The Effects of Codon Content and Context on Transcript Stability and Translation Efficiency 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:

Olivia Atkins

April 27th, 2023

Approved by:
Scarlet S. Shell, PhD

Abstract

Tuberculosis (TB) is a deadly infectious disease caused by *Mycobacterium tuberculosis* (Mtb). It primarily affects the lungs and has a 50% survival rate if left untreated. Treatment is difficult, costly, and can last up to six months, and drug resistance is a growing concern. The bacterium enters the body through inhalation and survives in the lungs inside a granuloma. In this environment, *M. tuberculosis* can slow its growth, become drug tolerant, and develop drug resistance. Altered gene expression is a key part of this slowed growth, but little is known about the mechanisms involved. Synonymous codon usage and adjacent codon context have emerged as important factors influencing gene expression. The focus of this study was to determine whether codon content and context have a causal effect on mRNA degradation rate and translation efficiency in the model mycobacterium *Mycolicibacterium smegmatis*. Through machine learning analysis, the arginine CGC (ArgCGC) codon was associated with decreased transcript stability in *M. smegmatis*. To investigate this relationship further, the ArgCGC codon, synonymous arginine codons, and frameshift codons with the same nucleotide content were each integrated into the *Mycolicibacterium smegmatis* chromosome in a series of four, as translational fusions to a gene encoding yellow fluorescent protein (YFP). Quantitative PCR (qPCR) was conducted on the various strains to determine the codon inserts' impact on mRNA abundance, and flow cytometry was used to determine the impact on protein abundance. Translation efficiency of the ArgCGC codon appears to be dependent on codon position relative to translation initiation and codon context. Overall, codons show a strong effect on translation efficiency in mycobacteria.

Introduction

Tuberculosis (TB) is an infectious, airborne disease primarily affecting the lungs and is caused by the pathogenic bacillus *Mycobacterium tuberculosis* (Mtb). Tuberculosis is incredibly deadly; prior to the COVID-19 pandemic, TB was the leading cause of death globally by a single infectious agent with an estimated 1.5 million deaths each year (World Health Organization, 2022). The COVID pandemic worsened outcomes for those with TB, as less cases were reported and drug access was limited (World Health Organization, 2022). Untreated TB has only a 50% survival rate, and treatment of this disease is difficult and costly even in countries of high socioeconomic status (World Health Organization, 2022). The current treatment regimen includes four to six months of multi-drug therapy. This timeline may be extended if the case becomes resistant to the antibiotics used to treat it. Likelihood of resistance increases if patients do not complete their full treatment course, are not provided the correct dosage or type of treatment, or if the drugs administered are not of good quality (Centers for Disease Control and Prevention, 2022). Multidrug-resistant (MDR) TB is considered a global health crisis and is defined by the Centers for Disease Control and Prevention (CDC) as a case in which the *M. tuberculosis* bacteria is resistant to two of the most important first-line drugs, isoniazid and rifampin. Deeper understanding of the pathophysiology of *M. tuberculosis* infection and drug resistance is necessary to improve patient outcomes.

Mtb enters the body through the inhalation of contaminated aerosols which deposit in the lower respiratory tract. In the lungs, the host immune system responds to the infection by encasing the bacteria in macrophages and other immune cells, forming a granuloma (Samuels et al., 2022). Researchers have found that the bacteria actively recruit these immune cells, indicating a possible evolutionary benefit to granuloma encasement (Volkman et al., 2010). Inside the granuloma, the bacteria encounter stress conditions that challenge Mtb survival (Samuels et al., 2022). Despite these conditions, the granuloma provides a niche environment for Mtb to survive, adapt, and evade host immune functions. Conditions encountered in the granuloma such as hypoxia, nutrient depravation, phagosome acidification, and presence of reactive oxygen or nitrogen species cause changes to metabolic flux and result in slowed Mtb growth (Samuels et al., 2022). In its slowed growth state Mtb can survive, become drug tolerant, and develop specific drug resistance . In addition to changes in metabolic flux, less understood

alterations in gene expression are observed in slowed Mtb growth (Reviewed in: Connolly et al., 2007).

Mycobacterial gene expression can be altered at several steps including transcription, post-transcriptional modification, mRNA degradation, translation, and protein degradation. There is lots of study surrounding what regulates transcription in mycobacteria (Park et al., 2003, Lin et al., 2017, Schnappinger et al., 2003); however, less is known about degradation and translation rates during infection and in infection-relevant stress conditions. When Mtb encounters stress conditions and enters these less metabolically active states, less protein is produced. This could be caused by direct changes to translation, changes in transcription and transcript degradation, or a combination of these factors. In non-growing mycobacteria, there is an observed global increase in mRNA stability (Rustad et al., 2013, Vargas-Blanco et al., 2019, Smeulders et al., 1999). Increasing mRNA half-life seems to be energetically favorable under these conditions (Vargas-Blanco et al., 2019). Other bacteria show similar behaviors. *E. coli* slows its mRNA turnover in hypoxic conditions (Georgellis et al., 1993) as well as during periods of starvation and non-growing phases (Esquerré et al., 2015, Chen et al., 2015). The underlying factors and mechanisms involved in slowed Mtb mRNA turnover and translation under physiological stress conditions remain to be thoroughly investigated.

Studies in yeast revealed mRNA stability is influenced by secondary structure, sequence, 5’ and 3’ UTRs, and transcript length (Parker, 2012; Neymotin et al., 2016; Cheng et al., 2017). In bacteria mRNA stability can be further impacted by alterations to the degradosome (Bernstein et al., 2004; Kido et al., 1996; Ow et al., 2000), RNase cleavage preferences (Khemici et al., 2015; Shahbabian et al., 2009; Martini et al., 2019), subcellular compartmentalization (Khemici et al., 2015; Moffitt et al., 2016), 5’ capping (Chen et al., 2009; Luciano et al., 2019), and association with regulatory proteins and sRNAs (Braun et al., 1998; Richards & Belasco, 2019).

There is a proven relationship between translation and mRNA degradation. Ribosomal association with a transcript, at several stages of translation, along with interactions with cis-regulatory elements strongly influence transcript stability. Bacterial mRNA half-lives can vary greatly from just seconds to well over an hour (Deana & Belasco, 2005). In *E. coli*, mRNA decay begins with endonucleolytic cleavage by RNase E or another RNase species. 3' exonucleases then attack the unprotected 3' end. The ribosomal binding site (RBS) (Jürgen et al., 1998;

Wagner et al., 1994) and 5' stem loop (Bricker & Belasco, 1999; Sa et al., 1992), both features of the 5' UTR, are key factors involved in transcript stability.

Further research suggests codon content may be an important factor impacting both mRNA stability and translation efficiency (Hanson & Coller, 2018). The redundant nature of the genetic code allows for 20 amino acids to be encoded by 61 different RNA triplets. Synonymous codon changes, those in which the same amino acid is encoded from different RNA sequences, have proven to confer important phenotypic changes. In *Bacillus subtilis* for example, one synonymous serine alteration allows the bacteria to form biofilms (Subramaniam et al., 2013). Codon usage in mycobacteria is influenced by additional factors such as high GC content (Cole et al., 1998), gene and amino acid conservation, and hydrophobicity (de Miranda, 2000).

Mutations in the Shine-Dalgarno (SD) ribosome binding site that weaken binding are shown to destabilize mRNA in *Bacillus subtilis* (Sharp & Bechhofer, 2003; Jürgen et al., 1998; Hambraeus et al., 2002). Beyond mutations, competitive binding to the RBS by regulatory proteins (Večerek et al., 2005), sRNA binding to or near the RBS (Gottesman, 2004), and the presence of a premature termination codon (Nilsson et al., 1987; Arnold et al., 1998; Braun et al., 1998) are all shown to decrease mRNA half-life. Ribosomal pausing can either stabilize or destabilize the transcript (Li et al., 2012; Deana & Belasco, 2005), with the latter thought to be caused by either a proline codon followed by a stop codon or clusters of rare codons within the coding sequence (Sunohara et al., 2004). Impaired translation appears to cause accelerated mRNA degradation possibly as a mechanism to limit the use of poorly or incorrectly translated mRNAs as templates for new proteins (Deana & Belasco, 2005).

Many studies have investigated the influence of codon content on translation efficiency. These studies are complicated by alterations that yield changes not only to the sequence but also features like tRNA usage, mRNA secondary structure, nucleotide content, and global codon usage (Boël et al., 2016). Deciphering the influence of these individual features reveals that codon content may be strongly correlated to protein expression (Boël et al., 2016; Tuller et al., 2010) and seemingly “silent” mutations can have an impact on translation speed (Chevance et al., 2014). Studies in yeast found that “optimal” codons were translated efficiently, increased mRNA half-lives, and were linked to highly abundant protein products (Carneiro, 2019, Presnyak, 2015). Similarly, another study showed that codon-optimized gene transcripts

increased both protein and mRNA concentrations in *E. coli*, suggesting that codon content plays a role in translation and its influence on mRNA stability (Boël et al., 2016).

Location within the transcript seems to determine which transcript features most influence translation efficiency. A study in *E. coli* revealed that rare codon usage at the N-terminus of bacterial transcripts correlated with increased protein expression due to reduced mRNA folding (Goodman et al., 2013). When researchers introduced synonymous changes to N-terminal codons there was an average 60-fold difference in protein abundance between the highest-expressing and lowest-expressing variants. There was a 14-fold increase in expression when rare codons, associated with reduced mRNA structure, were introduced as compared to commonly used codons. This was found to be due to rare codons being more A/U rich. In another study features like mRNA folding and individual codons were shown to influence translation most strongly within the first eighteen nucleotides, or the head, of the transcript (Boël et al., 2016). In the tail, overall codon content was the most influential in determining expression. Several studies in both prokaryotic and eukaryotic systems have demonstrated that reduced secondary structure at the ribosome binding site (RBS) generally caused higher protein expression, supporting the idea that initiation is the rate-limiting step for translation (Boël et al., 2016; Kudla et al., 2009; Bentele et al., 2013; Ingolia et al., 2009).

Specific codons can also impact the rate of translation across the genome. Codons decoded by wobble base pairing with a non-cognate tRNA greatly correlate to decreased expression (Boël et al., 2016), and slower translation (Spencer et al., 2012) particularly when grouped together in the transcript. These effects were only observed for a specific codon, not synonymous sequences, indicating a codon-specific influence rather than an effect of amino acid structure. In other cases, amino acid structure seemed influential as charged residues such as arginine, glutamine, aspartic acid, and lysine all correlated with greater expression (Boël et al., 2016). Codons with internal sequence motifs can also impact translation rates. SD-like sequences within a coding region have affinity for the anti-SD site on ribosomes, causing ribosomal pausing which slows translation speed (Li et al., 2012, Goodman et al., 2013, Chevance et al., 2014). Li et al. also postulated that because ribosomal pausing also reduces the number of free ribosomes able to initiate translation, widespread SD-like sequences in coding regions may slow bacterial growth rates. This could be particularly interesting to study in Mtb, which displays slowed growth *in vivo*.

Codon context, or neighboring codons, also affects translation rate in bacteria. By 1980, there was evidence to support the impact of transcript mutations on translation efficiency of an adjacent codon (Bossi & Roth, 1980). A more recent study measured translation speed independent of transcript or protein product stability (Chevance et al., 2014), using a histidine operon transcription attenuation feature (Johnston et al., 1980) and *lac* operon reporter. They used this system to study the impact of codon context 5' and 3' to the UCA codon which is only read by a single tRNA species and does not have affinity to the anti-SD site. They observed significant de-attenuation, a result of efficient translation, and β-galactosidase expression for 21 out of 64 codon pairs placed 5' or 3' to the UCA codon, confirming the importance of codon context on translation (Chevance et al., 2014). There is little research on the impact of codon context on translational efficiency in Mtb.

In this study, we sought to determine whether codon content and context have a causal effect on mRNA degradation rate and translation efficiency in *Mycolicibacterium smegmatis*, a model organism for *Mycobacterium tuberculosis*. Our lab has used machine learning to determine which features of endogenous *M. smegmatis* mRNA transcripts can predict degradation rates. Abundance of certain codons was found to be higher in fast-degrading genes and lower in slow-degrading genes. This corroborates that codon content may be impactful in mRNA stability in mycobacteria. The specific codon of interest in this study is arginine CGC (ArgCGC), which correlated with less stability. It is hypothesized that the ArgCGC codon is translated inefficiently due to wobble base pairing leading to faster mRNA degradation. To investigate this relationship further, the ArgCGC codon, synonymous arginine codons, and frameshift codons with the same nucleotide content were each integrated into the *M. smegmatis* chromosome in a series of four as translational fusions to a gene encoding yellow fluorescent protein (*yfp*). Quantitative PCR (qPCR) was conducted on the various strains to determine the codon inserts' impact on mRNA abundance, and flow cytometry was used to determine the impact on protein abundance. Translation efficiency of the ArgCGC codon appears to be dependent on codon position relative to translation initiation and codon context. Overall, codons have a strong effect on translation efficiency in mycobacteria.

Materials and Methods

Plasmid Design and Construction

Plasmid pSS303 (Nguyen et al., 2020), which uses the pMyc-1-tetO promoter and associated 5' UTR to encode YFP with a C-terminal His tag, was used as a template for the insertion of a first set of test sequences which included a FLAG tag (DYKDDDDK) and codon sequence of interest as an N-terminal fusion to YFP. These FLAG-containing insertions into pSS303 created pSS504 (FLAG only) and pSS517-pSS522 (Table 1) where test sequences were inserted at amino acid residues 10 through 13, relative to translation initiation. Plasmid pSS359 (Nguyen et al., 2020) was used as a template for the insertion of a second set of codon test sequences, which were inserted into the plasmid's 54 nucleotide *sigA* coding sequence beginning at nucleotide 52. ValineGUG (V) and SerineUCG (S) were chosen to flank the test codons due to their abundant use in mycobacteria (de Miranda et al., 2000) as well as their hydrophobicity and neutral charge, respectively. In these SigA-containing constructs the test sequences are located at amino acid residues 20 through 23, relative to translation initiation. This changed the codon context as compared to the FLAG-containing constructs, while maintaining uniform context that should not lead to lower translation efficiency within the SigA-containing constructs. SigA-containing constructs made plasmids pSS721-pSS725 and pSS730 (Table 2). Each of these plasmids contained *HygR*, a gene conferring genetic resistance to Hygromycin B. All of the plasmids integrated into the Giles site. Tables 3 and 4 show primers used for plasmid construction and verification of correct integration.

Table 1. *M. smegmatis* Strains, Plasmids, and Descriptions: FLAG-Containing Constructs

Strain	Plasmid	Plasmid Description
SS-M_0947, 0948, 0949	pSS504	*Pmyc1 tetO* promoter + *Pymc1* 5’ UTR + 1x 5' Flag tag + *yfp*
SS-M_0950, 0965, 0966	pSS517	*Pmyc1 tetO* promoter + *Pymc1* 5’ UTR + 1x 5' Flag tag + 4x AlaGCC+ *yfp*
SS-M_0951, 0952, 0967	pSS518	*Pmyc1 tetO* promoter + *Pymc1* 5’ UTR + 1x 5' Flag tag + 4x ArgCGC+ *yfp*
SS_M_0953, 0954, 0955	pSS519	*Pmyc1 tetO* promoter + *Pymc1* 5’ UTR + 1x 5' Flag tag + 4x ArgCGG+ *yfp*
SS-M_0956, 0957, 0958	pSS520	*Pmyc1 tetO* promoter + *Pymc1* 5’ UTR + 1x 5' Flag tag + 4x ArgCGT+ *yfp*
SS-M_0959, 0960, 0961	pSS521	*Pmyc1 tetO* promoter + *Pymc1* 5’ UTR + 1x 5' Flag tag + 4x GlyGGC+ *yfp*
SS-M_0962, 0963, 0964	pSS522	*Pmyc1 tetO* promoter + *Pymc1* 5’ UTR + 1x 5' Flag tag + 4x ProCCG+ *yfp*

This table outlines the *M. smegmatis* strain number along with the corresponding plasmid identification number and description for inserts in the plasmids. *yfp* indicates the gene encoding yellow fluorescent protein.

Table 2. *M. smegmatis* Strains, Plasmids, and Descriptions: SigA-Containing Constructs

Strain	Plasmid	Plasmid Description
SS-M_1391, 1392, 1393	pSS721	*Pmyc1 tetO* promoter + *Pmyc1* 5′ UTR+ first 54 nt of *sigA*+ VVSS +*yfp*
SS-M_1394, 1395, 1396	pSS722	*Pmyc1 tetO* promoter + *Pmyc1* 5′ UTR+ first 54 nt of *sigA*+ VV+ 4x AlaGCC+SS +*yfp*
SS-M_1397, 1398, 1399	pSS723	*Pmyc1 tetO* promoter + *Pmyc1* 5′ UTR+ first 54 nt of *sigA*+ VV+ 4x ArgCGC+SS +*yfp*+ *HygR*
SS-M_1400, 1401, 1402	pSS724	*Pmyc1 tetO* promoter + *Pmyc1* 5′ UTR+ first 54 nt of *sigA*+ VV+ 4x ArgCGG+SS +*yfp*
SS-M_1403, 1404, 1405	pSS725	*Pmyc1 tetO* promoter + *Pmyc1* 5′ UTR+ first 54 nt of *sigA*+ VV+ 4x GlyGGC+SS +*yfp*
SS-M_1406,1407, 1408	pSS730	*Pmyc1 tetO* promoter + *Pmyc1* 5′ UTR+ first 54 nt of *sigA*+ VV+ 4x ProCCG+SS +*yfp*

This table outlines the *M. smegmatis* strain number along with the appropriate plasmid insertion number and description for inserts into the pSS359 backbone. *sigA* indicates the gene encoding sigma factor A. VVSS indicates valine (GTGGTG) and serine (TCGTCG) control insert, which flanks either side of the test codon. *yfp* indicates the gene encoding yellow fluorescent protein.

Table 3. Plasmids and Primers for Amplification, Integration Verification, and Sequencing

Plasmid	Primers for Adding Test Sequences	Primers for Verifying Plasmid Integration	Primers for Sequencing
pSS504	SSS2225, SSS2226	Right Junction: SSS1172 F, 1174 R Left Junction: SSS1173 F, 1174 R	SSS1994 F, SSS834 R
pSS517	SSS2227, SSS2228	Right Junction: SSS1172 F, 1174 R Left Junction: SSS1173 F, 1174 R	SSS1994 F, SSS834 R
pSS518	SSS2229, SSS2230	Right Junction: SSS1172 F, 1174 R Left Junction: SSS1173 F, 1174 R	SSS1994 F, SSS834 R
pSS519	SSS2231, SSS2232	Right Junction: SSS1172 F, 1174 R Left Junction: SSS1173 F, 1174 R	SSS1994 F, SSS834 R
pSS520	SSS2233, SSS2234	Right Junction: SSS1172 F, 1174 R Left Junction: SSS1173 F, 1174 R	SSS1994 F, SSS834 R
pSS521	SSS2235, SSS2236	Right Junction: SSS1172 F, 1174 R Left Junction: SSS1173 F, 1174 R	SSS1994 F, SSS834 R
pSS522	SS2237, SSS2238	Right Junction: SSS1172 F, 1174 R Left Junction: SSS1173 F, 1174 R	SSS1994 F, SSS834 R
pSS721	SSS3102, SSS3103	Right Junction: SSS1172 F, 1174 R Left Junction: SSS1173 F, 1174 R	SSS1412 F, SSS834 R
pSS722	SSS3130, SSS3131	Right Junction: SSS1172 F, 1174 R Left Junction: SSS1173 F, 1174 R	SSS1412 F, SSS834 R
pSS723	SSS3132, SSS3133	Right Junction: SSS1172 F, 1174 R Left Junction: SSS1173 F, 1174 R	SSS1412 F, SSS834 R
pSS724	SSS3134, SSS3135	Right Junction: SSS1172 F, 1174 R Left Junction: SSS1173 F, 1174 R	SSS1412 F, SSS834 R
pSS725	SSS3138, SSS3139	Right Junction: SSS1172 F, 1174 R Left Junction: SSS1173 F, 1174 R	SSS1412 F, SSS834 R
pSS730	SSS3156, SSS3157	Right Junction: SSS1172 F, 1174 R Left Junction: SSS1173 F, 1174 R	SSS1412 F, SSS834 R

This table outlines the plasmids and their corresponding primers used for insert amplification, *M. smegmatis* integration verification, and sequencing in *E. coli*.

Table 4. Primer Descriptions and Sequences

Primer	Description	Sequence (5' → 3')
SSS2225	Forward primer to amplify 1x 5' Flag tag (Giles site) + *yfp* and pSS303	AAGGACGATGACGATAAGGCCA GCGATAGCACTGAG
SSS2226	Reverse primer to amplify 1x 5' Flag tag (Giles site) + *yfp* and pSS303	ATCGTCATCGTCCTTATAGTCCA TGATGTATATCTCCTTCTTAAT
SSS2227	Forward primer to amplify 1x 5' Flag tag (Giles site) + 4x AlaGCC+ *yfp* and pSS303	GCCGCCGCCGCCGCCAGCGA TAGCACTGAG
SSS2228	Reverse primer to amplify 1x 5' Flag tag (Giles site) + 4x AlaGCC+ *yfp* and pSS303	GGCGGCGGCGGCCTTATCGT CATCGTCCTTATAGTC
SSS2229	Forward primer to amplify 1x 5' Flag tag (Giles site) + 4x ArgCGC+ *yfp* and pSS303	CGCCGCCGCCGCGCCAGCGA TAGCACTGAG
SSS2230	Reverse primer to amplify 1x 5' Flag tag (Giles site) + 4x ArgCGC+ *yfp* and pSS303	GCGGCGGCGGCGCTTATCGT CATCGTCCTTATAGTC
SSS2231	Forward primer to amplify 1x 5' Flag tag (Giles site) + 4x ArgCGG+ *yfp* and pSS303	CGGCGGCGGCGGGCCAGCG ATAGCACTGAG
SSS2232	Reverse primer to amplify 1x 5' Flag tag (Giles site) + 4x ArgCGG+ *yfp* and pSS303	CCGCCGCCGCCGCTTATCGT CATCGTCCTTATAGTC
SSS2233	Forward primer to amplify 1x 5' Flag tag (Giles site) + 4x ArgCGT+ *yfp* and pSS303	CGTCGTCGTCGTGCCAGCGA TAGCACTGAG

SSS2234	Reverse primer to amplify 1x 5' Flag tag (Giles site) + 4x ArgCGT+ *yfp* and pSS303	ACGACGACGACGCTTATCGTCATCGTCCTTATAGTC
SSS2235	Forward primer to amplify 1x 5' Flag tag (Giles site) + 4x GlyGGC+ *yfp* and pSS303	GGCGGCGGCGGCGCCAGCGATAGCACTGAG
SSS2236	Reverse primer to amplify 1x 5' Flag tag (Giles site) + 4x GlyGGC+ *yfp* and pSS303	GCCGCCGCCGCCCTTATCGTCATCGTCCTTATAGTC
SS2237	Forward primer to amplify 1x 5' Flag tag (Giles site) + 4x ProCCG+ *yfp* and pSS303	CCGCCGCCGCCGGCCAGCGATAGCACTGAG
SSS2238	Reverse primer to amplify 1x 5' Flag tag (Giles site) + 4x ProCCG+ *yfp* and pSS303	CGGCGGCGGCGGCTTATCGTCATCGTCCTTATAGTC
SSS3102	Forward primer to amplify VVSS and pSS359	GTGGTGTCGTCGGCTGCCAGCGATAGCACTG
SSS3103	Reverse primer to amplify VVSS and pSS359	CGACGACACCACGGTGCGCTTCACCGGCTC
SSS3130	Forward primer to amplify VV AlaGCC SS and pSS359	GCCGCCGCCGCCTCGTCGGCTGCCAGCGATAGCACTG
SSS3131	Reverse primer to amplify VV AlaGCC SS and pSS359	GGCGGCGGCGGCCACCACGGTGCGCTTCACCGGCTC
SSS3132	Forward primer to amplify VV ArgCGC SS and pSS359	CGCCGCCGCCGCTCGTCGGCTGCCAGCGATAGCACTG
SSS3133	Reverse primer to amplify VV	GCGGCGGCGGCGCACCACGGTGCGCTTCACCGGCTC

	ArgCGC SS and pSS359	
SSS3134	Forward primer to amplify VV ArgCGG SS and pSS359	CGGCGGCGGCGGTCGTCGGCTGCCAGCGATAGCACTG
SSS3135	Reverse primer to amplify VV ArgCGG SS and pSS359	CCGCCGCCGCCGCACCACGGTGCGCTTCACCGGCTC
SSS3138	Forward primer to amplify VV GlyGGC SS and pSS359	GGCGGCGGCGGCTCGTCGGCTGCCAGCGATAGCACTG
SSS3139	Reverse primer to amplify VV GlyGGC SS and pSS359	GCCGCCGCCGCCCACCACGGTGCGCTTCACCGGCTC
SSS3156	Forward primer to amplify VV ProCCG SS and pSS359	GCTGCCAGCGATAGCACTG
SSS3157	Reverse primer to amplify VV ProCCG SS and pSS359	ATCGCTGGCAGCCGACGACGGCGGCGGCGGCACCACGGTGCGCTTCACCGGCTC
SSS1172	Forward primer for checking Giles site right junction insertion in *M. smegmatis*	CTCCGAACTCCTCCGAAACC
SSS1173	Forward primer for checking Giles site left junction insertion in *M. smegmatis*	ACATATCTGTCGAAGCGCCC
SSS1174	Reverse primer for checking Giles site right junction insertion in *M. smegmatis*	TGACGATCAACTCCGCGGGGCCGGGCCA
SSS1175	Reverse primer for checking Giles site left junction insertion in *M. smegmatis*	CGGTGGATCCGCGCAACCTG

JR273	qPCR primer for *M. smegmatis sigA*	GACTACACCAAGGGCTACAAG
JR274	qPCR primer for *M. smegmatis sigA*	TTGATCACCTCGACCATGTG
SSS833	Forward qPCR primer for YFP	GATAGCACTGAGAGCCTGTT
SSS834	Reverse qPCR primer for YFP, Reverse primer for sequencing of constructs	CTGAACTTGTGGCCGTTTAC
SSS1994	Forward primer for sequencing of FLAG-containing constructs	TAACTACGTCGACATCGATACTCGCTGGTCC AGAACTGAT
SSS1412	Forward primer for sequencing of SigA-containing constructs	GGAAAAGAGGTCATCCAGGAAGAAATATTG GATCGTCGGC

This table provides a list of all of the primers used, their descriptions, and sequences.

Q5 PCR for amplification and insertion

pSS359 plasmid backbone was diluted to 4.72 ng/μL. For each insert, a separate PCR tube was prepared to 50 μL final volume as follows: 22.5 μL ultra-pure water, 10 μL Q5 buffer, 1 μL 10 μM dNTPs, 2.5 μL primer 1, 2.5 μL primer 2, 1 μL diluted pSS359, and 0.5 μL Q5 polymerase. Primers 1 and 2 for each insert are delineated in Table 4. In the thermocycler, the samples were 1) first brought to 98℃ for 2 minutes, 2) then were at 98℃ for 20 seconds. 3) Samples were cooled to 70℃ to anneal the primers for 30 seconds and 4) heated at 72℃ for 3.5 minutes longer to elongate the primers. 5) Steps 2 through 4 were repeated 34 times, 6) samples were kept at 72°C for 5 minutes, and 7) kept at 4°C for storage.

Gel

100 mL of TAE buffer and 1.00 g of agarose (Apex BioResearch, Catalog No. 20-102QD) were mixed and heated in the microwave for 2.5 minutes. After five minutes of cooling, 2.5 μL of EtBr was added and the solution was mixed by gently swirling the flask. After another five minutes, the gel was poured into the apparatus to solidify. After solidification, the gel was placed in a gel electrophoresis apparatus and covered in a TAE buffer. 5 μL of 6x DNA dye was added

to each PCR sample and mixed thoroughly. PCR samples were then loaded into the gel along with a 1 kb ladder. Gels were run at 140 volts for typically 20 minutes.

DNA extraction and purification

DNA extraction and purification from gel bands was completed with use of the Zymoclean Gel DNA Recovery Kit (Genesee Scientific Catalog No. 11-301). 500 μL of ADB was added to the 1.5 mL tube with gel band. The tubes were incubated at 55°C for 10 minutes or until gel bands were completely dissolved. Melted agarose was transferred to the Zymo spin column and centrifuged in a collection tube at max speed for 1 minute. Flowthrough was discarded. 200 μL of DNA wash buffer was added to the column and centrifuged at max speed for 30 seconds. Flowthrough was discarded and this wash step was repeated twice more. After washing, 8 μL of ultra-pure water was pipetted directly on the column and this incubated for five minutes at room temperature. The column was placed in a 1.5 mL tube and centrifuged at max speed for one minute to elute DNA.

HiFi assembly

2.5 μL of 2x HiFi DNA Assembly Master Mix (New England Biolabs Catalog No. E2621S) was added to 2.5 μL of purified PCR product, typically containing 0.02 to 0.5 pmols of DNA fragments. The mixture then was incubated in a thermocycler for 2.5 hours at 50°C, then stored at 4°C.

***E. coli* Transformation and Culture**

HiFi cloning mix from the assembly step above was added to 35 μL of NEB 5-alpha competent *E. coli* cells (New England Biolabs Catalog No. C2987U) and incubated together for 30 minutes at 4°C. The mixture was then heat shocked at 42°C for 30 seconds. Tubes were then placed back in ice for five minutes. At room temperature, 500 μL of recovery SOC Outgrowth Medium was added to the tubes and they were incubated for one hour at 37°C in a shaker at 200 rpm. After incubation, the tubes are centrifuged for one minute at 14000 rpm to pellet. Supernatant was poured off and the pellet was resuspended in the remaining supernatant. The transformed *E. coli* cells were then plated on LB plates with 200 μg/mL hygromycin B and incubated at 37°C for 24 hours. After 24 hours, individual colonies were picked from positive control plates and grown in

a liquid culture of LB broth and 200 μg/mL hygromycin B. These liquid cultures grew for 24 hours in a shaker at 37°C.

Plasmid Mini-Prep

Plasmid mini-preps were completed for *E. coli* culture of two colonies from each insert using the ZR Plasmid Mini-Prep-Classic Kit according to the manufacturer's instructions (Genesee Scientific Catalog No. 11-308AB).

Sequencing

Sanger sequencing was completed to confirm desired insertion in the *E. coli* strain.

Bacterial strains and growth conditions

Mycolicibacterium smegmatis mc^2155 constructed strains were grown in Difco Middlebrook 7H9 medium (BD Catalog No. 271310) with 1x albumin dextrose catalase (ADC), 0.2% glycerol, and 0.05% Tween 80. Cultures were shaken at 200 rpm and 37°C to an optical density at 600 nm (OD600) between 0.5 to 0.8 at the time of harvest. Cultures grown with 150 μg/mL hygromycin B were wrapped in aluminum foil to protect the photosensitive antibiotic.

Transformation by electroporation

Transformation by electroporation was used to integrate plasmid constructs into the competent *M. smegmatis* mc^2155 strain. Following electroporation, cells were suspended in 7H9 media and incubated at 37°C, 200 rpm for three hours.

Plating

Transformed *M. smegmatis* strains were plated on 7H10 agar (Millipore Middlebrook 7H10 Agar Base Catalog No. M0303-500G) plates with 150 ug/mL hygromycin. Plates were incubated for 96-120 hours at 37°C.

Colony checking PCR

Individual colonies were picked into 10 μL of water and boiled for 10 minutes at 95°C. Individual boiled templates were used in PCR reactions to confirm plasmid integration into the

M. smegmatis genome. Master mixes for the right and left integration junctions were made with the following volumes correlating to a single 10 μL PCR reaction: 0.5μL DMSO, 0.2μL primer 1 (10 μM), 0.2 μL primer 2 (10 μM), 0.2 μL dNTPs (10 mM), 1.0 μL Standard Taq Buffer (New England Biolabs Catalog No. M0273), 0.05 μL Taq polymerase (New England Biolabs Catalog No. M0273), 6.85 μL ultra-pure H_2O. For right junction verification, SSS1172 (forward) and SSS174 (reverse) were used. For left junction verification, SSS1173 (forward) and SSS1175 (reverse) were used. 1 μL of boiled template was added to 9 μL of each master mix. The thermocycler settings were as follows: 1) 95°C for 4:00 2) 95°C for 0:20 3) 55°C for 0:30 to anneal 4) 68°C for 1:00 for elongation 5) go to step 2 35 times 6) 68°C for 5:00 7) 4°C forever. PCR products were visualized on a 1% agarose gel with right junction products expected to be 387 base pairs (bp) and left junction products to be 582 bp.

Fluorescence Microscopy

Sample cultures, including SS-M_0486 (positive control) and SS-M_0497 (negative control), were grown and normalized to 0.6 OD. Slides were prepared by first pipetting about 50 μL of melted agarose onto each slide. 5 μL of each culture was then pipetted onto the agar of its respective slide and the coverslip was placed on top. Images were taken using fluorescence microscopy (Zeiss ApoTome Imager.Z1). White light normalization was set to 4000 for all images.

RNA extraction

Cultures were grown and pellets were flash frozen using liquid nitrogen at an OD of 0.5 to 0.8 and samples within each batch were within 0.1 OD of each other. Flash frozen cultures were stored at -80°C until day of RNA extraction. 1mL of Tri Reagent (Molecular Research Center, Inc. Catalog No. TR118) was added to each culture tube to resuspend the pellet then was transferred to corresponding 100 μm Zirconium bead-filled tube (OPS Diagnostics Catalog No. PFMB-100-100-12) on ice. The tubes were placed in the MP Fast Prep-24 (MP Biomedicals, LLC Catalog No. 6005.0) where they were shaken for three 30 second cycles of 7m/s with two minutes on ice after each cycle. 300 μL of chloroform was pipetted into each tube. RNA extraction was completed using the Direct-Zol mRNA miniprep (Zymo Research, catalogue R2052) following the manufacturer's instructions. The tubes were vortexed for 15 seconds

centrifuged for 20 minutes at 15,000 rpm and 4°C. After centrifuging, 500 μL of the clear top phase in the tube was pipetted into a corresponding 1.5 mL tube with 500 μL of 100% ethanol. The top phase and ethanol were mixed by pipetting up and down then loaded onto a Directzol column. The columns in a collection tube were centrifuged for 30 seconds at 15,000 rpm and 21°C. Flow through was pipetted into waste. 400 μL of RNA wash buffer (Zymo Research Catalog No. R1003-3-48) was pipetted on to each column and centrifuged for 30 seconds at 15,000 rpm 21°C. All following spins are done at the same conditions unless otherwise specified. Flowthrough was discarded. A DNase master mix was prepared using 75 μL DNase digestion buffer (Zymo Research Catalog No. E1010-1-16) and 5 μL DNase I (Zymo Research Catalog No. E1011-A) per sample. 80 μL of master mix was added to each sample, directly on the column matrix and samples incubated 30 minutes at room temperature. After 30 minutes, 400 μL of Directzol RNA pre-wash was added to the column and spun. Flowthrough was discarded and this step was repeated once more. Then 700 μL of RNA wash buffer was added to the column and spun down for 2 minutes. RNA was diluted using 50 μL of RNase-free water. RNA concentrations and absorbance ratios were measured on the Nanodrop. RNA samples were stored at -80°C.

cDNA synthesis and cleanup

To make single-stranded cDNA from the isolated RNA, a cDNA synthesis and cleanup procedure was completed. 600 ng of RNA was used from each sample in a total volume of 5.25 μL. For each sample, two tubes were prepared: one for reverse transcriptase (RT) and one with no RT. A primer master mix was made using 0.5 μL 100mM Tris pH 7.5, 0.17 μL of 3 μg/μL random primers, and 0.33 μL ultra-pure water per sample. 1 μL of master mix was added to each sample and incubated in the thermocycler for 10 minutes at 70°C. Samples were cooled for five minutes on ice. RT and no RT master mixes were then prepared. The RT master mix was as follows (per sample): 2 μL 5x Protoscript buffer, 0.5 μL 10 mM dNTPs, 0.5 μL 100 mM DTT, 0.25 μL RNase inhibitor, and 0.5 μL Protoscript II RT. The no RT master mix was the same except 0.5 μL of RT was replaced with 0.5 μL of ultra-pure water. 3.75 μL of master mix was added to each sample and incubated for 10 minutes at 25°C, 42°C for 18 hours, then 4°C forever.

For cDNA cleanup, an RNA degradation master mix was made using 0.5 M EDTA and 1 M NaOH is equal volume. 10 μL of master mix was added to each sample and mixed. Samples

were then incubated for 15 minutes at 65°C. Following this incubation cDNA was purified using the Monarch PCR and DNA Clean-Up Kit (New England Biolabs Catalog No. T1030S) and the manufacturer's instructions were followed.

Quantitative PCR (qPCR)

qPCR was conducted to measure the relative abundance of mRNA among the different strains. cDNA samples for qPCR were obtained from cDNA synthesis and cleanup, as described above. cDNA was first diluted to 1 ng/mL using 2 µL of the appropriate sample in pure water, then diluted to the desired concentration of 200 pg/µL also in pure water. All cDNA dilutions were done freshly on the day qPCR was to be performed. A 2.5 mM primer mix was made for each *sigA* and *yfp*. For *sigA*, the final primer mixture consisted of 2.5 µM JR273 and 2.5 µM JR274. For *yfp*, the final mixture consisted of 2.5 µM SSS833 and 2.5 µM SSS834 (see Table 4). A mastermix was made for each *sigA* and *yfp* and included 1µL of each respective primer mix, 5µL of iTaq SYBR Green Supermix (Bio-Rad: Catalog No. 172-5124), and 2 µL of ultra-pure water. iTaq was added last and the mixtures were kept on ice until use. 2 µL of cDNA were loaded into each designated well of a 96-well PCR plate (Axygen Catalog No. PCR-96M2-HS-C). 8 µL of appropriate master mix was then added to the wells and mixed gently by pipetting. Water controls were included on the plate, 3 for each primer mix. These controls were made by adding 2 µL of ultra-pure water with 8 µL of master mix.

After preparing the plate, it was covered with a sealing film. The plate was run on the QuantStudio 6 Pro Real-Time PCR System (ThermoFisher Scientific Catalog No. A43180). The reaction parameters were as follows: 1) 50℃ for 2 minutes, 2) 95℃ for 10 minutes, 3) 95℃ for 15 seconds, 4) 61℃ for 1 minute, 5) repeated steps 1 through 4 for 40 cycles. In each 61℃ step the SYBR Green fluorescence was recorded. The Quantification Cycle (C_q) threshold was set to 0.2. *yfp* samples were normalized to the corresponding *sigA* samples as follows: the difference, ΔC_q, was calculated, and then used to calculate relative expression, $2^{-\Delta Cq}$, of *yfp*. GraphPad Prism 9 was used for figure creation.

Flow Cytometry

Cultures were grown in biological triplicates of each strain to an OD of 0.5. Within 24 hours prior to the experiment, 7H9 media was filtered to remove precipitates that can otherwise appear

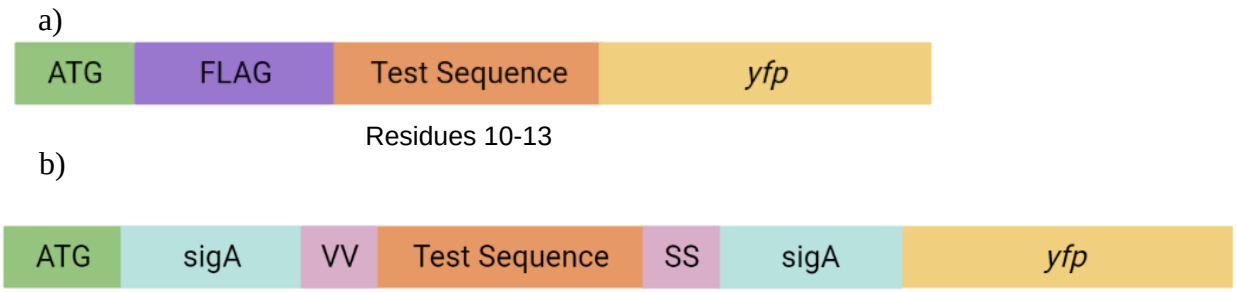
as events with similar forward and side scatter properties to *M. smegmatis* cells. On the day of experimentation, cultures were diluted to OD_{600} of 0.025 in 1 mL of the freshly filtered 7H9 media. The diluted cultures were then filtered with 5um filter needles to remove clumps. Samples were run on the Cytoflex Flow Cytometer with gain of FSC 500 and Violet SSC 50. Thresholds were set to Violet SSC-H 100,000 and FSC-H 40,000. 7H9 was run in between samples until the baseline event starting threshold was achieved (~50 events/second). To ensure uniform events to analyze, 5000 events within a small gate of the densest cells were captured for each sample.

Results and Discussion

Constructing *Mycolicibacterium smegmatis* strains with codon inserts embedded into YFP

In order to determine the impact of the ArgCGC codon content and context on gene expression in mycobacteria, gene constructs had to first be designed. A total of 13 plasmids were created based on one of two plasmid backbones: pSS303 and pSS359 (Nguyen et al., 2020). Two distinct plasmid backbones were used in order to test two different positions and sequence contexts for each of the test sequences. pSS303 encodes YFP with a C-terminal 6x His tag, driven by the pMyc1-tetO promoter and pMyc1-associated 5' UTR. pSS359 is identical except for the addition of the first 54 nt of the *M. smegmatis sigA* coding sequenced fused to the N-terminus of YFP. Although the promoter contains a tet operator, there is no Tet repressor present in the strains used and strong constitutive expression is therefore expected from these plasmids. The plasmids integrated into the Giles site within the *M. smegmatis* chromosome. A 1x FLAG tag sequence was fused to the N-terminus of pSS303, and test sequences of interest were placed between the FLAG sequence and the YFP coding sequence. The neutral sequence VVSS was inserted after the sequence encoding residue 16 of the *sigA* coding sequence, and test sequences of interest were placed between VV and SS. One plasmid was created for each of the following: no extra sequence controls (1x FLAG only or VVSS sequence only), the ArgCGC test codon, synonymous Arg codons, and frameshift codons with the same nucleotide content. Schematics of these gene constructs are shown in Figure 1. Control and test sequences are delineated in Table 5. Representative examples of pSS303 with FLAG and pSS359 with VVSS in Figures 2 through 5 show the complete nucleotide and amino acid sequences of the promoter, 5' UTR, and coding sequences. Plasmid integration into the *M. smegmatis* chromosome was confirmed by left and right junction checking gels (see Figure 6).

Figure 1. Gene Construct Schematics



This figure has coding sequence schematics for the a) FLAG-containing and b) SigA-containing constructs. In the FLAG-containing constructs, the test sequence is preceded by a common 1x FLAG sequence and is fused to *yfp*. The test codons comprise amino acid residues 10 through 13. In the SigA-containing constructs, the test sequence is inserted at nucleotide 54 of the *sigA* coding sequence, between a threonine and alanine, and fused to *yfp*. This alters the positioning of the test codons relative to translation initiation compared to the first set; test sequence in now at residues 20 through 23.

Table 5. Gene Construct Test Sequences, Plasmids, & *M. smegmatis* Strains

This table provides a description of the gene constructs including the description of each control and test sequence with actual sequences also provided. The sequences shaded in blue indicate FLAG-containing constructs and the pink shaded sequences are the SigA-containing constructs. Each codon is 4x in the test sequence. The 1x FLAG and VVSS control sequences are also included. The new plasmid created from the construct is denoted in the plasmid column and the assigned strain numbers are in the strain column.

Description	Sequence	Plasmid	Strain
FLAG	**GAC TAT AAG GAC GAT GAC GAT AAG**	pSS504	SS-M_0947, 0948, 0949
AlaGCC	GCC GCC GCC GCC	pSS517	SS-M_0950, 0965, 0966
ArgCGC	CGC CGC CGC CGC	pSS518	SS-M_0951, 0952, 0967
ArgCGG	CGG CGG CGG CGG	pSS519	SS-M_0953, 0954, 0955
ArgCGT	CGT CGT CGT CGT	pSS520	SS-M_0956, 0957, 0958
GlyGGC	GGC GGC GGC GGC	pSS521	SS-M_0959, 0960, 0961
ProCCG	CCG CCG CCG CCG	pSS522	SS-M_0962, 0963, 0964
VVSS	**GUG GUG UCG UCG**	pSS721	SS-M_1391, 1392, 1393
AlaGCC	GCC GCC GCC GCC	pSS722	SS-M_1394, 1395, 1396
ArgCGC	CGC CGC CGC CGC	pSS723	SS-M_1397, 1398, 1399
ArgCGG	CGG CGG CGG CGG	pSS724	SS-M_1400, 1401, 1402
GlyGGC	GGC GGC GGC GGC	pSS725	SS-M_1403, 1404, 1405
ProCCG	CCG CCG CCG CCG	pSS730	SS-M_1406, 1407, 1408

Figure 2. Promoter, 5' UTR, FLAG Tag, & YFP Nucleotide Sequence of FLAG-Containing Constructs

AGAAATATTGGATCGTCGGCACCGTCACGGCCGTGGGAGGCGGCACGATCCGCGACGTGATGATCGGCCGCATCCCCACGGTGCTGCGCAGT
GAGCTCTACGCCATCCCGGCGTTGATCTGTGCGTTCGCACGCACAGGCCCGGTGTGAGAAGGGTCTCTGACGAGCGGGAGAACTCCCTATCA
GTGATAGAGTTTGTCCTCCCTATCAGTGATAGATAGGCTCTGGGAGTACCCGTGTGTACGACCAGCACGGCATACATCATTTCGACGCCGAGAG
ATTCGCCGCCCGAAATGAGCACGATCCGCATGCTTAATTAAGAAGGAGATATACATCATGGACTATAAGGACGATGACGATAAGXXXXXXXXXXX
XGCCAGCGATAGCACTGAGAGCCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTG
CGCGGCGAGGGCGAGGGCGATGCCACCAACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACC
CTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGA
AGGCTACGTCCAGGAGCGCACCATCACCTTCAAGGACGACGGCACCTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTG
AACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTTCAACAGCCACAACGTCTAT
ATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACGTGGAGGACGGCGGCGTGCAGCTCGCCGACCACT
ACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAAC
GAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCACGGCAGCTCCGGCTCCAGCGGGTGTCATCACCATCA
CCATCACTAG

This figure shows the nucleotide sequence of the promoter highlighted in green, the 5' UTR in blue, the start codon in dark green, the FLAG Tag in purple and bolded, and a placeholder for the test sequence indicated by Xs highlighted in pink and underlined, as well as YFP in yellow for the FLAG-containing constructs.

Figure 3. FLAG Tag and YFP Amino Acid Sequence of FLAG-Containing Constructs

MDYKDDDDKXXXXASDSTESLFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKLICTTGKLPVP
WPTLVTTLGYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTITFKDDGTYKTRAEVKFEGDTLVNRIELK
GIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGGVQLADHYQQNTPIGDGPVLLPD
NHYLSYQSKLSKDPNEKRDHMVLLEFVTAAGITHGSSGSSGCHHHHHH

This figure shows the amino acid sequence of the methionine in dark green, the FLAG Tag highlighted in purple and bolded, and a placeholder for the 4x test sequence indicated by Xs highlighted in pink and underlined, as well as YFP in yellow for the FLAG-containing constructs.

Figure 4. Promoter, 5' UTR, SigA, VVSS, & YFP Nucleotide Sequence of SigA-Containing Constructs

AGAAATATTGGATCGTCGGCACCGTCACGGCCGTGGGAGGCGGCACGATCCGCGACGTGATGATCGGCCGCATCCCCACGGTGCTGCGCAGT
GAGCTCTACGCCATCCCGGCGTTGATCTGTGCGTTCGCACGCACAGGCCCGGTGTGAGAAGGGTCTCTGACGAGCGGGAGAACTCCCTATCAG
TGATAGAGTTTGTCCTCCCTATCAGTGATAGATAGGCTCTGGGAGTACCCGTGTGTACGACCAGCACGGCATACATCATTTCGACGCCGAGAGATT
CGCCGCCCGAAATGAGCACGATCCGCATGCTTAATTAAGAAGGAGATATACATCGTGGCAGCGACAAAGGCAAGCCCGGCAACCGAAGAGCCG
GTGAAGCGCACCGTGGTGXXXXXXXXXXXXTCGTCGGCTGCCAGCGATAGCACTGAGAGCCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAG
CTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGCGCGGCGAGGGCGAGGGCGATGCCACCAACGGCAAGCTGACCCTGAAGCTGATCT
GCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATG
AAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCACCTTCAAGGACGACGGCACCTACAAGACCCGC
GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAA
GCTGGAGTACAACTTCAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACGTG
GAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAG
CTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCACGGCAG
CTCCGGCTCCAGCGGGTGTCATCACCATCACCATCACTAG

This figure shows the nucleotide sequence of the promoter highlighted in green, the 5' UTR highlighted in blue, the start codon in dark green, SigA highlighted in purple, the VVSS control sequence bolded and in teal, and a placeholder for the test sequence indicated by Xs highlighted in pink and underlined, as well as YFP in yellow for the SigA-containing constructs.

Figure 5. SigA, VVSS, & YFP Amino Acid Sequence of SigA-Containing Constructs

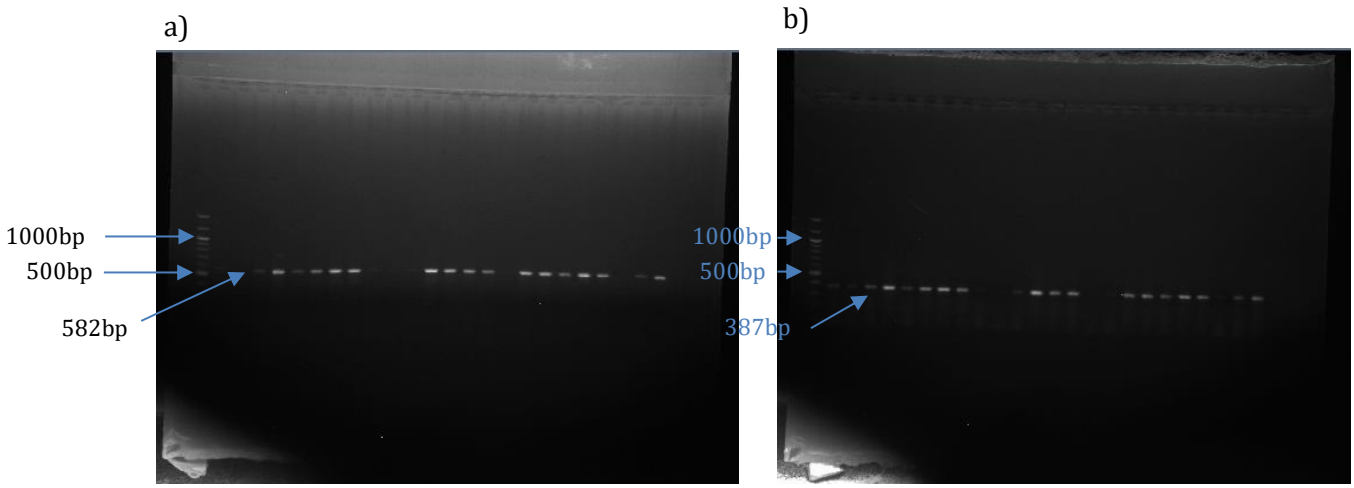
MAATKASPATEEPVKRTVVXXXXSSAASDSTESLFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLT
LKLICTTGKLPVPWPTLVTTLGYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTITFKDDGTYKTRAEVK
FEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGGVQLADHYQQ
NTPIGDGPVLLPDNHYLSYQSKLSKDPNEKRDHMVLLEFVTAAGITHGSSGSSGCHHHHHH

This figure shows the amino acid sequence of the methionine in dark green, SigA highlighted in purple, the VVSS control sequence bolded and in teal, and a placeholder for the test sequence indicated by Xs highlighted in pink and underlined, as well as YFP in yellow for the SigA-containing constructs.

Fluorescence microscopy was completed to ensure protein expression was detectable using the YFP fusion and to gather initial qualitative data for expression of the various FLAG-containing strains (see Figure 7). All of the strains showed fluorescence at 40x magnification, indicating the YFP fusion was expressed at detectable levels. The observed fluorescence corresponds to the amount of protein expression for the YFP gene, which in turn reports the relative amount of expression from each codon insert construct. The ArgCGT and ProCCG constructs showed

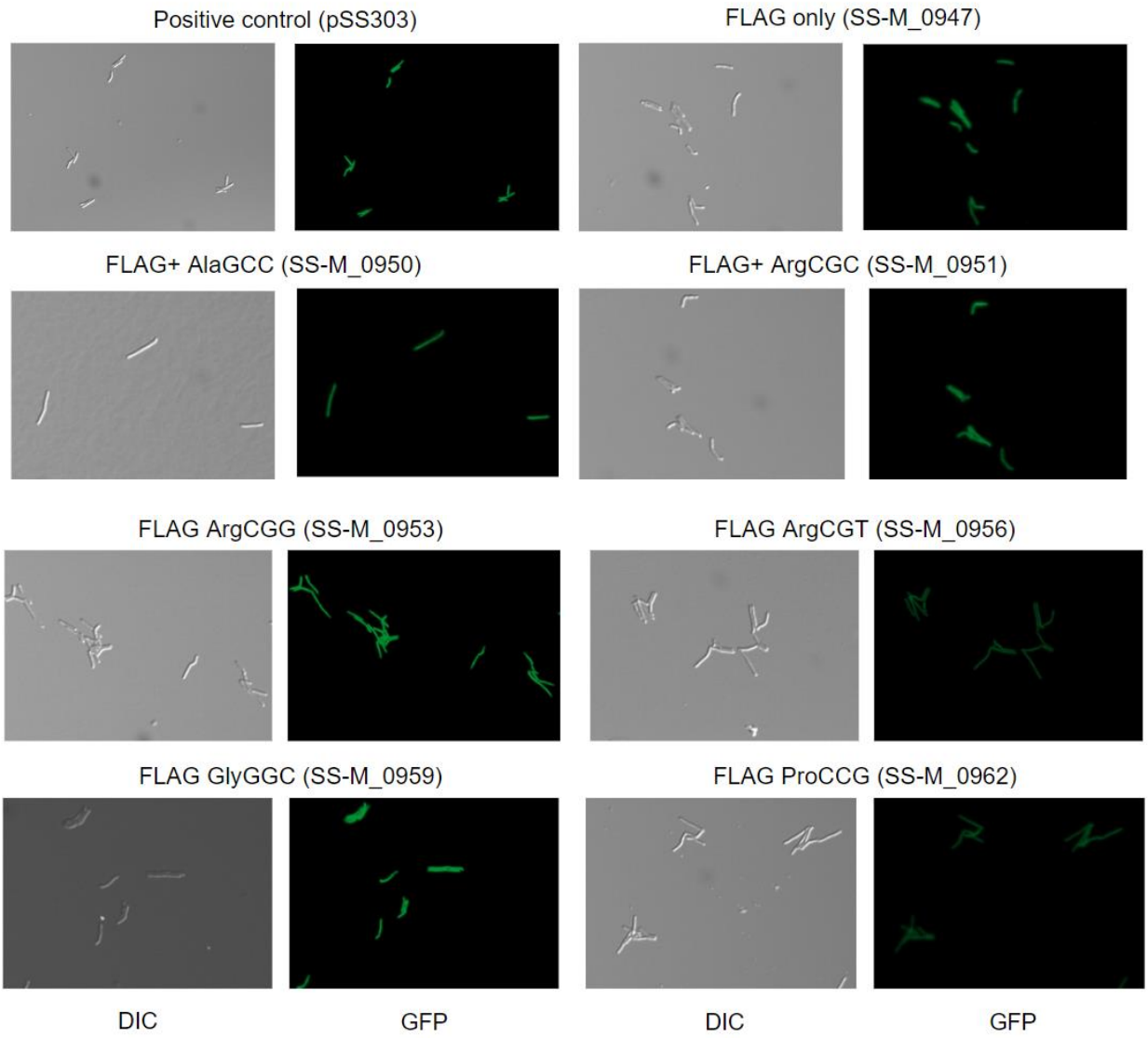
lower levels of fluorescence, indicating lower translation efficiency due to these inserts, as compared to the positive control and other strains.

Figure 6. *M. smegmatis* Integration Checking Gels



This figure shows the a) left junction Giles site integration checking gel and b) right junction Giles site integration checking gel used to confirm plasmid integration into *M. smegmatis* chromosome at the Giles site. The 100bp ladder is in the first well of each gel. For the left junction, the expected product is 582bp and for the right, 387bp.

Figure 7. Fluorescence Microscopy Images of Transformed *M. smegmatis* Strains



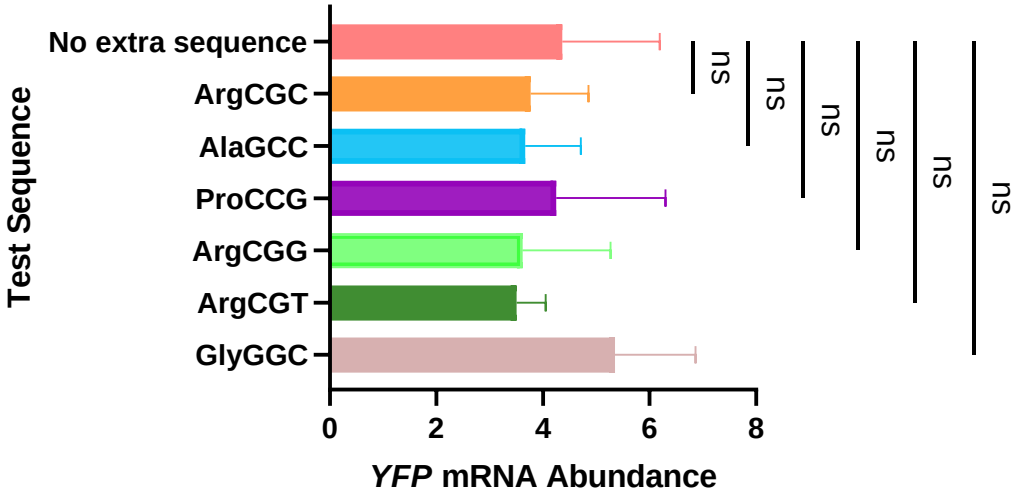
This figure shows the fluorescence microscopy images from the first set of constructs. Cells are shown with differential interference contrast (DIC) and the GFP fluorescence channel. The positive control is included (non-clumping SS-M_0486 with pSS303), showing maximum fluorescence and was used as a point of comparisons for the test sequence strains.

qPCR data reveal codon content may not have a significant impact on transcript stability

Following strain validation, strains were grown between OD 0.55 and 0.65 and flash frozen. RNA extraction and cDNA synthesis were completed. Quantitative reverse transcriptase PCR (qPCR) was conducted on cDNA of biological triplicate strains to test the impacts of codon

content on transcript abundance. Only the strains with FLAG insertions were analyzed by qPCR. There was no significant difference in *YFP* mRNA abundance detected for any of the strains compared to the no extra sequence control (see Figure 8). There was large error associated with all of the samples, so it is possible that there are differences not detectable here. For these specific samples, Hygromycin B was not used in liquid cultures throughout the entire workflow. Without this selective pressure, it is possible the plasmid was lost to some extent in the strains. This may explain the large error among triplicate strains.

Figure 8. qPCR Results for FLAG-Containing Constructs



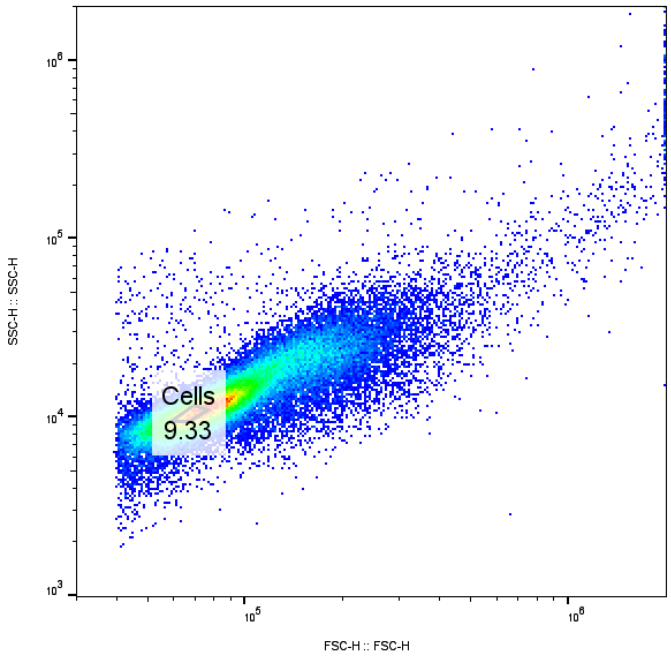
This figure displays *yfp* mRNA abundance measured by qPCR for each test sequence. Biological triplicate strains were run and error bars are shown for each construct. Each test sequence's mRNA abundance shows no significant difference, as denoted by the "ns" bars on the right, from the control FLAG only construct (no extra sequence).

Flow cytometry reveals that codon content and context have strong effects on translation efficiency in *M. smegmatis*

Flow cytometry was used to measure the relative protein abundance, which provides insight into translation efficiency of the various constructs. Biological triplicates of strains were grown to an

OD between 0.4 and 0.5, harvested, diluted, and filtered prior to flow cytometry. A preliminary gate was used to ensure a similar number of events was captured for each sample, and during analysis a new gate was drawn. An example of this cell size gating is shown in Figure 9.

Figure 9. Cell Size Gating for Flow Cytometry



This figure shows the forward and side scatter of *M. smegmatis* cells obtained in a single sample from flow cytometry. In the FlowJo analysis software, the densest portion of cells (in red) was gated for further analysis. This gate is indicated by a black outline and label "cells" box.

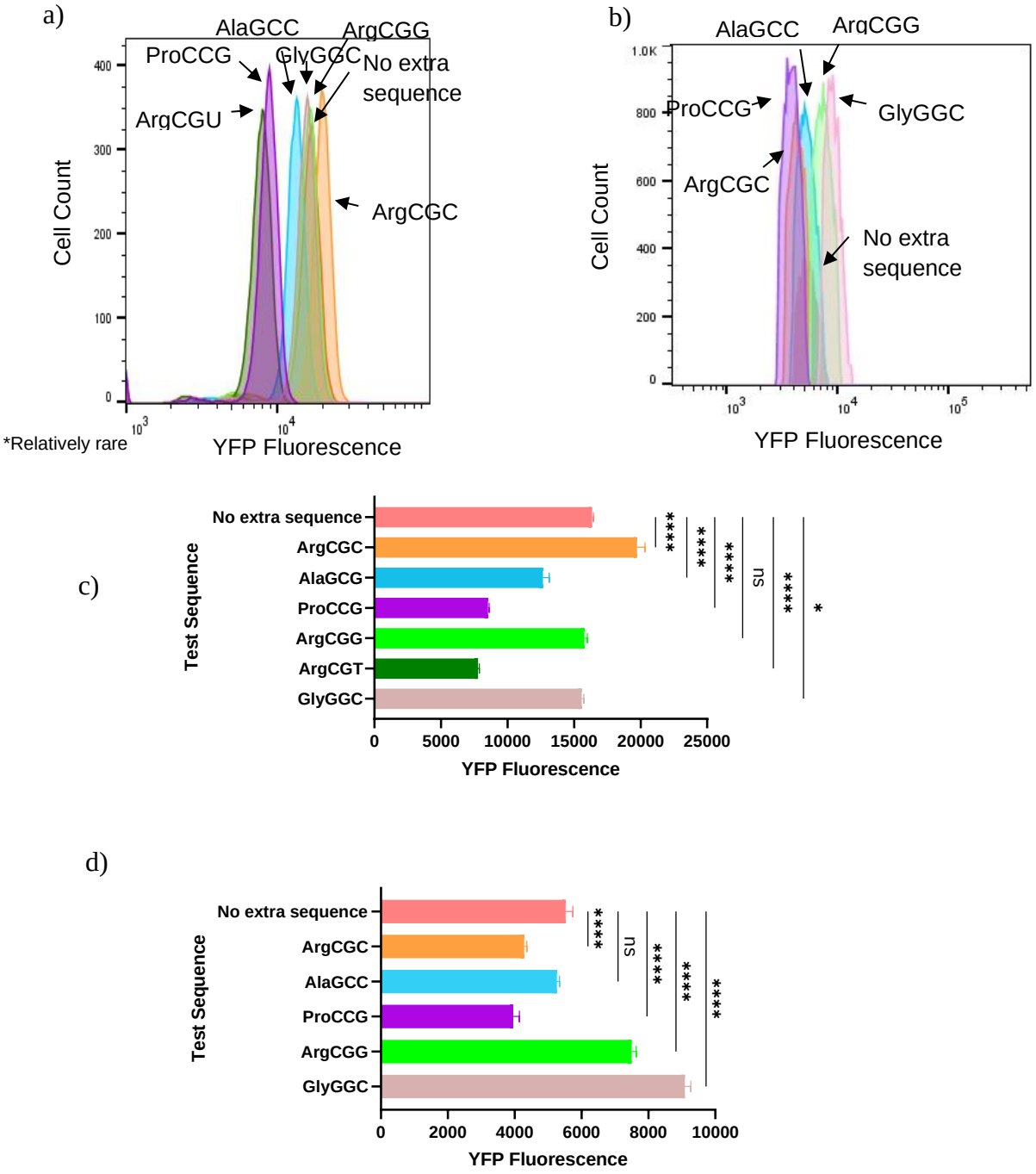
After gating, the YFP fluorescence was plotted against the cell count to produce histograms representing the relative YFP expression (Figure 10a & 10b). To more clearly visualize the differences in protein abundance among strains, bar charts were created with the average of the median YFP fluorescence of each biological triplicate for each construct (Figure 10c and 10d). Significantly less fluorescence was observed for the SigA-containing constructs compared to the FLAG-containing constructs, consistent with differences previously reported for the parental

plasmids of pSS359 and pSS303 (Nguyen et al., 2020). For the FLAG-containing constructs (see Figure 10c) ArgCGC is translated more efficiently than any of the other codons and this value is significantly different from the control sequence. This was unexpected because the ArgCGC codon was associated with fast degradation which we hypothesized would correlate to poor translation efficiency due to wobble base pairing. In this construct, the test sequence is located at amino acid residues +10 to +13 downstream of the translation initiation site. The AlaGCC codon was translated less efficiently than ArgCGC and the control sequence with a detected fluorescence of about 13000. ProCCG and ArgCGU showed very similar fluorescence values which are the lowest of all the samples in this set; they were both significantly different from the control. This corroborated the microscopy data that showed these two constructs with the lowest brightness of all the strains. We would expect this result because it is known that ArgCGU is relatively rare in mycobacteria (de Miranda et al., 2000) and proline is generally translated inefficiently due to its distinct side chain structure (Krafczyk et al., 2021). ArgCGG and GlyGGC showed similar fluorescence as compared to the control; however, the GlyGGC codon showed a slightly significant difference from the control while ArgCGG showed no significant difference.

The protein abundance from the SigA-containing constructs, cloned in the backbone of pSS359, is shown in Figure 10d. When the position of the codons relative to translation initiation and the codon context changed, there were significant changes in the protein abundance for the same codon. For example, in the FLAG-containing constructs, ArgCGC showed the greatest protein abundance of all the constructs, however, in this altered positioning and context the ArgCGC codon showed significantly less expression than the control sequence. This result would be expected based on the hypothesis that wobble base pairing causes decreased translation

efficiency, therefore, faster mRNA degradation. The AlaGCC codon resulted in protein expression that was not significantly different from that of the control sequence. This differed from the FLAG-containing construct where AlaGCC led to decreased translation efficiency compared to the control. ProCCG was, once again, translated the least efficiently of the entire set. ArgCGG showed significantly increased fluorescence compared to the control. This was different from the FLAG-containing construct where it showed no significant difference from the control. The new context and/or positioning improved translation efficiency for the ArgCGG codon. The GlyGGC codon similarly showed increased translation efficiency in this new context and positioning of the SigA-containing construct. ArgCGU was not assessed in this plasmid format.

Figure 10. Flow Cytometry Results for FLAG-Containing and SigA-Containing Constructs



This figure shows histograms of the YFP fluorescence for each test sequence for a) FLAG-containing and b) SigA-containing constructs. Peaks are labeled with arrows and the associated test sequence. Bar plots and statistical tests were done using GraphPad Prism 9. A one-way ANOVA was performed with multiple comparisons for the mean of each construct with the mean of the control "no extra sequence" construct (FLAG only for FLAG-containing constructs and VVSS for SigA-containing constructs). The degree of significance compared to the control is shown by the bars and asterisks to the right. "ns" denotes non-significant differences between constructs. c) For the first set of constructs ArgCGC shows increased YFP fluorescence relative to the FLAG only control, indicating increased

translation efficiency. d) For the second set, ArgCGC shows decreased YFP fluorescence relative to the VVSS only control, indicating decreased translation efficiency in this context and positioning.

Overall, these results suggest that codon content, as well as codon context and positioning relative to translation initiation impact translation efficiency in *M. smegmatis*. Translation of the ArgCGC codon appears to be dependent on codon position and context, as in the two construct sets very different translational profiles are observed. For future directions it would be important to measure the effects of position and context on mRNA abundance using qPCR for the SigA-containing constructs. This would provide more information on the relationship between translation efficiency and mRNA stability in mycobacteria. Furthermore, measuring mRNA degradation rates for both sets would further explain the transcript stability of the constructs as it may relate to codon content and context. Further assessing the impact of position and context may be informative. This could be completed by having the context remain the same but adding or deleting residues at the N-terminus to adjust the position of the test sequences relative to translation initiation. This may elucidate whether the differences shown between FLAG-containing and SigA-containing constructs was due to position, context, or a combination of the two factors. Finally, analyzing the ArgCGC locations in native *M. smegmatis* genes from which the hypothesis was generated would provide insight into the role position has on transcript stability and translation efficiency. This would be particularly important because in the machine learning program the ArgCGC codon correlated to fast mRNA degradation, which was not shown experimentally for each positioning.

Acknowledgements

A special thank you to Scarlet Shell, PhD for her continued guidance and support in this project. Additional thanks to Huaming Sun, M. Carla Martini, PhD, M. Natalia Alonso, PhD, Julia Ryan, and the remaining members of the Shell Lab for their instruction and assistance throughout my time in the lab. This project was sponsored in part by the WPI Arts and Sciences Advisory Board through the Summer Training in the Arts and Sciences (STAR) fellowship, as well as NIH and NSF grant funding.

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(3), 319–330.

Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., & Blüthgen, N. (2013). Efficient translation initiation dictates codon usage at gene start. *Molecular Systems Biology*, 9(1), 675. https://doi.org/10.1038/msb.2013.32

Bernstein, J. A., Lin, P.-H., Cohen, S. N., & Lin-Chao, S. (2004). Global analysis of Escherichia coli RNA degradosome function using DNA microarrays. *Proceedings of the National Academy of Sciences*, 101(9), 2758–2763. https://doi.org/10.1073/pnas.0308747101

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

Bossi, L., & Roth, J. R. (1980). The influence of codon context on genetic code translation. *Nature*, 286(5769), 123–127. https://doi.org/10.1038/286123a0

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. *The EMBO Journal*, 17(16), 4790–4797. https://doi.org/10.1093/emboj/17.16.4790

Bricker, A. L., & Belasco, J. G. (1999). Importance of a 5′ Stem-Loop for Longevity of papA mRNA in Escherichia coli. Journal of Bacteriology, 181(11), 3587–3590.

Carneiro, R. L., Requião, R. D., Rossetto, S., Domitrovic, T., & Palhano, F. L. (2019). Codon stabilization coefficient as a metric to gain insights into mRNA stability and codon bias

and their relationships with translation. *Nucleic Acids Research*, 47(5), 2216–2228. https://doi.org/10.1093/nar/gkz033

Centers for Disease Control and Prevention. (2022). "Tuberculosis (TB)—Drug-Resistant TB". https://www.cdc.gov/tb/topic/drtb/defaμLt.htm

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

Chen, Y. G., Kowtoniuk, W. E., Agarwal, I., Shen, Y., & Liu, D. R. (2009). LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nature Chemical Biology*, 5(12), Article 12. https://doi.org/10.1038/nchembio.235

Cheng, J., Maier, K. C., Avsec, Ž., Rus, P., & Gagneur, J. (2017). Cis-regulatory elements explain most of the mRNA stability variation across genes in yeast. *RNA*, 23(11), 1648–1659. https://doi.org/10.1261/rna.062224.117

Chevance, F. F. V., Guyon, S. L., & Hughes, K. T. (2014). The Effects of Codon Context on In Vivo Translation Speed. *PLOS Genetics*, 10(6), e1004392. https://doi.org/10.1371/journal.pgen.1004392

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., … Barrell, B. G. (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, 393(6685), Article 6685. https://doi.org/10.1038/31159

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

de Miranda, A. B., Alvarez-Valin, F., Jabbari, K., Degrave, W. M., & Bernardi, G. (2000). Gene expression, amino acid conservation, and hydrophobicity are the main factors shaping codon preferences in Mycobacterium tuberculosis and Mycobacterium leprae. *Journal of Molecular Evolution*, 50(1), 45–55. https://doi.org/10.1007/s002399910006

Deana, A., & Belasco, J. G. (2005). Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes & development*, 19(21), 2526–2533. https://doi.org/10.1101/gad.1348805

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), 275. https://doi.org/10.1186/s12864-015-1482-8

Georgellis, D., Barlow, T., Arvidson, S., & von Gabain, A. (1993). Retarded RNA turnover in Escherichia coli: A means of maintaining gene expression during anaerobiosis. *Molecular Microbiology*, 9(2), 375–381. https://doi.org/10.1111/j.1365-2958.1993.tb01698.x

Goodman, D. B., Church, G. M., & Kosuri, S. (2013). Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science*, 342(6157), 475–479. https://doi.org/10.1126/science.1241934

Gottesman, S. (2004). The Small RNA Regulators of Escherichia coli: Roles and Mechanisms. *Annual Review of Microbiology*, 58(1), 303–328. https://doi.org/10.1146/annurev.micro.58.030603.123841

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 (Reading, England)*, 148(Pt 6), 1795–1803. https://doi.org/10.1099/00221287-148-6-1795

Hanson, G., & Coller, J. (2018). Codon optimality, bias and usage in translation and mRNA decay. *Nature Reviews Molecular Cell Biology*, 19(1), Article 1. https://doi.org/10.1038/nrm.2017.91

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., & Weissman, J. S. (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science*, 324(5924), 218–223. https://doi.org/10.1126/science.1168978

Johnston, H. M., Barnes, W. M., Chumley, F. G., Bossi, L., & Roth, J. R. (1980). Model for regulation of the histidine operon of Salmonella. *Proceedings of the National Academy of Sciences*, 77(1), 508–512. https://doi.org/10.1073/pnas.77.1.508

Jürgen, 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. *Molecular & General Genetics: MGG*, 258(5), 538–545. https://doi.org/10.1007/s004380050765

Khemici, V., Prados, J., Linder, P., & Redder, P. (2015). Decay-Initiating Endoribonucleolytic Cleavage by RNase Y Is Kept under Tight Control via Sequence Preference and Sub-

cellular Localisation. *PLOS Genetics*, 11(10), e1005577. https://doi.org/10.1371/journal.pgen.1005577

Kido, M., Yamanaka, K., Mitani, T., Niki, H., Ogura, T., & Hiraga, S. (1996). RNase E polypeptides lacking a carboxyl-terminal half suppress a mukB mutation in Escherichia coli. *Journal of Bacteriology*, 178(13), 3917–3925. https://doi.org/10.1128/jb.178.13.3917-3925.1996

Krafczyk, R., Qi, F., Sieber, A., Mehler, J., Jung, K., Frishman, D., & Lassak, J. (2021). Proline codon pair selection determines ribosome pausing strength and translation efficiency in bacteria. Communications Biology, 4(1), Article 1. https://doi.org/10.1038/s42003-021-02115-z

Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. *Science* (New York, N.Y.), 324(5924), 255–258. https://doi.org/10.1126/science.1170160

Li, G.-W., Oh, E., & Weissman, J. S. (2012). The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*, 484(7395), Article 7395. https://doi.org/10.1038/nature10965

Lin, W., Mandal, S., Degen, D., Liu, Y., Ebright, Y. W., Li, S., Feng, Y., Zhang, Y., Mandal, S., Jiang, Y., Liu, S., Gigliotti, M., Talaue, M., Connell, N., Das, K., Arnold, E., & Ebright, R. H. (2017). Structural Basis of Mycobacterium tuberculosis Transcription and Transcription Inhibition. *Molecular Cell*, 66(2), 169-179.e8. https://doi.org/10.1016/j.molcel.2017.03.001

Luciano, D. J., Levenson-Palmer, R., & Belasco, J. G. (2019). Stresses that Raise Np4A Levels Induce Protective Nucleoside Tetraphosphate Capping of Bacterial RNA. *Molecular Cell*, 75(5), 957-966.e8. https://doi.org/10.1016/j.molcel.2019.05.031

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. https://www.frontiersin.org/articles/10.3389/fmicb.2019.00591

Moffitt, J. R., Pandey, S., Boettiger, A. N., Wang, S., & Zhuang, X. (2016). Spatial organization shapes the turnover of a bacterial transcriptome. *ELife*, 5, e13065. https://doi.org/10.7554/eLife.13065

Neymotin, B., Ettorre, V., & Gresham, D. (2016). Multiple Transcript Properties Related to Translation Affect mRNA Degradation Rates in Saccharomyces cerevisiae. *G3 Genes|Genomes|Genetics*, 6(11), 3475–3483. https://doi.org/10.1534/g3.116.032276

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

Nilsson, G., Belasco, J. G., Cohen, S. N., & von Gabain, A. (1987). Effect of premature termination of translation on mRNA stability depends on the site of ribosome release. *Proceedings of the National Academy of Sciences*, 84(14), 4890–4894. https://doi.org/10.1073/pnas.84.14.4890

Ow, M. C., Liu, Q., & Kushner, S. R. (2000). Analysis of mRNA decay and rRNA processing in Escherichia coli in the absence of RNase E-based degradosome assembly. *Molecular Microbiology*, 38(4), 854–866. https://doi.org/10.1046/j.1365-2958.2000.02186.x

Park, H.-D., Guinn, K. M., Harrell, M. I., Liao, R., Voskuil, M. I., Tompa, M., Schoolnik, G. K., & Sherman, D. R. (2003). Rv3133c/dosR is a transcription factor that mediates the hypoxic response of Mycobacterium tuberculosis. *Molecular Microbiology*, 48(3), 833–843. https://doi.org/10.1046/j.1365-2958.2003.03474.x

Parker, R. (2012). RNA Degradation in *Saccharomyces cerevisae*. Genetics, 191(3), 671–702. https://doi.org/10.1534/genetics.111.137265

Presnyak, V., Alhusaini, N., Chen, Y.-H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K. E., Graveley, B. R., & Coller, J. (2015). Codon Optimality Is a Major Determinant of mRNA Stability. *Cell*, 160(6), 1111–1124. https://doi.org/10.1016/j.cell.2015.02.029

Richards, J., & Belasco, J. G. (2019). Obstacles to Scanning by RNase E Govern Bacterial mRNA Lifetimes by Hindering Access to Distal Cleavage Sites. *Molecular Cell*, 74(2), 284-295.e5. https://doi.org/10.1016/j.molcel.2019.01.044

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

Sa, E., P, B., & Jg, B. (1992). A 5'-terminal stem-loop structure can stabilize mRNA in Escherichia coli. Genes & Development, 6(1). https://doi.org/10.1101/gad.6.1.135

Samuels, A. N., Wang, E. R., Harrison, G. A., Valenta, J. C., & Stallings, C. L. (2022). Understanding the contribution of metabolism to Mycobacterium tuberculosis drug

tolerance. *Frontiers in Cellular and Infection Microbiology*, 12. https://www.frontiersin.org/articles/10.3389/fcimb.2022.958555

Schnappinger, D., Ehrt, S., Voskuil, M. I., Liu, Y., Mangan, J. A., Monahan, I. M., Dolganov, G., Efron, B., Butcher, P. D., Nathan, C., & Schoolnik, G. K. (2003). Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: Insights into the Phagosomal Environment. *Journal of Experimental Medicine*, 198(5), 693–704. https://doi.org/10.1084/jem.20030846

Shahbabian, K., Jamalli, A., Zig, L., & Putzer, H. (2009). RNase Y, a novel endoribonuclease, initiates riboswitch turnover in Bacillus subtilis. *The EMBO Journal*, 28(22), 3523–3533. https://doi.org/10.1038/emboj.2009.283

Sharp, J. S., & Bechhofer, D. H. (2003). Effect of Translational Signals on mRNA Decay in Bacillus subtilis. *Journal of Bacteriology*, *185*(18), 5372–5379. https://doi.org/10.1128/JB.185.18.5372-5379.2003

Smeulders, M. J., Keer, J., Speight, R. A., & Williams, H. D. (1999). Adaptation of Mycobacterium smegmatis to Stationary Phase. Journal of Bacteriology, 181(1), 270–283. https://doi.org/10.1128/JB.181.1.270-283.1999

Spencer, P. S., Siller, E., Anderson, J. F., & Barral, J. M. (2012). Silent Substitutions Predictably Alter Translation Elongation Rates and Protein Folding Efficiencies. *Journal of Molecular Biology*, 422(3), 328–335. https://doi.org/10.1016/j.jmb.2012.06.010

Subramaniam, A. R., DeLoughery, A., Bradshaw, N., Chen, Y., O'Shea, E., Losick, R., & Chai, Y. (2013). A serine sensor for multicellularity in a bacterium. *ELife*, 2, e01501. https://doi.org/10.7554/eLife.01501

Sunohara, T., Jojima, K., Tagami, H., Inada, T., & Aiba, H. (2004). Ribosome stalling during translation elongation induces cleavage of mRNA being translated in Escherichia coli. *The Journal of Biological Chemistry*, 279(15), 15368–15375. https://doi.org/10.1074/jbc.M312805200

Tuller, T., Waldman, Y. Y., Kupiec, M., & Ruppin, E. (2010). Translation efficiency is determined by both codon bias and folding energy. *Proceedings of the National Academy of Sciences*, 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

Večerek, B., Moll, I., & Bläsi, U. (2005). Translational autocontrol of the Escherichia coli hfq RNA chaperone gene. *RNA*, 11(6), 976–984. https://doi.org/10.1261/rna.2360205

Volkman, H. E., Pozos, T. C., Zheng, J., Davis, J. M., Rawls, J. F., & Ramakrishnan, L. (2010). Tuberculous Granuloma Induction via Interaction of a Bacterial Secreted Protein with Host Epithelium. *Science*, 327(5964), 466–469. https://doi.org/10.1126/science.1179663

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. Journal of Bacteriology, 176(6), 1683–1688. https://doi.org/10.1128/jb.176.6.1683-1688.1994

World Health Organization. (2022). Global Tuberculosis Report 2022. Retrieved January 16, 2023, from https://www.who.int/publications-detail-redirect/9789240061729