Differential translation of mRNA isoforms transcribed with distinct sigma factors

DYLAN M. MCCORMICK,1,5 JEAN-BENOÎT LALANNE,1,2,4,5 TAMMY C.T. LAN,3 SILVI ROUSKIN,3 and GENE-WEI LI1

1Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
2Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
3Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA

ABSTRACT

Sigma factors are an important class of bacterial transcription factors that lend specificity to RNA polymerases by binding to distinct promoter elements for genes in their regulons. Here we show that activation of the general stress sigma factor, σB, in Bacillus subtilis paradoxically leads to dramatic induction of translation for a subset of its regulon genes. These genes are translationally repressed when transcribed by the housekeeping sigma factor, σA, owing to extended RNA secondary structures as determined in vivo using DMS-MaPseq. Transcription from σB-dependent promoters excludes the secondary structures and activates translation, leading to dual induction. Translation efficiencies between σB- and σA-dependent RNA isoforms can vary by up to 100-fold, which in multiple cases exceeds the magnitude of transcriptional induction. These results highlight the role of long-range RNA folding in modulating translation and demonstrate that a transcription factor can regulate protein synthesis beyond its effects on transcript levels.

Keywords: sigma factor; RNA structure; translation efficiency; B. subtilis; dual induction

INTRODUCTION

Transcriptional regulation by sigma factors is a hallmark of bacterial gene expression. Sigma factors bind to the core RNA polymerases, forming holoenzymes that can initiate transcription at sites with well-defined sequences. In Bacillus subtilis, most genes are transcribed by the housekeeping sigma factor σA, and some are additionally or exclusively transcribed by alternative sigma factors that control specific processes such as sporulation and motility (Haldenwang 1995; Helmann 2019). The alternative sigma factor σB is involved in the general stress response (Haldenwang and Losick 1979; Haldenwang 1995; Hecker et al. 2007; Price 2014) and initiates transcription for over two hundred genes with well-defined promoter sequences (Petersohn et al. 1999; Nicolas et al. 2012; Zhu and Stülke 2018). Induction of transcription leads to corresponding increases in RNA levels (Fig. 1A).

Translational regulation is also widespread in B. subtilis, although it is not typically thought to be controlled by transcription factors. Differential translation among genes in the same operon is largely driven by differences in mRNA secondary structure (Burkhardt et al. 2017) and is important for stoichiometric production of proteins in the same complex or metabolic pathway (Li et al. 2014; Lalanne et al. 2018). Translation can be additionally regulated by RNA-binding proteins or riboswitches that modulate the accessibility of the ribosome binding sites on the mRNA (Yakhnin et al. 2004, 2007; Breaker 2018).

Operons are often controlled both transcriptionally and translationally (Fig. 1A), but seldomly by the same regulator (Hollands et al. 2012; Chauvier et al. 2017; Bastet et al. 2018).

Here we show that the transcription factor σB not only activates transcription, but also derepresses translation for a subset of its regulon genes. Using Rend-seq (end-enriched RNA-seq) (Lalanne et al. 2018) and ribosome profiling, we identified 12 genes whose apparent translation efficiency is increased substantially during σB activation. Most of them are transcribed from a σB-dependent promoter as well as at least one σA-dependent promoter, generating multiple transcript isoforms. By modulating σB activities, we found that each transcript isoform is associated with a

© 2021 McCormick et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see http://majournal.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
distinct translation efficiency, with strongly repressed translation for $\sigma^A$-driven isoforms and elevated translation for $\sigma^B$-driven isoforms. These were orthogonally confirmed using a fluorescent reporter in a subset of examples. Both computational RNA folding and in vivo structural probing by DMS-MaPseq (Zubradt et al. 2016) indicate that the repressed $\sigma^A$-driven isoforms possess extended RNA secondary structures that sequester the ribosome binding sites. On the other hand, $\sigma^B$-driven isoforms have shorter 5' UTRs that only include the regions corresponding to the second halves of the extended stem–loops in the longer $\sigma^A$-driven isoforms. Therefore, $\sigma^B$ can simultaneously activate both transcription and translation by modulating isoform-specific secondary structures.

RESULTS

$\sigma^B$ activates translation for a subset of its regulon

We first observed translational activation of $\sigma^B$ regulon genes while profiling gene expression for a $B. subtilis$ strain with an elevated general stress response during steady-state growth due to a genetic modification (Materials and Methods). Rend-seq and ribosome profiling data were generated to quantify the mRNA levels and protein synthesis rates, respectively, for both the wild type ("$\sigma^B$ inactive") and the genetically modified strain ("$\sigma^B$ active"). The density of ribosome footprints for a gene provides an estimate for the relative rate of protein synthesis, provided that most ribosomes complete translation to yield full-length polypeptides and that the elongation time averaged across the entire transcript is constant (Ingolia et al. 2009; Li 2015; Li et al. 2014; Lalanne et al. 2018). Translation efficiency (TE), defined as the rate of protein production per mRNA molecule, can then be estimated from Rend-seq and ribosome profiling data by calculating the per-gene ribosome profiling coverage over Rend-seq coverage, that is, the ribosome density along a transcript (Li 2015; Li et al. 2014). Given $\sigma^B$'s well-understood role in transcription initiation, we expected its regulon members to change in mRNA levels and not TE.

Surprisingly, we found that several genes in the $\sigma^B$ regulon showed far greater increases in protein synthesis rate (ribosome profiling) than in mRNA levels (Rend-seq).
Between the two conditions, 25% of the annotated \( \sigma^B \) regulon genes (Zhu and Stülke 2018) had substantially different expression levels (56/225 with >3.7-fold change, Fig. 1B,C). Although most genes showed concordant changes in mRNA levels and protein synthesis rates, a notable population (21%, 12/56) exhibited a considerably greater increase in protein synthesis rates than mRNA levels (>2.7-fold), suggesting an increase in apparent translation efficiency (Fig. 1D). Among these translationally activated \( \sigma^B \) regulon genes, the magnitude of TE increases often exceeded the rise in mRNA levels, as most genes (75%, 9/12) exhibited a fold change in apparent TE accounting for >50% of the observed fold change in protein synthesis rate (Fig. 1E, compared to purely transcriptionally activated genes in Fig. 1F, Materials and Methods). Hence, translational induction contributes to the majority of the increase in expression of a subset of the \( \sigma^B \) regulon, suggesting a yet-unknown strategy for activating translation following \( \sigma^B \) induction.

\( \sigma^B \)-dependent alternative mRNA isoforms drive translational up-regulation

To identify the regulatory features that could drive translational up-regulation, we examined the transcript architecture of translationally activated \( \sigma^B \) regulon genes using Rend-seq. Through sparse fragmentation of input RNAs, Rend-seq enriches for the 5' and 3'-mapped read counts, respectively, and the black scale bars correspond to 0.5 kb. Fold changes (FC) for Rend-seq and ribosome profiling between \( \sigma^B \) active and \( \sigma^B \) inactive conditions are shown. Rend-seq 5' ends corresponding to the \( \sigma^A \)-dependent transcription start sites are marked by red arrows. Putative \( \sigma^B \)-dependent promoter sequences are listed for each gene (+1 corresponds to the 5' end of the \( \sigma^B \)-dependent isoform mapped by Rend-seq). The consensus sequences for the −10 and −35 regions of \( \sigma^B \)-dependent promoters are GTTTaa and GGG(A/T)A(A/T) (Petersohn et al. 1999). For ctc specifically, the additional 5'/3' peak pair ('*') in the \( \sigma^B \) active condition corresponds to a spurious RNase A cleavage site that likely occurred post-lysis. See also Supplemental Figures S1, S2.

Between the two conditions, 25% of the annotated \( \sigma^B \) regulon genes (Zhu and Stülke 2018) had substantially different expression levels (56/225 with >3.7-fold change, Fig. 1B,C). Although most genes showed concordant changes in mRNA levels and protein synthesis rates, a notable population (21%, 12/56) exhibited a considerably greater increase in protein synthesis rates than mRNA levels (>2.7-fold), suggesting an increase in apparent translation efficiency (Fig. 1D). Among these translationally activated \( \sigma^B \) regulon genes, the magnitude of TE increases often exceeded the rise in mRNA levels, as most genes (75%, 9/12) exhibited a fold change in apparent TE accounting for >50% of the observed fold change in protein synthesis rate (Fig. 1E, compared to purely transcriptionally activated genes in Fig. 1F, Materials and Methods). Hence, translational induction contributes to the majority of the increase in expression of a subset of the \( \sigma^B \) regulon, suggesting a yet-unknown strategy for activating translation following \( \sigma^B \) induction.
long, σ^A-dependent isoforms. By estimating the relative prevalence of short and long isoforms across Rend-seq and ribosome profiling data sets with different levels of σ^B induction, we could infer the individual translation efficiency for each isoform (Fig. 3A; Supplemental Fig. S3, Materials and Methods), hereafter referred to as the isoform-specific translation efficiency. Compared to the σ^A-dependent isoforms, we found that the TE for the σ^B isoform was three- to 100-fold larger (median = 8.4, Fig. 3C). The σ^A isoform-specific TEs were all below the median TE across the transcriptome (5/8 in the bottom quartile, Fig. 3B), whereas the σ^B isoform-specific TEs were all above the median (7/8 in the top quartile). These results indicate that these σ^A-dependent isoforms are translationally repressed compared to most genes, whereas the σ^B-dependent isoforms are translationally activated.

In contrast to the σ^B regulon genes that display complex isoform architectures, genes with predominantly simple isoforms (highlighted in cyan in Fig. 1B–D, Materials and Methods, Rend-seq and ribosome profiling traces for a subset shown in Supplemental Fig. S5) showed largely unchanged translation efficiency (Fig. 1D), consistent with pure transcriptional activation (Fig. 1F). Interestingly, we also found three genes with robust transcriptional activation but little increase in protein synthesis rate (highlighted in yellow in Fig. 1B–D), corresponding to a large decrease in apparent translation efficiency in the σ^A active condition. Two of them (csbA and ywjA) exhibit an isoform arrangement that is converse to the translationally activated ones, with long, σ^B-dependent and short, σ^A-dependent isoforms (Supplemental Fig. S4). In the remaining case (yfkJ), the σ^B-dependent isoform has a truncated Shine–Dalgarno sequence, explaining a nearly 10-fold reduction in translation.

Focusing on the translationally activated σ^B regulon genes, we confirmed that TE was isoform-specific using fluorescent reporter constructs for ctc and yvrE (Fig. 3D). Specifically, we fused the fluorescent protein mNeon-Green to the carboxy-terminal end of each gene. For each fusion protein (ctc-mNeon, yvrE-mNeon), two distinct isoform-specific 5′ untranslated region (5′ UTR) variants were placed under the control of an ectopic promoter: (i) a short-isoform variant (S) that included each gene’s native 5′ UTR corresponding to the σ^A-dependent isoform (as identified by Rend-seq), and (ii) a long-isoform variant (L) that included ~100 additional nucleotides in the upstream region, which covers a portion of the coding sequence (CDS) of the upstream gene in the operon. Additionally, a start codon and nonnative ribosome binding site (RBS) were inserted directly upstream to enable translation of the truncated upstream CDS in the long-isoform variant. We then quantified the isoform-specific TE for each construct by normalizing relative protein expression (from fluorescence, Materials and Methods) to relative mRNA levels (from RT-qPCR, Materials and Methods). We found that these isoform-specific TEs qualitatively recapitulated our sequencing-based measurements (Fig. 3D). Specifically, the isoform-specific TE of the long-isoform constructs was roughly four- to sixfold lower than that of the short-isoform constructs, although any further decreases were difficult to
Extensive secondary structure is associated with translationally repressed, σ^A-dependent isoforms

To understand the possible role of mRNA secondary structures in setting isoform-specific translation efficiency, we computationally folded for the σ^A-dependent isoforms of ctc and yvrE. By mapping the putative Shine–Dalgamo (SD) sequences that recruit ribosome binding (Shine and Dalgamo 1974) onto minimum free energy (MFE) structures (Materials and Methods), we found that the majority of bases in the SD sequences were sequestered deep in stable, long-range structures (Fig. 4A). Strikingly, in both cases the σ^A-dependent 5′ ends were located inside the loop of the long RNA stems, such that the short, σ^A-generated isoforms have their 5′ UTRs entirely liberated from these extended secondary structures. The likelihood of SD sequestration was further supported by calculating the base-pairing probability for each position in the SD sequences, which revealed that the majority of positions were predicted to be paired across the thermodynamic ensemble (base-pairing probability ≈1). Given that SD sequences facilitate ribosome recruitment to mRNA to initiate translation, we expected that the presence of extensive secondary structure at and around these elements in the σ^A-dependent isoforms could plausibly repress translation of the downstream σ^B regulon gene. However, numerous factors in the cellular microenvironment affect the folding dynamics of RNAs, yielding in vivo structures that can differ substantially from their in silico counterparts (Rouskin et al. 2014; Spitale et al. 2015; Burkhardt et al. 2017; Mustoe et al. 2018). Accordingly, we decided to experimentally validate these computationally predicted structures for the σ^A-dependent isoforms of ctc and yvrE.

We used the RNA structure probing method DMS-MaPseq to quantify mRNA structures in vivo. This technique involves treating RNA with the methylating agent dimethyl sulfate (DMS) to modify the base-pairing faces of accessible adenine and cytosine nucleobases. These modifications are subsequently encoded as mutations during reverse transcription using a specialized thermostable group II intron reverse transcriptase, generating a mutational signal that is detectable using high-throughput sequencing and has been shown to correlate with base accessibility (Zubrady et al. 2016; Tomezsko et al. 2020). We used a targeted version of DMS-MaPseq to specifically reverse transcribe and amplify the predicted structural region in the σ^A-dependent isoforms of ctc and yvrE following DMS treatment in vivo (Fig. 4B). After sequencing these amplicons, we examined the per-base mutational fractions against a control without DMS treatment and confirmed that DMS induced a characteristic signal at amino bases (Fig. 4C).

We refolded the σ^A-dependent isoforms of ctc and yvrE using DMS signal as a constraint (Materials and Methods) and found strong agreement with the earlier MFE structures (Fig. 4D). In particular, the regions containing the SD sequences were indeed highly structured in vivo and thus less accessible to the translation machinery. Additionally, the inferred RNA secondary structure in these regions was robust to the choice of the folding window size (Materials and Methods). These extended structures that occlude the ribosome binding sites are consistent with the repressed translation of the long, σ^A-dependent isoforms.

After validating the computationally predicted secondary structures by DMS-MaPseq, we extended our computational analysis to additional translationally activated σ^B regulon genes and found a consistent pattern of characteristic structures in the σ^A-dependent isoforms that sequester the sequence elements required for translation initiation (Fig. 5). Similar to ctc and yvrE, the remaining six genes for which we estimated isoform-specific TE all displayed MFE structures with the SD sequences located in extended stem–loops, and base-pairing probabilities indicated that the SD sequences were predominantly paired. These results suggest that these other σ^A-dependent long isoforms are also translationally repressed by extensive secondary structures, like the orthogonally validated instances of ctc and yvrE.

Internal σ^B promoters liberate mRNA secondary structure and activate translation

In contrast to being repressed in the σ^A-dependent isoforms, genes in the short, σ^B-dependent isoforms had above-normal levels of translation (Fig. 3C). The single-nucleotide resolution afforded by Rend-seq data revealed a common feature among this group of genes: the TSSs of the σ^B-dependent isoforms were located within the extended secondary structure, often inside the loop region or in the downstream stem (Figs. 4A, 5, magenta and arrow). Therefore, σ^B-driven transcription generates isoforms with 5′ UTRs that lack the upstream portion of the stem sequestering the SD sequence in the long, σ^A-dependent
isoforms, thereby freeing up the ribosome binding site for efficient translation initiation.

The prevalence of this regulation suggests an alternative configuration for $\sigma^B$-dependent gene expression that does not entirely rely on its canonical role as acting at the transcriptional level. In this operonic architecture, $\sigma^A$-driven promoters produce long, polycistronic mRNAs containing stable structures that impede translation initiation for $\sigma^B$ regulon genes located at the ends of these transcripts (Fig. 6). When activated by stress, however, $\sigma^B$ initiates transcription from alternative promoters directly upstream of its regulon genes, bypassing the inhibitory secondary structures and thereby promoting ribosome binding on these shorter mRNAs. The resulting increase in protein expression predominantly arises from a greater ribosome flux on these transcripts, demonstrating a novel function for $\sigma^B$ in regulating gene expression in a simultaneous transcriptional-translational induction.

**FIGURE 4.** $\sigma^A$-dependent mRNA isoforms have extended secondary structures in vivo. (A) Minimum free energy (MFE) structures of the $\sigma^A$-dependent isoforms of ctc and yvrE near the ribosome binding site. The transcription start sites of $\sigma^B$-dependent isoforms (indicated with arrows), putative Shine–Dalgarno (SD) sequences, and start codons are highlighted in magenta, blue, and green, respectively. The stop codon of the upstream gene in the operon is indicated with an orange box. Computationally determined base-pairing probabilities for individual bases in the SD sequences are shown beside each structure. (B) DMS-MaPseq workflow for in vivo RNA structure determination of $\sigma^A$-dependent isoforms. (C) Cumulative distributions of the per-base mutational fractions for the $\sigma^A$-dependent isoforms of ctc and yvrE. Solid and dashed lines indicate conditions with and without DMS treatment. (D) DMS-constrained MFE structures of representative transcripts for $\sigma^A$-dependent isoforms of ctc and yvrE colored by normalized DMS-MaPseq mutation rate (DMS signal), where values correspond to increased base accessibility. Structured regions containing putative SD sequences are magnified.
DISCUSSION

Bacterial sigma factors have long been studied as quintessential examples of gene regulation. Mechanistically, their direct effects on transcription initiation are well-understood (Paget 2015). We expand this view by demonstrating that the alternative sigma factor \( \sigma^B \) in \( B. \) subtilis can also influence translation initiation for several of its regulon genes. Translation activation is accomplished by modulating isoform-specific RNA secondary structures that normally impede translation initiation. mRNA isoform-specific modulation of translation efficiency has been noted before in other species such as the classic example of the galactose operon in \( E. \) coli (Queen and Rosenberg 1981), as well as in eukaryotic systems (Floor and Doudna 2016), but not previously in the mechanistic...
context of alternative sigma factor induction. This multifunctional control of transcription and translation by a single trans-acting factor serves as a strategy to enable massive up-regulation of gene expression under specific cellular conditions.

The RNA secondary structures that impede translation in the long, σ^A-dependent isoforms often include regions of the upstream open reading frames (ORFs), raising questions about whether ribosomes translating the upstream ORFs may perturb the formation of the inhibitory secondary structures. Ribosomes are known to unwind structured regions of RNA as they elongate over coding sequences (Takyar et al. 2005; Wen et al. 2008). We observed that the stop codon of the upstream gene in the operon was typically located within the large stem-loop (Figs. 4, 5). This places ribosomes in proximity to the critical structural elements if the upstream message is actively translated. However, the results from our fluorescent reporter assay show that this configuration is not capable of fully restoring translation for either ctc or yvrE, despite the upstream gene being driven by an exogenous ribosome binding site with the consensus SD sequence. These data suggest that translation of the upstream gene is insufficient to fully derepress downstream genes, presumably because the ribosome footprint does not extend sufficiently downstream to disrupt RNA structure, or possibly due to rapid refolding of secondary structures after ribosomes pass through.

What is the utility of this regulatory strategy? From an evolutionary perspective, it seems counterintuitive for these genes to be found within larger operons despite being lowly translated. We could instead imagine a transcription terminator evolving in the region between the upstream genes in the operon and the σ^B-dependent TSS, which would ensure that the σ^B regulon gene is only induced upon activation of the general stress response. One potential explanation for multifunctional regulation is to allow fine-tuned expression of some σ^B regulon genes during non-stress conditions. On the one hand, this transcript architecture enables these genes to be transcribed during exponential growth. On the other hand, translation may have been selected against in the same condition to avoid fitness defects from overexpression. In this case, the observed basal expression from the σ^A-driven isoforms would be sufficient for their functions during nonstress conditions.

Another possible explanation for this regulatory strategy could be that small amounts of these proteins are necessary for coping with general stress during transitional periods where σ^B has already been activated but synthesis of general stress proteins is still ongoing. A fitness benefit would be challenging to identify except in specific conditions where the cell relies on one of these particular σ^B regulon genes for survival. Indeed, extensive phenotyping of σ^B-regulon member deletions under varied stresses has demonstrated the limited impact of individual proteins on cell fitness (Höper et al. 2005). Given a lack of characterization for most of these genes, we did not find functional commonalities among them beyond their association with general stress. Identifying the exact stress conditions in which this regulatory strategy confers a fitness advantage constitutes an interesting future direction.

Regardless of the function of σ^B-dependent translational activation, our characterization of sigma factor-mediated dual induction (Fig. 6) expands our view of the regulatory roles of sigma factors and reveals an intriguing principle of bacterial genome organization that could be further investigated in similar organisms. Indeed, inspection of the intergenic regions for the operons considered above among other Bacilli revealed evidence of conserved RBS sequestration in long isoforms which were liberated in short isoforms originating from putative σ^B promoters (Supplemental Data 1; Supplemental Table S2; Materials and Methods), suggesting functional roles of dual transcription-translation activation. Beyond the general stress regulon, this observed principle could be at play for other alternative sigma factors, as many of their regulon genes (33%, excluding σ^B-dependent genes) have at least half of their RNA levels derived from longer isoforms with upstream transcription start sites (Supplemental Data 2; Materials and Methods).
TABLE 1. Strains and plasmids used in this study

| Name     | Genotype                                    | Origin     |
|----------|---------------------------------------------|------------|
| GLB115   | BS168, wild-type Bacillus subtilis subsp. 168 | J. Wang    |
| GLB572   | BS168 levB::Pveg-ctc-S-mNeon kanR           | This study |
| GLB573   | BS168 levB::Pveg-ctc-L-mNeon kanR           | This study |
| GLB574   | BS168 levB::Pveg-yvrE-S-mNeon kanR          | This study |
| GLB575   | BS168 levB::Pveg-yvrE-L-mNeon kanR          | This study |
| pDMM001  | pJBL044 ctc-S-mNeon                         | This study |
| pDMM002  | pJBL044 ctc-L-mNeon                         | This study |
| pDMM003  | pJBL044 yvrE-S-mNeon                        | This study |
| pDMM004  | pJBL044 yvrE-L-mNeon                        | This study |

MATERIALS AND METHODS

Strains and strain construction

Strains used to generate new data in this study are listed in Table 1. Strains pertaining to matched Rend-seq and ribosome profiling data sets retrieved from GEO accession GSE162169 (Lalanne et al. 2021) are listed in Supplemental Table S1.

To construct the strains for the fluorescent reporter assay, the genes ctc and yvrE (with variable upstream regions) were fused to the fluorescent protein mNeonGreen with a carboxy-terminal linker and cloned into pJBL044 under the constitutive promoter Pveg using Gibson assembly (New England Biolabs). The original pJBL044 plasmid was constructed using isothermal assembly from a fragment of pDR160 (Bose and Grossman 2011), a kanR cassette (Guérout-Fleury et al. 1995), levB homology regions, the Pveg promoter, and the strong efp terminator. The assembled plasmids were transformed into Mix and Go! E. coli DH5 Alpha Competent Cells (Zymo Research) per the manufacturer’s instructions and isolated using a QiAprep Spin Miniprep Kit (QIAGEN). The fusion constructs were then integrated into BS168 at the levB locus using standard cloning techniques (Harwood and Cutting 1990), and successful recombinants were verified by colony PCR. All plasmids and recombinants (see Table 1) were further validated by Sanger sequencing (Quintara Biosciences).

Growth conditions

Unless indicated otherwise, all strains were grown at 37°C with shaking (250 rpm) in LB supplemented with carbenicillin (100 µg/mL for E. coli) and/or kanamycin (50 µg/mL for E. coli, 5 µg/mL for B. subtilis) when appropriate. For overnight cultures, LB liquid media was inoculated with single colonies from LB agar plates.

For matched Rend-seq/ribosome profiling data sets, strains were grown in LB or conditioned MCC medium (Parker et al. 2020; Lalanne et al. 2021) with various inducer (xylose, IPTG) concentrations (see Supplemental Table S1). For these data sets, cells were grown in exponential phase for at least 10 doublings before harvesting at OD600 ≈ 0.3.

Existing Rend-seq and ribosome profiling data sets

Matched Rend-seq and ribosome profiling data sets used to identify genes with increased TE (Fig. 1) and to estimate the short isoform fraction and corresponding apparent TE (Fig. 3; Supplemental Fig. S3) were obtained from GEO accession GSE162169 (Lalanne et al. 2021). These data sets display a range of σB activation due to a diverse set of genetic modifications and growth media. In particular, we previously identified that tuning the expression of translation termination factors RF2 and PrmC activates σB to varying degrees (Lalanne et al. 2021). For example, the σB active data presented in Figures 1 and 2 correspond to a CRISPRi knockdown of RF2, while σB inactive corresponds to wild-type. Importantly, although it is possible that different RF2 levels could affect translation initiation (and therefore TE) of genes (Lalanne et al. 2021), none of the genes that show a substantial increase in TE (Fig. 1) have a UGA stop codon or are cotranscribed with a gene ending with UGA stop (UGA being the stop codon cognate to RF2). Hence, the molecular causes of σB activation are distinct and independent from the mechanisms leading to translational activation characterized here.

Quantification of mRNA level, ribosome footprint density, and translation efficiency

From pile-up files (.wig format), the mRNA level corresponding to a gene was quantified as the 1% winsorized average read density for 3'-end mapped Rend-seq reads across the body of the gene, excluding a 40 nt region the start and end of the gene (start+40 nt to end−40 nt for averaging). Ribosome footprint read density was similarly calculated (1% winsorized density from start+40 nt to end−40 nt). Read densities were then normalized to rpkm (reads per kilobase per million reads mapped) using the total number of reads mapping to non-rRNA or tRNAs. For all genes, bootstrap (randomly sampling from the distribution of read counts per position across the body of the gene and calculating the corresponding resampled density and downstream quantities) was used as a measure of technical and read count variability. Error bars in Figure 3A and Supplemental Figure S3 correspond to the standard deviation across bootstrap subsamplings. Large error bars correspond to large counting noise (regions with few reads mapped). The translation efficiency of each gene was calculated as the ribosome profiling rpkm divided by the Rend-seq rpkm. Only genes with >50 reads mapped were considered to identify candidates with substantially elevated TE (Fig. 1).
Changes in translation efficiency for induced \( \sigma^B \) regulon genes

To identify genes with increased translation efficiency, we used a threshold of a \( >2.7\)-fold increase in apparent translation efficiency in the \( \sigma^B \) active vs. inactive conditions.

A \( >2.7\)-fold decrease in apparent TE was used to mark genes with repressed translation upon \( \sigma^B \) induction. Among the four genes with repressed translation (cspA, yklJ, ywiA, sigB), the gene for sigB itself was excluded from further consideration because of the large overlap of its open reading frame with the upstream gene and to avoid interpretation difficulty arising from the translation termination defect in the RF2 knockdown condition. The three remaining genes are highlighted in yellow in Figure 1B–D, with Rend-seq and ribosome profiling traces shown in Supplemental Figure S4.

To identify induced \( \sigma^B \) regulon genes \( (>3.7\)-fold increase in Rend-seq and/or ribosome profiling read density and classification as a member of the annotated \( \sigma^B \) regulon, \( n = 50 \) with simple mRNA isoforms, we leveraged our deeply sequenced Rend-seq data set in LB (Lalanne et al. 2018) to exclude genes which were not the first gene of their mRNA (e.g., second gene in a polycistronic transcript), displayed multiple upstream transcription start sites in addition to the \( \sigma^B \)-dependent start site, or had substantial transcription from long isoforms. The resulting ”simple isoform” \( \sigma^B \) regulon genes (highlighted in cyan in Fig. 1B–D) displayed a much more restricted range in fold-change in apparent TE across the \( \sigma^B \) active vs. inactive conditions. A subset have their transcriptional and translational responses separately displayed in Figure 1F, and Rend-seq/ribosome profiling traces are shown for some examples in Supplemental Figure S5.

The above analyses are summarized in Supplemental Table S3.

**Determination of isoform-specific TE**

To estimate the isoform-specific TE for particular genes, we assume that each individual mRNA isoform has a distinct TE, and that the total ribosome footprint density for a gene with multiple mRNA isoforms is equal to the sum of the isoform-specific TEs weighted by the mRNA abundance of each isoform.

Specifically, consider a two-gene operon with a long isoform that includes both gene 1 and gene 2 as well as a short isoform that contains gene 2 exclusively (schematically illustrated in Supplemental Fig. S3A). Denote overall mRNA level for genes 1 and 2 by \( m_1 \) and \( m_2 \), and overall ribosome footprint density \( r_1 \) and \( r_2 \) for the two genes, respectively. Further, let \( m_{short} \) and \( m_{long} \) be the level of the short and long isoform, respectively, and \( TE_{short} \) and \( TE_{long} \) the corresponding isoform-specific TE. Note that the overall mRNA level for genes 1 and 2 are related to isoform mRNA levels by: \( m_1 = m_{long} + m_{short} \) and \( m_2 = m_{short} + m_{long} \). Hence, from the total mRNA level for both genes, we can infer the isoform mRNA levels: \( m_{short} = m_1, \) and \( m_{short} = m_2 - m_1 \).

By assumption, for the ribosome density on gene 2: \( r_2 = m_{short} TE_{short} + m_{long} TE_{long} \). For the apparent TE of gene 2, we thus have:

\[
TE_{apparent}^{2} = \frac{r_2}{m_2} = \frac{m_{short} TE_{short} + m_{long} TE_{long}}{m_2}.
\]

Reorganizing the equation leads to:

\[
TE_{apparent}^{2} = f_{short} TE_{short} + (1 - f_{short}) TE_{long}.
\]

For the fluorescence reporter assay, the strains GLB115, GLB573, GLB574, and GLB575 were grown to \( OD_{600} \approx 1-2 \) and then back-diluted 200-fold into fresh media. Three technical replicates per culture were grown at 37°C for 12 h in a BioTek Synergy H1 microplate reader, and absorbance (600 nm) and fluorescence intensity (EX 485/20 nm, EM 520/20 nm) were measured every 5 min. Fluorescence was normalized by absorbance at each time point, and any background signal from cellular/media autofluorescence was removed by subtracting the mean normalized fluorescence values of the wild-type BS168 replicates. These quantities were then converted to relative values by normalizing proportionally to the signal for the S reporters.

For reverse transcription-qPCR (RT-qPCR), overnight cultures of the same strains were back-diluted to \( OD_{600} \approx 2 \times 10^{-4} \) and regrown for roughly 10 generations. At \( OD_{600} \approx 0.3 \), 5 mL of cells were harvested and mixed with 5 mL of chilled methanol, spun down at 4°C for 10 min, and frozen at −80°C after removing the supernatant. Thawed cell pellets were treated with 100 µL of 10 mg/mL lysozyme in TE, and total RNA was extracted using an RNeasy Mini Kit (QIAGEN). DNA was removed using TURBO DNase (Thermo Fisher Scientific), and RNA was purified using isopropanol precipitation. Reverse transcription was performed.
using Random Hexamer Primer (Thermo Fisher Scientific) and M-MuLV Reverse Transcriptase (New England Biolabs) per the manufacturer's instructions. RNA levels were measured on a Roche LightCycler 480 Real-Time PCR system using two primer sets for mNeon and one primer set each for the loading controls gyrA and sigA (mNeon F1, mNeon R1, mNeon F2, mNeon R2, gyrA F, gyrA R, sigA F, sigA R, see Table 2). The fold change in mNeon RNA levels relative to the S reporters was calculated by taking the average of three technical replicates across each combination of primer sets (mNeon1/gyrA, mNeon1/sigA, mNeon2/gyrA, mNeon2/sigA).

Isoform-specific TE was subsequently calculated by normalizing mean relative fluorescence by mean fold change in mNeon RNA levels, and the standard deviation was propagated from each measurement type.

**RNA secondary structure prediction**

Minimum free energy (MFE) structures were predicted using the RNAdfold program of the ViennaRNA Package (Lorenz et al. 2011) with default parameters. Base-pairing probabilities were determined by constraining each position in a sequence individually as unpaired and then calculating the partition function from the ensemble free energy computed by RNAdfold. The probability of each position being unpaired was calculated by dividing the partition function for the constrained sequence by the partition function for an unconstrained sequence, and the base-pairing probabilities were simply the probabilities of the complements. Putative Shine–Dalgamo (SD) sequences were identified as the region upstream of the start codon that forms the strongest duplex with the anti-Shine–Dalgamo (asD, 5′-TACACCTCCT-3′) sequence in the 16S ribosomal RNA. RNA secondary structures determined using RNAdfold were visualized using VARNA v3.93 (Visualization Applet for RNA) (Darty et al. 2009). The structures sequestering the ribosome binding sites shown in Figures 4 and 5 were confirmed to be robust to the specific regions computationally folded, both at the level of secondary structure and base-pairing probabilities of the SD sequences.

**DMS-MaPseq**

In vivo DMS treatment was performed as previously described (Zubradt et al. 2016; Tomeszko et al. 2020). Specifically, an overnight culture of BS168 was split two ways and back-diluted to OD600 ≈ 0.2, 15 mL of each culture was incubated at 37°C for 2 min with shaking (1000 rpm) after treating one with 750 µL of dimethyl sulfate (DMS, ~5% final concentration). The reaction was stopped by adding 30 mL of chilled stop solution (30% β-mercaptoethanol, 25% isoamyl alcohol) to each sample, after which they were immediately transferred to ice and spun down at 4°C for 8 min. The cell pellets were washed with 8 mL of chilled wash solution (30% β-mercaptoethanol), resuspended in residual wash solution, and frozen at ~80°C. Thawed cell pellets were treated with 100 µL of 10 mg/mL lysozyme in TE, and total RNA lysis buffer (10 mM EDTA, 50 mM sodium acetate) was added to 650 µL. Total RNA was extracted using hot acid-phenol:chloroform and isopropanol precipitation.

For library preparation, established protocols (Zubradt et al. 2016; Tomeszko et al. 2020) were followed. DNA was removed using TURBO DNase, and RNA >200 nt was purified using an RNA Clean & Concentrator-5 Kit per the manufacturer's instructions (Zymo Research). Ribosomal RNA was depleted using a MICROBExpress Bacterial mRNA Enrichment Kit (Thermo Fisher Scientific), and RNA >200 nt was again purified using an RNA Clean & Concentrator-5 Kit. Reverse transcription was performed at 64°C for 90 min using 70 ng of RNA from each sample and TGI RT-III (Ingenix). The RT primers were specific to each gene (ctc R, yvrE R, see Table 2). The RT reaction was treated with 1 µL RNase H (New England Biolabs) and incubated at 37°C for 20 min to remove RNA. Roughly 1/10 of the resulting volume was used as template for a two-step PCR amplification with Phusion High-Fidelity DNA Polymerase (New England Biolabs) per the manufacturer's specifications, which was run for 15–25 cycles with the RT primer serving as the reverse primer (ctc F, yvrE F, see Table 2). PCR products (~240–290 bp) were purified by gel extraction on an 8% TBE polyacrylamide gel (Thermo Fisher Scientific) and isopropanol precipitation. Samples with particularly low dsDNA concentrations (as measured on an Invitrogen Qubit 4 Fluorometer) were reamplified for 7–20 additional cycles and purified in the same manner. After adding adapters via PCR, the libraries were sequenced on an Illumina MiSeq (2 × 250 nt reads).

To determine the DMS signal, FASTQ files were processed and analyzed using the DREEM (Detection of RNA folding Ensembles using Expectation-Maximization clustering) pipeline with the “–fastq” and “–struct” options (Tomezsko et al. 2020). In brief, paired-end reads were filtered for quality and trimmed using FASTQC v.0.11.8 and TrimGalore 0.4.1, respectively. Reads were aligned to target sequences in the reference genome NC_000964.3 (the NCBI RefSeq database (O’Leary et al. 2016) using Bowtie2 2.3.4.1 with the options “–local” and “–no-unal” and “–no-discordant” and “–no-mixed” and “–X 1000 -L 12.” Mapped reads were represented as bit vectors and clustered by their mutational signatures using the DREEM algorithm with standard parameters (Tomezsko et al. 2020). Per-base mutational fractions were initially quantified using the population-average fraction of mismatches and deletions. Following expectation-maximization (EM)
clustering, the DMS reactivity was taken as the mutation rates of the bases in the cluster K = 1. After normalizing to the median of the top 5% of positions (with the upper limit set to 1.0), the DMS signal was used as a folding constraint for predicting RNA secondary structures with the program RNAstructure v.6.0.1 (Reuter and Mathews 2010). Additionally, the folding windows were expanded symmetrically by 50, 100, 150, and 200-nt in either direction to assess the robustness of the predicted folds. RNA secondary structures were visualized using VARNA v3.93 (Darty et al. 2009). The sequencing data sets for DMS-MaPseq are available online using the GEO accession GSE168393.

Conservation analysis

To assess whether σ^A-σ^B isoform configurations and RNA secondary structures in the long isoforms were conserved in other species from the Bacillus genus, we extracted and annotated intergenic regions for the σ^A regulon genes with marked TE induction displaying both short and long isoforms (Fig. 2; Supplemental Fig. S1; ctc, yphD, yhfR, yocK, ycdB, yhfH, yxaB, csbX). Analysis was restricted to Bacillus species (genus taken from the GTDB taxonomy [Parks et al. 2018]) within the reference and representative bacterial genomes from RefSeq (O’Leary et al. 2016) with an identified homolog of the rsbV-rsbW-sigB operon (Lalanne et al. 2021), leading to 26 species analyzed (listed in Supplemental Table S2). For all these Bacillus species, homologs of pairs of genes involving the σ^A-σ^B isoform configurations from B. subtilis (RefSeq protein accession listed in Supplemental Table S2) were taken as query for a blastp search (Ye et al. 2006) with an E-value cutoff of 1 × 10^{-7}. Bacillus species in which the two genes were conserved, found in the same order, and separated by <400 bp were retained for further analysis.

Most operons considered were not widely conserved in the other Bacillus species, with all but the prs-ctc operon (conserved in 18/25 Bacillus) conserved in up to two more species (Bacillus atropheus and Bacillus amyloiquefaciens, see Supplemental Table S2). The sequences of the intergenic regions for conserved operons were extracted, folded for minimum free energy RNA secondary structures using RNAfold (Lorenz et al. 2011), and annotated for putative Shine-Dalgarno sequences (as described above), σ^A-dependent promoters (region of maximum local alignment using the Smith-Waterman algorithm to the consensus motif GTTAA(13X)GGGWAW or GTTAA(14X)GGGWAW using the nuc44 scoring matrix), and possible intervening terminators (from the list of high-confidence, bioinformatically identified terminators in Johnson et al. 2020). Supplemental Data 1 summarizes the analysis. We found evidence for RNA secondary structures sequestering Shine-Dalgarno sequences in the σ^A-dependent isoforms, and freed in the σ^B-dependent isoforms, for the overwhelming majority of intergenic regions in other Bacillus species (Supplemental Table S2; Supplemental Data 1) despite changes in sequence. Interestingly, most of the examples with weaker structures had evidence for a strong intrinsic terminator upstream of the σ^B-dependent promoter.

Analysis of isoform architecture for genes associated with other alternative sigma factors

To assess whether genes under the control of other alternative sigma factors also pervasively displayed transcription from longer upstream mRNA isoforms, we analyzed a deep Rend-seq data set from B. subtilis in LB (Lalanne et al. 2018). For all annotated promoters in DBTBS (Sierra et al. 2008) associated with alternative sigma factors with positional information (n = 319, excluding σ^B-dependent genes), we computed the Rend-seq read density in windows −115 to −15 (TSS-upstream) and +15 to +115 (TSS-downstream) positions relative to the annotated transcription start site of the promoter. Promoters with downstream read density lower than 0.1 reads/nt were not considered further (below expression cutoff, n = 122). For the remaining 197 promoters, we calculated the ratio of TSS-downstream to TSS-upstream read densities and retained instances in which the ratio was larger than 0.5 (i.e., 50% of the expression coming from a putative long isoform). To exclude cases where the signal arose from a separate upstream transcript as opposed to a bona fide long isoform, instances with a mapped 3′ end (3′ peak z-score >12) in the region −115 to +115 were further excluded. In fine, 33% (65/197) of expressed genes downstream from annotated alternative sigma factor promoters had evidence for most of their transcription coming from a long upstream isoform in LB (summarized in Supplemental Data 2). This suggests that the isoform-specific translational activation described in the present work could be applicable to other sigma factors in B. subtilis.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank members of the Li laboratory and the Rouskin laboratory for critical discussions. We thank S. McKeithen-Mead for providing the mNeon template. This research was supported by National Institutes of Health (NIH) grant R35GM124732, the National Science Foundation (NSF) CAREER Award, the Smith Odyssey Award, the Pew Biomedical Scholars Program, a Sloan Research Fellowship, the Searle Scholars Program, the Smith Family Award for Excellence in Biomedical Research, a Natural Sciences and Engineering Research Council of Canada (NSERC) doctoral Fellowship, and a Howard Hughes Medical Institute (HHMI) International Student Research Fellowship (to J.-B.L.).

Received March 7, 2021; accepted April 23, 2021.

REFERENCES

Bastet L, Turcotte P, Wade JT, Lafontaine DA. 2018. Maestro of regulation: riboswitches orchestrate gene expression at the levels of translation, transcription and mRNA decay. RNA Biol 15: 679–682. doi:10.1080/15476286.2018.1451721

Bhattacharyya S, Jacobs WM, Adkar B V, Yan J, Zhang W, Shakhnovich EI. 2016. Accessibility of the Shine-Dalgarno sequence dictates N-terminal codon bias in E. coli. Mol Cell 70: 894–905.e5. doi:10.1016/j.molcel.2018.05.008

Boël G, Letto R, Neely H, Price WN, Wong KH, Su M, Luff JD, Valecha M, Everett JK, Acton TB, et al. 2016. Codon influence on protein expression in E. coli correlates with mRNA levels. Nature 529: 358–363. doi:10.1038/nature16509

Borujeni AE, Cetnar D, Farasat I, Smith A, Lundgren N, Salis HM. 2017. Precise quantification of translation inhibition by mRNA structures
Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. 2014. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**: 701–705. doi:10.1038/nature12894

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* **71**: 1342–1346. doi:10.1073/pnas.71.4.1342

Sierro N, Makita Y, de Hoon M, Nakai K. 2008. DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information. *Nucleic Acids Res* **36**: D93–D96. doi:10.1093/nar/gkm910

Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung JW, Kuchelmeister HY, Batista PJ, Torre EA, Kool ET, et al. 2015. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**: 486–490. doi:10.1038/nature14263

Takyar S, Hickerson RP, Noller HF. 2005. mRNA helicase activity of the ribosome. *Cell* **120**: 49–58. doi:10.1016/j.cell.2004.11.042

Tomezsko PJ, Corbin VDA, Gupta P, Swaminathan H, Glasgow M, Persad S, Edwards MD, McIntosh L, Papenfuss AT, Emery A, et al. 2020. Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature* **582**: 438–442. doi:10.1038/s41586-020-2253-5

Wen JD, Lancaster L, Hodges C, Zeri AC, Yoshimura SH, Noller HF, Bustamante C, Tinoco I. 2008. Following translation by single ribosomes one codon at a time. *Nature* **452**: 598–603. doi:10.1038/nature06716

Yakhnin H, Zhang H, Yakhnin AV, Babitzke P. 2004. The trpRNA-binding attenuation protein of *Bacillus subtilis* regulates translation of the tryptophan transport gene trpP(yhaG) by blocking ribosome binding. *J Bacteriol* **186**: 278–286. doi:10.1128/JB.186.2.278-286.2004

Yakhnin H, Pandit P, Petty TJ, Baker CS, Romeo T, Babitzke P. 2007. CsrA of *Bacillus subtilis* regulates translation initiation of the gene encoding the flagellin protein (flag) by blocking ribosome binding. *Mol Microbiol* **64**: 1605–1620. doi:10.1111/j.1365-2958.2007.05765.x

Ye J, McGinnis S, Madden TL. 2006. BLAST: improvements for better sequence analysis. *Nucleic Acids Res* **34**: W6–W9. doi:10.1093/nar/gkl164

Zhu B, Stülke J. 2018. SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism *Bacillus subtilis*. *Nucleic Acids Res* **46**: D743–D748. doi:10.1093/nar/gkx908

Zubradt M, Gupta P, Persad S, Lambowitz AM, Weissman JS, Rouskin S. 2016. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat Methods* **14**: 75–82. doi:10.1038/nmeth.4057