Discovery of Small RNAs and Characterization of Their Regulatory Roles in Mycobacterium Tuberculosis

Discovery of small RNAs and characterization of their regulatory roles in *Mycobacterium tuberculosis*

A dissertation presented

by

Elias Roth Gerrick

to

The Committee on Higher Degrees in Biological Sciences in Public Health

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological Sciences in Public Health

Harvard University

Cambridge, Massachusetts

April, 2018

Discovery of small RNAs and characterization of their regulatory roles in *Mycobacterium tuberculosis*

Abstract

Mycobacterium tuberculosis is an important global health pathogen and is the leading cause of death due to an infectious disease. The pathogen resides inside of human macrophages and is exposed to a wide array of different bactericidal stresses, yet manages to subvert or adapt to each of these in order to grow, divide, and cause disease in humans. Therefore, there is a need for a deeper understanding of how *M. tuberculosis* adapts to the stress conditions that are imposed by the human host in order to better understand how to combat this deadly disease.

One important way by which bacteria respond to stress and rapidly adapt to changing environments is through the use of *trans*-encoded small RNAs (sRNAs). These short RNA molecules become highly induced in specific conditions, where they bind directly to a set of mRNA targets to regulate their expression. The interactions between sRNAs and their targets is generally mediated by a protein accessory factor such as Hfq, which acts as an RNA chaperone to allow for sRNA-mRNA binding. Although much is known about how sRNAs function in model bacterial species including *Escherichia coli*, comparatively little is understood about these regulators in mycobacteria. For example, sRNA discovery studies in *M. tuberculosis* have focused on sRNAs present in the absence of stress, and not a single sRNA-target interaction has been experimentally validated in mycobacteria. Additionally, mycobacteria contain no obvious homologue of any known accessory factor, and no other accessory factor has been identified in this lineage.

After a review of the literature on what is currently known about *M. tuberculosis* stress adaptation and sRNAs in Chapter 1, we perform large scale sRNA discovery in *M. tuberculosis*

during exposure to host-like stress conditions in Chapter 2. By creating a computational sRNA search tool, we generate a master set of 189 *M. tuberculosis* sRNA candidates and profiles of their expression. In Chapter 3, we perform the most in-depth characterization of an sRNA in mycobacteria to date by focusing on one sRNA that becomes highly abundant in multiple stress conditions in *M. tuberculosis*. We show that this sRNA, renamed MrsI, acts as an iron sparing sRNA and binds directly to an mRNA target. We additionally provide evidence that MrsI acts in an anticipatory manner during exposure to oxidative stress to prime *M. tuberculosis* to rapidly enter an iron sparing state. Finally, in Chapter 4 we perform a variety of screens towards the identification of a mycobacterial sRNA accessory factor. Although these screens do not identify a candidate accessory factor, they do provide important insights into characteristics of the as-of-yet unidentified protein.

Taken together, these projects greatly increase our understanding of sRNAs in mycobacteria and how they function in the *M. tuberculosis* stress response. We provide a compendium of sRNAs and their expression patterns during exposure to host-like stressors, and provide in-depth characterization of how one sRNA, MrsI, adapts the pathogen to iron-limited conditions. The results presented here will be critical for future studies on stress responses in *M. tuberculosis*, particularly through the pathogen's use of sRNAs.

Table of Contents

Chapter 5: Discussion

List of Figures and Tables

Chapter 2

Chapter 3

Chapter 4

Appendices

Acknowledgements

I thank my advisor, Dr. Sarah Fortune, for her guidance throughout my thesis work. Your unbridled enthusiasm has been incredibly helpful throughout my Ph.D., and you never fail to push me to accomplish more than I thought I could. You have taught me invaluable lessons for my scientific development, and have encouraged me to never doubt my own abilities. Power pose!

I also thank my dissertation advisory committee members Dr. Eric Rubin, Dr. Simon Dove, and Dr. Marcia Goldberg for your excellent suggestions on my science and the directions of my projects. I am incredibly grateful to have benefited from the input from such amazing scientists. I would also like to thank Eric and Simon for your inspirational help and advice regarding professional development. In addition, I thank my dissertation committee members Dr. Simon Dove, Dr. Jonathan Livny, Dr. Yonatan Grad, and Dr. Manoj Duraisingh for your time, support, and feedback.

I would like to thank past and current members of the Fortune lab who have helped me throughout my Ph.D. training. In particular, I thank my friend and awesome mentor Dr. Jeremy Rock for teaching me so much about how to be a scientist, and I owe much of my successes in graduate school to your constant help. I am fairly certain that your patience for my incessant questions is the 8th wonder of the world. Dr. Thibault "The Muscles from Brussels" Barbier, thank you for being the most optimistic and positive member of Team Small RNA and an overall amazing friend. I also thank Dr. Dr. Allison Carey, Dr. Mike Chase, Forrest Hopkins, Dr. Dr.

Lenette Lu, Dr. Caitlin Spaulding, Josie François, Vincent Lin, Raylin Xu, Skye Fishbein, Shoko Wakabayashi, Katie Wu, and everyone else in both the Fortune and Rubin labs that has helped me so much throughout graduate school, you're all awesome scientists and even better people.

I have been incredibly fortunate to have an amazing group of friends. Jeremy, Thibault, Skye, Gabe, Rachel, Steph, Allison, Andie, Erika, Steve, Karina, Sheena, Deepali, Vanessa, Yama, Tyler, and Vincent, you all made graduate school such a fun experience and were the best group of friends anyone could ask for. Skye, thanks for always being there with emotional and scientific support. Long stories get short, our frequent hangouts and nacho, beer, and science dinners were the best.

I would like to thank my family: My mother Sara, my father Andy, my twin sister Sophie, my big sister Emily, and my extended family for their love and encouragement and for never doubting me.

Finally, my fiancée Kimberline Yang. Your unwavering love and support have made me better in every respect. Thank you for always being there for me despite the distance.

Chapter 1

Introduction

1.1 *Mycobacterium tuberculosis*: a global health threat

Tuberculosis, the infectious disease caused by the bacterium *Mycobacterium tuberculosis*, is a devastating global health disease that currently infects one third of the world's population and is responsible for the deaths of 1-2 million people per year, making it the leading cause of death due to an infectious disease (1). *M. tuberculosis* infection is difficult to treat, as diagnostic tools are severely outdated, the only vaccine against the pathogen (Bacille Calmette-Guerin) has limited efficacy, and the antibiotic treatment regimen lasts as long as 9 months. In addition, strains of *M. tuberculosis* that are resistant to antibiotics are becoming increasingly prevalent, necessitating a better understanding of the bacterium and how it has become such a successful human pathogen.

1.2 *M. tuberculosis* in the host: a stressful life

When *M. tuberculosis* infects humans, it enters the lung and is engulfed by macrophages, where the pathogen resides. However, macrophages are immune cells that are equipped to kill microbes and therefore assault *M. tuberculosis* with an onslaught of stresses. These host-mediated stresses that *M. tuberculosis* must adapt to include membrane attacks, low pH, oxidative radicals, nutrient starvation and iron deprivation. One of the first stresses encountered by the bacteria are attacks on the membrane by antimicrobial peptides, which are present in the airway and produced by the macrophage upon *M. tuberculosis* uptake (2, 3). These peptides form pores in the bacterial membrane and thereby disrupt the integrity of the microbial envelope. In addition, macrophages fuse the *M. tuberculosis*-containing phagosome with acidic lysosomes, creating phagolysosomes, in an attempt to kill the pathogen with acid

stress. To combat this, *M. tuberculosis* blocks phagosome maturation and fusion and is naturally resistant to acid (4, 5). The macrophage also produces oxidative radicals to damage the bacteria, which *M. tuberculosis* responds to by producing high levels of catalase and superoxide dismutase proteins (6, 7). In addition to these direct attacks against the mycobacterial cell, the host immune system also attempts to kill the pathogens via starvation by sequestering nutrients (8, 9). One of these nutrients, iron, is an essential micronutrient for both the host and pathogen. Humans have several mechanisms to keep iron from the bacteria, whereas *M. tuberculosis* has evolved complex machineries to steal it from human proteins.

1.3 *M. tuberculosis* in the host: a Fe-rocious struggle

Iron is an essential micronutrient that is required for a wide array of processes for both *M. tuberculosis* and the host. Due to its ability to donate and accept electrons and thereby switch between ferrous (Fe^{2+}) and ferric (Fe^{3+}) states organisms have evolved to rely on iron for processes such as electron transport, neutralization of reactive oxygen species, and DNA synthesis (10). However, despite its importance in all domains of life and its high abundance as an element, useable iron is a scarce resource. At physiological pH iron is highly insoluble, and therefore inaccessible. Thus, access to iron has become a key point of contention in many host-pathogen interactions whereby the host attempts to sequester iron away from the pathogen while the pathogen endeavors to subvert these measures (10, 11). However, despite the dependence on iron and the arms race that has evolved to acquire it, obtaining too much iron can be detrimental or even lethal. In the oxidizing environment in which many organisms live, iron undergoes the Fenton reaction in which it catalyzes the formation of oxidative radicals that go on to damage proteins, nucleic acids, and lipids in the cell (12). Thus, while pathogens

struggle to obtain enough iron to survive they must also carefully monitor and store the nutrient to prevent toxic levels of accumulation of free intracellular iron. To this end, many bacteria have developed elaborate mechanisms of monitoring, acquiring, and storing iron. *M. tuberculosis* has several regulatory pathways dedicated to iron acquisition and allocation, and new components of the mycobacterial iron regulation machinery are still being discovered.

1.3.1 Sequestration of iron by the host

As an obligate intracellular human pathogen, the first challenge *M. tuberculosis* faces in regards to iron is acquisition. Upon sensing an infection, the human body and immune system take measures to sequester iron away from bacteria, thereby restricting growth of the pathogen. For example, during bacterial infection the immune cells such as neutrophils increase the concentration of iron-sequestering host proteins such as transferrin and lactoferrin in the blood in an effort to prevent pathogens from acquiring the nutrient (13, 14). Additionally, bacterial infection stimulates the production of the hormone hepcidin, which decreases the levels of circulating iron and thereby reduces growth of bacterial pathogens (15, 16). Finally, host immune cells such as macrophages have developed direct measures to prevent iron acquisition by internalized pathogens. For example, infected macrophages produce the protein siderocalin, which binds bacterial siderophores, or iron scavenging proteins, to prevent them from providing the pathogen with iron (17). Infected macrophages also produce the iron transporter Nramp1, which reduces the levels of iron in phagosomes and in the macrophage itself, thereby depriving the bacteria of iron (11, 18, 19). These mechanisms of iron sequestration are

extremely important for limiting the growth of *M. tuberculosis*, as an excess of iron in the host has been shown to exacerbate tuberculosis disease in both mice and humans (20, 21).

1.3.2 Acquisition of host iron by M. tuberculosis

In order to establish a successful infection in humans, *M. tuberculosis* has evolved several mechanisms to thwart these multifaceted efforts by the host and obtain iron for its own use. One way in which mycobacteria obtain iron is through the use of siderophores, of which mycobacteria produce two major types: membrane-associated and secreted (22). Mycobactins are nonpolar iron chelators that remain associated with the mycobacterial cell envelope, whereas carboxymycobactins and exochelins are secreted outside the cell, where they then scavenge iron from the extracellular milieu (23). However, because most iron that *M. tuberculosis* encounters is bound to host proteins, carboxymycobacterin and exochelin have evolved high enough affinity that they can steal iron from high-affinity iron binding host proteins such as transferrin, lactoferrin, and ferritin (23–25). The secreted mycobacterial siderophores then transfer their pilfered iron to either the envelope-associated mycobactins or the ABC transporter IrtAB, which bring the iron into the bacterial cell for incorporation into the pathogen's proteins (17).

The tug-of-war for iron between *M. tuberculosis* and humans has also necessitated additional iron-acquisition pathways by the bacteria. Of the total iron content in the human body, less than 1% is in transferrin (the major iron-source for siderophores) while more than 80% is in heme (26). Intracellular *M. tuberculosis* can encounter host heme when macrophages

internalize and degrade erythrocytes, whereas extracellular *M. tuberculosis* can encounter heme by destroying red blood cells with its hemolysin proteins (27, 28). The bacteria then use a unique secreted hemophore to acquire either free heme or take heme from hemoglobin proteins, and transfer it into the cell where a heme degrading enzyme MhuD liberates the iron (26, 29, 30). Thus, *M. tuberculosis* has developed multiple elaborate mechanisms to acquire iron from the host.

1.3.3 Regulation of iron uptake and storage by M. tuberculosis

Due to the dangerous nature of iron, *M. tuberculosis* must also tightly regulate and control the import and storage of this essential nutrient. The focal point by which mycobacteria sense and respond to changing iron levels is the iron responsive transcription factor IdeR. IdeR, an essential protein in *M. tuberculosis*, is an iron-dependent transcriptional activator and repressor (31–33). In iron replete conditions, IdeR exists in an iron bound, active state. When active, IdeR predominantly acts as a transcriptional repressor by binding to the -10 box of many genes involved in iron acquisition and metabolism, such as the *mbt* gene cluster involved in siderophore biosynthesis (33, 34). Preventing the synthesis of siderophores during growth in iron-rich conditions prevents the bacteria from acquiring toxic levels of iron, which would undergo the Fenton reaction inside the microbial cytosol. Active IdeR additionally acts as a transcriptional activator for the genes *bfrA* and *bfrB,* which encode the iron storage proteins bacterioferritin and ferritin, respectively (31). Upon iron limitation, *apo*-IdeR becomes inactive and de-regulates these genes, allowing for strong induction of iron acquisition systems and decreasing the levels of iron storage proteins in the cell. This regulation has been shown to be

critical for *M. tuberculosis* pathogenesis, as *ideR* mutants are unable to proliferate within a macrophage (33).

1.4 The unusual suspects: sRNAs in bacterial stress adaptation

Bacteria have many mechanisms of adapting to the stresses they encounter, but one recently-appreciated method is through the use of *trans*-encoded small RNAs (sRNAs). sRNAs are short 50-250nt long transcripts that fine-tune bacterial gene expression for rapid adaptation to stress conditions. sRNAs regulate bacterial responses to stresses such as iron starvation, membrane damage, oxidative stress, nutrient starvation, and virulence (35–40). Most sRNAs that have been characterized thus far in bacteria such as *Escherichia coli* repress expression of a specific set of target mRNAs with the help of the sRNA chaperone protein Hfq, although recent discoveries have identified alternative functions, mechanisms, and protein partners. Most of what is known about sRNAs and their regulatory functions comes from work done in bacteria such as *E. coli* and *Salmonella enterica*, and the lessons learned from these organisms are critical to the understanding of sRNA-mediated regulation in any bacterial system.

1.4.1 sRNAs: where do they come from

The first sRNAs, which were discovered fortuitously in *E. coli*, were encoded from intergenic regions (between protein coding genes) (41). Subsequently, dozens of sRNAs were discovered in *E. coli* by searching for conserved features of sRNAs, such as secondary structures, rho-independent terminators, and promoters, within the intergenic regions of bacterial genomes (42–44). However, sRNAs encoded from other locations have been discovered in recent years,

in part due to advances in sequencing technologies. For example, the 3' UTRs of many ORFs were found to contain sRNAs, which share a rho-independent terminator with the protein-coding gene (45–47). Interestingly, these sRNAs can either be transcribed from independent promoters (45) or generated by a processing event at the 3' end of the mRNA (47). Alternatively, sRNAs can originate from 5' UTRs of mRNAs. For example, the SreA and SreB riboswitches in *Listeria monocytogenes* were found to accumulate as prematurely terminated transcripts and act in *trans* to repress the translation of an mRNA encoding a virulence protein (48). In addition to being encoded from the UTRs of protein-coding genes, sRNAs have also been identified in the processed spacers in tRNA precursor transcripts (49). Thus, although most known sRNAs are encoded from intergenic loci, other regions of the genome are becoming appreciated as an understudied reservoir of these regulators.

1.4.2 sRNAs: what do they do

The most common function of an sRNA is to fine-tune bacterial gene expression in response to a specific stress. For example, the iron-sparing sRNA RyhB of *E. coli* was found to be highly induced during iron limitation. This induction is due to alleviation of repression by the *E. coli* iron-dependent transcriptional repressor, Fur (Ferric uptake regulator), during iron starvation (50). Upon Fur inactivation, RyhB rapidly accumulates and represses the translation of mRNAs encoding as many as 56 genes through direct binding of the sRNA to the mRNAs via limited complementarity (51). In addition to blocking translation, this binding interaction facilitates the degradation of the target transcripts by the single-strand-specific endoribonuclease RNaseE (52, 53). The genes targeted by RyhB predominantly encode nonessential iron storage proteins,

iron-containing enzymes, and iron metabolism proteins such as *bacterioferritin (bfr), aconitase (acn),* and *succinate dehydrogenase (sdhCDAB).* By repressing these nonessential iron-containing proteins, RyhB prioritizes the limited stores of iron for essential proteins and thereby allows the cell to grow and survive for longer periods of time in the absence of iron.

In addition to repression of mRNAs, sRNAs can also increase expression of target transcripts. For example, in addition to its inhibitory functions, RyhB has been found to activate the expression of the *shiA* mRNA, which encodes a shikimate permease. This sRNA-mediated activation occurs by RyhB binding directly to an inhibitory structure located at the 5' end of the *shiA* mRNA, which under iron replete conditions prevents the translation of the transcript (54). However, when RyhB is induced during iron limitation the sRNA binds and melts this inhibitory secondary structure, allowing for translation of the shikimate permease. Shikimate is then transported into the cell, where it is used for the synthesis of siderophores and thereby allows the bacteria to scavenge additional iron from the surrounding environment. In addition, RyhB is required for the translation of the siderophore biosynthesis genes encoded by the *entCEBAH* operon (55) as well as the siderophore transporter *cirA* (56), placing this sRNA at a central regulatory point in the response to iron deprivation.

Interestingly, studies on RyhB have revealed an additional function mediated by sRNAs. In an experiment identifying interactions between RyhB and other transcripts in the cell, a fragment of a precursor tRNA transcript was found to bind to RyhB (49). This approximately 50nt fragment, which was encoded from the 3' external transcribed spacer of the *leuZ* tRNA, binds

directly to RyhB to block the sRNA from interacting with its mRNA targets. This inhibition of RyhB action was found to set a threshold for the level that RyhB must accumulate to before target regulation can occur, which allows for even tighter control of sRNA-mediated post-transcriptional regulation. Excitingly, sequence conservation analysis of other tRNA spacers suggests that this novel class of sRNA, which has been termed an sRNA sponge, may be a widespread phenomenon.

1.4.3 sRNAs: how do they do it

sRNAs mediate regulation of mRNAs through direct binding interactions between the sRNA and the target. The sRNA usually binds to the 5' end of the mRNA and blocks ribosome binding, although sRNAs often stimulate degradation of the target as well (57–59). The interaction is initiated by a short sequence of perfect complementarity between the sRNA and target termed the seed region. Seed regions are generally 6-8nt long, and a single sRNA can have one seed region that regulates all of its targets or multiple seed regions that each regulate a subset of targets. Studies on RyhB have found that the energetics of the binding between the seed region and targets correlates with the strength of regulation, and that stronger binding leads to stronger regulation (60). Additionally, seed regions are generally highly conserved, and mutations to the seed region lead to complete abrogation of target regulation (60, 61). In order to facilitate intermolecular interactions with target mRNAs, seed regions are usually single stranded in the folded sRNA and disruption of the sRNA secondary structure can drastically reduce sRNA function (60–62). For example, the RyhB seed region that binds to targets such as *sodB* and *shiA* is in the single-stranded loop of a hairpin, and disruption of the folding of the

hairpin stem reduces RyhB activity (61). However, seed regions alone are generally not sufficient to mediate target binding, and most sRNAs characterized to date rely on the assistance of an sRNA chaperone protein.

The sRNA chaperone that the majority of characterized sRNAs require is host factor for phage Qβ (Hfq). Hfq is a short, homohexameric toroid protein that binds to sRNAs and their mRNA targets, and is required for both the stability and function of its sRNAs (63–65). For example, RyhB is degraded more rapidly without Hfq, and the chaperone is essential for RyhB-mediated regulation of both repressed and activated targets (53, 54, 66). Although Hfq has been found to be required for the interactions of most characterized sRNAs with their targets, recently other sRNA chaperone proteins have been discovered. For example, the iron sparing response of *Bacillus subtilis* relies on the action of an iron sparing sRNA named FsrA, which performs a function analogous to that of RyhB. However, although *B. subtilis* encodes an Hfq homologue, Hfq is dispensable for FsrA function which instead relies on the activity of three small proteins named FbpA, FbpB, and FbpC (67, 68). In addition, recent studies have also elucidated a large set of sRNAs in enterobacteria that rely on the novel sRNA accessory factor ProQ instead of Hfq for their stability and function (69, 70).

Through interactions with Hfq, sRNAs have also been found to associate with other proteins that assist in mediating regulatory effects. For example, Hfq in *E. coli* was found to interact with RNase E, which is stimulated by the sRNAs to degrade their mRNA targets (71–73). Through this

interaction, RyhB stimulates the degradation of its target *sodB,* and other sRNAs have been found to use the same pathway (57, 74, 75).

1.5 Mycobacterial sRNAs: little understood regulators

In stark contrast to what is known about sRNAs in bacteria such as *E. coli* and *S. enterica*, very little is understood about sRNAs and how they function in the important human pathogen *M. tuberculosis* and related species. For example, sRNAs in *E. coli* have been identified and validated in a multitude of stress conditions and the functions of many of these regulators have been determined. However, in mycobacteria only a few sRNA discovery studies have been performed and these have not searched for sRNAs during exposure to stress conditions (76–81). Due to the tight regulation of sRNAs many stress-responsive sRNAs likely remain undiscovered. In addition, although some limited sRNA characterization has been performed in mycobacteria, experimental validation of sRNA-target interactions has never been demonstrated and therefore the mechanism by which any mycobacterial sRNA functions remains uncertain (82). Furthermore, mycobacteria do not encode any obvious homologues of the RNA chaperone proteins Hfq or ProQ, and no other sRNA accessory factor has been identified in mycobacteria (83). Due to the pivotal role that sRNAs have been found to play in stress responses and virulence in other bacteria and the impact that *M. tuberculosis* poses as a global health threat, it is important to remedy this lack of knowledge on mycobacterial sRNAs through both large-scale sRNA discovery studies and in-depth characterization of specific sRNAs.

1.6 Summary of aims

The work described in this dissertation aims to broaden our understanding of sRNA-mediated post-transcriptional riboregulation in *M. tuberculosis*, in order to better understand how this important global health pathogen adapts to the stressful conditions it encounters during pathogenesis. In Chapter 2, we create a computational sRNA search tool for use on sRNA-Sequencing datasets in order to identify novel sRNAs in *M. tuberculosis*. We then perform sRNA-Sequencing on *M. tuberculosis* after exposure to stresses that the bacterium encounters in the human host to identify novel sRNAs and characterize their expression patterns during stress, providing valuable insights into their potential functions. Interestingly, we identify a subset of sRNAs with strong induction in multiple stresses. In Chapter 3, we characterize one of these promiscuously induced sRNAs, which we rename MrsI. We demonstrate that MrsI acts as an iron sparing sRNA in *M. smegmatis* and *M. tuberculosis*, and that it regulates its targets by a direct binding interaction. Additionally, we provide evidence that MrsI induction in *M. tuberculosis* during exposure to certain stresses serves to prime the cell to enter an iron sparing state more rapidly upon iron deprivation, consistent with a model of anticipatory regulation. Finally, in Chapter 4 we perform bioinformatic, biochemical, and genetic screens in an effort to identify the accessory factor that assists sRNA function in mycobacteria. Although none of these screens revealed a candidate accessory factor, the results provide important insights into potential characteristics of this undiscovered protein.

1.7 References

1. WHO Global Tuberculosis Report Available at: http://www.who.int/tb/publications/global_report/Exec_Summary_13Nov2017.pdf?ua=

1 [Accessed March 18, 2018].

2. Alonso S, Pethe K, Russell DG, Purdy GE (2007) Lysosomal killing of Mycobacterium mediated by ubiquitin-derived peptides is enhanced by autophagy. *Proc Natl Acad Sci USA* 104(14):6031-6036.

3. Liu PT, Stenger S, Tang DH, Modlin RL (2007) Cutting Edge: Vitamin D-Mediated Human Antimicrobial Activity against Mycobacterium tuberculosis Is Dependent on the Induction of Cathelicidin. *J Immunol* 179(4):2060–2063.

4. Macgurn JA, Cox JS (2007) A Genetic Screen for Mycobacterium tuberculosis Mutants Defective for Phagosome Maturation Arrest Identifies Components of the ESX-1 Secretion System. *Infect Immun* 75(6):2668–2678.

5. Pethe K, et al. (2004) Isolation of Mycobacterium tuberculosis mutants defective in the arrest of phagosome maturation. *Proc Natl Acad Sci USA* 101(37): 13642-7

6. Ng VH, Cox JS, Sousa AO, MacMicking JD, McKinney JD (2004) Role of KatG catalase-peroxidase in mycobacterial pathogenisis: Countering the phagocyte oxidative burst. *Mol Microbiol* 52(5):1291–1302.

7. Cooper AM, et al. (2000) Transient Loss of Resistance to Pulmonary Tuberculosis in p47 phox − / − Mice Transient Loss of Resistance to Pulmonary Tuberculosis. *Infect Immun* 68(3):1231–1234.

8. Dahl JL, et al. (2003) The role of RelMtb-mediated adaptation to stationary phase in long-term persistence of Mycobacterium tuberculosis in mice. *Proc Natl Acad Sci* 100(17):10026–10031.

9. Primm TP, et al. (2000) The Stringent Response of Mycobacterium tuberculosis Is Required for Long-Term Survival. *J Bacteriol* 182(17):4889–4898.

10. Schaible UE, Kaufmann SHE (2004) Iron and microbial infection. *Nat Rev Microbiol* 2(12):946–953.

11. Ganz T (2018) Iron and infection. *Int J Hematol* 107(1):7–15.

12. Imlay JA (2013) The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nat Publ Gr* 11. doi:10.1038/nrmicro3032.

13. Kontoghiorghes GJ, Weinberg ED (1995) Iron: Mammalian defense systems, mechanisms of disease, and chelation therapy approaches. *Blood Rev* 9(1):33–45.

14. Ratledge C (2004) Iron, mycobacteria and tuberculosis. *Tuberculosis*, pp 110–130.

15. Arezes JO, et al. (2015) Hepcidin-Induced Hypoferremia Is a Critical Host Defense

Mechanism against the Siderophilic Bacterium Vibrio vulnificus. *Cell Host Microbe* 17:47–57.

16. Drakesmith H, Prentice AM (2012) Hepcidin and the iron-infection axis. *Science (80-)* 338(6108):768–772.

17. Rodriguez GM, Smith I (2006) Identi cation of an ABC Transporter Required for Iron Acquisition and Virulence in. *Society* 188(2):424–430.

18. Fritsche G, et al. (2008) Modulation of macrophage iron transport by Nramp1 (Slc11a1). *Immunobiology* 212:751–757.

19. Ganz T (2009) Iron in innate immunity: starve the invaders. *Curr Opin Immunol* 21(1):63–67.

20. Schaible UE, Collins HL, Priem F, Kaufmann SHE (2002) Brief Definitive Report Correction of the Iron Overload Defect in ᴺʟ -2-Microglobulin Knockout Mice by Lactoferrin Abolishes Their Increased Susceptibility to Tuberculosis. *J Exp Med* 12700(11):1507–1513.

21. Gangaidzo IT, et al. (2001) Association of pulmonary tuberculosis with increased dietary iron. *J Infect Dis* 184(7). doi:10.1086/323203.

22. Fang Z, Sampson SL, Warren RM, Gey Van Pittius NC, Newton-Foot M (2015) Iron acquisition strategies in mycobacteria. *Tuberculosis* 95(2):123–130.

23. Sritharan M (2016) Iron homeostasis in Mycobacterium tuberculosis: Mechanistic insights into siderophore-mediated Iron uptake. *J Bacteriol* 198(18):2399–2409.

24. Gobin J, Horwitz MA (1996) Exochelins of Mycobacterium tuberculosis remove iron from human iron-binding proteins and donate iron to mycobactins in the M. tuberculosis cell wall. *J Exp Med* 183(4):1527–32.

25. Boradia VM, et al. (2014) Mycobacterium tuberculosis acquires iron by cell-surface sequestration and internalization of human holo-transferrin. *Nat Commun* 5. doi:10.1038/ncomms5730.

26. Tullius M V., et al. (2011) Discovery and characterization of a unique mycobacterial heme acquisition system. *Proc Natl Acad Sci* 108(12):5051–5056.

27. Ganz T (2012) Macrophages and Systemic Iron Homeostasis. *J Innate Immun* 4:446–453.

28. Deshpande RG, Khan MB, Bhat DA, Navalkar RG (1997) Isolation of a contact-dependent haemolysin from Mycobacterium tuberculosis. *J Med Microbiol -Vol* 46:233–238.

29. Owens CP, et al. (2013) The mycobacterium tuberculosis secreted protein Rv0203 transfers heme to membrane proteins MmpL3 and MmpL11. *J Biol Chem* 288(30):21714–

21728.

30. Nambu S, Matsui T, Goulding CW, Takahashi S, Ikeda-Saito M (2013) A new way to degrade heme: The mycobacterium tuberculosis enzyme MhuD catalyzes heme degradation without generating CO. *J Biol Chem* 288(14):10101–10109.

31. Rodriguez GM, Voskuil MI, Gold B, Schoolnik GK, Smith I (2002) ideR, An essential gene in mycobacterium tuberculosis: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infect Immun* 70(7):3371–81.

32. Gold B, Marcela Rodriguez G, Marras SAE, Pentecost M, Smith I (2001) The Mycobacterium tuberculosis ideR is a dual functional regulator that controls transcription of genes involved in iron acquisition, iron storage and survival in macrophages. *Mol Microbiol* 42(3):851–865.

33. Pandey R, Rodriguez GM (2014) IdeR is required for iron homeostasis and virulence in Mycobacterium tuberculosis. *Mol Microbiol* 91(1):98–109.

34. Prakash P, Yellaboina S, Ranjan A, Hasnain SE (2005) Computational prediction and experimental verification of novel IdeR binding sites in the upstream sequences of Mycobacterium tuberculosis open reading frames. *Bioinforma Discov NOTE* 21(1010):2161–2166.

35. Oglesby-Sherrouse AG, Murphy ER (2013) Iron-responsive bacterial small RNAs: variations on a theme. *Metallomics* 5(5):1756–5901.

36. Papenfort K, Sun Y, Miyakoshi M, Vanderpool CK, Vogel J (2013) Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell* 153(2):426–437.

37. Bouvier M, Sharma CM, Mika F, Nierhaus KH, Vogel J (2008) Small RNA Binding to 5' mRNA Coding Region Inhibits Translational Initiation. *Mol Cell* 32(6):827–837.

38. Argaman L, Altuvia S (2000) fhlA repression by OxyS RNA: Kissing complex formation at two sites results in a stable antisense-target RNA complex. *J Mol Biol* 300(5):1101–1112.

39. Chabelskaya S, Gaillot O, Felden B (2010) A Staphylococcus aureus small RNA is required for bacterial virulence and regulates the expression of an immune-evasion molecule. *PLoS Pathog* 6(6). doi:10.1371/journal.ppat.1000927.

40. Pfeiffer V, Papenfort K, Lucchini S, Hinton JCD, Vogel J (2009) Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat Struct Mol Biol* 16(8):840–846.

41. Gottesman S (2004) The Small RNA Regulators of *Escherichia coli* : Roles and Mechanisms. *Annu Rev Microbiol* 58(1):303–328.

42. Chen S, et al. (2002) A bioinformatics based approach to discover small RNA genes in the Escherichia coli genome. *BioSystems* 65:157–177.

43. Argaman L, et al. (2001) Novel small RNA-encoding genes in the intergenic regions of Escherichia coli. *Curr Biol* 11:941–950.

44. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S (2001) Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev* 15(13):1637–1651.

45. Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J (2012) An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J* 31(20):4005–4019.

46. Miyakoshi M, Chao Y, Vogel J (2015) Regulatory small RNAs from the 3' regions of bacterial mRNAs. *Curr Opin Microbiol* 24:132–139.

47. Chao Y, Vogel J (2016) A 3' UTR-Derived Small RNA Provides the Regulatory Noncoding Arm of the Inner Membrane Stress Response. *Mol Cell* 61(3):352–363.

48. Loh E, et al. (2009) A trans-Acting Riboswitch Controls Expression of the Virulence Regulator PrfA in Listeria monocytogenes. *Cell* 139(4):770–779.

49. Lalaouna D, et al. (2015) A 3' external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. *Mol Cell* 58(3):393–405.

50. Massé E, Gottesman S (2002) A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc Natl Acad Sci U S A* 99(7):4620–4625.

51. Massé E, Vanderpool CK, Gottesman S (2005) Effect of RyhB small RNA on global iron use in Escherichia coli. *J Bacteriol* 187(20):6962–6971.

52. Massé E, Escorcia FE, Gottesman S (2003) Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. *Genes Dev* 17(19):2374–2383.

53. Desnoyers G, Morissette A, Pré K, Massé E (2009) Small RNA-induced differential degradation of the polycistronic mRNA iscRSUA. *EMBO J* 28:1551–1561.

54. Prévost K, et al. (2007) The small RNA RyhB activates the translation of shiA mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. *Mol Microbiol* 64(5):1260–1273.

55. Salvail H, et al. (2010) A small RNA promotes siderophore production through transcriptional and metabolic remodeling. *Proc Natl Acad Sci* 107(34):15223–15228.

56. Salvail H, Caron M-P, Bé Langer J, Massé E (2013) Antagonistic functions between the

RNA chaperone Hfq and an sRNA regulate sensitivity to the antibiotic colicin. *EMBO J* 32:2764–2778.

57. Prévost K, Desnoyers G, Jacques JF, Lavoie F, Massé E (2011) Small RNA-induced mRNA degradation achieved through both translation block and activated cleavage. *Genes Dev* 25(4):385–396.

58. Feng L, et al. (2015) A Qrr noncoding RNA deploys four different regulatory mechanisms to optimize quorum-sensing dynamics. *Cell* 160(1–2):228–240.

59. Papenfort K, Bouvier M, Mika F, Sharma CM, Vogel J (2010) Evidence for an autonomous 5’ target recognition domain in an Hfq-associated small RNA. *Proc Natl Acad Sci* 107(47):20435–20440.

60. Hao Y, et al. (2011) Quantifying the sequence-function relation in gene silencing by bacterial small RNAs. *Proc Natl Acad Sci* 108(30):12473–12478.

61. Peterman N, Lavi-Itzkovitz A, Levine E (2014) Large-scale mapping of sequence-function relations in small regulatory RNAs reveals plasticity and modularity. *Nucleic Acids Res* 42(10):12177–12188.

62. Richter AS, Backofen R (2012) Accessibility and conservation: General features of bacterial small RNA-mRNA interactions? *RNA Biol* 9(7):954–965.

63. Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. *Nat Rev Microbiol* 9(8):578–589.

64. Sauer E, Weichenrieder O (2011) Structural basis for RNA 3’-end recognition by Hfq. *Proc Natl Acad Sci* 108(32):13065–13070.

65. Soper TJ, Doxzen K, Woodson SA (2011) Major role for mRNA binding and restructuring in sRNA recruitment by Hfq. *Rna* 17(8):1544–1550.

66. Andrade JM, Pobre V, Matos AM, Arraiano CM (2012) The crucial role of PNPase in the degradation of small RNAs that are not associated with Hfq. *Rna* 18(4):844–855.

67. Gaballa A, et al. (2008) The Bacillus subtilis iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. *Proc Natl Acad Sci* 105(33):11927–11932.

68. Smaldone GT, Antelmann H, Gaballa A, Helmann JD (2012) The FsrA sRNA and FbpB protein mediate the iron-dependent induction of the Bacillus subtilis lutABC iron-sulfur-containing oxidases. *J Bacteriol* 194(10):2586–2593.

69. Smirnov A, et al. (2016) Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc Natl Acad Sci* 113(41):11591–11596.

70. Smirnov A, Wang C, Drewry LL, Vogel J (2017) Molecular mechanism of mRNA repression in trans by a ProQ-dependent small RNA. *EMBO J* 36:1029–1045.

71. Bandyra KJ, et al. (2012) The Seed Region of a Small RNA Drives the Controlled Destruction of the Target mRNA by the Endoribonuclease RNase E. *MOLCEL* 47:943–953.

72. Guillier M, Gottesman S (2008) The 5' end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. *Nucleic Acids Res* 36(21):6781–6794.

73. Marie-Pier Caron DAL& EM (2010) Small RNA-mediated regulation at the level of transcript stability. *RNA Biol*. doi:10.4161/rna.7.2.11056.

74. Morita T, Maki K, Aiba H (2005) RNase E-based ribonucleoprotein complexes: Mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev* 19(18):2176–2186.

75. Afonyushkin T, Večerek B, Moll I, Bläsi U, Kaberdin VR (2005) Both RNase E and RNase III control the stability of sodB mRNA upon translational inhibition by the small regulatory RNA RyhB. *Nucleic Acids Res* 33(5):1678–1689.

76. Arnvig KB, et al. (2011) Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis. *PLoS Pathog* 7(11). doi:10.1371/journal.ppat.1002342.

77. Arnvig KB, Young DB (2009) Identification of small RNAs in Mycobacterium tuberculosis. *Mol Microbiol* 73(3):397–408.

78. Pelly S, Bishai WR, Lamichhane G (2012) A screen for non-coding RNA in Mycobacterium tuberculosis reveals a cAMP-responsive RNA that is expressed during infection. *Gene* 500(1):85–92.

79. Li S-K, et al. (2013) Identification of small RNAs in Mycobacterium smegmatis using heterologous Hfq. *Rna* 19(1):74–84.

80. Tsai CH, et al. (2013) Identification of novel sRNAs in mycobacterial species. *PLoS One* 8(11). doi:10.1371/journal.pone.0079411.

81. Miotto P, et al. (2012) Genome-Wide Discovery of Small RNAs in Mycobacterium tuberculosis. *PLoS One* 7(12). doi:10.1371/journal.pone.0051950.

82. Solans L, et al. (2014) The PhoP-Dependent ncRNA Mcr7 Modulates the TAT Secretion System in Mycobacterium tuberculosis. *PLoS Pathog* 10(5). doi:10.1371/journal.ppat.1004183.

83. Jousselin A, Metzinger L, Felden B (2009) On the facultative requirement of the bacterial

RNA chaperone, Hfq. *Trends Microbiol* 17(9):399–405.

Chapter 2

High-throughput discovery and characterization of small RNAs in *Mycobacterium tuberculosis*

Elias R. Gerrick[a], Thibault Barbier[a], Michael A. DeJesus[b], Thomas R. Ioerger[b], Sarah M. Fortune[a,c]

[a]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA; [b]Department of Computer Science and Engineering, Texas A&M University, College Station, Texas, USA; [c] The Ragon Institute of MGH, Harvard and MIT, Cambridge, Massachusetts 02139, USA

Author contributions: E.R.G. and T.B. performed small RNA-Sequencing. E.R.G. performed sRNA-Sequencing analysis and sRNA discovery. M.A.D. and T.R.I. performed Tn-Seq metaanalysis and essentiality calling. E.R.G., T.B., M.A.D., T.R.I., and S.M.F. designed experiments. The work related to the BS_Finder tool and sRNA essentiality is published, with modifications, in "DeJesus MA, Gerrick ER, Xu W, Park SW, *et al*. (2017). Comprehensive essentiality analysis of the Mycobacterium tuberculosis genome via saturating transposon mutagenesis. *mBio* 8:e02133-16." The work on sRNA discovery in stress conditions is part of a manuscript that has been submitted at the time of writing.

Abstract

Small RNAs (sRNAs) are regulatory RNA molecules in prokaryotes that are particularly important for bacterial stress responses and rapid adaptation to changing environments. Although much is known about the identities and functions of sRNAs in bacteria such as *Escherichia coli*, very little is known about sRNAs in the important human pathogen *Mycobacterium tuberculosis*. Limited sRNA discovery has been performed in *M. tuberculosis,* but these studies have focused on sRNAs that are highly expressed during logarithmic phase growth without added stressors. Therefore, the majority of stress responsive sRNAs in *M. tuberculosis* likely remain undiscovered. Here, we create a computational sRNA discovery tool, BS_Finder, which we show accurately identifies sRNAs in small RNA-Sequencing datasets. We apply this sRNA sequencing and discovery approach to *M. tuberculosis* after exposure to five stress conditions that mimic those the bacteria are exposed to within the human host. From this, we generate a master dataset of 189 sRNA candidates in *M. tuberculosis* along with their expression profiles in each stress condition. These data provide invaluable insights into the regulatory roles that these sRNAs impart on the cell, and will prove useful for dissecting the mechanisms of sRNA action in *M. tuberculosis*.

Introduction

Small noncoding RNAs (sRNAs) are short RNA molecules involved in post transcriptional regulation of gene expression, and have become recognized as important regulatory molecules for bacterial growth and stress adaptation. These regulators, which range in size from approximately 50-250nt in length, have been identified across diverse lineages of prokaryotic

species (1–4). sRNAs have been found to be important for normal growth, stress adaptation, and virulence in the species in which they have been extensively studied (5, 6). Although there are several mechanisms of sRNA action, the most common is to repress the translation of a set of target genes to fine-tune bacterial gene expression. This occurs through an interaction between the sRNA and its mRNA targets, usually at the 5' end of the mRNA. For example, the membrane stress responsive sRNA MicL represses the expression of the membrane lipoprotein Lpp to modulate membrane stability (7), and the RyhB sRNA is induced during iron starvation to repress the expression of nonessential iron containing proteins (8). However, although much is known about sRNAs from work done in organisms such as enteric bacteria, comparatively little is known about these important regulators in several pathogenic species, including *Mycobacterium tuberculosis*.

M. tuberculosis is an important global health pathogen that currently infects one third of the world's population and is the leading cause of mortality due to an infectious disease. *M. tuberculosis* causes disease by infecting the human lung, where it resides inside of alveolar macrophages. Due to its intracellular niche of a host immune cell, the bacteria are faced with a wide array of deadly stresses that they must overcome to establish an infection (9–11). However, very little is currently known regarding sRNAs and their regulatory effects in *M. tuberculosis.* Although sRNA discovery studies have been performed in *M. tuberculosis*, searches have been limited to logarithmic- and stationary-phases of growth, and therefore stress-adaptation sRNAs likely remain undiscovered (12–14).

Given the relative paucity of information on the identity and function of sRNAs in mycobacteria, we used an sRNA-Sequencing protocol on RNA from *M. tuberculosis* and created a computational sRNA search tool, BS_finder. We used this search tool to mine the data for sRNAs, and validated that BS_finder accurately identifies and annotates sRNAs. We then used this sRNA-discovery pipeline on RNA harvested from *M. tuberculosis* exposed to five pathogenically relevant stress conditions. Using this approach, we identifed a reference dataset of 189 sRNA candidates and their expression profiles in *M. tuberculosis*.

Results

Creation of an sRNA-discovery pipeline

In order to comprehensively identify and annotate sRNAs in *M. tuberculosis* we first used a modified version of a bacterial small RNA-Sequencing (sRNA-Seq) protocol that utilizes size-fractionation to enrich for small transcripts and ligates adapters to their natural ends (15). To identify novel candidate sRNAs in these datasets we next created a computational search tool, BS_finder (Bacterial sRNA finder). BS_finder utilizes a sliding window approach to identify small transcripts encoded completely or partially from within an intergenic region with significantly higher read depth than the surrounding area, as well as sharp 5' and 3' boundaries. We employed stringent thresholding criteria to identify 62 high-confidence sRNA candidates in *M. tuberculosis*, ranging in size from 40nt-268nt with an average size of 101nt (Fig. 2.1A). These sRNAs were named using the nomenclature proposed by Lamichhane et al., 2013, which names the sRNA based on the numerical identifier for the closest upstream coding gene (16). There was modest overlap between the sRNAs identified here and the sRNAs previously identified in

studies in *M. tuberculosis* enumerating intergenic transcripts or putative sRNAs (12, 13). Many of the previously published sRNA candidates failed to reach our depth and boundary criteria designed to distinguish them from mRNA degradation products, but differences also likely reflect alterations in culture conditions, growth phase, and methods of library preparation. However, where the 5' and 3' boundaries of *M. tuberculosis* sRNAs have been experimentally determined by 5' and 3' RACE there was excellent concordance with our analysis (13, 14, 17, 18). Of the 8 experimentally mapped *M. tuberculosis* sRNA 5' ends, BS_finder mapped all of them within 3bp of the experimentally defined end (Fig. 2.1B, left). Annotation of 3' boundaries by BS_finder was also highly accurate, albeit less accurate than annotation of 5' ends, with 4 out of 6 ends mapped within 3bp of the experimentally defined end (Fig. 2.1B, right).

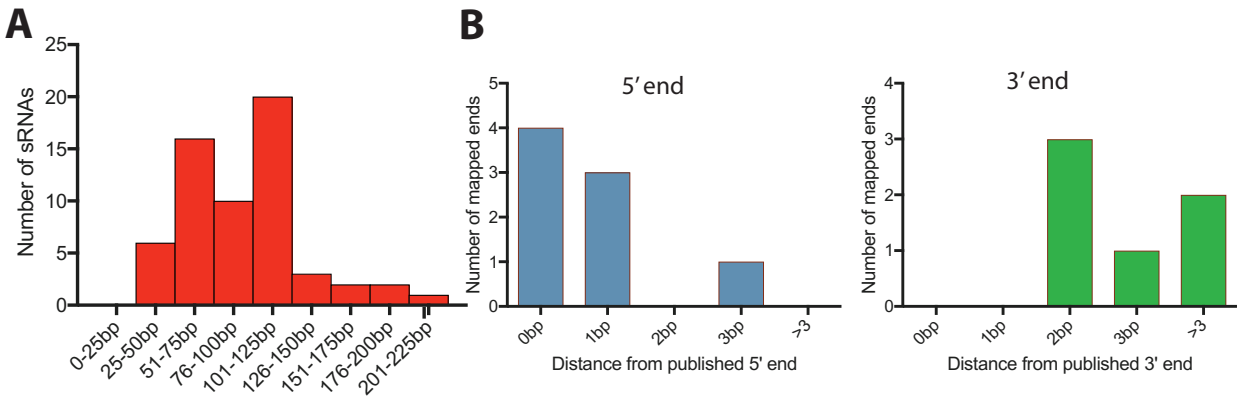


Figure 2.1: BS_finder identifies sRNAs in *M. tuberculosis* with highly accurate boundaries. (A) Size distribution of sRNAs identified by BS_finder run on sRNA-Sequencing datasets generated from *M. tuberculosis* grown in rich medium at log phase growth. (B) Distance between BS_finder-annotated ends and experimentally defined ends for a subset of *M. tuberculosis* sRNAs.

Essentiality calling of *M. tuberculosis* sRNAs using saturating transposon mutagenesis

In order to assess essentiality of the sRNAs discovered by BS_Finder, we used Deep-sequencing of transposon insertion libraries (TnSeq). TnSeq has traditionally been insensitive to sRNAs and other small genes due to their small size and the resulting relatively small number of transposon insertion sites in these features. This, combined with the low saturation levels of TnSeq libraries causes an inability to determine the essentiality of small features. To overcome this limitation, we used a large-scale analysis of 14 independent TnSeq libraries in *M. tuberculosis* generated during growth in nutrient rich medium. Using this approach, 5 high confidence novel sRNA candidates (ncRv0638, ncRv0810c, ncRv12783c, ncRv13418cA, ncRv13418cB), and one previously identified sRNA, ncRv3583A, were found to be essential for *in vitro* growth in nutrient rich medium. All five of the essential sRNAs identified in this analysis appear to share transcriptional start sites (TSS) with essential ORFs based on TSS mapping data (19), suggesting that they reflect the processed 5'UTRs of longer transcripts. Processing of 5' and 3' UTRs to generate independent trans-acting sRNAs have been described in several organisms (20, 21). Thus, the association of these RNAs with essential ORFs does not eliminate the possibility that each represents an independently essential element. For example, ncRv12783c shares a TSS with the ORF Rv2783c and 1 of the 2 TA sites between the annotated 3' end of ncRv12783c and the translational start of Rv2783c tolerated a small number of insertions, supporting the possibility that Rv2783c and ncRv12783c are independently essential RNA species. It is also possible that other small transcripts within this list are essential but cannot be confidently identified as such because they have only a single TA site. This is exemplified by the 4.5S RNA which has a single TA site that tolerated no insertions. 4.5S RNA is the RNA component of the signal recognition particle (SRP), which is essential in *E. coli* (22).

Importantly, the 4.5S RNA shares a TSS with Rv3722c in *M. tuberculosis* (19) exhibiting a similar pattern as the other putatively essential sRNAs identified here.

Discovery of sRNAs in *M. tuberculosis* involved in stress responses

We sought to define the compendium of sRNAs in *M. tuberculosis* that are both constitutively expressed and responsive to stress conditions. Although multiple sequencing-based discovery studies in *M. tuberculosis* have sought to define sRNAs, only two have used sRNA-Seq approaches, and those only searched for constitutively expressed sRNAs (12, 23). To define a comprehensive set of sRNAs in *M. tuberculosis* involved in both housekeeping and stress adaptation, we performed sRNA-Seq on RNA harvested from *M. tuberculosis* grown in nutrient rich media as well as 5 different host-like stress conditions: iron limitation, oxidative stress with *tert*-Butyl hydroperoxide, detergent-induced membrane stress, acid stress, and nutrient starvation. We used BS_finder to identify sRNAs in the sRNA-Seq datasets and to define 5' and 3' boundaries (23). We defined a master set of 189 *M. tuberculosis* candidate sRNAs (Appendix Table A1.1). Of these, 103 were not identified by previous sequencing-based sRNA discovery studies, highlighting the value of performing sRNA discovery in stress conditions (12, 23, 24). Moreover, our dataset provides accurate coordinates of sRNA 5' and 3' ends and thus can serve as a reference source for the field.

Many of the sRNAs were regulated by stress, with 82 having greater than 6-fold differential expression in at least one condition (Fig. 2.2). We distinguished 2 groups of conditionally expressed sRNAs, inducible and repressible (Fig. 2.2), and a third group of sRNAs for which we

did not find evidence of robust conditional regulation. As expected, most inducible sRNAs demonstrated strong induction upon exposure to a single stress. For example, expression of ncRv13661A, which has been proposed to be a homologue of the 6S RNA (25) increased 48-fold in nutrient limitation. 6S RNA in bacteria such as *E. coli* has been shown to accumulate in stationary phase where it acts as a decoy for RNA polymerase bound to the log-phase growth sigma factor, thus preferentially allowing for transcription by alternative sigma factor-containing RNA polymerases. Thus, induction of ncRv13661A during starvation is in agreement with 6S expression patterns in other bacteria (26, 27). We also identified other sRNAs with strong induction in specific stresses, such as ncRv11429 which was induced greater than 300-fold in SDS-mediated membrane stress. Interestingly, a small set of sRNAs was induced in multiple stresses, including three sRNAs, ncRv11803, ncRv11846, and ncRv12659, which were highly induced in three different stress conditions (Fig. 2.2).

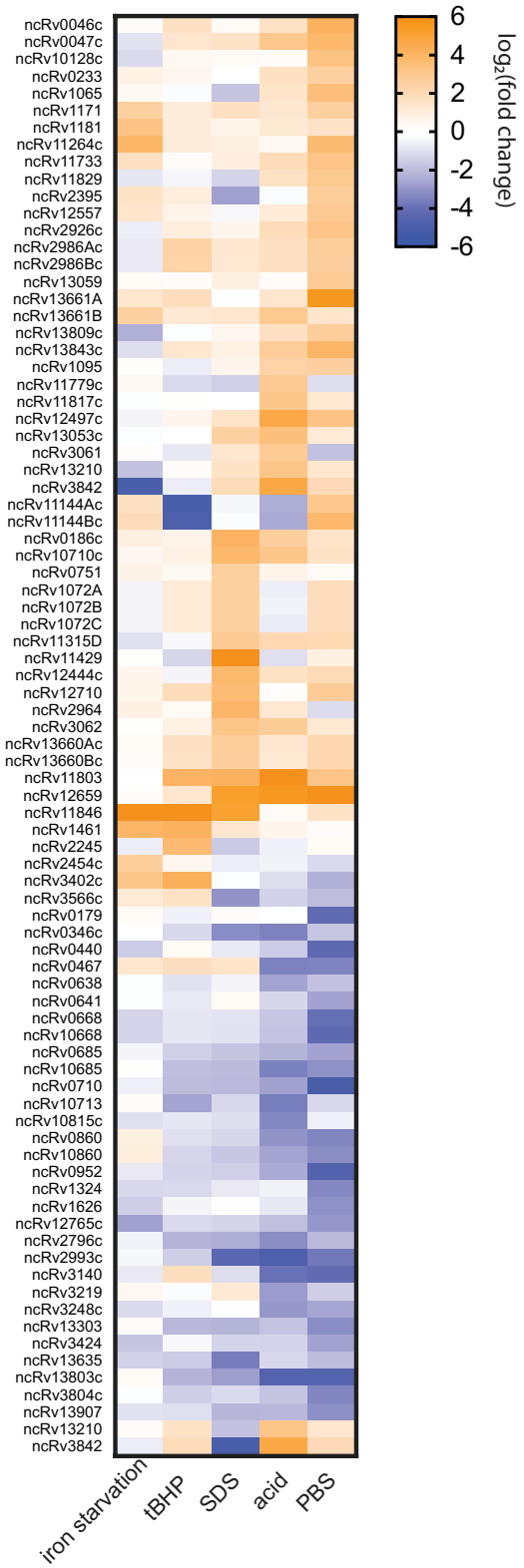


Figure 2.2. Discovery of sRNAs in *M. tuberculosis* involved in stress responses. Heat map of the 82 *M. tuberculosis* sRNAs significantly differentially expressed across 3 biological replicates ($p<0.05$, fold change ≥ 6) in at least one stress condition.

Discussion

Although much is known about sRNAs in enteric bacteria, relatively little is understood about these regulators in mycobacteria. The first requirement for the study of sRNAs in any bacterial species is discovery, and high-throughput methods for this essential step have been lacking. Here, we applied an sRNA-Sequencing technique to generate large, sRNA-rich datasets and created the computational tool BS_finder to mine these datasets for novel sRNAs. We showed that BS_finder identifies sRNAs and their boundaries with high levels of accuracy, which is important for downstream analysis. To gain insights into sRNA function and expand the compendium of known mycobacterial sRNAs we additionally performed essentiality calling and sRNA identification and profiling under diverse stress conditions.

Through a large-scale analysis of TnSeq datasets, we were able to ascribe essentiality to 6 sRNA candidates in *M. tuberculosis*, five of which had not been identified in previous studies. Interestingly, all 6 of these putative sRNAs share a TSS with an essential protein-coding gene. This operonic nature of the sRNAs complicates the assigned essentiality, as transposons in these sRNAs would also likely disrupt expression of the downstream essential gene. Thus, the lack of transposon insertions in each of these sRNAs could be due to either an independently essential function of the sRNA or polar effects of the transposon on the protein coding gene in the operon. Future experiments will be necessary to dissect the essentiality and function of each of these sRNA candidates.

By applying our sRNA-discovery pipeline to *M. tuberculosis* exposed to stress, we identified a large number of previously undiscovered sRNA candidates and additionally gained important information regarding their potential functions. Most sRNAs have been found to regulate the bacterial response to a single stress condition, and therefore knowledge about the induction pattern of an sRNA provides key insights into its most likely regulatory role. For example, we identified one sRNA, ncRv11429 that is induced greater than 300-fold during exposure to detergent-mediated membrane stress (Fig. 2.2). This sRNA is not strongly induced in any other stress condition tested, and thus its robust, specific induction implicates ncRv11429 in the membrane stress response pathway. Membrane stress responses in other bacteria, particularly the Gammaproteobacteria *E. coli* and *Salmonella*, are well characterized and rely on several different sRNAs such as RybB, MicL, and CpxQ regulating the translation of membrane proteins (7, 21, 28). It will thus be interesting to determine whether ncRv11429 plays a similar role in the mycobacterial envelope stress response, and the extent to which conservation of expression profiles for sRNAs can be used to infer function. Importantly, our work also provides a template for conducting similar systematic studies of sRNAs in other prokaryotic species to discover novel sRNAs and gain valuable insights into their regulatory roles. This will likely prove particularly useful for other important human pathogens for which there is a dearth of information on sRNAs.

Materials and Methods

Bacterial strains and growth conditions. *Mycobacterium tuberculosis* H37Rv was cultured in 7H9 medium (Difco Laboratories) with 2% v/v glycerol, 0.05% v/v tween-80, and 10% oleic acid-albumin-dextrose-catalase to mid-log phase (OD_{600}=0.2-0.8), prior to harvesting RNA (sRNA identification and essentiality calling) or transfer to stress media (sRNA identification and profiling in stress). For oxidative stress and membrane stress, cells were pelleted and then resuspended in 7H9 medium with 2% v/v glycerol, 0.05% v/v tyloxapol, and 10% albumin-dextrose-NaCl, in addition to either *tert*-Butyl hydroperoxide (tBHP, oxidative stress) or sodium dodecyl sulfate (SDS, membrane stress). Acid stress was stimulated by adjusting the pH of the above 7H9 medium to pH 4.5. Starvation was stimulated by transferring the cells into PBS with 0.05% tyloxapol. . Iron starvation was performed as described previously (29). Briefly, cells were grown to OD_{600}=0.5-1.0, then washed once with an equal volume of low iron medium and diluted to OD_{600}=0.1 in the same media. Cells were then grown to ~OD_{600}=1.0, diluted back to OD_{600}=0.1 in low iron media, and grown to OD_{600}=0.2-0.4 before adding 50µg/mL of the iron chelator deferoxamine (DFO). Cells were exposed to stress for 4 hours (tBHP, SDS, acid) or 24 hours (PBS, iron starvation) prior to harvesting RNA for sRNA-Sequencing.

RNA extraction, library construction, and sequencing. Cell pellets were resuspended in TRIzol reagent (Ambion, Life Technologies) and lysed using a FastPrep 24. RNA was purified using the Direct-Zol miniprep kit (Zymo Research) and DNA was digested using TURBO DNase (Ambion, Life Technologies).

Size selection to achieve an approximately 300nt cutoff was performed using RNA Clean and Concentrator-25 columns (Zymo Research) according to manufacturer's instructions with the following modification: preparation of the adjusted RNA Binding Buffer was done by adding an equal volume of RNA binding buffer and 70% ethanol. The small RNA fraction was then depleted of rRNA using the RiboZero-Bacteria kit (Illumina). The small RNA library was then prepared as described previously (15) using a modified TruSeq Small RNA-Sequencing kit (Illumina). Briefly, RNA was dephosphorylated using RppH enzyme (New England Biolabs) followed by an ethanol precipitation step. 3' and 5' adapters were then ligated to the RNA followed by reverse transcription and PCR amplification. The libraries were then purified using Agencourt AMPure XP beads (Beckman Coulter). 150bp paired-end sequencing was then performed on a MiSeq sequencer. Sequenced reads were aligned to the Mycobacterium tuberculosis NC_000962.3 genome (National Center for Biotechnology Information database) using the Burrows-Wheeler alignment tool (30).

sRNA identification Per-base coverage of the genome was obtained using the genomecov tool of the BEDtools suite (31). sRNAs were then identified using BS_finder (Bacterial sRNA finder), which uses a sliding window approach to sRNA identification. Briefly, a sliding window scans the per-base read depth and searches for positive slope (the intensity of which can be modified to adjust stringency) across the window, which demarcates a 5' end (when searching the plus strand) or a 3' end (when searching the minus strand) if the position passes a modifiable read depth threshold. A nested sliding window then searches for decreases in slope across the window, which demarcates a 3' or 5' end, respectively. If the discovered feature is not entirely

within an ORF and has a 5' end at least 50bp away from the nearest ORF translational start site then it is output as a candidate sRNA. BS_finder was run on the dataset using default parameters: a sliding window size of 2bp with a slope threshold of 3 and read depth threshold of 500.

Statistical Analysis of Essentiality

A hierarchical Bayesian model that was conditioned on local sequence (matching the NP motif or not) was developed for analyzing essentiality. Genes are modeled as a mixture of two classes of essentiality, represented by an indicator variable z_i: z_i=1 for essential genes (which are represented by insertions in very few, if any, libraries), and z_i=0 for non-essential genes (which are highly saturated in most libraries). The model uses Binomial distributions for each class to assess the likelihood of having observed the number of replicates with insertions, summed across all TA sites in the gene. As sites matching the non-permissive motif have a significantly lower probability of insertion compared to other sites, the likelihood of the observations is decomposed into a product of two different Binomial distributions, each with their own set of parameters. In this way non-essential genes are not penalized for containing unoccupied non-permissive sites, as these sites have a higher probability of being empty compared to other sites.

The parameters of the Binomial distributions (θ) are assigned Beta priors (which in turn are assigned weakly-informative hyperpriors of their own) and their values are estimated by sampling from distributions conditioned on the data. Ultimately, we are interested in the probability that a given gene is essential (i.e. that the true unobservable state of essentiality for

the *i*th gene, z_i, equals one). This posterior probability effectively involves solving the following integral (marginalizing over the unknown parameters):

$$p(z_i = 1|Y^{(i)}, \Phi_1, \omega_1)$$
$$\propto \text{Bernoulli}(z_i = 1|\omega_1) \times \int \text{Binomial}(k_m^{(i)}|n_m^{(i)}, \theta_m^{(i)}) \, \text{Beta}(\theta_m^{(i)}|\Phi_{1,m}) \, d\theta_m^{(i)}$$
$$\times \int \text{Binomial}(k_{\neg m}^{(i)}|n_{\neg m}^{(i)}, \theta_{\neg m}^{(i)}) \, \text{Beta}(\theta_{\neg m}^{(i)}|\Phi_{1,\neg m}) \, d\theta_{\neg m}^{(i)}$$

where $Y^{(i)} =< k^{(i)}, n^{(i)} >$ represents the insertion data for the *i*th gene (*k* insertions at *n* TA sites), Φ_1 represents a vector of all the parameters for essential genes, and ω_1 represents the mixture probability (i.e. probability of any given gene being essential). The first integral considers TA sites in gene *i* that match (m) the NP motif, while the second integral considers those sites that do not match ($\neg m$) the motif. As this formula does not have a simple analytical solution, we use a Markov-Chain Monte Carlo (MCMC) sampling procedure (see Supplemental Material for details) to generate estimates of the posterior probability, effectively integrating over the unknown parameters. Finally, to set a significance threshold that controls the false discovery rate (FDR) at 5%, we apply a multiple-tests correction akin to the Benjamini-Hochberg procedure, but applied to posterior probabilities rather than p-values.

References

1. Wagner EG, Rompy P(2015) Small RNAs in Bacteria and Archaea: Who They Are, What They Do, and How They Do It. *Adv Genetics* 90: 133-208

2. Livny J, Waldor MK (2007) Identification of small RNAs in diverse bacterial species. *Curr Opin Microbiol* 10(2):96–101.

3. Toledo-Arana A, et al. (2009) The Listeria transcriptional landscape from saprophytism to

virulence. *Nature* 459(7249):950–956.

4. Gottesman S (2004) The Small RNA Regulators of *Escherichia coli* : Roles and Mechanisms. *Annu Rev Microbiol* 58(1):303–328.

5. Chabelskaya S, Gaillot O, Felden B (2010) A Staphylococcus aureus small RNA is required for bacterial virulence and regulates the expression of an immune-evasion molecule. *PLoS Pathog* 6(6). doi:10.1371/journal.ppat.1000927.

6. Papenfort K, et al. (2006) σE-dependent small RNAs of Salmonella respond to membrane stress by accelerating global omp mRNA decay. *Mol Microbiol* 62(6):1674–1688.

7. Guo MS, et al. (2014) MicL, a new σE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev* 28(14):1620–1634.

8. Jacques JF, et al. (2006) RyhB small RNA modulates the free intracellular iron pool and is essential for normal growth during iron limitation in Escherichia coli. *Mol Microbiol* 62(4):1181–1190.

9. Rohde KH, Abramovitch RB, Russell DG (2007) Mycobacterium tuberculosis Invasion of Macrophages: Linking Bacterial Gene Expression to Environmental Cues. *Cell Host Microbe* 2(5):352–364.

10. Schnappinger D, et al. (2003) Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages. *J Exp Med* 198(5):693–704.

11. Stallings CL, Glickman MS (2010) Is Mycobacterium tuberculosis stressed out? A critical assessment of the genetic evidence. *Microbes Infect* 12(14–15):1091–1101.

12. Wang M, et al. (2016) An automated approach for global identification of sRNA-encoding regions in RNA-Seq data from Mycobacterium tuberculosis. *Acta Biochim Biophys Sin (Shanghai)* 48(6):544–553.

13. Arnvig KB, et al. (2011) Sequence-Based Analysis Uncovers an Abundance of Non-Coding RNA in the Total Transcriptome of Mycobacterium tuberculosis. *PLoS Pathog* 7(11)

14. Arnvig KB, Young DB (2009) Identification of small RNAs in Mycobacterium tuberculosis. *Mol Microbiol* 73(3):397–408.

15. Gómez-Lozano M, Marvig RL, Molin S, Long KS (2014) Identification of Bacterial Small RNAs by RNA Sequencing. *Methods in Molecular Biology (Clifton, N.J.)*, pp 433–456.

16. Lamichhane G, Arnvig KB, McDonough KA (2013) Definition and annotation of (myco)bacterial non-coding RNA. *Tuberculosis* 93(1):26–29.

17. DiChiara JM, et al. (2010) Multiple small RNAs identified in Mycobacterium bovis BCG are also expressed in Mycobacterium tuberculosis and Mycobacterium smegmatis. *Nucleic Acids Res* 38(12):4067–4078.

18. Miotto P, et al. (2012) Genome-Wide Discovery of Small RNAs in Mycobacterium tuberculosis. *PLoS One* 7(12). doi:10.1371/journal.pone.0051950.

19. Shell SS, et al. (2015) Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape. *PLOS Genet* 11(11):e1005641.

20. Lalaouna D, Simoneau-Roy M, Lafontaine D, Massé E (2013) Regulatory RNAs and target mRNA decay in prokaryotes. *Biochim Biophys Acta - Gene Regul Mech* 1829(6–7):742–747.

21. Chao Y, Vogel J (2016) A 3' UTR-Derived Small RNA Provides the Regulatory Noncoding Arm of the Inner Membrane Stress Response. *Mol Cell* 61(3):352–363.

22. Brown S, Maurillf ANI, Fourniek J (1984) The 4-S S RNA Gene of Escherichia coli is Essential for Cell Growth. *Mol Hid* 178:533–550.

23. Dejesus MA, et al. (2017) Comprehensive essentiality analysis of the Mycobacterium tuberculosis genome via saturating transposon mutagenesis. *MBio* 8(1). doi:10.1128/mBio.02133-16.

24. Arnvig KB, et al. (2011) Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis. *PLoS Pathog* 7(11). doi:10.1371/journal.ppat.1002342.

25. Hnilicová J, et al. (2014) Ms1, a novel sRNA interacting with the RNA polymerase core in mycobacteria. *Nucleic Acids Res* 42(18):11763–11776.

26. Wassarman KM (2012) 6S RNA: A regulator of transcription. *Regul RNAs Prokaryotes* 65(August):109–130.

27. Gottesman S, et al. (2006) Small RNA regulators and the bacterial response to stress. *Cold Spring Harb Symp Quant Biol* 71:1–11.

28. Klein G, Raina S (2017) Small regulatory bacterial RNAs regulating the envelope stress response. *Biochem Soc Trans* 45(2):417–425.

29. Kurthkoti K, et al. (2017) The capacity of mycobacterium tuberculosis to survive iron starvation might enable it to persist in iron- deprived microenvironments of human granulomas. *MBio* 8(4):e01092-17.

30. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinforma Appl NOTE* 25(1610):2078–2079.

31. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma Appl NOTE* 26(6):841–84210.

Chapter 3

Discovery and characterization of the mycobacterial small RNA MrsI, which mediates an anticipatory iron sparing response

Elias R. Gerrick[a], Thibault Barbier[a], Michael R. Chase[a], Raylin Xu[a], Josie François[a], Vincent H. Lin[a], Matthew Szucs[b], Jeremy M. Rock[a], Rushdy Ahmad[b], Brian Tjaden[c], Jonathan Livny[b], Sarah M. Fortune[a,d]

[a]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA; [b]The Broad Institute of MIT and Harvard, Cambridge, USA; [c]Computer Science Department, Wellesley College, Wellesley, Massachusetts 02481, USA; [d]The Ragon Institute of MGH, Harvard and MIT, Cambridge, Massachusetts 02139, USA

Author contributions: E.R.G. performed all experiments except proteomics, and T.B., R.X., J.F., and V.L. assisted with experiments. M.S. and R.A. performed and analyzed proteomics experiment. E.R.G., T.B., M.R.C., J.M.R., B.T., J.L., and S.M.F. designed research. E.R.G., T.B., and S.F. wrote the manuscript. This chapter comprises work that has been submitted at the time of writing, with modifications.

Abstract

One key to the success of *Mycobacterium tuberculosis* as a pathogen is its ability to reside in the hostile environment of the human macrophage. Bacteria adapt to stress through a variety of mechanisms including the use of small regulatory RNAs (sRNAs), which post-transcriptionally regulate bacterial gene expression. However, very little is currently known about mycobacterial sRNA-mediated riboregulation. To date, mycobacterial sRNA discovery has been performed primarily in log phase growth and no direct interaction between any mycobacterial sRNA and its targets has been validated. Previously, we performed large-scale sRNA discovery and expression profiling in *M. tuberculosis* during exposure to five pathogenically relevant stresses. From these data, we identified a subset of sRNAs that are highly induced in multiple stress conditions. We focused on one of these sRNAs, ncRv11846, here re-named Mycobacterial regulatory sRNA in Iron (MrsI). We characterized the regulon of MrsI and show for the first time in mycobacteria that it regulates one of its targets, *bfrA*, through a direct binding interaction. MrsI mediates an iron-sparing response that is required for optimal survival of *M. tuberculosis* under iron limiting conditions. However, MrsI is induced by multiple host-like stressors, which appear to trigger MrsI as part of an anticipatory response to impending iron deprivation in the macrophage environment.

Introduction

Mycobacterium tuberculosis survives in macrophages where it is exposed to an array of stresses including iron restriction, nutrient limitation, oxidative stress, low pH, and membrane stress (1–5). Bacteria adapt to these stresses through both transcriptional and post-transcriptional

responses, including the regulatory functions of *trans*-encoded small noncoding RNAs (sRNAs). Despite the global impact of *M. tuberculosis*, mycobacteria are part of a broad group of bacterial pathogens for which very little is known about sRNA-mediated regulation. For example, mycobacterial sRNAs have primarily been identified during growth in rich medium, and there have been few efforts to identify sRNAs involved in stress-responses (6–12). Additionally, there has been only minimal characterization of the few validated mycobacterial sRNAs. Indeed, just a few putative targets of a single sRNA in *M. tuberculosis* have been described and no direct interaction between an sRNA and predicted mRNA target has been validated (13). It is postulated that the rules for sRNA function defined in other bacteria extend to mycobacteria, although they lack key elements of the sRNA machinery including any obvious homologues of the sRNA chaperone proteins Hfq and ProQ (14, 15).

In other prokaryotes in which sRNAs have been characterized, sRNAs most commonly act by binding to the 5' end of *trans*-encoded mRNA targets thereby repressing translation of the mRNA and often facilitating target mRNA degradation (16). The interaction between sRNAs and their targets is initiated through a short 5-7nt sequence of perfect complementarity termed the "seed region" and extends to include a longer region with limited complementarity. Many sRNAs are critical for bacterial stress adaptation and pathogenesis, becoming strongly induced during stress exposure in order to regulate a set of targets (17–20). For example, expression of the iron-sparing sRNAs RyhB and PrrF of enteric bacteria and *Pseudomonas aeruginosa*, respectively, is highly upregulated during iron starvation and these sRNAs repress mRNAs encoding nonessential iron containing proteins (21, 22).

Given the relative paucity of information on the identity and function of sRNAs in mycobacteria, we previously used high-throughput methodologies to comprehensively identify sRNAs expressed in *M. tuberculosis* during exposure to five *in vivo* relevant stress conditions (Chapter 2). Interestingly, we found that a subset of sRNAs is highly induced in multiple stresses. One of these, here renamed MrsI or ncRv11846, is induced during exposure to iron starvation, oxidative stress, and membrane stress. We identify the regulon of MrsI in both pathogenic and nonpathogenic species and determine direct targets by experimentally validating its seed region and the complementary binding site in the 5' UTR of regulated mRNAs. MrsI is critical for normal adaptation to iron deprivation in *M. tuberculosis,* and its induction in oxidative stress leads to more rapid regulation of the target *bfrA* upon iron limitation. This supports a model by which *M. tuberculosis* uses oxidative and membrane stress responses as gateways to enter an anticipatory iron sparing state.

Results

ncRv11846 is a promiscuously induced sRNA in *M. tuberculosis*

In our previous sRNA discovery experiments, we identified 82 *M. tuberculosis* sRNAs with robust differential expression in stress. Most of the induced sRNAs were specific to a single stress but a smaller set of sRNAs was induced in multiple stresses, including ncRv11803, ncRv11846, and ncRv12659, which were highly induced in three different stress conditions (Chapter 2, Fig. 2.2). We focused on one of the promiscuously induced sRNAs, ncRv11846, which is the most highly induced sRNA during iron starvation and oxidative stress, and is also

highly induced in membrane stress (Fig. 2.2, Fig. 3.1A). ncRv11846 is approximately 100nt long and is predicted to be highly structured (Fig. 3.1B). Transcription of the sRNA starts approximately 100bp upstream of Rv1847, a gene of unknown function that appears to be an independent transcript, given that we have previously identified its distinct transcriptional start site (Fig. 3.1A)(25). In addition, ncRv11846 has a predicted rho-independent terminator at the 3' end of the sRNA (Fig. 2B)(26)**.** We used Northern blot analysis to confirm that ncRv11846 is a small independent transcript and is induced in iron starvation and oxidative stress (Appendix Fig. A2.1). ncRv11846 is highly conserved across *Mycobacteriaceae* and *Nocardiaceae*, particularly in a 30nt stretch preceding the terminator (Fig. 3.1C). A binding site for the mycobacterial iron-dependent transcription factor IdeR is present near the transcriptional start site of this sRNA in both *M. tuberculosis* and the nonpathogenic relative *Mycobacterium smegmatis* (27). This is consistent with the increased abundance of the sRNA in growth in iron limiting conditions, suggesting that ncRv11846 might be involved in the bacterial response to iron deprivation.

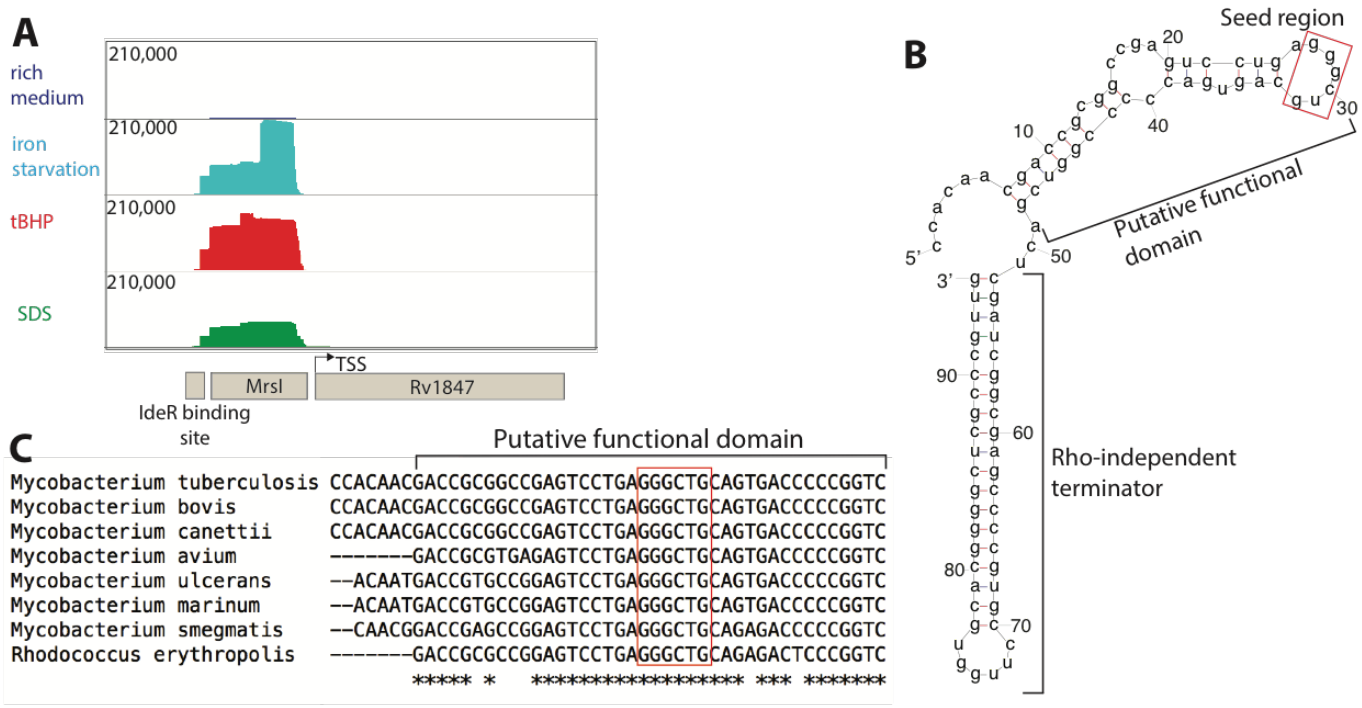


Figure 3.1. ncRv11846/MrsI is a promiscuously induced, highly structured and conserved sRNA. (A) Read distribution around the ncRv11846/MrsI locus in sRNA-Seq data in rich medium and MrsI-inducing stress conditions. The linear scale (shown on the left) is the numbers of reads. (B) Predicted minimum free energy secondary structure of ncRv11846/MrsI (mFold). The putative functional domain and rho-independent terminator are marked. The red box denotes the 6nt seed region. (C) Alignment of the putative functional domain of ncRv11846/MrsI homologues from selected actinobacterial species. The red box denotes the sRNA seed region.

The ncRv11846 homologue in *M. smegmatis* is an iron-sparing sRNA

To develop *M. smegmatis* as a model for dissecting the mechanism of ncRv11846 action, we defined the expression pattern of its homologue in *M. smegmatis*. This homologue had previously been identified in a screen for sRNAs in *M. smegmatis*, and both ends were mapped by RACE analysis (10). The promoter of the sRNA in *M. smegmatis*, defined here as the 200bp upstream of the 5' end, was fused to a luciferase reporter to measure induction in each stress condition. Iron deprivation resulted in high levels of induction but, unlike in *M. tuberculosis*, neither oxidative stress nor membrane stress induced expression of the sRNA in *M. smegmatis*

(Fig. 3.2A). This suggests that the core function of the sRNA across mycobacterial species is mediated during iron deprivation. Consequently, we named the sRNA mycobacterial regulatory sRNA in iron (MrsI).

To investigate the function of MrsI on mycobacterial gene expression and adaptation, we constructed a deletion mutant (Δ*mrsI*) in *M. smegmatis*. When the cells were grown in iron rich medium, no growth difference was observed (Fig. 3.2B, left). However, in iron-limited medium, *M. smegmatis* Δ*mrsI* reached a lower final optical density, while MrsI was strongly induced in wild type cells (Fig. 3.2B, right, Appendix Fig. A2.2A). Complementation on an episome restored both MrsI levels and growth of the deletion mutant in iron limiting conditions (Fig. 3.2B, right, Appendix Fig. A2.2B).

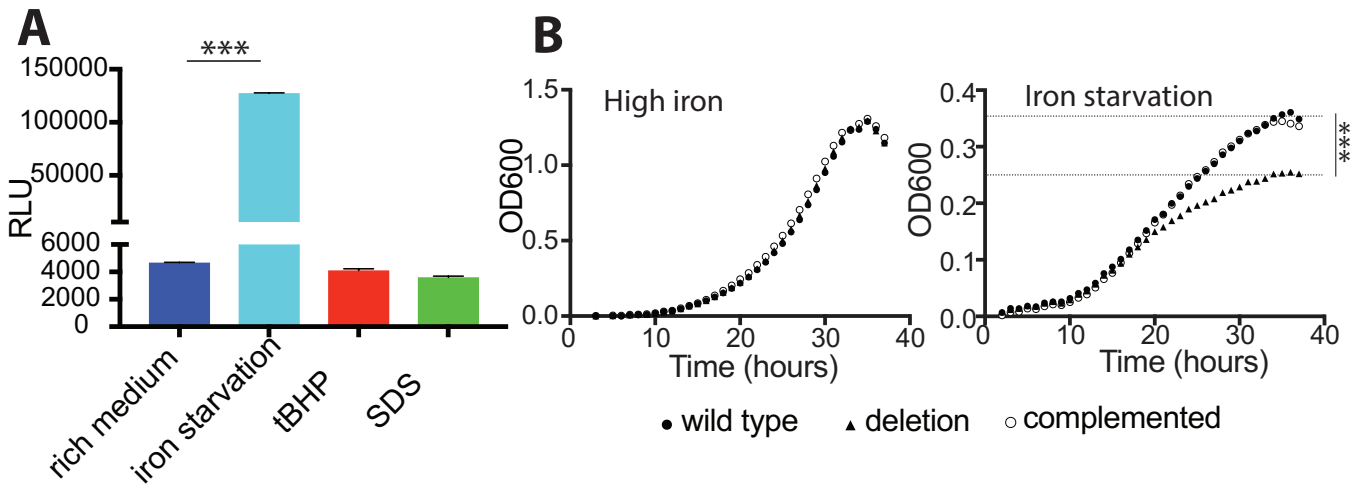


Figure 3.2. ncRv11846/MrsI is induced during iron limitation in *M. smegmatis* and is critical for normal adaptation to iron starvation. (A) The promoter of the ncRv11846/MrsI homologue in *M. smegmatis* was fused to a luciferase reporter and cells were exposed to stress before measuring luciferase activity. ***p<0.001 (Unpaired T-test). Error bars represent SD of 3 replicates. (B) Growth curves of MrsI strains in medium with (left) and without (right) iron. Dashed lines indicate the final optical density reached for strains with and without *mrsI*. ***p< 0.0001 (Unpaired T-test).

This phenotype is reminiscent of the growth defect observed for deletion of the iron-sparing sRNA RyhB *in Escherichia coli* and PrrF1 and PrrF2 in *Pseudomonas aeruginosa* (22, 28). We thus hypothesized that MrsI functions as an iron sparing sRNA during iron limitation, repressing the expression of nonessential iron-containing proteins to restrict iron for essential functions (28, 29). To test this hypothesis, we used transcriptional profiling to identify genes whose expression is regulated by MrsI during iron deprivation in *M. smegmatis*. 20 genes, organized in 12 transcription units, had significantly higher abundance in the *mrsI* deletion strain as compared to the wild type and complemented strains, consistent with repression by MrsI (Fig. 3.3A, red data points, Appendix Fig. A2.3). Of these 12 transcripts, 8 code for nonessential proteins that are predicted to bind iron or be involved with iron metabolism, including the NiFe hydrogenase maturation factor HypF, the bacterioferritin BfrA, and the ferredoxin reductase FprA (Appendix Table A2.1). We confirmed the higher transcript levels of *bfrA* and *hypF* in the *mrsI* deletion during iron limitation by RT-qPCR (Appendix Fig. A2.4A). Additionally, we performed proteomics on the same strains as above during iron limitation. Gene Set Enrichment Analysis confirmed significant enrichment of the differentially expressed proteins within the gene set of targets identified by transcriptomics (30). These results support a model in which MrsI acts as an iron sparing sRNA in *M. smegmatis*, whereby it represses the expression of nonessential iron containing proteins to reserve the dwindling iron stores for essential proteins.

MrsI regulates *bfrA* through a direct interaction

We hypothesized that MrsI directly regulates transcripts encoding iron-containing proteins. To test this, we first used the sRNA target prediction softwares TargetRNA2 and CopraRNA to agnostically predict direct mRNA interactors of MrsI in *M. smegmatis* (31, 32). However, of the 115 direct MrsI targets predicted by these tools, none appeared to be regulated in a MrsI-dependent fashion in the expression analysis. We therefore sought to validate our experimentally identified putative targets by manually identifying regions involved in MrsI-mRNA interaction. sRNAs initiate contact with their targets through a short 6-8nt sequence of perfect complementarity termed the seed region. Because seed regions are usually highly conserved, we focused on the stretch of nucleotides encompassing the 5' hairpin (Fig. 3.1C) (33, 34). The 7nt apical loop of this hairpin was the most promising site, as seed regions are generally single stranded to allow for intermolecular base pairing with target mRNAs (Fig. 3.1B , red box) (33). Importantly, the 5' ends of 7 of the 12 differentially expressed transcriptional units have perfect complementarity to 6nt of the loop (Appendix Table A2.1, Appendix Fig. A2.4B).

To test if MrsI directly binds to its mRNA targets through the putative 6nt seed region we performed reciprocal mutation of *mrsI* and one candidate target, *bfrA* (Fig. 3.3B, C). The 5' UTR of *bfrA* was fused to the exogenous gene *zeoR,* creating a *bfrA-zeoR* expression reporter while *mrsI* was inducibly expressed from an episomal plasmid. We measured *bfrA-zeoR* expression during iron limitation by RT-qPCR. We then introduced point mutations into the *mrsI* seed region and the putative binding site in the *bfrA* 5' UTR and assessed the effect on *bfrA-zeoR* levels. In a strain containing $mrsI^{WT}bfrA^{WT}$-*zeoR*, induction of MrsI repressed *bfrA-zeoR*

expression as expected (Fig. 3.3B, C). Introduction of a G41A mutation into the putative seed region of *mrsI* abrogated regulation of the target (Fig. 3.3C). To determine whether this loss of regulation was due to true disruption of a seed sequence, we introduced a compensatory C→T mutation to the predicted binding site in the 5' UTR of *bfrA-zeoR* (*bfrA*$^{C\rightarrow T}$-*zeoR*). The *bfrA*$^{C\rightarrow T}$-*zeoR* mutation restored regulation by *mrsI*^{G41A} (Fig. 3.3B, C), whereas *mrsI*^{WT} failed to regulate *bfrA*$^{C\rightarrow T}$-*zeoR* (Fig. 3.3C). These data demonstrate that MrsI regulates *bfrA* by direct interaction between a 6nt seed sequence and a perfectly complementary region in the *bfrA* 5' UTR. This is the first validated direct sRNA target in mycobacteria.

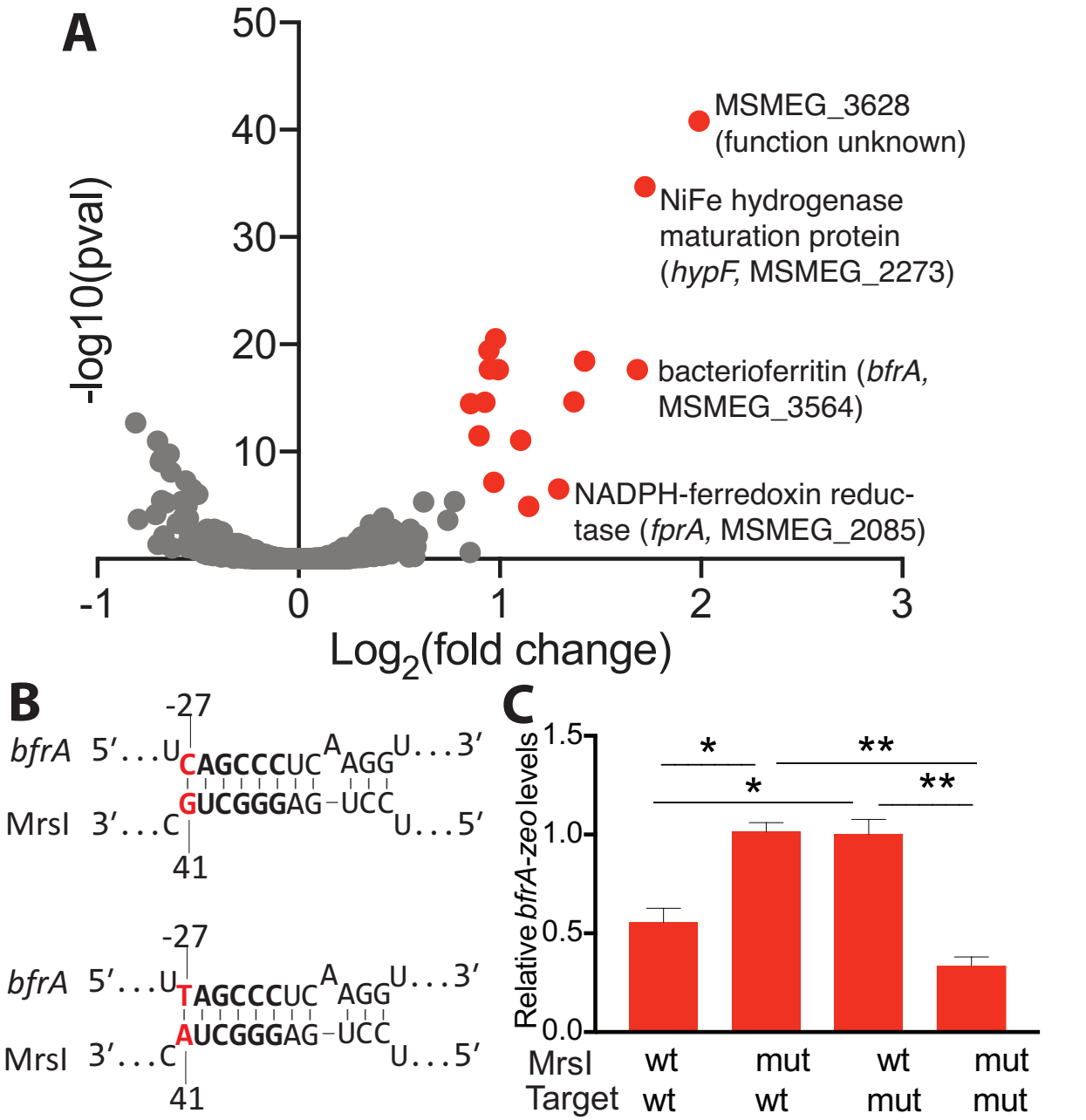


Figure 3.3. MrsI is an iron sparing sRNA in *M. smegmatis* and binds directly to mRNA targets through a 6nt seed region. (A) Volcano plot of transcriptomics for WT and Δ*mrsI M. smegmatis* after 6 hours of iron starvation. Red data points indicate genes with elevated levels in the deletion strain compared to both wild type and complemented strains (fold change greater than 1.5, p<0.05). (B) Schematic of the wt-wt (top) and mut-mut (bottom) binding interaction between MrsI and the target *bfrA*. The MrsI seed region is in bold, and the bases mutated for the compensatory mutation assay are in red. (C) MrsI regulates *bfrA* directly. The promoter and 5′ UTR of *bfrA* were fused to the *zeoR* gene, and reciprocal mutations were made in the putative interaction sites on MrsI and *bfrA-zeoR*. Levels of *bfrA-zeoR* in each strain were measured by RT-qPCR. *p<0.05, **p<0.005 (Unpaired T-test). Error bars represent SD of 3 replicates.

MrsI in *M. tuberculosis* mediates extensive transcriptome changes during iron deprivation

We next used transcriptional profiling to define the effects of MrsI on the *M. tuberculosis* transcriptome during iron starvation, oxidative stress and membrane stress (Fig. 3.4A). We used the mycobacterial CRISPR interference system (CRISPRi) to inducibly knock down MrsI expression (35). Knock-down of MrsI resulted in the increased expression of 118 genes, consistent with repression by MrsI (Fig. 3.4A). 106 of these genes were differentially expressed during iron deprivation, whereas 5- and 12-genes were differential in oxidative stress and membrane stress, respectively. Thus, while MrsI affects gene expression in three stresses in *M. tuberculosis*, its effects are most extensive during iron deprivation. We next defined the phenotypic consequences of MrsI-mediated riboregulation in *M. tuberculosis* during iron starvation. While no growth defect was observed upon MrsI knockdown in the absence of stress, knock-down of MrsI attenuated growth of *M. tuberculosis* during iron deprivation, similar to the phenotype observed in *M. smegmatis* (Fig. 3.4B).

Two of the genes regulated by MrsI in *M. smegmatis*, *bfrA* and *fprA,* were also regulated in *M. tuberculosis* during iron deprivation. The MrsI binding sites in both of these transcripts are perfectly conserved between *M. smegmatis* and *M. tuberculosis*, suggesting preservation of targeting regions. Additionally, both encode iron-containing proteins consistent with a conserved role for MrsI in iron sparing. To distinguish between direct and indirect effects, we identified other putative direct targets in *M. tuberculosis* by scanning the 5' ends of each of the regulated genes for MrsI binding sites. Including *bfrA* and *fprA*, 20 genes organized into 9

transcription units contained potential MrsI binding sites in the 5’ UTR. These are therefore putative direct targets (Fig. 3.4A, green box), and 17 of these genes encode predicted iron-binding proteins. Of the 20 genes, two were differential during oxidative stress and membrane stress. The more robust changes to the transcriptome during MrsI knockdown in iron limitation and the fact that 17 of the 20 putative direct MrsI targets encode nonessential iron containing proteins support a model in which MrsI functions as an iron sparing sRNA in both *M. smegmatis* and *M. tuberculosis*.

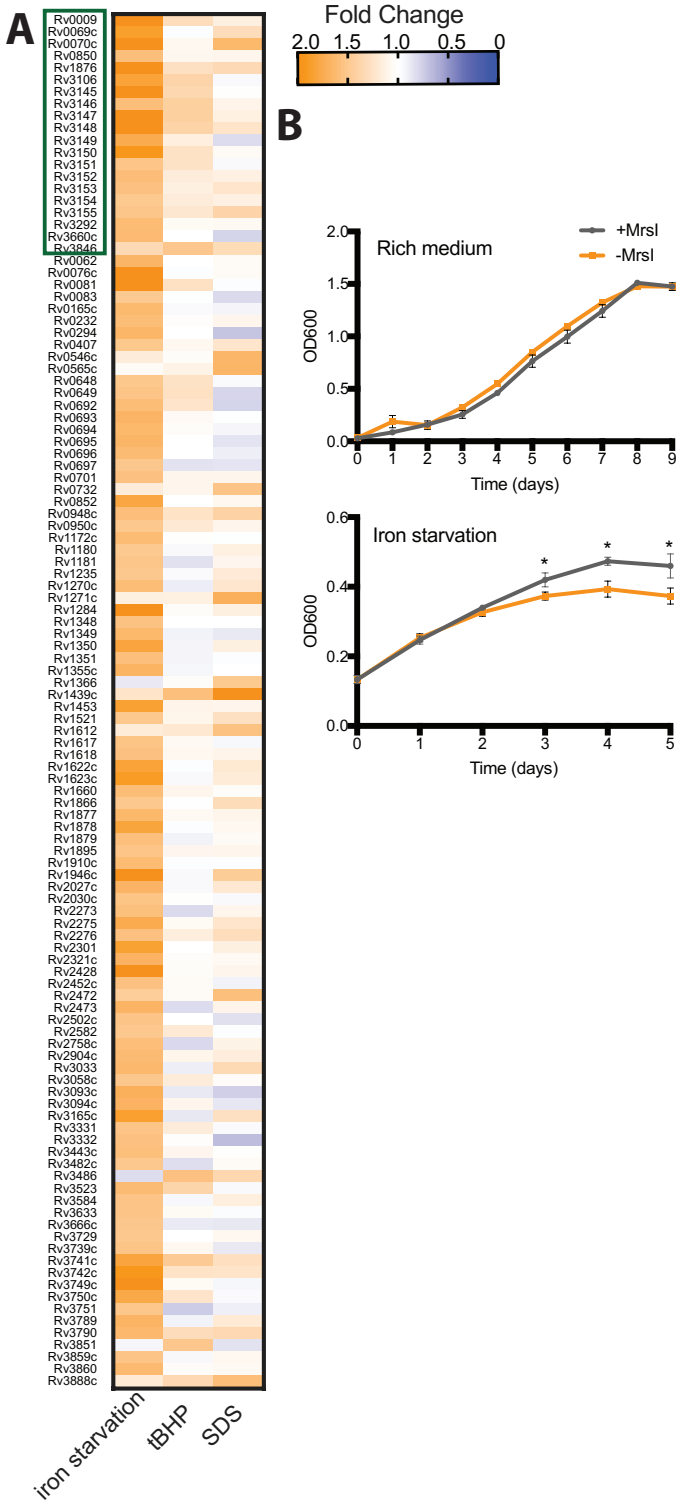


Figure 3.4. MrsI mediates an iron sparing response in *M. tuberculosis*. (A) Heat map of genes putatively directly regulated by MrsI in MrsI-inducing stress conditions. Green box indicates putative direct targets. (B) Growth curves of *M. tuberculosis mrsI* knockdown strain with and without CRIPSRi induction in rich medium (top) and during iron starvation (bottom). *p<0.05 (Unpaired T-test). Error bars represent SD of 3 replicates.

MrsI in *M. tuberculosis* mediates an anticipatory iron sparing response

That MrsI predominantly regulates iron metabolism while it is induced by multiple stresses may be indicative of a stress adaptation program that integrates multiple signals to anticipate iron starvation and facilitate a more rapid iron-sparing response. We reasoned that pre-exposure of *M. tuberculosis* to oxidative stress prior to iron starvation would lead to more rapid MrsI-mediated repression of its targets. In line with the proposed model, cells pre-exposed to oxidative stress repressed *bfrA* expression more quickly under iron limitation than cells not pre-exposed to oxidative stress (Fig. 3.5 top, time 8h). This early repression of *bfrA* is abrogated with MrsI knockdown, confirming that the effect is MrsI-dependent (Fig. 3.5 bottom, time 8h). Interestingly, with MrsI knockdown, exposure to oxidative stress alone led to increased levels of *bfrA,* suggesting that this effect is dampened by MrsI in wildtype cells (Fig. 3.5, time 4h). By 24 hours of iron starvation, *bfrA* repression was the same with or without pre-exposure to oxidative stress (Fig. 3.5, time 28h). MrsI expression is therefore induced by multiple stresses which allows for faster regulation of its targets, consistent with a model of an anticipatory response to iron starvation.

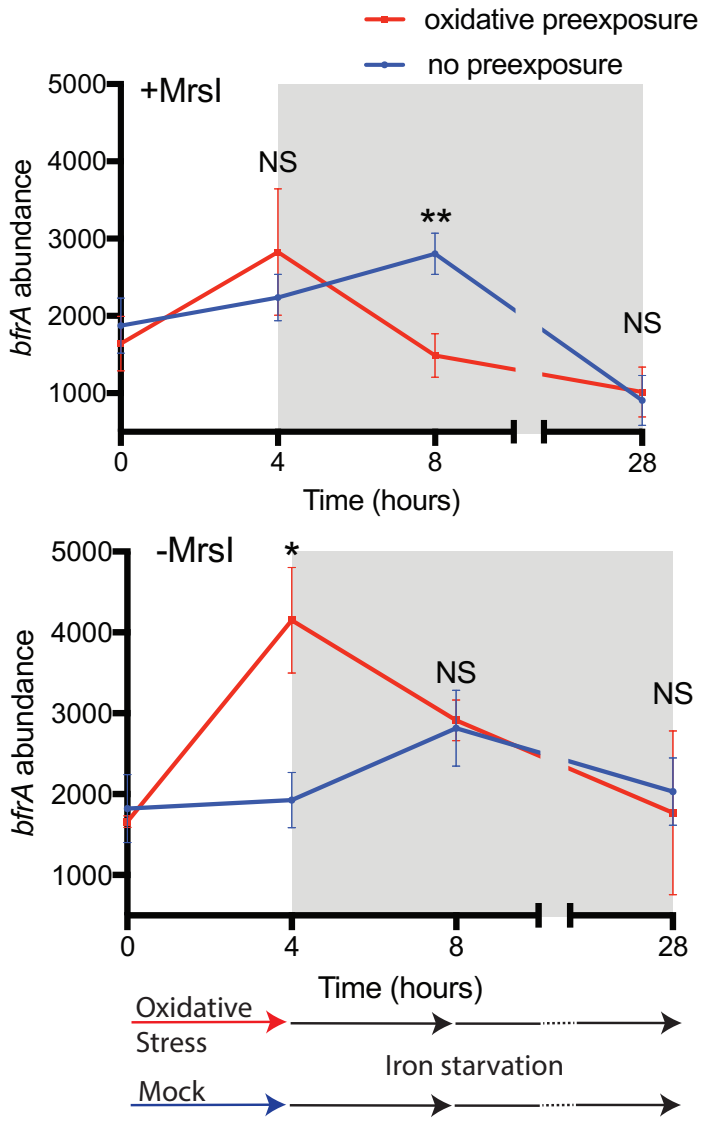


Figure 3.5. MrsI mediates an anticipatory iron sparing response in *M. tuberculosis.* Levels of the MrsI target *bfrA* with and without pre-exposure to oxidative stress in *M. tuberculosis*. Three biological replicates of the MrsI knockdown strain were grown with or without oxidative stress for 4 hours prior to iron starvation for 24 hours (gray region). Levels of *bfrA* were measured using Nanostring. Samples with and without MrsI knockdown (bottom and top, respectively) are shown. *p<0.05, **p<0.005 (Unpaired T-test). Error bars represent the SD of 3 biological replicates.

Discussion

We validated the first direct interaction between a mycobacterial sRNA and an mRNA target, which provides important insights into target recognition by mycobacterial sRNAs (Fig. 3.3C). Data from MrsI suggest that the rules for target binding do not precisely mirror those that have been defined in other prokaryotes. Indeed, the established bioinformatic sRNA target prediction tools TargetRNA2 and CopraRNA did not predict that MrsI would directly regulate any of the experimentally identified targets in *M. smegmatis*. Running TargetRNA2 individually on each of these experimentally identified MrsI targets revealed that only 2 have favorable binding energies (Appendix Fig. A2.5). The predicted binding energies for the other MrsI-target pairs are substantially higher (less favorable) than those of the enterobacterial interactions on which the software is trained, and thus they are omitted from the list of predicted direct targets (Appendix Fig. A2.5). Our results thus suggest a space for lineage-specific tuning of sRNA target binding parameters. It is also interesting to note that sRNAs in enterobacteria require RNA chaperone proteins such as Hfq and ProQ to mediate mRNA regulation, despite their comparatively high mRNA affinity (36–38). Although mycobacteria lack a known Hfq or ProQ homologue, the less favorable binding energy of mycobacterial sRNAs suggests that noncanonical RNA chaperones exist in this lineage. Interestingly, studies in *Bacillus subtilis* have revealed a set of three noncanonical sRNA chaperones involved in the iron-sparing response (39, 40). Further studies will be necessary to identify the sRNA chaperone(s) that mediate MrsI regulation of its mRNA targets and if the binding energies of MrsI-mRNA interactions are representative of all mycobacterial sRNAs.

Although MrsI becomes highly induced in three stresses in *M. tuberculosis*, its effects on the transcriptome are more extensive during iron starvation (Dataset S4, Fig. 3.4A). Here, we show that pre-exposure of *M. tuberculosis* to oxidative stress results in more rapid MrsI-mediated repression of *bfrA* during iron starvation (Fig. 3.5). This suggests that *M. tuberculosis* may take advantage of the predictable pattern of its stresses by using oxidative and membrane stresses as warning signals that it has entered a macrophage and will soon become deprived of iron. Therefore, sensing oxidative stress and membrane stress cause *M. tuberculosis* to enter an anticipatory iron-sparing state, priming MrsI to repress translation of targets such as *bfrA*. We cannot rule out the possibility that MrsI-mediated repression of these nonessential iron containing proteins could be adaptive in oxidative stress. However, studies in other bacteria suggest that this response alone could be detrimental in oxidative stress where iron sequestration by iron storage proteins has been shown to prevent Fenton reactions from occurring.

Anticipatory responses to predictable sequences of environmental conditions have become increasingly recognized and highlight the intricate adaptation mechanisms bacteria have acquired during co-evolution with hosts. For example, an anticipatory metabolic response was recently described for *M. tuberculosis*, in which exposure to hypoxia induces a metabolic remodeling of cell surface glycolipids and prepares the cell to re-initiate peptidoglycan biosynthesis upon exit from hypoxia (41). Additionally, predictive regulation has been described in *E. coli*, which resides in the human gut (42, 43). *E. coli* encounters lactose during gut transit at an earlier point than maltose, and exposure to lactose causes pre-induction of a set of

maltose utilization operons, preemptively preparing the cells to use maltose as a carbon source (42). Further studies will be necessary to determine the prevalence and functional importance of this anticipatory regulation for pathogenesis of *M. tuberculosis* and other human-associated bacterial species.

Materials and Methods

Bacterial strains and growth conditions

All *M. tuberculosis* strains used are derivatives of H37Rv and all *M. smegmatis* strains are derivatives of mc^2155. Strains are listed in Table Appendix Table A2.2. Bacterial plasmids used in this study are listed in Appendix Table A2.3 and details of their construction are in SI materials and methods. Oligonucleotides used are in Appendix Table A2.4. *M. tuberculosis* and *M. smegmatis* strains were recovered from frozen stocks in Middlebrook 7H9 medium at 37˚. Kanamycin (20μg/mL), hygromycin B (50μg/mL), and anhydrotetrayclcine (ATc, 100-200ng/mL) were added when appropriate. For iron starvation experiments, cells were grown in minimal medium with or without 50μM $FeCl_3$, as described previously (44). The iron starvation medium, which has no added iron, was additionally treated with Chelex-100 to remove residual iron. For oxidative stress and membrane stress cells were grown in 7H9 supplemented with *tert*-Butyl hydroperoxide (tBHP) or sodium dodecyl sulfate (SDS), respectively. For nutrient starvation cells were grown in PBS with 0.05% v/v tyloxapol. Details of growth conditions for sRNA-Seq and transcriptomics can be found in SI Materials and Methods.

RNA Extraction, RT-qPCR, sRNA-Sequencing, total RNA-Sequencing

RNA extraction and RT-qPCR were performed as described previously (35). Details of treatment of cells for transcriptomics can be found in SI Materials and Methods. RNA for total RNA-Sequencing was prepared from two biological replicates using the KAPA RNA Hyperprep kit (KAPA Biosystems) following manufacturer instructions. For both sRNA-Sequencing and total RNA-Sequencing in *M. tuberculosis*, oxidative stress was induced with 1mM tBHP and membrane stress was induced with 0.5% SDS. Details of RNA-Seq data analysis can be found in SI Materials and Methods.

Luciferase Assays

M. smegmatis cells were grown to mid-log phase (OD_{600}=0.2-0.8) before exposure to 0.06mM tBHP (oxidative stress) or 0.02% SDS (membrane stress) for 2 hours, or growth in medium without iron for 24 hours (iron starvation), and luciferase production was measured using the RenillaGlo Luciferase Assay kit (Promega) following manufacturer instructions.

Proteomics

Three replicates of Δ*mrsI,* complemented, and wild type *M. smegmatis* were expanded to mid-log phase in minimal medium supplemented with iron and hygromycin B. Cells were washed once with iron starvation medium and grown in the same medium with hygromycin B and ATc for 10 hours. Proteomics was performed using quantitative LC-MS/MS. Details of proteomic sample preparation and analysis can be found in SI Materials and Methods. Gene Set Enrichment Analysis (30) was performed using the Wilcoxon test. The gene set used was the genes identified by transcriptomics as regulated by MrsI in *M. smegmatis* (Table A1).

Anticipatory response experiments

Three biological replicates of *M. tuberculosis* were grown in 7H9 medium, with or without 24 hours of ATc-mediated induction of CRISPRi, to mid-log phase before transfer to 7H9 medium with or without 1mM tBHP four hours. Samples with CRISPRi induced were supplemented with ATc. Cells were pelleted, washed once with iron starvation medium, then transferred to iron starvation medium with or without ATc for 24 hours. RNA was harvested before transfer into each medium, and after 4 and 24 hours of iron starvation for gene expression analysis by Nanostring (Nanostring Technologies). Details of Nanostring methodology and analysis can be found in SI Materials and Methods.

Acknowledgements

We thank Eric Rubin and all members of the Rubin laboratory for insightful discussions and suggestions. We thank G. Marcela Rodriguez for advice with iron starvation procedures. This work was supported by the National Institutes of Health [U19 AI107774]. E.R.G. is supported by a fellowship from the National Science Foundation (DGE1144152). B.T. is supported by the grant R15 GM102755. T.B. was supported by an EMBO fellowship (ALTF 617-2015).

References

1. Stallings CL, Glickman MS (2010) Is Mycobacterium tuberculosis stressed out? A critical assessment of the genetic evidence. *Microbes Infect* 12:1091–1101.

2. Modlin Philip Liu LT, Stenger S, Tang DH, Liu PT, Modlin RL (2007) the Induction of Cathelicidin Is Dependent on Mycobacterium tuberculosis Antimicrobial Activity against Cutting Edge: Vitamin D-Mediated Human Cutting Edge: Vitamin D-Mediated Human Antimicrobial Activity against Mycobacterium tuberculosis Is Dependent on the Induction of Cathelicidin 1. *J Immunol Ref* 179(4):2060–2063.

3. Purdy GE, Niederweis M, Russell DG (2009) Decreased outer membrane permeability protects mycobacteria from killing by ubiquitin-derived peptides. *Mol Microbiol* 73(5):844–57.

4. Schaible U, Kaufmann S (2004) Iron and microbiall infection. *Nat Rev Microbiol*.

5. Honda K, et al. (2008) Epithelium Mycobacterial Growth in Alveolar Lipocalin 2-Dependent Inhibition of Lipocalin 2-Dependent Inhibition of Mycobacterial Growth in Alveolar Epithelium. *J Immunol Ref* 181:8521–8527.

6. Li S-K, et al. (2013) Identification of small RNAs in Mycobacterium smegmatis using heterologous Hfq. *Rna* 19(1):74–84.

7. Arnvig KB, et al. (2011) Sequence-Based Analysis Uncovers an Abundance of Non-Coding RNA in the Total Transcriptome of Mycobacterium tuberculosis. *PLoS Pathog* 7(11).

8. Arnvig KB, Young DB (2009) Identification of small RNAs in Mycobacterium tuberculosis. *Mol Microbiol* 73(3):397–408.

9. Dejesus MA, et al. (2017) Comprehensive essentiality analysis of the Mycobacterium tuberculosis genome via saturating transposon mutagenesis. *MBio* 8(1). doi:10.1128/mBio.02133-16.

10. Tsai C-H, et al. (2013) Identification of Novel sRNAs in Mycobacterial Species. *PLoS One* 8(11). doi:10.1371/journal.pone.0079411.

11. Miotto P, et al. (2012) Genome-Wide Discovery of Small RNAs in Mycobacterium tuberculosis. *PLoS One* 7(12). doi:10.1371/journal.pone.0051950.

12. Pelly S, Bishai WR, Lamichhane G (2012) A screen for non-coding RNA in Mycobacterium tuberculosis reveals a cAMP-responsive RNA that is expressed during infection. *Gene* 500(1):85–92.

13. Solans L, et al. (2014) The PhoP-Dependent ncRNA Mcr7 Modulates the TAT Secretion System in Mycobacterium tuberculosis. *PLoS Pathog* 10(5).

doi:10.1371/journal.ppat.1004183.

14. Haning K, Cho SH, Contreras LM, Lathem WW, Bai G (2014) Small RNAs in mycobacteria: an unfolding story. *Front Cell Inf Microb* 4(96).

15. Smirnov A, et al. (2016) Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc Natl Acad Sci* 113(41):11591–11596.

16. Storz G, Rg Vogel J, Wassarman KM (2011) Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Mol Cell* doi:10.1016/j.molcel.2011.08.022.

17. Gerhart E, Wagner H, Romby P (2015) Small RNAs in Bacteria and Archaea: Who They Are, What They Do, and How They Do It. *Adv Genet* 90:133–208.

18. Hebrard M, et al. (2012) sRNAs and the virulence of Salmonella enterica serovar Typhimurium. *RNA Biol*.

19. Chabelskaya S, Gaillot O, Felden B (2010) A Staphylococcus aureus Small RNA Is Required for Bacterial Virulence and Regulates the Expression of an Immune-Evasion Molecule. *PLoS Pathog* doi:10.1371/journal.ppat.1000927.

20. Toledo-Arana A, et al. (2009) The Listeria transcriptional landscape from saprophytism to virulence. *Nature* 459. doi:10.1038/nature08080.

21. Massé E, Vanderpool CK, Gottesman S (2005) Effect of RyhB Small RNA on Global Iron Use in Escherichia coli. *J Bacteriol* 187(20):6962–6971.

22. Reinhart AA, et al. (2015) The prrF-encoded small regulatory RNAs are required for iron homeostasis and virulence of Pseudomonas aeruginosa. *Infect Immun* 83(3):863–875.

23. Wang M, et al. (2016) An automated approach for global identification of sRNA-encoding regions in RNA-Seq data from Mycobacterium tuberculosis. *Acta Biochim Biophys Sin* 48(6):544–553.

24. Gómez-Lozano M, Marvig RL, Molin S, Long KS (2014) Identification of Bacterial Small RNAs by RNA Sequencing. *Methods in Molecular Biology (Clifton, N.J.)*, pp 433–456.

25. Shell SS, et al. (2015) Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape. *PLOS Genet* 11(11):e1005641.

26. Czyz A, Mooney RA, Iaconi A, Landick R (2014) Mycobacterial RNA Polymerase Requires a U-Tract at Intrinsic Terminators and Is Aided by NusG at Suboptimal Terminators. *mBio* 5(2).

27. Prakash P, Yellaboina S, Ranjan A, Hasnain SE (2005) Computational prediction and experimental verification of novel IdeR binding sites in the upstream sequences of

Mycobacterium tuberculosis open reading frames. *Bioinforma Discov NOTE* 21(1010):2161–2166.

28. Jacques JF, et al. (2006) RyhB small RNA modulates the free intracellular iron pool and is essential for normal growth during iron limitation in Escherichia coli. *Mol Microbiol* 62(4):1181–1190.

29. Oglesby-Sherrouse AG, Murphy ER (2013) Iron-responsive bacterial small RNAs: variations on a theme. *Metallomics* 5(4):276.

30. Subramanian A, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102(43): 15545-15550.

31. Kery MB, Feldman M, Livny J, Tjaden B (2014) TargetRNA2: identifying targets of small regulatory RNAs in bacteria. *Nucleic Acids Res* 42(Web Server issue):W124-9.

32. Wright PR, et al. (2013) Comparative genomics boosts target prediction for bacterial small RNAs. *Proc Natl Acad Sci USA* 110(37): E3487-E3496.

33. Richter A, Backofen R (2012) Accessibility and conservation: General features of bacterial small RNA–mRNA interactions? *RNA Biol*. doi:10.4161/rna.20294.

34. Papenfort K, Podkaminski D, Hinton JCD, Vogel J (2012) The ancestral SgrS RNA discriminates horizontally acquired Salmonella mRNAs through a single G-U wobble pair. *Proc Natl Acad Sci U S A* 109(13):E757-64.

35. Rock JM, et al. (2017) Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nat Microbiol*

36. Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. *Nat Rev Microbiol* 9(8):578-589.

37. Papenfort K, Förstner KU, Cong J-P, Sharma CM, Bassler BL (2015) Differential RNA-seq of *Vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation. *Proc Natl Acad Sci* 112(7):E766–E775.

38. Feng L, et al. (2015) A Qrr Noncoding RNA Deploys Four Different Regulatory Mechanisms to Optimize Quorum-Sensing Dynamics. *Cell* 160(2):228-240.

39. Smaldone GT, Antelmann H, Gaballa A, Helmann JD (2012) The FsrA sRNA and FbpB protein mediate the iron-dependent induction of the Bacillus subtilis lutABC iron-sulfur-containing oxidases. *J Bacteriol* 194(10):2586–2593.

40. Gaballa A, et al. (2008) The *Bacillus subtilis* iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. *Proc Natl Acad Sci U S A* 105(33):11927–32.

41. Eoh H, et al. (2017) Metabolic anticipation in Mycobacterium tuberculosis. *Nat Microbiol* 2(May):17084.

42. Mitchell A, et al. (2009) Adaptive prediction of environmental changes by microorganisms. *Nature* 460. doi:10.1038/nature08112.

43. Tagkopoulos I, Liu Y-C, Tavazoie S (2008) Predictive Behavior Within Microbial Genetic Networks. *Science (80-)* 320(5881):1313–1317.

44. Kurthkoti K, et al. (2017) The Capacity of *Mycobacterium tuberculosis* To Survive Iron Starvation Might Enable It To Persist in Iron-Deprived Microenvironments of Human Granulomas. *MBio* 8(4):e01092-17.

Chapter 4

Bioinformatic, biochemical, and genetic screens towards the identification of a mycobacterial small RNA accessory factor

Elias R. Gerrick[a], Ryan Hughes[c], Raylin Xu[a], Josie Francois[a], Vincent Lin[a], James C. Sacchettini[c], Sarah M. Fortune[a,b]

[a]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA; [b]The Ragon Institute of MGH, Harvard and MIT, Cambridge, Massachusetts 02139, USA; [c]Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas, United States of America

Author contributions: E.R.G. performed all experiments, and R.X., J.F., and V.L. assisted with experiments. E.R.G., R.H., J.C.S., and S.M.F. designed experiments. This chapter comprises unpublished work.

Abstract

Small RNAs are regulatory RNAs in prokaryotes that, with the help of protein accessory factors, play a major role in post-transcriptional regulation of gene expression. The most well-understood accessory factor by far is the protein Hfq, which has been identified in approximately half of all bacterial species and is essential for the action of most sRNAs that have been characterized to date. Recently additional sRNA accessory factor proteins have been discovered, such as ProQ in enteric bacteria and FbpA-C in *Bacillus subtilis*. The important global health pathogen *M. tuberculosis*, however, has no obvious homologues of any of these important regulatory proteins, and no other sRNA accessory factor has been identified in mycobacteria. Here, we perform bioinformatic, biochemical, and genetic screens in order to identify an accessory factor involved in mycobacterial sRNA-mediated riboregulation. However, while we do identify a putative structural homologue of Hfq in mycobacteria that appears to have evolved an orthologues function, we do not identify any candidate sRNA accessory factors. However, the results of our suppressor screen provide useful information regarding likely characteristics of any accessory factor for the sRNA-mediated iron sparing response in *M. smegmatis*.

Introduction

Small RNAs (sRNAs) alone are most frequently not sufficient for regulation of their target mRNAs in bacteria such as *Escherichia coli* and *Salmonella*, and require an RNA chaperone protein, or accessory factor, to mediate the sRNA-target interaction (1). Indeed, the majority of sRNAs characterized to date rely on the chaperone protein Hfq for both their stability and

target-binding. Hfq is a small protein of approximately 70-90 amino acids with a highly conserved secondary structure of $\alpha_1\beta_5$, and forms homohexameric toroid quaternary structures. This ring structure of Hfq is responsible for giving the protein two RNA-binding faces, which have different binding properties (2–5). In a canonical sRNA-target interaction, Hfq binds the sRNA on one face and the mRNA target on another, allowing the two RNA species to interact along the rim of the Hfq hexamer. Hfq most frequently recognizes sRNAs through binding to the poly-uridine track of their rho-independent terminators, which sRNAs ubiquitously use to terminate transcription (6–8). Hfq has additionally been found to interact with the dominant RNA degradation enzyme in *E. coli*, RNaseE and this interaction facilitates target mRNA degradation following sRNA binding (9–11).

However, despite the importance of Hfq for sRNA-target interactions in bacteria such as *E. coli* and *Salmonella*, another core accessory factor, ProQ, has recently been discovered. ProQ is present in α–, β–, and γ-proteobacteria and was found, like Hfq, to bind to many sRNAs in *S. enterica* (12). ProQ was subsequently shown to be required for the action of an sRNA, and therefore is hypothesized to represent a novel class of core sRNA chaperone proteins (13). Additionally, other putative sRNA chaperones have been discovered in *Bacillus subtilis*; a set of 3 small, basic proteins named FbpA-C were found to be required for the iron sparing response mediated by the sRNA FsrA (14, 15). Interestingly, *B. subtilis* and other low-GC gram positive bacteria have Hfq proteins, but Hfq has not been found to have an important role in sRNA-mediated riboregulation in these species (16–18). These results illuminate the possibility that other sRNA chaperones remain to be discovered, and could explain the lack of reliance on Hfq,

or the lack of an Hfq protein all-together, in certain bacterial lineages. Indeed, as Hfq proteins have been identified in approximately half of all bacterial species, it is possible that completely novel sRNA accessory factors exist in those lineages lacking Hfq.

Mycobacteria are one lineage for which no known sRNA accessory factor exists. We reasoned that mycobacteria may have a noncanonical chaperone protein that mediates sRNA-mRNA interactions and used bioinformatic , biochemical, and genetic approaches to screen for possible candidates. However, while one of these screens did reveal a mycobacterial structural homologue of Hfq, no sRNA accessory factor was identified.

Results

Bioinformatic screen for structural homologues of Hfq in *M. tuberculosis*

Interestingly, the cyanobacterial genera *Anabaena* and *Synechocystis* were predicted to lack Hfq proteins based on a lack of homologues in BLAST searches, but two noncanonical Hfq proteins were identified in these organisms (19). These homologues lack amino acid conservation with other Hfq proteins, but like canonical Hfq have the classical $\alpha_1\beta_5$ structural motif. Additionally, crystallography confirmed that they form homohexameric toroids and are virtually identical to other Hfq proteins structurally, despite the lack of amino acid conservation. This raised the possibility that structural homologues of Hfq that lack sequence similarity exist in other lineages. To test if this were the case in mycobacteria, we created a bioinformatic screen to search for proteins with structures similar to Hfq in the *M. tuberculosis* proteome. First, we created a training set of 19 Hfq alleles from diverse bacterial lineages, and

performed secondary structure prediction on these using the webtool NetSurfP (20). We used this Hfq secondary structure training set to train a position weight matrix (PWM), which incorporates the per-position probability of an Hfq homologue forming an α-helix, β-strand, or coil as its secondary structure. We then performed secondary structure prediction on every *M. tuberculosis* protein that is within the size range of being an Hfq protein (defined here as 120 amino acids or fewer). This resulted in a list of 522 *M. tuberculosis* proteins and their resulting secondary structure predictions. Next, we scanned this set of *M. tuberculosis* protein secondary structure predictions with the PWM. Each protein was scored based on the percent similarity of its secondary structure prediction to known Hfq secondary structure predictions, with the highest scoring window of each protein being used to calculate the percent similarity (Fig. 4.1).

The bioinformatic screen revealed one protein, Rv3208A, a gene of unknown function in *M. tuberculosis* that is highly conserved across mycobacterial species, with a high score of 83.2% structural similarity to known Hfq proteins. Closer examination revealed that Rv3208A is predicted to have 2 β-strands preceding the α-helix at the N-terminus, and 4 β-strands at the C-terminus instead of 5 (Fig. 4.1B). However, aside from these differences the predicted secondary structures are highly similar. These promising results prompted us to use the Phyre2 tool to model Rv3208A onto existing protein tertiary structures in the Protein Data Bank (21). Importantly, one of the highest scoring templates used by Phyre2 to model Rv3208A was the structural homologue of Hfq identified in the cyanobacterium *Synechocystis* sp. PCC 6803. Additionally these promising bioinformatic results prompted us to study the protein further in *M. smegmatis* and rename the protein MhfQ (mycobacterial Hfq).

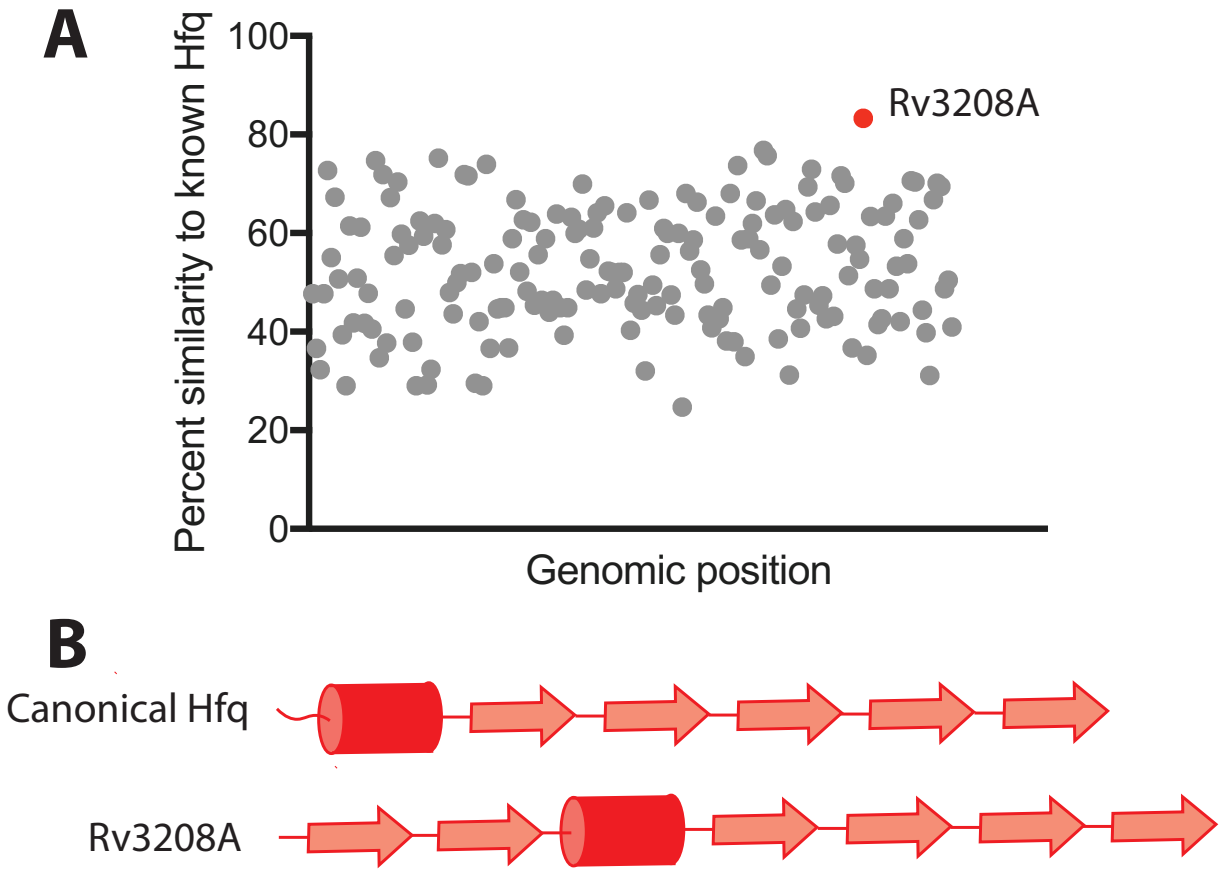


Figure 4.1. Bioinformatic screen for structural homologues of Hfq in *M. tuberculosis* identifies Rv3208A. (A) The percent-similarity of each *M. tuberculosis* small protein's secondary structure to that of known Hfq alleles. The highest scoring region for each protein is shown. The highest-scoring protein (Rv3208A) is in red. (B) Representation of the secondary structure prediction of canonical Hfq proteins and Rv3208A. Alpha helices are represented by cylinders, beta strands are represented by arrows.

MhfQ in *M. smegmatis* appears to be involved in stress adaptation but mediates minimal effects on the transcriptome

To investigate the function of MhfQ in mycobacteria, we constructed a deletion mutant (Δ*mhfQ*) in *M. smegmatis*. Deletion of *hfq* in many bacteria such as *E. coli* results in pleiotropic growth defects and expansive changes to the transcriptome, whereas deletion of *hfq* in other

organisms results in growth defects in only some or no growth conditions (1, 17, 22–24). We found that deletion of *hfq* in *M. smegmatis* resulted in no growth defect during logarithmic phase growth, osmotic stress, acid stress, or nutrient starvation (data not shown). However, the Δ*mhfQ* strain had a mild growth defect during ethanol stress and a more severe growth defect during exposure to the envelope-stress inducing antibiotic polymyxin B (Fig. 4.2A, B). Importantly, the structural homologue of Hfq from the cyanobacterial species *Anabaena* PCC 7120 complemented the growth defect observed during polymyxin B stress, consistent with MhfQ being a homologue of Hfq (Fig. 4.2B). Interestingly, the *E. coli* Hfq allele did not complement the polymyxin B growth defect of the Δ*mhfQ* strain (data not shown) suggesting that these noncanonical Hfq alleles may have altered RNA binding properties or functions.

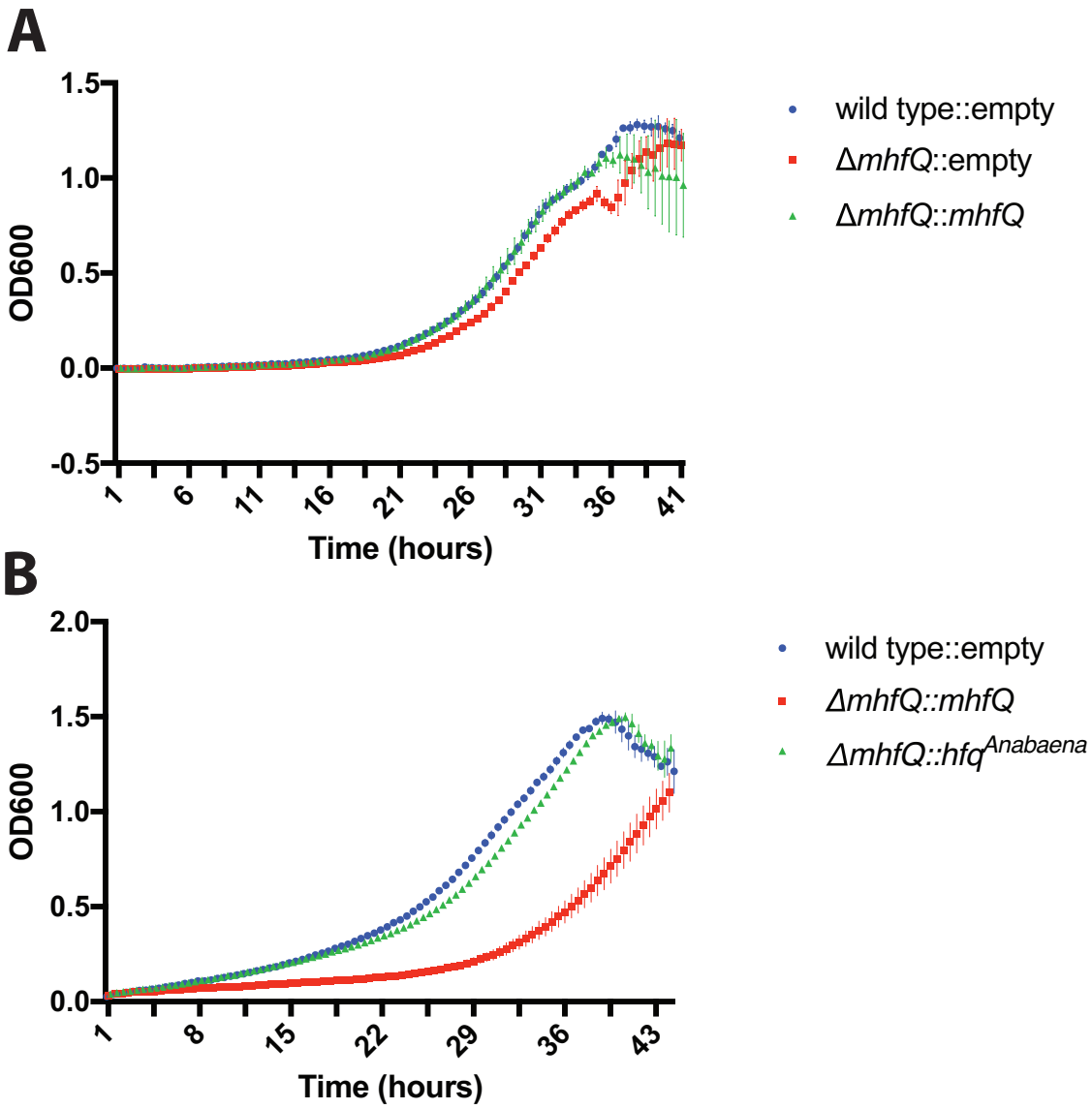


Figure 4.2. Deletion of *mhfQ* in *M. smegmatis* results in stress-specific growth defects that are complemented by *Anabaena* Hfq. (A) Growth of the wild type, Δ*mhfQ*,and complemented *M. smegmatis* strains in ethanol stress. (B) Growth of the wild type, Δ*mhfQ*, and Δ*mhfQ::hfq*^{Anabaena} strains of *M. smegmatis* in polymyxin B stress.

Given the results above, we performed transcriptomics on wild type and Δ*mhfQ* strains of *M. smegmatis* during exposure to polymyxin B stress. However, the effects of MhfQ on the transcriptome appeared to be minimal (Fig. 4.3). Only 4 genes, not including *mhfQ* itself, had greater than 2-fold changes in gene expression between the wild type and Δ*mhfQ* strains.

Additionally, all 4 genes are genes of unknown function. Deletion of Hfq in other bacteria results in much more substantial transcriptomic changes due to the central role that Hfq plays in post transcriptional regulation of gene expression in these organisms, with the expression of hundreds of genes being altered (25). In fact, the effects of the *mhfQ* deletion in *M. smegmatis* during polymyxin B exposure were even less extensive than those observed for deletion of *hfq* in *B. subtilis* during logarithmic growth, an organism in which Hfq has not been shown to play a major role in sRNA-mediated riboregulation (16, 17). Thus, we reasoned that MhfQ may not function as an sRNA chaperone in mycobacteria.

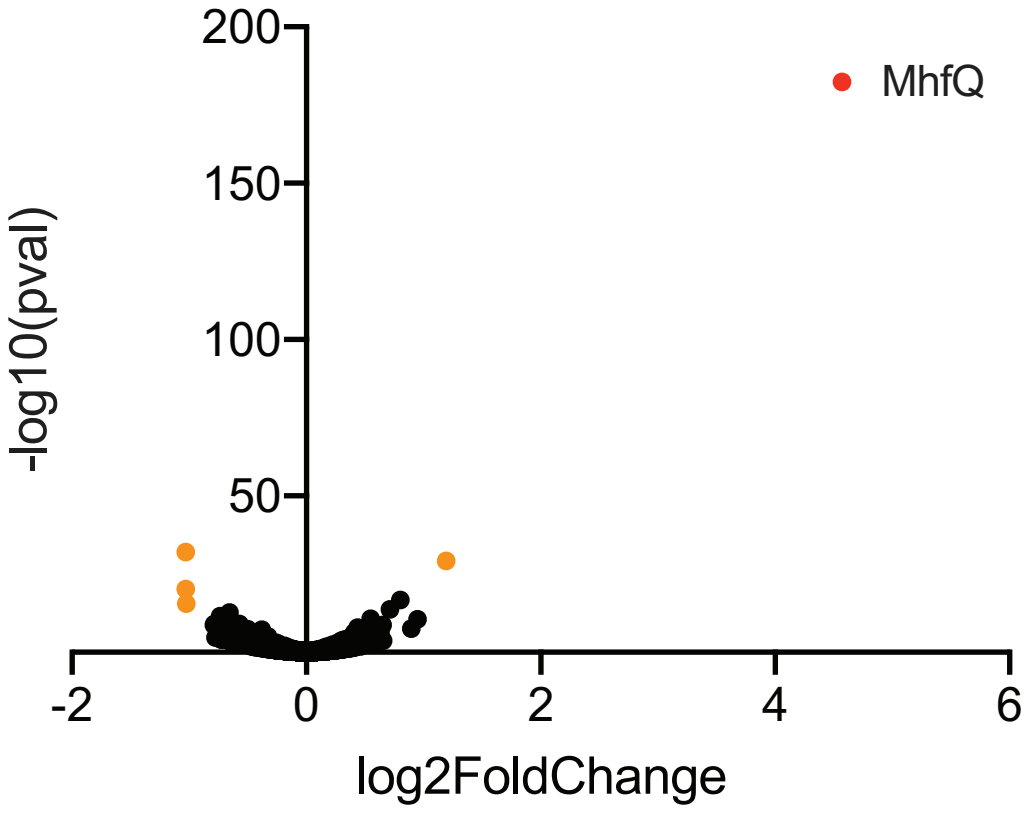


Figure 4.3. Trancriptomic effects of loss of *mhfQ* during exposure to polymyxin B stress in *M. smegmatis*. *M. smegmatis* strains were exposed to polymyxin B prior to harvesting RNA for RNA-sequencing. Red data point signifies MhfQ, orange data points signify significantly differential expressed genes ($p<0.05$, fold change > 2.0).

MhfQ in *M. smegmatis* interacts with proteins involved in fatty acid metabolism

Due to the minor effects of MhfQ on the transcriptome, we sought to determine which proteins MhfQ interacts with in *M. smegmatis*. We reasoned that if MhfQ acts as a functional homologue of Hfq it should interact with RNA metabolism proteins, similar to Hfq in *E. coli*. To test this, we performed Co-immunoprecipitation of MhfQ after crosslinking with formaldehyde. The samples were then sent for Mass spectrometry analysis for protein identification (Table 4.1). Interestingly, the proteins we found to be associated with MhfQ were predominately involved in fatty acid metabolism, such as enoyl CoA hydratases and Acyl-CoA hydratase. Additionally, no RNA-binding proteins were identified as associated with MhfQ in the assay. Therefore, we concluded that while MhfQ appears to be a structural homologue of Hfq, the sRNA-chaperone function is likely not conserved in mycobacteria.

Table 4.1. Protein interacting partners of MhfQ in *M. smegmatis*

Protein	Function
ClpP (both)	Protease
RibAB (RibA2)	Riboflavin biosynthesis pathway
MSMEG_4709	Enoyl CoA Hydratase
EchA8	Enoyl CoA Hydratase
MSMEG_4715 (FadE19)	Acyl-CoA dehydrogenase
MSMEG_0464	Phosphomethylpyrimidine kinase

sRNA-protein pulldowns towards the discovery of a mycobacterial sRNA accessory factor

Due to the apparent lack of sRNA-chaperone activity of MhfQ, we reasoned that mycobacteria may have a novel class of accessory factor that lacks similarity to any other known sRNA chaperone proteins. To test this, we sought to perform an agnostic screen for sRNA-binding proteins in *M. smegmatis* by using an sRNA-protein pulldown. (12, 26, 27). We focused on the sRNA MrsI (characterized in Chapter 3), as this is the most highly characterized sRNA in mycobacteria. We first inserted the streptavidin-binding S1 aptamer into MrsI (28), and found that incorporation of a short structurally-insulating stem flanked by flexible polyA stretches did not disrupt the predicted secondary structure (Fig. 4.4A). Additionally, this tagged allele of MrsI complemented the growth defect of the deletion mutant during iron starvation (Fig. 4.4B). We then performed streptavidin-mediated pulldowns of the S1 aptamer-tagged MrsI followed by mass spectrometry analysis. However, mass spectrometry revealed no protein that passed thresholding criteria in the eluate of the S1-tagged samples that was not found in untagged MrsI control samples.

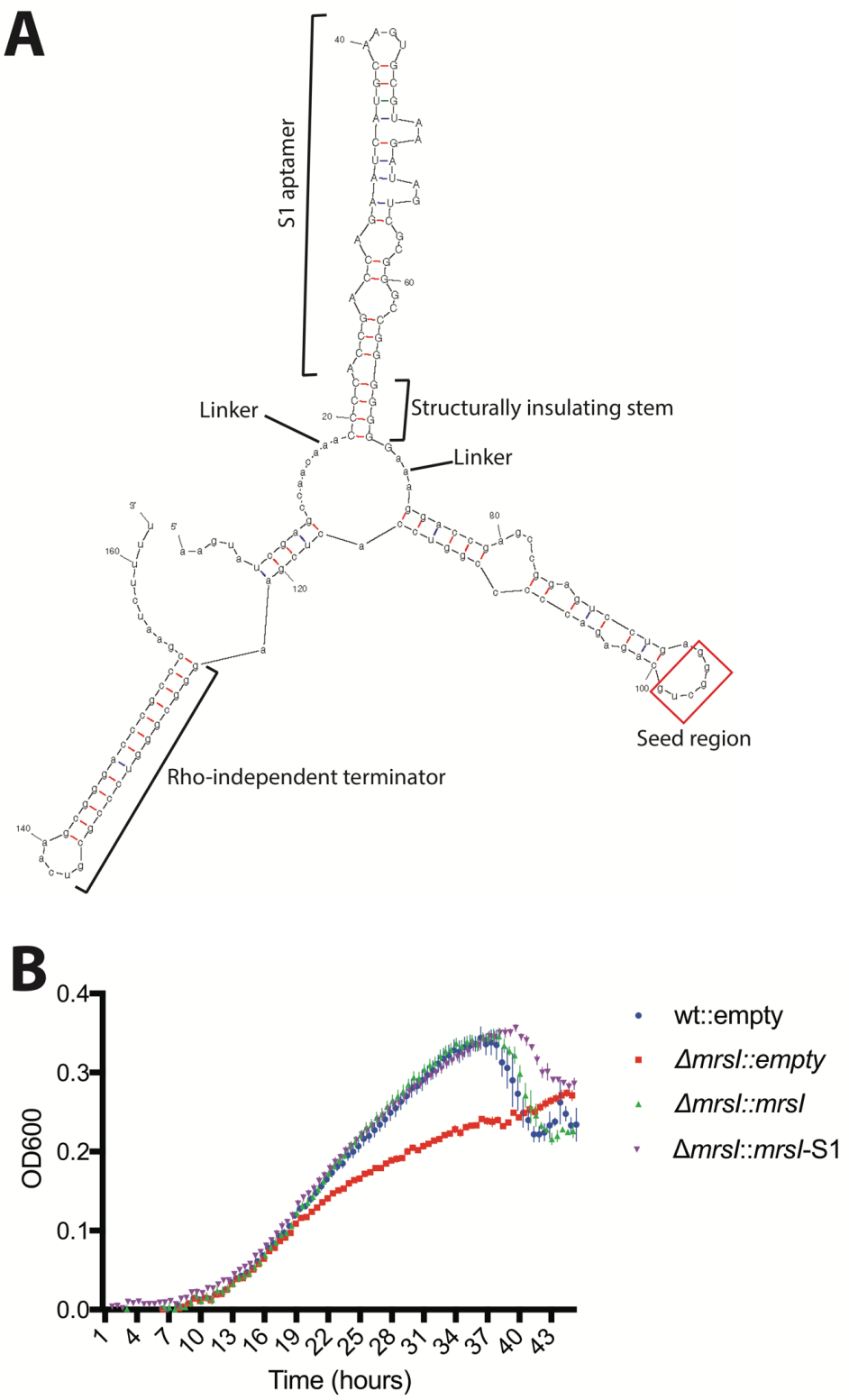


Figure 4.4. Aptamer tagging of MrsI preserves sRNA structure and function. (A) Secondary structure prediction of the S1-tagged allele of MrsI (mfold). (B) Growth curves of *M. smegmatis mrsI* strains during iron starvation

Implementation of a suppressor screen toward the discovery of a mycobacterial sRNA accessory factor

We next sought to identify the mycobacterial sRNA accessory factor using a genetic approach, for which we used a suppressor screen. In order to implement a suppressor screen, we first made a strain of *M. smegmatis* in which MrsI is conditionally lethal by fusing the promoter and 5' UTR of one MrsI target, *bfrA*, to the resistance gene for the antibiotic zeocin (*zeoR)*. Zeocin was chosen due the mechanism of resistance mediated by the ZeoR protein; while many antibiotic resistance proteins are enzymatic, ZeoR inactivates zeocin by binding the antibiotic in a one-to-one ratio. This makes zeocin an optimal antibiotic for an sRNA suppressor screen, as modest levels of repression by the sRNA should result in increased susceptibility in a linear fashion. We confirmed that MrsI, expressed from a strong promoter on an integrating plasmid, mediates zeocin susceptibility in this strain by monitoring growth with and without the sRNA (Fig. 4.5). Next, we plated the cells on agar containing zeocin to obtain suppressor mutants. To confirm zeocin resistance, we patched the resulting colonies onto fresh zeocin plates. We next performed Sanger Sequencing of the *mrsI* gene for each suppressor mutant and eliminated colonies that had obtained resistance through mutation of MrsI. Next, to eliminate mutants that became resistant to zeocin in a MrsI-independent manner, we performed growth curves of each suppressor during iron starvation. Any suppressor that had no growth defect during iron starvation was eliminated as a candidate accessory factor mutant, as an accessory factor mutant would be expected to have a growth defect similar to that of the *mrsI* deletion strain during iron starvation.

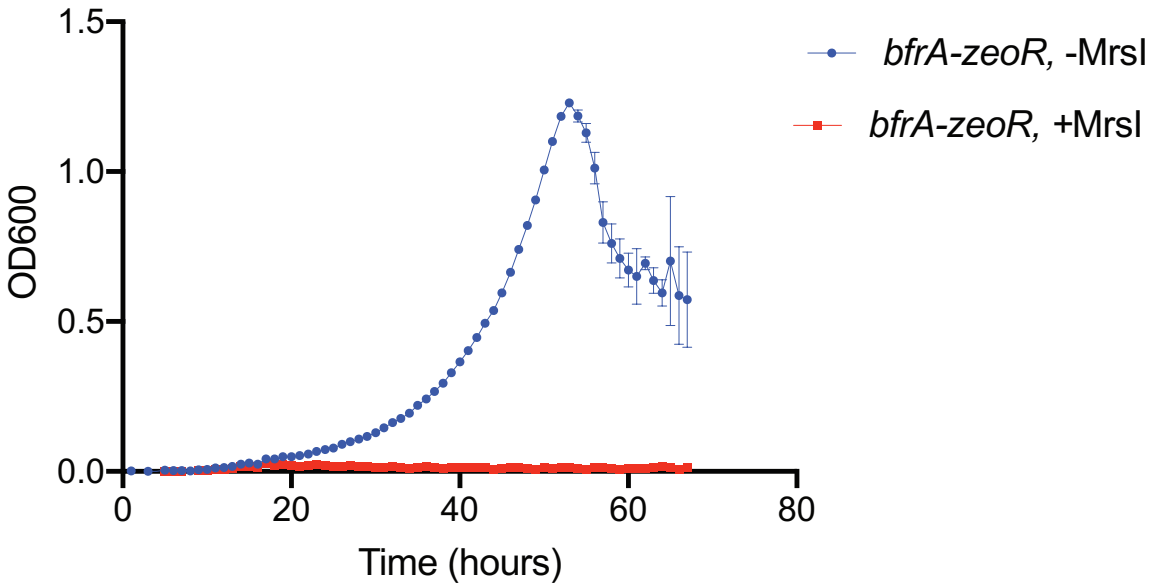


Figure 4.5. MrsI causes conditional susceptibility to zeocin in *M. smegmatis* containing *bfrA-zeoR*. Growth curve of *M. smegmatis* strains with the promoter and 5' UTR of *bfrA* fused to the *zeoR* gene, with and without MrsI. Cells were grown in the presence of zeocin. Error bars represent the SD of 3 replicates.

This resulted in 6 suppressors with growth defects during iron starvation and intact *mrsI* genes (Figure 4.6). Of these, one suppressor phenocopied the Δ*mrsI* strain of *M. smegmatis* (Fig. 4.6B, suppressor 4.1), while the others exhibited phenotypes that were intermediate between the deletion and wild type strains (Figure 4.6).

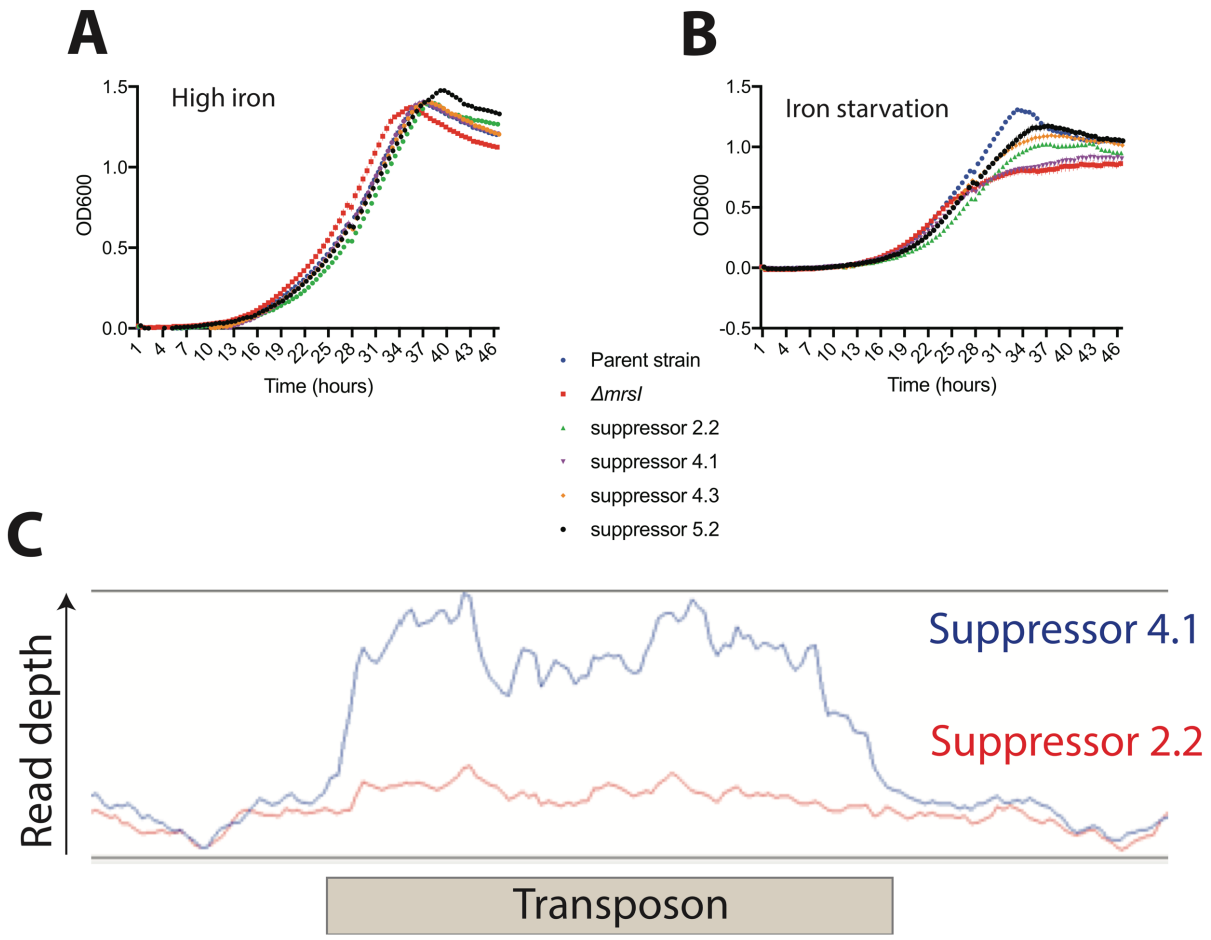


Figure 4.6. Phenotypes of a subset of suppressors during growth in iron replete and iron starvation conditions. (A) Growth of the suppressors, the parent strain, and Δ*mrsI* strain in high iron medium. (B) Growth of the suppressors, the parent strain, and Δ*mrsI* strain in iron starvation medium. For clarity, a representative 3 suppressor strains with intermediate phenotypes are shown (C) Visualization of the transposon gene duplication in suppressor 4.1, compared to suppressor 2.2 as a reference. Read depth is shown on the Y axis. The sequencing trace for suppressor 4.1 is shown in Blue, and the trace for suppressor 2.2, which does not have duplication of transposons, is shown in red.

We next sought to identify the causative mutations associated with the suppressors by performing whole-genome sequencing on the 6 suppressor strains, followed by variant calling. We searched for convergent mutations within the strains, but found that only one gene, the alpha subunit of RNA polymerase (*rpoA*) had convergent single nucleotide polymorphisms (SNPs). These convergent mutations in RpoA were present in 3 of the suppressors with

intermediate phenotypes during iron starvation. Interestingly, while one of the mutations was a silent mutation and two were missense mutations, all three were in the beginning of the C-terminal domain of RpoA. However, RpoA was deemed an unlikely candidate for an sRNA accessory factor, and these mutations likely suppressed zeocin resistance in an alternative manner.

Interestingly, while we did not identify SNPs that appeared to cause the observed phenotypes, we observed duplications of naturally occurring transposon genes in the suppressor strain that perfectly phenocopied the *mrsI* deletion strain (suppressor 4.1) (Fig. 4.6C, blue). This suggests that in suppressor 4.1 transposons had duplicated themselves and possibly inserted into other sites of the genome, perhaps in response to the zeocin-mediated stress. We therefore considered the possibility that the phenotype causing the zeocin resistance and growth defect during iron starvation in suppressor 4.1 was due to a novel transposon-mediated disruption of the accessory factor gene, and not a SNP. To test this hypothesis, we mined the sequencing data for novel transposon-genome junctions. However, we were unable to identify any novel transposon insertion sites in the genome by this method.

Application of proximity-dependent biotinylation toward the discovery of a mycobacterial sRNA accessory factor

We next sought to identify the accessory factor of MrsI by using proximity-dependent biotinylation. This technique uses a mutant form of the BirA biotin ligase protein from *E. coli* (BirA*) which produces reactive biotin species (29, 30). These reactive biotin molecules diffuse

away from the BirA* protein to selectively biotinylate proteins in close proximity to BirA*. By fusing BirA* to a protein of interest and then purifying biotinylated proteins, interactors can be identified. To adapt this system for the study of sRNA-protein interactions, we inserted the λN22 RNA aptamer into MrsI (31). We then fused the λN22-aptamer binding protein (λABP) to BirA*. Thus, the λN22-λABP interaction should bring MrsI and BirA* together, along with any MrsI-associated proteins. We first verified that the insertion of the λN22 aptamer into MrsI did not disrupt the function of the sRNA by testing for complementation of the growth defect during iron starvation by the aptamer-tagged MrsI allele (Fig. 4.7A). We then grew the MrsI-λN22; λABP-BirA* containing strain in iron starvation medium supplemented with excess levels of biotin before harvesting protein for further analysis. By probing the samples with streptavidin linked to a fluorescent marker we detected increased levels of protein biotinylation in samples with the BirA* fusion than without (Fig. 4.7B). However, we observed no differential biotinylation signal between a control strain expressing untagged MrsI and the strain expressing MrsI-λN22. To test the possibility that differences existed between the samples with untagged versus tagged MrsI that were not detected by the blot, due to sensitivity or high levels of background, we purified the biotinylated proteins using monomeric avidin-coated beads, then performed mass spectrometry analysis on the resulting eluates. However, similar to the MrsI-S1 aptamer pulldowns, we detected no proteins that passed our thresholding criteria in the tagged samples that were not found in the untagged controls.

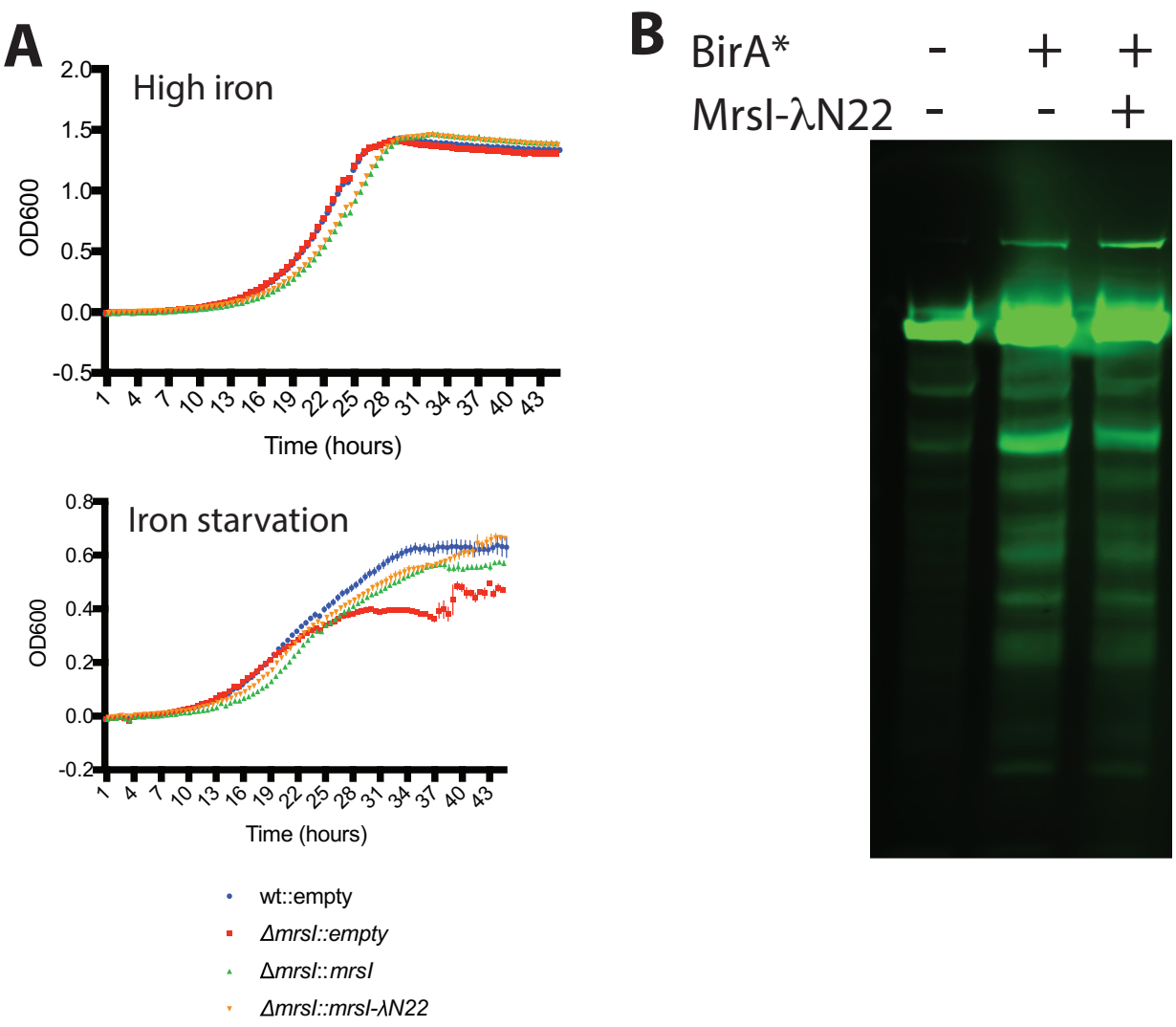


Figure 4.7. Aptamer tagging of MrsI for proximity-dependent biotinylation of MrsI-associated proteins. (A) Growth curves of *M. smegmatis* strains with or without MrsI, and with a λN22-aptamer-tagged allele of MrsI. (B) Western Blot using a streptavidin probe to detect biotinylated proteins in samples with and without BirA*-λABP and λN22-tagged MrsI.

Discussion

In this work, we provide evidence that mycobacteria, which were thought to not encode a homologue of the sRNA chaperone protein Hfq, possess a noncanonical structural homologue of Hfq. MhfQ has substantial structural similarities to Hfq proteins from other bacterial species but lacks sequence identity with other Hfq alleles. This is similar to what was observed for the

Hfq alleles identified in the cyanobacterial genera *Anabaena* and *Synechocystis*, which likewise are structural homologues of canonical Hfq without sequence similarity (19). However, MhfQ is unique in that it appears to have an N-terminal β_2 motif which Phyre2 analysis predicts to be a β-grasp fold. β-grasp folds are a common and highly versatile family of protein domains involved in diverse functions including protein-binding, RNA binding, cofactor binding, and post-translational protein modification (32). Thus, the presence of the β-grasp fold on the N-terminus of MhfQ could provide any of a wide array of functions for the protein.

Interestingly, while MhfQ shares structural similarity to all known Hfq alleles, the closest structural homologue appears to be that of the cyanobacterium *Synechocystis* sp. strain PCC 6803 based on Phyre2 modeling. However, *Synechocystis* Hfq has not been shown to bind to RNA. In fact, the residues found to be important for RNA binding in canonical Hfq proteins are not conserved in *Synechocystis* Hfq (19). Additionally, while the surface of canonical Hfq proteins is positively charged in order to facilitate RNA binding, the surface of Hfq from Synechocystis is negatively charged (19). These results suggest that *Synechocystis* Hfq may have evolved an orthologous function. Consistent with this, an *hfq* deletion mutant in this species was found to be defective for motility and it was found that the Hfq allele binds to the type IV pilus base (33, 34). Although deletion of *hfq* from Synechocystis does cause differential expression of a small number of genes, including a subset involved in pilus assembly, the results raise the possibility that Hfq in this cyanobacterial species has lost RNA-chaperone function. This is consistent with the results presented here on MhfQ, in which MhfQ appears to have minimal effects on the transcriptome of *M. smegmatis* and binds to proteins involved in fatty

acid metabolism. It is interesting to consider, if these Hfq structural homologues indeed lack RNA chaperone activity, what the evolutionary driver was for this change in protein function. Additionally, this raises the question of what the true sRNA accessory factor is in these lineages.

It is also interesting to consider the implications of the lack of accessory factor candidates from the suppressor screen. There are two major biological reasons that could explain this outcome: essentiality and redundancy. If the MrsI accessory factor is an essential protein, a suppressor screen would be much less likely to succeed, due to the fact that mutations that disrupt the accessory factor could be lethal to the cell. However, in our suppressor screen we relied on natural mutation instead of a more disruptive mutation such as transposon mutagenesis, which theoretically should allow for mutations that disrupt the chaperone activity, but not the essential function, of a theoretical essential accessory factor. However, the limitation to the possible range of mutations that an essential target would impose would substantially decrease the effective target size, thereby decreasing the probability of identifying these mutations in the screen. The other major possibility is that the MrsI accessory factor is either duplicated or redundant. Indeed, the *M. smegmatis* strain used in these studies has large genomic duplications (35). If the accessory factor were present in a duplicated region then identification by a suppressor screen would be infeasible.

Another question raised by the results of the suppressor screen is how mutations in *rpoA* conferred resistance to zeocin. While it is possible that these mutations caused resistance by a MrsI-independent mechanism, the fact that the mutants have a growth defect during iron

starvation makes this less likely. One possible explanation could be that the mutations in *rpoA,* which encodes the α-subunit of RNA polymerase, abrogate the ability of RNA polymerase to bind to the promoter of MrsI. The α-subunit of RNA polymerase has been shown to play an important role in promoter binding for certain promoters, due to both direct binding to promoter UP elements and interactions with transcription factors (36, 37). However, in the suppressor screen presented here two different promoters drove MrsI expression. The two exogenous copies of *mrsI* inserted into the cells were both under the control of the strong mycobacterial promoter UV15, but the strain was built in a wild type background and therefore also retained its endogenous copy of *mrsI* under the control of its native promoter. The endogenous copy of *mrsI* is unlikely to have been under selective pressure during plating on zeocin due to the fact that this step of the screen was performed under iron replete conditions, so this copy of *mrsI* would not be expected to be expressed. Therefore, it is unlikely that the *rpoA* mutations conferred resistance and an iron starvation growth defect due to deficient transcription of MrsI, as this would require the mutation to impact transcription of both the constitutive and stress-inducible promoters of MrsI present in these strains, and only the former would have been under selective pressure. However, the C-terminal domain of the α-subunit of RNA polymerase, where the convergent mutations occurred, has also been shown in *E. coli* to interact with the termination factor NusA. NusA stimulates pausing of the polymerase and thereby facilitates rho-independent transcriptional termination (38). In mycobacteria the termination factor NusG was shown to assist RNA polymerase in termination of rho-independent termination (39). Because termination of MrsI transcription uses a rho-independent terminator, it is therefore possible that the *rpoA* mutations conferred MrsI-

dependent resistance to zeocin by interfering with polymerase binding to transcriptional termination factors, which would impact MrsI length, structure, and function. Interestingly, studies on the sRNA RyhB from *E. coli* have demonstrated that the bacteria inefficiently terminate transcription of the sRNA in certain growth conditions to regulate its activity (40).

The combined lack of an accessory factor candidate from the above experiments raises questions as to the not only the identity of the mycobacterial accessory factor, but also the presence of one. One possible explanation for the results of the above experiments is that mycobacterial sRNAs do not require the help of a chaperone protein. However, the two proposed factors that could obviate the need for an accessory factor in certain bacterial lineages, GC content and extended base-pairing, make this possibility seem unlikely (41). Low GC content was found to correlate with a lack of reliance on Hfq in bacteria such as the firmicutes, and the weaker RNA structures in low GC organisms was proposed to remove any need for a chaperone in sRNA-mRNA interactions. However, mycobacteria have a higher GC content than most of the organisms that rely heavily on sRNA chaperones such as *E. coli* and *Salmonella.* The second factor, extended base-pairing, also seems to not apply given that the interactions between MrsI and its targets appears to be largely contained to the seed region. Mycobacteria therefore likely have an accessory factor involved in sRNA-mediated riboregulation, and future experiments will be required to uncover its identity.

Materials and Methods

Bioinformatic screen for mycobacterial Hfq homologues

A training set for the position weight matrix of 19 Hfq alleles from different bacterial lineages was created, and secondary structure prediction was performed using the NetSurfP webtool (20). This process was repeated with the list of all proteins smaller than 120 amino acids from *M. tuberculosis* strain H37Rv. The position weight matrix was then used to scan the secondary structure prediction from each small *M. tuberculosis* protein. The score given to each protein was the highest scoring window of the protein's secondary structure. To obtain a percent identity to known Hfq for each protein, a random sample-out approach was taken, whereby 3 alleles from the training set were randomly removed from the training set and given a score by the PWM. This "true Hfq" score was then used as the reference to calculate the 'percent similarity to known Hfq' for each *M. tuberculosis* protein.

Bacterial strains and growth conditions

All bacterial strains were derivatives of *M. smegmatis* strain mc^2155. Bacteria were cultures in 7H9 medium (difco laboratories) or iron starvation medium as described in chapter 3.

MrsI-S1 pulldowns

Cells were grown to mid-log phase (optical density = 0.2-0.8) in 7H9 medium with 50µg/mL hygromycin B before pelleting, washing once with iron starvation medium, then resuspending to an optical density of 0.01 in iron starvation medium supplemented with 50µg/mL of hygromycin B and 100ng/mL of anhydrotetracycline to induce MrsI-S1 expression. Cells were expanded overnight. The next day, cells were pelleted and washed once with lysis buffer (50mM HEPES pH 7.4, 10mM $MgCl_2$, 100mM NaCl, 1mM DTT, 0.1% Triton X-100, 10% glycerol),

then pelleted and resuspended inlysis buffer with protease inhibitors. Samples were then lysed by bead-beating 5x30seconds. Lysate was cleared by centrifugation (14,000g for 20 minutes at 4˚C). Lysate was blocked using egg white avidin, then streptavidin-conjucated magnetic beads were incubated with the clarified lysate for one hour at 4˚C. Beads were then washed 5 times with 500µL lysis buffer, and S1-aptamer tagged MrsI was eluted by incubating the beads with lysis buffer supplemented with 5µM biotin for 30 minutes at 4˚C.

MrsI suppressor screen

Cells were expanded in 7H9 medium before plating on 7H10 agar supplemented with 25ng/mL zeocin. After 3-4 days colonies were patched onto fresh 7H10 agar with zeocin. Cultures were inoculated from the patches, and growth curves in high iron and iron starvation conditions were performed as described in chapter 3. Whole genome sequencing was performed using the Nextera XT DNA library prep kit (Illumina) according to manufacturer instructions. Variant calling was performed using Pilon (42).

Proximity dependent biotinylation

Cells were expanded in 7H9 medium supplemented with 50µg/mL hygromycin B before being pelleted, washed once with iron starvation medium, then resuspended in iron starvation medium supplemented with 50µg/mL hygromycin B and 50µg/mL biotin at an optical density of 0.01. Cells were grown overnight in these medium. For pulldowns of biotinylated protein, cell pellets were were washed 3 times with PBS, then resuspended in lysis buffer (50mM Tris, 150mM NaCl pH7.4) and lysed by bead beating. An equal volume of 2x Wash Buffer (50mM Tris,

150mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton-X 100, pH 7.5) was added to the lysate. Lysate was clarified by centrifuging at 14,000*g* for 15 minutes at 4˚C. Monomeric avidin beads (Bioclone) were then added to the clarified lysates and incubated for 1 hour at room temperature, with agitation. Biotinylated protein was then purified from the beads according to manufacturer instructions. Blotting for biotinylated proteins was performed as described in Chapter 3, with the exception that a fluorescently labeled streptavidin protein was used to probe the blot (Thermo Fisher).

References

1. Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. *Nat Rev Microbiol* 9(8):578–589.
2. Link TM, Valentin-Hansen P, Brennan RG (2009) Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. *Proc Natl Acad Sci* 106(46):19292–19297.
3. Schumacher MA, Pearson RF, Mùller T, Valentin-Hansen P, Brennan RG (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq– RNA complex: a bacterial Sm-like protein. *EMBO J*. 21(13):3546-3556.
4. Soper TJ, Doxzen K, Woodson SA (2011) Major role for mRNA binding and restructuring in sRNA recruitment by Hfq. *Rna* 17(8):1544–1550.
5. Schu DJ, Zhang A, Gottesman S, Storz G (2015) Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. *EMBO J* 34:2557–2573.
6. Otaka H, Ishikawa H, Morita T, Aiba H, Susan Gottesman (2011) PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proc Natl Acad Sci* 108(32):13059-13064.
7. Sauer E, Weichenrieder O (2011) Structural basis for RNA 3'-end recognition by Hfq. *Proc Natl Acad Sci* 108(32):13065–13070.
8. Ishikawa H, Otaka H, Maki K, Morita T, Aiba H (2012) The functional Hfq-binding module of bacterial sRNAs consists of a double or single hairpin preceded by a U-rich sequence and followed by a 3' poly(U) tail. *Rna* 18(5):1062–1074.
9. Ikeda Y, Yagi M, Morita T, Aiba H (2011) Hfq binding at RhlB-recognition region of RNase

E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in Escherichia coli. *Mol Microbiol* 79(2):419–432.

10. Morita T, Maki K, Aiba H (2005) RNase E-based ribonucleoprotein complexes: Mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev* 19(18):2176–2186.

11. Bandyra KJ, et al. (2012) The Seed Region of a Small RNA Drives the Controlled Destruction of the Target mRNA by the Endoribonuclease RNase E. *Mol Cell* 47(6):943–953.

12. Smirnov A, et al. (2016) Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc Natl Acad Sci* 113(41):11591–11596.

13. Smirnov A, Wang C, Drewry LL, Vogel J (2017) Molecular mechanism of mRNA repression in trans by a ProQ-dependent small RNA. *EMBO J* 36:1029–1045.

14. Gaballa A, et al. (2008) The Bacillus subtilis iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. *Proc Natl Acad Sci* 105(33):11927–11932.

15. Smaldone GT, Antelmann H, Gaballa A, Helmann JD (2012) The FsrA sRNA and FbpB protein mediate the iron-dependent induction of the Bacillus subtilis lutABC iron-sulfur-containing oxidases. *J Bacteriol* 194(10):2586–2593.

16. Hämmerle H, et al. (2014) Impact of Hfq on the Bacillus subtilis transcriptome. *PLoS One* 9(6). doi:10.1371/journal.pone.0098661.

17. Rochat T, et al. (2015) Tracking the elusive function of Bacillus subtilis Hfq. *PLoS One* 10(4). doi:10.1371/journal.pone.0124977.

18. Bohn C, Rigoulay C, Bouloc P (2007) No detectable effect of RNA-binding protein Hfq absence in Staphylococcus aureus. *BMC Microbiol* 7. doi:10.1186/1471-2180-7-10.

19. Bøggild A, Overgaard M, Valentin-Hansen P, Brodersen DE (2009) Cyanobacteria contain a structural homologue of the Hfq protein with altered RNA-binding properties. *FEBS J* 276(14):3904–3915.

20. Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C (2009) BMC Structural Biology A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct Biol* 9(9).

21. Mezulis S, Yates CM, Wass MN, E Sternberg MJ, Kelley LA (2015) The Phyre2 web portal for protein modeling, prediction and analysis. doi:10.1038/nprot.2015.053.

22. Christiansen JK, et al. (2004) The RNA-Binding Protein Hfq of Listeria monocytogenes:

Role in Stress Tolerance and Virulence. *J Bacteriol* 186(11):3355–3362.

23. Gao M, Barnett MJ, Long SR, Teplitski M (2010) Role of the Sinorhizobium meliloti global regulator Hfq in gene regulation and symbiosis. *Mol Plant Microbe Interact* 23(4):355–365.

24. Boudry P, et al. (2014) Pleiotropic role of the RNA chaperone protein Hfq in the human pathogen Clostridium difficile. *J Bacteriol* 196(18):3234–3248.

25. Valentin-Hansen P, Eriksen M, Udesen C (2004) The bacterial Sm-like protein Hfq: A key player in RNA transactions. *Mol Microbiol* 51(6):1525–1533.

26. Zhang X, et al. (2011) Identification of RNAIII-binding proteins in Staphylococcus aureus using tethered RNAs and streptavidin aptamers based pull-down assay. *BMC Microbiol* 15(1).

27. Said N, et al. (2009) In vivo expression and purification of aptamer-tagged small RNA regulators. *Nucl Acid Research* 37(20).

28. Srisawat C, Engelke DR (2001) Streptavidin aptamers: Affinity tags for the study of RNAs and ribonucleoproteins. *RNA* 7(4):632-641.

29. Roux KJ, Kim DI, Burke B (2013) BioID: A Screen for Protein-Protein Interactions. *Curr Protoc Protein Sci* 7419232314:1–19.

30. Roux KJ, Kim DI, Raida M, Burke B (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* 196(6):801–10.

31. Daigle N, Ellenberg J (2007) l N -GFP: an RNA reporter system for live-cell imaging. *Nat Methods* 4(8):633-636.

32. Burroughs AM, Balaji S, Iyer LM, Aravind L (2007) Small but versatile: The extraordinary functional and structural diversity of the β-grasp fold. *Biol Direct* 2. doi:10.1186/1745-6150-2-18.

33. Schuergers N, et al. (2014) Binding of the RNA chaperone Hfq to the type IV pilus base is crucial for its function in Synechocystis sp. PCC 6803. *Mol Microbiol* 92(4):840–852.

34. Dienst D, et al. (2008) The cyanobacterial homologue of the RNA chaperone Hfq is essential for motility of Synechocystis sp. PCC 6803. *Microbiology* 154(10):3134–3143.

35. Galamba A, et al. (2001) Disruption of adhC reveals a large duplication in the Mycobacterium smegmatis mc 2 155 genome. *Microbiology* 147:3281–3294.

36. Ross W, et al. (1993) A Third Recognition Element in Bacterial Promoters: DNA Binding by the Subunit of RNA Polymerase. *Science* 262(5138):1407–1413.

37. Finney AH, Blick RJ, Murakami K, Ishihama A, Stevens AM (2002) Role of the C-Terminal Domain of the Alpha Subunit of RNA Polymerase in LuxR-Dependent Transcriptional Activation of the lux Operon during Quorum Sensing. *J Bacteriol* 184(16):4520–4528.

38. Liu K, Zhang Y, Severinov K, Das A, Hanna MM (1996) Role of Escherichia coli RNA polymerase alpha subunit in modulation of pausing, termination and anti-termination by the transcription elongation factor NusA. *EMBO J* 15(1):150–61.

39. Czyz A, Mooney RA, Iaconi A, Landick R (2014) Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *MBio* 5(2). doi:10.1128/mBio.00931-14.

40. Morita T, Ueda M, Kubo K, Aiba H (2015) Insights into transcription termination of Hfq-binding sRNAs of Escherichia coli and characterization of readthrough products. *Rna* 21(8):1490–1501.

41. Jousselin A, Metzinger L, Felden B (2009) On the facultative requirement of the bacterial RNA chaperone, Hfq. *Trends Microbiol* 17(9):399–405.

42. Walker BJ, et al. (2014) Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11). doi:10.1371/journal.pone.0112963.

Chapter 5

Discussion

5.1 sRNAs in mycobacteria

5.1.1 sRNA discovery and expression profiling in mycobacteria

Here, we report the most extensive sRNA discovery study in *M. tuberculosis* to date, and discover approximately 100 novel sRNA candidates that were not found in previous sequencing-based sRNA discovery studies (1–3). We additionally create the computational sRNA search tool BS_Finder, which identifies accurate boundaries of sRNAs and can be applied to sRNA-Sequencing datasets from any prokaryotic species (4). We identified many candidate sRNAs that have differential expression in response to stress conditions that are relevant to *M. tuberculosis* pathogenesis in the host: membrane stress, oxidative stress, low pH, nutrient limitation, and iron starvation (5–12). Each of the conditions tested resulted in both increases and decreases in expression of sRNAs, but some conditions had more highly differential candidates than others. For example, the most highly induced sRNA we identified was ncRv11429, which is induced greater than 300-fold during SDS-mediated envelope stress. This highlights the importance of performing sRNA discovery screens during stress conditions, since the sRNA is not expressed during normal growth and therefore was not identified in previous studies. Furthermore, the induction of an sRNA during specific stresses provides invaluable clues regarding the sRNA's function and conditions under which to mechanistically probe its regulatory role. For instance, envelope stress-responsive sRNAs are a well-studied class of sRNA (13–18) and it will be interesting to determine whether ncRv11429 performs similar functions, particularly given the differences between the mycobacterial and gram negative cell envelopes. Additional in-depth studies on more *M. tuberculosis* sRNAs will be vital for our understanding of post-transcriptional gene regulation in this important pathogen.

Only a few of the *M. tuberculosis* sRNA candidates are highly conserved across both pathogenic and nonpathogenic mycobacterial lineages. Of the most highly differentially expressed sRNA candidates, only MrsI and the mycobacterial homologue of the 6S RNA have obvious homologues in the saprophytic *M. smegmatis*. This could either be because selective pressure to preserve the sequence of these sRNAs is low across the majority of the transcript, or because these sRNAs are indeed specific to pathogenic mycobacterial species. Indeed, previous studies have shown that the seed region is under stronger selective pressure to remain conserved than other portions of sRNAs (19), and due to the small size of seed regions searching for homologues by overall sequence conservation could therefore be relatively insensitive. Although MrsI is highly conserved in its seed region, an approximately 35 nucleotide hairpin encompassing the seed is also very highly conserved (85% identity between *M. tuberculosis* and *M. smegmatis,* compared to 35% identity for the rho-independent terminator). This large conserved domain allowed for identification of homologues of MrsI in many mycobacterial species. However, the reason for the conservation of this hairpin is still unknown, and if it is a MrsI-specific reason, such as acting as a recognition site for a MrsI-specific accessory factor, then one would expect that conservation analysis of other *M. tuberculosis* sRNAs could be less able to identify homologues outside of more closely related mycobacterial species. Alternatively, if most sRNAs in *M. tuberculosis* do in fact lack true homologues outside of the *M. tuberculosis* complex it could indicate that these sRNAs have evolved in response to host-specific stresses or to facilitate virulence. Further experiments will be necessary in order to

better understand these results, including characterization of targets of the non-conserved sRNAs and sRNA discovery in other mycobacterial species during stress.

5.1.2 A mycobacterial iron sparing small RNA

Iron is an essential nutrient for mycobacterial growth and pathogenesis, and the arms race between the bacterium and host for this metal has resulted in complex machineries to both acquire and regulate the use of the metal (20–22). Although much of the regulatory machinery governing iron acquisition and use have been well studied in mycobacteria, by performing the most in-depth characterization to date of a mycobacterial sRNA we identified a mycobacterial iron-sparing sRNA. We show that ncRv11846/MrsI, which we initially identified as an sRNA that was highly induced in multiple stress conditions, primarily functions as an iron sparing sRNA in both *M. smegmatis* and *M. tuberculosis*. We identify the set of transcripts regulated by MrsI in both pathogenic and nonpathogenic mycobacteria and validate that MrsI negatively regulates the *bacterioferritin A* (*bfrA)* mRNA by directly binding to the 5' UTR of the transcript, which is the first validation of a direct interaction between a mycobacterial sRNA and a target. Our results provide important insights into sRNA-mediated riboregulation in mycobacteria; for example, two established sRNA target prediction tools did not predict any of our experimentally identified MrsI targets to be regulated by the sRNA (23, 24). Closer examination revealed that this may be due to the less energetically favorable predicted binding energies for MrsI and its targets in *M. smegmatis* compared to the sRNA-target interactions on which these programs are trained, of which none are mycobacterial. Thus, it appears that traditional target prediction algorithms may need to be tuned to be accurate for mycobacteria, although validation of sRNA-

target pairings of other mycobacterial sRNAs will be necessary to determine the extent of these differences.

5.1.3 Promiscuous sRNA induction and anticipatory responses

Iron sparing sRNAs have been identified in several bacterial species, and the *E. coli* iron sparing sRNA RyhB is one of the most well-characterized (25–28). However, MrsI in *M. tuberculosis* is unique among iron sparing sRNAs in that it is highly induced in three separate stress conditions: iron starvation, oxidative stress, and membrane stress. Despite its strong induction in three stresses in *M. tuberculosis,* MrsI appears to mediate robust regulatory effects only during iron starvation. We demonstrate that exposure of the bacteria to oxidative stress prior to iron starvation primes MrsI to mediate a more rapid iron sparing response, consistent with MrsI induction during oxidative stress and membrane stress serving to anticipate iron starvation. This predictive regulation could occur in *M. tuberculosis* due to the fact that the stresses that the bacteria encounter are due to an immune program by the host and therefore occur in a predictable fashion. Anticipatory regulation in response to predictable variations in environments has been documented in other microbes, such as *E. coli* (29, 30). For example, when *E. coli* travels through the gut it encounters the carbon source lactose at an earlier point than maltose. Growth of *E. coli* in the presence of lactose results in increased expression of the maltose utilization operon, whereas growth on maltose does not result in increased expression of lactose utilization genes, and this pre-induction of maltose utilization genes provides a fitness advantage (30). Additional examples of anticipatory regulation have been found for *E. coli*, as well as in *Saccharomyces cerevisiae*, but this is the first evidence of an sRNA being

involved in a predictive response. Future work will be necessary to determine the extent to which predictive regulation occurs in *M. tuberculosis*, both through the use of sRNAs and protein-mediated responses. It will be particularly interesting to characterize the regulons of the other two promiscuously induced sRNAs we identified, ncRv11803 and ncRv12659, in the conditions in which they accumulate. Whether these two sRNAs have evolved to help the bacteria adapt to multiple stress conditions or if they are involved in predictive regulatory networks will be an exciting avenue of research.

5.2 A mycobacterial sRNA accessory factor

Here, we identified MhfQ, a mycobacterial structural homologue of the sRNA chaperone protein Hfq which mycobacteria were predicted to lack. However, our results suggest that MhfQ does not function as an sRNA accessory factor in mycobacteria. We additionally performed biochemical and genetic screens toward the identification of novel accessory factors in *M. smegmatis*, but these experiments failed to produce any candidates. Given the results of these studies, one has reason to question whether mycobacteria contain an sRNA chaperone protein at all, and if sRNAs in this lineage mediate their regulation without assistance from any other factors. However, multiple pieces of evidence suggest that this is not the case. First, the MrsI-target interactions we identified are predicted to be less energetically favorable than Hfq-dependent sRNA-target interactions from other bacteria, which suggests that mycobacterial sRNAs require an accessory factor to mediate binding (Chapter 3). Additionally, two hypotheses have been put forth to explain why certain lineages may be less dependent on sRNA-chaperone proteins: extended base pairing and low GC content (31). It was suggested that if sRNAs have

extensive base pairing with their target an accessory factor may not be necessary. However, we found little-to-no evidence of extensive base pairing outside the seed region between MrsI and its targets (Chapter 3). Furthermore, other mRNAs besides the MrsI targets we identified also have the 6nt MrsI binding site at their 5' ends, and if no accessory factor mediates the interaction between MrsI and mRNAs one would expect that these genes would be MrsI-regulated as well. Additionally, low GC content was proposed to obviate the requirement for chaperone proteins, as in the low GC gram positive bacteria such as Firmicutes (32, 33). However, mycobacteria are one of the more GC-rich bacterial lineages (GC content ~65%), and based on this hypothesis they would be at least as dependent on accessory factors as *E. coli* (GC content ~50%). Therefore, it seems highly likely that mycobacterial sRNA-mediated riboregulation depends on an unknown accessory element.

The lack of an obvious accessory factor candidate from the suppressor screen suggests that the protein or proteins will be either essential or redundant, as these are the two major potential explanations for the suppressor screen results. Either of these possibilities would make a genetic approach unlikely to yield candidates, and therefore further biochemical approaches will likely be necessary. One promising possibility for the identification of the accessory factor is the approach that identified ProQ as a novel accessory factor in *S. enterica*, termed Grad-Seq (34). This technique separates sRNAs based on their biochemical properties, which are almost completely dictated by the protein-binding partners of the transcripts. Therefore, Grad-Seq identifies clusters of sRNAs that bind the same accessory factor, and tandem aptamer-based pulldowns of a set of sRNAs in the same cluster allowed the authors to identify ProQ as a novel

chaperone protein. If applied to mycobacteria, this approach could identify the accessory factor(s) involved in sRNA action as well as cluster the sRNAs based on which protein they require, and would provide extremely useful information for the field.

Taken together, the results presented here are both the most comprehensive sRNA discovery approach in *M. tuberculosis* to date, as well as the most detailed characterization of an sRNA from *M. tuberculosis*. This combination provides crucial information to the field that was previously lacking. The sRNA discovery and expression profiling in *M. tuberculosis* during stress conditions provides novel, exciting sRNA candidates with potential roles in *M. tuberculosis* pathogenesis. Our characterization of MrsI-mediated riboregulation and its promiscuous induction gives important insights into the iron starvation stress response of the bacterium. Additionally, these results elucidate potential disparities in the rules governing sRNA-mediated regulation in mycobacteria compared to other bacteria, and furthermore suggest that *M. tuberculosis* uses predictive regulation involving sRNAs to rapidly adapt to host-mediated stress. These results will guide future work on *M. tuberculosis* adaptation to stress and the crucial role that sRNAs play.

References

1. Arnvig KB, et al. (2011) Sequence-Based Analysis Uncovers an Abundance of Non-Coding RNA in the Total Transcriptome of Mycobacterium tuberculosis. *Plos Pathog* 7(11):1-16.
2. Arnvig KB, Young DB (2009) Identification of small RNAs in Mycobacterium tuberculosis. *Mol Microbiol* 73(3):397–408.
3. Wang M, et al. (2016) An automated approach for global identification of sRNA-encoding

regions in RNA-Seq data from Mycobacterium tuberculosis. *Acta Biochim Biophys Sin (Shanghai)* 48(6):544–553.

4. Dejesus MA, et al. (2017) Comprehensive essentiality analysis of the Mycobacterium tuberculosis genome via saturating transposon mutagenesis. *MBio* 8(1). doi:10.1128/mBio.02133-16.

5. Alonso S, Pethe K, Russell DG, Purdy GE Lysosomal killing of Mycobacterium mediated by ubiquitin-derived peptides is enhanced by autophagy. Available at: http://www.pnas.org.ezp-prod1.hul.harvard.edu/content/pnas/104/14/6031.full.pdf [Accessed March 8, 2018].

6. Pethe K, et al. Isolation of Mycobacterium tuberculosis mutants defective in the arrest of phagosome maturation. Available at: http://www.pnas.org.ezp-prod1.hul.harvard.edu/content/pnas/101/37/13642.full.pdf [Accessed March 8, 2018].

7. Liu PT, Stenger S, Tang DH, Modlin RL (2007) Cutting Edge: Vitamin D-Mediated Human Antimicrobial Activity against Mycobacterium tuberculosis Is Dependent on the Induction of Cathelicidin. *J Immunol* 179(4):2060–2063.

8. Macgurn JA, Cox JS (2007) A Genetic Screen for Mycobacterium tuberculosis Mutants Defective for Phagosome Maturation Arrest Identifies Components of the ESX-1 Secretion System. *Infect Immun* 75(6):2668–2678.

9. Ng VH, Cox JS, Sousa AO, MacMicking JD, McKinney JD (2004) Role of KatG catalase-peroxidase in mycobacterial pathogenisis: Countering the phagocyte oxidative burst. *Mol Microbiol* 52(5):1291–1302.

10. Cooper AM, et al. (2000) Transient Loss of Resistance to Pulmonary Tuberculosis in p47 phox − / − Mice Transient Loss of Resistance to Pulmonary Tuberculosis. *Infect Immun* 68(3):1231–1234.

11. Dahl JL, et al. (2003) The role of RelMtb-mediated adaptation to stationary phase in long-term persistence of Mycobacterium tuberculosis in mice. *Proc Natl Acad Sci* 100(17):10026–10031.

12. Ratledge C (2004) Iron, mycobacteria and tuberculosis. *Tuberculosis*, pp 110–130.

13. Balbontín R, Fiorini F, Figueroa-Bossi N, Casadesús J, Bossi L (2010) Recognition of heptameric seed sequence underlies multi-target regulation by RybB small RNA in salmonella enterica. *Mol Microbiol* 78(2):380–394.

14. Bouvier M, Sharma CM, Mika F, Nierhaus KH, Vogel J (2008) Small RNA Binding to 5' mRNA Coding Region Inhibits Translational Initiation. *Mol Cell* 32(6):827–837.

15. Pfeiffer V, Papenfort K, Lucchini S, Hinton JCD, Vogel J (2009) Coding sequence targeting

by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat Struct Mol Biol* 16(8):840–846.

16. Castillo-keller M, Vuong P, Misra R (2006) Novel Mechanism of Escherichia coli Porin Regulation Novel Mechanism of Escherichia coli Porin Regulation. *J Bacteriol* 188(2):576–586.

17. Guo MS, et al. (2014) MicL, a new σE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev* 28(14):1620–1634.

18. Papenfort K, et al. (2006) σE-dependent small RNAs of Salmonella respond to membrane stress by accelerating global omp mRNA decay. *Mol Microbiol* 62(6):1674–1688.

19. Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J (2012) An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J* 31(20):4005–4019.

20. Schaible UE, Kaufmann SHE (2004) Iron and microbial infection. *Nat Rev Microbiol* 2(12):946–953.

21. Rodriguez GM (2006) Control of iron metabolism in Mycobacterium tuberculosis. *Trends Microbiol* 14(7):320–327.

22. Ganz T (2009) Iron in innate immunity: starve the invaders. *Curr Opin Immunol* 21(1):63–67.

23. Kery MB, Feldman M, Livny J, Tjaden B (2014) TargetRNA2: Identifying targets of small regulatory RNAs in bacteria. *Nucleic Acids Res* 42(W1):W124-9.

24. Wright PR, et al. (2013) Comparative genomics boosts target prediction for bacterial small RNAs. *Proc Natl Acad Sci* 110(37):E3487–E3496.

25. Massé E, Vanderpool CK, Gottesman S (2005) Effect of RyhB small RNA on global iron use in Escherichia coli. *J Bacteriol* 187(20):6962–6971.

26. Reinhart AA, et al. (2015) The prrF-encoded small regulatory RNAs are required for iron homeostasis and virulence of Pseudomonas aeruginosa. *Infect Immun* 83(3):863–875.

27. Gaballa A, et al. (2008) The Bacillus subtilis iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. *Proc Natl Acad Sci* 105(33):11927–11932.

28. Oglesby-Sherrouse AG, Murphy ER (2013) Iron-responsive bacterial small RNAs: variations on a theme. *Metallomics* 5(5):1756–5901.

29. Tagkopoulos I, Liu YC, Tavazoie S (2008) Predictive behavior within microbial genetic networks. *Science (80-)* 320(5881):1313–1317.

30. Mitchell A, et al. (2009) Adaptive prediction of environmental changes by microorganisms. *Nature* 460(7252):220–224.

31. Jousselin A, Metzinger L, Felden B (2009) On the facultative requirement of the bacterial RNA chaperone, Hfq. *Trends Microbiol* 17(9):399–405.

32. Hämmerle H, et al. (2014) Impact of Hfq on the Bacillus subtilis transcriptome. *PLoS One* 9(6). doi:10.1371/journal.pone.0098661.

33. Rochat T, et al. (2015) Tracking the elusive function of Bacillus subtilis Hfq. *PLoS One* 10(4). doi:10.1371/journal.pone.0124977.

34. Smirnov A, et al. (2016) Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc Natl Acad Sci* 113(41):11591–11596.

Appendices

Appendix 1. Discovery and profiling of sRNAs in *M. tuberculosis*

Table A1.1. Master list of sRNAs in *M. tuberculosis* and their expression during stress

name	strand	start	end	log_2fold change low iron	log_2fold fold change tBHP	log_2fold fold change SDS	log_2fold fold change acid	log_2fold fold change PBS
ncRv10006	+	9824	9877	1.58	0.34	-0.36	-0.31	-1.42
ncRv0046c	-	49956	50033	0.25	1.72	0.36	1.63	4.23
ncRv0047c	-	51723	51779	-0.86	1.40	1.64	3.09	3.87
ncRv0053	+	58117	58193	-0.85	-1.17	-0.25	-1.48	0.46
ncRv10071	+	80239	80441	-0.61	-0.99	-0.77	1.65	-1.52
ncRv10071c	-	80254	80344	-0.66	-0.42	0.29	0.13	0.37
ncRv10128	+	156452	156568	-0.85	-1.78	-0.82	-0.22	-1.83
ncRv10128c	-	156519	156568	-1.11	0.48	0.31	0.23	3.40
ncRv10150c	-	177236	177285	1.10	-0.10	0.19	-0.22	0.04
ncRv0179	+	209683	209800	0.35	-0.49	0.19	0.00	-4.14
ncRv0186	+	218318	218381	-1.74	0.35	-0.93	0.00	-0.41
ncRv0186c	-	218320	218379	0.90	0.72	4.14	2.67	1.50
ncRv0233	+	279379	279584	0.83	0.55	0.03	1.66	2.60
ncRv0236c	-	286844	286901	-2.55	-0.22	-1.06	-0.72	-1.24
ncRv10243A	+	293527	293667	1.64	0.35	1.09	0.90	1.36
ncRv10243B	+	293603	293663	1.90	0.18	1.09	0.99	1.56
ncRv10243c	-	293665	293712	-0.49	1.69	2.12	-1.16	-0.54
ncRv0314	+	382877	382979	-0.93	0.44	0.92	1.02	1.82
ncRv0346c	-	415475	415547	0.04	-1.19	-3.14	-3.43	-1.67
ncRv0440	+	530097	530288	-1.51	0.41	-0.68	-1.50	-4.21
ncRv10440c	-	530246	530295	1.56	0.51	-0.33	-0.73	1.67
ncRv0441Ac	-	530246	530353	1.42	0.33	-0.23	-0.71	1.58
ncRv0441Bc	-	530246	530425	1.41	0.37	-0.19	-0.57	1.59
ncRv0467	+	558801	558885	1.38	1.79	1.52	-3.46	-3.39
ncRv0475	+	566256	566454	-0.58	1.16	1.78	0.63	-1.18
ncRv10475	+	566402	566455	0.08	1.06	1.88	0.63	-0.74
ncRv0490	+	579290	579378	0.21	-0.88	-0.75	1.54	0.02
ncRv10609	+	704185	704246	1.09	0.52	0.56	0.62	0.71
ncRv10637	+	733361	733459	-0.05	-0.71	1.17	-0.07	0.72
ncRv0638	+	734118	734244	-0.07	-0.85	-0.41	-2.60	-1.76
ncRv0641	+	736002	736284	-0.08	-0.66	0.41	-1.21	-2.61
ncRv10666	+	759480	759610	1.17	-0.34	-1.41	-1.83	0.58
ncRv0668	+	767311	767378	-1.30	-0.73	-0.83	-1.67	-3.96
ncRv10668	+	767334	767378	-1.29	-0.75	-0.87	-1.73	-4.18
ncRv0685	+	785795	786071	-0.39	-1.38	-1.69	-2.18	-2.61
ncRv10685	+	786021	786071	0.11	-1.86	-1.99	-3.47	-3.04
ncRv10699	+	800268	800314	-0.31	-1.00	0.48	-1.91	2.13
ncRv0710	+	806130	806220	-0.53	-1.93	-2.00	-2.68	-5.08
ncRv10710c	-	806180	806282	0.53	0.81	3.91	3.16	1.60

Table A1.1 (continued)

ncRv10713	+	811228	811296	0.24	-2.60	-1.19	-3.51	-1.17
ncRv0751	+	842236	842354	0.75	0.44	2.62	0.71	0.37
ncRv10815c	-	909328	909369	-0.95	-0.75	-0.97	-3.26	-0.52
ncRv0820	+	913454	913530	0.02	-0.64	-1.85	1.04	-2.36
ncRv0822c	-	916265	916424	2.31	0.31	0.03	-0.58	0.37
ncRv0860	+	958344	958509	0.93	-0.95	-1.21	-2.97	-3.33
ncRv10860	+	958459	958509	1.04	-1.24	-1.65	-2.59	-3.13
ncRv0897A	+	1000569	1000826	-0.39	-0.27	-1.27	-0.15	-0.75
ncRv0897B	+	1000719	1000826	-0.30	-0.16	-0.84	-0.12	-0.42
ncRv10932Ac	-	1041129	1041165	0.18	-0.90	-0.56	2.57	-1.40
ncRv0952	+	1063969	1064101	-0.70	-1.28	-1.41	-2.43	-4.66
ncRv10996	+	1113599	1113673	-0.68	-0.42	-0.82	0.04	0.02
ncRv1008	+	1127875	1127999	-0.09	-0.29	1.34	2.12	-1.43
ncRv11008	+	1127949	1127999	0.65	-0.51	1.37	2.27	-2.08
ncRv11013	+	1133262	1133316	-0.43	0.57	-0.83	-0.94	-0.94
ncRv11042c	-	1165548	1165613	1.86	0.32	0.10	0.33	1.81
ncRv1065	+	1187351	1187589	0.47	-0.16	-1.71	1.49	3.57
ncRv1072A	+	1197046	1197179	-0.42	1.13	2.60	-0.55	1.83
ncRv1072B	+	1197046	1197200	-0.42	1.14	2.59	-0.48	1.86
ncRv1072C	+	1197082	1197179	-0.41	1.13	2.60	-0.62	1.84
ncRv11092c	-	1220388	1220486	-0.60	-0.21	0.68	-0.28	0.00
ncRv1094	+	1222774	1222976	-1.14	-0.28	0.23	1.31	-2.15
ncRv1095	+	1224272	1224347	0.15	-0.59	0.65	2.42	2.69
ncRv11144Ac	-	1271918	1271961	1.67	-4.92	-0.31	-2.35	3.09
ncRv11144Bc	-	1271918	1272169	1.90	-4.78	-0.15	-2.43	3.84
ncRv11147c	-	1275610	1275674	1.62	-0.78	-0.33	-1.17	-0.20
ncRv11160A	+	1287126	1287201	0.99	0.03	-0.28	1.10	1.56
ncRv1171	+	1301253	1301327	2.59	1.09	1.71	1.33	2.61
ncRv11179c	-	1313344	1313452	1.34	0.35	0.48	-0.93	2.12
ncRv1181	+	1315110	1315241	3.35	1.11	0.73	1.26	1.55
ncRv11199	+	1342888	1342941	0.63	-0.60	0.26	-1.77	-1.02
ncRv1220c	-	1363447	1363506	-0.10	-0.39	-0.70	-0.53	0.05
ncRv11264c	-	1413105	1413227	3.96	1.08	0.93	0.48	3.74
ncRv1298	+	1455386	1455461	-1.32	1.06	-0.54	0.57	0.88
ncRv11298	+	1455406	1455461	-1.35	1.07	-0.57	0.59	0.91
ncRv11315A	+	1471620	1471865	-0.56	-0.46	1.49	2.33	1.16
ncRv11315B	+	1473385	1473503	0.10	-1.04	-0.24	0.66	-0.59
ncRv11315C	+	1476586	1476885	-2.07	-0.42	0.21	-0.37	0.93
ncRv11315D	+	1476976	1477061	-0.92	-0.27	2.93	2.14	2.10
ncRv1324	+	1487997	1488133	-1.18	-1.15	-0.70	-0.44	-3.25
ncRv1329	+	1497132	1497220	-0.46	-0.96	-0.79	-0.57	-2.38
ncRv1389	+	1564297	1564499	-0.17	-0.36	-0.95	0.18	-1.29
ncRv1404	+	1581047	1581140	-0.88	0.83	1.94	1.42	1.93
ncRv11429	+	1606254	1606319	0.12	-1.24	8.38	-0.94	0.89
ncRv1461	+	1646930	1647224	4.06	4.22	1.31	0.61	0.22
ncRv11534	+	1735508	1735687	-0.24	-0.46	-1.86	-1.65	1.78

Table A1.1 (continued)

ncRv11565c	-	1773845	1773907	-1.99	1.52	-0.25	-1.01	-0.50
ncRv1617	+	1816131	1816235	2.45	-0.42	-0.87	-0.57	-0.45
ncRv1621Ac	-	1821646	1821753	1.47	0.14	0.94	0.65	-0.71
ncRv1621Bc	-	1821646	1821881	0.93	0.26	1.09	0.94	-1.05
ncRv1626	+	1828761	1828849	-1.44	-0.39	0.07	-0.71	-3.05
ncRv11689Ac	-	1914963	1915028	-0.57	0.75	0.68	-0.16	1.55
ncRv11689Bc	-	1915034	1915091	-0.53	0.71	0.96	0.37	0.57
ncRv11733	+	1960667	1960774	1.65	0.24	0.94	1.93	3.26
ncRv11779c	-	2014582	2014627	0.45	-1.09	-1.40	2.97	-1.02
ncRv11793	+	2030986	2031034	0.40	0.84	-2.05	-0.39	1.54
ncRv1796	+	2033679	2033904	0.01	-0.29	-1.91	-1.47	0.30
ncRv11803	+	2046922	2046995	-0.07	4.13	4.17	6.26	3.25
ncRv11817c	-	2061060	2061143	-0.07	0.12	0.02	3.17	1.30
ncRv1821	+	2068863	2068962	-0.43	-0.72	-0.88	-1.01	-0.32
ncRv11829	+	2074745	2074827	-0.76	-0.36	-1.30	1.60	3.02
ncRv11846	+	2096758	2096863	8.09	6.05	5.14	0.34	1.59
ncRv12023	+	2268164	2268231	-0.58	0.02	0.77	0.01	-0.29
ncRv2165Ac	-	2429342	2429502	-0.52	-2.18	1.50	-1.62	0.72
ncRv12196	+	2461328	2461464	0.06	0.05	-0.09	0.65	-1.82
ncRv2197c	-	2461394	2461511	0.41	0.38	-0.52	-0.10	0.28
ncRv2213c	-	2479780	2479905	-0.71	0.99	-0.26	0.88	-1.84
ncRv2214c	-	2479780	2480069	-0.15	0.85	-0.47	0.73	-1.90
ncRv12220A	+	2489055	2489252	0.91	-0.23	-0.34	-0.13	-1.53
ncRv12220B	+	2489205	2489252	1.00	-0.20	-0.16	-0.85	-1.42
ncRv2237	+	2510162	2510341	-0.21	-0.31	-0.93	-0.58	0.94
ncRv12237c	-	2510260	2510339	1.83	-0.04	-1.63	1.30	1.90
ncRv2245	+	2518022	2518132	-0.57	3.76	-1.57	-0.50	0.41
ncRv2247	+	2522152	2522224	0.43	2.27	-0.76	-1.04	-0.68
ncRv2346c	-	2625801	2625981	1.45	-0.19	-0.18	0.50	0.98
ncRv2395	+	2692172	2692404	1.57	0.99	-2.67	-0.14	2.77
ncRv12444c	-	2744989	2745080	0.61	-0.39	3.79	1.58	2.01
ncRv2454c	-	2753582	2753692	2.71	0.55	-0.58	-0.45	-1.13
ncRv12497c	-	2812101	2812274	-0.40	0.65	1.55	4.70	3.34
ncRv12557	+	2877751	2877807	1.46	0.73	-0.28	1.14	2.98
ncRv2624	+	2950409	2950584	-0.06	-0.05	0.48	-0.57	0.61
ncRv12641	+	2966405	2966450	1.24	-0.18	0.12	1.10	-0.53
ncRv12649c	-	2973672	2973791	-1.41	0.24	1.73	0.62	2.02
ncRv12659	+	2980911	2981083	0.21	1.39	5.27	5.53	7.62
ncRv12674	+	2991119	2991174	-1.04	-0.17	0.11	0.44	-0.36
ncRv12710	+	3023435	3023497	0.67	1.94	3.66	0.18	2.89
ncRv12765c	-	3075379	3075433	-2.66	-1.10	-1.26	-1.84	-2.88
ncRv12783c	-	3092761	3092889	-0.79	-0.94	-0.57	-0.72	-1.87
ncRv2789c	-	3097653	3097719	0.91	1.42	-0.22	0.41	0.52
ncRv2791Ac	-	3100156	3100341	-2.74	0.22	-1.08	0.11	-0.41
ncRv2791Bc	-	3100156	3100377	-2.75	0.26	-1.11	0.12	-0.38
ncRv2796c	-	3104990	3105073	-0.44	-2.15	-2.34	-3.11	-1.99

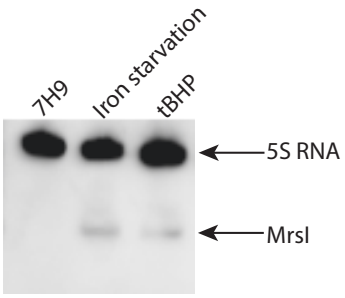
Table A1.1 (continued)

ncRv12803	+	3112324	3112402	0.01	0.14	-1.62	0.01	-0.47
ncRv2840c	-	3148103	3148316	-0.02	-0.45	-0.07	0.00	-0.71
ncRv2841	+	3148341	3148436	-0.68	-1.24	-0.63	-1.82	-0.02
ncRv2861c	-	3173101	3173170	-1.23	-0.67	-2.44	-0.87	-1.38
ncRv2885	+	3194139	3194249	0.58	-0.09	0.48	1.48	1.22
ncRv2926c	-	3240464	3240614	-0.55	1.00	0.60	1.94	3.23
ncRv12950c	-	3302595	3302692	-0.60	0.31	0.58	-0.78	0.28
ncRv2964	+	3316459	3316614	0.86	0.40	3.94	1.29	-1.05
ncRv2986Ac	-	3343113	3343216	-0.69	2.43	1.35	1.75	2.79
ncRv2986Bc	-	3343113	3343246	-0.70	2.41	1.28	1.70	2.69
ncRv2993c	-	3350950	3351071	-0.32	-1.44	-4.32	-4.83	-3.68
ncRv13003c	-	3363019	3363152	0.46	-0.41	0.71	-1.64	-0.14
ncRv3009	+	3367227	3367347	-0.41	0.75	-0.02	-0.24	-0.66
ncRv13053c	-	3415195	3415325	-0.13	-0.06	2.58	3.51	1.20
ncRv13059	+	3421005	3421100	0.35	0.30	0.93	0.29	2.91
ncRv3061	+	3423239	3423349	0.17	-0.72	1.37	2.95	-1.77
ncRv3062	+	3425499	3425632	0.11	0.84	3.20	2.85	1.27
ncRv3140	+	3507781	3508051	-0.70	1.83	-1.03	-3.88	-4.08
ncRv13191	+	3558709	3558799	-0.90	-0.10	1.53	0.50	-1.50
ncRv13210	+	3587651	3587785	-1.75	0.28	1.57	3.20	1.37
ncRv3213c	-	3590641	3590875	-0.94	-1.00	-2.03	-0.59	-2.11
ncRv3219	+	3595951	3596019	0.48	-0.14	1.22	-2.77	-1.45
ncRv3248c	-	3628095	3628180	-1.11	-0.50	0.08	-2.86	-2.56
ncRv3285	+	3666307	3666546	-1.01	0.18	-0.08	1.99	-1.08
ncRv13303	+	3690941	3691060	0.25	-2.06	-2.19	-1.74	-3.13
ncRv3312c	-	3700701	3700749	0.66	0.60	-1.01	1.17	2.19
ncRv13317	+	3704917	3704981	0.02	-0.62	-0.04	0.30	0.17
ncRv13319c	-	3707580	3707640	-1.31	-1.25	-1.09	-1.18	-2.41
ncRv3402c	-	3820566	3820714	3.16	4.28	-0.13	-1.02	-2.25
ncRv13418Ac	-	3837297	3837458	-0.42	0.50	-0.42	0.51	-0.81
ncRv13418Bc	-	3837346	3837458	-0.41	0.47	-0.41	0.49	-0.77
ncRv3424	+	3841611	3841741	-1.69	-0.26	-1.28	-1.30	-2.65
ncRv3555	+	3994763	3994831	-0.25	-0.35	-0.63	-0.53	-2.14
ncRv3566c	-	4008416	4008489	1.22	1.60	-2.97	-1.33	-1.89
ncRv3575	+	4017088	4017168	-0.73	1.76	0.42	1.10	0.58
ncRv3582c	-	4024269	4024488	-0.44	0.83	0.29	0.37	-1.07
ncRv13635	+	4075410	4075455	-1.35	-1.54	-3.56	-1.18	-1.93
ncRv3648c	-	4088267	4088350	-0.03	1.32	0.71	2.54	1.08
ncRv13660Ac	-	4099381	4099442	0.35	1.70	2.71	1.32	2.23
ncRv13660Bc	-	4099384	4099477	0.23	1.56	2.65	1.25	2.22
ncRv13661A	+	4100683	4100981	1.37	1.84	0.09	1.39	5.60
ncRv13661B	+	4101158	4101232	2.61	1.22	1.41	3.03	1.46
ncRv13707c	-	4151043	4151091	2.56	-0.36	-1.07	-0.61	-1.50
ncRv13709A	+	4153538	4153607	0.70	0.94	-0.22	0.66	2.10
ncRv13709B	+	4153620	4153691	-0.83	1.37	-0.07	1.31	-0.02
ncRv13709C	+	4153620	4153713	-0.70	1.32	-0.05	1.37	0.09

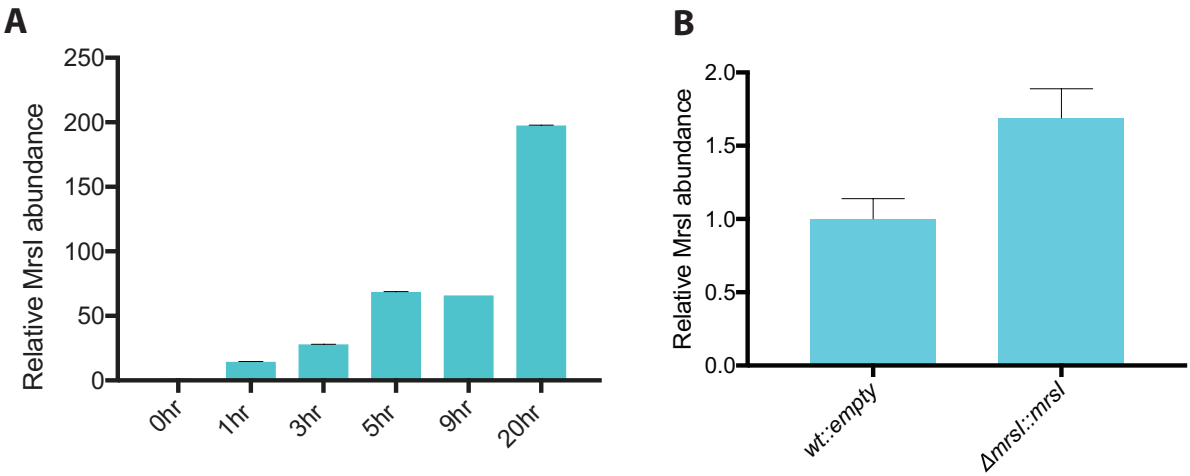
Table A1.1 (continued)

ncRv3714c	-	4159796	4159886	-1.86	-0.40	0.88	0.19	0.50
ncRv13722c	-	4168192	4168281	1.04	-0.76	0.27	-0.11	0.45
ncRv13803c	-	4265583	4265630	0.33	-2.15	-2.74	-4.56	-4.58
ncRv3804c	-	4265583	4265765	-0.13	-1.42	-1.10	-1.69	-3.33
ncRv13809c	-	4273549	4273615	-2.27	-0.13	0.53	1.67	2.74
ncRv3842	+	4314704	4314984	-4.92	-0.59	1.91	4.77	2.02
ncRv13843c	-	4316635	4316724	-0.98	1.38	0.81	2.75	4.03
ncRv3848	+	4322250	4322330	2.41	-2.01	-0.51	-1.45	1.67
ncRv3850	+	4323899	4324126	0.88	-0.10	-0.89	-1.05	-0.19
ncRv3875	+	4352801	4352974	0.84	0.08	0.49	1.34	1.01
ncRv3890c	-	4373660	4373787	-0.93	0.60	-1.79	-0.94	0.72
ncRv13907	+	4393229	4393358	-0.87	-0.95	-2.07	-2.05	-3.08

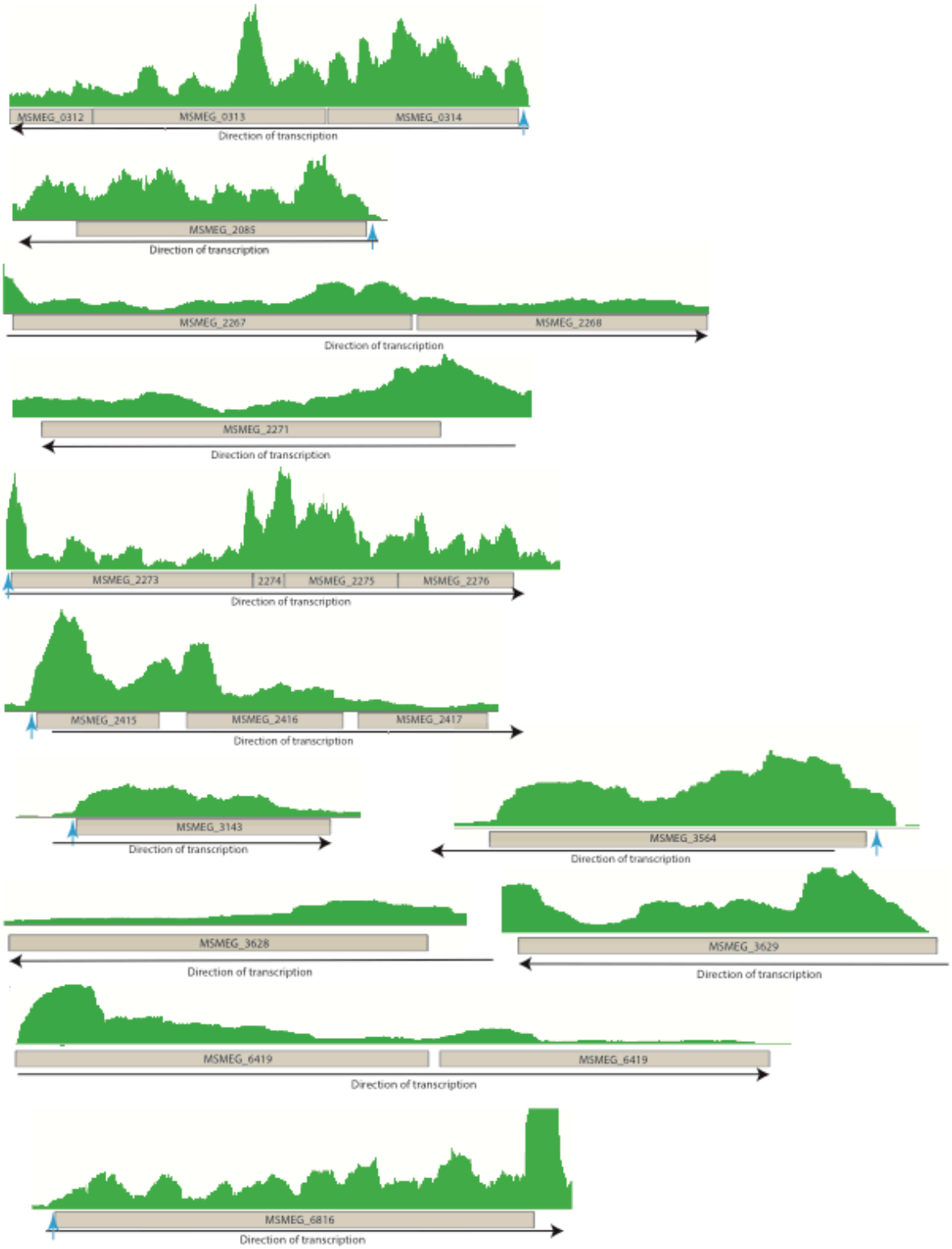
Appendix 2. Supplemental materials for Chapter 3



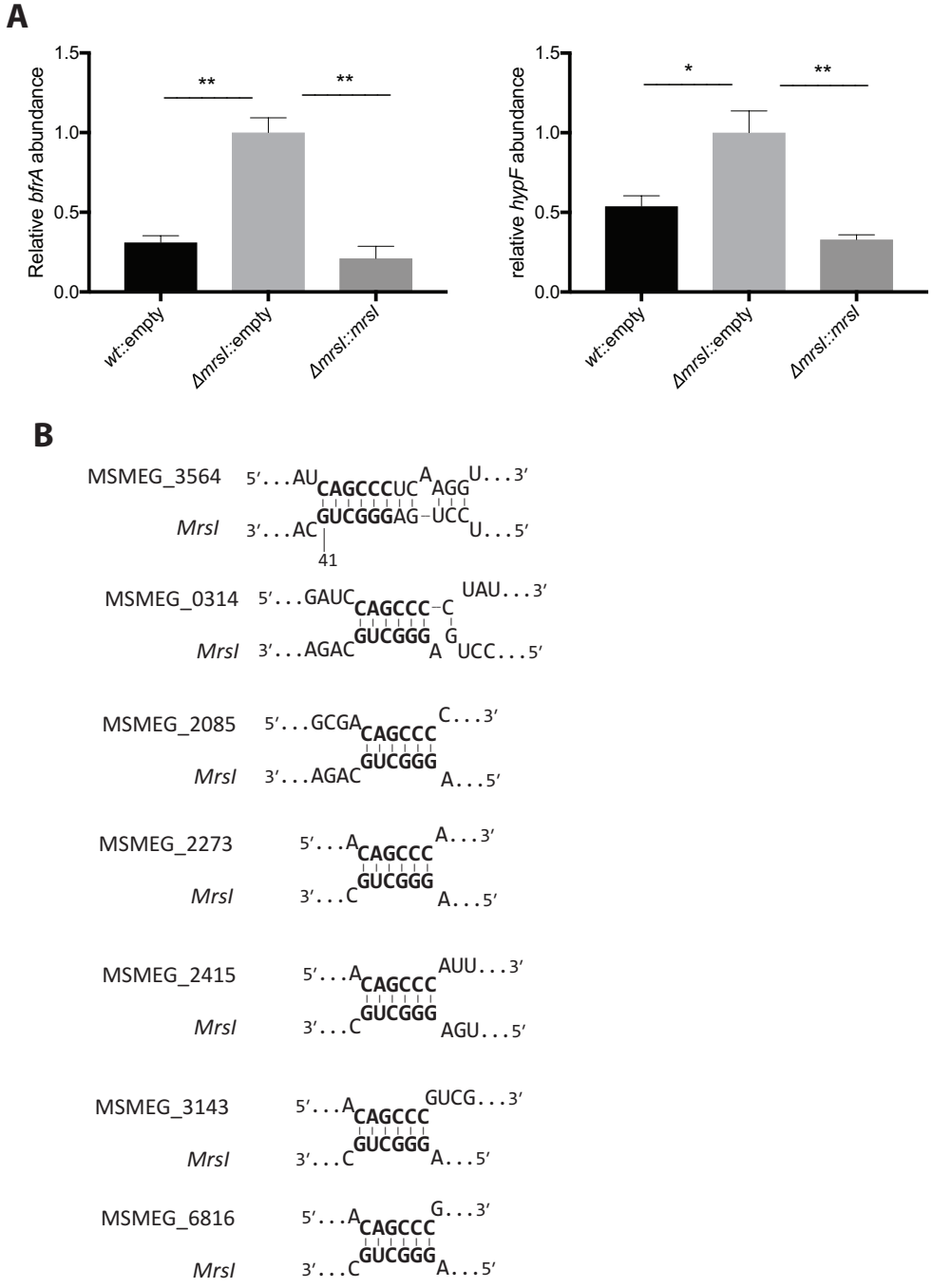
Appendix Figure A2.1: Northern blot analysis of MrsI from *M. tuberculosis* after growth in 7H9, iron starvation, and oxidative stress. Arrows indicate the 5S RNA and MrsI bands.



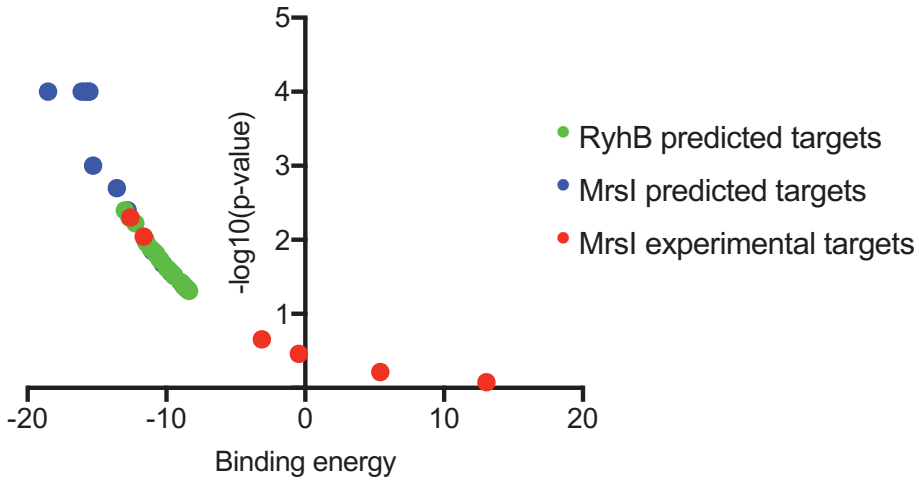
Appendix Figure A2.2: Expression levels of MrsI in *M. smegmatis* during iron starvation. A) Wild type *M. smegmatis* was starved of iron for 20 hours and the levels of MrsI were measured at multiple time points during growth by RT-qPCR. Error bars represent the SD of 3 replicates. B) Expression levels of MrsI were measured by RT-qPCR in the *wt::empty* and *ΔmrsI::mrsI* strains after 20 hours of iron starvation.



Appendix Figure A2.3: Transcription units putatively regulated by MrsI in *M. smegmatis*. Visualization of each of the transcription units found by RNA-Seq to be regulated by MrsI. Direction of transcription is indicated (black arrow), as is the putative MrsI binding site in the 5' UTR of each transcript for which MrsI is predicted to bind (blue arrow).



Appendix Figure A2.4: Regulation of putative targets by MrsI in *M. smegmatis*. A) RT-qPCR validation of the regulation of *bfrA* and *hypF* by MrsI after 5 hours of iron starvation. *p<0.05, **p<0.005 (Unpaired T-test). Error bars represent SD of 3 replicates. B) Predicted binding interactions between MrsI and the putative targets identified by RNA-Seq. Nucleotides in bold are the seed region of the sRNA.



Appendix Figure A2.5. Bioinformatic and manual prediction of MrsI direct targets in *M. smegmatis*. TargetRNA2 target prediction using default parameters of MrsI from *M. smegmatis* (blue) and RyhB from *E. coli* (green), and the experimentally identified MrsI targets (red). For experimentally defined targets, the seed region length was changed to 6nt and the 'Single Target' option was used.

Appendix Table A2.1. Genes regulated by MrsI in *M. smegmatis*

gene number	Protein function	Product binds iron	MrsI binding site
MSMEG_0312-0314	Glucose-6-phosphate dehydrogenase	Yes	Yes
MSMEG_2085	NADPH-ferredoxin reductase (*fprA*)	Yes	Yes
MSMEG_2267-2268	Unknown	Yes	No
MSMEG_2273-2276	NiFe hydrogenase maturation factors	Yes	Yes
MSMEG_2415-2417	Hemerythrin binding proteins	Yes	Yes
MSMEG_3143	Aconitase (*acnA*)	Yes	Yes
MSMEG_3564	bacterioferritin (*bfrA*)	Yes	Yes
MSMEG_3629	Unknown	Unknown	No
MSMEG_6419-6420	Unknown	Unknown	No
MSMEG_6816	molybdopterin oxidoreductase	Yes	Yes

Appendix Table A2.2: Bacterial strains used in this study

Organism	Strain #	Genotype
M. smegmatis	sEG1	wil-type mc^{2}155
M. smegmatis	sEG64	*mrsI::loxP*
M. smegmatis	sEG73	mc^{2}155 p$P_{UV15\text{-}Tet}$-null
M. smegmatis	sEG74	*mrsI::loxP* p$P_{UV15\text{-}Tet}$-null
M. smegmatis	sEG80	*mrsI::loxP* p$P_{UV15\text{-}Tet}$-*mrsI*
M. smegmatis	sEG110	*mrsI::loxP* p$P_{UV15\text{-}Tet}$-*mrsI* L5::P_{bfrA}-*zeoR*
M. smegmatis	sEG244	*mrsI::loxP* p$P_{UV15\text{-}Tet}$-*mrsI* L5::P_{bfrA}-*zeoR*(C27T)
M. smegmatis	sEG245	*mrsI::loxP* p$P_{UV15\text{-}Tet}$-*mrsI*(G41A) P_{bfrA}-*zeoR*
M. smegmatis	sEG246	*mrsI::loxP* p$P_{UV15\text{-}Tet}$-*mrsI*(G41A) P_{bfrA}-*zeoR*(C27T)
M. smegmatis	sEG257	mc^{2}155 L5::P_{mrsI}-luciferase
M. tuberculosis	tEG1	Wild type H37Rv
M. tuberculosis	tEG16	*L5::dcas9* sgRNA-*mrsI*
M. tuberculosis	TB965	*L5::dcas9* sgRNA-null
Escherichia coli		BL21- Gold(DE3) E. coli B F- ompT hsdS(r8 - m8 -) dcm+ Tetr gal λ(DE Hte

Appendix Table A2.3: Plasmids used for this study

Plasmid name	Description
pEG66	p$P_{UV15\text{-}Tet}$-null (Hyg)
pEG76	null::L5(Kan)
pEG79	Δ*mrsI::lox-Hyg-lox*
pEG101	p$P_{UV15\text{-}Tet}$-*mrsI*(Hyg)
pEG125	P_{bfrA}-zeoR::L5(Kan)
pEG164	p$P_{UV15\text{-}Tet}$-*mrsI*(G41A)(Hyg)
pEG168	P_{bfrA}-zeoR(C27T)::L5(Kan)
pEG170	P_{mrsI}-luciferase::L5(Kan)
pTB-Sth#1_P1A2	$P_{UV15\text{-}Tet}$*dcas9* sgRNA-*mrsI*::L5(Kan)
pJR965	$P_{UV15\text{-}Tet}$*dcas9* sgRNA-null::L5(Kan)

Appendix Table A2.4: Oligos used for this study

Oligo name	Sequence	Description
EG236	gactacaccaagggctacaag	RT-qPCR for *sigA*
EG237	ttgatcacctcgaccatgtg	RT-qPCR for *sigA*
EG493	AGCAATTGACGAGCGAACT	RT-qPCR for MSMEG_3564 *(bfrA)*
EG494	ATTCGGCACGCGTATGTT	RT-qPCR for MSMEG_3564 *(bfrA)*
EG627	ccttttgcgtttaatactgtttTTCTCACACGGCCGGTCGG	*mrsI* deletion plasmid construction
EG628	tatggcgcgcGCGGGTCCCGCGTCAAGC	*mrsI* deletion plasmid construction
EG629	cgggacccgcGCGCGCCATAACTTCGTA	*mrsI* deletion plasmid construction
EG630	gaagccttgcAGTGGATCCATAACTTCGTATAATG	*mrsI* deletion plasmid construction
EG631	tggatccactGCAAGGCTTCCCTAATTTAGC	*mrsI* deletion plasmid construction
EG632	agcgagacgaaatacgcgatCGGGTGATTCCGCTGTTG	*mrsI* deletion plasmid construction
EG634	ttctcacacggccggtcggc	*mrsI* deletion cassette amplification
EG635	cgggtgattccgctgttggt	*mrsI* deletion cassette amplification
EG686	tagataggctctgcacAAGTATCGAGCCAACGGAC	plasmid construction
EG687	agccgtgaacgacacAAAAGATTCGGGCGGGTC	plasmid construction
EG785	ccttttgcgtttaatactgtttTTCACCGGGCCTTTCCGC	plasmid construction
EG788	gctagagccgtgaacgaccaCTAGTCCTGCTCCTCGGC	plasmid construction
EG789	acttggccatGGTCACTCCTAGACACCTTGAG	plasmid construction
EG790	aggagtgaccATGGCCAAGTTGACCAGTG	plasmid construction
EG795	ACTTCGTGGAGGACGACTT	RT-qPCR for *zeoR*
EG796	CAGGCCAGGGTGTTGTC	RT-qPCR for *zeoR*
EG929	cgggggtctctgtagccctcaggac	site directed mutagenesis
EG930	gtcctgagggctacagagacccccg	site directed mutagenesis
EG935	ccttgagggctaatcaagtggtgctgtttgccg	site directed mutagenesis
EG936	cggcaaacagcaccacttgattagccctcaagg	site directed mutagenesis
EG995	ccttttgcgtttaatactgtttCACCGCGATGTGGCACCT	plasmid construction
EG996	ggctgccgtgCGATAATAGGCAAGGCTTCCCTAATTTAGC	plasmid construction
EG997	cctattatcgCACGGCAGCCCGGTGAAG	plasmid construction
EG998	gctagagccgtgaacgaccaTTACTGCTCGTTCTTCAGCACGC	plasmid construction
EG1057	TAATACGACTCACTATAGGGtgcaccaaggcacggg	Northern probe template construction
EG1058	acaacgaccgcggcc	Northern probe template construction
EG1097	TAATACGACTCACTATAGGGTACTTTTCCACC	
EG1098	ttacggcggccacagc	
Sth#1_2T	GGGAaccgggggtcactgcagccc	plasmid construction
Sth#1_2B	AAACgggctgcagtgacccccggt	plasmid construction

Supplemental Materials and Methods for Chapter 3

Plasmid construction

A list of all plasmids used in this study is located in *SI Appendix, Table S3*. Plasmid pEG79 for creating the *mrsI* deletion in *M. smegmatis* was generated by amplifying the lox-Hyg-lox cassette and the regions upstream and downstream of *M. smegmatis mrsI* using oligos EG629/630, EG627/628, and EG631/632 (Table S4), respectively. These 3 amplicons were then assembled into pEG66 linearized with PmlI using Gibson assembly master mix (NEB). Plasmid pEG101 for inducible *M. smegmatis mrsI* overexpression was generated by amplifying the *mrsI* gene from *M. smegmatis* genomic DNA using oligos EG686/687 and Gibson assembly was used to place the amplicon into pEG66 linearized with PmlI. Plasmid pEG125 for measuring *bfrA-zeoR* repression by MrsI was created by amplifying the *bfrA* promoter and 5' UTR with oligos EG785/789 and the *zeoR* gene with oligos EG788/790 and assembling into pEG76 linearized with PmeI and NdeI with Gibson assembly. Plasmid pEG164 was generated by site directed mutagenesis on plasmid pEG101 using oligos EG929/930. Plasmid pEG168 was created using site directed mutagenesis on plasmid pEG125 with oligos EG935/936. Plasmid pEG170, for measuring MrsI induction in *M. smegmatis*, was created by amplifying the promoter of *mrsI* from *M. smegmatis* using oligos EG995/996 and the luciferase gene using oligos EG997/998. The amplicons were then assembled into pEG76 linearized with PmeI and NdeI using Gibson assembly. Plasmid pTB-Sth#1_P1A2 for inducible CRISPRi knockdown of *mrsI* in *M. tuberculosis* was created by annealing oligos Sth#1_2T/ Sth#1_2B and annealing with pJR965 linearized with BsmBI (1).

Construction of the *Mycobacterium smegmatis* Δ*mrsI* and Δ*mrsI::mrsI* strains

To delete the *mrsI* gene from *M. smegmatis*, the regions flanking *mrsI* were amplified as described above to create pEG79. The deletion cassette was then amplified using oligos EG634/635 and transformed into the *M. smegmatis* recombineering strain (2). The deletion replaced the sequence of *mrsI* upstream of the rho-independent terminator with a *lox-Hyg-lox* cassette, and deletion mutants were selected for on Hygromycin B. The lox-Hyg-lox cassette was then removed using Cre-recombinase to create a clean deletion strain.

The complemented strain was created by transforming pEG101 (above) into the deletion strain. pEG101 contains the *mrsI* gene downstream of an ATc-inducible variant of the strong UV15 promoter, a strong mycobacterial expression promoter.

***in vitro* stress of *M. tuberculosis* for sRNA-Seq**

Wild type *M. tuberculosis* was expanded in 7H9 medium (Difco) supplemented with 2% v/v glycerol, 0.05% v/v tween-80, and 10% v/v oleic acid albumin dextrose catalase (OADC, Sigma Aldrich) (*M. tuberculosis*) to mid-log phase (OD_{600} =0.5-1.0). Prior to exposure cells were pelleted and resuspended in 7H9 media supplemented with 10% vol/vol albumin dextrose NaCl (ADN), glycerol, and tyloxapol supplemented with 1mM *tert*-butyl Hydroperoxide (tBHP)(oxidative stress), 0.05% v/v SDS (membrane stress), or adjusted to pH 4.5 (acid). For carbon starvation the pelleted cells were resuspended in PBS supplemented with 0.05% tyloxapol. Iron starvation was performed as described previously (3). Briefly, cells were grown to OD_{600}=0.5-1.0, then washed once with an equal volume of low iron medium and diluted to OD_{600}=0.1 in the same media. Cells were then grown to ~OD_{600}=1.0, diluted back to OD_{600}=0.1

in low iron media, and grown to OD_{600}=0.2-0.4 before adding 50µg/mL of the iron chelator deferoxamine (DFO). Cells were exposed to the stresses for 4 hours (oxidative stress, SDS stress, acid stress) or 24 hours (PBS starvation, DFO exposure during iron starvation) before RNA harvest. No-stress control conditions (7H9, iron-supplemented minimal media) were harvested after 24 hours.

MrsI regulon identification

For transcriptomics in *M. smegmatis* to identify the MrsI regulon, wild type::empty, *M. smegmatis* Δ*mrsI*::empty, and complemented strains were grown in 7H9 medium with ADC, tween-80, and glycerol supplemented with Hygromycin B (100µg/mL) to mid log phase (OD_{600}=0.5-1.0), then pelleted, washed once with low iron media, and resuspended in low iron media supplemented with hygromycin B and ATc to OD_{600}=0.1. Cells were grown in low iron media for 6 hours before harvesting RNA for transcriptomics. For RT-qPCR validation of *bfrA*, cells were grown as described above, with a RNA being extracted at the indicated timepoints.

For transcriptomics in Mtb to identify the MrsI regulon, the two biological replicates of the empty guide control strain and MrsI knockdown strain were grown in 7H9+OADC supplemented with glycerol, kanamycin, and tween-80 to mid-log phase (OD_{600}=0.5-1). For iron deprivation transcriptomics, cells were pelleted, washed one time with iron starvation medium, then resuspended in iron starvation medium to OD_{600}=0.1. The cultures for both strains were then split into plus and minus CRISPRi induction cultures, with 200ng/mL ATc added to the plus induction cultures. Cells were then grown to late log phase (OD_{600}=1.0) and diluted back to

OD_{600}=0.1 in low iron media supplemented with 50µg/mL DFO, and ATc in the plus induction cultures. Cells were grown for 24 hours before harvesting RNA for transcriptomics. For oxidative stress and SDS stress transcriptomics, the empty guide control strain and MrsI knockdown strain were grown to early log phase (OD_{600}=0.2) and then were split into plus and minus induction cultures, with 200µg/mL ATc being added to the plus induction cultures. After 24 hours of pre-induction, cultures were pelleted and resuspended in an equal volume of 7H9 media supplemented with ADN, tyloxapol, and tBHP or SDS at the same concentrations used for sRNA discovery and expression profiling. ATc was also added to the plus induction cultures. Cultures were exposed to oxidative and SDS stress for 4 hours before harvesting RNA.

RNA-Seq data analysis

Reads were aligned using the BWA-MEM algorithm (4). Samtools was used to sort and split the reads into plus and minus strands, and the bedtools genomecov tool was used to generate per-base read depth files (5, 6). sRNAs were identified using the BS_Finder default parameters, as described previously (7). Feature counts were calculated using a custom python script and differential expression analysis was performed using DESeq2 (8). Total RNA-Seq data analysis was performed identically, with the exception that the NC_000962 (Mtb) or NC_008596 genome feature files (gff) were used for feature calling and differential expression analysis.

For Mtb differential expression analysis during MrsI regulon identification, transcriptomics was performed on both an empty guide control strain and a MrsI knockdown strain, with and without induction of CRISPRi. DESeq2 was performed on each strain with and without

induction. To determine the effects of MrsI knockdown without nonspecific effects of CRISPRi induction, normalized fold-changes were determined by subtracting the fold change of the empty guide control from the fold change of the knockdown strain. Normalized p-values were calculated using the geometric mean of the two individual adjusted p-values. To define a gene as differentially expressed in both *M. smegmatis* and *M. tuberculosis*, a 1.5-fold change cutoff was used in addition to an adjusted p-value cutoff of $p<0.05$.

Bacterial phenotyping experiments

For growth curves in *M. smegmatis*, cells were grown to mid log phase in minimal medium supplemented with 50µM $FeCl_3$ and pelleted for 10min at 4000 rpm. Cell pellets were washed with an equal volume of low iron medium, pelleted and resuspended in the same medium to OD_{600}=0.01 in low iron medium with or without 50µM $FeCl_3$, with hygromycin and ATc. Growth was monitored using an automated plate reader (Growth Curves USA). In *M. tuberculosis*, growth curves in low iron were performed by growing cells to mid log phase in 7H9 medium, pelleting cells and washing one time with low iron medium. Cells were resuspended in low iron medium supplemented with 5µg/mL DFO to an OD of 0.1, with and without 200ng/mL ATc. For all *M. tuberculosis* phenotyping, growth of three biological replicates was monitored by OD_{600} measurements.

Global Proteome Analysis

Three replicates of Δ*mrsI (ΔmrsI::empty)*, three replicates of complemented (Δ*mrsI::mrsI*) and three replicates of wild type (*wt::empty*) were detected by quantitative LC-MS/MS methods. Cells

were grown to mid log phase (OD_{600} =0.5-1.0) in iron-supplemented minimal media with hygromycin B (50µg/mL). Cells were then pelleted, washed once with low iron minimal media supplemented with hygromycin B (50µg/mL) and ATc (100ng/mL), and resuspended in the same media. Cells were then grown in low iron media for 10 hours before pelleting and harvesting protein. All samples were lysed in 6M urea with protease inhibitors and clarified. The denatured protein was reduced, and alkylated, and double digested with both Lys-C and Trypsin overnight. Equivalent amount of tryptic peptides from each sample were labeled with TMT-10 reagent (Thermo Fisher Scientific) and the individual label incorporation was checked via LC-MS/MS. All samples, as expected, had greater than 95% label incorporation. The labeled digests were combined and basic reverse phase (bRP) fractionated into 24 fractions to decrease sample complexity and increase the dynamic range of detection. This global proteome detection and quantification method was developed at the Broad Institute(9). The proteome data was acquired on a Q-Exactive+ mass spectrometer (Thermo Fisher Scientific). Peptide spectrum matching and protein identification was performed using Spectrum Mill (Agilent). Peptide identification false discovery rates (FDRs) were calculated at 3 different levels: spectrum, distinct peptide, and distinct protein. Peptide FDRs were calculated in SM using essentially the same pseudo-reversal strategy evaluated by Elias and Gygi (10) and shown to perform the same as library concatenation. A false distinct protein ID occurs when all the distinct peptides that group together to constitute a distinct protein have a deltaForwardReverseScore ≤ 0. We adjust settings to provide peptide FDR of 1-2% and protein FDR of 0-1%. SM also carries out sophisticated protein grouping using the methods previously described (11). Only proteins with >2 peptides and at least 2 TMT ratios in each replicate are counted as being identified and quantified. From

the initial protein report generated by Spectrum Mill, the report was first filtered by species ensuring that each protein identified was from the organism Mycobacterium smegmatis str. MC2 155. The entries were then filtered by unique peptides ensuring that each of the proteins had >= 2 unique peptides positively identified. A total of **4704** proteins were confidently identified with >=2 unique peptides. Next, the median reporter ion intensity ratios were median-normalized to ensure that the distributions were centered on zero. The three normalized median reporter ion intensity ratios that corresponded to each of the biological replicates were processed using a 1-sample moderated T-test to generate the differential list of proteins.

TargetRNA2 prediction of sRNA targets

For agnostic prediction of MrsI targets, we ran *M. smegmatis* MrsI in TargetRNA2 using default parameters against the NC_008596 genome (12). For forced interactions between MrsI and experimentally determined targets, we changed the seed region length parameter to 6nt and used the 'Single Target' option.

For CopraRNA, the MrsI alleles from *M. tuberculosis*, *M. smegmatis*, and *M. bovis* were used as input.

Oxidative stress pre-exposure experiments and Nanostring gene expression analysis

25ng of RNA was used as input for Nanostring nCounter assays using a SPRINT profiler (Nanostring Technologies) using custom designed probes. Data was analyzed using nSolver version 4 by normalizing raw counts to internal positive controls and three housekeeping genes (Rv1568, Rv1538c, Rv1915). Furthermore, the effects of ATc were accounted for by using a non-

targeting sgRNA control strain (TB965, Table S2). The counts of the *mrsI*-knockdown strain with ATc were normalized using the effects of ATc on the counts for each gene in the non-targeting control strain.

Northern Blots

Northern blot analysis was performed using DIG labeled probes. Probe templates for MrsI and 5S RNA for *M. tuberculosis* were created by PCR amplifying *M. tuberculosis* H37Rv gDNA with PCR primers EG1057/8 and EG1097/8, respectively. DIG-labeled probes were synthesized using the DIG RNA Labeling Kit (Sigma-Aldrich). RNA samples were run on Novex TBE-Urea 6% gels (Thermo Fisher Scientific) and transferred to nitrocellulose membranes using the Invitrogen iBlot DNA Transfer Stacks (Thermo Fisher Scientific). Membranes were prehybridized with ULTRAhyb buffer (Thermo Fisher Scientific) for 2 hours at 65˚C before addition of probe and overnight incubation at 65˚C. Membranes were then washed with 2x SSC, 0.1% SDS solution at room temperature followed by 0.5x SSC, 0.1% SDS solution at 65˚C. Membranes were then washed and blocked with the DIG Wash and Block Buffer Set (Roche) and bound with anti-DIG-AP antibody (Sigma-Aldrich), according to manufacturer's instructions. Blots were developed using CDP-Star (Roche).

SI References

1. Rock JM, et al. (2017) Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. doi:10.1038/nmicrobiol.2016.274.

2. van Kessel JC, Hatfull GF Recombineering in Mycobacterium tuberculosis. doi:10.1038/NMETH996.

3. Kurthkoti K, et al. (2017) The Capacity of *Mycobacterium tuberculosis* To Survive Iron Starvation Might Enable It To Persist in Iron-Deprived Microenvironments of Human Granulomas. *MBio* 8(4):e01092-17.

4. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 0(0):1–3.

5. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma Appl NOTE* 26(6):841–84210.

6. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinforma Appl NOTE* 25(1610):2078–2079.

7. Dejesus MA, et al. (2017) Comprehensive essentiality analysis of the Mycobacterium tuberculosis genome via saturating transposon mutagenesis. *MBio* 8(1). doi:10.1128/mBio.02133-16.

8. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15. doi:10.1186/s13059-014-0550-8.

9. Mertins P, et al. (2013) Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nat Methods* 10(7):634–637.

10. Elias JE, Gygi SP (2010) Target-Decoy Search Strategy for Mass Spectrometry-Based Proteomics. *Methods Mol Biol* (SUPP.60):1–16.

11. Nesvizhskii AI, Aebersold R (2005) Interpretation of Shotgun Proteomic Data. *Mol Cell Proteomics* 4(10):1419–1440.

12. Kery MB, Feldman M, Livny J, Tjaden B (2014) TargetRNA2: identifying targets of small regulatory RNAs in bacteria. *Nucleic Acids Res* 42(Web Server issue):W124-9.