) *186*, 3663-3669.

Hsu, T., Hingley-Wilson, S.M., Chen, B., Chen, M., Dai, A.Z., Morin, P.M., Marks, C.B., Padiyar, J., Goulding, C., Gingery, M.*, et al.* (2003). The primary mechanism of attenuation of bacillus Calmette–Guérin is a loss of secreted lytic

function required for invasion of lung interstitial tissue. Proceedings of the National Academy of Sciences *100*, 12420-12425.

Kamada, K., Hanaoka, F., and Burley, S.K. (2003). Crystal Structure of the MazE/MazF Complex. Molecular Cell *11*, 875-884.

Keren, I., Shah, D., Spoering, A., Kaldalu, N., and Lewis, K. (2004). Specialized Persister Cells and the Mechanism of Multidrug Tolerance in Escherichia coli. Journal of Bacteriology *186*, 8172-8180.

Kimman, T.G., Smit, E., and Klein, M.R. (2008). Evidence-Based Biosafety: a Review of the Principles and Effectiveness of Microbiological Containment Measures. Clinical Microbiology Reviews *21*, 403-425.

Komoda, T., Sato, N.S., Phelps, S.S., Namba, N., Joseph, S., and Suzuki, T. (2006). The A-site Finger in 23 S rRNA Acts as a Functional Attenuator for Translocation. Journal of Biological Chemistry *281*, 32303-32309.

Korch, S.B., Contreras, H., and 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*, 1618-1630.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Leviev, I., Levieva, S., and Garrett, R.A. (1995). Role for the highly conserved region of domain IV of 23S-like rRNA in subunit-subunit interactions at the peptidyl transferase centre. Nucleic Acids Res *23*, 1512-1517.

Lew, J.M., Kapopoulou, A., Jones, L.M., and Cole, S.T. (2011). TubercuList--10 years after. Tuberculosis (Edinburgh, Scotland) *91*, 1-7.

Lewis, K. (2013). Platforms for antibiotic discovery. Nat Rev Drug Discov *12*, 371-387.

Mackie, G.A. (1992). Secondary Structure of themRNA for Ribosomal Protein 520. Journal of Biological Chemistry *267*, 1054-1061.

Maisonneuve, E., Shakespeare, L.J., Jørgensen, M.G., and Gerdes, K. (2011). Bacterial persistence by RNA endonucleases. Proceedings of the National Academy of Sciences of the United States of America *108*, 13206-13211.

Malherbe, S.T., Shenai, S., Ronacher, K., Loxton, A.G., Dolganov, G., Kriel, M., Van, T., Chen, R.Y., Warwick, J., Via, L.E.*, et al.* (2016). Persisting PET-CT lesion activity and M. tuberculosis mRNA after pulmonary tuberculosis cure. Nature medicine *22*, 1094-1100.

Marianovsky, I., Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (2001). The Regulation of the Escherichia coli mazEF Promoter Involves an Unusual Alternating Palindrome. Journal of Biological Chemistry *276*, 5975-5984.

Masuda, H., Tan, Q., Awano, N., Wu, K.-P., and Inouye, M. (2012). YeeU enhances the bundling of cytoskeletal polymers of MreB and FtsZ, antagonizing the CbtA (YeeV) toxicity in Escherichia coli. Molecular Microbiology *84*, 979-989.

Masuda, Y., Miyakawa, K., Nishimura, Y., and Ohtsubo, E. (1993). chpA and chpB, Escherichia coli chromosomal homologs of the pem locus responsible for stable maintenance of plasmid R100. Journal of Bacteriology *175*, 6850-6856.

Merryman, C., Moazed, D., Daubresse, G., and Noller, H.F. (1999). Nucleotides in 23S rRNA protected by the association of 30S and 50S ribosomal subunits. Journal of molecular biology *285*, 107-113.

Mets, T., Lippus, M., Schryer, D., Liiv, A., Kasari, V., Paier, A., Maiväli, Ü., Remme, J., Tenson, T., and Kaldalu, N. (2017). Toxins MazF and MqsR cleave Escherichia coli rRNA precursors at multiple sites. RNA Biology *14*, 124-135.

Mwinga, A., and Bernard Fourie, P. (2004). Prospects for new tuberculosis treatment in Africa. Tropical Medicine & International Health *9*, 827-832.

Nakajima, H. (1993). Tuberculosis: a global emergency. World Health *46*, 3.

Pellegrini, O., Mathy, N., Gogos, A., Shapiro, L., and Condon, C. (2005). The Bacillus subtilis ydcDE operon encodes an endoribonuclease of the MazF/PemK family and its inhibitor. Mol Microbiol *56*, 1139-1148.

Petrov, A.S., Bernier, C.R., Hershkovits, E., Xue, Y., Waterbury, C.C., Hsiao, C., Stepanov, V.G., Gaucher, E.A., Grover, M.A., Harvey, S.C*., et al.* (2013). Secondary structure and domain architecture of the 23S and 5S rRNAs. Nucleic Acids Res *41*, 7522-7535.

Provvedi, R., Boldrin, F., Falciani, F., Palu, G., and Manganelli, R. (2009). Global transcriptional response to vancomycin in Mycobacterium tuberculosis. Microbiology (Reading, England) *155*, 1093-1102.

Prysak, M.H., Mozdzierz, C.J., Cook, A.M., Zhu, L., Zhang, Y., Inouye, M., and Woychik, N.A. (2009). Bacterial toxin YafQ is an endoribonuclease that associates with the ribosome and blocks translation elongation through sequence-specific and frame-dependent mRNA cleavage. Mol Microbiol *71*, 1071-1087.

Ramage, H.R., Connolly, L.E., and Cox, J.S. (2009). Comprehensive Functional Analysis of Mycobacterium tuberculosis Toxin-Antitoxin Systems: Implications for Pathogenesis, Stress Responses, and Evolution. PLoS genetics *5*, e1000767.

Rand, L., Hinds, J., Springer, B., Sander, P., Buxton, R.S., and Davis, E.O. (2003). The majority of inducible DNA repair genes in Mycobacterium tuberculosis are induced independently of RecA. Molecular Microbiology *50*, 1031-1042.

Raviglione, M.C., and Uplekar, M.W. (2006). WHO's new Stop TB Strategy. Lancet (London, England) *367*, 952-955.

Sala, A., Bordes, P., and Genevaux, P. (2014). Multiple toxin-antitoxin systems in Mycobacterium tuberculosis. Toxins (Basel) *6*, 1002-1020.

Sambandamurthy, V.K., Wang, X., Chen, B., Russell, R.G., Derrick, S., Collins, F.M., Morris, S.L., and Jacobs, W.R. (2002). A pantothenate auxotroph of Mycobacterium tuberculosis is highly attenuated and protects mice against tuberculosis. Nat Med *8*, 1171-1174.

Sat, B., Hazan, R., Fisher, T., Khaner, H., Glaser, G., and Engelberg-Kulka, H. (2001). Programmed Cell Death in Escherichia coli: Some Antibiotics Can Trigger mazEF Lethality. Journal of Bacteriology *183*, 2041-2045.

Sauert, M., Wolfinger, M.T., Vesper, O., Muller, C., Byrgazov, K., and Moll, I. (2016). The MazF-regulon: a toolbox for the post-transcriptional stress response in Escherichia coli. Nucleic Acids Res *44*, 6660-6675.

Schatz, A., Bugle, E., and Waksman, S.A. (1944). Streptomuycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. Experimental Biology and Medicine *55*, 66-69.

Schifano, J.M., Cruz, J.W., Vvedenskaya, I.O., Edifor, R., Ouyang, M., Husson, R.N., Nickels, B.E., and Woychik, N.A. (2016). tRNA is a new target for cleavage by a MazF toxin. Nucleic Acids Research, gkv1370.

Schifano, J.M., Edifor, R., Sharp, J.D., Ouyang, M., Konkimalla, A., and Husson, R.N. (2013). Mycobacterial toxin MazF-mt6 inhibits translation through cleavage of 23S rRNA at the ribosomal A site. Proceedings of the National Academy of Sciences *110*, 1-6.

Schifano, J.M., Vvedenskaya, I.O., Knoblauch, J.G., Ouyang, M., Nickels, B.E., and Woychik, N.A. (2014). An RNA-seq method for defining endoribonuclease cleavage specificity identifies dual rRNA substrates for toxin MazF-mt3. Nature Communications *5*, 1-11.

Schifano, J.M., and Woychik, N.A. (2014). 23S rRNA as an a-Maz-ing new bacterial toxin target. RNA Biology *11*, 101-105.

Sharp, J.D., Cruz, J.W., Raman, S., Inouye, M., Husson, R.N., and Woychik, N.A. (2012). Growth and Translation Inhibition through Sequence-specific RNA

Binding by Mycobacterium tuberculosis VapC Toxin. The Journal of biological chemistry *287*, 12835-12847.

Shell, S.S., Wang, J., Lapierre, P., Mir, M., Chase, M.R., Pyle, M.M., Gawande, R., Ahmad, R., Sarracino, D.A., Ioerger, T.R.*, et al.* (2015). Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape. PLoS genetics *11*, e1005641.

Short, F.L., Pei, X.Y., Blower, T.R., Ong, S.-L., Fineran, P.C., Luisi, B.F., and Salmond, G.P.C. (2013). Selectivity and self-assembly in the control of a bacterial toxin by an antitoxic noncoding RNA pseudoknot. Proceedings of the National Academy of Sciences of the United States of America *110*, E241-E249.

Temmel, H., Muller, C., Sauert, M., Vesper, O., Reiss, A., Popow, J., Martinez, J., and Moll, I. (2017). The RNA ligase RtcB reverses MazF-induced ribosome heterogeneity in Escherichia coli. Nucleic Acids Res *45*, 4708-4721.

Tiemersma, E.W., van der Werf, M.J., Borgdorff, M.W., Williams, B.G., and Nagelkerke, N.J.D. (2011). Natural History of Tuberculosis: Duration and Fatality of Untreated Pulmonary Tuberculosis in HIV Negative Patients: A Systematic Review. PloS one *6*, e17601.

Tsuchimoto, S., and Ohtsubo, E. (1993). Autoregulatiou by cooperative binding of the PemI and PemK proteins to the promoter region of the pem operon. Molecular and General Genetics MGG *237*, 81-88.

Van Melderen, L., and De Bast, M.S. (2009). Bacterial Toxin – Antitoxin Systems: More Than Selfish Entities? PLoS genetics *5*, e1000437.

Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A.C., Engelberg-Kulka, H., and Moll, I. (2011). Selective Translation of Leaderless mRNAs by Specialized Ribosomes Generated by MazF in Escherichia coli. Cell *147*, 147-157.

Wang, X., Lord, D.M., Cheng, H.-Y., Osbourne, D.O., Hong, S.H., Sanchez-Torres, V., Quiroga, C., Zheng, K., Herrmann, T., Peti, W.*, et al.* (2012). A Novel Type V TA System Where mRNA for Toxin GhoT is Cleaved by Antitoxin GhoS. Nature chemical biology *8*, 855-861.

WHO (2016). Global Tuberculosis Report 2016.

Winther, K., Tree, J.J., Tollervey, D., and Gerdes, K. (2016). VapCs of Mycobacterium tuberculosis cleave RNAs essential for translation. Nucleic Acids Research *44*, 9860-9871.

Winther, K.S., Brodersen, D.E., Brown, A.K., and Gerdes, K. (2013). VapC20 of Mycobacterium tuberculosis cleaves the sarcin-ricin loop of 23S rRNA. Nature communications *4*, 2796.

Yamaguchi, Y., and Inouye, M. (2011). Regulation of growth and death in Escherichia coli by toxin–antitoxin systems. Nature Reviews Microbiology *9*, 779-790.

Yamaguchi, Y., Park, J.-h., and Inouye, M. (2011). Toxin-Antitoxin Systems in Bacteria and Archaea. Annual Review Genetics *45*, 61-79.

Zhang, Y., Zhang, J., Hara, H., Kato, I., and Inouye, M. (2005). Insights into the mRNA cleavage mechanism by MazF, an mRNA interferase. Journal of Biological Chemistry *280*, 3143-3150.

Zhang, Y., Zhang, J., Hoeflich, K.P., Ikura, M., Qing, G., and Inouye, M. (2003). MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli. Molecular Cell *12*, 913-923.

Zhu, L., Phadtare, S., Nariya, H., Ouyang, M., Husson, R.N., and Inouye, M. (2008). The mRNA interferases, MazF-mt3 and MazF-mt7 from Mycobacterium tuberculosis target unique pentad sequences in single-stranded RNA. Molecular Microbiology *69*, 559-569.

Zhu, L., Zhang, Y., Teh, J.S., Zhang, J., Connell, N., Rubin, H., and Inouye, M. (2006). Characterization of mRNA interferases from Mycobacterium tuberculosis. Journal of Biological Chemistry *281*, 18638-18643.