



The Impact of Leadered and Leaderless Gene Structures on Translation Efficiency, Transcript Stability, and Predicted Transcription Rates in *Mycobacterium smegmatis*

Tien G. Nguyen,^{a*} Diego A. Vargas-Blanco,^a Louis A. Roberts,^a Scarlet S. Shell^{a,b}

^aDepartment of Biology and Biotechnology, Worcester Polytechnic Institute, Worcester, Massachusetts, USA

^bProgram in Bioinformatics and Computational Biology, Worcester Polytechnic Institute, Worcester, Massachusetts, USA

Tien G. Nguyen and Diego A. Vargas-Blanco contributed equally to this work. Author order was determined on a basis of who initiated the project.

ABSTRACT Regulation of gene expression is critical for *Mycobacterium tuberculosis* to tolerate stressors encountered during infection and for nonpathogenic mycobacteria such as *Mycobacterium smegmatis* to survive environmental stressors. Unlike better-studied models, mycobacteria express ~14% of their genes as leaderless transcripts. However, the impacts of leaderless transcript structures on mRNA half-life and translation efficiency in mycobacteria have not been directly tested. For leadered transcripts, the contributions of 5' untranslated regions (UTRs) to mRNA half-life and translation efficiency are similarly unknown. In *M. tuberculosis* and *M. smegmatis*, the essential sigma factor, SigA, is encoded by a transcript with a relatively short half-life. We hypothesized that the long 5' UTR of *sigA* causes this instability. To test this, we constructed fluorescence reporters and measured protein abundance, mRNA abundance, and mRNA half-life and calculated relative transcript production rates. The *sigA* 5' UTR conferred an increased transcript production rate, shorter mRNA half-life, and decreased apparent translation rate compared to a synthetic 5' UTR commonly used in mycobacterial expression plasmids. Leaderless transcripts appeared to be translated with similar efficiency as those with the *sigA* 5' UTR but had lower predicted transcript production rates. A global comparison of *M. tuberculosis* mRNA and protein abundances failed to reveal systematic differences in protein/mRNA ratios for leadered and leaderless transcripts, suggesting that variability in translation efficiency is largely driven by factors other than leader status. Our data are also discussed in light of an alternative model that leads to different conclusions and suggests leaderless transcripts may indeed be translated less efficiently.

IMPORTANCE Tuberculosis, caused by *Mycobacterium tuberculosis*, is a major public health problem killing 1.5 million people globally each year. During infection, *M. tuberculosis* must alter its gene expression patterns to adapt to the stress conditions it encounters. Understanding how *M. tuberculosis* regulates gene expression may provide clues for ways to interfere with the bacterium's survival. Gene expression encompasses transcription, mRNA degradation, and translation. Here, we used *Mycobacterium smegmatis* as a model organism to study how 5' untranslated regions affect these three facets of gene expression in multiple ways. We furthermore provide insight into the expression of leaderless mRNAs, which lack 5' untranslated regions and are unusually prevalent in mycobacteria.

KEYWORDS 5' UTR, leaderless translation, *Mycobacterium tuberculosis*, mRNA stability, posttranscriptional control mechanisms, sigma factors, *smegmatis*, transcription

Citation Nguyen TG, Vargas-Blanco DA, Roberts LA, Shell SS. 2020. The impact of leadered and leaderless gene structures on translation efficiency, transcript stability, and predicted transcription rates in *Mycobacterium smegmatis*. J Bacteriol 202:e00746-19. <https://doi.org/10.1128/JB.00746-19>.

Editor Tina M. Henkin, Ohio State University

Copyright © 2020 Nguyen et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Scarlet S. Shell, sshell@wpi.edu.

* Present address: Tien G. Nguyen, The Broad Institute, Cambridge, Massachusetts, USA.

Received 29 November 2019

Accepted 19 February 2020

Accepted manuscript posted online 24 February 2020

Published 9 April 2020

The pathogen *Mycobacterium tuberculosis* has evolved numerous strategies to survive in different niches within the human host. Bacterial adaptation to these harsh environments is usually achieved by gene regulation, both transcriptionally and post-transcriptionally. While promoters play critical roles in gene regulation, other gene features and mechanisms have additional important regulatory roles. One such important gene feature is the 5′ untranslated region (5′ UTR), which contains the Shine-Dalgarno (SD) sequence within the ribosome binding site (RBS) and, therefore, can serve as a translation regulator (1–5). For example, 5′ UTR interactions with *cis* and *trans* elements, such as complementary sequences within the UTR or coding sequence, small RNAs (sRNAs), and RNA-binding proteins, can modulate protein synthesis by blocking or improving accessibility to the RBS (6–9). Importantly, it has been shown in *Escherichia coli* and other bacteria that transcription and translation are physically coupled, and thus 5′ UTR-mediated modulation of translation could have repercussions on transcription rate as well (10–14). Translation blocks in *Mycobacterium smegmatis* have been shown to decrease transcription as well (15), suggesting that transcription-translation coupling occurs in mycobacteria, although the extent and consequences are unknown.

The 5′ UTRs can also regulate gene expression by altering mRNA turnover rates. This can be a consequence of altered translation rates, as impairments to translation often lead to faster mRNA decay (16–22). In other cases, mRNA stability is directly affected by sRNA binding to 5′ UTRs or by UTR secondary structure (9, 23–28). In *E. coli*, the half-life of the short-lived transcript *bla* can be significantly increased when its native 5′ UTR is replaced with the 5′ UTR of *ompA*, a long-lived transcript (29–31). Conversely, deletion of *ompA*’s native 5′ UTR decreased its half-life by 5-fold (30). The longevity conferred by the *ompA* 5′ UTR was attributed to the presence of a nonspecific stem-loop as well as the specific RBS sequence (30–32). Secondary structure formation in 5′ UTRs has been shown to play a major role in transcript stability in other bacteria as well, such as for *ermC* in *Bacillus subtilis* (33, 34) and *pufBA* in *Rhodobacter capsulatus* (35–37). Moreover, obstacles that hinder the linear 5′ scanning function of RNase E (a major RNase in *E. coli* and mycobacteria) can prevent access to downstream cleavage sites, increasing transcript half-life (38). Such obstacles include the 30S ribosomal subunit bound to an SD-like site far upstream of the translation start site in one case (39). UTRs can also contain binding sites for the global regulator CsrA, which can both promote and prevent mRNA decay in *E. coli* (40). Although effects of 5′ UTRs on mRNA stability, translation, and transcription rate have been widely studied in more common bacterial systems, there is a paucity of information of the regulatory effects of 5′ UTRs in mycobacteria.

Compared to *E. coli* and most other well-studied bacteria, mycobacteria possess a large number of leaderless transcripts; approximately 14% of annotated genes are leaderless in both *M. smegmatis* and *M. tuberculosis* (41–43). Studies in *E. coli* have shown that translation of leadered and leaderless transcripts is functionally distinct (44–50), suggesting fundamental differences in their mechanisms of regulation. In contrast to *E. coli*, where leaderless transcripts are generally translated less efficiently (42, 51–54), leaderless transcripts in mycobacteria appear to be translated robustly (42, 43). However, direct comparisons of translation rates for leadered versus leaderless transcripts in mycobacteria have yet to be reported.

Among leadered transcripts, 5′ UTR lengths vary. We hypothesized that longer 5′ UTRs were more likely to play regulatory roles through modulation of translation, transcription rate, and mRNA turnover. One such long-leadered transcript in both *M. tuberculosis* and *M. smegmatis* encodes sigma factor alpha (*sigA*), the primary sigma factor in mycobacteria (55, 56). Here, we used the mycobacterial model *M. smegmatis* and a series of yellow fluorescent protein (YFP) reporters to investigate the effects of the *sigA* 5′ UTR as well as leaderless gene structures on transcription, translation, and mRNA half-life. We found that the *sigA* 5′ UTR caused lower translation efficiency, reduced mRNA half-life, and a higher predicted transcript production rate compared to those of a control 5′ UTR. Leaderless transcripts were translated at similar rates as those of transcripts bearing the *sigA* 5′ UTR and had similar half-lives but appeared to be

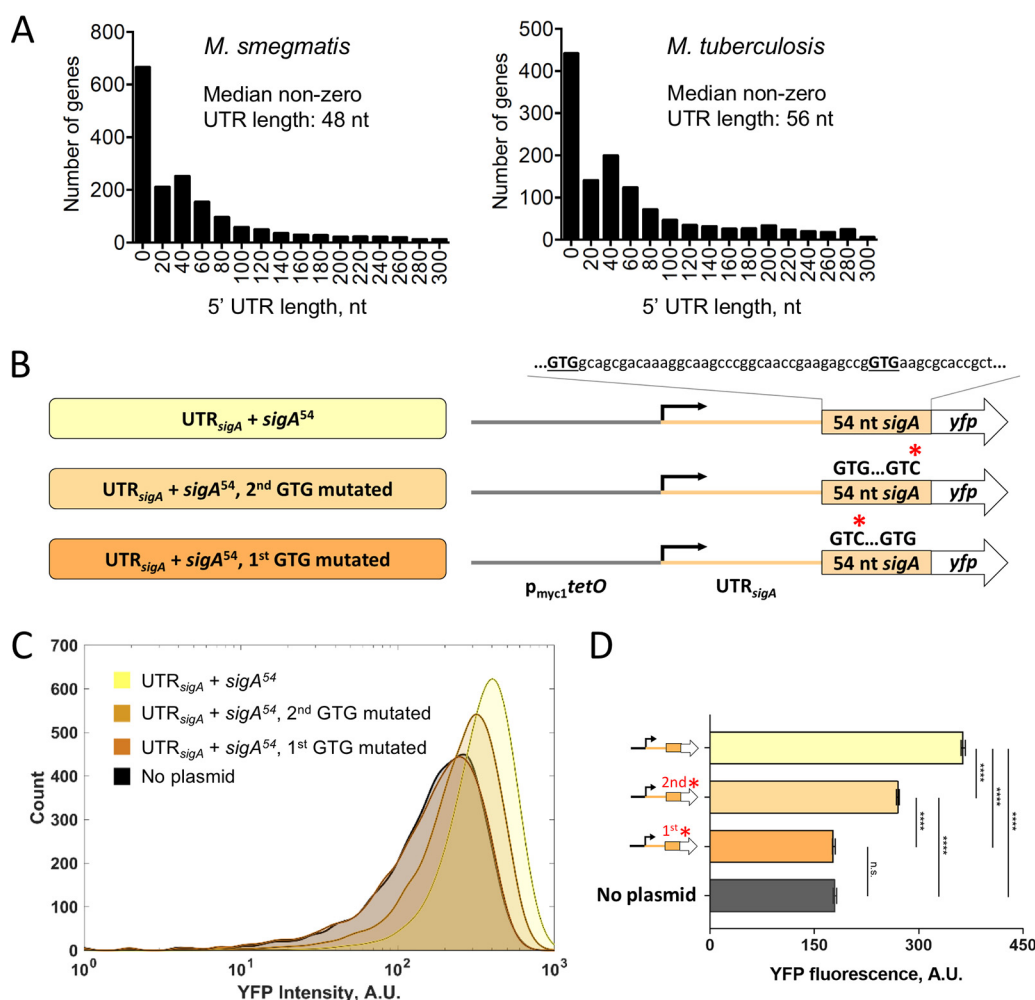


FIG 1 The *M. smegmatis* *sigA* gene has a longer-than-typical 5' UTR. (A) Distributions of 5' UTR lengths for *M. smegmatis* and *M. tuberculosis* genes reported to be transcribed from a single TSS (41, 42). (B) Constructs to confirm the predicted *sigA* translation start site. p_{myc1} tetO was described in reference 57. UTR_{sigA} denotes the 123-nt sequence between the experimentally determined TSS (41) and the annotated translation start site. (C) Flow cytometry with YFP-expressing constructs diagrammed in panel B. (D) Median fluorescence intensities determined by flow cytometry. Error bars denote 95% confidence interval (CI). Fluorescence intensities were compared by Kruskal-Wallis test followed by Dunn's multiple-comparison test. ****, $P < 0.0001$; ns, $P > 0.05$.

transcribed less efficiently, leading to lower steady-state mRNA and protein abundances. Our results highlight the potential of 5' UTRs to affect transcription efficiency as well as translation and mRNA half-life and support the idea that leaderless translation can be either more or less efficient than leadered translation in mycobacteria, depending on the characteristics of the leader. Alternative interpretations of our data are possible and would lead to different conclusions, particularly with respect to the efficiency of leaderless translation. These will be discussed.

RESULTS

Validation of the *sigA* 5' UTR boundaries. Transcription start site mapping has defined the 5' ends of 5' UTRs on a transcriptome-wide basis in both *M. smegmatis* and *M. tuberculosis* (41, 42). Using annotated translation start sites to define the 3' ends of the 5' UTRs, the median 5' UTR lengths in *M. smegmatis* and *M. tuberculosis* are 48 and 56 nucleotides (nt), respectively, after excluding leaderless genes (Fig. 1A; see also Table S1 in the supplemental material) (41, 42). The 5' UTR length distributions are skewed, with a mode of approximately 40 nt (Fig. 1A). We hypothesized that longer-than-average 5' UTRs are more likely to have regulatory roles and sought to investigate the

role of the 5' UTR of the *M. smegmatis* *sigA* gene. The *M. tuberculosis* *sigA* 5' UTR (128 nt) is also predicted to be longer than the median. To ensure that the predicted 5' UTR boundaries of *M. smegmatis* *sigA* were correct, we experimentally validated the predicted start codon at genome coordinate 2827625 in GenBank accession number [NC_008596](#), which resulted in a 123-nt UTR. A second GTG codon 39 nt downstream at 2827586 also had an appropriately positioned Shine-Dalgarno-like sequence and could conceivably be used as a start codon. We therefore made reporter constructs in which the strong constitutive promoter $p_{myc1tetO}$ (57) drove expression of a transcript containing the *sigA* 5' UTR and the sequence encoding YFP, with a C-terminal 6 $\times$ His tag and an N-terminal fusion of the sequence encoded by the first 54 nt of the annotated *sigA* coding sequence. We then individually mutated each of the two putative GTG start codons to GTC (Fig. 1B). Mutations of the first GTG to GTC reduced fluorescence to levels indistinguishable from autofluorescence in a strain that lacked the *yfp* gene (Fig. 1C and D, no plasmid). In contrast, mutation of the second GTG to GTC reduced fluorescence to an intermediate level (Fig. 1C and D). We therefore concluded that the first GTG is likely to be the predominant site of translation initiation, while the second GTG may affect expression levels but is not by itself sufficient to produce above-background expression. For subsequent experiments, we considered the first GTG to be the most likely start of the coding sequence and thereby define the *sigA* 5' UTR as 123 nt in length.

Assumptions made in subsequent data analysis. In subsequent sections, we will report data on mRNA abundance, mRNA half-life, and protein abundance for a series of reporter constructs. We will also report predicted transcription rates and apparent translation efficiencies, which are calculated from the abundance and half-life data. These calculated values rest upon a key assumption that most of the mRNA synthesized in the cell contributes to the measured abundance and half-life values. If this is not true, the data could be interpreted differently and different conclusions reached (58). These alternate interpretations are offered in Discussion.

The initial portion of the *sigA* coding sequence affects mRNA half-life and predicted transcript production rate. To capture 5' UTR-dependent effects on transcription, mRNA stability, and translation, we sought to investigate the role of the *sigA* 5' UTR (UTR_{*sigA*}) in the context of a *yfp* transcript. UTR-mediated regulation of translation sometimes involves base pairing of 5' UTR sequences with elements in the early portion of the coding sequence. Thus, we decided to include in our investigation the first 54 nt from the coding region of *sigA* (*sigA*⁵⁴). To determine if *sigA*⁵⁴ alone affected expression, we compared fluorescence from our YFP reporters with or without the *sigA*⁵⁴ N-terminal extension, independent from UTR_{*sigA*}. Transcription was driven by the $p_{myc1tetO}$ promoter for these and all constructs used in this study. While this semisynthetic promoter contains TetR binding sites, the strains used in this study did not encode the corresponding Tet repressor, and the promoter was therefore constitutively active. Where indicated, constructs included the $p_{myc1tetO}$ -associated 5' UTR (UTR _{$p_{myc1tetO}$}) as initially described in reference 57. To ensure that expression initiated only from the annotated promoter and not from spurious promoter-like sequences in UTR _{$p_{myc1tetO}$} or *sigA*⁵⁴, we built a control strain in which nt –53 through –1 of the promoter were deleted ($\Delta p_{myc1tetO}$) (Fig. 2A).

We first tested the impact of *sigA*⁵⁴ on YFP fluorescence intensity using UTR _{$p_{myc1tetO}$} . Interestingly, the *sigA*⁵⁴ strain was ~9-fold less fluorescent than the strain in which YFP lacked this N-terminal extension (Fig. 2B). To confirm that the reduced YFP fluorescence in the presence of *sigA*⁵⁴ indeed reflected reduced protein levels rather than altered YFP structure or intrinsic fluorescence, we measured protein levels directly by Western blotting. The Western blotting data were consistent with the flow cytometry result, showing an approximately 16-fold reduction of YFP levels with the inclusion of *sigA*⁵⁴ compared to the no-*sigA*⁵⁴ strain (Fig. 2C; see also Fig. S1 in the supplemental material).

To assess if the presence of *sigA*⁵⁴ affected *yfp* transcript levels, we conducted quantitative PCR (qPCR) for the same set of strains. Indeed, *sigA*⁵⁴ *yfp* levels were

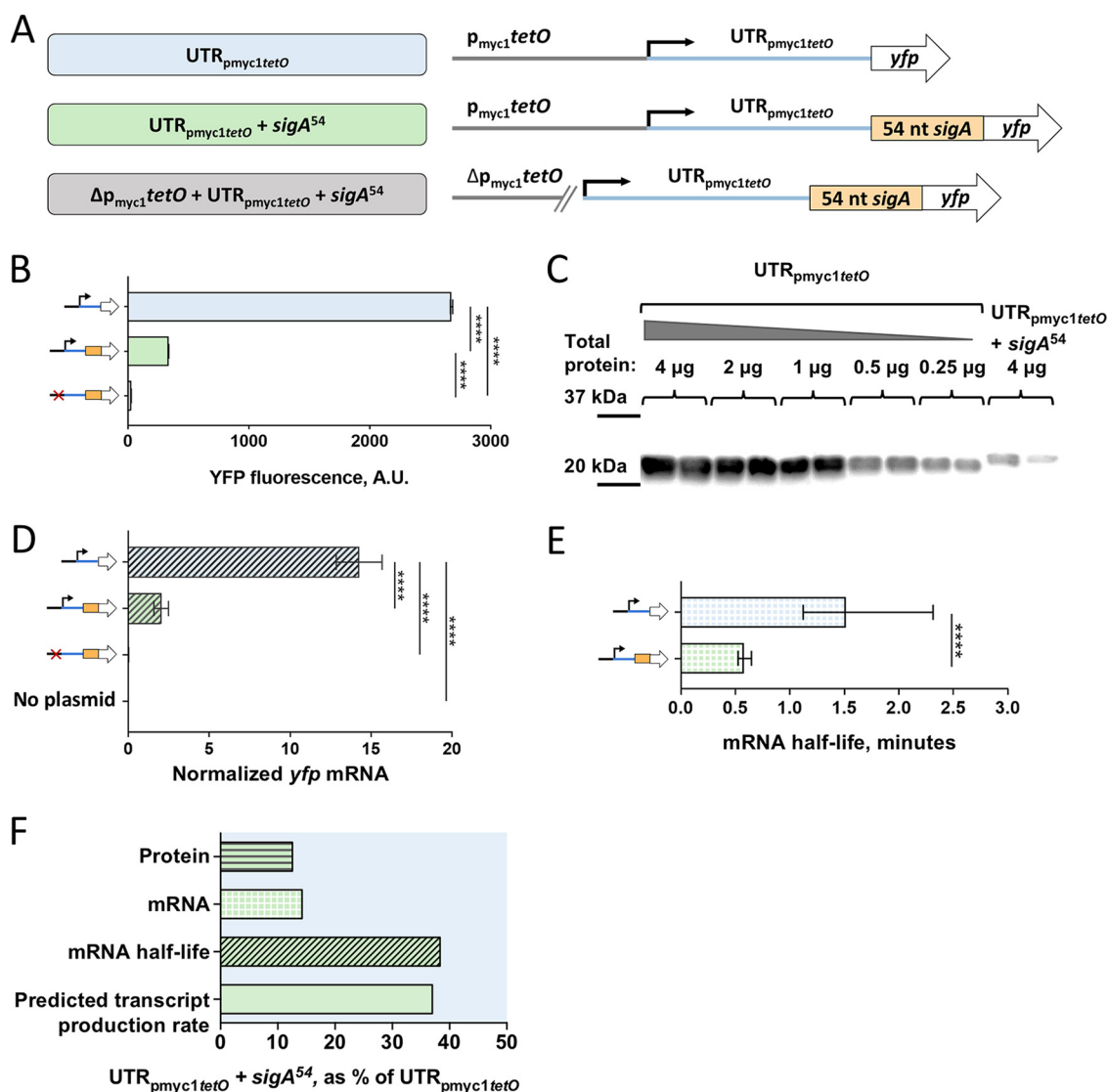


FIG 2 The first 54 nt of the *sigA* coding sequence affects transcript production rate and mRNA half-life. (A) Constructs transformed into *M. smegmatis* to determine the impact of the first 54 nt of the *sigA* coding sequence (*sigA*⁵⁴) on expression of a YFP reporter. (B) Median YFP fluorescence of strains bearing the constructs in panel A, determined by flow cytometry. Error bars denote 95% CI. Strains were compared by Kruskal-Wallis test followed by Dunn's multiple-comparison test. (C) Lysates from strains bearing constructs with and without *sigA*⁵⁴ were subject to Western blotting to detect the C-terminal 6 $\times$ His tag on the YFP. The mass of total protein loaded per lane is stated. (D) *yfp* mRNA abundance for strains bearing the indicated constructs, determined by qPCR and normalized to expression of endogenous *sigA*. Error bars denote standard deviation. Strains were compared by analysis of variance (ANOVA) with Tukey's honestly significant difference (HSD) test. (E) The half-lives of *yfp* mRNA produced from the indicated constructs were measured. Error bars denote 95% CI. Half-lives were compared using linear regression analysis ($n = 3$). (F) Protein abundance, mRNA abundance, mRNA half-life, and calculated transcript production rate for the construct containing *sigA*⁵⁴ are shown as a percentage of the values produced by a construct that lacks *sigA*⁵⁴ but is otherwise identical. ****, $P < 0.0001$.

approximately 6-fold lower than those of the *yfp* strain (Fig. 2D). This suggested that the decrease in YFP protein levels could be due to a reduction in *yfp* mRNA levels. Alternatively, the *sigA*⁵⁴ *yfp* transcript could be translated less efficiently, leading to reduced mRNA stability and thus lower steady-state abundance.

We then wondered if *sigA*⁵⁴ affected transcript abundance by increasing the rate of transcript decay or by decreasing the rate of transcription. Thus, we determined mRNA half-life for *yfp* with and without *sigA*⁵⁴. As shown in Fig. 2E, we estimated the half-life of *yfp* alone to be ~ 1.5 min and the half-life of *yfp* plus *sigA*⁵⁴ to be ~ 0.6 min. We concluded that the first 54 nt of *sigA* made the *yfp* transcript more susceptible to degradation. Knowing the abundance and decay rate of a transcript, the rate of

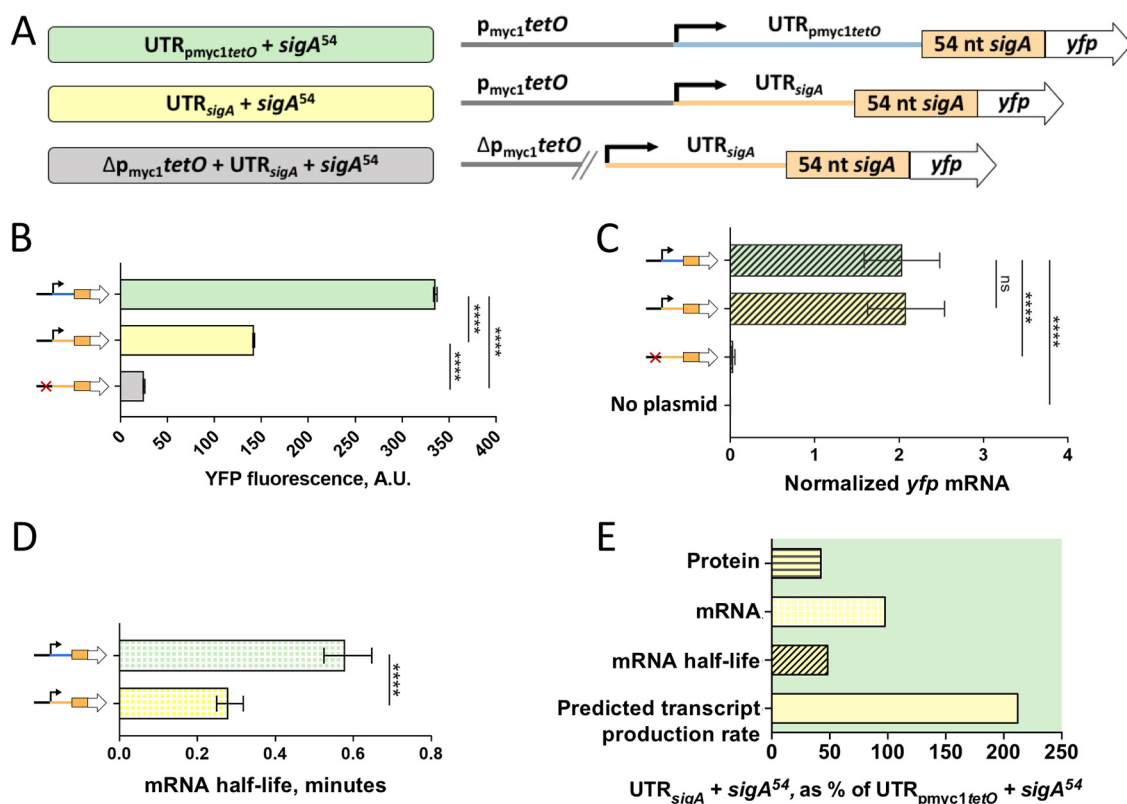


FIG 3 The *sigA* 5' UTR affects translation efficiency, mRNA half-life, and transcript production rate. (A) Constructs transformed into *M. smegmatis* to determine the impact of the *sigA* 5' UTR on expression of a YFP reporter. (B) Median YFP fluorescence of strains bearing the constructs in panel A, determined by flow cytometry. Error bars denote 95% CI. Strains were compared by Kruskal-Wallis test followed by Dunn's multiple-comparison test. (C) *yfp* mRNA abundance for strains bearing the indicated constructs, determined by qPCR and normalized to expression of endogenous *sigA*. Error bars denote standard deviation. Strains were compared by ANOVA with Tukey's HSD. (D) The half-lives of *yfp* mRNA produced from the indicated constructs were measured. Error bars denote 95% CI. Half-lives were compared using linear regression analysis ($n = 3$). (E) Protein abundance, mRNA abundance, mRNA half-life, and calculated transcript production rate for the construct containing the *sigA* 5' UTR are shown as a percentage of the values produced by a construct that contains the p_{myc1tetO}-associated 5' UTR. Note that some data shown in Fig. 2 are reproduced here to facilitate comparisons. ****, $P < 0.0001$; ns, $P > 0.05$.

transcription can be predicted mathematically (59). This predicted transcription rate encompasses initiation, elongation, and termination, and changes in the apparent transcription rate could therefore theoretically result from changes in any of those three facets. We will henceforth refer to this calculated rate as the predicted transcript production rate. The insertion of *sigA*⁵⁴ as an N-terminal extension for YFP appeared to reduce the *yfp* transcript production rate by approximately 60% (Fig. 2F).

The *sigA* 5' UTR affects transcript half-life, translation, and predicted transcript production rate. In order to assess the effects of UTR_{sigA} on transcription, mRNA stability, and translation, we replaced UTR_{pmyc1tetO} with UTR_{sigA} in our *sigA*⁵⁴ *yfp* reporters as shown in Fig. 3A. The presence of UTR_{sigA} led to an approximately 2-fold reduction in YFP fluorescence intensity when compared to the UTR_{pmyc1tetO} reporter strain (Fig. 3B). We wondered if the reduction in fluorescence attributed to UTR_{sigA} was associated with reduced *yfp* transcript abundance. However, qPCR revealed equivalent transcript levels for strains with UTR_{sigA} and UTR_{pmyc1tetO} (Fig. 3C), indicating that the reduced protein levels were more likely a consequence of reduced translation efficiency. Interestingly, *yfp* mRNA half-life was reduced to 0.28 min by the presence of UTR_{sigA} (Fig. 3D), suggesting that a higher transcript production rate is required to maintain the steady-state mRNA abundance that we observed (Fig. 3E). Taken together, our findings suggest that UTR_{sigA} may affect transcription, transcript decay, and translation. In Fig. 3E, we summarize these results as percentages of *yfp* transcript produc-

tion rate, mRNA abundance, mRNA half-life, and YFP protein levels relative to the UTR_{pmyc1tetO} *sigA*⁵⁴ strain.

We analyzed the sequences and predicted secondary structures of UTR_{sigA} and UTR_{pmyc1tetO} to investigate possible causes of the apparent difference in translation efficiency. The ribosome binding sites (RBSs) of these two UTRs have similar degrees of identity to a theoretically perfect mycobacterial SD sequence (the reverse complement of the 3' end of the 16S rRNA) (see Fig. S2A in the supplemental material). We noted that the spacing between the SD and start codon differed between the two UTRs (see Fig. S2A). However, both spacings are common among native *M. smegmatis* transcripts harboring these SD sequences (Fig. S2B), suggesting that neither spacing is particularly extreme. Secondary structure predictions by Sfold (60, 61) suggested that the UTR_{pmyc1tetO} SD is likely to be in a single-stranded loop while the UTR_{sigA} SD is likely to be partially base-paired (Fig. S2C and D), suggesting that there may be differences in SD accessibility for ribosome binding. Either the differences in SD start codon spacing or the differences in predicted secondary structure could potentially be responsible for the observed differences in apparent translation efficiency.

Leaderless mRNAs may be transcribed less efficiently. Leaderless transcripts are common in mycobacteria and were found to be associated with reduced protein abundance compared to that of leadered transcripts with near-consensus Shine-Dalgarno sites (43), suggesting that leaderless translation may be generally less efficient as was shown in *E. coli* (51–53). However, this hypothesis was not experimentally tested in mycobacteria. We therefore built two leaderless *yfp* reporters under the control of the p_{pmyc1tetO} promoter, with and without the *sigA*⁵⁴ N-terminal extension (Fig. 4A). When we compared YFP fluorescence between the leadered and leaderless reporters, we found that the leaderless fusions were substantially less fluorescent than those containing either 5' UTR, regardless of the presence of *sigA*⁵⁴ (Fig. 4B). The leaderless constructs also had reduced *yfp* mRNA levels compared to those of all of the leadered constructs (Fig. 4C). When comparing the leaderless constructs to the UTR_{pmyc1tetO} construct, protein levels were decreased to a greater extent than mRNA levels, (Fig. 4D), suggesting that the leaderless mRNAs were indeed translated less efficiently than mRNAs bearing UTR_{pmyc1tetO}. However, the difference in protein abundance from constructs without leaders and with UTR_{sigA} could be largely explained by the difference in mRNA levels (Fig. 4D), suggesting that leaderless and UTR_{sigA}-leadered mRNAs are translated with similar efficiencies.

To further evaluate the relationship between leader status and translation efficiency, we compared the relative abundances of proteins and mRNAs in *M. tuberculosis* using published quantitative proteomics data (62) and transcriptome sequencing (RNA-seq) data (42). For both leaderless transcripts and transcripts with 5' UTRs ≥15 nt in length, mRNA abundance and protein abundance were significantly correlated ($P < 0.0001$, Spearman's ρ) (Fig. 4E). We omitted transcripts with 1- to 14-nt UTRs because it is unknown if these behave more like leadered transcripts or more like leaderless transcripts with respect to the mechanisms of translation initiation. Linear regression of these correlations revealed that they were statistically indistinguishable for leadered versus leaderless transcripts, consistent with the idea that variability in translation efficiency among mycobacterial genes is largely driven by factors other than the presence or absence of a leader.

We wondered if the reduced abundance of the leaderless *yfp* transcripts relative to the UTR_{pmyc1tetO}-leadered transcripts was associated with reduced mRNA stability. Indeed, *yfp* half-lives for the UTR_{pmyc1tetO} leadered transcripts were longer than their leaderless counterparts (Fig. 4F). In contrast, the leaderless transcripts had half-lives similar to the transcript bearing UTR_{sigA} (Fig. 4F). Interestingly, leaderless transcripts with and without the *sigA*⁵⁴ N-terminal extension had equivalent half-lives. Taken together, the data indicate that the destabilizing effect of *sigA*⁵⁴ observed in Fig. 2F is dependent on the UTR_{pmyc1tetO} present in those constructs.

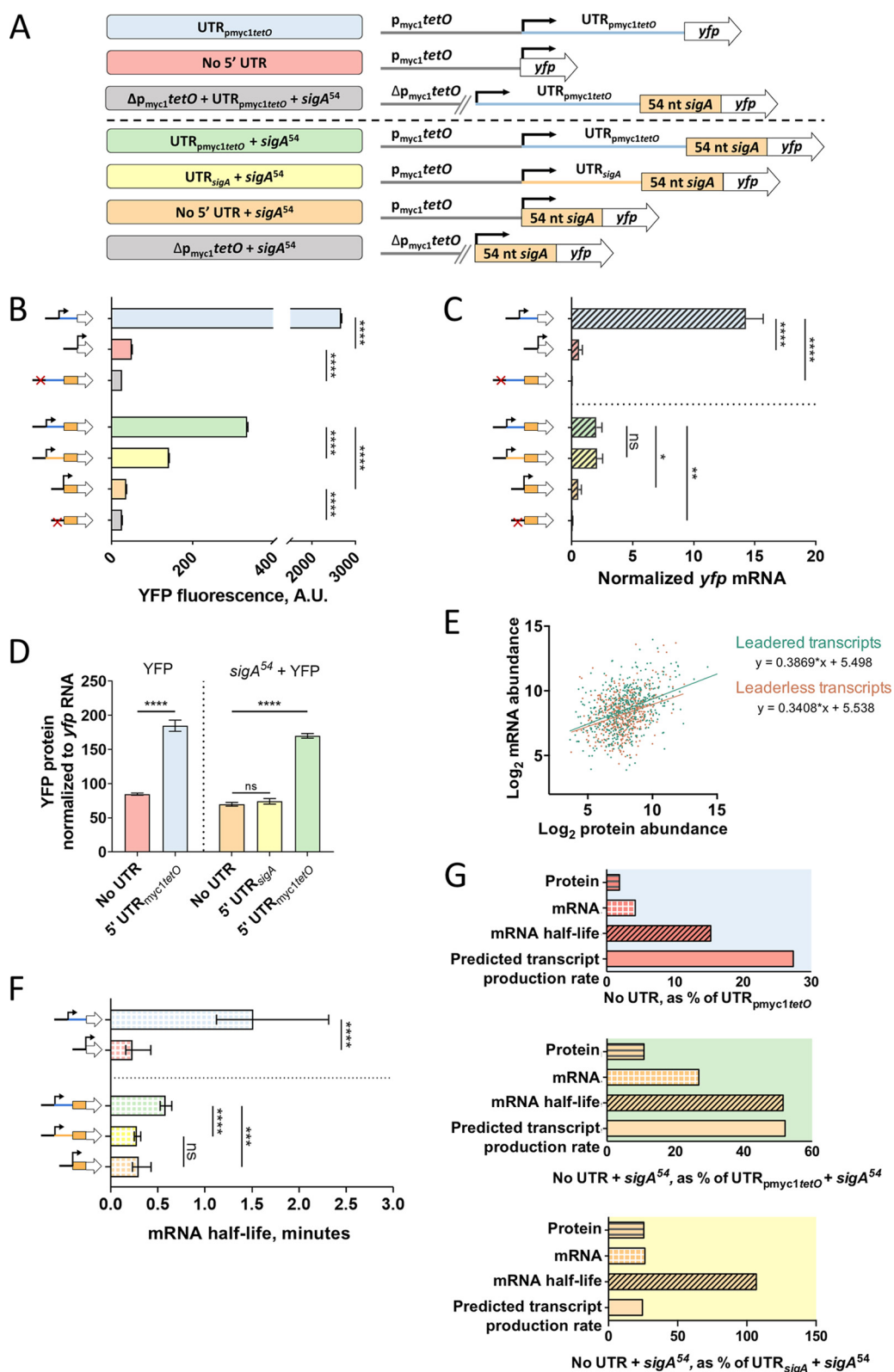


FIG 4 Leaderless transcripts have altered translation efficiencies, mRNA half-lives, and predicted transcript production rates compared to those of leadered controls. (A) Constructs transformed into *M. smegmatis* to compare leaderless versus leadered gene structures. (B) Median YFP fluorescence of strains bearing the constructs in panel A, determined by flow cytometry. Error bars denote 95% CI. Strains were compared by Kruskal-Wallis test followed by Dunn's multiple-comparison test. (C) *yfp* mRNA abundance for strains bearing the indicated constructs, determined by qPCR and normalized to expression of endogenous *sigA*. Error bars denote standard deviation. Strains were compared by ANOVA with Tukey's HSD.

(Continued on next page)

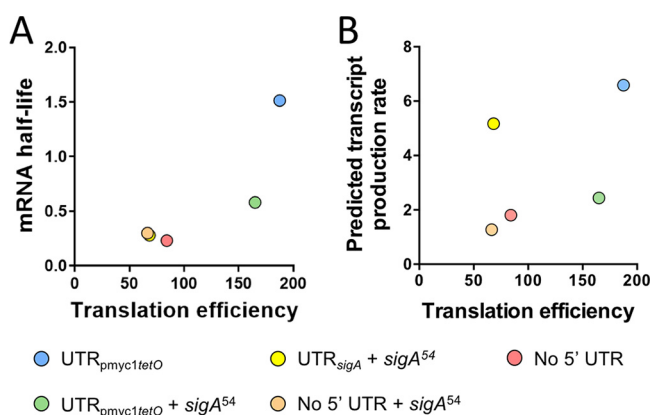


FIG 5 Translation efficiency is poorly correlated with mRNA half-life and predicted transcript production rate. Translation efficiency was defined as the ratio of protein abundance to mRNA abundance (arbitrary units). (A) Variability in mRNA half-life is largely not explained by variability in translation efficiency. (B) Variability in predicted transcript production rate is uncorrelated with translation efficiency.

The predicted *yfp* transcript production rates of the leaderless constructs were lower than those of their leadered counterparts (Fig. 4G). This is consistent with the idea that transcription-translation coupling can cause transcription rates to be altered as a function of translation efficiency (14). However, the *UTR_{sigA}*-leadered transcript appeared to be translated with a similar efficiency as the leaderless constructs (Fig. 4D) and yet had a substantially higher transcript production rate.

Translation efficiency is a poor predictor of mRNA half-life and transcript production rate. The five constructs described above had identical promoters but produced strains that varied widely with respect to protein abundance, mRNA abundance, mRNA half-life, translation efficiency, and predicted transcript production rates. Given the reported impacts of translation efficiency on mRNA stability in bacteria (16–22), we wondered to what extent the differences in half-life among our constructs were explained by differences in translation efficiency. We defined translation efficiency as follows:

$$\text{translation efficiency} = \frac{\text{protein abundance}}{[\text{mRNA}]}$$

However, the relationship between translation efficiency and mRNA half-life was weak (Fig. 5A), indicating that the variability in mRNA half-life was largely due to other factors. Translation rate has also been reported to affect transcription rate (14), but these two properties did not appear to be correlated in our constructs (Fig. 5B), suggesting that the differences in transcript production rate were not a consequence of differences in translation rate. Alternative interpretations of these and other analyses are discussed below.

DISCUSSION

The *sigA* transcripts in both *M. tuberculosis* and *M. smegmatis* were reported to have relatively short half-lives (15, 63), and we hypothesized that this property was conferred in part by the 5' UTR. We therefore sought to determine the impacts of the *M.*

FIG 4 Legend (Continued)

(D) Transcripts containing the *p_{myc1tetO}*-associated 5' UTR are translated more efficiently than leaderless transcripts or those containing the *sigA* 5' UTR. (E) Published *M. tuberculosis* mRNA abundance (42) and protein abundance (62) levels for genes that have a single TSS and are leaderless or have 5' UTRs of ≥ 15 nt. Protein and mRNA abundance were significantly correlated for both gene structures ($P < 0.0001$, Spearman's ρ). Linear regression analysis revealed that the slopes were statistically indistinguishable ($P = 0.44$). (F) The half-lives of *yfp* mRNA produced from the indicated constructs were measured. Error bars denote 95% CI. Half-lives were compared using linear regression analysis ($n = 3$). (G) Protein abundance, mRNA abundance, mRNA half-life, and calculated transcript production rate for leaderless transcripts compared to transcripts with 5' UTRs. Note that some data shown in Fig. 2 and 3 are reproduced here to facilitate comparisons. ****, $P < 0.0001$; ***, $P < 0.001$; ns, $P > 0.05$.

smegmatis *sigA* 5' UTR on expression and mRNA stability. Compared to a 5' UTR associated with high levels of protein expression and commonly used in mycobacterial expression vectors (57), the *sigA* UTR indeed conferred a shorter half-life as well as reduced translation efficiency (which could be the cause of the reduced half-life). However, the half-life of a *sigA*-leadered transcript was similar to that of a leaderless transcript. Insertion of part of the *sigA* coding sequence as an N-terminal translational fusion to our YFP reporter also caused a reduction in mRNA half-life. These findings suggest that the relative instability of the native *sigA* transcript is a product of multiple features, including the 5' UTR and regions of the coding sequence. However, this effect was not observed for a leaderless version of the translational fusion, indicating the effect is context dependent.

Our mRNA abundance and half-life data allowed us to calculate predicted transcript production rates. These calculated transcript production rates reflect the combined contributions of transcription initiation, elongation, and termination (with initiation and termination likely being the largest contributors), and our methodology did not allow us to distinguish between these processes. It is important to note that these rates were not directly measured but rather inferred from direct measurements of mRNA abundance and half-life; other interpretations of the data are therefore possible, as described below. Interestingly, the *sigA* 5' UTR appeared to increase transcript production rates compared to the p_{myc1} *tetO*-associated UTR or leaderless transcripts. The promoter sequence upstream of the transcription start site (TSS) was identical for all constructs. The effect of the 5' UTR on transcript production rate could therefore be mediated by at least three possible mechanisms. First, the sequence downstream of the TSS could affect RNA polymerase binding and therefore initiation as has been reported in *E. coli* (64). This is consistent with the finding that the *E. coli* RNA polymerase footprint extends 20 nt downstream of the TSS in both the open and closed initiation complexes (65) and advises caution when using TSS data to predict promoters. Second, the composition of the 5' UTR could affect rates of premature termination and therefore rates of production of full-length transcript as has also been reported (19, 66). We also note that the qPCR primers that we used to measure transcript abundance and half-life anneal to the coding sequence, so we only quantify transcripts that extend at least 89 nt into the coding sequence. Transcripts that terminate before this point are not detected in our experimental setup. Third, elongation rates could vary among constructs due to presence or absence of pause sites (67) or DNA binding proteins that form roadblocks (68). Viewed broadly, this result highlights the complexity of bacterial transcription.

An alternative interpretation of these data is possible if one considers the idea that newly synthesized mRNAs may be degraded nearly instantaneously if not immediately engaged by the translation machinery. To our knowledge first proposed in detail in reference 58, this hypothesis invokes a population of "dark matter" mRNAs that are synthesized but decay so rapidly that they are not detected by abundance and half-life measurements. In this model, measured half-lives reflect primarily the decay of those transcripts that are quickly engaged by ribosomes and therefore long enough lived to contribute to steady-state abundance. The model implies that steady-state abundance is a function not only of transcription rate and measured half-life, as assumed in our transcript production rate calculations, but also of partitioning between the translated pool and the "dark matter" pool. If more mRNA is partitioned into the translated pool, greater steady-state abundance could be achieved without increased transcription rates and vice versa. Importantly, changes in partitioning between these pools need not correspond with predictable changes in measured mRNA half-life or protein levels. For example, transcripts containing the *sigA* 5' UTR could conceivably engage ribosomes more quickly than transcripts containing the p_{myc1} *tetO*-associated UTR and therefore enter the translated pool at higher rates yet be more susceptible to degradation during or after translation, resulting in shorter half-lives and thus similar steady-state mRNA abundance as we observed (Fig. 3).

The relative efficiency of leaderless versus leadered translation in mycobacteria has

not been experimentally established. Proteomics data from *M. tuberculosis* suggested that proteins encoded on leaderless transcripts were less abundant than those encoded on leadered transcripts with evident SD sequences, but this difference appeared to be explained by differences in mRNA levels (43). A subsequently reported quantitative proteomics data set (62) allows for a more rigorous assessment of the relationship between mRNA abundance and protein abundance in *M. tuberculosis*. When comparing leaderless genes to leadered genes with a single TSS, we found that leaderless genes indeed on average had slightly but significantly lower levels of both mRNA and protein (Mann-Whitney tests for both, $P < 0.01$). However, the relationships between mRNA abundance and protein abundance did not differ for these two groups.

There are at least two ways to interpret protein/mRNA ratios. In Results and the figures, we refer to these ratios as a measure of translation efficiency on the grounds that they reflect the number of protein molecules produced per mRNA molecule. This interpretation rests upon the assumption that most mRNAs are stable enough to contribute to measurements of steady-state abundance. There are various reports of mutations or modifications that affect mRNA levels and protein levels differently (18, 19, 54, 66, 69), supporting the idea that translation efficiency can indeed be estimated by comparing protein levels to mRNA levels. Using this assumption and definition of translation efficiency, the *M. tuberculosis* data imply that there is no global difference in translation efficiency for leadered versus leaderless transcripts. Using the same assumptions, the small number of controlled comparisons that we report here support that idea; transcripts with the $p_{\text{myc1}}\text{tetO}$ -associated UTR were translated more efficiently than leaderless transcripts, but a transcript with the *sigA* UTR was translated with similar efficiency as its leaderless counterpart. Notably, the difference in translation efficiency between the two leadered transcripts might be attributable to differences in secondary structure rather than differences in favorability of the SD sequences.

A positive correlation between mRNA half-life and translation efficiency was reported for *E. coli* (70), consistent with the idea that translation may protect mRNAs from degradation. We did not observe such a correlation within our set of five transcripts, indicating that translation efficiency is not the primary driver of the variability in half-lives that we observed. However, a broad analysis of this relationship in mycobacteria is warranted.

The protein and mRNA abundance data must be interpreted differently if there is a “dark matter” mRNA pool that does not contribute to the mRNA abundance measurements. In this model, steady-state mRNA abundance is affected by translation efficiency as a consequence of partitioning between the translated pool and the “dark matter” pool. Translation efficiency would therefore affect steady-state mRNA abundance to a greater extent than that predicted by measured mRNA half-lives. Importantly, in this model, the relationship between protein abundance and mRNA abundance is not a reliable indicator of translation efficiency. In *M. tuberculosis*, lower average levels of both mRNA and protein from leaderless genes could therefore be a consequence of slower engagement of ribosomes and greater partitioning of mRNAs into the undetected “dark matter” pool.

It is prudent to note the assumptions underlying our definition of mRNA half-life. We assume that following transcription block by rifampin, the initial rapid decrease in *yfp* transcript abundance that we observe (see Fig. S3 in the supplemental material) indeed reflects the degradation rate for most *yfp* transcripts produced in an unperturbed cell. This assumption could be invalid if rifampin immediately alters mRNA degradation rates. If there are populations of “dark matter” mRNAs, the measured half-lives presumably reflect only the decay rate of those mRNAs that are engaged by ribosomes. We also note that, for most constructs, we observed a second, slow phase of mRNA decay (see Fig. S3 and Materials and Methods). This could reflect changes in mRNA decay in response to rifampin or reflect the presence of a second pool of transcripts with an inherently slower decay rate. As we could not distinguish between these possibilities, and the slow-decaying pool appeared to comprise at most 10% of the total, we did not

TABLE 1 Strains and plasmids

Plasmid	Strain (reference)	Characteristics
	<i>ΔMSMEG_2952</i> strain (71)	mc ² 155 <i>MSMEG_2952::Hyg^r</i>
pSS303	SS-M_0486	p _{myc1} <i>tetO</i> promoter + p _{myc1} 5' UTR + <i>yfp-6</i> ×His
pSS309	SS-M_0489	p _{myc1} <i>tetO</i> promoter + <i>sigA</i> 5' UTR + first 54 nt of <i>sigA</i> + <i>yfp-6</i> ×His
pSS310	SS-M_0493	p _{myc1} <i>tetO</i> promoter + no 5' UTR + <i>yfp-6</i> ×His
pSS314	SS-M_0497	p _{myc1} <i>tetO</i> promoter with a deletion of nt −53 through −1 + <i>yfp-6</i> ×His
pSS316	SS-M_0521	p _{myc1} <i>tetO</i> promoter + Δ1GTG <i>sigA</i> + first 54 nt of <i>sigA</i> + <i>yfp-6</i> ×His
pSS335	SS-M_0524	p _{myc1} <i>tetO</i> promoter + Δ2GTG <i>sigA</i> + first 54 nt of <i>sigA</i> + <i>yfp-6</i> ×His
pSS359	SS-M_0623	p _{myc1} <i>tetO</i> promoter + p _{myc1} 5' UTR + first 54 nt of <i>sigA</i> + <i>yfp-6</i> ×His
pSS360	SS-M_0626	p _{myc1} <i>tetO</i> promoter + no 5' UTR + first 54 nt of <i>sigA</i> + <i>yfp-6</i> ×His
pSS365	SS-M_0629	p _{myc1} <i>tetO</i> promoter with a deletion of nt −53 through −1 + first 54 nt of <i>sigA</i> + <i>yfp-6</i> ×His
pSS384	SS-M_0636	p _{myc1} <i>tetO</i> promoter with a deletion of nt −53 through −1 + <i>sigA</i> 5' UTR + first 54 nt of <i>sigA</i> + <i>yfp-6</i> ×His
pSS385	SS-M_0639	p _{myc1} <i>tetO</i> promoter with a deletion of nt −53 through −1 + p _{myc1} 5' UTR + first 54 nt of <i>sigA</i> + <i>yfp-6</i> ×His

further analyze it in this study. However, we cannot exclude the possibility that it is real and physiologically relevant.

MATERIALS AND METHODS

Strains and culture conditions. All experiments were done using a *Mycobacterium smegmatis* *ΔMSMEG_2952* strain (71), which is less prone to aggregation (clumping) than its parent strain mc²155 and therefore permits higher confidence measurements by flow cytometry. This strain and its derivatives (Table 1) were grown in Middlebrook 7H9 medium with albumin-dextrose-catalase (ADC) supplementation (final concentrations, 5 g/liter bovine serum albumin fraction V, 2 g/liter dextrose, 0.85 g/liter NaCl, and 3 mg/liter 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 (OD₆₀₀) of ~0.8 at the time of harvest.

Plasmid construction. Plasmid pSS303 was built on a backbone derived from pGH1000A (72) by inserting a *yfp* cassette containing the gene sequence of a YFP reporter (sfYFP, obtained from Ivy Fitzgerald and Benjamin Glick) with a 6×His tag at the C terminus (complete amino acid sequence, MASDSTESLFTGVVPILVELDGDGVNGHKFSVRGEGEGDATNGKLTCLKICTTGKLPVPWPTLVTLGYGVQCFARYPDH MKQHDFFKSAPEGVVQERTITFKDDGTYKTRAEVKEFGDTLVNRIELKGIDFKEDGNILGHKLEYNFNHNVYITADK QKNGIKANFKIRHNVEDGGVQLADHYQNTPIGDGPVLLPDNHYLSYQSKLSKDPNEKRDMVLLEFVTAAGITHGS SGSSGCHHHHHH). Two synthetic transcriptional terminators were inserted flanking the cassette as follows: *tsyA* (73) upstream and *ttsB* (74) downstream. Transcription was initiated by the p_{myc1} *tetO* promoter, which was constitutively active in our strains due to the absence of the corresponding *tet* repressor (57). All constructs (pSS303 and derivatives noted in Table 1) were built using NEBuilder HiFi DNA assembly master mix (catalog number E2621). Each assembled plasmid was integrated in *M. smegmatis* *ΔMSMEG_2952* (71) at the Giles phage site and selected with 200 μg/ml hygromycin.

Cell fixation and flow cytometry. Several 1.5-ml aliquots of *M. smegmatis* cultures were pelleted, resuspended in 500 μl 2% paraformaldehyde in phosphate-buffered saline (PBS), and incubated at room temperature for 30 min. Cells were rinsed twice using 900 μl PBS + 0.1% Tween 20 and resuspended to a calculated OD₆₀₀ of 15. Prior to flow cytometry analysis, cells were filtered using an 18-gauge 5-μm filter needle and diluted with Middlebrook 7H9 to an OD₆₀₀ of 0.015. YFP fluorescence intensity was measured per manufacturer's instructions using a BD Accuri C6 flow cytometer collecting 100,000 events per sample (Fig. 1B and C) or a BD LSR II flow cytometer collecting 50,000 events per sample (Fig. 2B, 3B, and 4B) using appropriate controls and thresholds. FlowJo v10.6 was used to draw tight forward scatter and side scatter gates to limit analysis to similarly sized cells, and GraphPad Prism 8 was used for statistical analysis.

RNA extraction and determination of mRNA abundance and stability. RNA extraction, measurement of mRNA abundance, and mRNA stability analyses from *M. smegmatis* cultures were conducted in biological triplicates as described in reference 15. Briefly, mRNA abundance was measured by quantitative PCR (qPCR) using iTaq SYBR green (Bio-Rad) on an Applied Biosystems 7500 with 400 pg of cDNA and 0.25 μM each primer in 10-μl reaction mixtures. Cycle parameters were 95°C for 15 s and 61°C for 60 s. Primers used to determine mRNA abundance are listed in Table 2.

For mRNA stability analysis, 5-ml *M. smegmatis* cultures were treated with rifampin at a final concentration of 150 μg/ml to halt transcription and snap-frozen in liquid nitrogen after 0, 1, 2, or 4 min. Abundance over time was determined for *sigA* and *yfp* using qPCR and used to estimate mRNA half-lives essentially as in reference 15. For each sample, the negative of the threshold cycle (C_T) represents

TABLE 2 Primers for qPCR

Primer name (reference)	Gene	Directionality	Sequence 5' → 3'
JR273 (75)	<i>sigA</i> (<i>MSMEG_2758</i>)	Forward	GACTACACCAAGGGCTACAAG
JR274 (75)	<i>sigA</i> (<i>MSMEG_2758</i>)	Reverse	TTGATCACCTCGACCATGTG
SSS833	<i>yfp</i>	Forward	GATAGCACTGAGAGCCTGTT
SSS834	<i>yfp</i>	Reverse	CTGAACCTGTGGCCGTTTAC

transcript abundance on a $\log_2$ scale. For each strain and gene, linear regression was performed on a plot of $-C_T$ versus time. Half-life was defined as $-1/\text{slope}$. *sigA* half-lives were equivalent in all strains and not shown. As we have observed for many other genes in mycobacteria, plotting $\log_2$ abundance over time produced a biphasic decay curve consistent with a period of faster exponential decay, followed by a period of much slower exponential decay (see Fig. S3A in the supplemental material). Similar biphasic decay curves have been reported by others for some *E. coli* genes (76–79). The initial rapid decay phase reflects the rate of decay for at least 90% of the *yfp* RNA present in our samples. We therefore used only this initial phase for mRNA half-life calculations (0, 1, and 2 min for strain SS-M_0489 and 0 and 1 min for strains SS-M_0493 and SS-M_0626) (see Fig. S3B). The slower decay phase could reflect the presence of minority transcript species that inherently decay more slowly or could reflect perturbation of cellular physiology due to rifampin.

Calculation of transcript production rates. The rate of transcript production was estimated as described in reference 59. Briefly, transcript production rate (V_t) is described as follows:

$$V_t = k \cdot [\text{mRNA}] + \mu \cdot [\text{mRNA}]$$

Where [mRNA] is a given transcript's concentration, μ is the growth rate of the cells ($\ln_2/\text{doubling time}$), and k is the degradation rate constant ($\ln_2/\text{half-life}$). Note that because [mRNA] is derived from our qPCR data and is therefore a relative value rather than absolute value, the calculated transcript production rate is also a relative rather than absolute value.

Protein extraction and BCA assay. *M. smegmatis* cells were pelleted; rinsed three times with Middlebrook 7H9, 0.2% glycerol, and 0.05% Tween 80 at 4°C; resuspended in PBS + 2% SDS + protease inhibitor cocktail (VWR; catalog number 97063-972); and transferred to 2-ml disruption tubes (OPS Diagnostics; 100- μm zirconium lysing matrix, molecular grade). Cultures were lysed using a FastPrep-24 5G instrument (MP Biomedical) using four cycles of 6.5 m/s for 30 s with 1 min on ice between cycles. Samples were clarified by centrifugation at $21,130 \times g$ at 4°C for 10 min, and the supernatant containing protein was recovered and stored at -20°C . Protein concentrations were calculated using the Pierce BCA protein assay (Thermo Scientific; catalog number 23225) according to the manufacturer's instructions.

Western blotting. Protein was normalized to the indicated masses in a final volume of 9 μl combined with 4 μl of 4 $\times$ protein loading dye (200 mM Tris-HCl [pH 6.8], 400 mM dithiothreitol [DTT], 8% SDS, 0.4% bromophenol blue, 40% glycerol) and heated to 95°C for 10 min. Using gradient gels (4 to 15% Mini-Protein TGX precast protein gels; Bio-Rad; catalog number 4561086), the samples were electrophoresed for 60 min at 140 V and then transferred to a polyvinylidene difluoride (PVDF) membrane. The membrane was incubated in blocking solution (PBS plus 5% nonfat milk) for 30 min and washed once for 5 min using washing buffer (PBS 1 $\times$ buffer plus 0.1% Tween 20). The membrane was probed with 1 $\mu\text{g}/\text{ml}$ His tag antibody (polyclonal antibody, rabbit; GenScript; catalog number A00174) in blocking solution for 60 min at room temperature. The membrane was then rinsed twice with wash buffer and once with 1 $\times$ PBS and incubated with anti-rabbit IgG-peroxidase (Sigma-Aldrich; catalog number A4914), 1:30,000 in blocking solution, for 60 min at room temperature. The membrane was rinsed as previously described and incubated with horseradish peroxidase (HRP) substrate (Radiance Q; Azure Biosystems; catalog number AC2101) as recommended by the manufacturer. Imaging was done using an Azure C200 imaging system (Azure Biosystems).

Software. GraphPad Prism was used for all linear regressions and comparisons (GraphPad Software, La Jolla, CA). The Srna program within Sfold was used for RNA secondary structure predictions (60, 61).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, XLSX file, 0.4 MB.

SUPPLEMENTAL FILE 2, PDF file, 1.4 MB.

ACKNOWLEDGMENTS

This work was supported by NSF CAREER award 1652756 to S.S.S.

Thanks go to Ivy Fitzgerald and Benjamin Glick for providing the sfYFP construct. We thank all members of the Shell lab for technical assistance and helpful discussions.

T.G.N., L.A.R., and S.S.S. conceived and designed the experiments. T.G.N. and D.A.V.-B. performed the experiments. T.G.N., D.A.V.-B., and S.S.S. analyzed the data. T.G.N., D.A.V.-B., and S.S.S. wrote the manuscript.

REFERENCES

- Shine J, Dalgarno L. 1974. The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci U S A* 71:1342–1346. <https://doi.org/10.1073/pnas.71.4.1342>.
- Ringquist S, Shinedling S, Barrick D, Green L, Binkley J, Stormo GD, Gold L. 1992. Translation initiation in Escherichia coli: sequences within the ribosome-binding site. *Mol Microbiol* 6:1219–1229. <https://doi.org/10.1111/j.1365-2958.1992.tb01561.x>.
- Chen H, Bjerknes M, Kumar R, Jay E. 1994. Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs. *Nucleic Acids Res* 22:4953–4957. <https://doi.org/10.1093/nar/22.23.4953>.
- Sterk M, Romilly C, Wagner E. 2018. Unstructured 5'-tails act through ribosome standby to override inhibitory structure at ribosome binding sites. *Nucleic Acids Res* 46:4188–4199. <https://doi.org/10.1093/nar/gky073>.

5. Habib NF, Jackson MP. 1993. Roles of a ribosome-binding site and mRNA secondary structure in differential expression of Shiga toxin genes. *J Bacteriol* 175:597–603. <https://doi.org/10.1128/jb.175.3.597-603.1993>.
6. Lease RA, Belfort M. 2000. A trans-acting RNA as a control switch in *Escherichia coli*: DsrA modulates function by forming alternative structures. *Proc Natl Acad Sci U S A* 97:9919–9924. <https://doi.org/10.1073/pnas.170281497>.
7. Mutalik VK, Qi L, Guimaraes JC, Lucks JB, Arkin AP. 2012. Rationally designed families of orthogonal RNA regulators of translation. *Nat Chem Biol* 8:447–454. <https://doi.org/10.1038/nchembio.919>.
8. Takahashi MK, Lucks JB. 2013. A modular strategy for engineering orthogonal chimeric RNA transcription regulators. *Nucleic Acids Res* 41:7577–7588. <https://doi.org/10.1093/nar/gkt452>.
9. Jagodnik J, Chiaruttini C, Guillier M. 2017. Stem-loop structures within mRNA coding sequences activate translation initiation and mediate control by small regulatory RNAs. *Mol Cell* 68:158–170. <https://doi.org/10.1016/j.molcel.2017.08.015>.
10. Fan H, Conn AB, Williams PB, Diggs S, Hahm J, Gamper HB, Hou Y-M, O'Leary SE, Wang Y, Blaha GM. 2017. Transcription-translation coupling: direct interactions of RNA polymerase with ribosomes and ribosomal subunits. *Nucleic Acids Res* 45:11043–11055. <https://doi.org/10.1093/nar/gkx719>.
11. Zhang Y, Mooney RA, Grass JA, Sivaramakrishnan P, Herman C, Landick R, Wang JD. 2014. DksA guards elongating RNA polymerase against ribosome-stalling-induced arrest. *Mol Cell* 53:766–778. <https://doi.org/10.1016/j.molcel.2014.02.005>.
12. Miller OL, Jr, Hamkalo BA, Thomas CA, Jr. 1970. Visualization of bacterial genes in action. *Science* 169:392–395. <https://doi.org/10.1126/science.169.3943.392>.
13. Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, Gottesman ME, Rosch P. 2010. A NusE:NusG complex links transcription and translation. *Science* 328:501–504. <https://doi.org/10.1126/science.1184953>.
14. Proshkin S, Rahmouni AR, Mironov A, Nudler E. 2010. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* 328:504–508. <https://doi.org/10.1126/science.1184939>.
15. Vargas-Blanco DA, Zhou Y, Zamalloa LG, Antonelli T, Shell SS. 2019. mRNA degradation rates are coupled to metabolic status in *Mycobacterium smegmatis*. *mBio* 10:e00957-19. <https://doi.org/10.1128/mBio.00957-19>.
16. Braun F, Le Derout J, Régnier P. 1998. Ribosomes inhibit an RNase E cleavage which induces the decay of the rpsO mRNA of *Escherichia coli*. *EMBO J* 17:4790–4797. <https://doi.org/10.1093/emboj/17.16.4790>.
17. Pato ML, Bennett PM, von Meyenburg K. 1973. Messenger ribonucleic acid synthesis and degradation in *Escherichia coli* during inhibition of translation. *J Bacteriol* 116:710–718. <https://doi.org/10.1128/JB.116.2.710-718.1973>.
18. Sharp JS, Bechhofer DH. 2003. Effect of translational signals on mRNA decay in *Bacillus subtilis*. *J Bacteriol* 185:5372–5379. <https://doi.org/10.1128/jb.185.18.5372-5379.2003>.
19. Hambræus G, Karhamaa K, Rutberg B. 2002. A 5' stem-loop and ribosome binding but not translation are important for the stability of *Bacillus subtilis* aprE leader mRNA. *Microbiology* 148:1795–1803. <https://doi.org/10.1099/00221287-148-6-1795>.
20. Jürgen B, Schweder T, Hecker M. 1998. The stability of mRNA from the gsjB gene of *Bacillus subtilis* is dependent on the presence of a strong ribosome binding site. *Mol Gen Genet* 258:538–545. <https://doi.org/10.1007/s004380050765>.
21. Arnold TE, Yu J, Belasco JG. 1998. mRNA stabilization by the ompA 5' untranslated region: two protective elements hinder distinct pathways for mRNA degradation. *RNA* 4:319–330.
22. Wagner LA, Gesteland RF, Dayhuff TJ, Weiss RB. 1994. An efficient Shine-Dalgarno sequence but not translation is necessary for lacZ mRNA stability in *Escherichia coli*. *J Bacteriol* 176:1683–1688. <https://doi.org/10.1128/jb.176.6.1683-1688.1994>.
23. Moll I, Afonyushkin T, Vytvytska O, Kabardin VR, Blasi U. 2003. Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. *RNA* 9:1308–1314. <https://doi.org/10.1261/ma.5850703>.
24. Skorski P, Leroy P, Fayet O, Dreyfus M, Hermann-Le Denmat S. 2006. The highly efficient translation initiation region from the *Escherichia coli* rpsA gene lacks a Shine-Dalgarno element. *J Bacteriol* 188:6277–6285. <https://doi.org/10.1128/JB.00591-06>.
25. Link TM, Valentin-Hansen P, Brennan RG. 2009. Structure of *Escherichia coli* Hfq bound to polyribadenylate RNA. *Proc Natl Acad Sci U S A* 106:19292–19297. <https://doi.org/10.1073/pnas.0908744106>.
26. Pedersen M, Nissen S, Mitarai N, Lo Svenningsen S, Snekken K, Pedersen S. 2011. The functional half-life of an mRNA depends on the ribosome spacing in an early coding region. *J Mol Biol* 407:35–44. <https://doi.org/10.1016/j.jmb.2011.01.025>.
27. Anderson KL, Dunman PM. 2009. Messenger RNA turnover processes in *Escherichia coli*, *Bacillus subtilis*, and emerging studies in *Staphylococcus aureus*. *Int J Microbiol* 2009:525491. <https://doi.org/10.1155/2009/525491>.
28. Unniraman S, Prakash R, Nagaraja V. 2001. Alternate paradigm for intrinsic transcription termination in eubacteria. *J Biol Chem* 276:41850–41855. <https://doi.org/10.1074/jbc.M106252200>.
29. Belasco JG, Nilsson G, von Gabain A, Cohen SN. 1986. The stability of *E. coli* gene transcripts is dependent on determinants localized to specific mRNA segments. *Cell* 46:245–251. [https://doi.org/10.1016/0092-8674\(86\)90741-5](https://doi.org/10.1016/0092-8674(86)90741-5).
30. Emory SA, Belasco JG. 1990. The ompA 5' untranslated RNA segment functions in *Escherichia coli* as a growth-rate-regulated mRNA stabilizer whose activity is unrelated to translational efficiency. *J Bacteriol* 172:4472–4481. <https://doi.org/10.1128/jb.172.8.4472-4481.1990>.
31. Chen LH, Emory SA, Bricker AL, Bouvet P, Belasco JG. 1991. Structure and function of a bacterial mRNA stabilizer: analysis of the 5' untranslated region of ompA mRNA. *J Bacteriol* 173:4578–4586. <https://doi.org/10.1128/jb.173.15.4578-4586.1991>.
32. Emory SA, Bouvet P, Belasco JG. 1992. A 5'-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*. *Genes Dev* 6:135–148. <https://doi.org/10.1101/gad.6.1.135>.
33. Bechhofer DH, Dubnau D. 1987. Induced mRNA stability in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* 84:498–502. <https://doi.org/10.1073/pnas.84.2.498>.
34. Hambræus G, Persson M, Rutberg B. 2000. The aprE leader is a determinant of extreme mRNA stability in *Bacillus subtilis*. *Microbiology* 146:3051–3059. <https://doi.org/10.1099/00221287-146-12-3051>.
35. Heck C, Rothfuchs R, Jäger A, Rauhut R, Klug G. 1996. Effect of the pufQ-pufB intercistronic region on puf mRNA stability in *Rhodobacter capsulatus*. *Mol Microbiol* 20:1165–1178. <https://doi.org/10.1111/j.1365-2958.1996.tb02637.x>.
36. Chen CY, Belasco JG. 1990. Degradation of pufLMX mRNA in *Rhodobacter capsulatus* is initiated by nonrandom endonucleolytic cleavage. *J Bacteriol* 172:4578–4586. <https://doi.org/10.1128/jb.172.8.4578-4586.1990>.
37. Heck C, Evguenieva-Hackenberg E, Balzer A, Klug G. 1999. RNase E enzymes from *Rhodobacter capsulatus* and *Escherichia coli* differ in context- and sequence-dependent in vivo cleavage within the polycistronic puf mRNA. *J Bacteriol* 181:7621–7625. <https://doi.org/10.1128/JB.181.24.7621-7625.1999>.
38. Richards J, Belasco JG. 2019. Obstacles to scanning by RNase E govern bacterial mRNA lifetimes by hindering access to distal cleavage sites. *Mol Cell* 74:284–295. <https://doi.org/10.1016/j.molcel.2019.01.044>.
39. Agaisse H, Lereclus D. 1996. STAB-SD: a Shine-Dalgarno sequence in the 5' untranslated region is a determinant of mRNA stability. *Mol Microbiol* 20:633–643. <https://doi.org/10.1046/j.1365-2958.1996.5401046.x>.
40. Romeo T, Babitzke P. 2018. Global regulation by CsrA and its RNA antagonists. *Microbiol Spectr* 6:RWR-0009-2017. <https://doi.org/10.1128/microbiolspec.RWR-0009-2017>.
41. Martini MC, Zhou Y, Sun H, Shell SS. 2019. Defining the transcriptional and post-transcriptional landscapes of *Mycobacterium smegmatis* in aerobic growth and hypoxia. *Front Microbiol* 10:591. <https://doi.org/10.3389/fmicb.2019.00591>.
42. Shell SS, Wang J, Lapierre P, Mir M, Chase MR, Pyle MM, Gawande R, Ahmad R, Sarracino DA, Ioerger TR, Fortune SM, Derbyshire KM, Wade JT, Gray TA. 2015. Leaderless transcripts and small proteins are common features of the mycobacterial translational landscape. *PLoS Genet* 11:e1005641. <https://doi.org/10.1371/journal.pgen.1005641>.
43. Cortes T, Schubert OT, Rose G, Arnvig KB, Comas I, Aebersold R, Young DB. 2013. Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. *Cell Rep* 5:1121–1131. <https://doi.org/10.1016/j.celrep.2013.10.031>.
44. Moll I, Hirokawa G, Kiel MC, Kaji A, Blasi U. 2004. Translation initiation with 70S ribosomes: an alternative pathway for leaderless mRNAs. *Nucleic Acids Res* 32:3354–3363. <https://doi.org/10.1093/nar/gkh663>.
45. O'Donnell SM, Janssen GR. 2002. Leaderless mRNAs bind 70S ribosomes more strongly than 30S ribosomal subunits in *Escherichia coli*. *J Bacteriol* 184:6730–6733. <https://doi.org/10.1128/jb.184.23.6730-6733.2002>.
46. Udagawa T, Shimizu Y, Ueda T. 2004. Evidence for the translation initiation of leaderless mRNAs by the intact 70 S ribosome without its dissociation into subunits in eubacteria. *J Biol Chem* 279:8539–8546. <https://doi.org/10.1074/jbc.M308784200>.

47. Grill S, Gualerzi CO, Londei P, Blasi U. 2000. Selective stimulation of translation of leaderless mRNA by initiation factor 2: evolutionary implications for translation. *EMBO J* 19:4101–4110. <https://doi.org/10.1093/emboj/19.15.4101>.
48. Grill S, Moll I, Giuliodori AM, Gualerzi CO, Blasi U. 2002. Temperature-dependent translation of leaderless and canonical mRNAs in *Escherichia coli*. *FEMS Microbiol Lett* 211:161–167. <https://doi.org/10.1111/j.1574-6968.2002.tb11219.x>.
49. Moll I, Grill S, Grundling A, Blasi U. 2002. Effects of ribosomal proteins S1, S2 and the DeaD/CsdA DEAD-box helicase on translation of leaderless and canonical mRNAs in *Escherichia coli*. *Mol Microbiol* 44:1387–1396. <https://doi.org/10.1046/j.1365-2958.2002.02971.x>.
50. Tedin K, Moll I, Grill S, Resch A, Graschopf A, Gualerzi CO, Blasi U. 1999. Translation initiation factor 3 antagonizes authentic start codon selection on leaderless mRNAs. *Mol Microbiol* 31:67–77. <https://doi.org/10.1046/j.1365-2958.1999.01147.x>.
51. O'Donnell SM, Janssen GR. 2001. The initiation codon affects ribosome binding and translational efficiency in *Escherichia coli* of cl mRNA with or without the 5' untranslated leader. *J Bacteriol* 183:1277–1283. <https://doi.org/10.1128/JB.183.4.1277-1283.2001>.
52. Van Etten WJ, Janssen GR. 1998. An AUG initiation codon, not codon-anticodon complementarity, is required for the translation of unleadered mRNA in *Escherichia coli*. *Mol Microbiol* 27:987–1001. <https://doi.org/10.1046/j.1365-2958.1998.00744.x>.
53. Beck HJ, Fleming IM, Janssen GR. 2016. 5'-Terminal AUGs in *Escherichia coli* mRNAs with Shine-Dalgarno sequences: identification and analysis of their roles in non-canonical translation initiation. *PLoS One* 11: e0160144. <https://doi.org/10.1371/journal.pone.0160144>.
54. Baumeister R, Flache P, Melefors O, von Gabain A, Hillen W. 1991. Lack of a 5' non-coding region in Tn1721 encoded tetR mRNA is associated with a low efficiency of translation and a short half-life in *Escherichia coli*. *Nucleic Acids Res* 19:4595–4600. <https://doi.org/10.1093/nar/19.17.4595>.
55. Gomez M, Doukhan L, Nair G, Smith I. 1998. sigA is an essential gene in *Mycobacterium smegmatis*. *Mol Microbiol* 29:617–628. <https://doi.org/10.1046/j.1365-2958.1998.00960.x>.
56. Manganelli R, Provvedi R, Rodrigue S, Beaucher J, Gaudreau L, Smith I. 2004. Sigma factors and global gene regulation in *Mycobacterium tuberculosis*. *J Bacteriol* 186:895–902. <https://doi.org/10.1128/jb.186.4.895-902.2004>.
57. Ehrt S, Guo XV, Hickey CM, Ryou M, Monteleone M, Riley LW, Schnappinger D. 2005. Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Res* 33:e21. <https://doi.org/10.1093/nar/gni013>.
58. Nogueira T, de Smit M, Graffe M, Springer M. 2001. The relationship between translational control and mRNA degradation for the *Escherichia coli* threonyl-tRNA synthetase gene. *J Mol Biol* 310:709–722. <https://doi.org/10.1006/jmbi.2001.4796>.
59. Esquerré T, Laguerre S, Turlan C, Carpusis AJ, Girbal L, Coccagn-Bousquet M. 2014. Dual role of transcription and transcript stability in the regulation of gene expression in *Escherichia coli* cells cultured on glucose at different growth rates. *Nucleic Acids Res* 42:2460–2472. <https://doi.org/10.1093/nar/gkt1150>.
60. Ding Y, Chan CY, Lawrence CE. 2004. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res* 32: W135–W141. <https://doi.org/10.1093/nar/gkh449>.
61. Ding Y, Chan CY, Lawrence CE. 2005. RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA* 11:1157–1166. <https://doi.org/10.1261/rna.2500605>.
62. Schubert OT, Ludwig C, Kogadeeva M, Zimmermann M, Rosenberger G, Gengenbacher M, Gillet LC, Collins BC, Rost HL, Kaufmann SH, Sauer U, Aebersold R. 2015. Absolute proteome composition and dynamics during dormancy and resuscitation of *Mycobacterium tuberculosis*. *Cell Host Microbe* 18:96–108. <https://doi.org/10.1016/j.chom.2015.06.001>.
63. Rustad TR, Minch KJ, Brabant W, Winkler JK, Reiss DJ, Baliga NS, Sherman DR. 2013. Global analysis of mRNA stability in *Mycobacterium tuberculosis*. *Nucleic Acids Res* 41:509–517. <https://doi.org/10.1093/nar/gks1019>.
64. Berg L, Lale R, Bakke I, Burroughs N, Valla S. 2009. The expression of recombinant genes in *Escherichia coli* can be strongly stimulated at the transcript production level by mutating the DNA-region corresponding to the 5'-untranslated part of mRNA. *Microb Biotechnol* 2:379–389. <https://doi.org/10.1111/j.1751-7915.2009.00107.x>.
65. Davis CA, Bingman CA, Landick R, Record MT, Jr, Saecker RM. 2007. Real-time footprinting of DNA in the first kinetically significant intermediate in open complex formation by *Escherichia coli* RNA polymerase. *Proc Natl Acad Sci U S A* 104:7833–7838. <https://doi.org/10.1073/pnas.0609888104>.
66. Lale R, Berg L, Stuttgart F, Netzer R, Stafsnes M, Brautaset T, Vee Aune TE, Valla S. 2011. Continuous control of the flow in biochemical pathways through 5' untranslated region sequence modifications in mRNA expressed from the broad-host-range promoter Pm. *Appl Environ Microbiol* 77:2648–2655. <https://doi.org/10.1128/AEM.02091-10>.
67. Davenport RJ, Wuite GJ, Landick R, Bustamante C. 2000. Single-molecule study of transcriptional pausing and arrest by *E. coli* RNA polymerase. *Science* 287:2497–2500. <https://doi.org/10.1126/science.287.5462.2497>.
68. Epshtein V, Toulme F, Rahmouni AR, Borukhov S, Nudler E. 2003. Transcription through the roadblocks: the role of RNA polymerase cooperation. *EMBO J* 22:4719–4727. <https://doi.org/10.1093/emboj/cdg452>.
69. Bhattacharyya S, Jacobs WM, Adkar BV, Yan J, Zhang W, Shakhnovich EI. 2018. Accessibility of the Shine-Dalgarno sequence dictates N-terminal codon bias in *E. coli*. *Mol Cell* 70:894–905. <https://doi.org/10.1016/j.molcel.2018.05.008>.
70. Boel G, Letso R, Neely H, Price WN, Wong KH, Su M, Luff J, Valecha M, Everett JK, Acton TB, Xiao R, Montelione GT, Aalberts DP, Hunt JF. 2016. Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature* 529:358–363. <https://doi.org/10.1038/nature16509>.
71. Yang Y, Thomas J, Li Y, Vilcheze C, Derbyshire KM, Jacobs WR, Jr, Ojha AK. 2017. Defining a temporal order of genetic requirements for development of mycobacterial biofilms. *Mol Microbiol* 105:794–809. <https://doi.org/10.1111/mmi.13734>.
72. Morris P, Marinelli LJ, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2008. Genomic characterization of mycobacteriophage Giles: evidence for phage acquisition of host DNA by illegitimate recombination. *J Bacteriol* 190:2172–2182. <https://doi.org/10.1128/JB.01657-07>.
73. Czyz A, Mooney RA, Iaconi A, Landick R. 2014. Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *mBio* 5:e00931-14. <https://doi.org/10.1128/mBio.00931-14>.
74. Huff J, Czyz A, Landick R, Niederweis M. 2010. Taking phage integration to the next level as a genetic tool for mycobacteria. *Gene* 468:8–19. <https://doi.org/10.1016/j.gene.2010.07.012>.
75. Rock JM, Hopkins FF, Chavez A, Diallo M, Chase MR, Gerrick ER, Pritchard JR, Church GM, Rubin EJ, Sassetti CM, Schnappinger D, Fortune SM. 2017. Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nat Microbiol* 2:16274. <https://doi.org/10.1038/nmicrobiol.2016.274>.
76. Chen H, Shioguchi K, Ge H, Xie XS. 2015. Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Mol Syst Biol* 11:781. <https://doi.org/10.15252/msb.20145794>.
77. Hambraeus G, von Wachenfeldt C, Hederstedt L. 2003. Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. *Mol Genet Genomics* 269:706–714. <https://doi.org/10.1007/s00438-003-0883-6>.
78. Brescia CC, Kaw MK, Sledjeski DD. 2004. The DNA binding protein H-NS binds to and alters the stability of RNA in vitro and in vivo. *J Mol Biol* 339:505–514. <https://doi.org/10.1016/j.jmb.2004.03.067>.
79. Sinha D, Matz LM, Cameron TA, De Lay NR. 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>.