March 2016

The Effects of Post-Transcriptional Processing on mRNA Stability in M. smegmatis

John Henry de Rivera
Worcester Polytechnic Institute

The Effects of Post-Transcriptional Processing on mRNA Stability in *M. smegmatis*

A Major Qualifying Project

Submitted to the Faculty of

Worcester Polytechnic Institute

In partial fulfillment of the requirements for the

Degree in Bachelor of Science

In

Biochemistry

By

John de Rivera

Date: 3/25/2016

Project Advisors:

Assistant Professor Scarlet Shell, Advisor

Professor José Argüello, Co-Advisor

Abstract

Mycobacterium tuberculosis kills over one million people each year globally and its mechanisms of virulence are still not fully understood. The heterodimeric EsxB and EsxA protein complex is thought to play a role in its pathogenicity. Following transcription of the genes encoding these two proteins, a processing event cleaves the portion encoding *esxB* and *esxA* from the upstream portion of the transcript. The processed esxB-esxA transcript has a 70 nucleotide, 5' untranslated region (5' UTR) that we hypothesized might increase the stability of the transcript through secondary structure folding resulting in two hairpin loops. We constructed novel reporter plasmids with deletions of the hairpin loops to study their role in transcript stability. These plasmids were transformed into *Mycobacterium smegmatis* and analyzed using fluorescent microscopy, flow cytometry, and quantitative PCR. Certain 5' UTR deletions resulted in increased protein abundance, whereas others resulted in decreased abundance. Quantitative PCR showed an increase of transcript abundance when the first and both hairpins were deleted and a decrease when just the second hairpin was deleted. It was concluded that the hairpins may not be necessary for transcript stability. Further experimentation is required for a complete understanding of the relationship between 5' UTR sequence, secondary structure, transcript stability, and translation.

Acknowledgements

The author would like to thank graduate students Paula Bertuso, Ying Zhou, and Diego Vargas Blanco for their grateful assistance and indispensable advice.

Introduction

Tuberculosis killed 1.5 million people and developed in an additional 9 million in 2013 (World Health Organization, 2014). However, challenges such as antibiotic resistance and the HIV-1 epidemic have retarded progress in eradicating tuberculosis. Despite advances in technology and medicine, the number of new cases continues to grow (Dye & Williams, 2010). Research into the mechanisms by which tuberculosis causes disease may lead to new and novel treatments that can reverse its current trend.

Tuberculosis is caused by the pathogenic bacteria *Mycobacterium tuberculosis*, an obligate aerobe that infects the pulmonary tracts of mammals (Champoux, Neidhardt, Drew, & Plorde, 2004). The genome of *M. tuberculosis* was sequenced in 1998 (Cole et al., 1998). Mycobacteria contain a type VII secretory system (T7SS) that is responsible for secreting the proteins EsxA (ESAT-6) and EsxB (CFP-10) among others (Abdallah et al., 2007). There are five known T7SS gene clusters that encode the five secretion systems termed ESX-1 through ESX-5. The first of these clusters ESX-1, is the most widely studied due to its pathological importance (Gey van Pittius et al., 2001). The ESX-3 gene cluster is vital for the transport of iron (Siegrist et al., 2009). ESX-5 secretes effector proteins that are involved in *M. tuberculosis* virulence (Abdallah et al., 2008). ESX-1 secretion system substrates include CFP-10 (culture filtrate protein 10; recently renamed to EsxB), ESAT-6 (early secreted antigen target 6, recently renamed esxA), EspA, EspB, and EspC (Guinn et al., 2004). Both of these proteins are required for full virulence of several pathogens including *Staphylococcus aureus* and *Mycobacterium tuberculosis* (Stanley, Raghavan, Hwang, & Cox, 2003) (Burts, Williams, DeBord, & Missiakas, 2005). EsxA and EsxB have been shown to form 1:1 heterodimeric complexes (Fortune et al., 2005). EsxB and EsxA are

encoded by the genes *esxB* and *esxA* respectively (Berthet, Ramuseen, Rosenkrands, Andersen, & Gicquel, 1998) (Sorensen, Nagai, Houen, Andersen, & Andersen, 1995). Mutations that affected the secretion or synthesis of EsxB and EsxA in murine models were shown to lessen the virulence of *M. tuberculosis* (Tan, Lee, Alexander, & Grinstein, 2006). It was found that these two proteins could potentially be responsible for inhibiting the action of phagosomes (Tan, Lee, Alexander, & Grinstein, 2006) and was suggested that the action of EsxA and EsxB are to facilitate the leakage of *M. tuberculosis* from a phagosome to the cytosol although this has been disputed (Chen et al., 2013) (Abdhalla et al., 2007).

The genes *esxA* and *esxB* are transcribed together in a single, shared mRNA transcript along with the genes PE35 and PPE68 shown below in figure 1a (Shell et al, submitted manuscript).

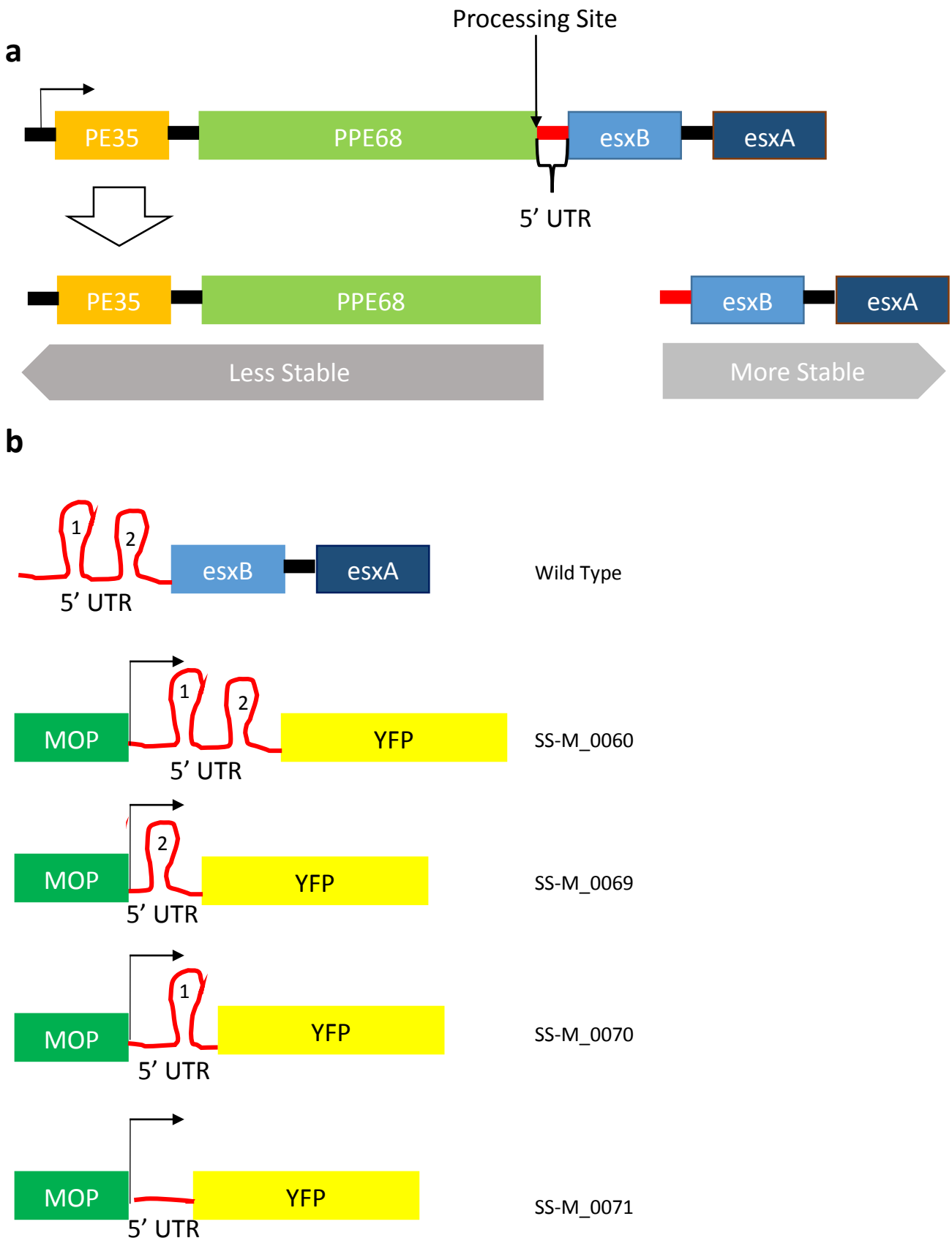


Figure 1. (a) The genes PE35, PPE68, esxB, and esxA are all transcribed onto the same mRNA transcript. Post transcriptional processing cleaves the strand in half resulting in two fragments of differing stability. (b) Schematic diagram of the strains SS-M 0060, SS-M 0069, SS-M 0070, and SS-M 0071.

There has been recent research into the *esxB* gene. *esxB* is a 303 bp gene located downstream of the *PE35* and *PPE68* genes and upstream of *esxA*. It has been observed that there is a processing site between *PPE68* and *esxB* that cleaves the mRNA strand resulting in a 70nt untranslated region upstream of *esxB* that will hereafter be referred to as the esxB 5' UTR (Shell et al, submitted manuscript). The function of this region is currently unknown and is the subject of this research.

Processing of mRNA is necessary in some prokaryotes because of the operon organization of their genome resulting in multiple ORFs within polycistronic transcripts. Also, unlike eukaryotes, transcription and translation are coupled by time and space resulting in relatively short half-lives in the span of minutes compared to hours in eukaryotes (Silva et al., 2011). There are several secondary structures and sequences of mRNA that can stabilize or destabilize the molecule. Some of the first studies into transcript stability showed that inverted repeats in repetitive extragenic palindromic sequences increased the stability of RNA (Newbury, Smith, & Higgins, 1987). New studies have found even more processes that prokaryotes utilize to regulate RNA stability. In 2009, Repoila and Darfeuille showed that ncRNA (non-coding) can base pair to mRNA, affecting its processing and stability. In 2011, Lasa et al suggested that genome wide antisense transcription in bacteria plays a role in RNA stability. One of the most prevalent conformations is the formation of a stem-loop. The presence of a stem-loop at the 5' end of an mRNA strand has been shown to inhibit the binding of exoribonucleases RNase J and PNPase (Condon, 2003).

It has been observed in *Bacillus subtilis* that mRNA cleavage resulted in stabilizing loops in the secondary structure of certain transcripts (Ludwig et al., 2001). One example is of the

glycolytic genes *cggR* and *gapA*. Both genes are transcribed into the same mRNA transcript but it was observed that the concentration of GapA protein was in 100-fold excess when compared to the concentration of the CggR protein (Meinke, Blencke, Ludwig, & Stulke, 2003). This was attributed to an endonucleolytic cleavage event by Rny that separated the cggR and gapA transcripts. The cggR mRNA is rapidly degraded (Half-life <30 sec) whereas the gapA mRNA is quite stable (half-life > 3 min) (Ludwig et al., 2001). The stability of the transcript was attributed to a stem-loop structure at the 5' end of the gapA mRNA (Meinke, Blencke, Ludwig, & Stulke, 2003). By processing mRNA post-transcriptionally, bacteria can regulate the concentration of proteins through differential mRNA stabilities.

Also in *Bacillus subtilis*, stem-loop structures help regulate the gene *aprE*, which encodes the extracellular, proteolytic enzyme subtilisin. *aprE* mRNA has a half-life of 25 minutes normally and is predicted to contain a hairpin-loop on the 5' end of the transcript. Mutations that resulted in the deletion or alteration of the 5' loop resulted in a five-fold decrease in transcript stability (Hambraeus, Persson, & Rutberg, 2000).

These examples show that there are several mechanisms in place to regulate protein production at the level of transcript stability. Because the virulence of tuberculosis has been shown to be strongly affected by the proteins EsxA and EsxB, it is important to determine the factors governing the expression of these proteins. There has been considerable research into determining the functions of these protein as well as the genes that code them; however, little study has been dedicated to the step between transcription and translation, the post-transcriptional processing of the mRNA strand. There have been other studies where processing

has affected the stability the resulting mRNAs but none have been focused on the 5' UTR for the esxB and esxA gene transcript.

This study explores the function of the 5' UTR region as it relates to mRNA strand stability and how that stability affects the resulting protein concentration. We hypothesize that the secondary structure of the 5' UTR increases the stability of the downstream mRNA strand by the formation of two 5' hairpin loops. Utilizing a commonly used related species to *M. tuberculosis,* the non-pathogenic species *M. smegmatis*, a novel reporter plasmid was constructed linking the 5' UTR sequence to a yellow fluorescent protein (YFP) gene in the pMV261-YFP plasmid. Several mutations were induced to the 5' UTR to disrupt the normal secondary structure and subsequent analysis done to determine the effects on transcript abundance and protein levels. The research shown in this experiment shows that the relationship between the 5' UTR and transcript abundance and translation is more complex than previously thought.

Methods

Plasmid Designs and Construction

The pMV261-msYFP plasmid was harvested from *E. coli* strain CT166 using the QIAprep Spin Miniprep Kit and protocol at room temperature and using ddH_2O to elute plasmid DNA from column. PCR was used to construct the desired fragment of DNA that would be inserted into the pMV261-msYFP plasmid to add the MOP promoter and 5' UTR upstream of the YFP gene. The PCR reaction was set up by adding 5µL 5X Q5 Reaction Buffer, 0.5 µL 10 mM each dNTP, 1.25 µL 10 µM forward primer (CTCAGTGCTATCGCTGGCCATTGGATTTCCCCTTATTTCCCTGTGTC), 1.25 µL 10 µM reverse primer (GCTTGACACTTTATGCTTCCGGCTCGTATAATGTGTGGCTTTCCGGTG

CACTCGCC) , 25 ng *M. smegmatis* mc^{2}155 genomic DNA, 0.25 µL Q5 High Fidelity DNA Polymerase, 5 µL 5X Q5 High GC Enhancer, and Nuclease Free Water to 25 µL. Thermocycler reaction steps were as follows: 1 cycle of 98C for 1 min; 30 cycles of 98C for 10 sec, 70C for 20 sec (annealing step), then 72C for 10 sec (extension step); 1 cycle of 72C for 2 min; then held at 6C indefinitely. The PCR products were then purified on standard Agarose gel with 1.5 µL Ethidium Bromide (10 mg/mL) added at 135V. The 129 bp insert DNA fragment was excised from the gel and purified using QIAquick Gel Extraction Kit and protocol. All centrifugation steps were carried out at 13,000 rpm at room temperature. A second PCR was run to add homologous overlapping tails to the product of the first PCR. The PCR reaction was set up and run identically to the first except the template was 25 ng of the purified product from the first PCR and the use of a different reverse primer (TACCAGATCTTTAAATCTAGACCCAGGCTTGACACTTTATGCTTCC GGC). The desired 155 bp fragment was purified and extracted identically as well. A third PCR reaction was run to delete the groEL promoter region from the pMV261-msYFP plasmid. The PCR reaction was set up and run identically to the first except the template was 25 ng of pMV261-msYFP plasmid and the use of a different forward primer (TCTAGATTTAAAGATCT GGTACCGCGGC) and reverse primer (ATGGCCAGCGATAGCACTGAG). Also, the annealing temperature was set to 71C and the extension step was lengthened to 2 min and 5 sec. The 4902 bp product was then gel purified and extracted using the same kits and protocol as the first two PCRs.

Gibson assembly was used to join the product of the second PCR and the product of the third PCR into a circularized plasmid. The Gibson Assembly Kit and Protocol were used in this process. 90 ng of vector DNA and 14.23 ng of insert DNA was added to 5 µL Gibson Assembly I

Master Mix (2x) along with ddH_2O to make 10 µL. The reaction was incubated at 50C for 1 hr then stored overnight in -20C.

NEB 5-α Competent *E. coli* cells were used for all procedures involving *E. coli*. The Gibson assembly product was transformed in *E. coli* cells using the NEB 5-α Competent *E. coli* transformation protocol except 2 µL of Gibson assembly mix was used during this transformation whereas 1 µL was used for all other transformations. All *E. coli* was plated on LB plates with 150 µg/mL hygromycin unless otherwise noted. Transformed *E. coli* was plated in 20 µL and 200 µL volume amounts respectively and incubated at 37C overnight. All *E. coli* were grown in liquid media of 5 mL LB Broth and 150 µg/mL hygromycin to an OD between 0.5 and 1.0 unless otherwise noted. *E. coli* colonies were grown in liquid media overnight at 37C. QIAquick DNA extraction kit and protocol was used to harvest the plasmid named pSS139. pSS139 was sent out for sequencing and the desired sequence was confirmed. The *E. coli* strains containing pSS139 was named SS-E_0013.

Four variants of pSS139 were designed to have various segments of the 5' UTR coding region deleted. Plasmid pSS158 had the MOP promoter deleted; plasmid pSS159 had the first 23 nucleotides of the 5' UTR deleted that form the upstream hairpin loop; plasmid pSS160 had the 25^{th} through 44^{th} nucleotides of the 5' UTR that form the downstream hairpin loop deleted; and plasmid pSS161 had the first 44 nucleotides of the 5' UTR that form both hairpins deleted. PCR-mediated deletion procedure was used to delete the desired segments. Originally, the primers for each of the four PCR reactions were designed as reverse compliments. These primers did not produce the desired products shown through gel electrophoresis of the PCR products. New primers were designed so that only the forward primer had a tail homologous to the reverse

primer. The length of the complementary between the two primers was reduced from 42 bp to 18 bp.

Each of the four PCR reactions contained 5µL 5X Q5 Reaction Buffer, 0.5 µL 10 mM dNTP's, 1.25 µL 10 µM forward primer, 1.25 µL 10 µM reverse primer, 25 ng pSS139, 0.25 µL Q5 Hot Start High Fidelity DNA Polymerase, 5 µL 5X Q5 High GC Enhancer, and Nuclease Free Water to 25 µL. pSS158, pSS159, pSS160, and pSS161 used the forward primers GCCGCGGTACCAGATCTTTAAAT CTAGACTTTCCGGTGCACTCGCC, CCGGCTCGTATAATGTGTGGACTCGCCATGGAATTGGTGA, GGTGC ACTCGCCGGAAGAGACACAGGGAAATAAGGGGAAATC, CCGGCTCGTATAATGTGTGGGACACAGGG AAATAAGGGGA, and the reverse primers TAAAGATCTGGTACCGCGGC, CCACACATTAT ACGAGCCGG, TCTTCCGGCGAGTGCACC, and CCAC ACATTATACGAGCCGG respectively. Thermocycler reaction steps for the four mutations were as follows: 1 cycle of 98C for 30 sec; 30 cycles of 98C for 10 sec, 30 sec at annealing temperature, then 72C for 2 min 5 sec (extension step); 1 cycle of 72C for 2 min; then held at 10C indefinitely. pSS158, pSS159, pSS160, and pSS161 had annealing temperatures of 67C, 67C, 65C, and 68C respectively. Gel electrophoresis was used to verify the size of the product and purify it. DNA was extracted from the gel using QIAquick gel extraction kit and protocol. All centrifugation steps were carried out at 13,000 rpm at room temperature. Gibson assembly kit was used to re-circularize the PCR products. 15ng of each extracted DNA was added to 2.5 µL Gibson Assembly I Master Mix (2x) and ddH_2O to make 5 µL. This was incubated at 50C for 1 hr then stored at -20C.

Plasmid pSS158, pSS159, pSS160, and pSS161 were transformed in *E. coli* cells using methods described for the transformation of pSS139. Plasmids were sent out for sequencing and

all deletions were confirmed. *E. coli* strains containing pSS139, pSS158, pSS159, pSS160, and pSS161 were named SS-E_0013, SS-E_0036, SS-E_0037, SS-E_0038, and SS-E_0039 respectively.

Plasmids pMV261-YFP and pMV762 were graciously provided by Scarlet Shell.

M. smegmatis strains and growth conditions

Mycobacterium smegmatis strain mc^2155 was used for all procedures unless specified differently. pMV261-YFP, pMV762, pSS139, pSS158, pSS159, pSS160, and pSS161 were transformed into competent *M. smegmatis* cells using electroporation. All *M. smegmatis* cells were plated on 7H10 agar with 150 µg/mL hygromycin and grown in liquid media in 7H9 agar with 150 µg/mL hygromycin unless specified otherwise. Transformed cells were plated in 20 µL and 200 µL amounts then incubated at 37C for 3 days. Colonies were picked and grown in liquid media overnight at 37C.

Z-stack acquisition and processing

Z- Stacks were acquired from 1 µL of culture on a glass slide with cell cultures grown to an OD between 0.68 and 0.88. The microscope used was an Axio Imager Ziess.Z1 with Apotome. Ten images were taken per slide with the first being the plane just out of focus of the cells and the remaining nine were the next focal planes .15 µM apart. Slides were used only once with a single Z-stack taken per slide. Stacks were taken in the DIC and Fluorescent channels. The brightness of each stack was set at a white level of 250 to correct for automatic picture adjustment. Stacks were exported as AVI files. ImageJ software was used to convert the images into grayscale and measure the mean gray value of each cell. The minimum threshold measurements were 50^2 pixels size and at least a gray value of 15 with a range from 0-256. Cells

partially in the image were excluded. The values for each cells Z-stacks were then summed and the average intensity for each strain calculated.

Flow Cytometry

Flow Cytometry was used to measure each strains fluorescence intensity using an Accuri C6. 5 mL cultures of each strain were grown to within a range of .70 OD and .84 OD. 1 mL of each sample was then filtered through a 5 µM filter. Each sample was tested at 100,000 events. Purified water was run in between samples.

Quantitative PCR

Strains SS-M_0059, SS-M_0060, SS-M_0068, SS-M_0069, SS-M_0070, and SS-M_0071 were grown in liquid media to an OD between 0.5 and 0.8. The RNA was extracted from each strain and the concentration measured using a nanodrop. The RNA concentrations varied between 163.3ng/µl and 505.5ng/µl. Contaminate DNA was degraded by combining 10µg of RNA, 1.25µl Ambion DNase Turbo, 5µl 10x Buffer, and water to a final reaction volume of 50µl. The reaction was incubated for one hour at 37C with agitation. A Qiagen RNeasy Mini Kit was used to clean up the RNA following kit protocol. Concentrations were measured again using the nanodrop and ranged between 159.0ng/µl and 237.3ng/µl.

cDNA synthesis was run in duplicate for each strain with one sample containing reverse transcriptase and the other sample containing none. 0.8µg RNA was added to water for a final volume of 5µl. A master mix of 7µl 100mM Tris pH 7.5 and 7µl 1mg/ml random primers was made and 1µl added to each reaction. This was incubated at 70C for ten minutes then snap cooled in an ice water bath. A master mix of 2µl 5x Protoscript II reaction buffer, 0.5µl 10mM DTT, 0.25µl

murine RNase inhibitor, and 0.25µl Protoscript II reverse transcriptase (water substituted for no-RT control). Master mix was added to each sample for a total reaction volume of 10µl and then incubated at 25C for ten minutes, then incubated overnight at 42C. 5µl 0.5M EDTA and 5µl 1M NaOH was added then the samples were incubated at 65C for 15 minutes. 12.5µl 1M Tris pH 7.5 was added after incubation. Samples were then purified using the Qiagen minElute Kit and SS-DNA concentrations were measured using the nanodrop. The concentrations for the reverse transcriptase samples ranged between 7.5ng/µl and 10.0ng/µl whereas the no-RT controls had concentrations ranging from 4.3ng/µl to 6.8ng/µl.

Three sets of primers to detect YFP transcripts were tested in triplicate against validation primers for sigmaA transcript. A dilution series of 200pg/µl, 40pg/µl, 8pg/µl, and 1.6pg/µl was used as well as a no-RT control and water control with each sample having 2 µl. Plate was run with a 20µl reaction volume, 61C annealing/extension step, and for 40 cycles. Primers SSS833 (GATAGCACTGAGAGCCTGTT) and SSS834 (CTGAACTTGTGGCCGTTTAC) had an efficiency of 0.89 and an R^2 value of 0.97 and were chosen for the qPCR experiment. cDNA from each strain was diluted to 40pg/µl (for a total of 80pg cDNA per reaction). Three reactions were run in triplicate for each strain and a water control: one with the YFP primers, one with the validation primers for sigma A, and one with the YFP primers but for the no-RT control. The plate was run with a 20µl reaction volume, 61C annealing/extension step, and for 40 cycles. Data was then exported to excel files.

Results

Construction of reporters to test the effects of 5' UTR secondary structure on esxB transcript stability

Protein production is a highly regulated pathway that is present in all cells. One regulatory feature that is the topic of this research is the stability of mRNAs. It has been observed in other organisms that secondary structures formed by mRNAs have an effect on their half-life. One of the most commonly observed mRNA stabilizing secondary structure are 5' hairpin loops which have been shown to inhibit the binding of exoribonuceases RNase J and PNPase (Condon, 2003). A similar stabilizing process is seen in *Mycobacterium tuberculosis* where the cleavage of an mRNA strand results in an increase in stability in the downstream sequence and a decrease in stability in the upstream sequence (Shell et al, submitted manuscript. The genes *PE35, PPE68, esxB,* and *esxB* are all transcribed as a polycistronic transcript. The transcript is then cleaved between the *PPE68* gene and the *esxB* gene. Figure 1a depicts the processing of the transcript. The 70 nt long region shown in red in figure 1 will be referred to as the 5' UTR.

To determine the likely secondary structure of the of the 5' UTR, the 70 nucleotide sequence was processed by Sfold, software from the Wadsworth Center of the NYS Department of Health for statistical folding of nucleic acids. Figure 2a shows the predicted secondary structure of the 5' UTR. Two hairpins are predicted to form near the 5' end of the transcript. To determine the function of these two loops, variants of the 5' UTR were created as shown in figures 1b and 2b-d. By deleting the sequences involved in forming the first, second, and both hairpins, the change in mRNA abundance could be observed. The cleavage event is mimicked by starting

transcription at the beginning of the 5’ UTR so that the transcription start site corresponds to the point where cleavage would occur. Figure 1b shows the structure of the transcript normally and the structure of the variants used in this experiment. A fluorescent reporter strategy was used to measure the abundance of the mRNA where a greater fluorescence intensity would correspond to a greater abundance of mRNA. YFP was used as the fluorescent reporter protein.

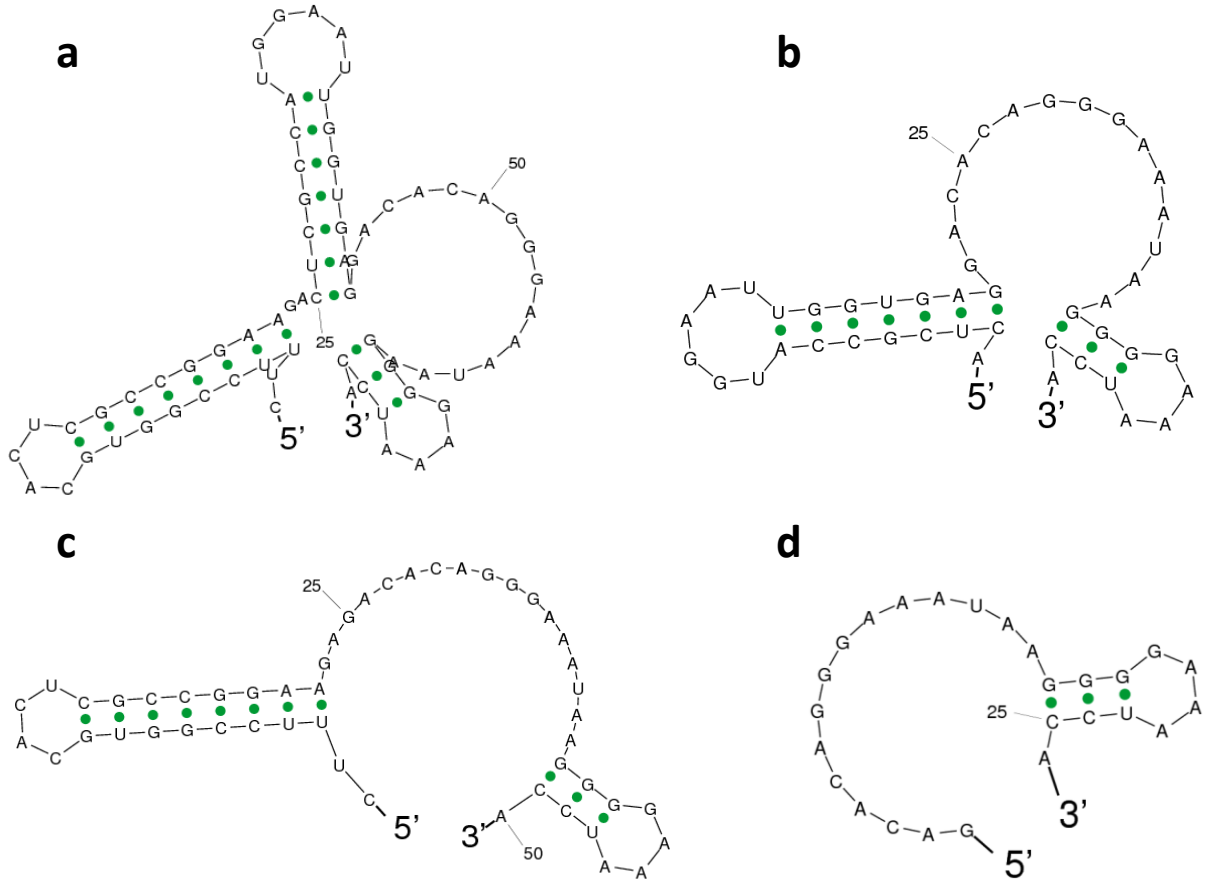


Figure 2. Predicted secondary structure of the 5’UTR calculated using Sfold software. Deletion of the hairpin loops was predicted not to affect ribosome binding or transcription because both the RBS and start codon are downstream of the second hairpin. (a) Wild Type. (b) First hairpin deleted. (c) Second hairpin deleted. (d) Both hairpins deleted.

Constructs were created from the pMV261-msYFP plasmid which contained a groEL promoter that regulated a YFP gene by deleting the promoter and replacing it with a MOP promoter and the 5' UTR. This sequence is depicted in figure 3. PCR was also the primary method used to delete the hairpin sequences from the 5' UTR.

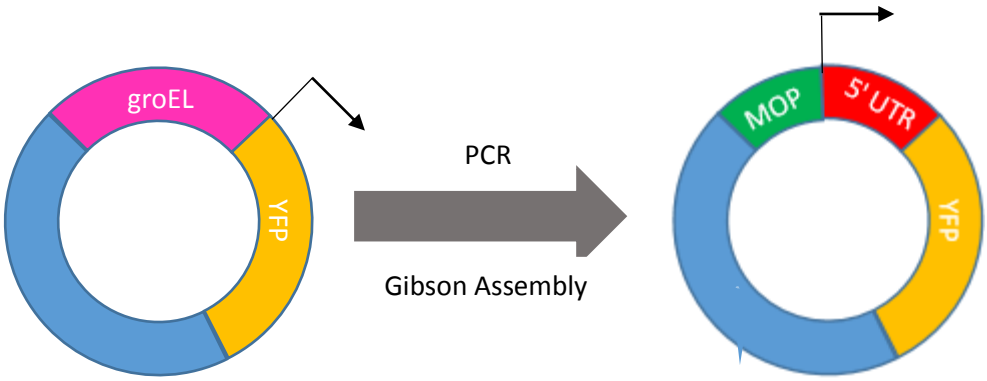


Figure 3. Plasmid pMV261-msYFP was used as the template to construct pSS139 using PCR and Gibson assembly techniques.

Primers were designed for the deleting PCR with overhanging tails so that the target sequence would be deleted and then the linearized plasmid could be rejoined into a circle by Gibson assembly. However, these primers were also noted to be reverse compliments of each other. The PCR did not work in any of the attempted deletions and the resulting gel test revealed large smears in all samples, shown in figure 4a.

The primers were redesigned as shown in figure 5 and when run in PCR, produced the desired deletions shown in figure 4b.

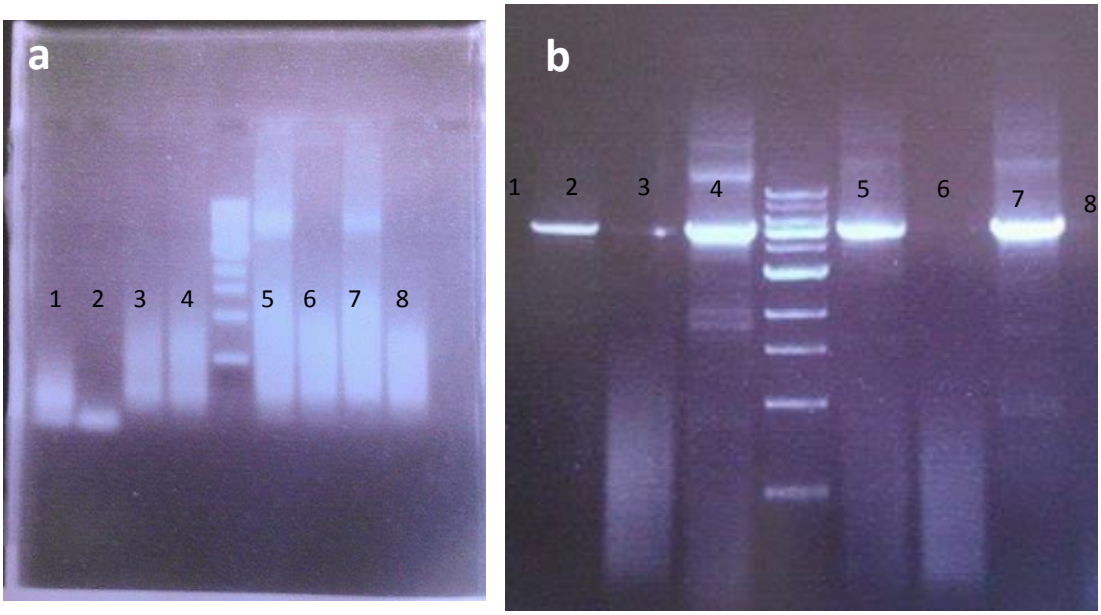


Figure 4. (a) Gel shows the results of the reverse compliment primers. Lanes 1 and 2 are the 1st hairpin deletion and no template control respectively. Lanes 3 and 4 are the 2nd hairpin deletion and no template control respectively. Lanes 5 and 6 are the both hairpin deletions and no template control respectively. Lanes 7 and 8 are the MOP promoter deletions and no template control respectively. (b) Gel shows the results of redesigned primers. Lanes 1 and 2 are the 1st hairpin deletion and no template control respectively. Lanes 3 and 4 are the 2nd hairpin deletion and no template control respectively. Lanes 5 and 6 are the both hairpin deletions and no template control respectively. Lanes 7 and 8 are the MOP promoter deletions and no template control respectively.

We speculate that the original primers dimerized rather than annealing to the template. The redesigned primers would likely still have some dimerization but clearly were not inseparable as shown by figure 4b. We hypothesize that the shortening of one of the primers reduced the

melting temperature of the primer dimer so that they would interact with the template more readily.

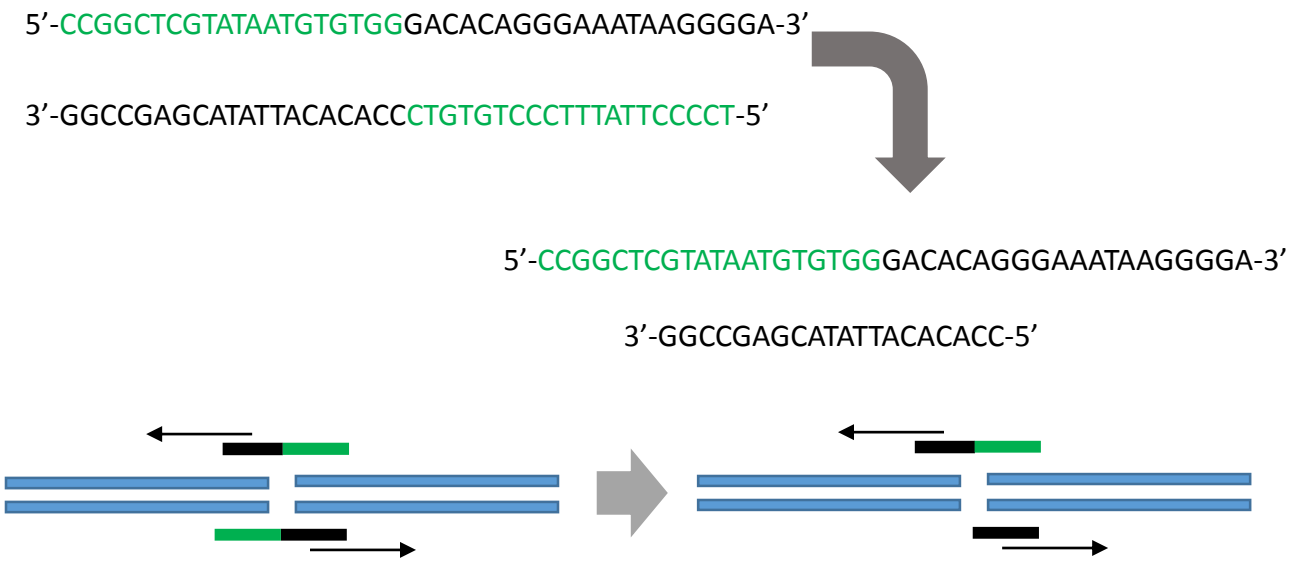


Figure 5. The first primers designed were reverse compliments of each other (green text represents homologous tails). The redesigned primers were still complimentary but one primer was shortened by the removal of the homologous tail.

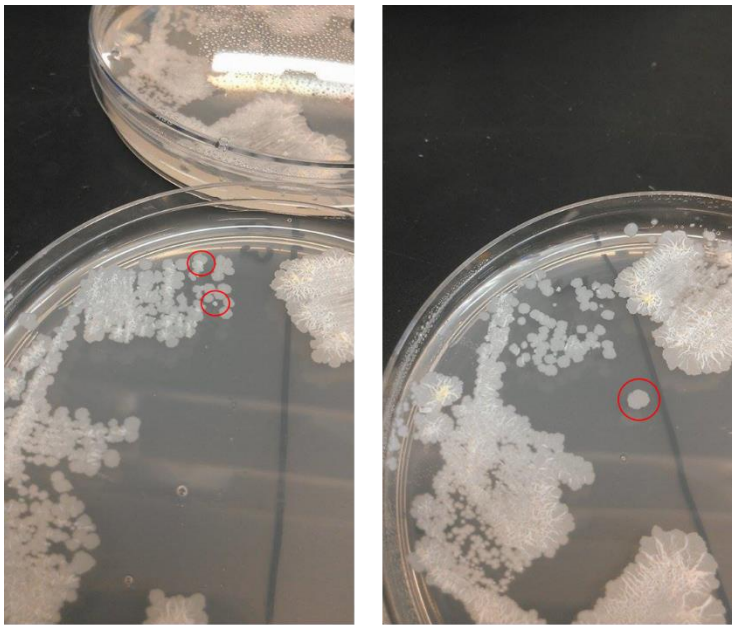
The variant plasmids were then transformed into *Mycobacterium smegmatis* along with several other plasmids used as controls. The names of each strain are shown in table 1 and will be used for the remainder of this paper.

Plasmid	*M. smegmatis* Strain	Description
pMV762	SS-M_0058	No fluorescence gene
pMV261-msYFP	SS-M_0059	groEL Promoter, YFP
pSS139	SS-M_0060	MOP Promoter, 5' UTR, YFP
pSS158	SS-M_0068	Same as pSS139 but with MOP Promoter Deletion
pSS159	SS-M_0069	Same as pSS139 but with first hairpin deletion in 5' UTR
pSS160	SS-M_0070	Same as pSS139 but with second hairpin deletion in 5' UTR
pSS161	SS-M_0071	Same as pSS139 but with both hairpins deletion in 5' UTR

Table 1. Lists the names of the plasmids and *M. smegmatis* strains used in this experiment as well as a description of each.

Colony Morphology

A difference in morphologies was noted in *M. smegmatis* while growing on 7H10media with 150 mg/ml hygromycin. Some colonies grew narrow and tall so that there was a noticeable mound on the plate (similar to a pimple in shape) whereas other colonies grew wide and flat. Both colonies were found on the same plates. There were more flat colonies that tall colonies and the two types are shown in figure 6. Streak tests showed that the difference in colony phenotype was not a result of contamination because when tall colonies were restreaked, they grew into both tall and flat colonies. The same was seen with flat colonies.



Tall Morphology Short Morphology

Figure 6. Shows the tall and flat colony morphology seen in *M. smegmatis*.

Fluorescence Microscopy

Each strain was measured for fluorescence intensity using fluorescence microscopy. Z-stack images of individual cells were taken then analyzed using ImageJ software for their average fluorescence intensity. These data are represented in figure 7. The data suggests that the deletion of the first hairpin loop increases the fluorescence compared to the no variance strain SS-M_0060 whereas a deletion of the second hairpin decreases fluorescence, although these differences were not statistically significant. The deletion of both hairpins has little or no effect. Strain SS-M_0058 showed no measurable fluorescence. Strain SS-M_0068 showed some fluorescent activity despite having the MOP promoter deleted. This was hypothesized to be the result of an unknown upstream promoter affecting the transcription of the YFP gene.

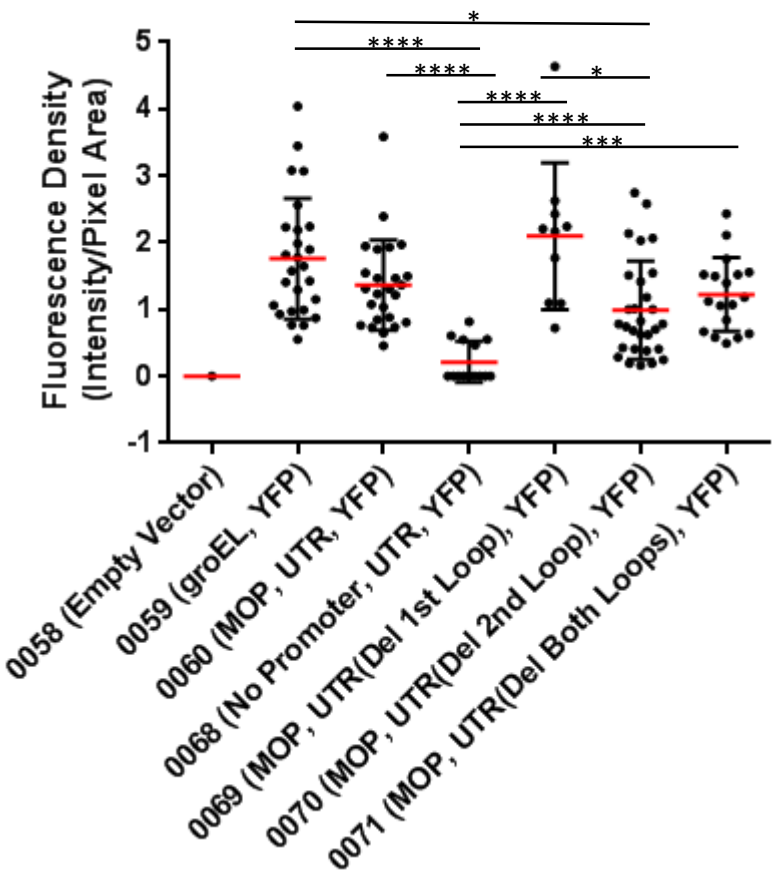


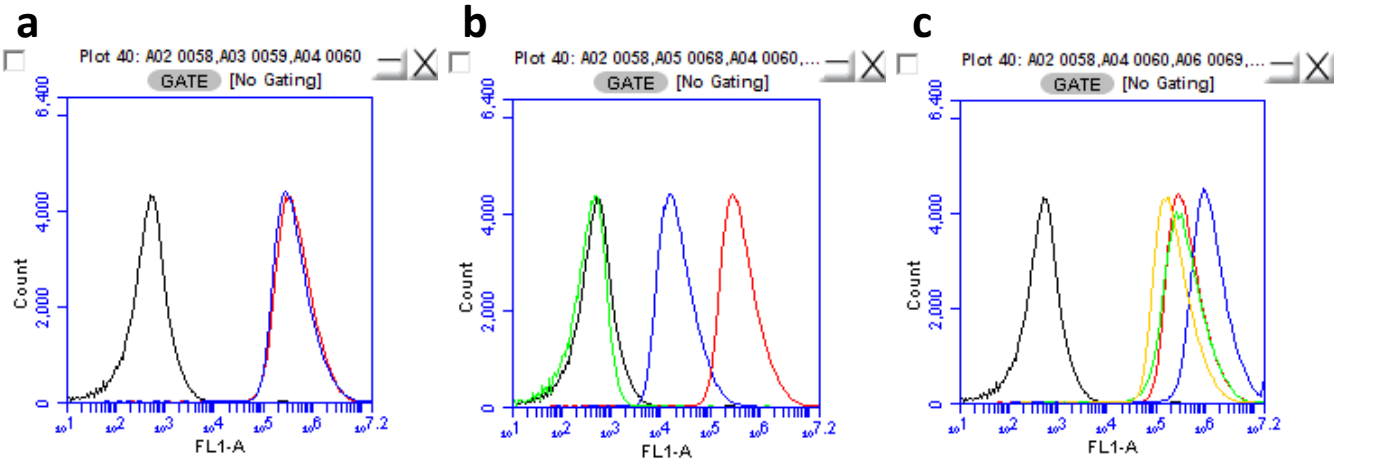
Figure 7. The average fluorescent density of each of the *M. smegmatis* strains is represented by the red line for each strain and the overhead line represent which strains data are significant when compared to each other with the more stars signifying a greater significance (* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$, and **** = $p \leq 0.0001$). Strain pairs without a bar had no significance between data. There is no comparison to SS-M_0058 because it had no measurable fluorescence and therefor no data collected.

Flow Cytometry

Flow cytometry was used to support the microscopy data with a much greater sample size. Figure 8 compares the fluorescence intensity of each of the seven strains. Figure 8a plots strains with no fluorescence against two similar strains each containing YFP but one with the groEL promoter and the other with the MOP promoter. It should be noted that the graphs are plotted logarithmically on the x-axis. The slight fluorescence measured in a strain containing no

fluorescent gene was thought to be autofluorescence. Figure 8b compares the strains from part A with SS-M_0068, the MOP promoter deletion. The fluorescence of SS-M_0068 is midway between strains emitting only autofluorescence and strains with active promoters for YFP. It is thought that there may be an unknown upstream promoter in the plasmid backbone that is influencing the expression of the YFP causing some fluorescence. Figure 8c compared all of the variant strains.

A similar pattern of changes in fluorescence intensity is seen in the flow cytometry data as with the data from fluorescence microscopy. Deletion of the first hairpin shown in blue on figure 8c slightly increased the fluorescence of the strain compared to the non-variant strain SS-M_0060 shown on figure 8c in red. Similarly, deleting the second hairpin decreased fluorescence, shown in yellow on figure 8c, and deleting both loops, shown in green on figure 8c had no noticeable effect.



Black- 0058 (Empty Vector)
Blue- 0059 (groEL, YFP)
Violet- 0060 (MOP, UTR, YFP)

Black- 0058 (Empty Vector)
Blue- 0068 (No Promoter, UTR, YFP)
Red- 0060 (MOP, UTR, YFP)

Black- 0058 (Empty Vector)
Yellow- 0070 (MOP, UTR(Del 2nd Loop), YFP)
Red- 0060 (MOP, UTR, YFP)
Green- 0071 (MOP, UTR(Del Both Loops), YFP)
Blue- 0069 (MOP, UTR(Del 1st Loop), YFP)

Figure 8. Flow cytometry data. Each strain had 100,000 individual cells analyzed for fluorescent intensity, represented on the horizontal axis. (a) Shows comparison between M. smegmatis strains SS-M_0058, SS-M_0059, and SS-M_0060. (b) Shows comparison between strains SS-M_0058, SS-M_0068, and SS-M_0060. (c) Shows comparison between strains SS-M_0058, SS-M_0070, SS-M_0060, SS-M_0071, and SS-M_0069.

Quantitative PCR

Quantitative PCR was used to directly measure the abundance of YFP mRNA in each strain without relying on fluorescence measurements. It was conducted on all strains except SS-M_0058. The levels of YFP mRNA were compared to the sigA housekeeping gene as shown in figure 9. In this experiment, the deletion of both loops substantially increased the abundance of the YFP transcript as compared to the non-variant strain SS-M_0060. This is also seen when the first hairpin is deleted. However, a modest decrease in transcript abundance is seen when the second hairpin is deleted.

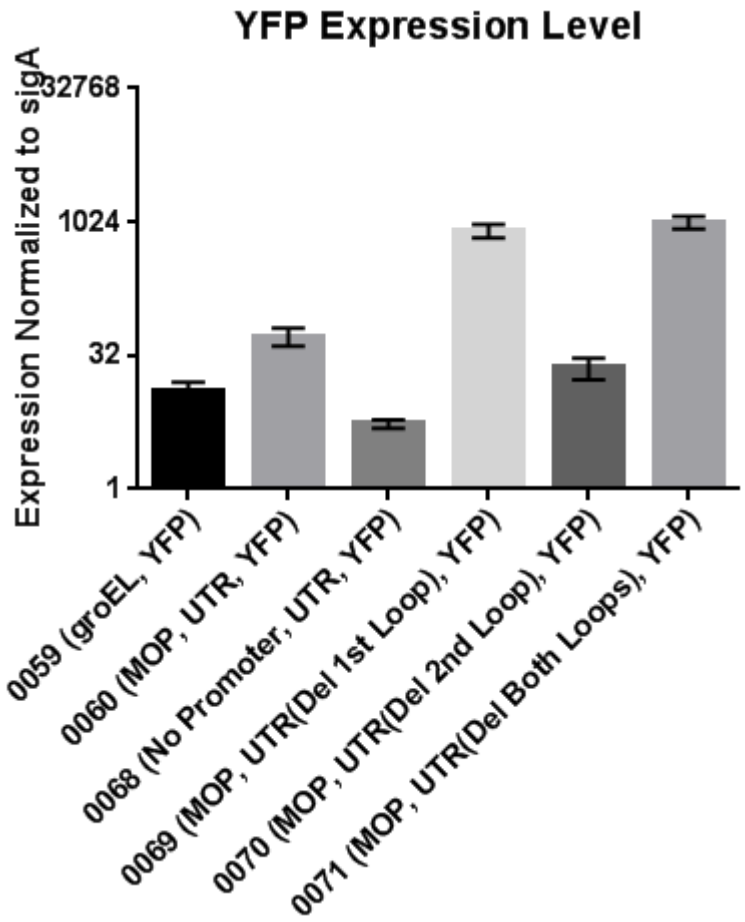


Figure 9. Compares the YFP expression levels of several strains compared to the levels of the sigA gene. Error bars represent the standard deviation of the technical triplicates.

Discussion

The ability for a cell to regulate the production of proteins is key to its survival and ability to adapt to different environments. For bacteria, specifically in this study *M. tuberculosis*, the ability to adapt quickly to a harsh environment such as the inside of a macrophage can be the difference between causing a potentially lethal infection or being destroyed by the body's defense mechanisms. In this study, we investigated the mechanistic basis of the observation that mRNA processing increases the stability of the esxB-esxA transcript in a related species, *M.*

smegmatis. We hypothesized that the formation of two hairpin loops at the 5' end of the strand were responsible for the increase in transcript stability post cleavage. It was found that deleting one or both of the hairpins did change abundance of the reporter protein mRNA it was attached to as well as the amount of reporter protein in the cells. The initial data from fluorescent measuring experiments (fluorescent microscopy and flow cytometry) measured only the abundance of YFP protein in the cell. It was predicted that the ribosome binding site and start codon were not affected and that the abundance of protein, as measured by the fluorescent intensity of the cell, was directly related to the amount of mRNA present as well as its stability. By controlling the transcription of the YFP gene with the MOP promoter, it was assumed that all constructs were transcribed equally. More stable transcripts would last longer in the cell and be able to be translated many times more than less stable transcripts that were degraded very rapidly resulting in a difference in fluorescence intensity. Our third experiment analyzed directly the steady-state abundance of the YFP transcript using quantitative PCR. By directly measuring the mRNA levels, any regulation at the ribosome will have no effect on the data. The difference between the fluorescence microscopy and flow cytometry data with the qPCR data suggests that there is some regulatory step occurring during translation. Additional experimentation is needed to determine the exact source of this inconsistency.

It is difficult to determine the affect an individual hairpin loop has on the overall transcript stability. The data does suggest however that the deletion of the first hairpin increases cell fluorescence and transcript abundance. The deletion of the second hairpin loop generated no significant data. Deletion of both loops increased transcript abundance. It can be concluded that the original hypothesis was false and that the hairpin loops may not be necessary for transcript

stability. Further analysis with qPCR at different time points after stopping transcription with rifampicin is needed to directly measure the stability of the mRNA.

Other experiments needed to fully understand the increase in transcript stability would have to focus on the nucleotide sequence and length of the 5' UTR to exclude or include them as possible features that would affect stability. Also, the relative distance of structural features in the 5' UTR should be examined for any effect of stability. It could be that the distance between hairpins or the length of unfolded nucleotides before the first hairpin also affects stability. Further investigation into the 5' UTR is needed to definitively understand the mechanism by which the stability is increased.

Understanding how *M. tuberculosis* is able regulate the esxB gene may lead to different and effective treatments of the disease. Current medicine is lagging behind the spread of tuberculosis as it continues to spread among the developing world. New treatment is needed to prevent new cases and research into the molecular level of tuberculosis pathology may provide the basis to eradicate the disease entirely.

References

Abdallah, A., Gey van Pittius, N., Cox, J., Vandenbroucke-Grauls, C., Appelmelk, B., & Bitter, W. (2007). Type VII secretion—mycobacteria show the way. Microbiol, 5(11), 883-891. doi:10.1038/nrmicro1773

Abdallah, A., Savage, N., Van Zon, M., Wilson, L., VandenbrouckeGrauls, C., … Bitter, W. (2008). The ESX-5 secretion system of Mycobacterium marinum modulates the macrophage response. *Journal of Immunology, 181*(10), 7166-75. doi:10.4049/jimmunol.181.10.7166

Berthet, F., Ramussen, P., Rosenkrands, I., Andersen, P., & Gicquel, P. (1998). A Mycobacterium tuberculosis operon encoding ESAT=6 and a novel low-molecularmass culture filtrate protein (CFP-10). *Microbiology, 144*, 3195-3203.

Brodin P., Majlessi L., Marsollier L., de Jonge I., Bottai D., Demangel C., … Brosch R. (2006). Dissection of ESAT-6 System 1 of Mycobacterium tuberculosis and Impact on Immunogenicity and Virulence. *Infection and Immunity,* 88-98. doi:10.1128/IAI.74.1.88-98.2006

Burts, M., Williams, W., DeBord, K., & Missiakas, D. (2005). EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of Staphylococcus aureus infections. Proceedings of the National Academy of Sciences of the United States of America, 102(4), 1169-1174.

Champoux, J., Neidhardt, F., Drew, W., & Plorde, J. (2004). *Sherris Medical Microbiology* (4th ed.) (K. Ryan & C. Ray, Eds.). McGraw Hill.

Chen, J., Zhang, M., Rybniker, J., Boy-Rottger, S., Dhar, N., Pojer, F., & Cole, S. (2013). Mycobacterium tuberculosis EspB binds phospholipids and mediates EsxA-independent virulence. *Molecular Microbiology, 89*(6), 1154-166. doi:10.1111/mmi.12336

Cole T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., … Barrell G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393(6685), 537–544. doi:10.1038/31159

Condon, C. (2003). RNA Processing and Degradation in Bacillus subtilis. *Microbiol Mol Biol Rev, 62*(2), 157-174. doi:10.1128/MMBR.67.2.157-174.2003

Dye, C., & Williams, B. (2010). The population dynamics and control of tuberculosis. *Science, 328*(5980), 856-861. doi:10.1126/science.1185449

Fortune, S., Jaeger, A., Sarracino, D., Chase, M., Sassetti, C., … Rubin, E. (2005). Mutually dependent secretion of proteins required for mycobacterial virulence. *Proceedings of the National Academy of Sciences of the United States of America, 102*(30), 10676-81. doi:10.1073/pnas.0504922102

Gey van Pittius, N., Gamieldien, J., Hide, W., Brown, G., Siezen, R., & Beyers, A. (2001). The ESAT-6 gene cluster of Mycobacterium tuberculosis and other high G C Gram-positive bacteria. *Genome Biol, 2*(10). doi:10.1186/gb-2001-2-10-research0044

Guinn, K., Hickey, M., Mathur, S., Zakel, K., Grotzke, J., Lewinsohn, D., … Sherman, D. (2004). Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of Mycobacterium tuberculosis. Mol Microbiol, 51(2), 359-370.

Hambraeus, G., Persson, M., & Rutberg, B. (2000). The aprE leader is a determinant of extreme mRNA stability in Bacillus subtilis. *Microbiology, 146*, 3051-59.

Lasa, I., Toledo-Arana, A., Dobin, A., Villanueva, M., Ruiz de los Mozos, I., … Gingeras, T. (2011). Genome-wide antisense transcription drives mRNA processing in bacteria. *Proceedings of the National Academy of Sciences of the United States of America, 108*(50), 20172-7. doi:10.1073/pnas.1113521108.

Ludwig, H., Homuth, G., Schmallsch, M., Dyka, F., Hecker, M., & Stulke, J. (2001). Transcription of glycolytic genes and operons in *Bacillus subtilis*: Evidence for the presence of multiple levels of control of the g*apA* operon. *Molecular Microbiology, 41*(2), 409-422. doi: 10.1046/j.1365-2958.2001.02523.x

Meher, A., Bal, N., Chary, K., & Arora, A. (2006). Mycobacterium tuberculosis H37Rv ESAT-6–CFP-10 complex formation confers thermodynamic and biochemical stability. *The FEBS Journal, 273*(7), 1445-1462. doi:10.1111/j.1742-4658.2006.05166.x

Meinke, C., Blencke, H., Ludwig, H., & Stulke, J. (2003). Expression of the glycolytic gapA operon in Bacillus subtilis: Differential synthesis of protein encoded by the operon. *Microbiology, 149*, 751-761. doi: 10.1099/mic.0.26078-0

Newbury, S., Smith, N., & Higgins, C. (1987). Differential mRNA stability controls relative gene expression within a polycistronic operon. *Cell, 51*(6), 1131-43. doi:10.1016/0092-8674(87)90599-X

Renshaw, P., Lightbody, K., Veverka, V., Muskett, F., Kelly, G., … Carr, M. (2005). Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *The EMBO Journal, 24*(14), 2491-2498. doi:10.1038/sj.emboj.7600732

Repoila, F., & Darfeuille, F. (2009). Small regulatory non-coding RNAs in bacteria: Physiology and mechanistic aspects. *Bio Cell, 101*(2), 117-131. doi:10.1042/BC20070137

Siegrist, S., Unnikrishnan, M., McConnell, M., Cheng, T., Siddiqi, N., … Rubin, E. (2009). Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. *Proceedings of the National Academy of Sciences of the United States of America, 106*(44), 18792-7. doi:10.1073/pnas.0900589106.

Silva, I., Saramago, M., Dressaire, C., Domingues, S., Viegas, S., & Arraiano, C. (2011). Importance and key events of prokaryotic RNA decay: The ultimate fate of an RNA molecule. *Wiley Interdisciplinary Reviews: RNA, 2*(6), 818-836. doi:10.1002/wrna.94

Stanley, S., Raghavan, S., Hwang, W., & Cox, J. (2003). Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system. *Proceedings of the National Academy of Sciences of the United States of America*, 100(22), 13001-13006.

Sorensen, A., Nagai, S., Houen, G., Andersen, P., & Andersen, A. (1995). Purification and characterization of a low-molecular-mass T-cell antigen secreted by Mycobacterium tuberculosis. *Infect Immun, 63*, 1710-1717.

Tan, T., Lee, W., Alexander, D., Grinstein, S., & Lui, J. (2006). The ESAT-6/CFP-10 secretion system of Mycobacterium marinum modulates phagosome maturation. *Cellular Microbiology, 8*(9), 1417-1429. doi:10.1111/j.1462-5822.2006.00721.x

World Health Organization. (2014). *Global Tuberculosis Report 2014*. Retrieved September 1, 2015, from http://apps.who.int/iris/bitstream/10665/137094/1/9789241564809_eng.pdf