Investigating RNase E Autoregulation in *Mycobacterium smegmatis*

A Major Qualifying Project

Submitted to the Faculty of

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Bachelor of Science

In

Biology & Biotechnology

By:

Karina Franca

May 18, 2020

Approved by:

Scarlet S. Shell, PhD

ABSTRACT

Mycobacterium tuberculosis (Mtb) is the pathogen that causes tuberculosis, which is currently the leading cause of death by a single infectious agent. Regulation of gene expression using the 5' Untranslated Regions (5' UTR) may influence the ability of Mtb to adapt and survive in stressful conditions within human hosts and remain in a non-replicating state for long periods of time. An enzyme involved in regulation is Ribonuclease E (RNase E), which has an important role in mRNA degradation and is responsible for cleaving the majority of *E. coli* mRNAs leading to transcript decay. In *E. coli*, RNase E has been observed to autoregulate its synthesis by targeting and cleaving its own 5' UTR. However, no studies have shown if RNase E is able to autoregulate in mycobacteria. RNase E is encoded by the gene *rne*. To understand if RNase E is able to autoregulate in mycobacteria, we determined the effects of the *rne* 5' UTR on expression of a reporter gene, mCherry, in *Mycobacterium smegmatis*. We found that mCherry mRNA and protein levels were reduced when the *rne* 5' UTR was present, in comparison to a leaderless construct. We then tested the importance of several endonucleolytic cleavage sites that had been found in the *rne* transcript. Protein levels decreased when we introduced a mutation predicted to abolish cleavage at one of these sites, and increased in a construct designed to recapitulate the cleaved transcript. These data suggest that cleavage at this position may be important for translation efficiency and/or transcript stabilization.

INTRODUCTION

Tuberculosis (TB), caused by the pathogen *Mycobacterium tuberculosis* (Mtb), is currently one of the top ten causes of death worldwide. In 2017, TB caused an estimated 1.3 million deaths and sickened roughly 10 million people (WHO, 2018). Although treatable in most cases, TB continues to be the leading cause of death by a single infectious agent, claiming more lives than HIV and AIDS (WHO, 2018). Treatment for drug-sensitive TB requires at least 6 months of combined antibiotics; however, if TB is multidrug resistant, the treatment becomes longer and more taxing (WHO, 2018). Multidrug resistant TB is becoming a prominent issue with the World Health Organization estimating 558,000 new cases that are resistant to the most effective antibiotic, rifampicin. However, not everyone infected with Mtb becomes sick. In 2014, an estimated 1.7 billion people had latent TB, in which they asymptomatically carried the bacteria (Houben & Dodd, 2016). Mtb is able to remain within the host for long periods of time in a non-replicating state, which may contribute to its survival (Sawyer, Grabowska, & Cortes, 2018). However, how Mtb is able to halt its replication and survive in non-optical conditions is unknown. Research on the biological functions of Mtb may provide crucial insight into novel ways to combat TB.

Regulation of gene expression may influence the ability of Mtb to adapt and survive in stressful conditions within human hosts. Bacterial mRNAs contain 5' Untranslated Regions (5' UTRs) which can contribute to gene regulation. A sequence within the 5' UTR that contributes to translation initiation and regulation of that process is called the Shine-Dalgarno sequence. The Shine-Dalgarno sequence functions as a ribosome-binding site to initiate translation by binding to a complementary sequence in the 16S rRNA. Although 5' UTRs have regulatory functions, almost one quarter of transcripts in mycobacteria are leaderless, meaning that they lack 5' UTRs and therefore also Shine-Dalgarno sequences (Cortes *et al.*, 2013; Shell *et al.*, 2015; Shine & Dalgarno, 1974). However, genes that encode proteins involved in active growth, such as ribosomal and energy generation proteins, were noticeably underrepresented in the leaderless transcriptome (Cortes *et al.*, 2013). Cortes *et al.* (2013) also observed more expression of leaderless mRNAs in starvation models. The results suggest that leaderless transcripts may have an important role in Mtb's non-replicating capabilities to assist in its survival in stressful conditions (Cortes *et al.*, 2013).

Ribonuclease E (RNase E) is an essential enzyme in *E. coli*, with an important role in mRNA degradation (Jain & Belasco, 1995). RNase E is responsible for cleaving the majority of *E. coli* mRNAs leading to transcript decay (Chao *et al.*, 2017; Hui, Foley, & Belasco, 2014). It has been observed that RNase E is responsible for the initial cleavage during mRNA degradation, demonstrating its essentiality within the process (Jain, Deana, & Belasco, 2002). Overproduction

or underproduction of RNase E can affect cell growth; therefore, an ideal balance is necessary to avoid too much RNase E, resulting in excessive RNA degradation, or too little, affecting RNA processing (Deutscher, 2015; Diwa, Bricker, Jain, & Belasco, 2000; Jain & Belasco, 1995). For example, one study reported that *E. coli* growth was impaired when RNase E levels were too low and that the minimum concentration needed for normal growth was 10-20% of the usual concentration found in wild type cells (Jain *et al.*, 2002). In addition, Jain *et al.* (2002) suggested that no other gene product can compensate for the lack of RNase E synthesis in *E. coli*, further demonstrating the enzyme's essentiality for cell functioning.

RNase E can contribute to mRNA degradation in bacteria by internally cleaving substrates after binding monophosphates at their 5' ends (Bandyra, Wandzik, & Luisi, 2018; Deutscher, 2015; Yajnik & Godson, 1993). However, RNase E has also been observed to cleave substrates without recognition of 5' ends, which may suggest that RNase E can function regardless of binding 5' monophosphates (Bandyra *et al.*, 2018). In addition to contributing to mRNA degradation, RNase E also controls gene expression through mRNA processing where different coding regions can become stabilized or destabilized resulting in one or more stable products rather than wholesale degradation (Joanny *et al.*, 2007; Nilsson & Uhtin, 1991; Taverniti, Forti, Ghisotti, & Putzer, 2011). Although RNase E is found in many bacteria, it is most studied in *E. coli* (Deutscher, 2015). Research has shown that the *rne* transcript, which encodes RNase E, is sensitive to RNase E levels in *E. coli* (Jain & Belasco, 1995). Based on this observation and the need for RNase E regulation, Jain & Belasco (1995) hypothesized that *E. coli* may have evolved to regulate the production of RNase E by targeting and degrading its own mRNA. Autoregulation of RNase E in *E. coli* involves changes in longevity of *rne* transcript in response to RNase E activity which allows RNase E expression to adjust to that of its substrates; increased abundance of substrates titrates RNase E away from degrading its own transcript (Diwa *et al.*, 2000; Sousa, Marchand, & Dreyfus, 2008).

The unique autoregulation ability of RNase E in *E. coli* is controlled by the 5' UTR of the *rne* gene (Diwa *et al.*, 2000; Schuck, Diwa, and Belasco, 2009). The Diwa *et al.* (2000) study observed that when the *rne* 5' UTR in addition to the first 28 nt of the *rne* coding sequence were fused with *lacz*, autoregulation still occurred despite RNase E levels not being dependent on *lacz*. This demonstrated that the *rne* 5' UTR is responsible for autoregulation. An evolutionarily conserved stem-loop (hp2) in the *rne* 5' UTR is most responsible for autoregulation while a second stem-loop (hp3) is also responsible to a lesser extent (Diwa *et al.*, 2000). The believed mechanism by which RNase E autoregulates in *E. coli* is by binding to the hp2 stem-loop within the *rne* 5' UTR which in turn facilitates cleavage at an undetermined site within the mRNA 5' UTR (Deutscher, 2015; Jain & Belasco, 1995). However, the hp3 stem-loop contains AU rich single stranded regions, which RNase E prefers to cleave in *E. coli*, suggesting that RNase E directly cleaves the hp3 stem-loop (Chao *et al.*, 2017; Diwa *et al.*, 2000). Autoregulation through RNase E binding to the hp2 stem-loop may increase *rne* mRNA degradation by increasing the affinity of RNase E to the mRNA (Schuck *et al.*, 2009). Schuck *et al.* (2009) proposed that RNase E binding to hp2 may change the conformation of RNase E, and as a result enhance its catalytic activity.

Another possibility may be that RNase E binding to hp2 inhibits translation which would in turn expose the mRNA to RNase E degradation (Schuck *et al.*, 2009).

Similar to *E. coli*, the RNase E protein was also observed to be essential in *M. smegmatis*, where growth was arrested when expression of *rne* was repressed (Taverniti *et al.*, 2011). In addition to observations of RNase E essentiality in *M. smegmatis*, RNase E is also predicted to be essential in Mtb. Transposon mutagenesis was used to predict essential genes for optimal growth in Mtb, and *rne* was among the genes identified (Dejesus *et al.*, 2017; Sassetti, Boyd, & Rubin, 2003). RNase E is a core component of the RNA degradosome complexes in Mtb and has been observed to interact with other RNA degradation proteins such as polynucleotide phosphorylase (PNPase) and ATP-dependent RNA helicase (RhlE) (Plocinski *et al.*, 2019). A truncated form of Mtb RNase E, which included the evolutionarily conserved catalytic domain, was purified and characterized *in vitro* (Zeller *et al.*, 2007). The truncated version of RNase E had the ability to form dimers and tetramers in solution, which is similar to what exists in *E. coli* homologues. Endoribonuclease activity was observed to be dependent on whether the 5' end of the RNA was phosphorylated, resulting in faster cleavage of 5'-monophosphorylated substrates compared to those that were non-phosphorylated (Zeller *et al.*, 2007).

Although there is evidence that RNase E is able to autoregulate its synthesis in *E. coli* by targeting and cleaving its own 5' UTR, no studies have shown if RNase E is able to autoregulate in mycobacteria (Jain & Belasco, 1995). Cleavage sites have been observed to be enriched in 5' UTRs of the *Mycobacterium smegmatis* transcriptome, with three cleavage sites mapped specifically in the *M. smegmatis rne* 5' UTR (Martini, Zhou, Sun, & Shell, 2019). Using the non-pathogenic Mtb model, *M. smegmatis*, we sought to understand if RNase E autoregulates its expression by cleaving its own 5' UTR in mycobacteria. We did this by observing the effects of the *rne* 5' UTR on expression of a reporter gene, mCherry. By determining if RNase E is able to autoregulate in mycobacteria, we can expand our knowledge on how Mtb regulates gene expression and how that may allow it to survive in non-optimal environments within human hosts for long periods of time.

Our results indicated that mCherry mRNA and protein levels were reduced when the *rne* 5'UTR was present, compared to a leaderless construct or a construct with a different 5' UTR commonly used in mycobacterial expression plasmids. Similar results were observed in constructs with mutations predicted to disrupt cleavage sites, with the exception of one where the mutation was predicted to interfere with cleavage at the -1 position relative to the translation start site. When cleavage was inhibited at -1, mCherry protein expression decreased approximately three-fold compared to the other constructs. However, mRNA levels were unaffected. Primer extension was used to quantify the cleavage products and resulted in a visible band corresponding with cleavage at -1. A construct designed to recapitulate the -1 cleavage product was expressed more highly than the full-length *rne* 5'UTR construct. Together, these observations suggest that cleavage at the -1 site may be important for transcript stabilization and/or translation.

MATERIALS & METHODS

Bacterial Culture Conditions. *E. coli* was cultured in LB broth and on LB agar that contained 50 μg/mL of kanamycin or 60 μg/mL of nourseothricine as needed. *M. smegmatis* was cultured in Middlebrook 7H9 supplemented with ADC (Albumin Dextrose Catalase, final concentrations 5 g/L bovine serum albumin fraction V, 2 g/L dextrose, 0.85 g/L sodium chloride, and 3 mg/L catalase), 0.2% glycerol and 0.05% Tween 80. *M. smegmatis* was also grown on 7H10 plates supplemented with ADC and 0.5% glycerol. *M. smegmatis* strains containing KAN-marked plasmids were grown in/on media that contained 25 μg/mL of kanamycin. Transformations with NAT-marked plasmids were plated on media containing or 30-60 μg/mL of nourseothricine. All plates and liquid cultures were incubated at 37°C. Liquid cultures were incubated in a 200 rpm shaking incubator.

PCR to make fragments for cloning. All plasmid constructs were created by HiFi assembly of backbones and inserts amplified by PCR with Q5 Polymerase (NEB). Each reaction contained the Q5 PCR master mix shown in Table 1. Samples were initially denatured at 98°C for 30 seconds then underwent 35 cycles of an additional denaturation at 98°C for 20 seconds, annealing at a temperature determined by the New England Biolabs Tm calculator based on the primers used for 30 seconds, and elongation at 72°C for 30 seconds per kb. Samples went through a final elongation step at 72°C for double the time of the first elongation, then were kept at 4°C until ready for use. The Q5 PCR products were run on a 1% TAE agarose gel and the correct sized bands were cut and purified following the instructions from the New England Biolabs Monarch DNA Gel Extraction kit (NEB).

Table 1: Q5 PCR Master Mix

Q5 PCR Mix	**Amount for 1X reaction (μL)**
5X Q5 Buffer	10 μL
Q5 GC Enhancer	10 μL
10 mM each dNTP mix	1 μL
10 μM Primer 1 (forward)	2.5 μL
10 μM Primer 2 (reverse)	2.5 μL
Q5 DNA Polymerase	0.5 μL
Template DNA (1ng/ μL)	1 μL
Ultrapure H_2O	22.5 μL
Total Volume	50 μL

HiFi Assembly. HiFi DNA Assembly master mix (NEB) was used to assemble the plasmids. Some plasmid assemblies were single-fragment assemblies while others consisted of adding an insert to a vector. For the vector and insert plasmids, approximately 50 ng of vector was used and the online New England Biolabs ligation calculator was used to determine how many ng of insert to use to achieve a 5:1 molar ratio. An equal volume of the HiFi DNA Assembly master mix was added to the total vector and insert volume. For single-fragment assemblies, 150-200 ng of the vector was combined with an equal volume of HiFi Assembly master mix. The reactions were incubated at 50°C for 1 hour and stored at -20°C until further use.

E. coli and M. smegmatis Transformations. The assembled plasmids were transformed into NEB-5-alpha Competent *E. coli* cells (NEB). The competent cells were thawed on ice and split into tubes with 20 µL in each for cloning reactions and 10 µL for a no-DNA control. The full HiFi Assembly reaction (5 µL) was added to the competent cells and the tubes were kept on ice for 30 minutes. The cells were heat shocked at 42°C for 30 seconds in a water bath and then placed on ice for 5 minutes. To recover the cells, 200 µL of SOC medium was added (NEB) and the cells were incubated at 37°C for 1 hour in a shaking incubator at 200 rpm. The cells were spread on LB plates that contained 50 µg/mL kanamycin and were incubated at 37°C for one day. Colonies were picked and grown in LB broth. The plasmids were extracted using the ZR Plasmid Miniprep-Classic kit (Zymo) following the procedure provided. Plasmids were sequence-verified.

Wild type mc^2155 *M. smegmatis* competent cells were thawed on ice and split into tubes with 20 µL in each and the remainder, approximately 5-10 µL, was used for a no-DNA control. Approximately 500 ng of purified plasmid was added to the cells and incubated for 10 minutes on ice. The cells were transferred to a cuvette and electroporated using a MicroPulser Electroporator (BioRad). The cells were rescued by addition of 500 µL of Middlebrook 7H9, transferred to a new tube and incubated at 37°C in a shaking incubator at 200 rpm for 2.5 hours. The cells were plated on 7H10 plates containing 25 µg/mL of kanamycin and incubated at 37°C for 3 days. Colony checking PCR was done to check for correct integration into the L5 site, and the expression cassettes from proper integrants were then amplified by PCR and sequenced to confirm that there were no mutations in the promoter, UTR, or coding sequence.

Plasmid Constructs. Table 2 shows the plasmid constructs and their replicates created with their corresponding *M. smegmatis* strains that were used in the experiments while Table 3 shows the primers used to create the plasmid constructs and check them using sequencing or checking PCR. All of the constructs are made from the pJEB402 backbone (Lee, Pascopella, Jacobs, & Hatfull, 1991) and contain a kanamycin resistance marker and an integrase and attP sequence for insertion in the L5 integration site. The MOP promoter sequence present in most of the constructs is 43 nt in length while the length of the *rne* 5' UTR is 236 nt. The MOP-associated 5' UTR present in one construct is 55 nt in length. The first 45 nt of the *rne* gene was incorporated in many of the plasmid constructs. The Δ9 mCherry gene with the first ATG deleted is 681nt in length.

Table 2: *M. smegmatis* Plasmids and Strains Used

Plasmid	Strain	Characteristics
pSS300	SS-M_0436	MOP promoter + Δ9 mCherry gene with pSS264 backbone
pSS301	SS-M_0437 SS-M_0438	MOP promoter + Δ16 mCherry gene with pSS264 backbone
pSS304	SS-M_0446 SS-M_0451	MOP promoter + No 5' UTR + first 45nt of *rne* gene + Δ9 mCherry gene with first ATG deleted
pSS305	SS-M_0448 SS-M_0449	No MOP promoter + *rne* 5' UTR + first 45nt of *rne* gene + Δ9 mCherry gene with first ATG deleted
pSS306	SS-M_0450 SS-M_0453	MOP promoter + *rne* 5' UTR + first 45nt of *rne* gene + Δ9 mCherry gene with first ATG deleted
pSS361	SS-M_0634 SS-M_0635	MOP promoter + MOP-associated 5' UTR + first 45nt of *rne* gene + Δ9 mCherry gene with first ATG deleted
pSS383	SS-M_0652 SS-M_0653	MOP promoter + *rne* 5' UTR + -1 C to G + first 45nt of *rne* gene + Δ9 mCherry gene with first ATG deleted
pSS388	SS-M_0646 SS-M_0647 SS-M_0648	MOP promoter + *rne* 5' UTR + -92 C to G + -93 C to G + first 45nt of *rne* gene + Δ9 mCherry gene with first ATG deleted
pSS389	SS-M_0654 SS-M_0655 SS-M_0656	MOP promoter + No 5' UTR + +27 C to G + +28 C to G + first 45nt of *rne* gene + Δ9 mCherry gene with first ATG deleted
pSS391	SS-M_0657 SS-M_0658 SS-M_0659	MOP promoter + *rne* 5' UTR + -195 C to G + first 45nt of *rne* gene + Δ9 mCherry gene with first ATG deleted
pSS393	SS-M_0649 SS-M_0650 SS-M_0651	MOP promoter + *rne* 5' UTR + +27 C to G + +28 C to G + first 45nt of *rne* gene + Δ9 mCherry gene with first ATG deleted
pSS423	SS-M_0758 SS-M_0759 SS-M_0760	MOP promoter + No 5' UTR + added -1 G + first 45nt of *rne* gene + Δ9 mCherry gene with first ATG deleted
pSS424	SS-M_0761 SS-M_0762 SS-M_0763	MOP promoter + No 5' UTR + added -1 C + first 45nt of *rne* gene + Δ9 mCherry gene with first ATG deleted
pJEB402	SS-M_0049	Empty vector

Table 3: Primers Used

Primer Name	Function	Sequence
JR273	Forward primer for *sigA* qPCR, use with JR274	GACTACACCAAGGGCTACAAG
JR274	Reverse primer for *sigA* qPCR, use with JR273	TTGATCACCTCGACCATGTG
SSS132	Reverse vector primer downstream of mCherry for sequencing and checking PCR	CCTGATTCTGTGGATAACCG
SSS138	Forward vector primer upstream of 5' UTR for sequencing and checking PCR	TCCTGAGTAGGACAAATCCG
SSS141	Forward primer to check for L5 integration (right junction), use with SSS142	GTGACAGGATTTGAACCTGC
SSS142	Reverse vector primer to check for L5 integration (right junction), use with SSS141	TAGAGCCGTGAACGACAGG
SSS143	Reverse primer to check for L5 integration (left junction), use with SSS144	ACAAAGTTGCGCTCGAACGC
SSS144	Forward vector primer to check for L5 integration (left junction), use with SSS143	TCGATGAGCCGCTTCTCGC
SSS241	Forward vector primer located farther upstream of SSS138. Used for sequencing and checking PCR when SSS138 results were inconclusive	CCGATTCATTAATCCGATGGTAGTGTGGGGTC
SSS254	Forward primer to amplify MOP promoter	GAGCGAGAAGCTTCCCAGGCTTGACACTTTATG
SSS563	Forward primer for mCherry qPCR, use with SSS564	GACTACTTGAAGCTGTCCTTCC
SSS564	Reverse primer in the middle of mCherry used for mCherry qPCR and sequencing and checking PCR	CGCAGCTTCACCTTGTAGAT
SSS870	Forward primer to amplify *rne* 5' UTR	GGGTGAGCGGCCTCAACC
SSS1079	Forward primer to amplify NAT cassette with overhangs for assembly into pSS291	CTAGCCAACAAAGCGACGGCCATCATGGCCGCGGTG
SSS1080	Reverse primer to amplify NAT cassette with overhangs for assembly into pSS291	CTGCCTCGTGAAGAAGGTTCAGGGGCAGGGCATGCTC

SSS1345	Reverse primer to amplify pSS291 backbone for insertion of the NAT cassette	TCTCCCCATGCGAGAGTAGG
SSS1346	Forward primer to amplify pSS291 backbone for insertion of the NAT cassette	CGTGCTCGAGCTAGCTGG
SSS1363	Reverse primer that anneals to the MOP promoter immediately upstream of the TSS to amplify pSS300, use with SSS1434	CCACACATTATACGAGCCGG
SSS1378	Forward primer to amplify pSS261 or pSS264 used to remove first 9 codons of mCherry, use with SSS1381	ATGGCCATCATCAAGGAGTT
SSS1381	Reverse primer to amplify pSS264 used to remove first 9 codons of mCherry with overlap to 16 nt of SSS1378, use with SSS1378	CCTTGATGATGGCCATCCACACATTATACGAGCCGG
SSS1382	Forward primer to amplify pSS261 or pSS264 to remove first 16 codons of mCherry, use with SSS1383	ATGCGCTTCAAGGTGCACAT
SSS1383	Reverse primer to amplify pSS261 or pSS264 to remove first 16 codons of mCherry with overlap to15 nt of SSS1382, use with SSS1382	GCACCTTGAAGCGCATCCACACATTATACGAGCCGG
SSS1434	Forward primer to amplify pSS300 excluding ATG, use with SSS1363 or 1441	GCCATCATCAAGGAGTT
SSS1435	Forward primer to amplify *rne* 5’ UTR + 45 nt *rne* coding sequence with overlap with MOP promoter, use with SSS1436	CCGGCTCGTATAATGTGTGGGGGTGAGCGGCCTCAACC
SSS1436	Reverse primer to amplify *rne* 5’ UTR+ 45 nt *rne* coding sequence with overlap with Δ9 mCherry, use with SSS1435 or SSS1440. Used for sequencing and checking PCR	AACTCCTTGATGATGGCGGGAGTCTGGGTTGATAGG
SSS1437	Reverse primer to amplify pSS300 with tail containing 30nt of the first 45nt of the *rne* gene, use with SSS1438	TAGGTCTTCGGTATGGGCATCTTCGGCCACCCACACATTATACGAGCCGG
SSS1438	Forward primer to amplify pSS300 with tail containing 33 nt of the first 45 nt of the *rne* gene, use with SSS1437	GCCCATACCGAAGACCTATCAACCCAGACTCCCGCCATCATCAAGGAGTTC
SSS1440	Forward primer to amplify the *rne* 5' UTR excluding MOP promoter and with overlap to the vector, use with SSS1436	CTGTCGTCATATCTAGACGGGTGAGCGGCCTCAACC

SSS1441	Reverse primer to amplify pSS300 excluding MOP promoter, use with SSS1434	GTCTAGATATGACGAC AG
SSS1505	Reverse primer to introduce C to G mutation at -195, use with SSS1582	GCACGTCGAACTCTTC ATTC
SSS1507	Reverse primer to introduce C to G mutations at -92 and -93, use with SSS1583	CTTGTTGTCCTCTCGCT CCG
SSS1508	Forward primer to introduce C to G mutation at -1, use with SSS1509	AGGAGAATTGGTGGCC GAAG
SSS1509	Reverse primer to introduce C to G mutation at -1, use with SSS1508	TTCGGCCACCAATTCT CCTC
SSS1511	Reverse primer to introduce C to G mutations at +27 and +28, use with SSS1584	GGGTTGATACCTCTTC GGTATG
SSS1582	Forward primer to introduce C to G mutation at -195, use with SSS1505	AAGAGTTCGACGTGCG CGACC
SSS1583	Forward primer to introduce C to G mutations at -92 and -93, use with SSS1507	AGAGGACAACAAGTCC GGGGAAC
SSS1584	Forward primer to introduce C to G mutation at +27 and +28, use with SSS1511	AGAGGTATCAACCAGA CTCC
SSS1833	Reverse primer to amplifying pSS304 with overlap for insertion of C at -1, use with SSS1834	TCGGCCACGCCACACA TTATACGAGC
SSS1834	Forward primer to amplifying pSS304 with overlap for insertion of C at -1, use with SSS1833	AATGTGTGGCGTGGCC GAAGATGCCC
SSS1835	Reverse primer to amplifying pSS304 with overlap for insertion of G at -1, use with SSS1836	TCGGCCACCCCACACA TTATACGAGC
SSS1836	Forward primer to amplifying pSS304 with overlap for insertion of G at -1, use with SSS1835	AATGTGTGGGGTGGCC GAAGATGCCC
SSS1845	Reverse primer to sequence the C or G insert at -1	TTTGAGTGAGCTGATA CCG
SSS1979	Reverse fluorescent primer labeled with Cy5 at the 5' end for primer extension. Binds to mCherry.	CCATGTGCACCTTGAA GCG

RNA Extraction and Purification. RNA was extracted from 5 mL *M. smegmatis* liquid cultures with an OD_{600} between 0.6-0.75 that were frozen with liquid nitrogen and stored at -80°C. The frozen cultures were thawed on ice and centrifuged at 3,900 rpm for 5 minutes at 4°C. The supernatant was discarded and the pellets were resuspended in 1 mL of TRIzol (Thermo Fisher)

and transferred to 100 µm zirconium lysing matrix beat-beating tubes (OPS Diagnostics). The cells were lysed using a FastPrep 5G machine for 3 cycles of 30 seconds at 7 m/sec followed by 2 min on ice after each cycle (MP Biomedicals). In a fume hood, 300 µL of chloroform was added to each sample. All samples were vortexed for 15 seconds and centrifuged at 15,000 rpm for 15 minutes at 4°C. 500 µL of the aqueous phase was mixed with 500 µL of 100% ethanol.

The RNA samples were purified using the Direct-zol RNA MiniPrep kit (Zymo). The RNA sample mix was transferred to the Direct-zol columns and centrifuged for 30 seconds at 15,000 rpm at room temperature. The flow through was discarded and 400 µL of RNA Wash Buffer (Zymo) was used to wash the samples. 80 µL of the DNase master mix, which consisted of 75 µL of DNase Digestion Buffer and 5 µL of DNase I, was added to each sample and incubated for 30 minutes at room temperature. Samples were washed twice with 400 µL of Direct-zol RNA Pre-Wash Buffer (Zymo). 700 µL of RNA Wash Buffer (Zymo) was used for a final wash and centrifuged for 2 minutes. The samples were eluted in 50 µL of RNase-free water in a clean 1.5 mL tube and vortexed to mix. RNA concentrations were measured using a Nanodrop (Thermo Fisher) and stored at -80°C.

cDNA Synthesis and Clean Up. RNA samples were thawed and diluted to have 600 ng in 5.25 µL total volume. Two sets of dilutions for each sample were made to have a reverse-transcriptase (RT) and no-RT batch. 1 µL of a random primers master mix, shown in Table 4, was added to each sample. The samples were incubated for 10 minutes at 70°C then snap-cooled in an ice-water bath for 5 minutes. While on ice, 3.75 µL of an RT or no-RT master mix shown in Table 5, were added to their respective samples. The samples were incubated for 10 minutes at 25°C followed by 5 hours at 42°C and then stored at 4°C until use.

After cDNA synthesis, RNA was degraded by adding 5 µL of 0.5 mM EDTA and 5 µL of 1 N NaOH to each sample and then incubating for 15 minutes at 65°C. Immediately following the 15 minutes, 12.5 µL of 1 M Tris HCL pH 7.5 was added to each sample to stop the reaction. The Monarch PCR & DNA Cleanup kit (NEB) was used for the cDNA clean up and all centrifugation steps were at 18,400 rcf for 1 min at room temperature. The samples were mixed with 325 µL of Binding Buffer (NEB) and transferred to a column. The samples were centrifuged and the flowthrough was discarded. 200 µL of Wash Buffer (NEB) was added to each sample and centrifuged, and this was repeated for a total of three washes. Samples were centrifuged again to remove excess ethanol from the column. Samples were eluted with 35 µL of RNase-free water in a new 1.5 mL tube and centrifuged. The samples were vortexed at low speed and the concentrations were determined using a Nanodrop (Thermo Fisher). The samples were stored at -20°C.

Table 4: Random Primers Master Mix

Reagent	1X
100 mM Tris pH 7.5	0.5 µL
Random Primers (3 mg/mL), NEB	0.17 µL
RNase-free Water	0.33 µL
Total Volume	1 µL

Table 5: RT and No-RT Master Mixes

Reagent	RT 1X	No-RT 1X
ProtoScript II Buffer (5X)	2 µL	2µL
dNTPs mix (10 mM each)	0.5 µL	0.5 µL
DTT (100 mM)	0.5 µL	0.5 µL
RNase Inhibitor, Murine, NEB (40,000 U/mL)	0.25 µL	0.25 µL
ProtoScript II Reverse Transcriptase, NEB (200,000 units/mL)	0.5 µL	0 µL
RNase-free Water	0 µL	0.5 µL
Total Volume	3.75 µL	3.75 µL

qPCR. All RT samples were diluted to 1 ng/µL using 2 µL of cDNA. The same amount of RNase-free water added to each RT sample was then added to 2 µL of the corresponding no-RT sample. All 1 ng/µL samples were further diluted to 200 pg/µL and 2 µL of each sample was added to a 96 well plate. The matched no-RT samples were again diluted using the same volumes as for the RT samples, and 2 µL of each was used for each qPCR reaction. Two qPCR master mixes were made, one for each set of primers used. The qPCR master mixes consisted of 1 µL of 2.5 µM each primer mix, 5 µL iTaq SYBR Green Supermix (BioRad), and 2 µL RNase-free water for each sample. The primer sets used were JR273 and JR274 for *sigA* and SSS563 and SSS564 for mCherry. 8 µL of the master mix was mixed with their corresponding diluted cDNA samples in the 96 well plate. Once mixed, the 96 well plate was covered with film and placed in the thermocycler. The thermocycler conditions were 50°C for 2 minutes, 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds and 61°C for 1 minute.

Flow Cytometry. 5 mL *M. smegmatis* liquid cultures were grown to an OD_{600} of ~0.75 and immediately placed on ice. Samples were diluted to an OD_{600} of 0.015 in a final volume of 1 mL with freshly filtered Middlebrook 7H9. We have found that media or buffer used for diluting

samples for flow cytometry should be filtered with a 0.22 μm filter to remove precipitate that can have bacteria-like forward and side scatter properties. 5-micron filter needles (BD) were used to filter the samples in order to remove clumps of bacteria. The filter needles were washed five times with Middlebrook 7H9 between samples and reused for three of the biological replicates for the empty vector strain (pJEB402) and the promoterless strain (pSS305). All other samples were filtered with fresh needles. Samples were run on a CytoFLEX S flow cytometer (Beckman Coulter). Forward and side scatter thresholds were lowered from the default settings to allow detection of bacteria. Events reflecting bacteria were distinguished from electronic noise by comparison of forward scatter vs side scatter plots from bacterial samples, Middlebrook 7H9, and filtered MiliQ water controls. A gate was drawn around the highest density region of a forward scatter vs side scatter plot. The same gate was applied to all samples, and 10,000 events were collected within the gate for each sample. The flow cytometer sip was rinsed with Middlebrook 7H9 between samples.

The FlowJo software was used to analyze the flow cytometry data. Forward scatter vs side scatter graphs, which represent cell size, were used to set a gate around the densest region of the size distribution, similarly to what was done during data collection. This population was used for the remaining analysis of the flow cytometry experiment. Analysis of populations tightly gated by cell size helps to minimize variability in fluorescence due to cell size differences. Fluorescence histograms with one representative replicate for each construct were created to visually compare differences in peaks. Median fluorescence and coefficient of variation were determined with the FlowJo software. GraphPad Prism 8 (GraphPad Software) was used to calculate the median fluorescence of all three replicates pooled together for each construct and to create graphical representations of the results.

Primer Extension analysis. Primer extension reactions were conducted as described by Schuster & Bertram (2014). 1 μL of a 2 μM Cy5 fluorescently labeled primer, SSS1979, that binds to mCherry was mixed with 5 μL of 1-2 μg/μL RNA for each reaction. The RT master mix that was used instead of the one described by Schuster & Bertram (2014) is shown in Table 6. After the samples were heated to 95°C for 2 minutes in order to stop the reaction, they were transferred to new 1.5 mL tubes that contained the same volume of sample loading dye (10 μL). The sample loading dye, containing bromophenol blue and xylene cyanol, were present in the 2X Novex TBE-Urea Sample Buffer (Thermo Fisher) which was used. The samples were placed on ice while the gel was set up. A 10% Mini-PROTEAN TBE-Urea Precast Gel (BioRad) was used with 1X TBE buffer. Samples were incubated at 70°C for 5 minutes in a heat block before being loaded onto the gel. The gel was run at 180 V until the dye was at the bottom of the gel and imaged on an Azure c600 (Azure Biosystems) gel doc.

Table 6: RT Master Mix for Primer Extension

Reagent	**1X**
ProtoScript II Buffer (5X)	2 µL
dNTPs mix (10 mM each)	1 µL
RNase Inhibitor, Murine, NEB (40,000 U/mL)	0.2 µL
ProtoScript II Reverse Transcriptase, NEB (200,000 units/mL)	0.5 µL
RNase-free Water	0.3 µL
Total Volume	4 µL

Fluorescence Microscopy. 2 µL of culture was mixed with 6 µL of mounting media to avoid photobleaching. 75 µL of melted 1% agar was placed on the microscope slides and approximately 3 µL of sample was loaded on top of the solidified agar pad to allow the samples to be in the same plane of focus. A drop of oil was placed on top of the coverslips and the slides were viewed with a Zeiss Axio microscope with ApoTome using a 40x oil objective (Zeiss). All samples were imaged using the same exposure time, and the same white balance settings were applied to all images before exporting as TIF files.

Statistical Analysis. Analysis of mean mRNA expression and median fluorescence values were performed using GraphPad Prism 8 (GraphPad Software). Statistical analysis was performed using a One-way ANOVA with a post-hoc Tukey Test for qPCR data and a Kruskal-Wallis with a post-hoc Dunn's Test for flow cytometry data.

RESULTS

mCherry translation begins at the annotated codon nine. In order to test the effects of the *rne* 5' UTR on gene expression, we made reporter constructs in which the MOP promoter (Lee, Pascopella, Jacobs, & Hatfull, 1991) drove expression of mCherry. However, we identified two additional potential translational start sites other than the annotated first ATG in mCherry (Figure 1A). The sequence before the true translational start site could potentially act as a leader and confound our studies of the effects of 5' UTRs. Therefore, we designed two constructs to identify the true start codon used by *M. smegmatis* to express mCherry. The two additional possible start codons were annotated codon 10 and annotated codon 17. The two constructs were designed such that the sequence upstream of the potential start codon was removed, creating delta 9 and delta 16 constructs (Figure 1B). The expression cassettes were placed in an integrating plasmid and

transformed into *M. smegmatis*. Fluorescence microscopy was used to compare the resulting strains. We observed mCherry expression from the delta 9 construct but none with the delta 16 construct, indicating that translation can initiate robustly from the annotated codon 10 but not from the annotated codon 17 (Figure 1C and D). This is consistent with the findings from Carroll *et al.* (2014).

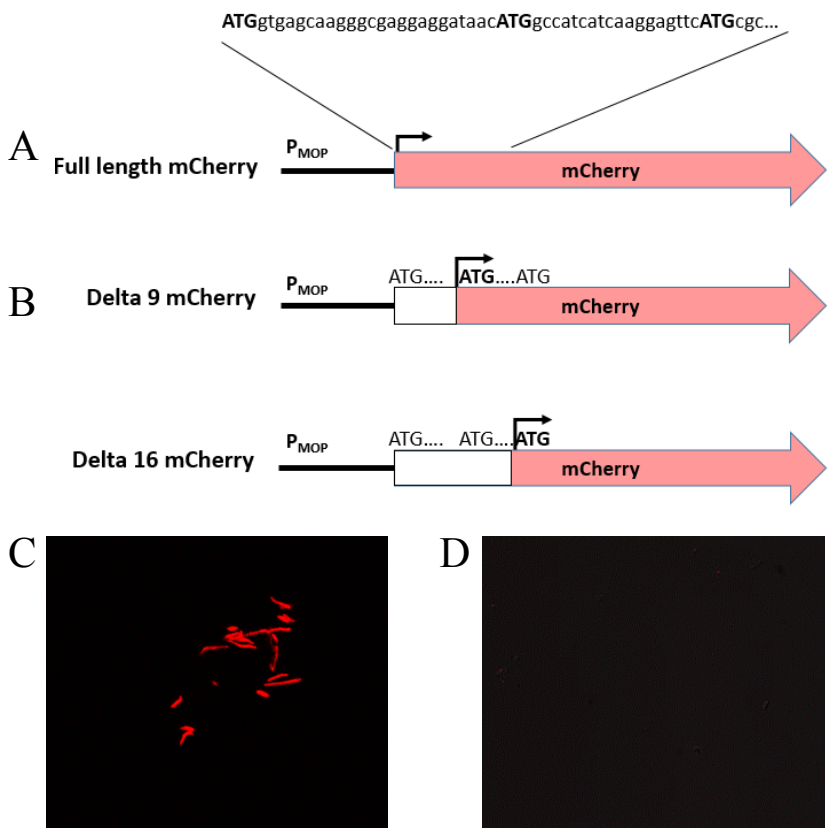


Figure 1. The true translational start site in mCherry is at annotated codon 10. A) Schematic of the three possible translational start sites in mCherry. B) Two constructs were created where the sequence upstream of the potential alternative start codons, located at codon 10 and codon 17, was removed resulting in delta 9 and delta 16 constructs. White boxed indicate deleted sequences. C) Fluorescence microscopy of *M. smegmatis* samples that contained a P_{MOP} + delta 9 plasmid construct. D) Fluorescence microscopy of *M. smegmatis* samples that contained a P_{MOP} + delta 16 plasmid construct.

Plasmid constructs for testing the impact of the rne 5' UTR on gene expression. In order to test the effects of the presence of the *rne* 5' UTR as well as inhibition of different cleavage sites on gene expression, 11 mCherry reporter plasmid constructs were created (Table 2). We included the first 45 nts of the *rne* coding sequence in all of the designed plasmid constructs to capture potential regulatory effects from interactions between that region and the 5' UTR, as well as to include a cleavage site previously identified in that region (Martini *et al.*, 2019) (Figure 2). Based on the results in Figure 1 indicating that the true start codon is at annotated codon 10, we designed all constructs to contain delta 9 mCherry. We removed the ATG start codon of delta 9 mCherry to ensure that translation would begin only at the GTG start codon at the beginning of the 45 nt of the *rne* coding sequence. Four cleavage sites were previously identified using an RNA-Seq transcript 5' end mapping approach (Martini *et al.*, 2019), three in the *rne* 5' UTR and one in the *rne* coding sequence. The cleavages were immediately upstream of nucleotides -195, -92, -1, and +27 relative to the translation start site. Like most of the cleavage sites identified in that paper, the *rne* cleavage sites were all immediately upstream of cytidines. We therefore hypothesized that the Cs were important for cleavage, and made C to G mutations at each location to attempt to disrupt

cleavage. We used these plasmid constructs as well as two controls: an empty vector plasmid and a plasmid construct that contained the MOP promoter + MOP-associated 5' UTR + delta 9 mCherry for comparison.

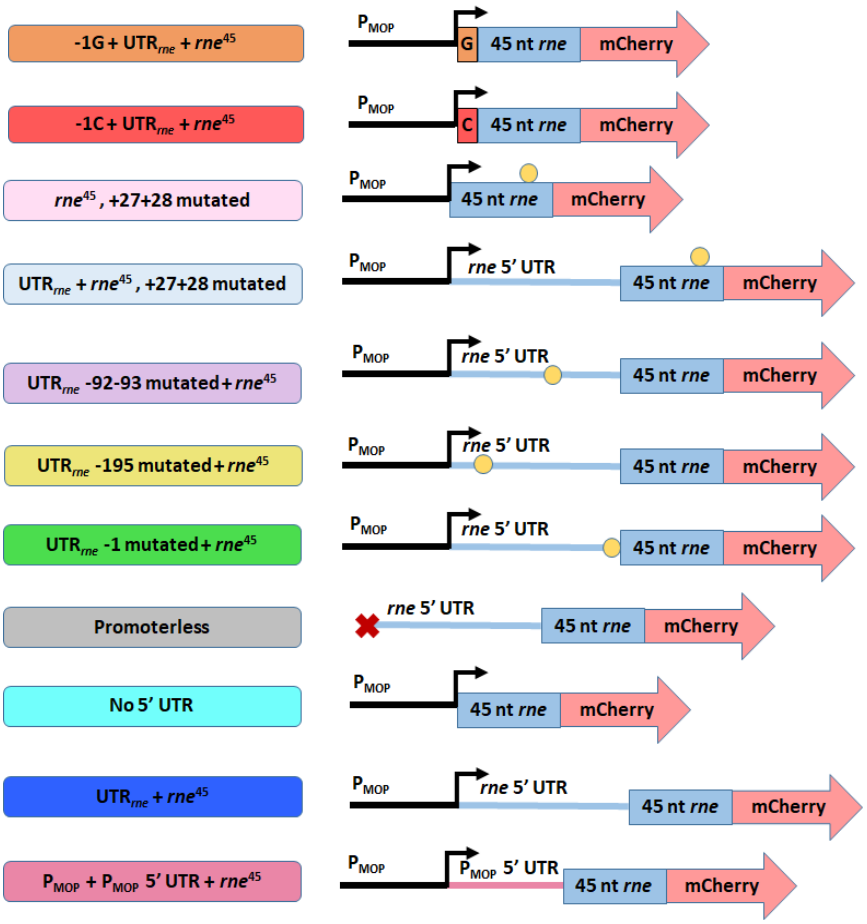


Figure 2. Plasmid constructs used in experiments. Plasmid constructs were transformed into *M. smegmatis* and used for flow cytometry and qPCR. All constructs contain delta 9 mCherry with the first ATG removed and the first 45 nt of the *rne* gene. All constructs contain the MOP promoter except for the promoterless control. The yellow circle represents locations of the C to G cleavage site mutations. The mutated nucleotides are numbered relative to the translation start site.

The rne 5' UTR decreases mCherry protein and mRNA levels. Fluorescence microscopy was used to qualitatively observe the effects of the *rne* 5' UTR on mCherry expression. We observed a decrease in fluorescence for the MOP Promoter + *rne* 5' UTR construct compared to the leaderless construct that contained the MOP promoter without a 5' UTR (Figure 3A and B). No fluorescence was observed for the construct that contained the *rne* 5' UTR and mCherry but lacked a promoter, as expected. Although there was a significant decrease in fluorescence when the *rne* 5' UTR was present, mCherry expression was higher than in the promoterless control, indicating that the observed fluorescence was not a result of background fluorescence. Fluorescence was quantitatively measured using flow cytometry and the results were consistent with the microscopy (Figure 3C). The median fluorescence of the leaderless construct was about two-fold higher than that of the construct that contained the *rne* 5' UTR (Figure 3D). However, the median fluorescence for the leaderless construct was lower compared to the MOP Promoter + MOP-associated 5' UTR construct. The leaderless construct therefore led to a decrease in mCherry protein production

compared to the MOP associated 5' UTR construct, while the presence of the *rne* 5' UTR resulted in an additional reduction in mCherry protein production.

In order to also determine the effects of the *rne* 5' UTR on mRNA levels, qPCR was used and expression means were determined and normalized to the housekeeping gene *sigA*. The presence of the *rne* 5' UTR resulted in lower expression compared to the leaderless construct lacking a 5' UTR (Figure 3E). As expected, the promoterless control had little to no expression. We concluded that the presence of the *rne* 5' UTR decreases both protein and RNA levels.

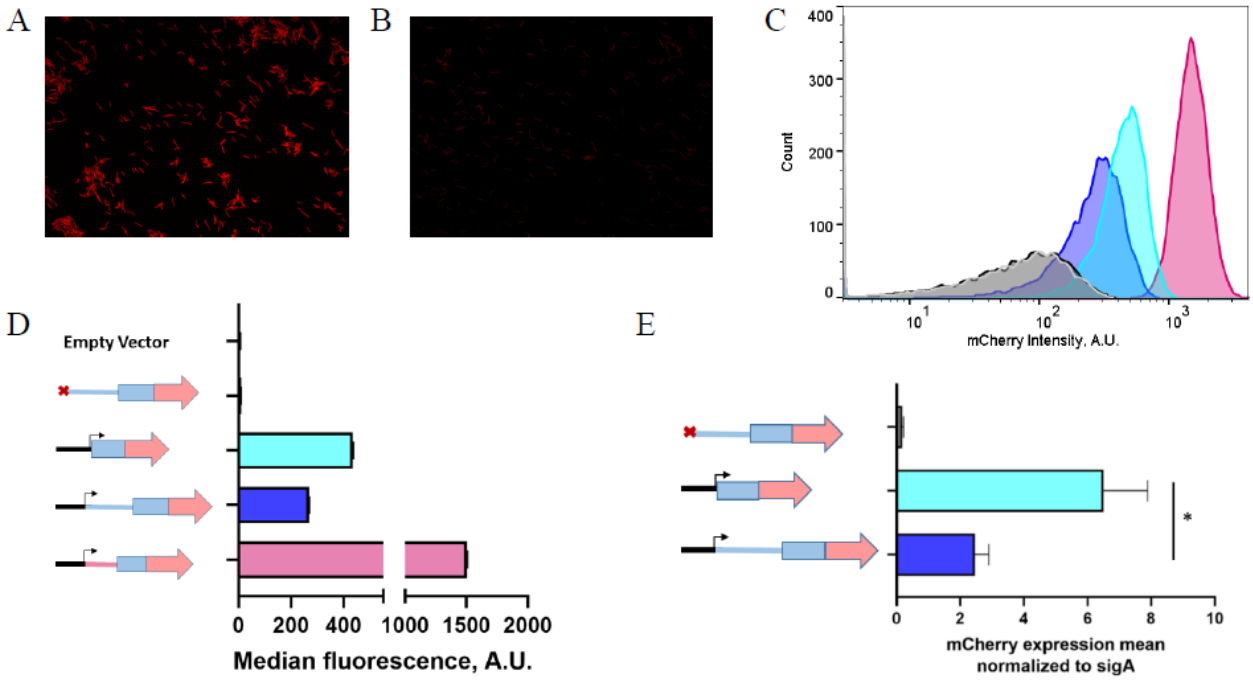


Figure 3. Negative effect of *rne* 5' UTR on mCherry expression and mRNA levels. A) Fluorescence microscopy of *M. smegmatis* samples that contained a leaderless plasmid construct with the MOP promoter. B) Fluorescence microscopy of *M. smegmatis* samples that contained a P_{MOP} + *rne* 5' UTR plasmid construct. C) Flow cytometry histogram comparing mCherry intensity in strains containing the *rne* 5' UTR containing construct, leaderless construct, P_{MOP} + MOP associated 5' UTR construct, and the promoterless and empty vector negative controls. The data shown are from one biological replicate of each strain. The colors correspond to the colors used in panel D for each construct. D) Median fluorescence from the flow cytometry data of all three biological replicates for each construct with schematics of the constructs shown on the y-axis. All constructs were determined to be statistically significantly different from each other, except that there was no significant difference between the promoterless construct and the empty vector. Kruskal-Wallis with a Dunn's multiple comparisons test. Error bars are medians of all three replicates pooled together with 95% confidence intervals (CI). E) Average mCherry mRNA expression normalized to *sigA* was determined by qPCR with all constructs shown in Figure 5. For clarity, here we show just the *rne* 5' UTR construct with the leaderless and promoterless control. The *rne* 5' UTR construct was determined to be statistically significantly different compared to the leaderless construct. One-way ANOVA followed by Tukey's post hoc test, * p=0.0126. Error bars are mean with SD.

Inhibition of cleavage at -1 negatively affects mCherry protein levels but not mRNA levels. The effects of mutations predicted to inhibit cleavage at the different cleavage sites were observed by flow cytometry to quantify mCherry protein levels. For each construct, three biological replicates were measured. The replicates were indistinguishable for all constructs (see Figure 8), and the median fluorescence of one replicate for each of the five cleavage site mutation constructs were compared in the histogram (Figure 4A). The median fluorescence of all three replicates for each of the five cleavage site mutation constructs were compared in Figure 4B. Mutation of most of the cleavage sites resulted in fluorescence levels similar to the unmutated *rne* 5' UTR construct (Figure 4A and B). However, the -1 C to G mutation resulted in an approximately three-fold decrease in fluorescence compared to the other cleavage site mutations and the *rne* 5' UTR construct.

To determine if inhibition of cleavage at -1 affects mRNA levels in addition to mCherry protein levels, we performed qPCR and determined expression means normalized to *sigA*. Mutation of the -1 cleavage site did not change mCherry mRNA levels relative to the other two cleavage site mutations within the *rne* 5' UTR or the unmutated 5' UTR (Figure 5). The UTR_{rne} + rne^{45}, +27+28 mutated construct and rne^{45}, +27+28 mutated construct were not included in this experiment because we chose to focus on the cleavage sites within the *rne* 5' UTR as potential sites for autoregulation. These results suggest that cleavage at the -1 site may be favorable for translation.

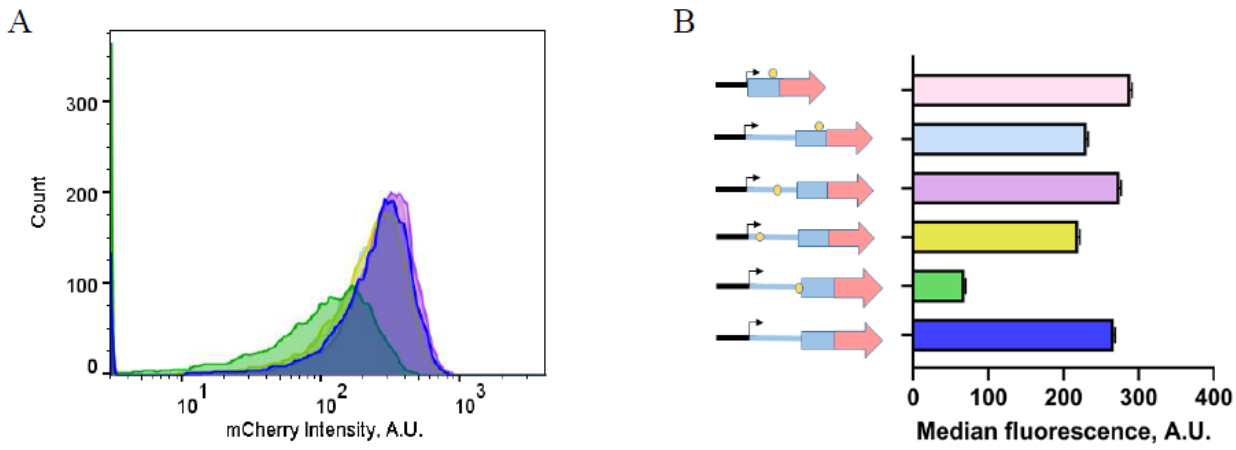


Figure 4. Effects of cleavage site mutations within the *rne* 5' UTR on mCherry fluorescence. A) Flow cytometry histogram of *M. smegmatis* samples that contained cleavage site C to G mutations within both the *rne* 5' UTR and the 45 nt of the *rne* gene compared to *rne* 5' UTR construct. The colors correspond to the colors used in panel B for each construct. B) Median fluorescence from the flow cytometry data of all three biological replicates for each construct with schematic of the constructs shown on the y-axis. Although all constructs were determined to be statistically significantly different from each other, the median fluorescence for the *rne* 5' UTR construct was very similar to most the cleavage site mutations, with the exception of the -1 C to G construct, which had substantially lower fluorescence. Kruskal-Wallis with a Dunn's multiple comparisons test. Stars indicating significance are not shown, since all constructs differed from each other. Error bars are median of all three replicates pooled together with 95% CI.

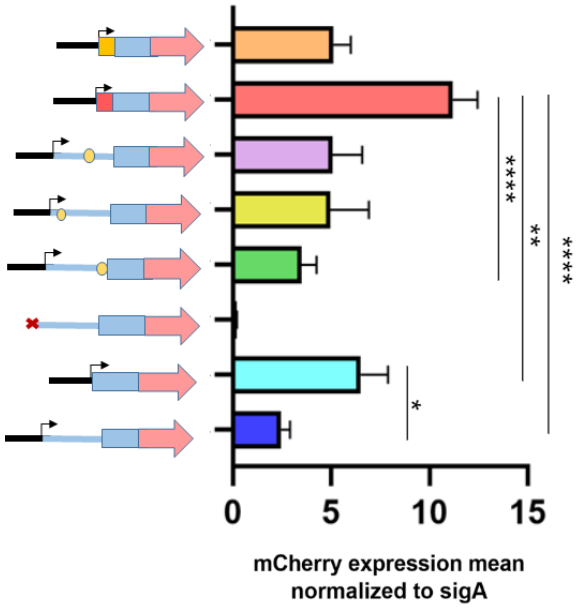


Figure 5. mCherry mRNA levels measured by qPCR. Average mCherry expression normalized to *sigA* was determined by qPCR. A yellow block immediately downstream of the transcription start site indicates G upstream of the translation start site, and a red block indicates C upstream of the translation start site. There was no significant difference between constructs bearing the -1 C to G mutation and either of the other cleavage site mutations or the unmutated *rne* 5' UTR construct. There was significance between the C upstream construct and all other constructs. The significance of some comparisons was omitted from the figure for clarity. One-way ANOVA followed by Tukey's post hoc test, * $p=0.0126$, ** $p=0.0036$, **** $p<0.0001$. Error bars are mean of three biological replicates with SD.

Adding a C nucleotide at the -1 position of an otherwise leaderless construct resulted in increased mCherry fluorescence and mRNA levels. If cleavage were to occur at the -1 position, it would result in a transcript with a C nucleotide upstream of the translational start side. In order to determine if the extra nucleotide upstream has an effect on translation, we created a plasmid construct to express a transcript similar to what would occur if the full-length *rne* 5' UTR was cleaved at the -1 position. The "C upstream at -1" construct contains a C upstream of the translational start site. We created an additional construct that contains a G upstream of the translational start site instead of C because transcription preferentially initiates with A and G nucleotides (Martini *et al.*, 2019). With both of these constructs, we were able to determine if just an extra nucleotide at -1 had an effect on mRNA abundance or translation, and specifically test the hypothesis that a C nucleotide present at -1 promotes translation.

The mCherry mRNA levels for the C upstream at -1 construct were significantly greater than the mRNA levels for the other cleavage site mutants, the unmutated *rne* 5' UTR, and the leaderless construct (Figure 5). In contrast, the G upstream construct resulted in mRNA levels similar to the other cleavage site mutants, the unmutated *rne* 5' UTR, and the leaderless construct. It was notable that the C upstream construct had higher mRNA levels than the leaderless construct, from which it differed by only one nucleotide. We also compared fluorescence in the C and G upstream constructs and the leaderless and the -1 mutation constructs using flow cytometry. The median fluorescence of the G upstream construct was similar to that of the -1 C to G mutation construct. The median fluorescence of the C upstream at -1 construct was higher than both the G upstream and -1 C to G mutations constructs but similar to the leaderless construct (Figure 6A and B). These observations suggest that cleavage at the -1 site, which results in a C nucleotide upstream of the translational start site at -1, may be important for transcript stabilization and/or translation.

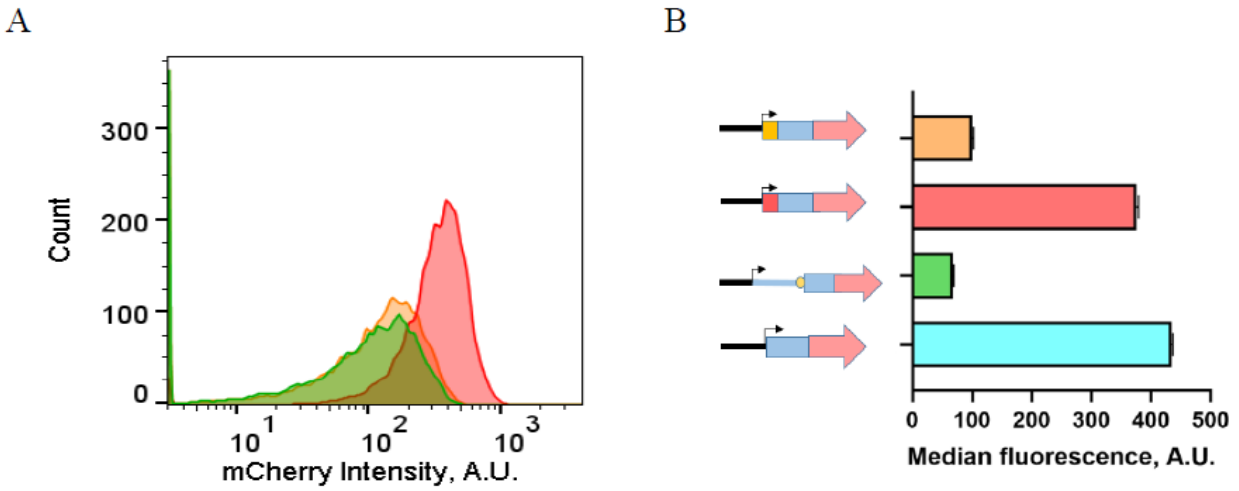


Figure 6. mCherry protein levels increase in the C upstream construct compared to the -1 mutated construct. A) Flow cytometry histogram of *M. smegmatis* samples that contained -1 C to G cleavage site mutation construct compared to constructs that are leaderless except for a C or G upstream of the translation start site. The colors correspond to the colors used in panel B for each construct. B) Median fluorescence from the flow cytometry data of all three biological replicates for each construct with schematics of the constructs shown on the y-axis. A yellow block immediately downstream of the transcription start site indicates G upstream of the translation start site, and a red block indicates C upstream of the translation start site. All constructs were determined to be statistically significantly different from each other. Kruskal-Wallis with a Dunn's multiple comparisons test. Stars indicating significance are not shown, since all constructs differed from each other. Error bars are median of all three replicates pooled together with 95% CI.

Fluorescence primer extension revealed bands corresponding to the full-length rne 5' UTR and UTR cleaved at -1. Fluorescence primer extension was used to quantify the specific cleavage products (Figure 7A). We observed large obscuring areas of fluorescence towards the bottom of the gel that was due to the presence of xylene cyanol in the loading dye which fluoresces in the same channel as Cy5 (Figure 7B). Two primer extension reactions were performed with RNA extracted from replicate cultures of the strain containing mCherry and the full-length *rne* 5' UTR. There was a bright band for mCherry reaction 1 that ran at a position consistent with the 321 nt primer extension product that would be made from the full length *rne* 5' UTR. The next brightest band we observed ran at a position consistent with the 86 nt primer extension product that would be made from mRNA cleaved at -1. This is the cleavage site that had the largest effect on mCherry expression when mutated (Figure 4A and B). The other three cleavage products would produce primer extension products of 280 nt, 178 nt, and 59 nt. Fainter bands were observed at positions consistent with these sizes. We observed an extremely faint band at approximately 120 nt that may be the 5S rRNA subunit. We did not observe any bands for the mCherry reaction 2, possibly because the RNA concentration was lower compared to the mCherry reaction 1. Reactions were also performed on RNA extracted from a WT strain using a primer annealing to the endogenous *rne* gene. However, no bands were seen.

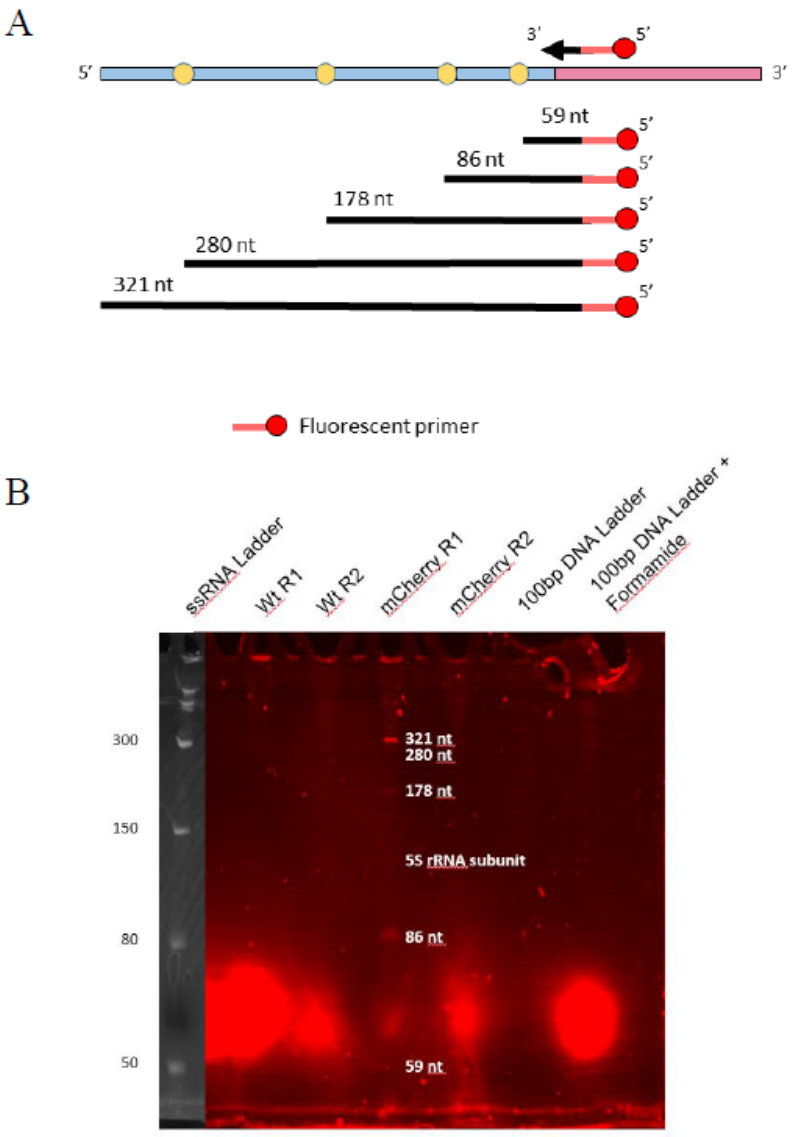
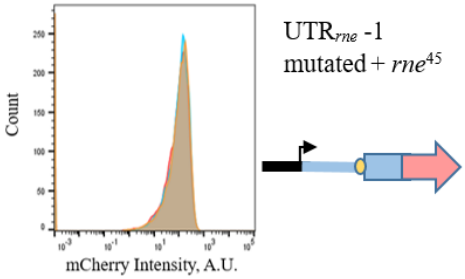


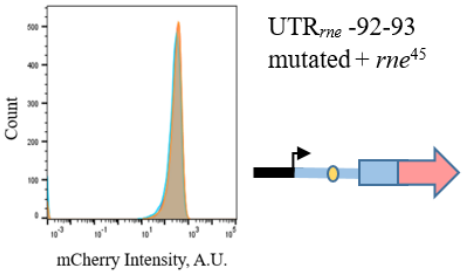
Figure 7. Fluorescence-based primer extension to quantify *rne* cleavage products. A) Schematic of the fluorescence-based primer extension method in which the fluorescent primer binds to mCherry mRNA and reverse transcription of RNA results in different cDNA product lengths corresponding with the mapped cleavage sites. B) Picture of the primer extension gel with annotated bands based on the expected cleavage products.

Flow cytometry histograms from replicate strains were overlapped to confirm similarities. Three biological replicates were used for each of the twelve plasmid constructs for flow cytometry. We overlaid the fluorescence histograms from the three replicates of each construct to confirm their similarities (Figure 8). Once determined that all replicates were similar, we chose the replicate with the median fluorescence value for each construct and used it for the flow cytometry histogram figures shown in Figure 9.



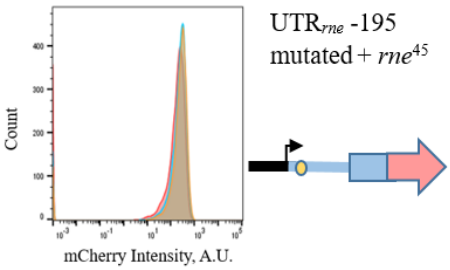
UTR$_{rne}$ -1 mutated + *rne*^{45}

	Strain	Count	Median mCherry	CV
	SS-M_0653	8288	72.3	99.9
	SS-M_0652-2	8456	72.7	99.6
	SS-M_0652-1	8792	58.8	107



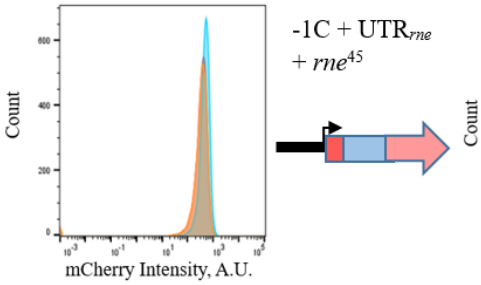
UTR$_{rne}$ -92-93 mutated + *rne*^{45}

	Strain	Count	Median mCherry	CV
	SS-M_0648	8631	290	49.3
	SS-M_0647	8971	250	53.2
	SS-M_0646	8632	285	49.0



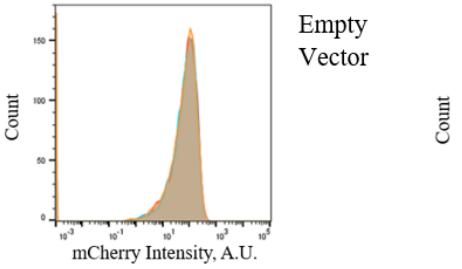
UTR$_{rne}$ -195 mutated + *rne*^{45}

	Strain	Count	Median mCherry	CV
	SS-M_0659	7942	253	54.1
	SS-M_0658	8633	232	56.0
	SS-M_0657	9015	182	65.6



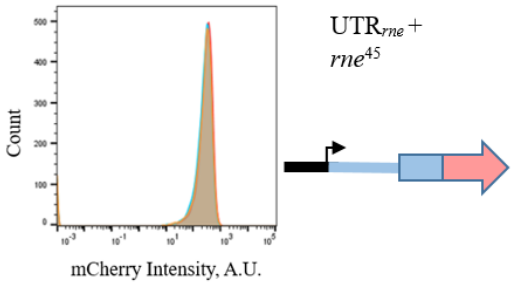
-1C + UTR$_{rne}$ + *rne*^{45}

	Strain	Count	Median mCherry	CV
	SS-M_0763	8715	337	46.6
	SS-M_0762	8970	453	38.2
	SS-M_0761	8884	340	46.4



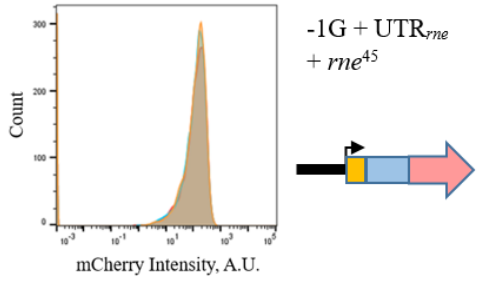
Empty Vector

	Strain	Count	Median mCherry	CV
	SS-M_0049-3	8822	4.70	144
	SS-M_0049-2	8825	4.59	143
	SS-M_049-1	8778	5.70	143



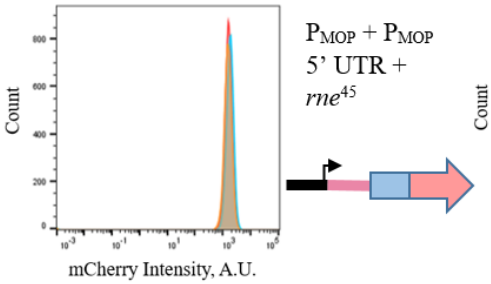
UTR$_{rne}$ + *rne*^{45}

	Strain	Count	Median mCherry	CV
	SS-M_0453	8344	267	52.6
	SS-M_0450-2	8680	253	52.6
	SS-M_0450-1	8560	283	49.4



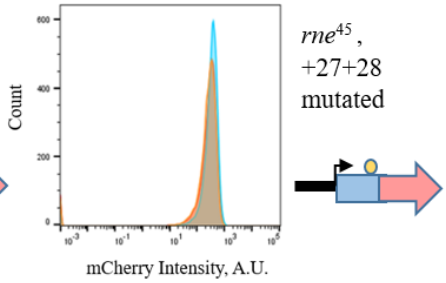
-1G + UTR$_{rne}$ + *rne*^{45}

	Strain	Count	Median mCherry	CV
	SS-M_0760	8755	102	86.5
	SS-M_0759	8600	101	87.8
	SS-M_0758	8488	97.8	89.0



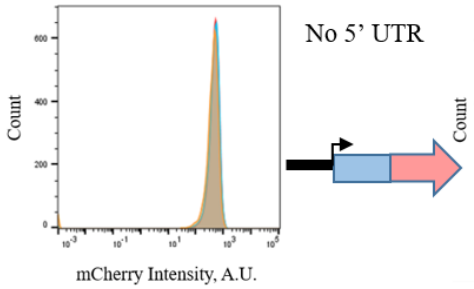
P_{MOP} + P_{MOP} 5' UTR + *rne*^{45}

	Strain	Count	Median mCherry	CV
	SS-M_0635-2	8566	1395	31.9
	SS-M_0635-1	8542	1677	29.9
	SS-M_0634	9016	1466	30.5



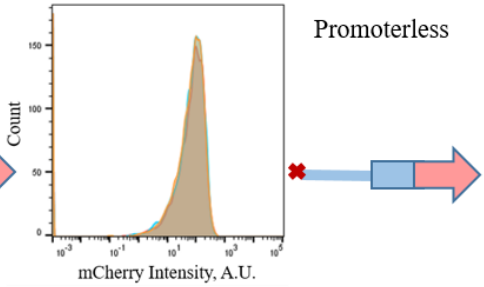
rne^{45}, +27+28 mutated

	Strain	Count	Median mCherry	CV
	SS-M_0656	8988	262	52.6
	SS-M_0655	9049	328	45.0
	SS-M_0654	8554	275	51.8



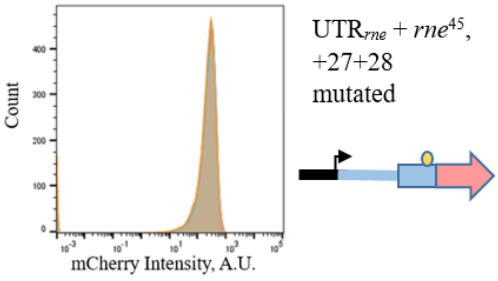
No 5' UTR

	Strain	Count	Median mCherry	CV
	SS-M_0451	8740	403	41.9
	SS-M_0446-2	8709	447	37.9
	SS-M_0446-1	8604	453	37.2



Promoterless

	Strain	Count	Median mCherry	CV
	SS-M_0449	8980	8.09	142
	SS-M_0448-2	8742	9.18	139
	SS-M_0448-1	8388	7.70	141



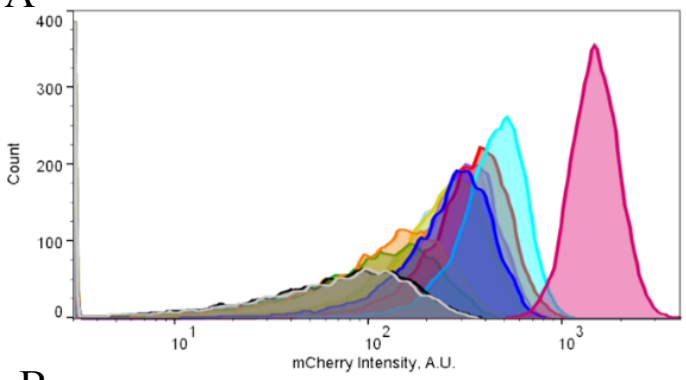
UTR$_{rne}$ + *rne*^{45}, +27+28 mutated

	Strain	Count	Median mCherry	CV
	SS-M_0651	8855	226	56.9
	SS-M_0650	8534	234	55.5
	SS-M_0649	8474	233	56.2

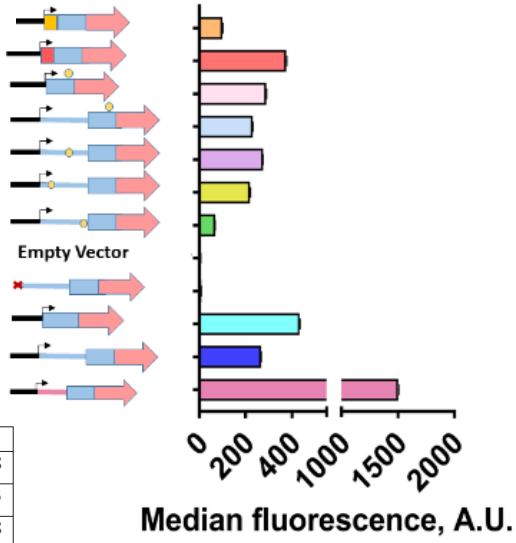
Figure 8. All three biological replicates for each construct have similar median fluorescence. Overlay of the median fluorescence histograms for each replicate for each construct. "Count" refers to the number of events recorded in a size-gated region.

Direct comparison of mCherry protein levels for all 12 constructs. Statistical analysis was performed on the data from all 12 constructs from flow cytometry Figure 9). All constructs were determined to be significantly different from each other, except there was no significance between the promoterless control and the empty vector (Kruskal-Wallis with a Dunn's multiple comparisons test). Due to it being difficult to visually compare specific constructs when all 12 were included in the histogram and bar graph, we chose to create smaller graphs using the data from 9A-C to compare specific constructs with each other as shown in Figures 3C and D, 4A and B, and 6A and B.

A



C



B

	Sample Name	Count	Median Fluorescence	CV
	G upstream at -1	8600	101	86.8
	C upstream at -1	8884	340	46.3
	No 5' UTR and +27 and +28 C to G mutations	8554	276	51.8
	+27 and +28 C to G mutations	8474	233	56.0
	-92 and -93 C to G mutations	8632	286	48.9
	-195 C to G cleavage site mutation	8633	232	55.9
	-1 C to G cleavage site mutation	8288	72.4	98.2
	Empty Vector	8822	4.70	137
	Promoterless control	8980	8.09	136
	No 5' UTR (leaderless)	8709	447	37.9
	rne 5' UTR	8344	266	52.6
	P_{MOP} 5' UTR + P_{MOP} 5' UTR	9016	1462	29.8

Figure 9. A) Flow cytometry histogram of all 12 constructs together. B) Table corresponding with information shown in the histogram in panel A. The number of events counted within the size-gated region for each representative replicate for each construct and the median fluorescence and CV can be seen. C) Median fluorescence of all 12 constructs. Statistical analysis using Kruskal-Wallis with a Dunn's multiple comparisons test was performed with the data of all 12 constructs together. All comparisons were statistically significant except the promoterless construct vs. the empty vector. For visual clarity, stars denoting p values are not shown.

Making constructs to test the impact of RNase E levels on reporter transcript cleavage. We hypothesized that RNase E is responsible for the cleavage events in the *rne* 5' UTR. In order to determine if RNase E protein levels affect cleavage of the *rne* 5' UTR, we planned to use a tet-off system in which the addition of Anhydrotetracycline (ATc) would repress *rne* transcription and allow us to subsequently observe 5' UTR cleavage by primer extension. Reduced 5' UTR cleavage during RNase E transcription inhibition may suggest RNase E is able to autoregulate in mycobacteria. Other lab members had constructed an *M. smegmatis* strain in which the native *rne* promoter was replaced with a promoter that is repressed by the tet repressor (TetR) in the presence of ATc. In this strain, TetR is expressed from a plasmid, pSS291, bearing a kanamycin resistance marker (KAN). We attempted to replace the KAN marker in pSS291 with a nourseothricine resistance marker (NAT) because all the *rne* reporter constructs contain a KAN marker and we needed different antibiotic resistance markers to select for the two plasmids. The plasmid was cloned in *E. coli* and sequence-verified. However, when attempting to transform this plasmid construct into *M. smegmatis*, the bacteria did not grow. The primers used to amplify NAT and the pSS291 backbone are shown in Table 3. The NAT cassette was amplified from pSS221 and the coding sequence highlighted in blue as well as the amino acid translation sequence can be seen in Table 7. The *E. coli* strain bearing the plasmid is SS-E_0374 and plasmid is pSS450. We wondered if the NAT cassette that was amplified was missing promoter elements important in *M. smegmatis* but not in *E. coli*. However, the region upstream of the coding sequence in the NAT cassette is 139 nt long, therefore is likely to contain a full promoter. We are unsure why this plasmid construct was able to confer NAT resistance in *E. coli* but in *M. smegmatis*.

Table 7. NAT cassette amplified sequence, coding sequence, and translation sequence

NAT cassette that was amplified with the NAT coding sequence highlighted in blue	CTAGCCAACAAAGCGACGGCCATCATGGCCGCGGTGATCAGCTAGAGGGGCGTCAGGC GCCGGGGGCGGTGTCCTGGCGCGCCTTGACACCGCTAGCTCGAGTGATATAATCTGGG AGGCATGCGAAGGAGATATACCTATGACCACTCTTGACGACACTGCTTACCGGTACCGC ACCAGTGTCCCGGGGGACGCCGAGGCCATCGAGGCACTGGATGGGTCCTTCACCACCG ACACCGTCTTCCGCGTCACCGCCACCGGGGACGGCTTCACCCTGCGGGAGGTGCCGGTG GACCCGCCCCTGACCAAGGTGTTCCCCGACGACGAATCGGACGACGAATCGGACGACG GGGAGGACGGCGACCCGGACTCCCGGACGTTCGTCGCGTACGGGGACGACGGCGACCT GGCGGGCTTCGTGGTCGTCTCGTACTCCGGCTGGAACCGCCGGCTGACCGTCGAGGACA TCGAGGTCGCCCCGGAGCACCGGGGGCACGGGGTCGGGCGCGCGTTGATGGGGCTCGC GACGGAGTTCGCCCGCGAGCGGGGCGCCGGGCACCTCTGGCTGGAGGTCACCAACGTC AACGCACCGGCGATCCACGCGTACCGGCGGATGGGGTTCACCCTCTGCGGCCTGGACA CCGCCCTGTACGACGGCACCGCCTCGGACGGCGAGCAGGCGCTCTACATGAGCATGCC CTGCCCCTGAACCTTCTTCACGAGGCAG
NAT translation	MTTLDDTAYRYRTSVPGDAEAIEALDGSFTTDTVFRVTATGDGFTLREVPVDPPLTKVFPD DESDDESDDGEDGDPDSRTFVAYGDDGDLAGFVVVSYSGWNRRLTVEDIEVAPEHRGHGV GRALMGLATEFARERGAGHLWLEVTNVNAPAIHAYRRMGFTLCGLDTALYDGTASDGEQ ALYMSMPCP*

DISCUSSION

RNase E is an essential ribonuclease that has been observed to autoregulate its synthesis is *E. coli*; however, the same has not yet been observed in mycobacteria (Jain & Belasco, 1995; Jain, Deana, & Belasco, 2002). We sought to investigate RNase E's potential autoregulatory capabilities in mycobacteria by observing the effects of the *rne* 5' UTR on expression of mCherry in *M. smegmatis*. The *rne* 5' UTR had a negative effect on mCherry, decreasing both protein and mRNA levels compared to the leaderless construct and MOP-associated UTR construct (Figure 3). The decrease in protein abundance when the *rne* 5' UTR was present may be due to differences in translation efficiency. The *rne* 5' UTR sequence may result in different mRNA folding compared to when there is no 5' UTR or when there is the MOP-associated UTR, affecting ribosome binding and translation efficiency. Half-life can be affected by translation efficiency, where a decrease in translation efficiency can lead to faster mRNA decay and therefore reduced steady-state mRNA levels (Arnold, Yu, Belasco, & 1998; Braun, Le Derout, & Régnier, 1998; Hambraeus, Karhumaa, & Rutberg, 2002; Jurgen, Schweder, & Hecker, 1998; Pato, Bennett, & von Meyenburg, 1973; Sharp, Bechhofer, 2003; Wagner, Gesteland, Dayhuff, & Weiss, 1994). Another possibility is that differences in protein abundance may be due to differences in mRNA abundance. Although we did not include the MOP-associated UTR construct in the qPCR experiment, we observed higher mRNA levels in the leaderless construct compared to the *rne* 5'UTR construct (Figure 3E). Therefore, the decrease in protein abundance in the *rne* 5'UTR construct may be due to a decreased mRNA abundance. Differences in fluorescence levels and mRNA levels may also be due to differences in transcription rate or differences in half-life. Transcript production rate can be calculated using mRNA abundance and half-life data in order to observe if the 5' UTR had an effect on the transcription rate (Nguyen, Vargas-Blanco, Roberts, & Shell, 2020). However, in order to determine transcription rates and to distinguish between these two possibilities, half-lives must first be measured.

After mutating the different cleavage sites within both the *rne* 5' UTR and the first 45 nt of the *rne* coding sequence, we observed that most of the cleavage site mutated constructs resulted in similar fluorescence levels compared to the *rne* 5'UTR construct (Figure 4). However, when the cleavage site at the -1 position was mutated, there was a large decrease in fluorescence. This difference may also be due to differences in translational efficiency, where cleavage at the -1 position may result in a more favorable transcript. The mRNA levels of all the cleavage site mutated constructs within the *rne* 5' UTR were similar to each other and to that of the unmutated *rne* 5' UTR construct (Figure 5). Therefore, the decreased protein levels in the -1 cleavage site mutant cannot be explained by decreased mRNA levels.

We predict that the cleavage site mutations inhibit cleavage; however, we did not directly test this. Therefore, we can't determine if the differences we observed for the -1 cleavage site mutant were due to lack of cleavage or to other properties of the transcript that were changed by

the mutations. In cases where mutating the cleavage site had no observed effect, we are unable to determine if that is due to cleavage at that position having no effect in general, or because the mutations didn't abolish cleavage. Another possibility to explain the lack of differences we observed for some of the cleavage site mutants is that there may be little cleavage at those positions to begin with. The primer extension data suggest that most of the transcript is not cleaved in the unmutated UTR construct, with the brightest band running at a position consistent with the 321 nt primer extension product that would be made from the full length *rne* 5' UTR (Figure 7). However, the next brightest band was observed to run at a position consistent with the 86 nt primer extension product that would be made from mRNA cleaved at -1. The transcript appears to be cleaved the most at the -1 position compared to the other cleavage sites within the *rne* 5' UTR. Future experiments should include the mutant constructs in the primer extension experiment to determine if the mutations in fact abolished cleavage at these locations.

We created the C and G upstream constructs to assess if having an extra nucleotide at -1 affects mRNA abundance or translation, and if any such differences are specifically due to the added C nucleotide. There was more C upstream transcript compared to any of the other constructs tested (Figure 5), which could be due to increased transcription or increased mRNA stability. Half-lives must again be measured in order to distinguish between these possibilities. There is also more protein for the C upstream construct which may be because there is more transcript. Having that extra C upstream may allow transcription to initiate more efficiently, resulting in more transcript and as a result, more protein. Another possibility is that translation is more efficient with the extra C upstream, resulting in more protein and potentially stabilizing the transcript. However, we did not confirm that the C upstream is actually present in the transcript. It is possible that transcription does not begin at the added nucleotide upstream and instead still begins at the first nucleotide of the coding sequence. In order to determine where transcription actually begins, future studies should include 5' RACE on a strain with the C upstream construct as well as the G upstream construct.

REFERENCES

Arnold, T. E., Yu, J., Belasco, J. G. (1998). mRNA stabilization by the ompA 5′ untranslated region: two protective elements hinder distinct pathways for mRNA degradation. *RNA 4*:319–330

Bandyra, K. J., Wandzik, J. M., & Luisi, B. F. (2018). Substrate recognition and autoinhibition in the central ribonuclease RNase E. *Molecular Cell, 72*(2). Doi: 10.1016/j.molcel.2018.08.039

Braun, F., Le Derout, J., Régnier, P. (1998). Ribosomes inhibit an RNase E cleavage which induces the decay of the rpsO mRNA of Escherichia coli. *EMBO J 17*:4790–4797. doi:10.1093/emboj/17.16.4790

Carroll, P., Muwanguzi-Karugaba, J., Melief, E., Files, M., & Parish, T. (2014). Identification of the translational start site of codon-optimized mCherry in Mycobacterium tuberculosis. *BMC research notes, 7*, 366. https://doi.org/10.1186/1756-0500-7-366

Chao, Y., Li, L., Girodat, D., Förstner, K. U., Said, N., Corcoran, C., … Vogel, J. (2017). In vivo cleavage map illuminates the central role of RNase E in coding and non-coding RNA pathways. *Molecular cell, 65*(1), 39–51. doi:10.1016/j.molcel.2016.11.002

Cortes, T., Schubert, O. T., Rose, G., Arnvig, K. B., Comas, I., Aebersold, R., & Young, D. B. (2013). Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in Mycobacterium tuberculosis. *Cell Reports, 5*(4), 1121–1131. doi: 10.1016/j.celrep.2013.10.031

Dejesus, M. A., Gerrick, E. R., Xu, W., Park, S. W., Long, J. E., Boutte, C. C., … Ioerger, T. R. (2017). Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis. *MBio, 8*(1). doi: 10.1128/mbio.02133-16

Deutscher, M. P. (2015). How bacterial cells keep ribonucleases under control. *FEMS Microbiology Reviews, 39*(3), 350–361. doi: 10.1093/femsre/fuv012

Diwa, A., Bricker, A. L., Jain, C., & Belasco, J. G. (2000). An evolutionarily conserved RNA stem-loop functions as a sensor that directs feedback regulation of RNase E gene expression. *Genes & development*, *14*(10), 1249–1260

Hambraeus, G., Karhumaa, K., Rutberg, B. (2002). A 5′ stem-loop and ribosome binding but not translation are important for the stability of Bacillus subtilis aprE leader mRNA. *Microbiology 148*:1795–1803. doi:10.1099/00221287-148-6-1795

Houben, R. M., & Dodd, P. J. (2016). The global burden of latent tuberculosis infection: A re-estimation using mathematical modelling. *PLoS medicine*, *13*(10), e1002152. doi:10.1371/journal.pmed.1002152

Hui, M. P., Foley, P. L., & Belasco, J. G. (2014). Messenger RNA degradation in bacterial cells. *Annual review of genetics*, *48*, 537–559. doi:10.1146/annurev-genet-120213-092340

Jain, C., & Belasco, J. G. (1995). RNase E autoregulates its synthesis by controlling the degradation rate of its own mRNA in Escherichia coli: unusual sensitivity of the rne transcript to RNase E activity. *Genes & Development, 9*(1), 84–96. doi: 10.1101/gad.9.1.84

Jain, C., Deana, A., & Belasco, J. G. (2002). Consequences of RNase E scarcity in Escherichia coli. *Molecular Microbiology, 43*(4), 1053–1064. doi: 10.1046/j.1365-2958.2002.02808.x

Joanny, G., Derout, J. L., Brechemier-Baey, D., Labas, V., Vinh, J., Regnier, P., & Hajnsdorf, E. (2007). Polyadenylation of a functional mRNA controls gene expression in Escherichia coli. *Nucleic Acids Research*, 35(8), 2494–2502. doi: 10.1093/nar/gkm120

Jurgen, B., Schweder, T., Hecker, M. (1998). The stability of mRNA from the gsiB gene of Bacillus subtilis is dependent on the presence of a strong ribosome binding site. *Mol Gen Genet 258*:538–545. doi:10.1007/s004380050765

Lee, M. H., Pascopella, L., Jacobs, W. R., Hatfull, G. F. (1991). Site-specific integration of mycobacteriophage L5: integration-proficient vectors for *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, and bacille Calmette-Guérin. *ProcNatl Acad Sci USA 88*: 3111–3115

Martini, M. C., Zhou, Y., Sun, H., & Shell, S. S. (2019). Defining the Transcriptional and Post-transcriptional Landscapes of Mycobacterium smegmatis in Aerobic Growth and Hypoxia. *Frontiers in Microbiology*, 10. doi: 10.3389/fmicb.2019.00591

Nilsson, P., & Uhtin, B. E. (1991). Differential decay of a polycistronic Escherichia coli transcript is initiated by RNaseE-dependent endonucleolytic processing. *Molecular Microbiology*, 5(7), 1791–1799. doi: 10.1111/j.1365-2958.1991.tb01928.x

Pato, M. L., Bennett, P. M., von Meyenburg, K. (1973). Messenger ribonucleic acid synthesis and degradation in Escherichia coli during inhibition of translation. *J Bacteriol 116*:710–718. doi:10.1128/JB.116.2.710-718.1973

Plocinski, P., Macios, M., Houghton, J., Niemiec, E., Płocińska, R., Brzostek, A., … Dziembowski, A. (2019). Proteomic and transcriptomic experiments reveal an essential

role of RNA degradosome complexes in shaping the transcriptome of Mycobacterium tuberculosis. *Nucleic Acids Research, 47*(11), 5892–5905. doi: 10.1093/nar/gkz251

Sassetti, C. M., Boyd, D. H., & Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology, 48*(1), 77–84. doi: 10.1046/j.1365-2958.2003.03425.x

Sawyer, E. B., Grabowska, A. D., & Cortes, T. (2018). Translational regulation in mycobacteria and its implications for pathogenicity. *Nucleic acids research, 46*(14), 6950–6961. doi:10.1093/nar/gky574

Schuck, A., Diwa, A., & Belasco, J. G. (2009). RNase E autoregulates its synthesis in Escherichia coli by binding directly to a stem-loop in the rne 5' untranslated region. *Molecular microbiology*, *72*(2), 470–478. doi:10.1111/j.1365-2958.2009.06662.x

Schuster, C. F., & Bertram, R. (2014). Fluorescence Based Primer Extension Technique to Determine Transcriptional Starting Points and Cleavage Sites of RNases In Vivo. *Journal of Visualized Experiments,* (92). doi: 10.3791/52134

Sharp, J. S., Bechhofer, D. H. (2003). Effect of translational signals on mRNA decay in Bacillus subtilis. *J Bacteriol 185*:5372–5379. doi:10.1128/jb.185.18.5372-5379.2003

Shell S. S., Wang J., Lapierre P., Mir M., Chase M. R., Pyle M. M., *et al.* (2015). Leaderless transcripts and small proteins are common features of the mycobacterial translational landscape. *PLoS Genet 11*(11): e1005641. https://doi.org/10.1371/journal.pgen.1005641

Shine, J., & Dalgarno, L. (1974). The 3-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proceedings of the National Academy of Sciences,* 71(4), 1342–1346. doi: 10.1073/pnas.71.4.1342

Sousa, S., Marchand, I., & Dreyfus, M. (2008). Autoregulation allows Escherichia coli RNase E to adjust continuously its synthesis to that of its substrates. *Molecular Microbiology, 42*(3), 867–878. doi: 10.1046/j.1365-2958.2001.02687.x

Taverniti, V., Forti, F., Ghisotti, D., & Putzer, H. (2011). Mycobacterium smegmatis RNase J is a 5′-3′ exo-/endoribonuclease and both RNase J and RNase E are involved in ribosomal RNA maturation. *Molecular Microbiology*, 82(5), 1260–1276. doi: 10.1111/j.1365-2958.2011.07888.x

Nguyen, T. G., Vargas-Blanco, D. A., Roberts, L. A., & Shell, S. S. (2020). The impact of leadered and leaderless gene structures on translation efficiency, transcript stability, and predicted transcription rates in Mycobacterium smegmatis. *Journal of Bacteriology. 10* (4) e00957-19. DOI: 10.1128/JB.00746-19

Wagner, L. A., Gesteland, R. F., Dayhuff, T. J., Weiss, R. B. (1994). An efficient Shine-Dalgarno sequence but not translation is necessary for lacZ mRNA stability in Escherichia coli. *J Bacteriol 176*:1683–1688. doi:10.1128/jb.176.6.1683-1688.1994

World Health Organization (2018). Global Tuberculosis Report 2018. Retrieved from https://www.who.int/tb/publications/global_report/en/

Yajnik V, Godson GN. (1993). Selective decay of Escherichia coli dnaG messenger RNA is initiated by RNase E. *Journal of Biological Chemistry*, 268(18): 13253–13260.

Zeller, M.-E., Csanadi, A., Miczak, A., Rose, T., Bizebard, T., & Kaberdin, V. R. (2007). Quaternary structure and biochemical properties of mycobacterial RNase E/G. *Biochemical Journal, 403*(1), 207–215. doi: 10.1042/bj20061530