Assessing the Effects of mRNA Abundance on

mRNA Degradation Rates in *Mycolicibacterium 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

Authored by:

Adrian Orszulak

and

Van Le

April 28th, 2022

Approved by:

Scarlet S. Shell, PhD

Abstract

Mycobacterium tuberculosis, the etiological agent of tuberculosis, is a difficult pathogen to treat, requiring a lengthy treatment course with numerous antibiotics. It is believed that a robust regulation of gene expression contributes to high tolerance to antibiotics and other stressors. mRNA concentration is one physical factor that may impact mRNA half-life, a contributing factor to overall gene expression. Previous work in *M. tuberculosis* and other bacteria indicates a lack of consensus regarding whether mRNA abundance and mRNA half-life show a strong, negative correlation or a weak, positive correlation. Additionally, mRNA abundance may impact protein abundance in a non-linear fashion. However, there is a lack of consensus regarding the relationship between mRNA abundance and protein abundance. By understanding the impact of mRNA abundance on regulating gene expression, we sought to gain a greater understanding of how *M. tuberculosis* is able to effectively respond to stress. Using a tetracycline-inducible gene expression system in the model organism *Mycolicibacterium smegmatis*, we tested various combinations of concentrations of anhydrotetracycline (aTc) and induction times to determine conditions that would provide the widest range of mRNA and protein expression levels. We established that a range of aTc concentrations from 0 ng/mL to 50 ng/mL at a 4-hour induction time provided a wide range of *gfpmut3* expression. The degradation data were too noisy to determine half-life and make meaningful conclusions regarding the relationship between mRNA abundance and mRNA half-life. Additionally, our system could not be used to investigate the relationship between mRNA abundance and protein abundance due to a loss of inducer-based expression for undetermined reasons.

Introduction

Tuberculosis (TB) is an upper respiratory infection that is a major cause of ill health globally (World Health Organization, 2020). The etiological agent for this disease is the pathogenic bacteria *Mycobacterium tuberculosis* (World Health Organization, 2020). In 2019 alone, 1.4 million deaths were attributed to *M. tuberculosis*, making it one of the leading causes of death by an infectious disease in the world (World Health Organization, 2020). While the number of cases in high-income countries has decreased significantly since the 1940s, *M. tuberculosis* is still a major problem in medium and low-income nations (CDC, 1990; World Health Organization, 2020). Therapies for TB are combinatorial and lengthy, incurring a physical, mental, and economic toll on the individual patient (World Health Organization, 2020). One reason for the lengthy course of treatment is that *M. tuberculosis* inside granulomas is tolerant to antibiotics and a number of stresses, such as hypoxia, nutrient starvation, low pH, and reactive oxygen species (ROS). The mechanisms of tolerance in *M. tuberculosis* stem from a rigorous and well-evolved regulation of gene expression that allows them to survive in such conditions (Reviewed in: Connolly et al., 2007; Reviewed in: Prax, M., & Bertram, R, 2014; Reviewed in: Boldrin et al., 2020).

Within stressful environments, *M. tuberculosis* is able to regulate its gene expression to adapt and persist. Regulation of the gene expression profile in *M. tuberculosis*, like any other bacteria, can occur at a select number of points: transcription of a gene, degradation of mRNA, translation of mRNA into protein, and degradation of the protein (Hausser et al., 2019). The degradation of mRNA is of key interest in studying the regulation of gene expression given the unstable nature of mRNA and the high energy cost of mRNA and protein synthesis within the cell (Pato et al., 1973; Reviewed in: Russel & Cook, 1995; Stouthamer, 1979; Dressaire et al., 2013). Additionally, studies examining the impacts of bacterial transcriptional and posttranscriptional regulatory mechanisms on the half-lives of mRNA in *M. tuberculosis* and *Mycolicibacterium smegmatis*, a model organism for *M. tuberculosis,* show that mRNA half-lives are extended in response to stress (Rustad et al., 2013; Vargas-Blanco et al., 2019). Understanding the mechanisms through which mRNA levels and degradation rates are regulated in response to resource and energy stress will yield a greater knowledge base of information regarding how *M. tuberculosis* is able to effectively respond to stress.

A number of features of mRNAs have been assessed to understand the mechanisms for regulating mRNA degradation in bacteria. These features include stem-loops (Emory & Belasco, 1992), leadered or leaderless gene transcripts (Chen et al., 1991; Nouaille et al., 2017; Nguyen et al., 2020), interaction with regulatory proteins and sRNAs (Arnvig & Young, 2012; Chen et al., 2015; Sinha et al., 2018), RNA-binding proteins (e.g., CsrA, Hfq) (Liu et al., 2010; Timmermans et al., 2010), poly-A tails (O'Hara et al., 1995), and codon content (Lenz, et al., 2011; Boël et al., 2016). In addition to these structural and mechanistic features, many studies have investigated the association between mRNA half-life and mRNA concentration. Bernstein et al. (2002) used DNA microarrays to identify an inverse relationship between mRNA half-lives and mRNA abundance in *E. coli*. Another study in *E.coli showed* a negative association between mRNA

half-lives and mRNA concentration (Esquerré et al., 2015). Using reporter systems with either arabinose-inducible or nisin-inducible promoters in log-phase, Nouaille et al. (2017) observed an inverse correlation between the half-life and concentration of *lacZ* mRNA in *E. coli* as well as *lacLM* mRNA in *Lactococcus lactis*. An inverse correlation between mRNA half-life and mRNA concentration in log-phase *M. tuberculosis* was shown by Rustad et al. (2013). A much weaker inverse correlation between mRNA half-life and mRNA concentration was observed in *M. smegmatis* (Sun et al., manuscript in preparation).

In contrast to studies that showed negative relationships, there were two studies that indicated a weakly positive correlation between mRNA half-life and mRNA abundance in *B. cereus* and *E. coli* through transcriptome-wide measurement of mRNA half-lives using RNA-seq (Kristoffersen et al., 2012; Chen et al., 2015). In a study conducted by Redon et al. (2005), *L. lactis* experiencing carbon starvation showed a positive correlation between mRNA half-life and mRNA concentration. There is a lack of consensus regarding the direction and extent of the correlation between mRNA half-life and mRNA concentration. Furthermore, the causality of this correlation is not well characterized, as concluded by Nouaille et al. (2017). Formulating a complete understanding of the relationship between mRNA half-life and mRNA concentration in mycobacteria would fill a key gap in our understanding of how mRNA degradation occurs and how it is regulated.

In addition, the relationship between mRNA abundance and protein abundance and the relationship's impacts on protein synthesis rate are not well characterized in bacteria. A number of studies have identified factors that affect translation rate and contribute to translational regulation in bacteria. These factors include codon adaptation index (cAI) (Tuller et al., 2010; Riba et al., 2019), tRNA adaptation index (tAI) (Lenz et al., 2010; Riba et al., 2019), 5' UTR (Chen et al., 1991; Kozak et al., 2005; Review: Ren et al., 2017), and Shine-Dalgarno affinity strength (Li et al., 2012; Saito et al., 2020; Tarai & Asai, 2020). The steady-state relationship between mRNA abundance and protein abundance has been studied in *Pseudomonas aeruginosa*, *E. coli*, and *Saccharomyces cerevisiae*. Kwon et al. (2014) found a strong positive correlation between protein abundance and mRNA abundance within two closely related *Pseudomonas aeruginosa* strains in log-phase using DNA microarrays and LC-MS/MS proteomics. Comparing the protein to mRNA ratios between these two strains also produced a positive correlation, indicating that the protein to mRNA ratios are evolutionarily conserved between closely related strains (Kwon et al., 2014). In de Sousa Abreu et al. (2009), a meta-analysis of protein and mRNA abundance in both *E. coli* and *S. cerevisiae* established a weakly positive correlation between protein abundance and mRNA abundance in the two species. Taniguichi et al. (2010) reported no direct correlation between steady-state mRNA concentration and protein concentration using a yellow fluorescent protein translationally fused to the C-terminus of proteins of interest in their native positions within *E. coli*. Similar conclusions were drawn in another study of *E. coli* examining mRNA concentration and protein concentration of a subset of native genes (Lee et al., 2003). Another study that evaluated the mean translation rate per mRNA in single living cells using fluorescence correlation spectrometry (FCS) showed a positive, linear

correlation between the mRNA concentration and protein concentration for dsRED in *E. coli* (Guet et al., 2008). There is a lack of consensus regarding the exact nature of the steady-state relationship of mRNA and protein concentration. These differences could be the result of different methods being used to assess and analyze the relationship between mRNA abundance and protein abundance. For example, most studies examined large numbers of different transcripts and proteins expressed at their native levels, while some examined a single transcript and protein expressed at different levels. Understanding this relationship further would address whether the efficiency of mRNA translation is affected by the mRNA concentration. Additionally, this relationship has not been assessed in *M. smegmatis*. Addressing this goal in *M. smegmatis* could yield information regarding the relationship between mRNA and protein concentration in the context of *M. tuberculosis*.

Guided by previous research, this study sought to answer questions related to understanding the impact of mRNA concentration on gene expression. First, we sought to further investigate the nature of the negative correlation between mRNA half-life and mRNA concentration by examining the impact of transcription rate on mRNA stability in *M. smegmatis*. Second, we sought to examine the relationship between the protein concentration and mRNA abundance in *M. smegmati*s. To accomplish these goals, a set of *M. smegmatis* strains were constructed containing a tet-ON inducible system for temporal regulation of fluorescent proteins. Due to high variability and noise between biological replicates observed in the degradation data, we were unable to determine a meaningful relationship between mRNA abundance and mRNA half-life. In addition, our tetracycline-inducible gene expression system could not be used to investigate the relationship between mRNA abundance and protein abundance as a result of a decrease in inducer-based protein expression for undetermined reasons.

Materials and Methods

Strains and culture conditions

$Mycolicibacterium$ $smegmatis$ mc^{2}155 strain and all constructed strains were grown in Difco^{TM} Middlebrook 7H9 medium with albumin dextrose catalase (ADC; final concentrations: 5 g/L bovine serum albumin fraction V (BSA), 2 g/L dextrose, 0.85 g/L NaCl, and 3 mg/L catalase), 0.2% glycerol, and 0.05% Tween 80. Cultures were shaken at 200 rpm and 37°C to an optical density at 600 nm (OD$_{600}$) between 0.5 to 0.8 at the time of harvest. Cultures grown and induced with anhydrotetracycline (aTc) were wrapped in aluminum foil to protect the photosensitive inducer.

Plasmid construction

Plasmid pSS303 (Nguyen et al., 2020) was used as a vector backbone and the yfp coding sequence in that plasmid was replaced with either a $gfpmut3$ or $mCherry$ fluorescent reporter gene to create pSS470 and pSS471 (Table 1). The plasmid contained a strong constitutive P$_{myc1}$ $2X$ $tetO$ promoter , which was repressed in the presence of tet repressor protein (TetR) and initiated with the addition of aTc, an antibiotic derivative of tetracycline, with an associated P$_{myc1}$ 5' UTR (Blokpoel et al., 2005; Carroll et al., 2005; Ehrt et al., 2005). Two synthetic terminators were present in the vector; $tsynA$ (Czyz et al., 2014) was upstream of the reporter gene, while $ttsbiB$ (Huff et al., 2010) was downstream of the reporter gene. A 6×Histidine tag was added at the C terminus for the GFPmut3 protein (complete amino acid sequence:MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKL PVPWPTLVTTFGYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAE VKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNI EDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGIT HGMDELYKCHHHHHH) and the mCherry protein (complete amino acid sequence:MAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFA WDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGH YDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELY KHHHHHH). The plasmids were integrated into the Giles phage site using an integrase promoter and protein sequence. Additionally, a tet $repressor$ gene along with its strong constitutive P$_{myc1}$ promoter and associated P$_{myc1}$ 5' UTR were used in conjunction with the pSS221 vector backbone to create pSS555. The plasmid was integrated into the L5 phage site using an integrase promoter and protein sequence.

All constructs were built using NEBuilder HiFi DNA assembly master mix (catalogue E2621) (Table 1). To create $M.$ $smegmatis$ mc^{2}155 strain with both $gfpmut3$ and $tetR$ genes, pSS470 was transformed using electroporation (Bio-Rad, catalogue 1652100) into $M.$ $smegmatis$ mc^{2}155 strain and integrated at the Giles phage site and selected with 250 µg/mL Hygromycin B. pSS555 was later transformed using electroporation and integrated in the same strain at the L5 phage site and selected with 40 µg/mL nourseothricin (Table 1). These steps were repeated to

create *M. smegmatis* mc^2_155 strain with both *mCherry* and *tetR* genes (Table 1). Additionally, the previously mentioned steps were used to create *M. smegmatis* mc^{2}155 strains with only pSS470, pSS471, or pSS555 (Table 1). To confirm successful plasmid integrations at Giles and L5 phage sites, several different primers were used (Table 2 and Table 3).

Table 1. A list of the *Mycolicibacterium smegmatis* strains and plasmids. *HygR* refers to a gene that produces a protein which confers resistance to Hygromycin B. *NAT* refers to a gene that produces nourseothricin N-acetyl transferase which confers resistance to nourseothricin.

Strain	Plasmid	Plasmid Description
SS-M_0836; SS-M_1097; SS-M_1098; SS-M_1099	pSS470	P$_{myc1}$ *2X tetO* promoter + P$_{myc1}$ 5′ UTR+ *gfpmut3-6xHis* + *HygR*
SS-M_0837; SS-M_0838	pSS471	P$_{myc1}$ *2X tetO* promoter + P$_{myc1}$ 5′ UTR+ *mCherry-6xHis* + *HygR*
SS-M_1062; SS-M_1063	pSS470 + pSS555	P$_{myc1}$ *2X tetO* promoter + P$_{myc1}$ 5′ UTR+ *gfpmut3-6xHis* + *HygR* (Giles site) and P$_{myc1}$ promoter + P$_{myc1}$ 5′ UTR+ *tetR* + *NAT* (L5 site)
SS-M_1073; SS-M_1074	pSS471 + pSS555	P$_{myc1}$ *2X tetO* promoter + P$_{myc1}$ 5′ UTR+ *mCherry-6xHis* + *HygR* (Giles site) and P$_{myc1}$ promoter + P$_{myc1}$ 5′ UTR+ *tetR* + *NAT* (L5 site)
SS-M_1071; SS-M_1072	pSS555	P$_{myc1}$ promoter + P$_{myc1}$ 5′ UTR+ *tetR* + *NAT*

Table 2. Primers used for plasmid construction, colony-checking PCR, and verifying plasmid integration into the Giles and L5 sites. *HygR* refers to a gene that produces a protein which confers resistance to Hygromycin B. *NAT* refers to a gene that produces nourseothricin N-acetyl transferase, which confers resistance to nourseothricin. Forward primers are denoted with an 'F' character whereas reverse primers are denoted with an 'R' character.

Plasmid	Primers for Amplification and Plasmid Creation	Primers for Colony-Checking PCR for fidelity	Primers for verifying plasmid integration
pSS470	Insert Amplification: SSS1992F,SSS1993R Vector Amplification: SSS1989F,SSS1800R	SSS1171F, SSS1851R	Giles Left integration: SSS1172F, SSS1174R Giles Right integration: SSS1173F, SSS1175R
pSS471	Insert Amplification: SSS1990F,SSS1991R Vector Amplification: SSS1989F,SSS1800R	SSS1171F, SSS1851R	Giles Left integration: SSS1172F, SSS1174R Giles Right integration: SSS1173F, SSS1175R
pSS555	Insert Amplification: SSS2159F,SSS2205R Vector Amplification SSS1488R,SSS1519F	SSS2315F, SSS1641R	L5 Left integration: SSS1103F, SSS142R L5 Right integration: SSS1104F, SSS144R

Table 3. Primer sequences. Listed in the table below are the sequences of primers listed in Table 2.

Primer	Description	Sequence (5' → 3')
SSS1992	Forward primer to amplify *gfpmut3* in pMV762	TTAAGAAGGAGATATACATCATGAGTAAAGGAGAAGAAC
SSS1993	Reverse primer to amplify *gfpmut3* in pMV762	GTGATGGTGATGGTGATGACATTTGTATAGTTCATCCATGC
SSS1990	Forward primer to amplify *mCherry* in pSS374	TTAAGAAGGAGATATACATCATGGCCATCATCAAGGAGTTC
SSS1991	Reverse primer to amplify *mCherry* in pSS374	TGATGGTGATGGTGATGACACTTGTACAGCTCGTCCATGC
SSS1989	Forward primer to amplify pSS303	TGTCATCACCATCACCAT
SSS1800	Reverse primer to amplify pSS303	GATGTATATCTCCTTCTTAAT
SSS2159	Forward primer to amplify P_{myc1} promoter + P_{myc1} 5′ UTR+ *tetR* in pSS221	CAAACTCTTCCTGTCGTCATATAGAAATATTGGATCGTCGG
SSS2205	Reverse primer to amplify P_{myc1} promoter + P_{myc1} 5′ UTR+ *tetR* in pSS221	GTTAACTACGTCGACATCGATATTAAGACCCACTTTCACATTTAAG
SSS1519	Forward primer to amplify pJEB402	TATCGATGTCGACGTAGTTAAC
SSS1488	Reverse primer to amplify pJEB402	ATATGACGACAGGAAGAGTT
SSS1171	Forward primer to amplify *gfpmut3* or *mCherry* for colony-checking PCR	GGAAAAGAGGTCATCCAGGA
SSS1851	Reverse primer to amplify *gfpmut3* or *mCherry* for colony-checking PCR	GGCAACGCCCTAGTGATGGTGATGGTGATGAC

SSS2315	Forward primer to amplify P_{myc1} promoter + P_{myc1} 5′ UTR+ partial *tetR* for colony-checking PCR	CGGTGAACGCTCTCCTG
SSS1641	Reverse primer to amplify P_{myc1} promoter + P_{myc1} 5′ UTR+ partial *tetR* for colony-checking PCR	TAGGCTGCTCTACACCA AGC
SSS1172	Forward primer to check left junction in Giles site in *M. smegmatis*	CTCCGAACTCCTCCGAA ACC
SSS1173	Forward primer to check right junction in Giles site in *M. smegmatis*	ACATATCTGTCGAAGCG CCC
SSS1174	Reverse primer to check left junction in Giles site in *M. smegmatis*	TGACGATCAACTCCGCG GGGCCGGGCCA
SSS1175	Reverse primer to check right junction in Giles site in *M. smegmatis*	CGGTGGATCCGCGCAA CCTG
SSS1103	Forward primer to check left junction in L5 site in *M. smegmatis*	TGGATTTGGTTTCAGCT CCC
SSS142	Reverse primer to check left junction in L5 site in *M. smegmatis*	TAGAGCCGTGAACGAC AGG
SSS1104	Forward primer to check right junction in L5 site in *M. smegmatis*	GACCTTGGTGCAGAAAT CGC
SSS144	Reverse primer to check right junction in L5 site in *M. smegmatis*	TCGATGAGCCGCTTCTC GC

Polymerase Chain Reaction (PCR) and DNA Recovery

Polymerase Chain Reaction (PCR) was performed in a 25 µL sample reaction volume. When using Q5, high-fidelity polymerase, the following volumes of necessary components were added to each reaction: 5.0 µL of 5X Q5 polymerase buffer (New England Biolabs, catalogue B9027S), 5.0 µL of 5X GC enhancer (New England Biolabs, catalogue B9208A), 0.50 µL of 10 µM forward primer, 0.50 µL of 10 µM reverse primer, 0.5 µL of 10 µM dNTPs, 0.25 µL of Q5 polymerase (New England Biolabs, catalogue M0491L), and 1.0 µL of the DNA to be amplified. UltraPure™ DNase/RNase-free distilled water was added for the remaining volume. When using Taq polymerase, the following volumes of necessary components were added to each reaction: 2.5 µL of 10X Taq polymerase buffer (New England Biolabs, catalogue B9014S), 1.0 µL of DMSO, 0.5 µL of 10 µM forward primer, 0.5 µL of 10 µM reverse primer, 0.5 µL of 10 µM dNTPs, 0.125 µL of Taq polymerase (New England Biolabs, catalogue M0273L), and approximately 5.0 µL of colony for colony-checking PCR. UltraPure™ DNase/RNase-free distilled water was added for the remaining volume. The forward and reverse primers for each strain can be found in Table 2. PCR was carried out at an initial denaturation step at 95℃ for 2 minutes, (i) a denaturation step of 95℃ for 20 seconds, (ii) an annealing step at an appropriate primer annealing temperature (℃) for 40 seconds, (iii) an elongation step at 72℃ for 2 minutes, and a final elongation step at 72℃ for 5 minutes. The steps labeled i, ii, and iii were repeated 35 times. Annealing temperatures were optimized for each primer set using the New England Biolabs Tm calculator. Additionally, the elongation time was based on the size of the PCR product following the precedent of 1 minute/kilobase.

All products were analyzed by gel electrophoresis using a 1.0% agarose gel with 0.2-0.5 µg/mL of ethidium bromide (EtBr) (depending on the mass of the gel) prepared in 1X Tris-acetate-EDTA (TAE) buffer. Gels were visualized using UV-light (Bio-Rad) and Quantity One software. Bands of interest were cut from the gel and purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, catalogue D4002) following the manufacturer's instructions. A NanoDrop One (Thermo Scientific) was used to measure sample concentrations of DNA. Any DNA sample intended for sequencing was sent to QuintaraBio following the company's instructions.

Flow cytometry

Cultures of *M. smegmatis* were grown in duplicate at a variety of induction times (1 hour, 4 hours, 24 hours, 26 hours, 28 hours, and 30 hours) and aTc induction concentrations (0 ng/mL, 1 ng/mL, 2 ng/mL, 5 ng/mL, 10 ng/mL, 20 ng/mL, 50 ng/mL, 100 ng/mL, and 200 ng/mL) at OD_{600} between 0.5 and 1.3. Harvested cultures were placed on ice, and diluted to form two, 1.0 mL cultures, which served as duplicates, with an OD_{600} of 0.025 freshly filtered 7H9 media and then filtered with a 5 µm filter needle to remove clumps . A CytoFlex flow cytometer was used to measure 5000 events of each culture with one universal gate drawn to encompass the densest region of cells on a forward scatter (FSC) vs violet side scatter (SSC) plot. The gain values were

500 for FSC and 50 for violet SSC. The thresholds were 100,000 for violet SSC-H and 40,000 for FSC-H. FlowJov10.8.0 was utilized to draw gates and analyze fluorescence data.

RNA extraction

RNA extractions were performed on triplicate cultures. Cell cultures were induced with one of the following aTc concentrations for 4 hours: 0 ng/mL, 2.5 ng/mL, 5 ng/mL, 10 ng/mL, 20 ng/mL, or 50 ng/mL. The cultures were then spun down to form a cell pellet. Any pellets that were not used immediately for extraction were stored at -80 °C and were thawed on ice before extraction. Cells were resuspended in 1.0 mL of TRIzol Reagent (VWR, catalogue MSPP-TR118), and pipetted into a 2.0 mL beating tube (OPS Diagnostics; 100-μm zirconium lysing matrix, molecular grade). The cells were then lysed using a FastPrep-24 5G instrument (MP Biomedical) (3 cycles of 7 m/s for 30 s, with 2 min on ice between cycles). Samples were treated with 300 μL of chloroform before being centrifuged for 15 minutes at 15,000 rpm and 4°C. The resulting aqueous layer was recovered from the sample, and extraction was completed using Direct-Zol^{TM} RNA miniprep (Zymo Research, catalogue R2052) following the manufacturer's instructions with an in-column DNase treatment. Sample concentrations and absorbance ratios were measured using a NanoDrop One (Thermo Scientific) before being stored at -80°C. RNA quality was assessed by gel electrophoresis. A volume containing 300 ng of extracted RNA was mixed with 2X RNA loading dye (New England Biolabs, catalogue 50-427-9) and heated at 65°C for 5 minutes using a heat block. Heated samples were loaded onto a 1.0% agarose gel with 0.2-0.5 μg/mL of EtBr (depending on the mass of the gel), and run with 1X Tris/Borate/EDTA (TBE) buffer. Gels were visualized using UV-light (Bio-Rad) and Quantity One software.

cDNA synthesis and cleanup

Each solution of extracted RNA was diluted into two, separate 5.25 μL, samples containing 600 ng of RNA using UltraPure™ DNase/RNase-free distilled water, one used as a negative control without reverse transcriptase (no RT) and one with reverse transcriptase (RT). A volume of 1.0 μL mix, containing 0.83 μL of 100 mM Tris, pH 7.5, and 0.17 μL of 3.0 mg/mL of random primers (Invitrogen: Catalogue No. 48190011), was pipetted to each of the diluted RNA samples. Samples were incubated at 70°C for 10 minutes before being snap-frozen in an ice water bath for 5 minutes and then transferred onto ice. A 3.75 μL mix was pipetted to the RT samples containing the following components: 2 μL of ProtoScript II RT Reaction Buffer (NEB: Catalogue No. B0368S), 0.5 μL of 10 mM each dNTPs, 0.5 μL of 100 nM DTT (NEB: Catalogue No. B1034A), 0.25 μL of RNase Inhibitor, Murine, New England Biolabs (40,000 U/mL), and 0.5 μL of ProtoScript® II Reverse Transcriptase, New England Biolabs (200,000 U/mL). A 3.75 μL mix was pipetted to the no RT samples containing the following components: 2 μL of ProtoScript II RT Reaction Buffer, 0.5 μL of 10 mM each dNTPs, 0.5 μL of 100 nM DTT, 0.25 μL of RNase Inhibitor, Murine, New England Biolabs (40,000 U/mL), and 0.5 μL of UltraPure™ DNase/RNase-free distilled water. Samples were incubated at 25°C for 10 minutes followed by 42°C for 2 hours. Samples were treated with 10 μL of master mix containing 5 μL of

0.5 M of EDTA and 5 µL of 1N NaOH to degrade any remaining RNA. Samples were then incubated at 65°C for 15 minutes before 12.5 µL of Tris HCl, pH 7.5, was added to neutralize the pH. All cDNA samples were cleaned up using the MinElute PCR Purification Kit (NEB #T1030L) following the manufacturer's instructions.

Quantitative PCR (qPCR)

All qPCR reactions were performed in a BSL-2 approved biosafety cabinet. qPCR was performed on 3 biological replicates. All samples of cDNA were obtained from RNA extraction and cDNA synthesis and cleanup as described above. Samples of cDNA were diluted firstly to 1.0 ng/µL from their original concentration in pure water. Those samples were then diluted to the desired concentration of 200 pg/µL. A 2.5 µM primer mix for *sigA* was created with a 1.12 sample error. The final primer mix contained 2.5 µM of JR273 and 2.5 µM of JR274 (Table 4). A similar 2.5 µM primer mix was created for *gfpmut3* resulting in a final primer mix with 2.5 µM of SSS306 and 2.5 µM of SSS308 (Table 4). For each of the targets, a mastermix was created using 1 µL of the appropriate primer mix, 2 µL of ultra-pure water, and 5 µL of iTaq Universal SYBR Green Supermix (Bio-Rad: Catalogue No. 172-5124). Mastermixes were kept on ice until use. Within a 96-well PCR microplate (Axygen™: Catalogue No. 14-222-334), 2 µL of cDNA was pipetted into each well. Once all wells were filled with cDNA, 8 µL of the appropriate mastermix for *sigA* or *gfpmut3* respectively was added to each well and mixed carefully using a micropipette. Water controls were made with the specific mastermixes but replacing the 2 µL of cDNA with 2 µL of ultra-pure water. The plate was then covered with a sealing film (Axygen: Catalogue No. UC-500). qPCR was run in a QuantStudio 6 Pro Real-Time PCR System (QuantStudio: Catalogue No. A43180). Samples in the microplate were incubated at (i) 50°C for 2 minutes, (ii) a 95°C for 1 minute, (iii) 95°C hold for 15 seconds, and (iv) 61°C hold for 2 minutes. In the final step (iv), the fluorescence of SYBR Green was recorded. Steps (iii) and (iv) were cycled through 40 times. To obtain information regarding transcript abundance and relative expression, the *gfpmut3* gene was normalized to the *sigA* housekeeping gene. For each sample, the number of cycles (C_t) of a gene of interest required to pass a threshold (C_T) of 0.1 was compared to *sigA*. The difference was used to calculate the ΔC_t for each sample. Relative expression was calculated as $2^{-\Delta Ct}$. Analysis and figure creation was completed using GraphPad Prism 9.

Table 4. Primers used for qPCR.

Primer	Description	Sequence (5' → 3')
JR273	Forward primer to amplify *sigA* cDNA in *M. smegmatis*	GACTACACCAAGGGCTACAAG
JR274	Reverse primer to amplify *sigA* cDNA in *M. smegmatis*	TTGATCACCTCGACCATGTG
SSS306	Forward primer to amplify *gfpmut3* cDNA	GAAGGTGATGCAACATACGG
SSS308	Reverse primer to amplify *gfpmut3* cDNA	TCCTGTACATAACCTTCGGG

Rifampicin Experiment for mRNA Half-Life Determination

Thirty-two mL cultures of *M. smegmatis* were grown to an OD_{600} of 0.5 and 0.7 in 250 mL Erlenmeyer flasks for each concentration of aTc for each biological replicate. Four hours before the step of adding rifampicin, aTc was added to achieve final concentrations of 0, 2.5, 5, 10, 20, 50 ng/mL. Three and a half hours later, 5 mL of bacterial culture was aliquoted from each flask into 5, 15 mL conical tubes, for five time points (0, 0.5 minute, 1 minute, 2 minutes, and 4 minutes). The conical tubes were placed on the tissue culture rotator and spun for 30 minutes. Cultures were then treated with rifampicin at a final concentration of 150 ug/mL to halt transcription and snap-frozen in liquid nitrogen after 0, 0.5 minute, 1 minute, 2 minutes, and 4 minutes. The cultures were then stored at -80 degrees Celsius for future RNA extractions, which was followed by cDNA synthesis and qPCR.

Transcript abundance of *sigA* and *gfpmut3* were used to determine mRNA half-lives. For each gene, the C_t was made negative, which represents transcript abundance on a $\log_2$ scale, and linear regression was performed on a plot of the negative C_t versus time using GraphPad Prism 9. Half-life was defined as the negative reciprocal of the best-fit slope (Equation 1).

$$\text{Half-life} = -\frac{1}{slope} \tag{1}$$

As seen previously in the context of mycobacteria, plotting $\log_2$ abundance over time produces a biphasic decay curve with a period of faster exponential decay followed by a period of much slower or undetectable exponential decay (Nguyen et al., 2020). Other studies in *E. coli* have observed similar biphasic curves for a variety of different genes (Blum et al., 1999; Brescia et al., 2004; Chen et al., 2015; Sinha et al., 2018).

Results & Discussion

Constructing *Mycolicibacterium smegmatis* strains with aTc-inducible GFPmut3 and mCherry

To determine the degree of correlation between mRNA half-life and concentration and whether transcription is causal in that relationship, we needed to measure mRNA half-life at a variety of mRNA concentrations for a single gene in *Mycolicibacterium smegmatis*. An anhydrotetracycline (aTc) induced expression system was selected to create a set of *M. smegmatis* strains given its strict temporal regulation of gene expression (Ehrt et al., 2005). We constructed three plasmids: one containing the two *tet-operator* P_{myc1} promoter and the associated P_{myc1} 5' UTR linked to a *gfpmut3* gene, one containing the two *tet-operator* P_{myc1} promoter with the associated P_{myc1} 5' UTR linked to an *mCherry* gene and another containing the same promoter and 5' UTR without the tet operators, linked to a *tet repressor* gene (Figure 1A, Figure 1B, and Figure 1C). We conducted fluorescence microscopy to verify the expected protein expression in the experimental strains containing *gfpmut3 + tetR*, *mCherry + tetR* and control strains with *gfpmut3* only and *tetR* only to validate that the system performed as expected in the presence and absence of aTc. Strains were induced with 0 ng/mL and 200 ng/mL of aTc and incubated for 24 hours for the *gfpmut3* only strain, the *tetR* only strain, and the *gfpmut3 + tetR* strain. Strains were induced with 0 ng/mL and 200 ng/mL of aTc and incubated for 4 hours for the *mCherry* only strain, the *tetR* only strain, and the *mCherry + tetR* strain. The *gfpmut3* and *mCherry* only strains fluoresced brightly in the presence and absence of aTc as expected (Figure 1C and Figure 2C). The *tetR* only strain did not fluoresce in either the presence and absence of aTc (Figure 1D and Figure 2D). The experimental strains containing *gfpmut3 + tetR* or *mCherry + tetR* only showed fluorescence in the presence of aTc (Figure 1E and Figure 2E). From these results, the performance of the tetracycline-inducible system was validated.
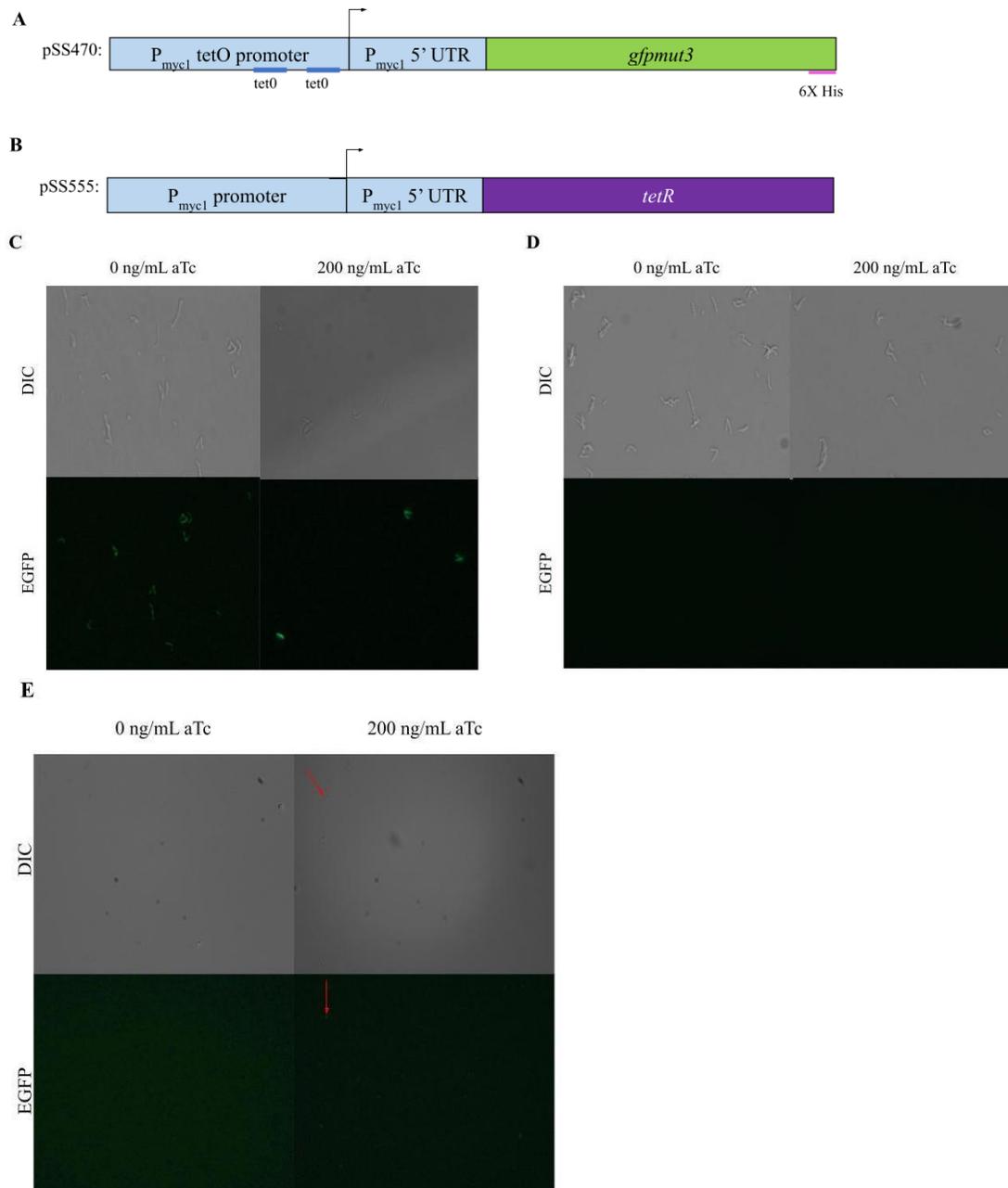
Figure 1. Constructing the aTc inducible system for GFPmut3 protein expression. **(A)** Schematic of the plasmid with the P_{myc1} promoter with associated P_{myc1} 5' UTR and two *tet-operator* regions and the *gfpmut3* gene inserted into the *M. smegmatis* Giles site **(B)** Schematic for the plasmid with the P_{myc1} promoter with associated P_{myc1} 5' UTR *tetR* gene inserted into the *M. smegmatis* L5 site **(C)** Fluorescence microscopy of *M. smegmatis* strains that contained a P_{myc1} promoter with associated P_{myc1} 5' UTR and two *tet-operator* regions linked *gfpmut3* gene inserted into the *M. smegmatis* Giles site **(D)** Fluorescence microscopy of *M. smegmatis* strains that contained a P_{myc1} promoter with associated P_{myc1} 5' UTR linked *tetR* gene **(E)** Fluorescence microscopy of *M. smegmatis* strains that contained a P_{myc1} promoter with associated P_{myc1} 5' UTR and two *tet-operator* regions linked *gfpmut3* gene inserted into the *M. smegmatis* Giles site and a P_{myc1} promoter with associated P_{myc1} 5' UTR linked *tetR* gene inserted into the *M. smegmatis* L5 site. Red arrows point out fluorescing cells.
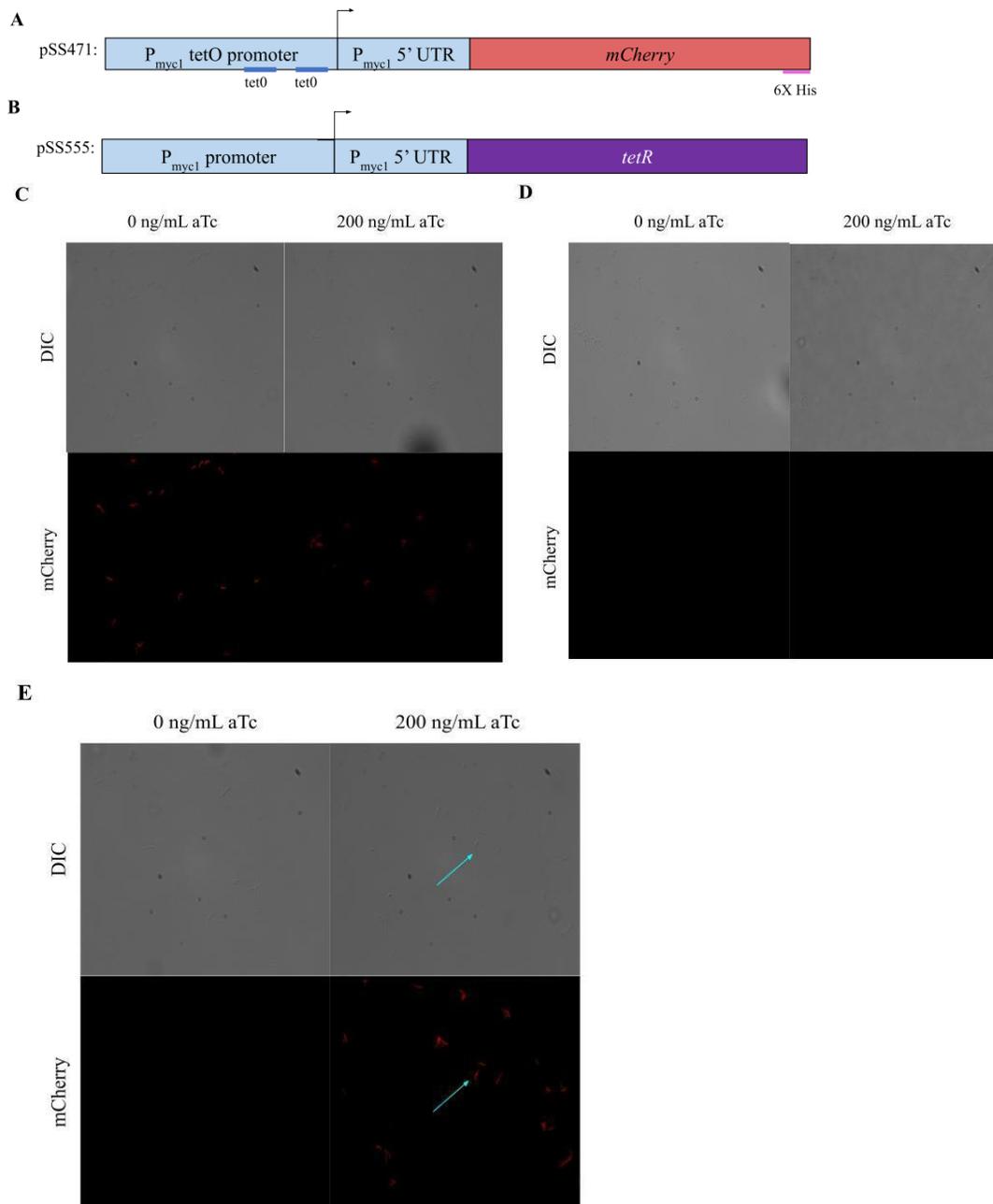
Figure 2. Constructing the aTc inducible system for mCherry protein expression. **(A)** Schematic of the plasmid with the P_{myc1} promoter with associated P_{myc1} 5' UTR and two *tet-operator* regions linked *mCherry* gene inserted into the *M. smegmatis* Giles site **(B)** Schematic of the plasmid with the P_{myc1} promoter with associated P_{myc1} 5' UTR *tetR* gene inserted into the *M. smegmatis* L5 site **(C)** Fluorescence microscopy of *M. smegmatis* strains that contained a P_{myc1} promoter with associated P_{myc1} 5' UTR and two *tet-operator* regions linked *mCherry* gene inserted into the *M. smegmatis* Giles site **(D)** Fluorescence microscopy of *M. smegmatis* strains that contained a P_{myc1} promoter with associated P_{myc1} 5' UTR linked *tetR* gene **(E)** Fluorescence microscopy of *M. smegmatis* strains that contained a P_{myc1} promoter with associated P_{myc1} 5' UTR and two *tet-operator* regions linked *mCherry* gene inserted into the *M. smegmatis* Giles site and a P_{myc1} promoter with associated P_{myc1} 5' UTR linked *tetR* gene inserted into the *M. smegmatis* L5 site. Blue arrows point out fluorescing cells.

17

Four hours induction generates the widest range of GFPmut3 expression for many aTc induction concentrations

Studies have reported mixed data regarding the relationship between steady-state mRNA and protein concentration (Lee et al., 2003; Guet et al., 2008; de Sousa Abreu et al., 2009; Taniguchi et al., 2011; Kwon et al., 2014). This relationship is not well-studied in the context of *M. smegmatis*. Using this tetracycline-inducible gene expression system, it was a goal of this study to understand the steady-state relationship between mRNA abundance and protein abundance in *M. smegmatis*. Guided by previous work in *M. smegmatis* and *Mycobacterium tuberculosis*, we found that the aTc concentrations of 0, 1, 5, 10, 20, 50, 100, 200 ng/mL were the most used and applicable to generate our desired wide range of mRNA levels (Sinha et al., 2007; Raghavan et al., 2008; Korch et al., 2009; Goyal et al., 2011; Minch et al., 2012). To determine the appropriate induction time with those aTc concentrations that would provide us with a wide range of GFPmut3 protein expression, we conducted flow cytometry after three different induction times: 1 hour, 4 hours, and 24 hours. After one hour or 24 hours of induction, samples exposed to higher aTc concentrations were fluorescent but those exposed to lower aTc concentrations did not consistently have above-background fluorescence (Figure 3A and 3C). In contrast, the 4 hour induction time point showed a greater range of GFPmut3 fluorescence (Figure 3B). Strains were also induced with the noted aTc concentrations and incubated for 24 hours, 26 hours, 28 hours, and 30 hours, to examine the potential of increasing the spread of GFPmut3 expression levels. Median fluorescence levels of GFPmut3 decreased as the incubation time increased, indicating a decrease in protein expression levels (Figure 4). From these results, we determined that 4 hours induction was the best time point that would allow us to establish the correlation between mRNA abundance and mRNA half-life.

Due to loss of aTc induced expression over time, we were not able to study the steady-state relationship between mRNA abundance and protein abundance in *M. smegmatis* in this paper. Instead, our results shed light on the use of a tetracycline-inducible gene expression system for studying the expression of fluorescent proteins and mRNA. When examining induction times greater than 24 hours, GFPmut3 expression decreased as induction time increased (Figure 4). This decrease in expression is most likely due to a loss of aTc. The inducer aTc has been shown to be temperature sensitive and especially photosensitive in L-broth (LB) and M9 media (Ehrt, et al., 2005; Politi et al., 2014; Baumschlager, et al., 2020). One study that worked to understand aTc degradation in the context of *E. coli* grown in LB and M9 through a linear model attributed nearly 42.5% of degradation to residual error unexplained with their linear model while 41.9% of degradation was due to temperature (Politi et al., 2014). Long-term exposure to higher temperatures, such as those used to grow bacterial cultures, appears to be a significant factor that accounts for aTc degradation (Politi et al., 2014). To better improve and use aTc in future *M. tuberculosis* and *M. smegmatis* experiments, a similar study as Politi et al. (2014) could be conducted in 7H9 media and at 37°C. Politi et al. (2014) also assessed the half-lives of other inducers beside aTc including IPTG and HSL and found that IPTG has the highest half-life out of three inducers and is stable over 32 hours. IPTG could be considered as a

potential inducer for future experiments, especially regarding longer induction times. Long GFPmut3 protein half-lives and loss of aTc induced expression (Figure 4) indicates that steady-state *gfpmut3* mRNA expression (see below) does not coincide with steady-state GFPmut3 protein expression.
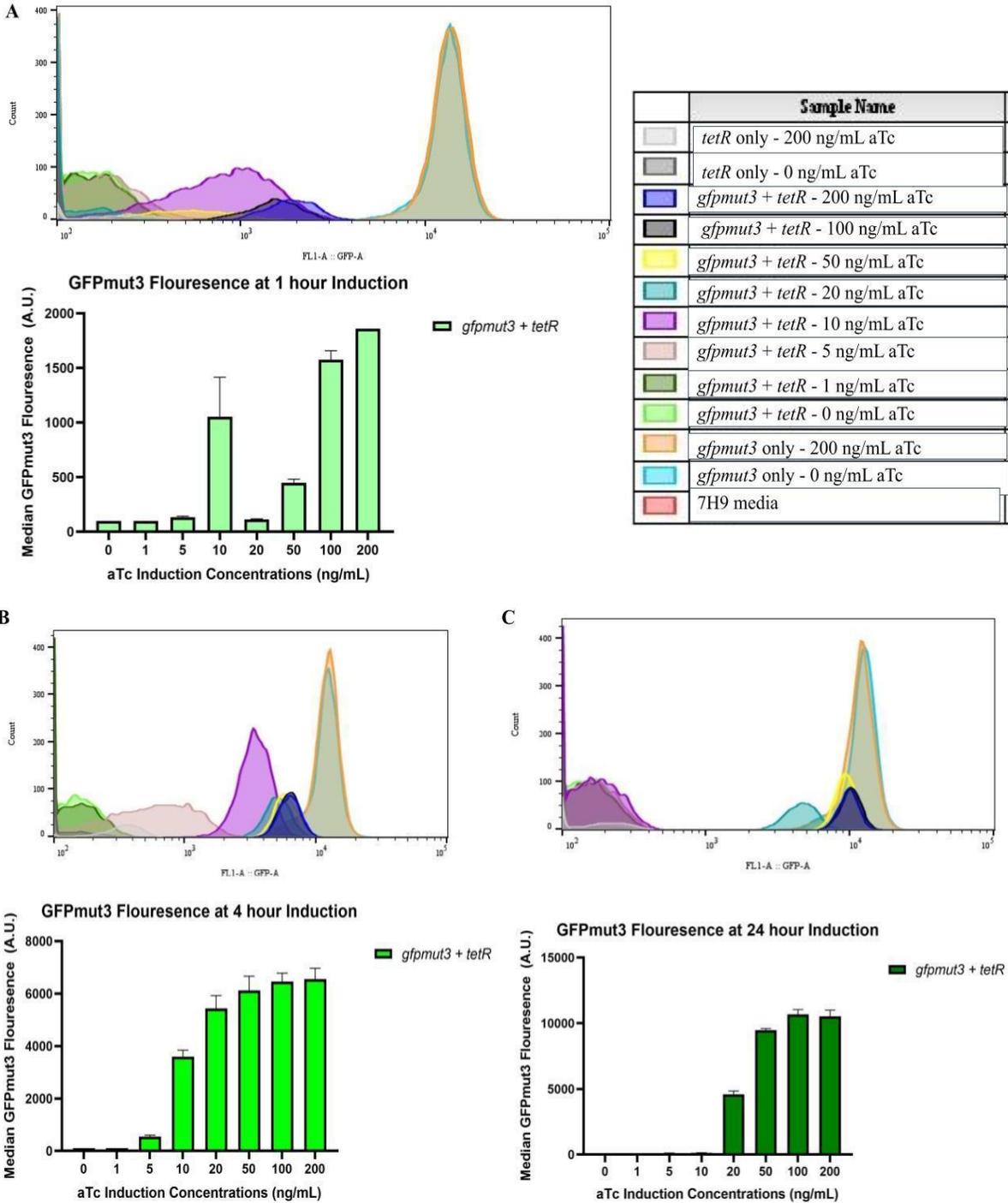
Figure 3. Validating the use of a 4 hour aTc induction time point for greatest spread of protein expression levels. Samples of the strains containing *gfpmut3 + tetR* were incubated at 0 ng/mL, 1 ng/mL, 5 ng/mL, 10 ng/mL, 20 ng/mL, 50 ng/mL, 100 ng/mL, or 200 ng/mL of aTc for different induction times. Samples containing *gfpmut3* only, and *tetR* only were incubated at 0 ng/mL and 200 ng/mL of aTc at different induction times. The fluorescence histogram from one duplicate is shown. The mean of median fluorescence of GFPmut3 expression of the duplicate samples and standard deviation were quantified using FlowJo. **(A)** 1-hour induction **(B)** 4-hours induction **(C)** 24-hours induction.
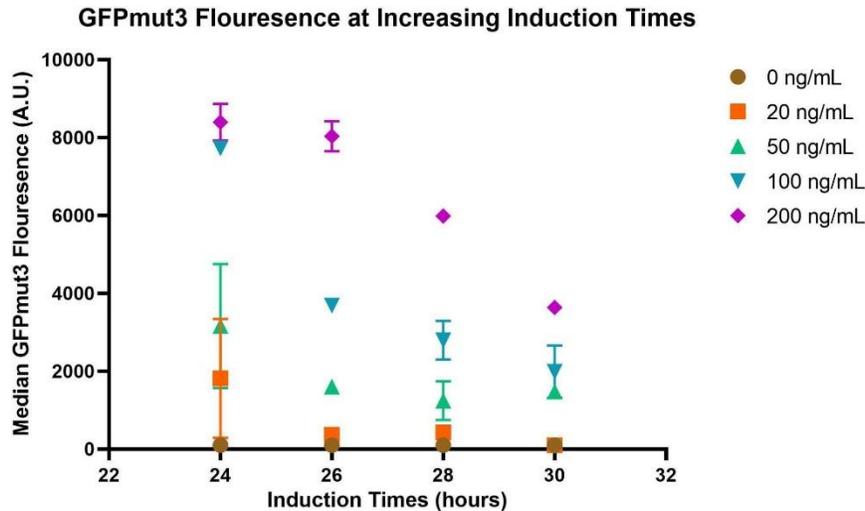
Figure 4. Induction times above 24 hours led to a decrease in GFPmut3 fluorescence. Samples of the strains containing *gfpmut3 + tetR* were incubated at 0ng/mL, 20 ng/mL, 50 ng/mL, 100 ng/mL, or 200 ng/mL of aTc for 24 hours, 26 hours, 28 hours, and 30 hours. Fluorescence was quantified by flow cytometry. The mean of median fluorescence of GFPmut3 expression of the duplicate samples and standard deviation were quantified using FlowJo.

TetR-controlled mCherry acts like an on-off inducible switch system for the variety of aTc concentrations

The initial goal of this research was to engineer plasmids with similar constructs but with two different genes, either *gfpmut3* or *mCherry*, to see if our findings held true for different genes. Based on the results of flow cytometry with the *gfpmut3* strains (Figure 3), we chose to test the *mCherry* strains after 4 and 24 hours of induction using the aTc concentrations of 0, 5, 10, 20, 50, 100, and 200 ng/mL. The 1-hour induction was excluded from testing the *mCherry* strains because we found that 1-hour induction was not long enough to give a wide range of GFPmut3 protein levels in the *gfpmut3 + tetR* strains (Figure 3A). In contrast to the *gfpmut3 + tetR* results, 4-hours induction was not long enough to allow strains containing *mCherry + tetR* with low concentrations from 5 ng/mL to 50 ng/mL of aTc to produce expression greater than the negative control strains (Figure 5A). In addition, there was a large difference between the fluorescence of samples induced with 100 ng/mL and 200 ng/mL of aTc compared to samples induced with 50 ng/mL of aTc. The *mCherry + tetR* induction system therefore seemed to behave as an on-off inducible switch system instead of a titratable system as seen in Figure 3B where a wide range of GFPmut3 fluorescence was observed. The 24 hours induction compared to 4 hours induction showed a sharp decrease in the median fluorescence levels of strains containing *mCherry + tetR* with 100 ng/mL and 200 ng/mL of aTc, which brought them closer to the median fluorescence levels of negative control strains (Figure 5B). Because of these observations, we decided that the *mCherry* strains were not able to establish a wide enough range of protein expression by using these aTc concentrations and time points of induction. As a result, these strains were excluded from the rest of this study.
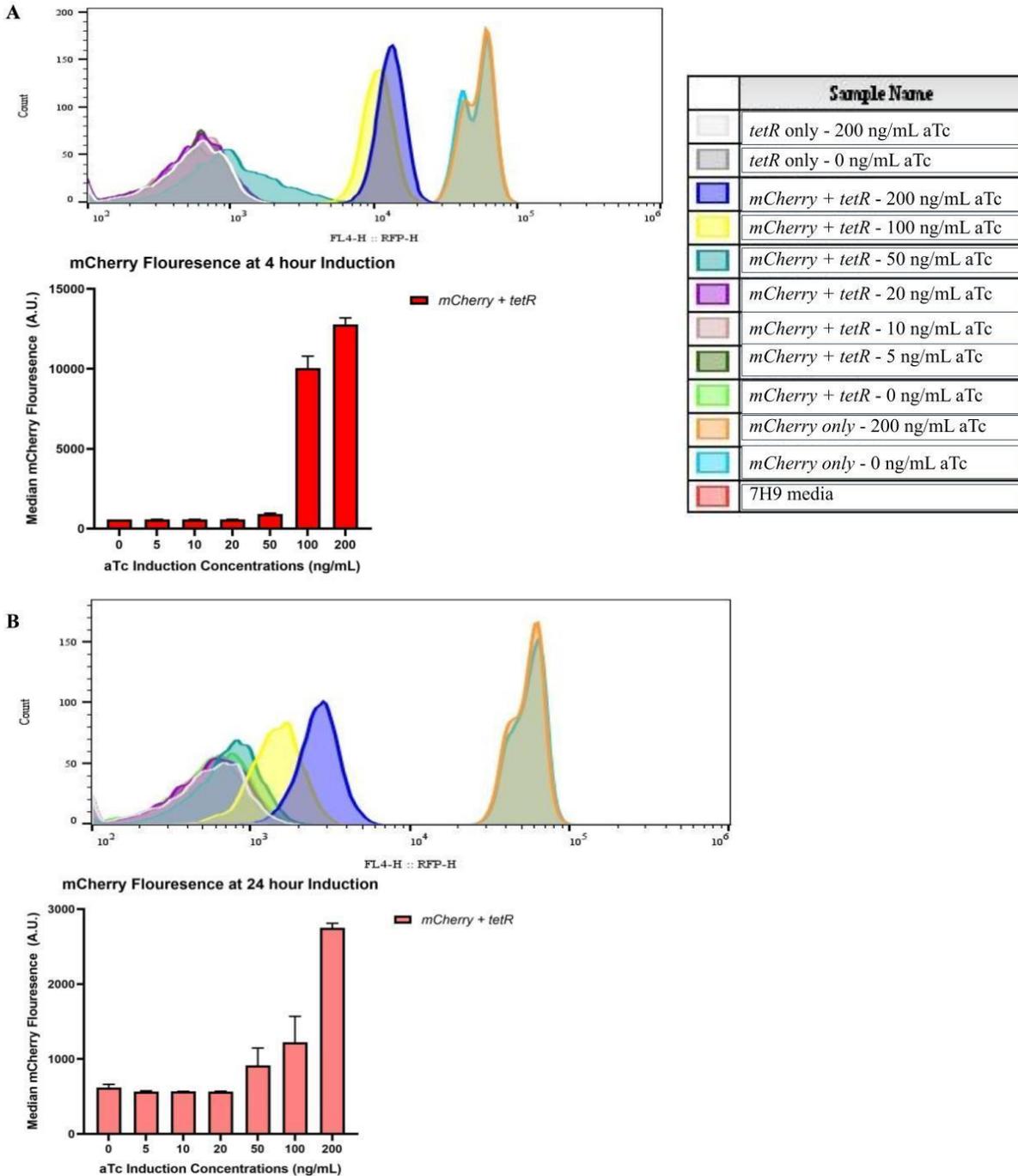
Figure 5. The tested range of aTc concentrations and induction times produced two mCherry fluorescence states: on and off. Samples of the strains containing *mCherry + tetR* were incubated at 0 ng/mL, 5 ng/mL, 10 ng/mL, 20 ng/mL, 50 ng/mL, 100 ng/mL, or 200 ng/mL of aTc for different induction times. Samples containing *mCherry* only, and *tetR* only were incubated at 0 ng/mL and 200 ng/mL of aTc at different induction times. The fluorescence histogram from one duplicate is shown. The mean of median fluorescence of GFPmut3 expression of the duplicate samples and standard deviation were quantified using FlowJo. **(A)** 1 hour induction **(B)** 4 hours induction **(C)** 24 hours induction.

A four-hours induction time point generates wide range of *gfpmut3* mRNA abundance for a range of aTc induction concentrations

qPCR was used to measure *gfpmut3* mRNA abundance four hours after induction with aTc. From previous flow cytometry results (Figure 3C), since there was a significant difference in GFPmut3 protein expression levels between 1 ng/mL and 5 ng/mL aTc (Figure 3B), we added an intermediate concentration of 2.5 ng/mL aTc in an attempt to increase the number of distinct expression levels in our study. We found that the differences in mRNA abundance in the 0-10 ng/mL aTc range were small (Figure 6). Furthermore, mRNA abundance at 50 ng/mL and 100 ng/mL aTc showed no discernible difference in expression from the 20 ng/mL of aTc (Figure 6) despite the differences in protein levels (Figure 3B). Therefore, we decided to exclude 1 ng/mL and 100 ng/mL of aTc from half-life experiments. Moving forward, we decided to induce *gfpmut3 + tetR* strains with 0, 2.5, 5, 10, 20, 50 ng/mL of aTc for 4 hours to determine mRNA half-lives.

Although we did not detect above-background fluorescence in the absence of aTc (Figure 3B), there was mRNA present at levels similar to the housekeeping gene *sigA* (Figure 6). The ability to tightly regulate expression by the *tet* repressor systems is not as clear or straightforward as may seem. Our data support the notion that there is some leakiness associated with the P_{myc1} 2X *tetO* promoter as there was some expression of *gfpmut3* mRNA in the absence of the aTc inducer (Figure 6). Future studies using this system should note this limitation.
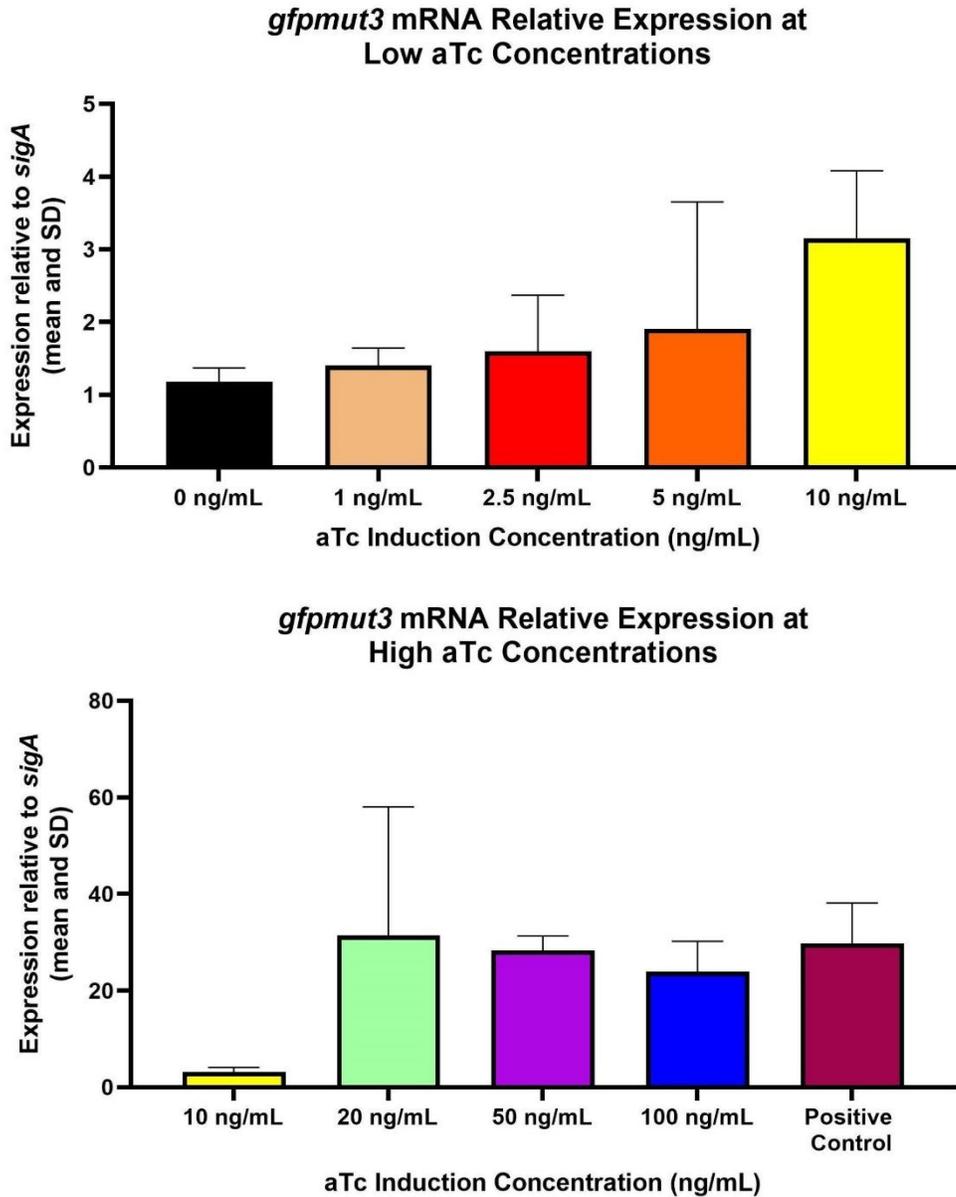
gfpmut3 mRNA Relative Expression at Low aTc Concentrations

gfpmut3 mRNA Relative Expression at High aTc Concentrations

Figure 6. qPCR to measure expression levels of *gfpmut3* mRNA relative to *sigA*. (Top) *gfpmut3 + tetR* strains samples were incubated for 4 hours with the lower aTc concentrations: 0 ng/mL, 1 ng/mL, 2.5 ng/mL, 5 ng/mL, and 10 ng/mL. (Bottom) *gfpmut3 + tetR* strains samples were incubated for 4 hours with the higher aTc concentrations: 20 ng/mL, 50 ng/mL, and 100 ng/mL. The 10 ng/mL concentration was included as a reference for comparison between the mRNA expression levels of the lower and higher aTc concentrations. The positive control of *gfpmut3* only was incubated for 4 hours in the absence of aTc.

mRNA half-lives were indeterminable due to high noise in the mRNA degradation curves

There is a lack of consensus regarding the direction and extent of the correlation between mRNA abundance and mRNA half-life. While some studies note the presence of a weak, positive correlation in non-growing conditions (Redon et al., 2005) and log-phase (Kristoffersen et al., 2012; Chen et al., 2015), others note an inverse correlation in log-phase growth (Bernstein et al., 2002; Redon et al., 2005; Rustad et al., 2013; Esquerré et al., 2015; Nouaille et al., 2017; Sun et al., manuscript in progress). Additionally, the causality for this correlation is unknown as noted by Nouaille et al. (2017). Using a tetracycline-inducible gene expression system of *gfpmut3*, we were interested in determining if mRNA abundance affected the rate of mRNA degradation and transcriptional causality within this relationship in *M. smegmatis*. Hence, we intended to determine the half-lives of *gfpmut3* mRNA at different induction levels. We prepared four biological replicate cultures (batches A-D) that were induced with 0, 2.5, 5, 10, 20, and 50 ng/mL of aTc for 4 hours. We then treated cultures with a high dose of rifampicin to block transcription and harvested RNA after different time points: 0, 0.5, 1, 2, and 4 minutes (Figure 7 and Figure 8). In all the degradation curves obtained for *gfpmut3* mRNA and *sigA* mRNA, there was high noise and high variability between the four batches across all aTc concentrations (Figure 7 and Figure 8). We found that there was little degradation of *gfpmut3* mRNA even after 4 minutes of treatment, implying that *gfpmut3* mRNA may have a longer half-life than expected (Figure 7). Unexpectedly, there appeared to be increased mRNA abundance following rifampicin treatment for some batches. In contrast to the *gfpmut3* degradation curves (Figure 7), there was degradation of *sigA* mRNA across all aTc concentrations (Figure 8). The overall rate of *sigA* degradation over time was slower than expected (Nguyen et al., 2020). Interestingly, we noticed that for all aTc concentrations, the trend of *sigA* degradation between the four batches was not consistent (Figure 8). Batches A and C were more similar to each other, with a faster rate of *sigA* mRNA degradation, whereas batches B and D were most similar to each other and had a slower rate of *sigA* mRNA degradation (Figure 8). Due to the high noise and variability in these mRNA degradation curves, we were not able to calculate *gfpmut3* mRNA half-lives with confidence and establish the correlation between mRNA abundance and mRNA half-lives.
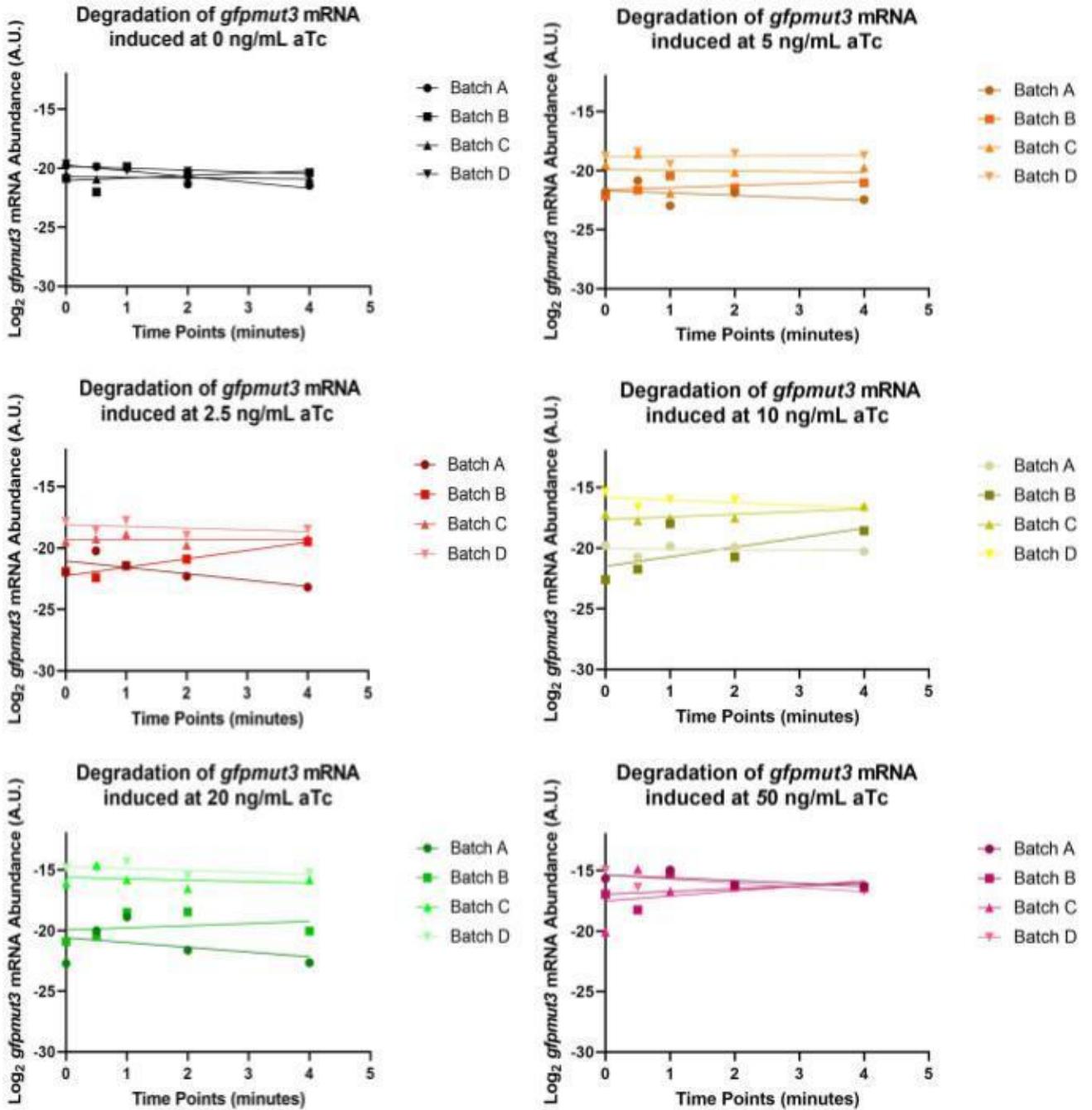
Figure 7. Degradation curves of *gfpmut3* mRNA induced for 4 hours at various concentrations of aTc. Four biological replicates, denoted as different batches, of *gfpmut3* + *tetR* strains were induced with 0, 2.5, 5, 10, 20, and 50 ng/mL of aTc for 4 hours and treated with rifampicin for 0, 0.5, 1, 2, and 4 minutes. mRNA abundance was measured by qPCR and $-C_t$ plotted on the y-axis as a unit of $\log_2$ abundance. Linear regression was performed on each batch using GraphPad Prism 9.
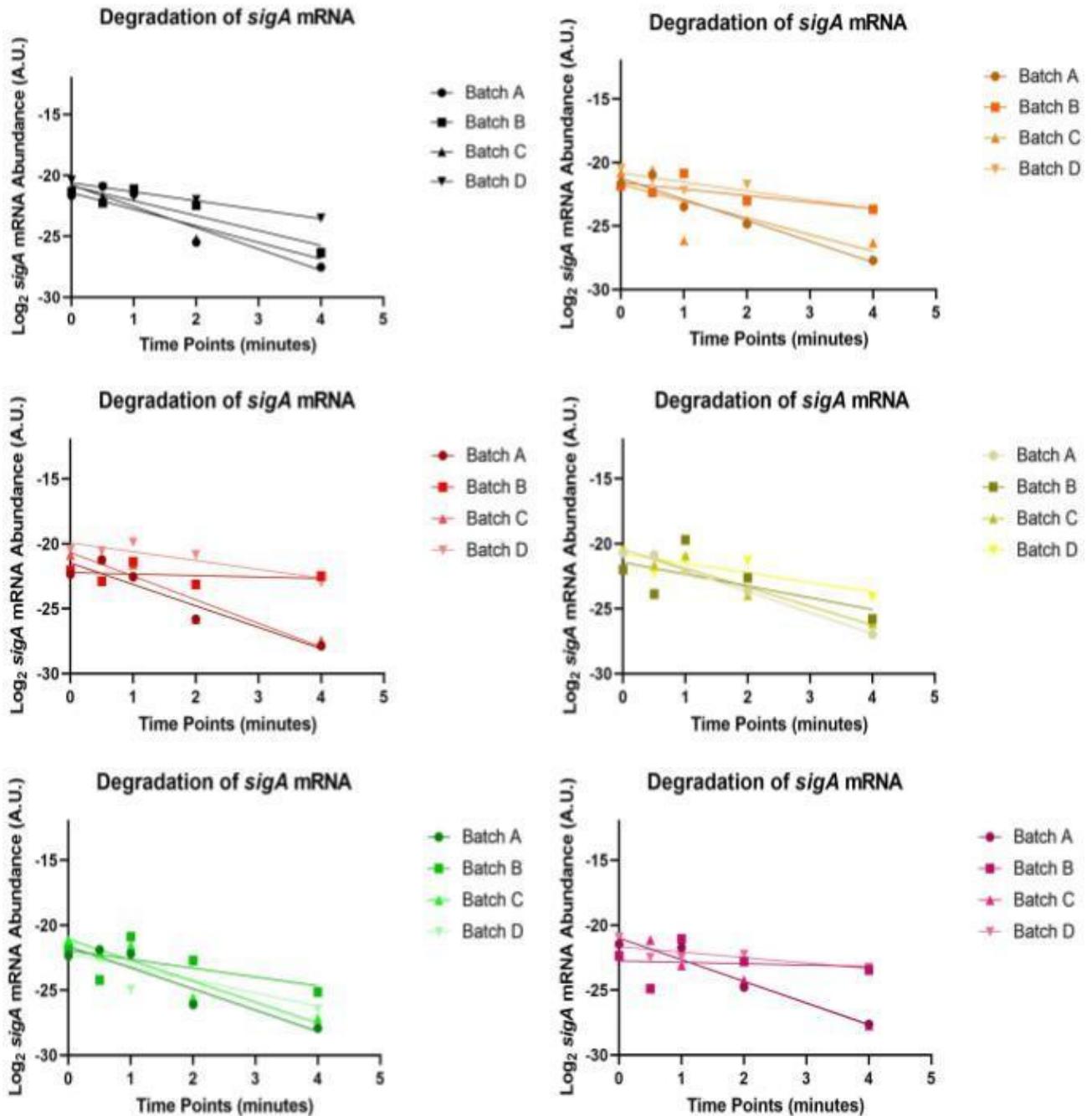
Figure 8. Degradation curves of *sigA* mRNA induced for 4 hours at various concentrations of aTc. Four biological replicates, denoted as different batches, of *gfpmut3* + *tetR* strains were induced with 0, 2.5, 5, 10, 20, and 50 ng/mL of aTc for 4 hours and treated with rifampicin for 0, 0.5, 1, 2, and 4 minutes. mRNA abundance was measured by qPCR and -C_t plotted on the y-axis as a unit of log_2 abundance. Linear regression was performed on each batch using GraphPad Prism 9.

The high variability and noise in degradation curves came from the RNA samples rather than qPCR error

To determine a probable cause of the high noise in the mRNA degradation curves among four biological replicates (Figure 7 and Figure 8), we repeated the qPCR for all samples induced with 20 ng/mL of aTc. Prior to repeating the qPCR, the concentrations of the cDNA samples were re-measured. The amplification plots, cDNA concentrations, and C_t values were compared between the two trials. Beginning with the amplification plots, we saw that Batch A and Batch B replicates had shallow curves in the first trial (Figure 9A). In the second trial, all curves were parallel to each other as expected (Figure 9B). Therefore, any technical errors in the qPCR were resolved. When looking at the amplification plot for the second trial, there appeared to be no significant difference between the *gfpmut3* curves at different time points (Figure 9B). Therefore, qPCR technique and technical errors were likely not the cause for the variability seen in the degradation curves (Figure 7 and Figure 8).
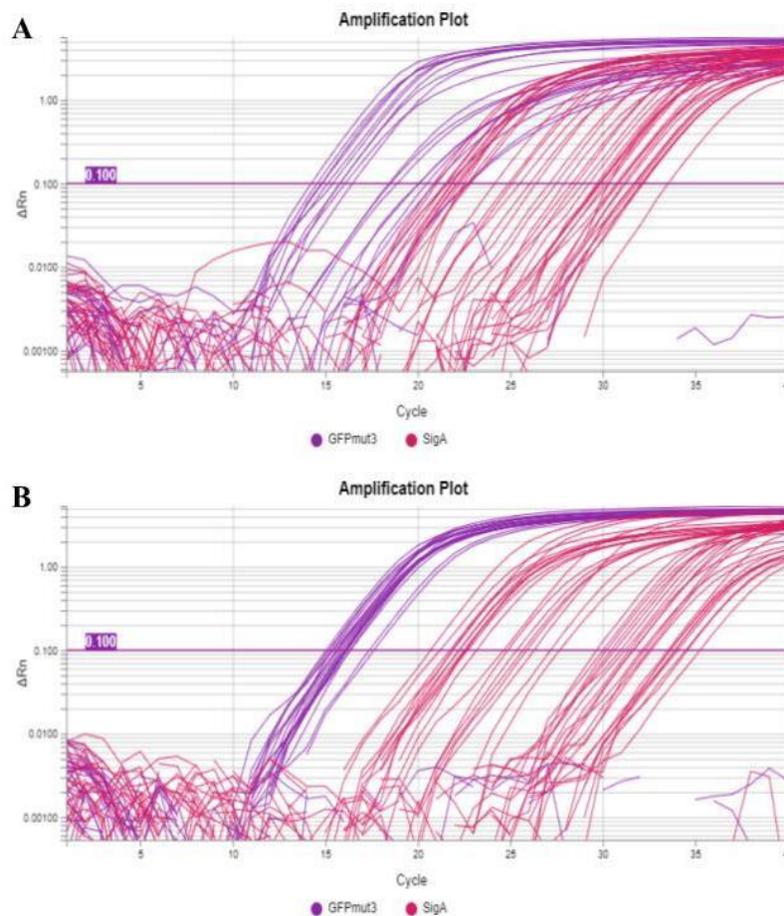

Figure 9. qPCR amplification plots of *gfpmut3* (purple) and *sigA* (red) with the C_t of 0.1 for all four replicates induced with 20 ng/mL of aTc for 4 hours. **(A)** The amplification plot from the first qPCR trial. **(B)** The amplification plot from the second qPCR trial.
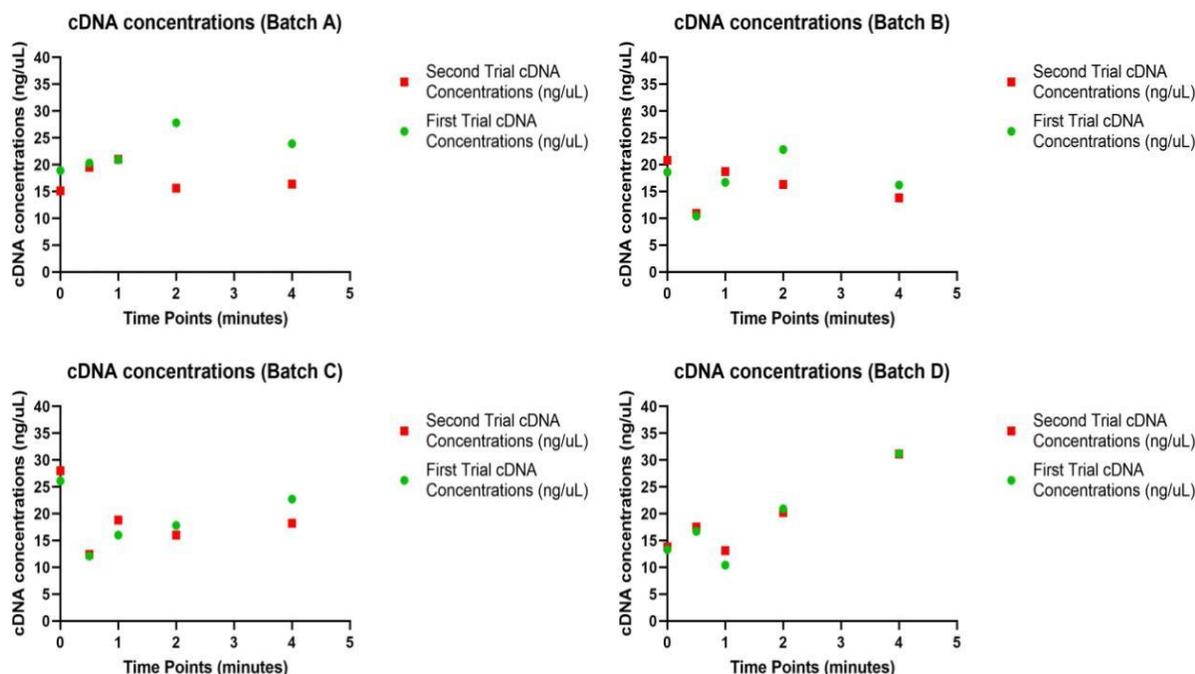
Figure 10. Comparison of the first trial cDNA concentrations (ng/uL) (green circle) and the second trial cDNA concentrations (ng/uL) (red square) at the different time points of rifampicin exposure for all replicates induced with 20 ng/mL of aTc for 4 hours. Each of the replicates was represented in their respective graphs: replicate batch A (top-left), replicate batch B (top-right), replicate batch C (bottom-left), replicate batch D (bottom-right).

To determine whether the cDNA samples were the cause of the variability in the degradation curves, we compared the cDNA concentrations and C_t values between the two qPCR trials for the 20 ng/mL aTc induction samples. To assess cDNA concentration quantification error, we compared the first trial cDNA measurements to the second trial cDNA measurements (Figure 10). Aside from two samples in replicate batch A that were significantly different, cDNA concentrations between the first and second qPCR trials did not vary substantially (Figure 10). Therefore, we believe the noise did not originate from quantifying cDNA.

To assess potential error in cDNA dilution or the qPCR, we plotted the first trial C_t values for *sigA* against the second trial C_t values for *sigA* (Figure 11). Our analysis was performed using *sigA* since other lab members have measured the half-life of *sigA* previously so we had an expectation for how the results should look. When comparing the first and second trial C_t values at each rifampicin exposure time point, there was no significant difference between them for all four replicates (Figure 11A and B). Therefore, the cDNA dilution and qPCR were not the cause of the high noise and variability we saw in the *sigA* degradation curves (Figure 8). When comparing the C_t values of the first trial against the C_t values of the second trial for *sigA*, we saw the different rifampicin exposure time points were out of expected order (Figure 11C). One would expect that as rifampicin exposure increased, the mRNA abundance for *sigA* would decrease. Thus, higher exposure time points should have higher C_t values. In our experiment,

however, all four replicates showed unexpected orders in the exposure time points (Figure 11C). Thus, the source of these cDNA samples, the RNA collected from the rifampicin experiment, are the likely source of the noise and variability in the *sigA* degradation curves (Figure 8).
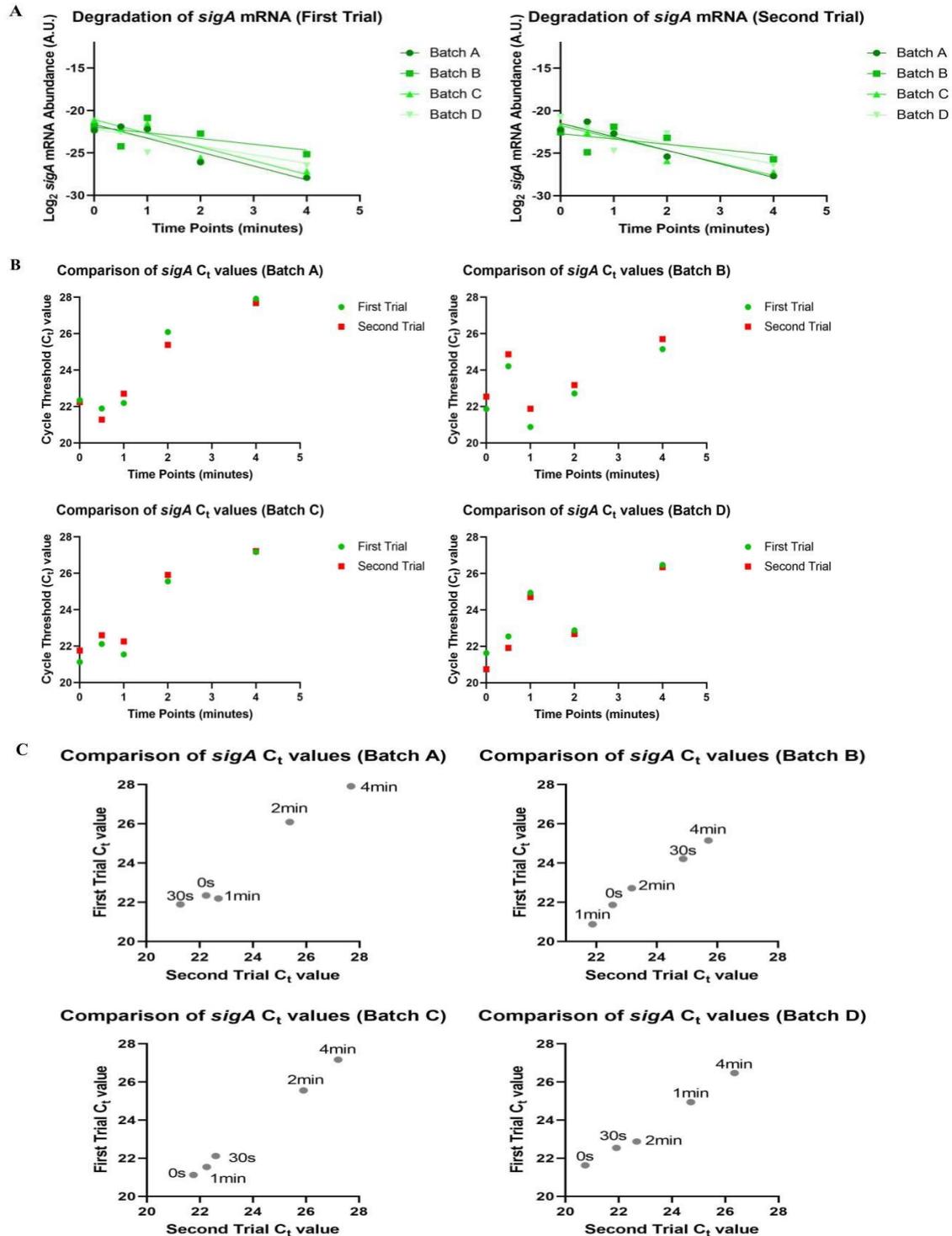
Figure 11. Comparing *sigA* C_t values between two qPCR trials done on the same set of cDNA samples. The data represent four replicates induced at 20 ng/mL of aTc for 4 hours. (**A**) The degradation curves of *sigA* in the first trial (left) and second trial (right) (**B**) Comparison of C_t values between the first (green circle) and second (red square) trials at different time points of rifampicin exposure (**C**) Comparison of C_t values of the first trial against the second trial. Labels at each data point denote the rifampicin exposure time point. The four batches of replicates were graphed separately.

The RNA samples are likewise the most likely cause of the noise seen in the *gfpmut3* mRNA degradation curves as well. The degradation curves for *gfpmut3* were more variable in the first trial than in the second trial (Figure 12A). However, both trials do not show any sign of mRNA degradation (Figure 12A). When comparing the C_t values between the first and second trials, there was substantial variability in replicates batch A and batch B (Figure 12B), which is consistent with the shallow curves observed in Figure 9A. We therefore focused on batch C and batch D replicates. When examining these two replicates, there was no significant difference in the C_t values between the first and second trials at the different rifampicin exposure time points. We observed that the time points were all clustered and in an unexpected order (Figure 12C). The unexpected order indicates that, as with *sigA*, the RNA samples are a cause for the noise in the degradation curves of *gfpmut3* (Figure 7). The clustering indicates that *gfpmut3* mRNA did not degrade as quickly as expected (Figure 12C). The time points for rifampicin exposure were selected based on a previous study using *yfp* in *M. smegmatis* (Nguyen et al., 2020). Our data potentially indicate that the half-life of *gfpmut3* mRNA is longer than *yfp* mRNA. Hence, we suggest performing a similar half-life experiment with longer rifampicin exposure time points such as 0, 1, 2, 4, 8, 10, and 12 minutes for each aTc concentration.

To summarize, the degradation curves were too noisy to determine half-lives of *gfpmut3* mRNA. Thus, we were unable to establish a relationship between mRNA abundance and mRNA half-life. We could not conclude whether transcription is causal in this relationship either. Additionally, loss of aTc-induced GFPmut3 protein expression over time prevented us from investigating the steady-state relationship between mRNA abundance and protein abundance. Future work should consider the recommendations for studying half-life and limitations of our gene expression system in *M. smegmatis*.
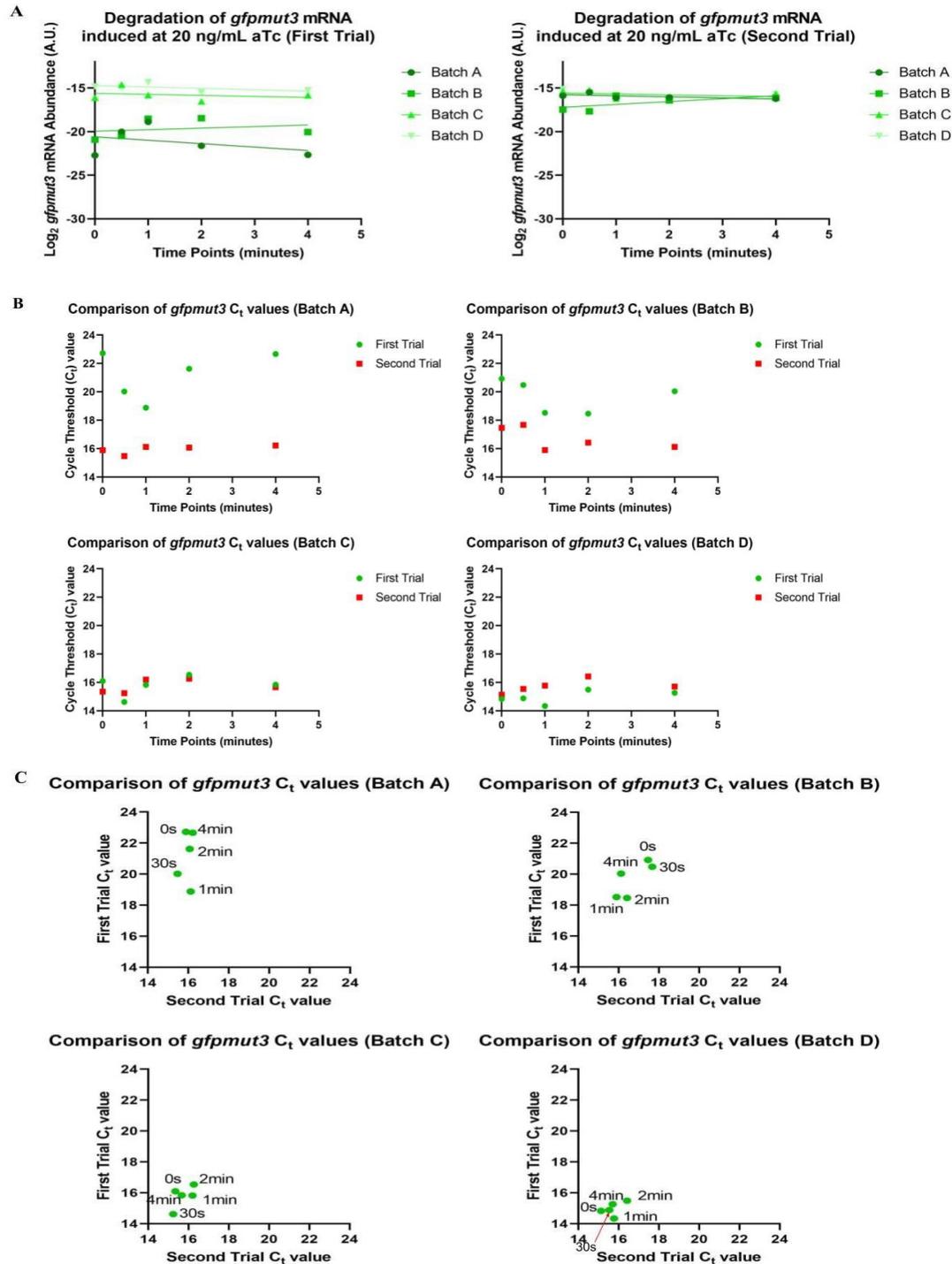
Figure 12. Comparing *gfpmut3* C_t values between the first and second qPCR trials done on the same set of cDNA samples. The data represent four replicates induced at 20 ng/mL of aTc for 4 hours. (**A**) The degradation curves of *gfpmut3* in the first trial (left) and second trial (right) (**B**) Comparison of C_t values between the first (green circle) and second (red square) trials at different time points of rifampicin exposure (**C**) Comparison of C_t values of the first trial against the second trial. Labels at each data point denote the rifampicin exposure time point. The four batches of replicates were graphed separately. In batch D, the red arrow denotes the 30 second rifampicin exposure time point.

Acknowledgements

First and foremost, we would like to thank our Biology and Biotechnology advisor Dr. Scarlet Shell. Her invaluable guidance, continuous support, and profound knowledge were invaluable to us.

We would also like to express our gratitude to Diego Vargas-Blanco, Ying Zhou, María Carla Martini, Julia Ryan, Mara Natalia Alonso, and all the members of the Shell Lab. Their help, guidance, and positivity within the lab was always helpful and appreciated.

References

Arnvig, K. & Young, D. (2012) Non-coding RNA and its potential role in *Mycobacterium tuberculosis* pathogenesis. *RNA Biology*, (9)4, 427-436. DOI: 10.4161/rna.20105

Baumschlager, A., Rullan, M., & Khammash, M. (2020) Exploiting natural chemical photosensitivity of anhydrotetracycline and tetracycline for dynamic and setpoint chemo-optogenetic control. *Nat Communications*, 11, 3834. https://doi.org/10.1038/s41467-020-17677-5

Bernstein, J. A., Khodursky, A. B., Lin, P. H., Lin-Chao, S., & Cohen, S. N. (2002). Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proceedings of the National Academy of Sciences*, *99*(15), 9697-9702. https://doi.org/10.1073/pnas.112318199

Blokpoel, M. C., Murphy, H. N., O'Toole, R., Wiles, S., Runn, E. S., Stewart, G. R., Young, D. B., & Robertson, B. D. (2005). Tetracycline-inducible gene regulation in mycobacteria. *Nucleic acids research*, *33*(2), e22-e22.

Blum, E., Carpousis, A. J., & Higgins, C. F. (1999). Polyadenylation promotes degradation of 3′-structured RNA by the *Escherichia coli* mRNA degradosome in vitro. *Journal of Biological Chemistry*, 274(7), 4009-4016.

Boël, G., Letso, R., Neely, H., Price, W. N., Wong, K. H., Su, M., Luff, D. J., Valecha, M., Everett, K. J., Acton, B. T., Xiao, R., Montelione, T. G., Aalberts, P. D., & Hunt, J. F. (2016). Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature*, *529*(7586), 358-363. https://doi.org/10.1038/nature16509

Boldrin, F., Provvedi, R., Cioetto Mazzabò, L., Segafreddo, G., & Manganelli, R. (2020). Tolerance and persistence to drugs: a main challenge in the fight against *Mycobacterium tuberculosis*. *Frontiers in Microbiology*, *11*, 1924.

Brescia, C. C., Kaw, M. K., & Sledjeski, D. D. (2004). The DNA binding protein H-NS binds to and alters the stability of RNA in vitro and in vivo. *Journal of molecular biology*, *339*(3), 505-514.

Carroll, P., Muttucumaru, D. N., & Parish, T. (2005). Use of a tetracycline-inducible system for conditional expression in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. *Applied and environmental microbiology*, *71*(6), 3077-3084.

Centers for Disease Control and Prevention (CDC). (1990, August 24). *Tuberculosis in developing countries*. Centers for Disease Control and Prevention. https://www.cdc.gov/mmwr/preview/mmwrhtml/00001729.htm.

Chen, L. H., Emory, S. A., Bricker, A. L., Bouvet, & Belasco, J. G. (1991). Structure and function of a bacterial mRNA stabilizer: analysis of the 5'untranslated region of ompA mRNA. *Journal of bacteriology*, *173*(15), 4578-4586. https://doi.org/10.1128/jb.173.15.4578-4586.1991

Chen, H., Shiroguchi, K., Ge, H., & Xie, X. S. (2015). Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Molecular systems biology*, *11*(1), 781. https://doi.org/10.15252/msb.20145794

Connolly, L. E., Edelstein, P. H., & Ramakrishnan, L. (2007). Why is long-term therapy required to cure tuberculosis?. *PLoS medicine*, 4(3), e120. https://doi.org/10.1371/journal.pmed.0040120

Czyz, A., Mooney, R. A., Iaconi, A., & Landick, R. (2014). Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *MBio*, *5*(2), e00931-14.

de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M., & Vogel, C. (2009). Global signatures of protein and mRNA expression levels. *Molecular BioSystems*, *5*(12), 1512-1526

Dressaire, C., Picard, F., Redon, E., Loubière, P., Queinnec, I., Girbal, L., & Cocaign-Bousquet, M. (2013). Role of mRNA stability during bacterial adaptation. *PloS one, 8*(3), e59059. https://doi.org/10.1371/journal.pone.0059059

Ehrt, S., Guo, X. V., Hickey, C. M., Ryou, M., Monteleone, M., Riley, L. W., & Schnappinger, D. (2005). Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic acids research*, *33*(2), e21-e21.

Emory, S. A., Bouvet, P., & Belasco, J. G. (1992). A 5'-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*. *Genes & development*, 6(1), 135-148. https://doi.org/10.1101/gad.6.1.135

Esquerré, T., Moisan, A., Chiapello, H., Arike, L., Vilu, R., Gaspin, C., Cocaign-Bousquet, M., & Girbal, L. (2015). Genome-wide investigation of mRNA lifetime determinants in *Escherichia coli* cells cultured at different growth rates. *BMC genomics*, *16*(1), 1-13. https://doi.org/10.1186/s12864-015-1482-8

Goyal, R., Das, A. K., Singh, R., Singh, P. K., Korpole, S., & Sarkar, D. (2011). Phosphorylation of PhoP protein plays direct regulatory role in lipid biosynthesis of *Mycobacterium tuberculosis*. *Journal of Biological Chemistry*, *286*(52), 45197-45208.

Guet, C. C., Bruneaux, L., Min, T. L., Siegal-Gaskins, D., Figueroa, I., Emonet, T., & Cluzel, P. (2008). Minimally invasive determination of mRNA concentration in single living bacteria. *Nucleic acids research*, 36(12), e73-e73. https://doi.org/10.1093/nar/gkn329.

Hausser, J., Mayo, A., Keren, L., & Alon, U. (2019). Central dogma rates and the trade-off between precision and economy in gene expression. *Nature communications*, *10*(1), 1-15. https://doi.org/10.1038/s41467-018-07391-8

Huff, J., Czyz, A., Landick, R., & Niederweis, M. (2010). Taking phage integration to the next level as a genetic tool for mycobacteria. *Gene*, *468*(1-2), 8-19.

Korch, S. B., Contreras, H., & Clark-Curtiss, J. E. (2009). Three *Mycobacterium tuberculosis* Rel toxin-antitoxin modules inhibit mycobacterial growth and are expressed in infected human macrophages. *Journal of bacteriology*, *191*(5), 1618-1630.

Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, 361, 13-37. https://doi.org/10.1016/j.gene.2005.06.037

Kristoffersen, S. M., Haase, C., Weil, M. R., Passalacqua, K. D., Niazi, F., Hutchison, S. K., Disney, B., Kolstø, A., Tourasse, N. J., Read, T. D., & Økstad, O. A. (2012). Global mRNA decay analysis at single nucleotide resolution reveals segmental and positional degradation patterns in a Gram-positive bacterium. *Genome biology*, *13*(4), 1-21. https://doi.org/10.1186/gb-2012-13-4-r30

Kwon, T., Huse, H. K., Vogel, C., Whiteley, M., & Marcotte, E. M. (2014). Protein-to-mRNA ratios are conserved between *Pseudomonas aeruginosa* strains. *Journal of proteome research*, *13*(5), 2370-2380.

Lee, P. S., Shaw, L. B., Choe, L. H., Mehra, A., Hatzimanikatis, V., & Lee, K. H. (2003) Insights Into the Relation Between mRNA and Protein Expression Patterns: II. Experimental Observations in *Escherichia coli*. *Wiley Periodicals Inc.*, 1, 2-9. DOI: 10.1002/bit.10841

Lenz, G., Doron-Faigenboim, A., Ron, E.Z., Tuller, T., & Gophna, U. (2011) Sequence Features of *E. coli* mRNAs Affect Their Degradation. *PLoS ONE*, 6(12): e28544. https://doi.org/10.1371/journal.pone.0028544

Li, G., Oh, E. & Weissman, J. (2012). The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*, 484, 538–541. https://doi.org/10.1038/nature10965

Liu, Y., Wu, N., Dong, J., Gao, Y., Zhang, X., Mu, C., Ningsheng, S.,& Yang, G. (2010). Hfq is a global regulator that controls the pathogenicity of Staphylococcus aureus. *PLoS One*, *5*(9), e13069.

Minch, K., Rustad, T., & Sherman, D. R. (2012). *Mycobacterium tuberculosis* growth following aerobic expression of the DosR regulon. *PLoS One*, *7*(4), e35935.

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*, 202(9), e00746-19. https://doi.org/10.1128/JB.00746-19

Nouaille, S., Mondeil, S., Finoux, A. L., Moulis, C., Girbal, L., & Cocaign-Bousquet, M. (2017). The stability of an mRNA is influenced by its concentration: a potential physical mechanism to regulate gene expression. *Nucleic acids research*, 45(20), 11711-11724. https://doi.org/10.1093/nar/gkx781

O'Hara, E. B., Chekanova, J. A., Ingle, C. A., Kushner, Z. R., Peters, E., & Kushner, S. R. (1995). Polyadenylylation helps regulate mRNA decay in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 92(6), 1807–1811. https://doi.org/10.1073/pnas.92.6.1807

Pato, M. L., Bennett, P.M., & Von Meyenburg, K. (1973) Messenger Ribonucleic Acid Synthesis and Degradation in *Escherichia coli* During Inhibition of Translation. *Journal of Bacteriology*, 116(2), 710-718. https://journals.asm.org/doi/epdf/10.1128/jb.116.2.710-718.1973.

Politi, N., Pasotti, L., Zucca, S., Casanova, M., Micoli, G., Cusella De Angelis, M. G., & Magni, P. (2014). Half-life measurements of chemical inducers for recombinant gene expression. *Journal of biological engineering*, 8(1), 1-10.

Prax, M., & Bertram, R. (2014). Review: Metabolic aspects of bacterial persisters. *Frontiers in cellular and infection microbiology*, 4, 148. https://doi.org/10.3389/fcimb.2014.00148

Raghavan, S., Manzanillo, P., Chan, K., Dovey, C., & Cox, J. S. (2008). Secreted transcription factor controls Mycobacterium tuberculosis virulence. *Nature*, *454*(7205), 717-721.

Redon, E., Loubière, P., & Cocaign-Bousquet, M. (2005). Role of mRNA stability during genome-wide adaptation of *Lactococcus lactis* to carbon starvation. *The Journal of biological chemistry,* 280(43), 36380–36385. https://doi.org/10.1074/jbc.M506006200

Ren, G. X., Guo, X. P., & Sun, Y. C. (2017). Regulatory 3' Untranslated Regions of Bacterial mRNAs. *Frontiers in microbiology*, 8, 1276. https://doi.org/10.3389/fmicb.2017.01276

Riba, A., Di Nanni, N., Mittal, N., Arhné, E., Schmidt, A., & Zavolan, M. (2019). Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *Proceedings of the national academy of sciences*, 116(30), 15023-15032. https://doi.org/10.1073/pnas.1817299116

Rustad, T. R., Minch, K. J., Brabant, W., Winkler, J. K., Reiss, D. J., Baliga, N. S., & Sherman, D. R. (2013). Global analysis of mRNA stability in *Mycobacterium tuberculosis*. *Nucleic acids research,* 41(1), 509-517. https://doi.org/10.1093/nar/gks1019

Russel, J. B., & Cook, G. M. (1995). Energetics of Bacterial Growth: Balance of Anabolic and Catabolic Reactions. *Microbiological Review*, 59(1), 48-62. https://doi.org/10.1128/mr.59.1.48-62.1995

Saito, K., Green, R., & Buskirk, A. R. (2020). Translational initiation in *E. coli* occurs at the correct sites genome-wide in the absence of mRNA-rRNA base-pairing. *eLife*, 9:e55002 DOI: 10.7554/eLife.55002

Sinha, K. M., Stephanou, N. C., Gao, F., Glickman, M. S., & Shuman, S. (2007). Mycobacterial UvrD1 is a Ku-dependent DNA helicase that plays a role in multiple DNA repair events, including double-strand break repair. *Journal of Biological Chemistry*, *282*(20), 15114-15125.

Sinha, D., Matz, L. M., Cameron, T. A., & De Lay, N. R. (2018). Poly (A) polymerase is required for RyhB sRNA stability and function in *Escherichia coli*. *Rna*, 24(11), 1496-1511. https://doi.org/10.1261/rna.067181.118

Stouthamer, A. H. (1979). The search for correlation between theoretical and experimental growth yields, p. 1–47. In J. R. Quayle (ed.), *International Review of biochemistry and microbial biochemistry*, vol. 21. University ParkPress, Baltimore.

Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H., Babu, M., Hearn, J. Emili, A., & Xie1,S. (2010). Quantifying *E. coli* Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *SCIENCE*, 329, 533-538. 10.1126/science.1188308.

Timmermans, J. & Van Melderen, L. (2010). Post-transcriptional global regulation by CsrA in bacteria. *Cellular and molecular life sciences*, 67(17), 2897-2908. https://doi.org/10.1007/s00018-010-0381-z

Terai, G. & Asai, K. (2020). Improving the prediction accuracy of protein abundance in *Escherichia coli* using mRNA accessibility. *Nucleic Acids Research*, 48(14), Page e81. https://doi.org/10.1093/nar/gkaa481

Tuller, T., Waldman, Y. Y., Kupiec, M., & Ruppin, E. (2010). Translation efficiency is determined by both codon bias and folding energy. *PNAS*, 107(8), 3645-3650. https://doi.org/10.1073/pnas.0909910107

Vargas-Blanco, D. A., Zhou, Y., Zamalloa, L. G., Antonelli, T., & Shell, S. S. (2019). mRNA degradation rates are coupled to metabolic status in *Mycobacterium smegmatis*. *MBio*, 10(4), e00957-19. https://doi.org/10.1128/mBio.00957-19

World Health Organization (WHO). (2020, October 14). *Tuberculosis (TB)*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/tuberculosis.