Systematic Quantification of Sequence and Structural Determinants Controlling mRNA stability in Bacterial Operons

Daniel P. Cetnar[1] and Howard M. Salis[1,2,3*]

[1]Department of Chemical Engineering, [2]Department of Biological Engineering, [3]Department of Biomedical Engineering. The Pennsylvania State University, University Park, PA, 16802.

* To whom correspondence should be addressed: salis@psu.edu

ABSTRACT

mRNA degradation is a central process that affects all gene expression levels, and yet the determinants that control mRNA decay rates remain poorly characterized. Here, we applied a synthetic biology, learn-by-design approach to elucidate the sequence and structural determinants that control mRNA stability in bacterial operons. We designed, constructed, and characterized 82 operons, systematically varying RNAse binding site characteristics, translation initiation rates, and transcriptional terminator efficiencies in the 5' UTR, intergenic, and 3' UTR regions, and measuring their mRNA levels using RT-qPCR assays. We show that introducing long single-stranded RNA into 5' UTRs reduced mRNA levels by up to 9.4-fold and that lowering translation rates reduced mRNA levels by up to 11.8-fold. We also found that RNAse binding sites in intergenic regions had much lower effects on mRNA levels. Surprisingly, changing transcriptional termination efficiency or introducing long single-stranded RNA into 3' UTRs had no effect on upstream mRNA levels. From these measurements, we developed and validated biophysical models of ribosome protection and RNAse activity with excellent quantitative correspondence. We also formulated design rules to rationally control a mRNA's stability, facilitating the automated design of engineered genetic systems with desired functionalities.

INTRODUCTION

Engineering sophisticated genetic systems requires the development of more comprehensive biophysical models that can predict how changes to sequence affect gene expression levels, taking into account transcription, translation, and mRNA degradation (1-4). mRNA decay rates are an important contributor to genetic circuit function, altering the circuit's dynamic and steady-state gene expression levels as well as controlling "turn off" times in response to changing transcriptional programs (5,6). While significant focus has been given to measuring and modeling transcriptional initiation rates (7-11), transcriptional control (12,13), and translation initiation rates (14-16), previous work on mRNA stability has primarily focused on discovering and characterizing the proteins and pathways responsible for mRNA decay (17-21).

A key challenge to modeling and predicting mRNA decay rates arises from the many participating enzymes and the high degree of coupling between transcription, translation, and RNAse activity (22,23). Predicting mRNA decay rates in multi-cistronic operons is further complicated by the potential for differential decay rates across coding sequences, called polarity (24,25). Here, we apply a learn-by-design approach to quantitatively measure how a bacterial mRNA's sequence and structural properties determine its stability through the several mechanisms of mRNA decay. From these results, we formulate quantitative design rules for controlling a mRNA's stability.

Several enzymes work together in a multi-pathway process to determine how quickly bacterial mRNAs are degraded (20,21), including endonucleases (*E. coli* RNAses E, G, III), exonucleases (*E. coli* RNAse II, PNPase, RNAse R), and the helper enzymes that modify the ends of mRNAs (*E. coli* RppH, DapF, PAPI) (**Figure 1**). These pathways are intimately coupled to both transcription and translation. Beginning with transcription initiation, the 5' ends of all bacterial mRNAs contain a triphosphate group that inhibits RNAse activity. However, over time, RppH hydrolyzes these 5' ends, yielding a monophosphate group that has higher affinity to RNAse E (26,27). RppH activity is promoted by the protein DapF and depends on the sequence and structure of at least the first three nucleotides of the transcript (**Figure 1A**) (28-30). As transcription elongation proceeds, co-transcriptional folding determines the types of mRNA structures that form and the portions that remain unstructured and accessible, which is also affected by RNA polymerases' transcriptional elongation rate (31,32). During this time, as start codons are transcribed, ribosomes will bind and begin to translate the mRNA, following RNA polymerase and covering the mRNA with protective ribosomes. Once the mRNA is fully transcribed, the translation initiation and elongation rates of its coding sequences will determine the number of protective ribosomes bound to the transcript (**Figure 1B**).

As transcription and translation proceeds, RNAses will interact with all accessible mRNA regions, binding and cleaving where possible. RNAse E and RNAse G cleave mRNAs at either their 5' ends via end-dependent entry (**Figure 1C**) or at internal sites that remain accessible and unstructured (direct entry) (33-35) (**Figure 1D**). Other RNAses (e.g. RNAse III) bind to specific

types of RNA structures, initiating their cleavage activity (36,37). Notably, an RNA endonuclease's first cleavage event will create two new ends in the transcript, a mono-phosphate 5' end and a 3' end, that can then acted upon by the RNA degradosome (38), a loosely associated multi-protein complex, that combines both exonucleases and endonucleases (e.g. oligo-RNase, RNAse R, PNPase) to rapidly reduce the transcript to short oligonucleotides and mononucleotides (39,40). These activities include endonucleases that bind nearby monophosphate 5' ends and cleave internally, releasing oligonucleotides (33) as well as exonucleases that bind to structurally accessible 3' ends and chew them back in a 3' to 5' direction, releasing mononucleotides (41). Helper enzymes (e.g. PAPI) improve the accessibility of 3' ends by adding unstructured polyA tails (39), accelerating exonuclease activity. Altogether, after the first cleavage event takes place, large regions of the mRNA transcript can be rapidly processed to mononucleotides. Overall, the rate of the first RNAse cleavage event is the slowest, making it a rate-limiting step in determining a mRNA's overall stability (42). In particular, a first cleavage event inside the 5' UTR or intergenic regions is often rapidly followed by destruction of start codons in nearby coding regions, eliminating protein expression.
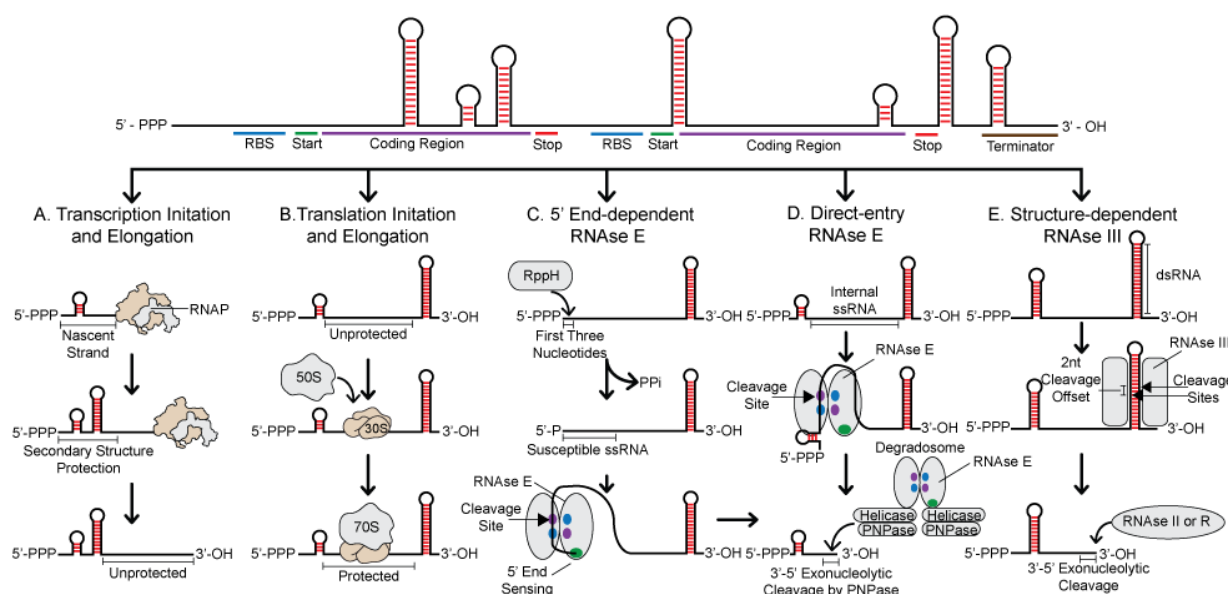
Figure 1: mRNA sequence determinants control mRNA Stability through coupling of transcription, translation, and RNAse activity. (A) Transcription elongation rates control mRNA structure and accessibility at 5' ends. (B) Translation initiation and elongation rates control the number of protective ribosomes bound to mRNA. (C) Both mRNA structure and the first three nucleotides at mRNA 5' ends controls RppH and RNAse E activity via the end-dependent decay pathway. (D) RNAse E binds and cleaves mRNA at single-stranded internal sites unprotected by structure. After initial cleavage, the RNA degradosome processively degrades mRNA. (E) RNAse III targets specific mRNA structures for cleavage. PNPase, RNAse II, and RNAse R degrade mRNA at 3' ends using 3' to 5' exonuclease activity.

Previous studies have qualitatively observed several factors that alter a mRNA's decay rate, including (i) mRNA sequence motifs that may specifically bind RNAses (43,44); (ii) specific sequences at a mRNA's 5' end that alter RppH binding (29,45); (iii) the co-transcriptional, temperature-sensitive formation of mRNA structures that block RNAse binding and increase

mRNA stability (46); (iv) the binding of ribosomes in 5' UTRs and coding regions that block RNAse binding, called protective ribosomes (24,47,48); (v) RNA structures at transcript 3' ends that block exonuclease activities (49); and (vi) changes in environmental growth conditions that differentially alter RNAse and helper enzyme levels (e.g. stress responses) (50). RNAse III is also known to cleave long RNA duplexes (51,52), while specific motifs, such as RAUUW (44), RNWUU (53), and RNAU (33,54), were suggested to specifically bind RNAse E. Overall, *E. coli* natural mRNAs have half-lives of about 1 to 10 minutes with the potential to vary protein expression levels by at least 10-fold (31,55).

However, the precise sequence determinants that alter mRNA stability have not been quantitatively elucidated. Prior studies suggesting specific binding motifs have relied on a small number of characterized mRNAs, while the relationship between RppH activity, translation rates, transcriptional termination efficiency, and transcript stability remains poorly characterized. Because natural mRNAs are often attacked by RNAses at multiple entry sites, it remains challenging to interpret natural sequence-stability assays and identify quantitative cause-effect relationships. Instead, we use a synthetic biology, learn-by-design approach to systematically alter a mRNA transcript's properties and measure its mRNA levels during sustained exponential growth conditions. Through this strategy, we designed, constructed, and characterized the mRNA levels from 82 mono- and bi-cistronic operons, systematically varying the mRNA's sequence and structural properties to modulate RppH, RNAse E/G, and RNAse III activities. We also systematically varied the mRNA's translation rates and transcriptional termination efficiencies to quantitatively determine how they protect mRNA transcripts from endonuclease or exonuclease activities. Altogether, our dataset provides a quantitative and comprehensive understanding of how mRNA sequence controls mRNA stability through several interactions.

MATERIAL AND METHODS

Plasmid Design and Cloning

A series of libraries of pFTV1-derived plasmids with a ColE1 origin of replication and Chloramphenicol resistance (Cm^R) were designed and constructed to express fluorescent protein reporters with the objective of testing specific design motifs' effect on mRNA stability. In **Supplementary Data 1**, the dataset name and sequence information are listed for each construct. For the datasets mRFP1 Expression Library and Terminator Efficiency, a base mono-cistronic mRFP1 pFTV1-derived plasmid was used that contained the σ^{70} constitutive promoter J23100 and a double terminator. To create the mRFP1 Expression Library, the 5' UTR was engineered using RBS Library Calculator to design a set of ribosome binding sites that varied the expression of mRFP1 1000-fold. The degenerate oligonucleotide (IDT) designed by RBS Calculator containing these designs was flanked by a 5' XbaI and 3' NdeI restriction site and primer binding sites, which allowed for PCR amplification. The plasmid was then assembled using cut and paste cloning. Similarly, the Terminator Efficiency set used construct 6 from the mRFP1 Expression Library as the base plasmid. The terminators tested were assembled by annealing two complementary oligonucleotides (IDT) with 5' FseI and 3' SphI restriction overhangs. The annealed oligonucleotides were ligated into the base plasmid replacing the existing terminators. Likewise, for the datasets sfGFP Expression Library and sfGFP Coupled

Expression, the base plasmid contained the σ^{70} constitutive promoter J23100 and a double terminator. RBS Libraries were designed and cloned in for both using a similar method as for the mRFP1 Expression Library, except that the restriction sites employed were 5' XbaI and 3' AgeI. For the datasets, 5' Variable Length Poly A, Intergenic Variable Length Poly A, RNAse Polarity, and RppH Set, a base bi-cistronic codon optimized mRFP1 and GFPmut3B pFTV1-derived plasmid was used that contained the σ^{70} constitutive promoter J23100 and a double terminator. At the 5' UTR, RBS Calculator v2.1 was used to design a ribosome binding site sequence with an XbaI restriction site on the 5' end and with a translation initiation rate of 9395 au on RBS Calculator's proportional scale. The newly designed 5' UTR and J23100 promoter was assembled by annealing two complementary oligonucleotides (IDT) with 5' BamHI and 3' SacI restriction overhangs and ligating them into the base plasmid. Likewise, RBS Calculator v2.1 was used to design a ribosome binding site sequence with an AatII restriction site on the 5' end with a translation initiation rate of 7083 au on RBS Calculator's proportional scale. The newly designed intergenic region was assembled by annealing two complementary oligonucleotides (IDT) with 5' EcoRI and 3' XhoI restriction overhangs and ligating them into the base plasmid. Additions to the base 5' UTR were made using a 5' BamHI site upstream of the promoter and the newly added 3' XbaI site. Sequence additions to the base intergenic UTR was made using an upstream 5' EcoRI site and the newly added 3' AatII site. For additions to the 3' UTR, a 5' PacI site and 3' SpeI site were used. All additions were made by annealing complementary oligonucleotides (IDT) with the correct restriction overhangs for the 5' UTR, intergenic region, and 3' UTR. All plasmids were transformed into *Escherichia coli* DH10B cells, followed by sequence verification of isolated clones.

Strains, Growth and Characterization

All measurements were conducted using *Escherichia coli* DH10B cells containing plasmids, cultured using M9 minimal media, and maintained in the exponential growth phase for at least 20 hours. For each construct, isogenic colonies were used to inoculate overnight cultures in 500 ul LB media supplemented with 50 ug/mL Chloramphenicol (Cm) in a 96-well deep-well plate. Overnight cultures were diluted 100-fold by diluting 2 uL of culture into 198 ul of M9 minimal media with Cm using a 96-well microtiter plate, and incubated at 37 ^{0}C with high orbital shaking inside a Spark spectrophotometer (TECAN). OD_{600} absorbance was taken every 10 minutes until the OD_{600} reached 0.15-.20, indicating the cells' entry into the mid-exponential phase of growth. At this time, a subsequent 96-well microtiter plate was prepared by serial dilution of the culture from the first plate into M9 minimal media with Cm maintaining the cells in the exponential phase of growth. For each culture, single-cell red florescent protein (RFP) or green florescent protein (GFP) fluorescence measurements were performed by collecting 10 ul from the end of the second dilution, transferring to a microtiter plate with 200 ul PBS solution with 2 mg/ml kanamycin, and recording 100,000 single-cell fluorescence levels with a Fortessa flow cytometer (BD Biosciences). All single-cell fluorescence distributions were unimodal. The mean of the distributions is calculated, and the background autofluorescence of *Escherichia coli* DH10B cells subtracted. All reported fluorescence levels are the average of at least two biological replicates, which are listed in **Supplementary Data 1**.

mRNA level measurements were performed on selected strains by inoculating a 5 mL culture of Cm supplemented LB media and incubated at 37 ^{0}C with 300 RPM shaking. Once cells reached an OD_{600} of 1.0, measured using a cuvette-based spectrophotometer (NanoDrop 2000C), they were diluted to an OD_{600} of 0.05 in a 5 mL culture of Cm supplemented M9 minimal media and incubated at 37 ^{0}C with 300 RPM shaking. Once cells reached an OD_{600} of between 0.4-0.6, they were diluted to OD 0.001 in a 5 mL culture of Cm supplemented M9 minimal media and incubated at 37 ^{0}C with 300 RPM shaking. Cells were harvested once they reached an OD_{600} of between 0.15-0.25 and their total RNA extracted using the Total RNA Purification kit (Norgen Biotek), followed by non-specific degradation of contaminant DNA using the Turbo DNAse kit (Ambion). Following extraction, cDNA was prepared using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems). Taqman based qPCR was performed using an ABI Step One Plus real-time thermocycler (Applied Biosystems), utilizing a Taqman probe targeting the constructs of interest mRFP1, codon optimized mRFP1, sfGFP, codon optimized GFPmut3B. Additionally, a custom 16S rRNA TaqMan probe was used as an endogenous control and was used to calculate relative mRNA levels from ΔC_t numbers. TaqMan probes sequence and validation are listed in Supplementary Data 2. Likewise, SYBR Green based qPCR was performed in a similar manner for the determination of the relative mRNA level.
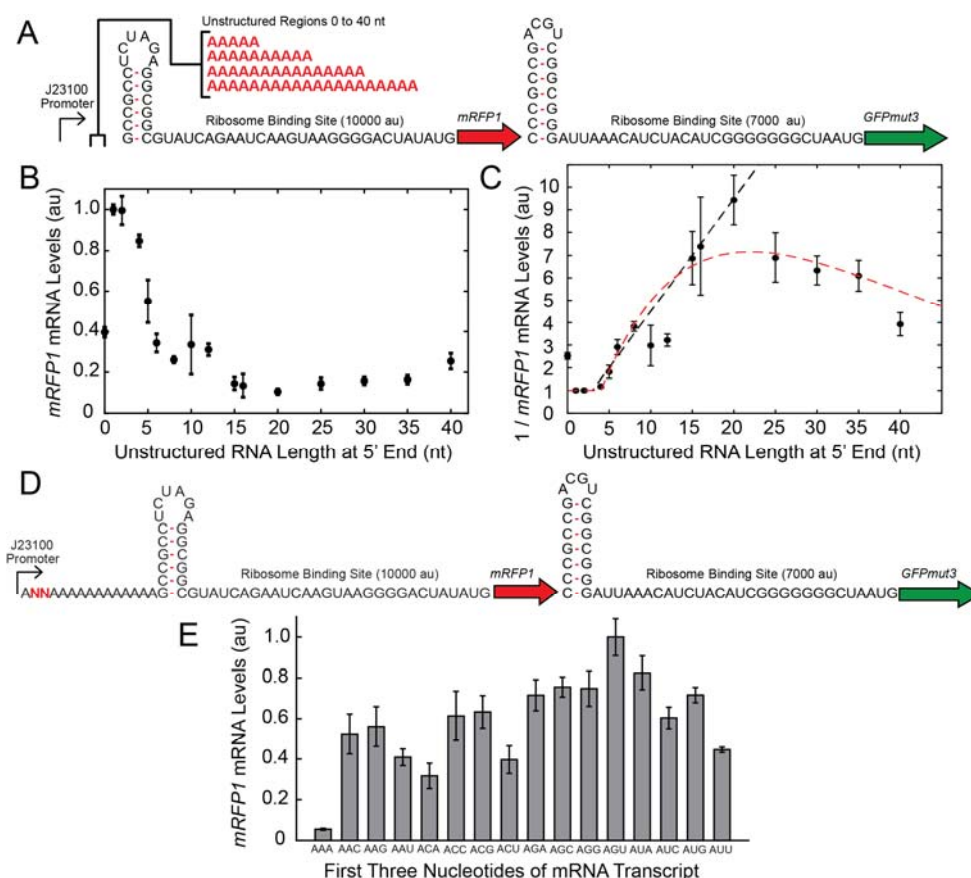
Figure 2: Sequence determinants of mRNA stability at 5' ends. (A) Schematic showing a mRFP1-sfGFP operon with systematically varied 5' UTR sequences with between 0 to 40 nucleotides of single-stranded RNA, inserted upstream of an insulating hairpin. (B) RT-qPCR measurements of *mRFP1* mRNA levels show a quantitative relationship between single-stranded RNA length and mRNA stability. Dots and error

bars are the mean and standard deviation of 2 biological and 6 technical replicates. (C) RT-qPCR measurements are replotted showing the relationship between the inverse mRNA level (proportional to decay rate) and single-stranded RNA length. Dashed black and red lines show predictions from two biophysical models. (D) Schematic showing a mRFP1-sfGFP operon with systematically varied dinucleotides at the 5' end. (E) RT-qPCR measurements of *mRFP1* RNA levels show how varied dinucleotides affected mRNA stability. Bar heights and error bars are the mean and standard deviation of 2 biological and 6 technical replicates.

RESULTS

Unstructured RNA in the 5' Untranslated Region Controls mRNA Stability

We first interrogated the sequence and structural determinants in the 5' untranslated region that affect a mRNA's stability in *E. coli*, combining systematic design, DNA assembly, and characterization using RT-qPCR and fluorescence measurements. To do this, we designed and constructed plasmid-encoded, bi-cistronic operons that utilize a constitutive sigma70 promoter (J23100), a rationally designed synthetic ribosome binding site, and two codon-optimized fluorescent protein reporters (mRFP1 and GFPmut3) (**Methods**, **Figure 2A**). The ribosome binding sites were designed by the RBS Calculator v2.1 to have moderate translation initiation rates (7000 to 10000 au). Using a sequence constraint, the RBS designs included a fast-folding, insulating mRNA hairpin ($\Delta G_{folding}$ = -12.61 kcal/mol) that prevents upstream sequence changes from changing the mRNA's translation initiation rate. Rationally designed upstream sequences were then inserted to replace the 5' untranslated region, beginning with the transcriptional start site and ending past the coding region's start codon. We then carried out long-time cultures of transformed strains, maintaining them in the exponential growth phase with periodic serial dilutions, followed by mRNA level measurements using RT-qPCR assays with TaqMan probes specific to the internal regions of reporter coding sequences (**Methods**). All sequences and measurements are located in **Supplementary Data 1**. All TaqMan probe sequences and measured efficiencies are located in **Supplementary Data 2**. The MIQE information for these experiments is located in **Supplementary Data 3**. By keeping the promoter unchanged and maintaining cells in the exponential growth phase for a long time period, changes in mRNA level can be primarily attributed to changes in mRNA stability, although we discuss potentially confounding factors.

In the first series of 16 operons, we designed and inserted 5' UTR sequences that contained between 0 to 40 polyA nucleotides, creating systematically longer single-stranded RNA (ssRNA) regions upstream of the ribosome binding site's insulating hairpin (**Figure 2A**). Using Vienna RNAfold (56), we verified that these polymeric A sequences do not fold into mRNA structures that could potentially block RNAse endonuclease activity. After characterizing these operons, we found a clear, quantitative relationship between the length of the ssRNA region and the resulting *mRFP1* mRNA level (**Figure 2B**). The highest mRNA levels occurred when the 5' UTR contained very short polyA sequences (2 or 3 As), followed by a precipitous decrease as the ssRNA region was lengthened. mRNA levels were reduced by about 2-fold when a ssRNA region was 5 nucleotides long. At its lowest, mRNA levels were reduced by 9.4-fold when 20 ssRNA nucleotides were added to the 5' UTR. However, further lengthening of the ssRNA region reversed the trend; we found that mRNA levels consistently increased when the ssRNA region was 20 to 40 nucleotides long. Surprisingly, without a polymeric A region (0 As), the

presence of an insulating hairpin at the transcriptional start site actually reduced mRNA levels by about 2.5-fold.

With further analysis, we then developed biophysical models to explain how lengthening the ssRNA region could alter mRNA levels in such a non-linear fashion. When cultures are maintained in the exponential phase of growth, mRNA levels will be inversely proportional to their decay or degradation rates. We therefore replotted our data to show how the inverse of mRNA levels are related to the lengths of the ssRNA region (**Figure 2C**). The trends in this relationship immediately suggested two types of biophysical models. In both models, RNAses E/G bind and cleave mRNA anywhere where is a minimally sized landing pad containing an unstructured ssRNA region. If the unstructured region is too short, RNAses E/G cannot bind and cleave mRNA. As the unstructured region is lengthened, the number of potential binding sites also increases, resulting in a proportional increase in RNAse activity.

In the first model, we calculate the change in a mRNA's decay rate $\Delta\delta_{mRNA}$ by only considering the length of the smallest possible binding site for RNAses E/G (N), the length of a ssRNA region (L), and an activity coefficient (C), according to the equation:

$$\Delta\delta_{mRNA} = C (L - N) \qquad\qquad \textbf{Equation 1}$$

With a minimum binding site of 2 nucleotides (N = 2 nt) and an activity coefficient of about one half (C = 0.528 1 / nt), Equation 1 can explain how mRNA decay rates depend on ssRNA regions from 1 to 20 nucleotides long (Pearson R^2 = 0.92, mean absolute error = 0.473) (**Figure 2C,** black dashed line). However, it's clear that this linear relationship is no longer true after the ssRNA region is more than 20 nucleotides long.

In the second model, we additionally take into account that untethered ssRNA molecules can dynamically form transient loop structures that limit their accessibility, and therefore may prevent RNAses from binding to them. From the wormlike chain model of polymer theory, the likelihood that ssRNA with length L will *not* form an intramolecular loop is proportional to exp($-$ L / P), where P is the persistence length of ssRNA, which is a measure of its bending stiffness. As the length of a ssRNA region exceeds the persistence length, it becomes highly likely that the ssRNA will fold into a disordered structure. We therefore incorporated this effect to calculate the change in a mRNA's decay rate according to the equation:

$$\Delta\delta_{mRNA} = C (L - N) \exp(- L / P) \qquad\qquad \textbf{Equation 2}$$

The persistence length of a ssRNA region depends on several properties, including RNA sequence, temperature, and the solution's salt composition, with previous measurements ranging from 1 to 7 nm or roughly 3 to 21 nucleotides (57-60). Here, with a minimum binding site of 2 nucleotides (N = 2 nt), an activity coefficient of about one (C = 0.975 1/nt), and a persistence length of 21 (P = 21 nt), we found that Equation 2 can explain how the mRNA's decay rate depended on ssRNA region's length from 1 to 40 nucleotides (Pearson R^2 = 0.874, mean absolute error = 0.767) (**Figure 2C**, red dashed line). Overall, the second model better predicts these mRNAs' decay rates across the entire dataset. However, when there is no single-stranded region (L = 0 nt), we do observe a discontinuity in both model predictions that could be caused by a confounding effect by another interaction, for example, a reduction in transcription rate.
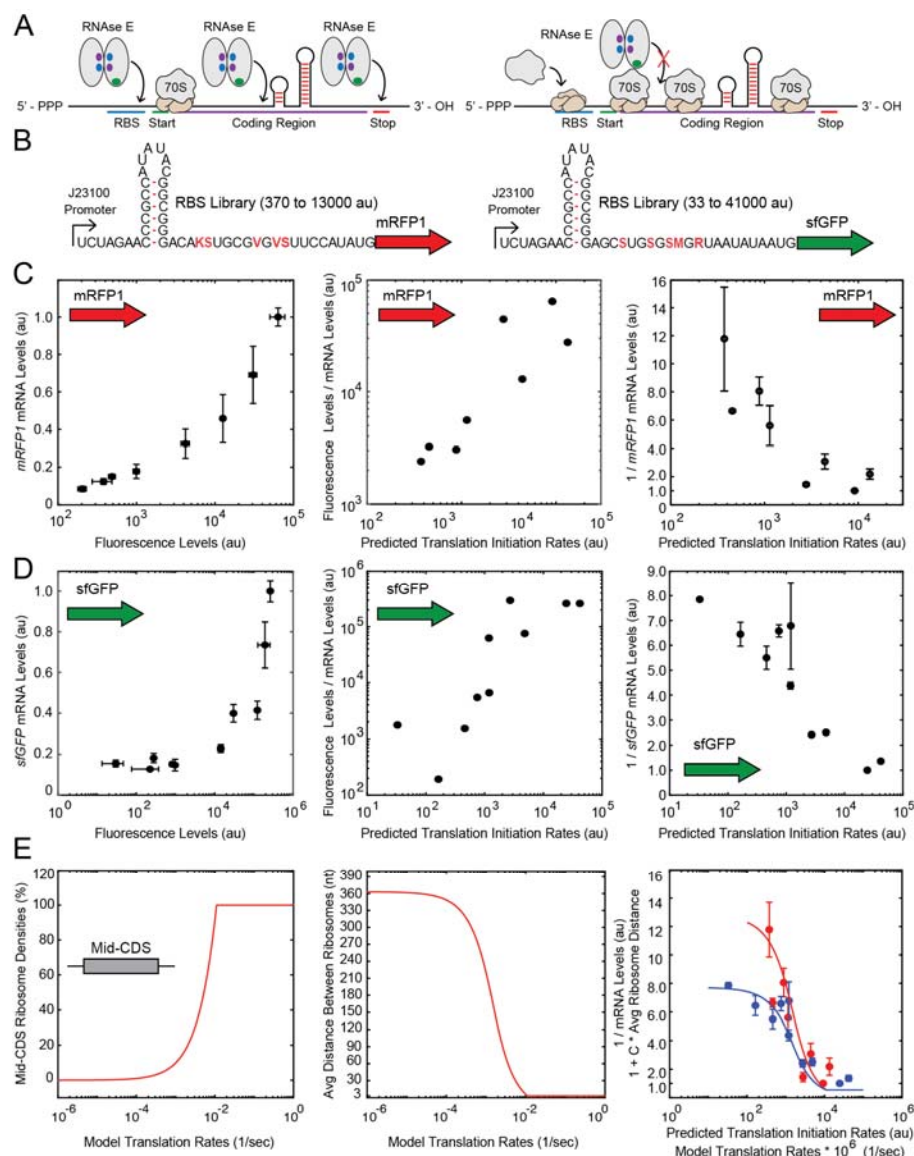
Figure 3: Translation rate and mRNA stability are coupled via ribosome protection. (A) Schematics illustrating how a mRNA's translation rate controls the accessibility of RNAse binding sites in an operon's coding regions. (B) Schematics showing the RBS library designs inside operons expressing either *mRFP1* or *sfGFP* reporter proteins. Highlighted nucleotides indicate the pinpoint mutations in RBS libraries. (C) Characterization of the *mRFP1* operons, showing (left) *mRFP1* mRNA levels compared to mRFP1 fluorescence levels, (middle) the apparent mRFP1 translaton rates (the ratio of mRFP1 fluorescence over *mRFP1* mRNA levels) compared to the *mRFP1* predicted translation initiation rates, and (right) the inverse *mRFP1* mRNA levels compared to the *mRFP1* predicted translation initiation rates. (D) Characterization of the *sfGFP* operons, showing (left) *sfGFP* mRNA levels compared to sfGFP fluorescence levels, (middle) the apparent *sfGFP* translation rates (the ratio of sfGFP fluorescence over *sfGFP* mRNA levels) compared to the *sfGFP* predicted translation initiation rates, and (right) the inverse *sfGFP* mRNA levels compared to the *sfGFP* predicted translation initiation rates. (E) Calculations from a biophysical model of ribosome protection show (left) how translation rate controls ribosome density, (middle) how translation rate controls the average distance between adjacent ribosomes, and (right) a

comparison between the average distance between ribosomes and the characterized mRNA decay rates (inverse mRNA levels) for (red) *mRFP1* and (blue) *sfGFP*. Dots and error bars showing mRNA level measurements are the mean and standard deviation from 2 biological and 6 technical replicates. Dots and error bars showing fluorescence level measurements are the mean and standard deviation from 2 to 4 biological replicates. Predicted translation initiation rates use the RBS Calculator v2.1 proportional scale.

First Three Nucleotides in the 5' Untranslated Region Control a mRNA's Stability

Previous *in vitro* measurements have shown that the first three nucleotides of a mRNA transcript have an effect on the rate of 5' end dephosphorylation by RppH or RppH-DapF complex (26), which will affect RNAse E/G's ability to initiate mRNA decay via the end-dependent pathway. Here, we carried out *in vivo* measurements to quantify how the first three nucleotides of a mRNA transcript altered its mRNA level. We designed and constructed a series of 16 bi-cistronic operons that express mRFP1 and GFPmut3 where the 5' UTR contains a 'ANN' 5' end, followed by a polymeric 15A region, an insulating RNA hairpin, and a designed ribosome binding site with a translation initiation rate of about 10000 au on the RBS Calculator v2.1 scale (**Figure 2D**). The polyA region prevents the formation of mRNA structures regardless of the first three nucleotides in the transcript, which was verified by using Vienna RNAfold (56). After maintaining cell cultures in the exponential growth phase, we carried out RNA extractions and RT-qPCR to measure *mRFP1* mRNA levels. We found that *mRFP1* mRNA levels varied by up to 18.2-fold when varying just the first three nucleotides of the transcript with the 5'-AAA-3' and 5'-AGU-3' trinucleotides have the most and least effect, respectively (**Figure 2E**). Even when excluding the most potent variant 5'-AAA-3', mRNA levels still varied by 3.2-fold, which is a potent effect for such a small sequence region. These results provide further support that the 5' end sequence affects how RppH and DapF interact with mRNA to accelerate end-dependent mRNA decay.

mRNA Translation Initiation Rate Controls mRNA Stability via Ribosome Protection

Previous studies have suggested that ribosomes can protect mRNA transcripts from RNAse activity, possibly by covering RNAse binding sites and limiting their accessibility (61,62) (**Figure 3A**). However, the relationship between a mRNA's translation rate and its decay rate has never been quantitatively studied. To determine such a quantitative relationship, we designed and constructed 18 mono-cistronic operons expressing either mRFP1 or sfGFP reporter, systematically varying their translation initiation rates by up to 1300-fold (**Figure 3B**). To do this, we inserted optimized RBS libraries into their 5' UTRs, each designed by the RBS Library Calculator to have an insulating hairpin, followed by a small number of pinpoint mutations that would greatly change the reporters' translation initiation rates (63). We then characterized the operons' expression levels, measuring both their mRNA and protein levels using RT-qPCR and flow cytometry, respectively (**Methods**).

We found that systematically increasing the mRNAs' translation initiation rates resulted in corresponding increases in both their mRNA and protein levels. For example, increasing the *mRFP1* translation initiation rate from 370 to 13000 au on the RBS Calculator v2.1 scale resulted in a 11.8-fold increase in *mRFP1* mRNA level and a 313-fold increase in mRFP1 fluorescence level (**Figure 3C**, left). Likewise, increasing the sfGFP translation initiation rate

from 33 to 41000 au resulted in a 7.85-fold increase in sfGFP mRNA level and a 8760-fold increase in sfGFP fluorescence level (**Figure 3D**, left).

Here, the higher protein levels originated from both higher translation initiation rates and higher mRNA levels. To distinguish these sources, we calculated the apparent translation rates by taking the ratio of fluorescence level over mRNA level, and compared them to the RBS Calculator's predicted translation initiation rates (**Figure 3CD**, middle). We confirmed excellent quantitative agreement between the predicted and measured translation rates for both *mRFP1* and *sfGFP* with strong statistical significance (*mRFP1*: Pearson R^2 = 0.782, p = 0.0035; *sfGFP*: Pearson R^2 = 0.706, p = 0.0023). We then replotted the same predictions and measurements to determine how the mRNAs' decay rates – the inverse of our mRNA level measurements – were affected by the mRNA's translation initiation rates (**Figure 3CD**, right). As expected, the lowest translation initiation rates yielded the highest mRNA decay rates, qualitatively consistent with a ribosome protection mechanism. However, quantitatively, our measurements could equally support a log-linear or sigmoidal relationship between translation rate and mRNA decay. To investigate further, we developed a biophysical model of ribosome protection that provides support for a sigmoidal relationship.

A Biophysical Model of Ribosome Protection Explains mRNA Stability

Consider a mRNA transcript that contains a protein coding sequence (CDS) with L_{mRNA} amino acids. The key question is to calculate how many of these nucleotides remain unprotected by elongating ribosomes. Here, we designate α as the ratio of the ribosome's translation initiation rate over its elongation rate. We also designate F as the physical footprint of each ribosome in units of trinucleotides (amino acids). Based on prior measurements, we specify that the ribosome's footprint F is about 10 trinucleotides (30 nt). According to a TASEP (totally asymmetric exclusion process) model of ribosome dynamics that includes the ribosome's footprint on the mRNA (64), designated as an "extended object", a single equation can be used to approximately calculate how the translation initiation rate controls the ribosome's average density. When translation initiation is the rate-limiting step (α is less than one), the steady-state ribosome density ρ_r is:

$$\rho_r = \min\left(\frac{F\alpha}{1+(F-1)\alpha}, 1\right)$$

Equation 3

when averaged over the middle portion of the CDS, away from the start and stop codons. The density of protected mRNA is $F\rho_r$, which includes the ribosome's footprint. By definition, the maximum number of bound ribosomes is L_{mRNA}/F with a density of one. The relationship defined by **Equation 3** illustrates how increasing a mRNA's translation initiation rate results in higher ribosome densities, up until the maximum possible value (**Figure 3E**, left).

We next consider the distances D between adjacent ribosomes, called the "headway distance" from studies of vehicular traffic. Here, the absence of a bound ribosome at a position along the mRNA is called a "hole". From the TASEP model with extended objects (65), the probability that a bound ribosome has *m* free trinucleotides in front of it is related to both the ribosome density ρ_r and hole density ρ_h according to:

$$P(D = m) = \frac{\rho_r}{\rho_r+\rho_h}\left(\frac{\rho_h}{\rho_r+\rho_h}\right)^m$$

Equation 4

where the hole density is determined by $\rho_h = 1 - F\rho_r$. The first part of **Equation 4** calculates the probability that a mRNA position is bound by a ribosome whereas the second part calculates the probability that the next m adjacent positions on the mRNA all contain a hole. We then substitute **Equation 3** and the definition of the hole density into **Equation 4**, yielding the probability distribution for the ribosomes' headway distances in terms of the ribosome density and footprint:

$$P(D = m) = \frac{\rho_r(1 - F\rho_r)^m}{(1 + \rho_r - F\rho_r)^{m+1}}$$ **Equation 5**

We then determine the ribosomes' average headway distance by calculating the first moment of the probability distribution in **Equation 5**, using:

$$\langle D \rangle = \sum_{m=1}^{m=L_{mRNA}} P(D = m)\, m$$ **Equation 6**

Altogether, for each selected translation initiation rate, we calculate the ribosome density using **Equation 3**, substitute it into **Equation 5** across a range of potential distances, and use this distribution and **Equation 6** to calculate the average distance between ribosomes (**Figure 3E**, middle). The overall relationship is a sigmoidal curve; at the highest translation initiation rates, the average distance plateaus to the smallest possible value (1 trinucleotide), whereas at the lowest translation initiation rates, the average distance plateaus at about $L_{mRNA}/2$.

Finally, according to a ribosome protection mechanism, we anticipated that the average distance between ribosomes – the amount of unprotected mRNA – is proportional to the mRNAs' decay rates. We therefore compared our inverse mRNA level measurements to the average distances between ribosomes across the 18 operons expressing *mRFP1* or *sfGFP* at varying translation initiation rates (**Figure 3E**, right) and found strong quantitative agreement with statistical significance (Pearson R^2 = 0.823, $p < 10^{-6}$), using proportionality constants C_{mRFP1} = 0.10 and C_{sfGFP} = 0.058. Here, the proportionality constant combines several factors, including the RNAse E/G activity per unprotected mRNA, the likelihood that unprotected mRNA forms protective structures, and the scale of the mRNA level measurements. The developed biophysical model of ribosome protection shows that the relationship between a mRNA's translation initiation rate and its decay rate is expected to be a sigmoidal curve, which is well supported by our measurements.

Location of RNAse Binding Sites Controls mRNA Stability in Bicistronic Operons

We next investigated how the locations of RNAse binding sites – inside either the 5' UTR, intergenic region, or 3' UTR regions – affected the mRNA levels of individual cistrons within a bicistronic operon. To do this, we designed and constructed 6 bicistronic operons expressing *mRFP1* and *GFPmut3*, inserting either a 20A ssRNA region (an RNAse E/G site) or a 25 bp RNA hairpin (a RNAse III site) into these three separate locations (**Figure 4A**). We applied Vienna RNA folding calculations to verify that the 20A region does not fold into a mRNA structure after being inserted into these locations. We performed similar calculations to verify that the 25 bp RNA hairpin does not fold into an alternative structure when inserted. We also constructed baseline operons that did not contain inserted RNAse binding sites, which we used as the reference for comparisons. Otherwise, all operons utilized the same promoters, ribosome binding sites, and coding regions. Based on prior studies of gene expression polarity (49,66-68),

we anticipated the possibility that *mRFP1* and *GFPmut3* mRNA levels could be differentially affected by these inserted RNAse binding sites.
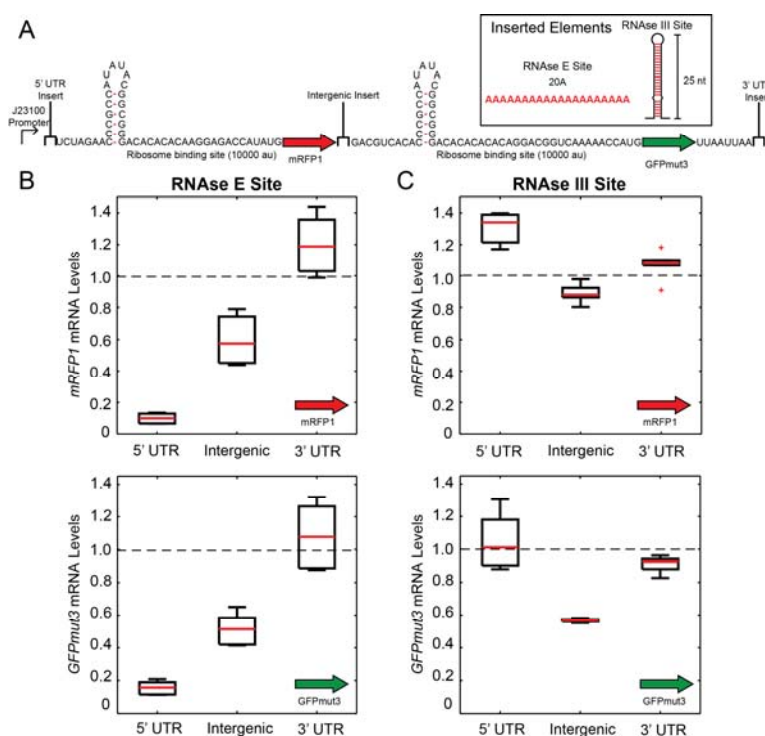

Figure 4: Positions of RNAse sites control mRNA stability. (A) Schematic illustrating the bistronic operons and the locations where RNAse E/G or RNAse III sites were inserted. (boxed) The two types of inserted RNAse binding sites are shown: either a 20A single-stranded RNA site or a 25 bp RNA hairpin structure. (B) Characterization of the operons, showing how the positions of inserted RNAse E sites affected the *mRFP1* and *sfGFP* mRNA levels. (C) Characterization of the operons, showing how the positions of inserted RNAse III sites affected the *mRFP1* and *sfGFP* mRNA levels. Dots and error bars are the mean and standard deviation from 2 biological and 6 technical replicates.

As before, we characterized these operons by extracting total RNA from cells maintained in the exponential growth phase and applying RT-qPCR assays to separately measure *mRFP1* and *GFPmut3* mRNA levels (**Methods**). Consistent with our prior results (**Figure 2B**), we found that inserting a 20A ssRNA region into the 5' UTR greatly reduced the *mRFP1* mRNA levels (by 11-fold) with a concomitant decrease in *GFPmut3* mRNA levels (by 7-fold) (**Figure 4B**). However, inserting the same 20A ssRNA region into the intergenic region had a much lower effect; *mRFP1* and *GFPmut3* mRNA levels were reduced by only 1.7-fold and 2-fold, respectively. Notably, inserting the 20A ssRNA region into the 3' UTR region had no appreciable effect on the mRNA levels of either cistron. Overall, we found that RNAse E/G sites had a potent effect on a mRNA's stability with a clear position-dependent trend; upstream sites accelerated mRNA decay more so than downstream sites. However, we did not observe any polarity effects; each RNAse E/G site similarly affected the mRNA levels of the operon's *mRFP1* and *GFPmut* cistrons.

In contrast, only one of the inserted RNAse III sites had an appreciable effect on the operons' mRNA levels, though it was accompanied by evidence of polarity (**Figure 4C**). Inserting a 25 bp

RNA hairpin into the intergenic region did not appreciably affect the *mRFP1* mRNA level (1.13-fold change), but did decrease the *GFPmut3* mRNA level by 1.8-fold. Inserting the same 25 bp RNA hairpin into either the 5' UTR or 3' UTR had no appreciable effect on either mRNA level. Overall, these results suggest that the directionality and processivity of mRNA decay depend on the RNAse that carries out the first cleavage event.
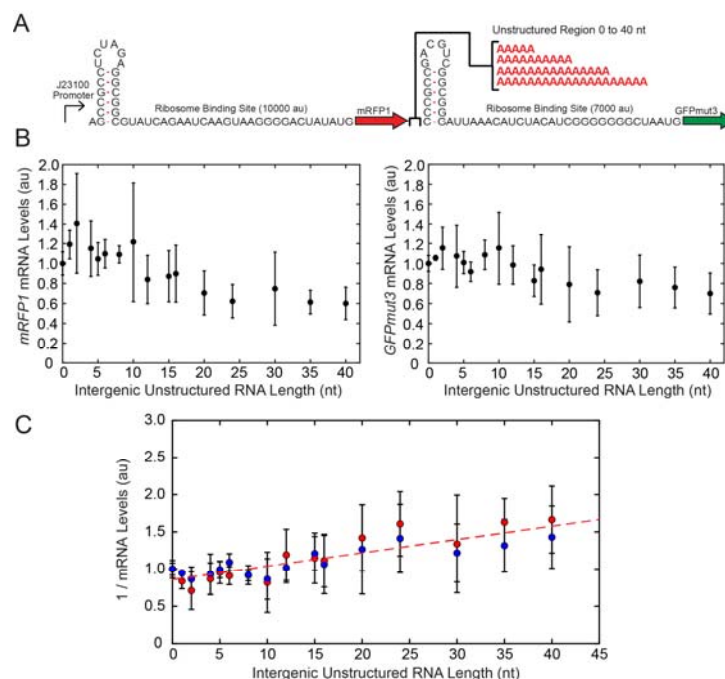

Figure 5: Sequence determinants of mRNA stability at intergenic sites. (A) Schematic illustrating bicistronic operons and the insertion of single-stranded RNA (RNAse E/G binding sites) with varying lengths into the intergenic region. (B) Characterization of the operons, showing (left) *mRFP1* mRNA levels compared to the single-stranded RNA lengths and (right) *GFPmut3* mRNA levels compared to the single-stranded RNA lengths. (C) The inverse mRNA levels for (red) *mRFP1* and (blue) *GFPmut3* are compared to the single-stranded RNA lengths alongside a (red dashed line) linear relationship with slope C = 0.018. Dots and error bars are the mean and standard deviation from 2 biological and 6 technical replicates.

Unstructured RNA in the Intergenic Region Modestly Affects mRNA Stability

Following the previous results, we carried out a systematic investigation to quantitatively determine how single-stranded RNA (ssRNA) sites inside the intergenic region control mRNA stability. We designed and constructed 16 bicistronic operons expressing *mRFP1* and *GFPmut3*, inserting polyA ssRNA sites inside their untranslated intergenic regions, ranging from 0 to 40 nt long (**Figure 5A**). We then carried out RNA extractions on cells maintained in the exponential growth phase, and measured both *mRFP1* and *GFPmut3* mRNA levels using RT-qPCR assays. Consistent with our previous results, we found that unstructured RNA inside the intergenic region has an appreciable, but modest, effect on mRNA stability, affecting both upstream and downstream cistrons similarly (**Figure 5B**). When the ssRNA length is small (0 to 2 nt long), there is small rising trend in *mRFP1* and *GFPmut3* mRNA levels, reaching up to a 1.4-fold increase at their maximum. However, further increases in ssRNA length (3 to 40 nt) reversed the trend; both mRNA levels decreased by modest amounts, reaching a minimum of 1.66-fold lower *mRFP1* and 1.43-fold lower *GFPmut3* mRNA levels.

As before, we replotted our measurements to illustrate how the inverse *mRFP1* and *GFPmut3* mRNA levels – proportional to the mRNA's decay rate – are controlled by the ssRNA length within intergenic regions (**Figure 5C**). Similar to 5' UTRs, we found that the ssRNA region must be at least 2 nt long before we observe any increase in mRNA decay. However, unlike 5' UTRs, we observed only a linear quantitative relationship between ssRNA length and mRNA decay with a small slope (activity coefficient C = 0.018). Overall, these results show that RNAse E/G is not able to strongly bind and cleave intergenic regions even with large ssRNA regions, resulting in much smaller changes in mRNA decay rates. Moreover, we do not observe appreciable amounts of differential mRNA decay across upstream and downstream cistrons (indicative of polarity) or attenuation of mRNA decay at higher ssRNA lengths (indicative of transient loop formation).
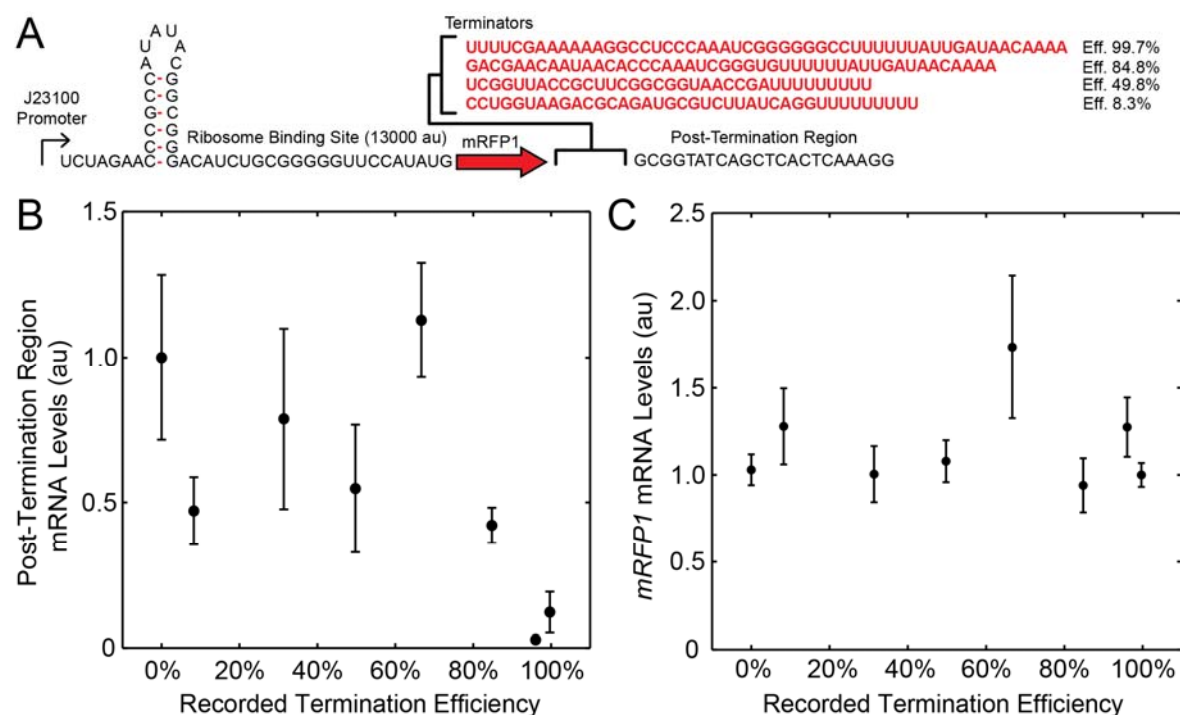
Figure 6: Transcriptional terminator efficiency does not impact mRNA stability. (A) Schematic illustrating *mRFP1* operons containing transcriptional terminators with varied efficiencies. (B) Characterization of the post-terminator mRNA levels using SYBR Green RT-qPCR assays, compared to prior measurements of the transcriptional terminators' efficiencies. (C) Characterization of the *mRFP1* mRNA levels using TaqMan RT-qPCR assays as compared to the transcriptional terminators' efficiencies. Dots and error bars are the mean and standard deviation from 2 biological and 6 technical replicates.

Transcriptional Terminator Efficiency Has No Effect on mRNA Stability

For our last dataset, we investigated how the efficiency of an operon's transcriptional terminator affected its mRNA stability. To do this, we designed and constructed 8 operons expressing mRFP1, inserting different intrinsic transcriptional terminators into their 3' UTRs (**Figure 6A**). Each terminator was selected from a toolbox of well-characterized synthetic terminators to have varying termination efficiencies from 8.3% to 99.7% (69). We also constructed a control operon that does not have a transcriptional terminator at the insertion position (0% efficiency). As an additional test, we also searched for any transcriptional terminator sequences downstream of

the insertion site. Using the FindTerm terminator finder (70), the closest transcriptional terminator candidate was at least 230 base pairs downstream of the *mRFP1* stop codon. Otherwise, all operons have identical promoters, ribosome binding sites, and protein coding sequences.

As before, we characterized the mRNA levels of these operons during exponential growth. For our first measurement, we applied SYBR Green RT-qPCR to measure the mRNA level of the region downstream of the inserted transcriptional terminators and to quantify the efficiencies of the transcriptional terminators (**Figure 6B**). We found that the levels of the post-terminator mRNA region were only greatly reduced once the previously recorded terminator efficiencies exceeded 80%. At the same time, we applied TaqMan RT-qPCR to measure the *mRFP1* mRNA levels. Surprisingly, we found no appreciable differences in *mRFP1* mRNA level across the full range of terminator efficiencies (**Figure 6C**). The *mRFP1* mRNA levels were similar whether a highly efficient terminator, a low efficiency terminator, or a non-terminating control sequence were inserted downstream of the *mRFP1* coding sequence. For the variants with little or no termination efficiency at the insertion site, it is likely that RNA polymerase continues to transcribe the mRNA for at least 230 nucleotides past the *mRFP1* stop codon, which would greatly lengthen the 3' UTR. These results suggest that the length and structure of the 3' UTR has little appreciable effect on the mRNA's decay rate, which is consistent with the prior insertion of unstructured RNA into the 3' UTR region (**Figure 4**).

DISCUSSION

We designed and characterized 82 operons to quantify the sequence and structural determinants controlling mRNA stability in *E. coli*. We introduced rationally designed sequences into the 5' UTR, intergenic, and 3' UTR regions within mono-cistronic or bi-cistronic operons to measure how they affected the operons' mRNA levels. Through this learn-by-design process, we systematically investigated how these determinants affected each mRNA decay pathway (**Figure 1**), including (i) the amount of unstructured, single-stranded RNA (ssRNA) in the 5' UTR, which controls the rate of RNAse E/G cleavage via the end-dependent pathway (**Figure 2B**); (ii) the first three nucleotides of the transcript, which controls RppH-DapF dephosphorylation at the 5' end (**Figure 2E**); (iii) the translation rate of coding sequences, which controls the amount of ribosome protection (**Figure 3**); (iv) the types and locations of RNAse binding sites, which controls decay rate and polarity (**Figure 4**); (iv) the amount of single-stranded RNA in the intergenic region, which controls the rate of RNAse E/G cleavage via the direct entry pathway **(Figure 5)**; and (v) the efficiency of the transcriptional terminator, which controls the sequence, structure, and length of the 3' UTR (**Figure 6**).

Overall, modifying the 5' UTR of a mRNA transcript had the largest effect on its stability. Inserting a 20 nucleotide single-stranded RNA (ssRNA) region into the 5' UTR led to a 9.4-fold decrease in mRNA level. Lowering the translation initiation rate of the first coding sequence by 35-fold – through redesigning its ribosome binding site – led to a 11.8-fold decrease in mRNA level. Modifying just the first three nucleotides of the transcript could lead to 18-fold change in mRNA level! In contrast, changing the intergenic region's sequence and structure had much less potent effects. Inserting a 20 nucleotide ssRNA region into the intergenic region reduced

mRNA levels by up to 2-fold, similarly affecting both upstream and downstream genes in a bi-cistronic operon (no polarity). Inserting a long RNA hairpin into the intergenic region mainly affected the downstream coding sequence's mRNA level (by 1.8-fold). Finally, we were genuinely surprised that our substantial changes to the 3' UTR sequence, structure, and length had no appreciable effect on the transcript's mRNA level. Based on our measurements, the transcriptional terminator does not alter a mRNA's decay rate and only facilitates decoupling of cistrons in adjacent operons. Overall, our results show that end-dependent decay – mediated by RNAse E/G and RppH-DapF – is the predominant pathway controlling mRNA stability. RNAse E/G likely bind internal unstructured sites with much lower affinity, requiring long stretches of unstructured RNA (e.g. entire coding regions) to achieve high cleavage rates.

Leveraging our measurements, we developed several biophysical models to explain how these sequence determinants controlled mRNA decay rates. Each model contains a small number of parameters, but can explain the observed quantitative trends. Previously, RNAse E/G were suggested to bind to specific binding motifs (e.g. RAUUW or RNWUU). Here, we propose that RNAse E/G simply bind unstructured RNA regions with a 2 nucleotide landing pad. Longer unstructured regions therefore provide a proportionally larger number of potential binding sites for RNAse E/G with an expected increase in cleavage rate (**Equation 1**). When this model is applied to the 5' UTR, it fits our mRNA decay rate measurements extremely well for when the ssRNA length is varied from 1 to 20 nucleotides (**Figure 2C**). However, when the 5' UTR ssRNA region is very long, we see an attenuation of this effect, suggesting that the region is no longer fully unstructured. We therefore turned to polymer theory, which provides well-established equations for calculating the likelihood that a polymer forms transient, non-specific structures (ie. wormlike chains), which determines the fraction of polymer that remains accessible. By incorporating polymer theory (**Equation 2**), a single model was able to explain how ssRNA length affects mRNA decay rates across the larger range of lengths from 1 to 40 nucleotides, using a persistence length that is consistent with prior measurements.

We applied the same biophysical modeling (**Equation 1**) to the intergenic region with a similarly consistent explanation. Longer ssRNA regions led to proportionally higher mRNA decay rates (**Figure 5**). However, the proportionality constant in this model was much lower, consistent with our observations that the direct entry pathway is much slower than the end-dependent pathway, when acting on the same amount of ssRNA. Notably, unlike 5' UTRs, intergenic regions with large amounts of ssRNA had the highest amounts of mRNA decay, which is expected because the formation of transient wormlike structures requires at least one freely movable (untethered) end. While the 5' end of the transcript can freely move, both ends of the intergenic region are physically constrained by the surrounding ribosomes engaged in translation elongation.

We also developed a biophysical model of ribosome protection that shows how systematically varying a coding sequence's translation rate controls its mRNA level (**Figure 3**), applying equations to calculate a mRNA transcript's ribosome density and ribosome-to-ribosome headway distance (**Equations 3-6**). The model provides a mechanistic basis for the observed sigmoidal-like relationship in our measurements with expected plateaus in mRNA level at either very low or very high translation rates, and with sharp changes in between. Here, we treated the CDS region as being equally accessible to RNAse E/G activity regardless of their RNA structure. We also assumed that all codons have the same translation elongation rates. In

return, the resulting equations can be analytically solved and remain remarkably simple to evaluate. However, the observed differences in proportionality constants for *mRFP1* and *sfGFP* could arise from these assumptions.

Collectively, our biophysical models suggest straight-forward rules for designing operons with controlled mRNA stabilities. For maximum stability, 5' UTRs should begin with 5'-AGN ends and support moderate-to-high translation initiation rates (10000 or higher on the RBS Calculator v2.1 scale), while limiting the amount of single-stranded RNA outside the ribosome binding site. For multi-cistronic operons, intergenic regions should also contain ribosome binding sites that support moderate-to-high translation initiation rates, though the amount of single-stranded RNA will not appreciably affect their stability. Our design rules here do not consider the presence of self-cleaving ribozymes, which introduce a 5'-hydroxyl group that slows down end-dependent mRNA decay (27). To achieve lower mRNA stabilities, transcripts can be designed to have 5'-AAN ends as well as single-stranded RNA regions in their 5' UTR regions, upstream of the ribosome binding site. Perhaps most surprisingly, the design of the 3' UTR does not affect the operon's expression levels and primarily controls read-through transcription into downstream operons. Altogether, our measurements and models provide the quantitative means to controlling mRNA stability in operons.

CONTRIBUTIONS

D.P.C. and H.M.S designed the study, analyzed results, and wrote the manuscript. D.P.C. conducted the experiments.

CONFLICT OF INTEREST STATEMENT

None.

REFERENCES

1. Dods, G., Gómez-Schiavon, M., El-Samad, H. and Ng, A.H. (2020) Accurate prediction of genetic circuit behavior requires multidimensional characterization of parts. *bioRxiv*.
2. Brophy, J.A. and Voigt, C.A. (2014) Principles of genetic circuit design. *Nature methods*, **11**, 508-520.
3. Taketani, M., Zhang, J., Zhang, S., Triassi, A.J., Huang, Y.-J., Griffith, L.G. and Voigt, C.A. (2020) Genetic circuit design automation for the gut resident species Bacteroides thetaiotaomicron. *Nature Biotechnology*, 1-8.
4. Aggarwal, N., Breedon, A.M.E., Davis, C.M., Hwang, I.Y. and Chang, M.W. (2020) Engineering probiotics for therapeutic applications: recent examples and translational outlook. *Current Opinion in Biotechnology*, **65**, 171-179.
5. Cheng, Y.-Y., Hirning, A.J., Josić, K.i. and Bennett, M.R. (2017) The timing of transcriptional regulation in synthetic gene circuits. *ACS synthetic biology*, **6**, 1996-2002.
6. Kurtz, C.B., Millet, Y.A., Puurunen, M.K., Perreault, M., Charbonneau, M.R., Isabella, V.M., Kotula, J.W., Antipov, E., Dagon, Y. and Denney, W.S. (2019) An engineered E. coli Nissle improves hyperammonemia and survival in mice and shows dose-dependent exposure in healthy humans. *Science Translational Medicine*, **11**.
7. Brewster, R.C., Jones, D.L. and Phillips, R. (2012) Tuning Promoter Strength through RNA Polymerase Binding Site Design in Escherichia coli. *PLOS Computational Biology*, **8**, e1002811.
8. Xue, X.-c., Liu, F. and Ou-Yang, Z.-c. (2008) A Kinetic Model of Transcription Initiation by RNA Polymerase. *Journal of Molecular Biology*, **378**, 520-529.
9. Tayara, H., Tahir, M. and Chong, K.T. (2020) Identification of prokaryotic promoters and their strength by integrating heterogeneous features. *Genomics*, **112**, 1396-1403.
10. Ireland, W.T., Beeler, S.M., Flores-Bautista, E., Belliveau, N.M., Sweredoski, M.J., Moradian, A., Kinney, J.B. and Phillips, R. (2020) Deciphering the regulatory genome of Escherichia coli, one hundred promoters at a time. *bioRxiv*, 2020.2001.2018.910323.

11.   Kong, L.-H., Xiong, Z.-Q., Song, X., Xia, Y.-J., Zhang, N. and Ai, L.-Z. (2019) Characterization of a Panel of Strong Constitutive Promoters from Streptococcus thermophilus for Fine-Tuning Gene Expression. *ACS Synthetic Biology*, **8**, 1469-1472.

12.   Agrawal, D.K., Tang, X., Westbrook, A., Marshall, R., Maxwell, C.S., Lucks, J., Noireaux, V., Beisel, C.L., Dunlop, M.J. and Franco, E. (2018) Mathematical Modeling of RNA-Based Architectures for Closed Loop Control of Gene Expression. *ACS Synthetic Biology*, **7**, 1219-1228.

13.   Westbrook, A., Tang, X., Marshall, R., Maxwell, C.S., Chappell, J., Agrawal, D.K., Dunlop, M.J., Noireaux, V., Beisel, C.L., Lucks, J. *et al.* (2019) Distinct timescales of RNA regulators enable the construction of a genetic pulse generator. *Biotechnology and Bioengineering*, **116**, 1139-1151.

14.   Espah Borujeni, A., Channarasappa, A.S. and Salis, H.M. (2013) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic acids research*, **42**, 2646-2659.

15.   Salis, H.M., Mirsky, E.A. and Voigt, C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotech*, **27**, 946-950.

16.   Espah Borujeni, A., Cetnar, D., Farasat, I., Smith, A., Lundgren, N. and Salis, H.M. (2017) Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. *Nucleic acids research*, **45**, 5437-5448.

17.   Misra, T.K. and Apirion, D. (1979) RNase E, an RNA processing enzyme from Escherichia coli. *Journal of Biological Chemistry*, **254**, 11154-11159.

18.   Robertson, H.D., Webster, R.E. and Zinder, N.D. (1968) Purification and properties of ribonuclease III from Escherichia coli. *Journal of Biological Chemistry*, **243**, 82-91.

19.   Donovan, W.P. and Kushner, S.R. (1986) Polynucleotide phosphorylase and ribonuclease II are required for cell viability and mRNA turnover in Escherichia coli K-12. *Proceedings of the National Academy of Sciences*, **83**, 120-124.

20.   Rauhut, R. and Klug, G. (1999) mRNA degradation in bacteria. *FEMS Microbiology Reviews*, **23**, 353-370.

21.   Hui, M.P., Foley, P.L. and Belasco, J.G. (2014) Messenger RNA Degradation in Bacterial Cells. *Annual Review of Genetics*, **48**, 537-559.

22.   Pfleger, B.F., Pitera, D.J., Smolke, C.D. and Keasling, J.D. (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nature Biotechnology*, **24**, 1027-1032.

23.   Hanson, G. and Coller, J. (2018) Codon optimality, bias and usage in translation and mRNA decay. *Nature reviews Molecular cell biology*, **19**, 20-30.

24.   Dar, D. and Sorek, R. (2018) Extensive reshaping of bacterial operons by programmed mRNA decay. *PLoS Genet*, **14**, e1007354.

25.   Belasco, J.G. and Biggins, C.F. (1988) Mechanisms of mRNA decay in bacteria: a perspective. *Gene*, **72**, 15-23.

26.   Vasilyev, N. and Serganov, A. (2015) Structures of RNA Complexes with the Escherichia coli RNA Pyrophosphohydrolase RppH Unveil the Basis for Specific 5′-End-dependent mRNA Decay. *Journal of Biological Chemistry*, **290**, 9487-9499.

27.   Celesnik, H., Deana, A. and Belasco, J.G. (2007) Initiation of RNA decay in Escherichia coli by 5′ pyrophosphate removal. *Molecular cell*, **27**, 79-90.

28.   Gao, A., Vasilyev, N., Serganov, A., Luciano, D.J., Richards, J., Belasco, J.G., Levenson-Palmer, R., Traaseth, N.J. and Marsiglia, W.M. (2018) Structural and kinetic insights into stimulation of RppH-dependent RNA degradation by the metabolic enzyme DapF. *Nucleic Acids Research*, **46**, 6841-6856.

29.   Gao, A., Vasilyev, N., Kaushik, A., Duan, W. and Serganov, A. (2020) Principles of RNA and nucleotide discrimination by the RNA processing enzyme RppH. *Nucleic Acids Research*, **48**, 3776-3788.

30.   Foley, P.L., Hsieh, P.-k., Luciano, D.J. and Belasco, J.G. (2015) Specificity and evolutionary conservation of the Escherichia coli RNA pyrophosphohydrolase RppH. *Journal of Biological Chemistry*, **290**, 9478-9486.

31.   Chen, H., Shiroguchi, K., Ge, H. and Xie, X.S. (2015) Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli. *Molecular Systems Biology*, **11**, 781.

32.   Makarova, O.V., Makarov, E.M., Sousa, R. and Dreyfus, M. (1995) Transcribing of Escherichia coli genes with mutant T7 RNA polymerases: stability of lacZ mRNA inversely correlates with polymerase speed. *Proceedings of the National Academy of Sciences*, **92**, 12250-12254.

33.   Garrey, S.M., Blech, M., Riffell, J.L., Hankins, J.S., Stickney, L.M., Diver, M., Hsu, Y.-H.R., Kunanithy, V. and Mackie, G.A. (2009) Substrate binding and active site residues in RNases E and G: the role of the 5'-sensor. *Journal of Biological Chemistry*.

34.   Clarke, J.E., Kime, L., Romero A., D. and McDowall, K.J. (2014) Direct entry by RNase E is a major pathway for the degradation and processing of RNA in Escherichia coli. *Nucleic Acids Research*, **42**, 11733-11751.

35.   Spickler, C., Stronge, V. and Mackie, G.A. (2001) Preferential cleavage of degradative intermediates of rpsT mRNA by the Escherichia coli RNA degradosome. *Journal of bacteriology*, **183**, 1106-1109.

36.   Aristarkhov, A., Mikulskis, A., Belasco, J.G. and Lin, E.C. (1996) Translation of the adhE transcript to produce ethanol dehydrogenase requires RNase III cleavage in Escherichia coli. *Journal of Bacteriology*, **178**, 4327-4332.

37.   Gordon, G.C., Cameron, J.C. and Pfleger, B.F. (2017) RNA Sequencing Identifies New RNase III Cleavage Sites in Escherichia coli and Reveals Increased Regulation of mRNA. *mBio*, **8**, e00128-00117.

38.   Miczak, A., Kaberdin, V.R., Wei, C.L. and Lin-Chao, S. (1996) Proteins associated with RNase E in a multicomponent ribonucleolytic complex. *Proceedings of the National Academy of Sciences*, **93**, 3865-3869.

39.   Cheng, Z.-F. and Deutscher, M.P. (2005) An Important Role for RNase R in mRNA Decay. *Molecular Cell*, **17**, 313-318.

40.   Spickler, C. and Mackie, G.A. (2000) Action of RNase II and Polynucleotide Phosphorylase against RNAs Containing Stem-Loops of Defined Structure. *Journal of Bacteriology*, **182**, 2422-2427.

41.   Deutscher, M.P. and Reuven, N.B. (1991) Enzymatic basis for hydrolytic versus phosphorolytic mRNA degradation in Escherichia coli and Bacillus subtilis. *Proceedings of the National Academy of Sciences*, **88**, 3277-3280.

42.   Jourdan, S.S. and McDowall, K.J. (2008) Sensing of 5ʹ monophosphate by Escherichia coli RNase G can significantly enhance association with RNA and stimulate the decay of functional mRNA transcripts in vivo. *Molecular Microbiology*, **67**, 102-115.

43.   Kim, K., Sim, S.-H., Jeon, C.O., Lee, Y. and Lee, K. (2011) Base substitutions at scissile bond sites are sufficient to alter RNA-binding and cleavage activity of RNase III. *FEMS Microbiology Letters*, **315**, 30-37.

44.   Ehretsmann, C.P., Carpousis, A.J. and Krisch, H.M. (1992) Specificity of Escherichia coli endoribonuclease RNase E: in vivo and in vitro analysis of mutants in a bacteriophage T4 mRNA processing site. *Genes & development*, **6**, 149-159.

45.   Luciano, D.J., Vasilyev, N., Richards, J., Serganov, A. and Belasco, J.G. (2017) A novel RNA phosphorylation state enables 5ʹ end-dependent degradation in Escherichia coli. *Molecular cell*, **67**, 44-54. e46.

46.   Del Campo, C., Bartholomäus, A., Fedyunin, I. and Ignatova, Z. (2015) Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. *PLoS Genet*, **11**, e1005613-e1005613.

47.   Deana, A. and Belasco, J.G. (2005) Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes & development*, **19**, 2526-2533.

48.   Eriksen, M., Sneppen, K., Pedersen, S. and Mitarai, N. (2017) Occlusion of the Ribosome Binding Site Connects the Translational Initiation Frequency, mRNA Stability and Premature Transcription Termination. *Frontiers in Microbiology*, **8**.

49.   Emory, S.A., Bouvet, P. and Belasco, J.G. (1992) A 5'-terminal stem-loop structure can stabilize mRNA in Escherichia coli. *Genes & Development*, **6**, 135-148.

50.   Zhang, Y., Zhang, J., Hoeflich, K.P., Ikura, M., Qing, G. and Inouye, M. (2003) MazF Cleaves Cellular mRNAs Specifically at ACA to Block Protein Synthesis in Escherichia coli. *Molecular Cell*, **12**, 913-923.

51.   Court, D.L., Gan, J., Liang, Y.-H., Shaw, G.X., Tropea, J.E., Costantino, N., Waugh, D.S. and Ji, X. (2013) RNase III: Genetics and Function; Structure and Mechanism. *Annual Review of Genetics*, **47**, 405-431.

52.   Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E. and Filipowicz, W. (2004) Single Processing Center Models for Human Dicer and Bacterial RNase III. *Cell*, **118**, 57-68.

53.   Chao, Y., Li, L., Girodat, D., Förstner, K.U., Said, N., Corcoran, C., Śmiga, M., Papenfort, K., Reinhardt, R., Wieden, H.-J. *et al.* (2017) In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. *Molecular Cell*, **65**, 39-51.

54.   Mackie, G.A. (2013) RNase E: at the interface of bacterial RNA processing and decay. *Nature reviews. Microbiology*, **11**, 45.

55.   Li, G.-W., Burkhardt, D., Gross, C. and Weissman, Jonathan S. (2014) Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *Cell*, **157**, 624-635.

56.   Lorenz, R., Bernhart, S.H., Zu Siederdissen, C.H., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, **6**, 26.

57.   Chen, H., Meisburger, S.P., Pabit, S.A., Sutton, J.L., Webb, W.W. and Pollack, L. (2012) Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proceedings of the National Academy of Sciences*, **109**, 799-804.

58.   Doose, S., Barsch, H. and Sauer, M. (2007) Polymer properties of polythymine as revealed by translational diffusion. *Biophys J*, **93**, 1224-1234.

59.   Laurence, T.A., Kong, X., Jäger, M. and Weiss, S. (2005) Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 17348-17353.

60.   Caliskan, G., Hyeon, C., Perez-Salas, U., Briber, R., Woodson, S. and Thirumalai, D. (2005) Persistence length changes dramatically as RNA folds. *Physical review letters*, **95**, 268303.

61.   Braun, F., Le Derout, J. and Régnier, P. (1998) Ribosomes inhibit an RNase E cleavage which induces the decay of the rpsO mRNA of Escherichia coli. *The EMBO Journal*, **17**, 4790-4797.

62.   Yarchuk, O., Jacques, N., Guillerez, J. and Dreyfus, M. (1992) Interdependence of translation, transcription and mRNA degradation in the lacZ gene. *Journal of Molecular Biology*, **226**, 581-596.

63.   Farasat, I., Kushwaha, M., Collens, J., Easterbrook, M., Guido, M. and Salis, H.M. (2014) Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Molecular Systems Biology*, **10**, 731.

64.   Shaw, L.B., Zia, R.K.P. and Lee, K.H. (2003) Totally asymmetric exclusion process with extended objects: A model for protein synthesis. *Physical Review E*, **68**, 021910.

65.   Sharma, A.K. and Chowdhury, D. (2011) Stochastic theory of protein synthesis and polysome: Ribosome profile on a single mRNA transcript. *Journal of Theoretical Biology*, **289**, 36-46.

66.   Luciano, D.J., Hui, M.P., Deana, A., Foley, P.L., Belasco, K.J. and Belasco, J.G. (2012) Differential Control of the Rate of 5ꞌ-End-Dependent mRNA Degradation in *Escherichia coli*. *Journal of Bacteriology*, **194**, 6233-6239.

67.   Mackie, G.A. (1998) Ribonuclease E is a 5ꞌ-end-dependent endonuclease. *Nature*, **395**, 720-724.

68.   Mackie, G.A. (2000) Stabilization of circular rpsT mRNA demonstrates the 5ꞌ-end dependence of RNase E action in vivo. *Journal of Biological Chemistry*, **275**, 25069-25072.

69.   Chen, Y.-J., Liu, P., Nielsen, A.A.K., Brophy, J.A.N., Clancy, K., Peterson, T. and Voigt, C.A. (2013) Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nature Methods*, **10**, 659.

70.   Salamov, V.S.A. and Solovyevand, A. (2011) Automatic annotation of microbial genomes and metagenomic sequences. *Metagenomics and its applications in agriculture, biomedicine and environmental studies. Hauppauge: Nova Science Publishers*, 61-78.