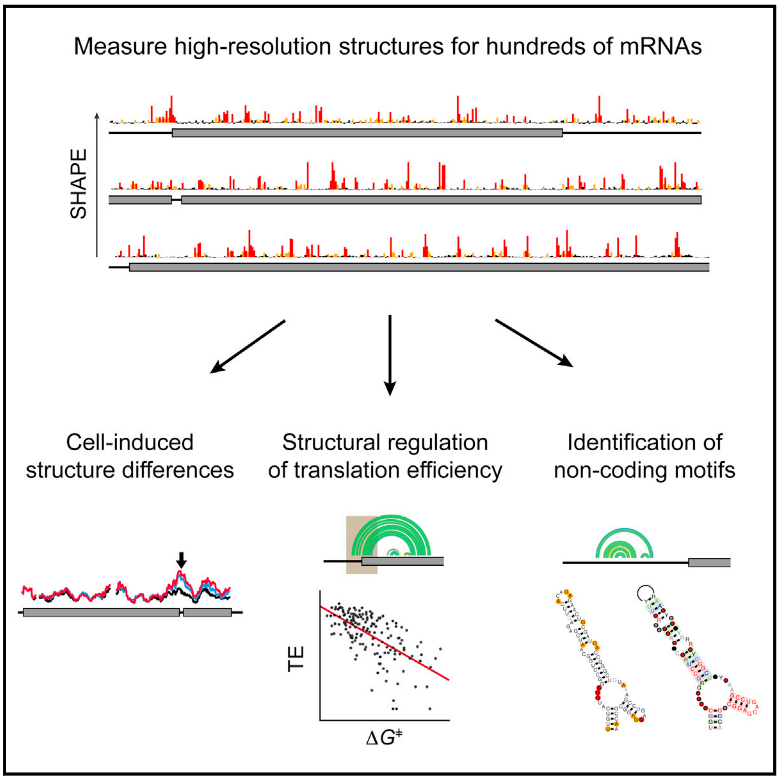
Cell

Article

Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing

Graphical Abstract


Measure high-resolution structures for hundreds of mRNAs
SHAPE
Cell-induced structure differences
Structural regulation of translation efficiency
Identification of non-coding motifs
TE
ΔG‡

Highlights

- *E. coli* mRNAs adopt highly diverse and complex structures
- Translation is the main source of mRNA structural destabilization in cells
- Translation efficiency is strongly correlated with ribosome-binding-site structure
- Conserved structured elements found in 35% of UTRs

Authors

Anthony M. Mustoe, Steven Busan, Greggory M. Rice, ..., Razvan Nutiu, Jeremy L. Baryza, Kevin M. Weeks

Correspondence

amustoe@unc.edu (A.M.M.),
weeks@unc.edu (K.M.W.)

In Brief

High-resolution probing of hundreds of genes in living *E. coli* cells reveals that bacterial mRNAs fold into highly diverse and complex structures and that these structures have widespread regulatory functions.

Mustoe et al., 2018, Cell *173*, 181–195
March 22, 2018 © 2018 Elsevier Inc.
https://doi.org/10.1016/j.cell.2018.02.034

Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing

Anthony M. Mustoe,[1,3,*] Steven Busan,[1,3] Greggory M. Rice,[1,2,3] Christine E. Hajdin,[2] Brant K. Peterson,[2] Vera M. Ruda,[2] Neil Kubica,[2] Razvan Nutiu,[2] Jeremy L. Baryza,[2] and Kevin M. Weeks[1,4,*]

[1]Department of Chemistry, University of North Carolina, Chapel Hill, NC, USA

[2]Novartis Institutes for Biomedical Research, Inc., Cambridge, MA, USA

[3]These authors contributed equally

[4]Lead Contact

*Correspondence: amustoe@unc.edu (A.M.M.), weeks@unc.edu (K.M.W.)

https://doi.org/10.1016/j.cell.2018.02.034

SUMMARY

mRNAs can fold into complex structures that regulate gene expression. Resolving such structures *de novo* has remained challenging and has limited our understanding of the prevalence and functions of mRNA structure. We use SHAPE-MaP experiments in living *E. coli* cells to derive quantitative, nucleotide-resolution structure models for 194 endogenous transcripts encompassing approximately 400 genes. Individual mRNAs have exceptionally diverse architectures, and most contain well-defined structures. Active translation destabilizes mRNA structure in cells. Nevertheless, mRNA structure remains similar between in-cell and cell-free environments, indicating broad potential for structure-mediated gene regulation. We find that the translation efficiency of endogenous genes is regulated by unfolding kinetics of structures overlapping the ribosome binding site. We discover conserved structured elements in 35% of UTRs, several of which we validate as novel protein binding motifs. RNA structure regulates every gene studied here in a meaningful way, implying that most functional structures remain to be discovered.

INTRODUCTION

Nearly all RNA molecules fold into structures that are stabilized by networks of base-pairing interactions. These structures mediate numerous functions, ranging from catalysis to ligand-responsive gene regulation (Cech and Steitz, 2014). In mRNAs, it is hypothesized that RNA structure broadly regulates gene translation efficiency (TE) (reviewed in Kozak, 2005), and numerous complex post-transcriptional regulatory structures have been identified in 5′ and 3′ UTRs (Cech and Steitz, 2014). However, efforts to understand the prevalence and role of mRNA structure-based regulatory mechanisms have been hampered by long-standing challenges in RNA structure modeling.

Recent transcriptome-wide structure-probing experiments have implied that mRNAs are frequently structured (Del Campo et al., 2015; Ding et al., 2014; Lu et al., 2016; Rouskin et al., 2014; Spitale et al., 2015; Sugimoto et al., 2015; Wan et al., 2014; Zubradt et al., 2017), but studies to date have lacked the resolution, quantitative accuracy, and comprehensive data coverage necessary to characterize structure at the level of individual mRNAs (Smola et al., 2015a; Weeks, 2015). In particular, there is no validated pathway for using dimethyl sulfate (DMS) probing data or ligation-dependent strategies to accurately model complex RNAs such as endogenous cellular mRNAs. Consequently, fundamental questions such as whether individual mRNAs adopt well-defined or dynamic structures, whether and why mRNA structure differs *in vivo* compared with *ex vivo*, and the extent to which RNA structures regulate gene expression have remained unresolved.

Reliable structure models are essential for understanding mRNA-regulatory mechanisms. A prime example concerns the role, if any, RNA structure plays in tuning gene TE—the amount of protein produced from a given mRNA transcript. TE is a precisely tuned quantity, varying over 100-fold between different genes, and is central to how cells maintain protein homeostasis (Li et al., 2014). Numerous studies have shown that RNA structural stability around the ribosome binding site (RBS) is a major determinant of TE for designed genes, primarily using reporter genes engineered to have specific compact structures in the vicinity of the translation start site (Goodman et al., 2013; Kudla et al., 2009; Salis et al., 2009). Indeed, for synthetic genes, quantitative models can predict and allow rational tuning of TE (Salis et al., 2009). However, studies of native, unmanipulated endogenous genes using poorly validated RNA structure models have observed poor correlations between TE and RBS structure (Boël et al., 2016; Guimaraes et al., 2014; Li et al., 2014; Tuller et al., 2010b). Several major studies have since proposed that TE is regulated via different mechanisms in endogenous genes (Boël et al., 2016; Burkhardt et al., 2017), but, in the absence of confident RNA structural models, it is premature to draw firm conclusions.

The ability to efficiently model accurate mRNA structures also has the potential to transform our understanding of the role of structure in mediating more complex forms of regulation.

To date, discovery of new functional non-coding motifs has been largely restricted to bioinformatics and genetics strategies. These strategies work well for identifying large, broadly conserved structures, such as riboswitches and ribozymes (Weinberg et al., 2015), but suffer from unacceptably high false positive rates when trying to identify smaller or less conserved motifs (Eddy, 2014). The prevalence of non-coding regulatory motifs genome-wide has therefore remained controversial, but it is likely that many functional motifs remain to be discovered. By comparison, starting with an accurate RNA structure model inverts the discovery problem and would potentially facilitate highly sensitive strategies for discovering novel RNA biology.

In this study, we harness recent technological advances to create the first ''no compromises'' RNA structure probing dataset on a transcriptome-wide scale. This conceptual advance allows us to dissect the mechanisms shaping in-cell RNA structure with unparalleled resolution and enables accurate structure modeling for hundreds of mRNA transcripts. These structure models, in turn, allow us to test key hypotheses regarding the prevalence and function of mRNA structure. Overall, our work establishes RNA structure as a pervasive and fundamental regulator of gene expression, likely directing the expression of every gene in *E. coli*.

RESULTS

High-Resolution Probing Reveals that mRNAs Adopt Highly Diverse Structures

We used the selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) (Siegfried et al., 2014; Smola et al., 2015b) chemical probing strategy to obtain quantitative, single-nucleotide resolution measurements of RNA structure across the *E. coli* transcriptome. SHAPE reactivities are proportional to local nucleotide flexibility and, thus, provide a direct measure of the extent of RNA structure. Using the extensively validated reagent 1-methyl-7-nitroisatoic anhydride (1M7), we probed RNA structure under three conditions: in living *E. coli* cells during mid-log growth in liquid culture; in living cells treated with the antibiotic kasugamycin, which inhibits translation initiation; and in protein- and ribosome-free extracts maintained in native-like buffers, which we refer to as cell-free (Figure 1A).

Critically for this study, we focused on studying the subset of native mRNAs in *E. coli* for which it was possible to acquire near-complete and very high-quality chemical probing data. This approach is thus distinct from prior transcriptome-scale studies, which used most or all collected data but, because of data sparseness and irregularity at the per-nucleotide level, required most chemical probing information to be averaged over many genes or averaged over large regions of an RNA. We applied an unbiased whole-transcriptome sequencing strategy that yielded high-quality structural data for 194 highly expressed transcripts, encoding approximately 400 genes, that met stringent read depth and completeness thresholds (Figure 1B; Siegfried et al., 2014). These datasets are of comparable quality as those collected in focused studies of individual RNAs (Figure 1C). 1M7 readily penetrates *E. coli* cells (McGinnis et al., 2015; Tyrrell et al., 2013; Watters et al., 2016), and we resolve precise nucleotide-resolution changes in SHAPE reactivity, reflective of protein binding in non-coding RNAs in cells (Figure S1). Reproducibility was confirmed by comparisons between biological replicates (Figure S2).

We initially characterized structural variation across different classes of RNA based on their cell-free SHAPE reactivities. Nucleotide-resolution SHAPE data immediately revealed the enormous diversity in RNA structure across *E. coli* genes (Figures 1B and 1C). This structural heterogeneity is obscured in meta-gene analyses (Figure S2), and, clearly, no individual RNA has a structure matching that of an averaged meta-gene. Non-coding RNAs (ncRNAs) and pre-tRNAs have low SHAPE reactivities (Figures 1B and 1C), consistent with these ncRNAs possessing stable, well-defined secondary and tertiary structures. By comparison, SHAPE reactivities of coding regions vary dramatically. Some genes exhibit very little stable structure, and others are structured to degrees similar to that of ncRNAs (Figures 1B and 1C). Within a given gene product category, there is again a wide diversity in mRNA structure (Figure 1C). There is no periodicity in the reactivity profiles of coding regions, indicating that, at least in *E. coli,* mRNA structure is not periodic (Figure S2). We suggest that periodicities observed in other studies may reflect sequence biases of non-MaP-based structure-probing methods and, for structures probed in cells, second-order effects of local ribosome-induced unfolding (STAR Methods). Overall, mRNA structures are diverse and largely orthogonal to gene identity and, thus, potentially able to exert heterogeneous and transcript-specific roles in regulating gene expression.

Translation Transiently Disrupts mRNA Structure in Cells

Comparisons between in-cell and cell-free datasets revealed that the cellular environment has a significant effect on mRNA structure. Specifically, coding regions are less structured (have higher SHAPE reactivities) in cells than under cell-free conditions (Figure 1D), consistent with observations from prior studies (Burkhardt et al., 2017; Ding et al., 2014; Rouskin et al., 2014; Spitale et al., 2015). We hypothesized that this structural destabilization was due to ribosome-induced mRNA unfolding during translation (Takyar et al., 2005) and, therefore, examined the relationship between in-cell SHAPE reactivity and gene TE, which is proportional to average ribosome occupancy (Li et al., 2014).

Three lines of evidence support that mRNA structural disruption observed in cells is primarily due to transient unfolding caused by active translation. First, we observe strong transcriptome-wide correlations between gene TE and in-cell SHAPE reactivity but not with cell-free SHAPE reactivity (Figure 2A). Second, in polycistronic transcripts, in-cell SHAPE reactivities increase precisely in highly translated genes, whereas genes on the same transcript with low TE have comparable in-cell and cell-free reactivities (Figure 2B). Third, compared with normal in-cell conditions, SHAPE reactivity decreases when translation is partially inhibited by the antibiotic kasugamycin, and the correlation between TE and SHAPE reactivity is sharply reduced (Figures 2A and 2B). By contrast, kasugamycin treatment has no effect on the structure of ncRNAs; any structural destabilization is constant across both in-cell conditions,

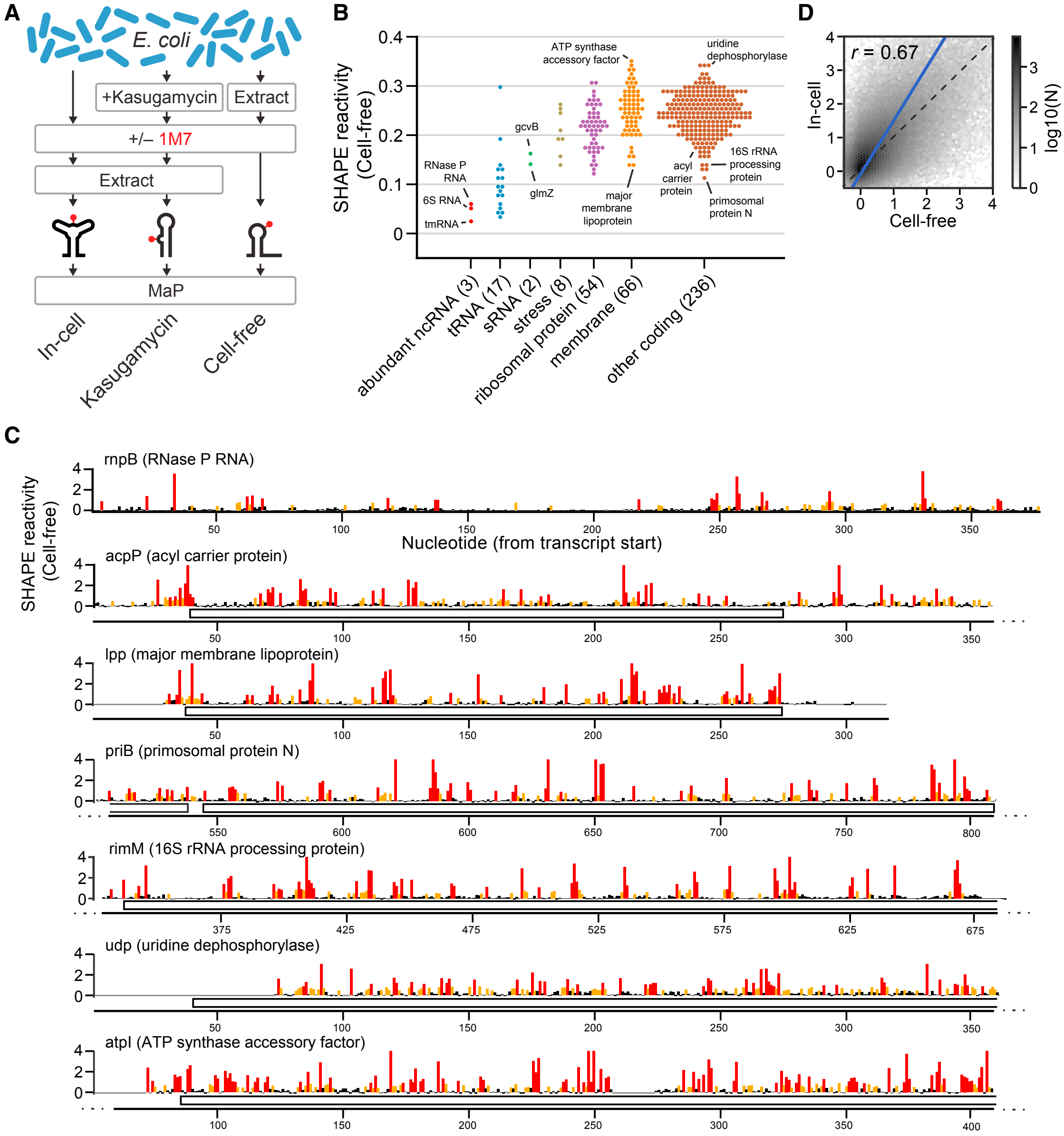


Figure 1. *E. coli* RNA Structure Overview

(A) Experimental strategy.

(B) Diversity of *E. coli* mRNA structures reflected by variation in median gene SHAPE reactivity.

(C) Nucleotide-resolution SHAPE profiles for selected genes. Genes are labeled in (B).

(D) Comparison of in-cell and cell-free SHAPE reactivities for coding regions shows that RNA structure is destabilized in cells but clearly correlated overall.

See also Figures S1 and S2.

consistent with the action expected of chaperone proteins such as Hfq (Figure 2C). Thus, although multiple cellular factors can remodel RNA structure *in vivo*, ribosome-induced unfolding is a primary cause of mRNA destabilization in cells, and this destabilization correlates with the translation level of individual genes.

Despite the destabilization caused by translation, SHAPE reactivities under in-cell, cell-free, and kasugamycin-treated conditions remain strongly correlated, suggesting that RNA structure is, on average, maintained in cells (Figures 1D and S2). A unique advantage of 1M7 SHAPE-MaP data is that they can be used to guide accurate secondary structure modeling using extensively validated strategies (Siegfried et al., 2014). Structural modeling was performed for all transcripts under each condition with sufficient SHAPE data, yielding both minimum free energy structure models and base-pairing probabilities. Consistent with the enormous diversity among SHAPE reactivity profiles, different transcripts exhibit highly

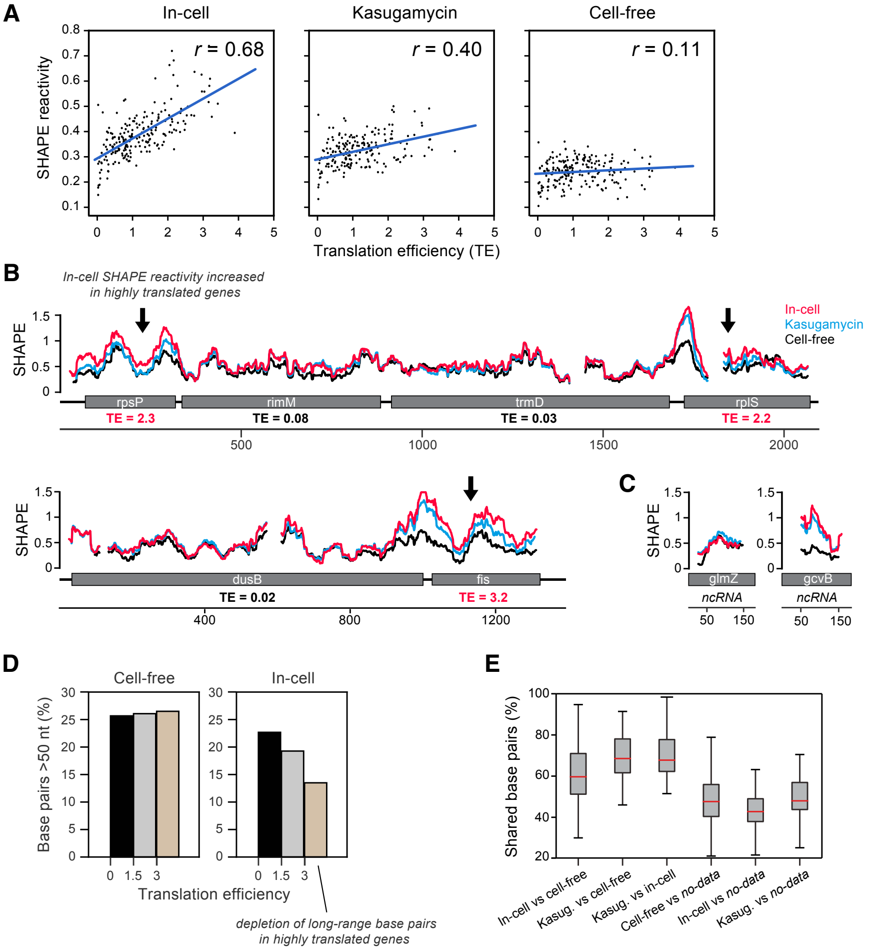


Figure 2. Translation Destabilizes Coding RNA Structure

(A) Gene median SHAPE reactivity versus translation efficiency (TE) (Li et al., 2014).

(B and C) SHAPE reactivity profiles for polycistronic mRNAs. Reactivities are shown as medians over 51-nt sliding windows (B). TE is shown beneath each gene. In-cell SHAPE reactivities increase specifically in highly translated genes. Kasugamycin treatment partially abrogates this increase in mRNAs, but (C) has no effect on noncoding RNAs glmZ and gcvB.

(D) Fraction of high-confidence (pairing probability > 98%) base pairs spanning greater than 50 nucleotides in cell-free and in-cell coding regions as a function of TE. Long-range base pairs are specifically disfavored in highly translated genes.

(E) Percentages of base pairs shared in minimum free energy RNA structure models. Boxes indicate the interquartile range (IQR), and whiskers indicate data within 1.5 × IQR of the top and bottom quartiles.

See also Figure S3.

variable degrees of structure (Figure S3). For some transcripts, 50% of nucleotides form high-probability base pairs, indicating that the mRNA adopts a well-defined global structure. For other transcripts, only ~10% of nucleotides form well-defined base pairs, indicating that the mRNA structure is highly dynamic. In-cell structure models have ~20% fewer base pairs than cell-free and kasugamycin structure models (Figure S3), consistent with translation-induced structural destabilization. Highly translated coding regions are selectively depleted of high-probability long-range base pairs in cells, implying that ribosome-induced unfolding specifically disfavors long-range pairing (Figure 2D). Nevertheless, more than 60% of minimum free energy and more than 70% of high-probability base pairs are shared between in-cell, cell-free, and kasugamycin structure models (Figures 2E and S3), and most structural differences are localized to dynamic regions (STAR Methods). By contrast, structure models predicted without SHAPE data deviate significantly from data-driven models (Figures 2E and S3).

In sum, RNA structure is destabilized in the cellular environment by active translation such that translation disfavors long-range base pairing. Nonetheless, in-cell RNA structure does not appear to undergo radical changes, leaving intact the potential for RNA structure to regulate cellular processes.

mRNA Structure Globally Tunes Gene TE

Our SHAPE-directed structure models provide an unparalleled resource for exploring hypotheses regarding the cellular functions of mRNA structure. One of the most important potential functions of mRNA structure is as a regulator of gene TE. Seminal studies of simplified model genes have shown that RNA structures that occlude the Shine-Dalgarno sequence and beginning of the coding sequence—collectively termed the RBS—impede loading of the gene into the mRNA binding channel of the 30S ribosomal subunit and, therefore, reduce TE (de Smit and van Duin, 1990; Goodman et al., 2013; Kudla et al., 2009; Salis et al., 2009). In contrast, studies of authentic native genes have reported that RBS structure is only weakly correlated with TE (Boël et al., 2016; Guimaraes et al., 2014; Li et al., 2014; Tuller et al., 2010b). More recently, it has been suggested that average structure across the entire coding sequence (CDS), rather than RBS structure, is the key determinant of TE for endogenous native genes (Burkhardt et al., 2017). Importantly, however, all of these studies relied on naive prediction or unvalidated RNA structure-modeling strategies.

Understanding TE in endogenous polycistronic transcripts is complicated by the phenomena of translational coupling, where translation of a downstream gene is dependent on and coupled to translation of upstream genes (Kozak, 2005). Because the mechanism of translation initiation likely differs in translationally coupled genes, we excluded possible translationally coupled genes from our analysis (Figure 3A). Genes were required to be either the first gene on the transcript or have more than a 2-fold different TE than the immediate upstream gene. The distinct role of RNA structure in translational coupling is discussed later.

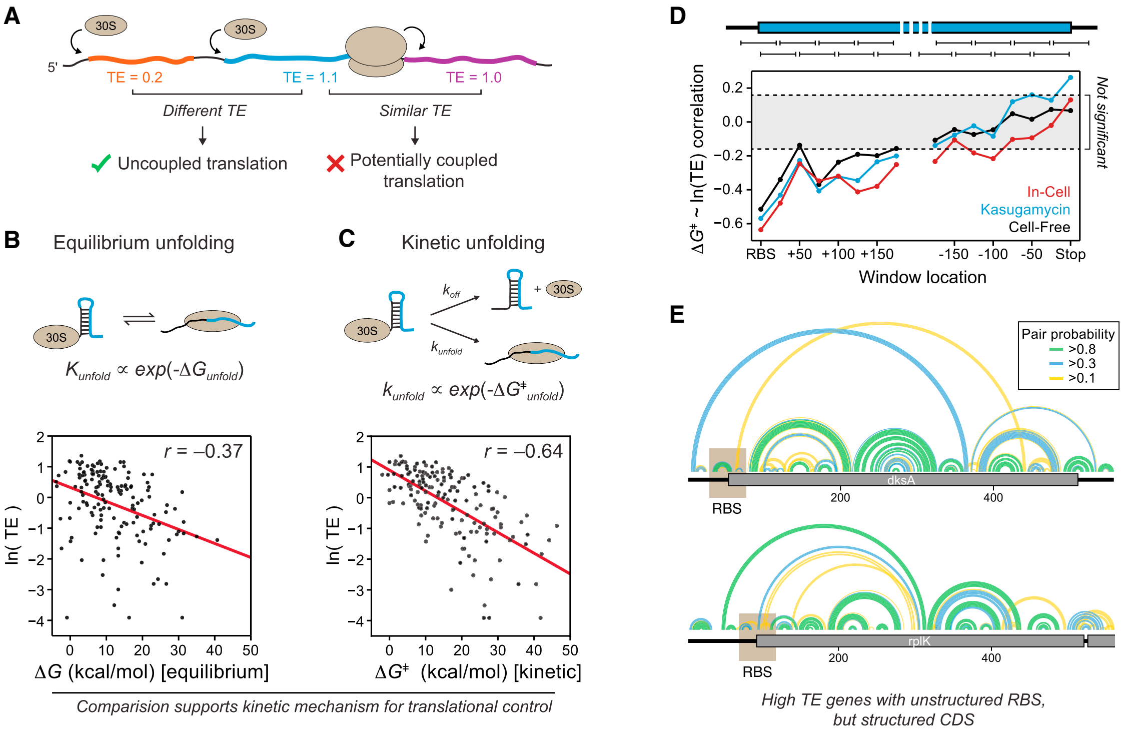


Figure 3. RBS Structure Regulates Translation

(A) Identification of potential translationally coupled genes, which were excluded from TE analysis.

(B) Equilibrium unfolding model for mRNA loading into the 30S mRNA channel (top) and correlation between TE and RBS ΔG_{unfold} for translationally uncoupled genes (bottom). n = 157.

(C) Kinetic unfolding model for mRNA loading into the 30S subunit (top) and correlation between TE and RBS $\Delta G^{‡}_{unfold}$ for translationally uncoupled genes (bottom).

(D) Correlation between gene TE and $\Delta G^{‡}_{unfold}$ computed for different coding sequence windows. The indicated significance cutoff corresponds to $p \approx 0.05$ (two-sided Wald test; precise cutoff varies between datasets).

(E) Example of two high TE genes with structured CDSs in-cell. Base pairs are shown as arcs colored by pairing probability. Both genes have unstructured RBSs and, hence, are predicted to have high TE by the RBS kinetic unfolding model (C) but not by models considering CDS structure.

See also Figure S4.

We used our SHAPE-directed structure models to examine two alternative biophysical mechanisms through which RBS structure may regulate mRNA loading onto the 30S subunit during translation initiation. If loading is an equilibrium process, then TE should vary with the equilibrium free energy of unfolding the RBS structure (ΔG_{unfold}) (Figure 3B; Salis et al., 2009). Alternatively, ribosome loading could be a non-equilibrium process, depending on a kinetic competition between RBS unfolding versus dissociation of the mRNA from the 30S subunit (de Smit and van Duin, 2003). In this kinetic scenario, TE should vary with the non-equilibrium free energy of unfolding, representative of the unfolding transition state, $\Delta G^{‡}_{unfold}$ (Figure 3C). Both ΔG_{unfold} and $\Delta G^{‡}_{unfold}$ can be computationally approximated but will only be accurate if the underlying RNA structure model is also accurate. Analysis of our SHAPE-directed models revealed that TE is weakly correlated with the equilibrium ΔG_{unfold} (r = –0.37) but strongly anticorrelated with $\Delta G^{‡}_{unfold}$ (r = –0.64), indicating that TE is strongly dependent on RBS unfolding kinetics (Figures 3B and 3C and S4; STAR Methods). Significantly, this r = –0.64 correlation between RBS structure and TE is comparable with that observed in prior studies of simplified engineered genes (Goodman et al., 2013; Kudla et al., 2009; Salis et al., 2009), suggesting that native endogenous genes regulate TE via similar mechanisms. (Note that prior studies have not attempted to resolve kinetic versus equilibrium mechanisms; Discussion.) This strong correlation is not inherent to our gene set. When we repeated our analysis using structures predicted without SHAPE data, we observed only a weak correlation between $\Delta G^{‡}_{unfold}$ and TE (r = –0.33; Figure S4), exactly consistent with prior studies of endogenous genes (Boël et al., 2016; Li et al., 2014). Thus, good structural models, as obtained by SHAPE-directed modeling, are essential for understanding the relationship between RNA structure and gene expression in native mRNAs and, in this case, inform a new understanding of regulation of native genes in *E. coli*.

It has also been proposed that RNA structures in the CDS can affect TE, potentially by modulating the rate of translation elongation (Burkhardt et al., 2017). We therefore examined the relationship between TE and $\Delta G^{‡}_{unfold}$ for windows downstream of the RBS (Figure 3D). $\Delta G^{‡}_{unfold}$ is weakly correlated with gene TE over the first 150 nucleotides of the CDS (r ≈–0.3; Figure 3D), suggesting that stable structures at the

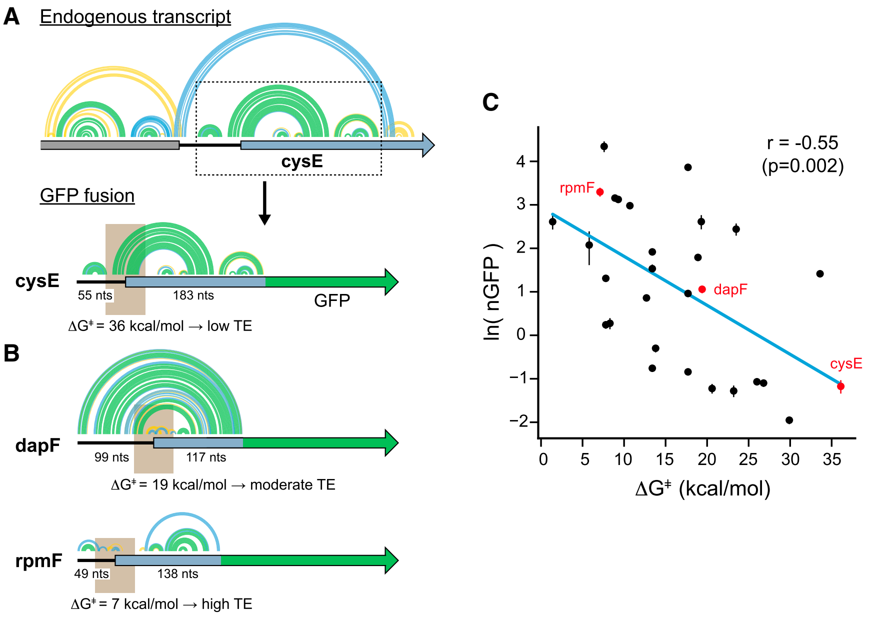


Figure 4. Reporter Gene Validation of the RBS Kinetic Unfolding Model

(A) Example parent endogenous transcript and fusion to GFP. Lengths of the fused non-coding and CDS segments are indicated. In-cell structures are shown as pairing probability arcs as in Figure 3. The RBS is highlighted in brown, with the computed $\Delta G^{\ddagger}_{unfold}$ shown underneath.

(B) Example fusions for endogenous genes predicted to have moderate and low $\Delta G^{\ddagger}_{unfold}$. Note that, despite being embedded in a larger hairpin structure, the *dapF* RBS is located in a relatively unstructured loop with moderate $\Delta G^{\ddagger}_{unfold}$ and, hence, is predicted to have moderate TE by the kinetic unfolding model.

(C) Fusion genes recapitulate the predicted trend between expression and RBS $\Delta G^{\ddagger}_{unfold}$. Protein expression was measured as GFP fluorescence normalized to a red fluorescent protein (RFP) reference encoded on the same plasmid (nGFP). Genes shown in (A) and (B) are highlighted in red. Data represent the mean ± SD from three replicates. n = 29. The p value was computed by two-sided Wald test.

5′ CDS can reduce TE, and consistent with this region playing an outsized role in determining the rate of translation elongation (Tuller et al., 2010a). However, the correlation is much weaker than that observed between RBS structure and TE. In addition, there is no correlation between TE and $\Delta G^{\ddagger}_{unfold}$ past this initial 5′ region (Figure 3D). Comparable results were observed for the equilibrium ΔG_{unfold} of CDS structure. Although translation destabilizes CDS structure, highly translated genes can be highly structured, and we identified many highly translated genes with stable, well-defined CDS structures (Figure 3E). Thus, our analysis indicates that, for the genes in this study, RNA structure primarily affects TE at the stage of translation initiation at the RBS, with TE relatively unaffected by downstream CDS structure.

To directly validate the kinetic RBS unfolding model of endogenous TE, we constructed translational fusions between endogenous genes and a GFP reporter (Figure 4). To preserve structures observed in our SHAPE-directed models, we included both the endogenous RBS and flanking regions encompassing self-contained structural elements upstream and downstream of the endogenous start codon. The TE of each fusion was then assessed as the normalized GFP fluorescence measured by flow cytometry. Critically, GFP expression was strongly anticorrelated with the expected $\Delta G^{\ddagger}_{unfold}$ of the RBS ($r = -0.55$; Figure 4C), supporting the fundamental importance of RBS structure in regulating TE. Consistent with the importance of the kinetic unfolding mechanism, GFP expression was less correlated with (equilibrium) ΔG_{unfold} ($r = -0.48$). Thus, even though native endogenous sequences are structurally complex and highly heterogeneous relative to each other, with accurate secondary structure models, it is possible to detect a strong relationship between RBS structure and TE, and this relationship is conserved across both native endogenous genes and heterologous reporter systems.

mRNA Structure Mediates Translational Coupling

Genes in polycistronic transcripts are often translationally coupled, meaning that translation of a downstream gene is modulated by translation of the preceding gene. Studies of several model transcripts have indicated that RNA structures can mediate translational coupling by acting as conformational switches that mask the RBS until unfolded by upstream ribosomes (Figure 5A; Kozak, 2005). Indeed, analysis of the "potentially translationally coupled" genes excluded from our analyses above revealed a much weaker relationship between RBS $\Delta G^{\ddagger}_{unfold}$ and TE ($r = -0.37$; data not shown), supporting that translationally coupled genes are regulated by different mechanisms. We therefore used our structure models to investigate the relevance of a structural switching mechanism transcriptome-wide.

We were immediately able to identify a potential broad role for RNA structure in mediating translational coupling. When adjacent genes have similar TEs, the RBS of the downstream gene tends to be base-paired to the coding sequence of the upstream gene (Figure 5B). Such "gene-linking" structures will be unfolded by movement of the ribosome during translation of the upstream gene, conditionally unmasking the downstream RBS (Figure 5A). In comparison, adjacent genes with different TEs tend to have self-contained structures with few gene-linking pairs, and, hence, the structural accessibility of the RBS should be relatively unperturbed by upstream translation (Figure 5C). Performing this analysis transcriptome-wide, we find that adjacent genes with many linking base pairs are significantly more likely to have similar TEs than those with few linking pairs ($p = 9 \times 10^{-5}$; Figure 5D). Thus, structural coupling between adjacent genes is a specific indicator of similar TEs, consistent with RNA structure mediating translational coupling. By comparison, we found that short intergenic distance is not a significant predictor of genes having similar TEs, even though intergenic distance is

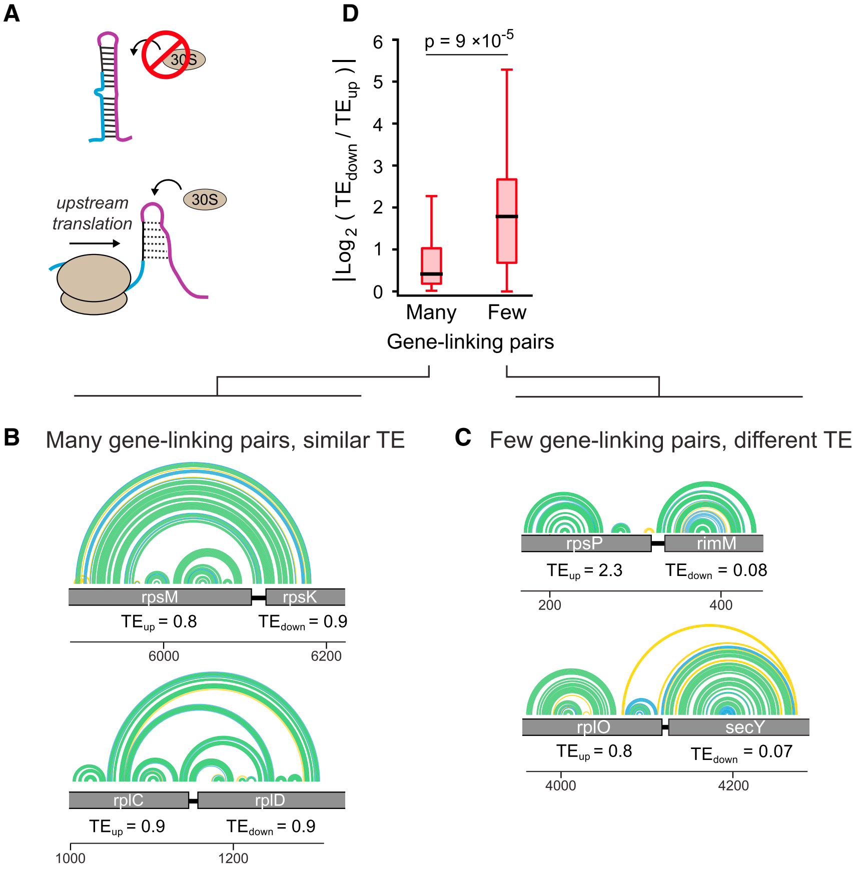


Figure 5. RNA Structure Mediates Translational Coupling

(A) Model of structure-mediated translational coupling in which upstream translation unfolds otherwise inhibitory RNA structures.

(B and C) Representative genes possessing many (B) or few (C) gene-linking base pairs. In-cell structures are shown as pairing probability arcs as in Figure 3. TE is shown beneath each gene.

(D) In-cell transcriptome-wide analysis reveals that having many gene-linking base pairs is a significant predictor that adjacent genes will have similar TEs. Gene pairs were classified as having few versus many linking pairs when they were in top and bottom quintiles of all gene pairs, respectively. The p value was computed by two-tailed Mann-Whitney *U* test. Boxes indicate the IQR, and whiskers indicate data within 1.5 × IQR of the top and bottom quartiles.

See also Figure S5.

typically thought to be a hallmark of translational coupling ($p = 0.1$; Figure S5). Indeed, we observe multiple cases where structure appears to mediate translational coupling of genes separated by more than 30 nucleotides (Figure S5). To further validate that gene-linking structures mediate translational coupling, we identified the top quintile of genes with the most gene-linking base pairs. Strikingly, 24% (8 of 33) of these most-linked genes, identified from RNA structure data alone, are known to be translationally coupled. RNA structure has been specifically shown to mediate translational coupling of *rplT* (Lesage et al., 1992), whereas *rpsK* and *rplD* (Figure 5B) and *rpsD*, *rplF*, *rpmD*, *rplW*, and *thrB* have been shown to be translationally coupled but via unknown mechanisms (Mattheakis and Nomura, 1988; Thomas et al., 1987; Yates and Nomura, 1980). We again note that high-quality structural data are essential—the relationship between structural coupling and TE is lost for structure predictions made in the absence of SHAPE data ($p = 0.09$). Combined, our data show that RNA structure-based switches comprised of gene-linking base pairs frequently and selectively couple translation of adjacent genes in *E. coli*.

Discovery and Validation of Novel RNA-Regulatory Motifs

Prior work has shown that experimentally supported RNA structure models can be used to identify novel RNA-regulatory elements *de novo* based on the fact that regulatory elements often have particularly well-determined structures (Mauger et al., 2015; Siegfried et al., 2014). We therefore searched for motifs in UTRs and intergenic regions (IGRs) with uncommonly stable (low SHAPE reactivity) and well-defined (low entropy) secondary structures (Figure 6A). Significantly, this unbiased low-SHAPE/low-entropy search returned 9 of 13 (69%) of the known functional RNA motifs covered by our SHAPE data. The majority of these known motifs are ribosomal protein autoregulatory elements (RAREs) located upstream of ribosomal protein genes. RAREs function by binding excess ribosomal protein to inhibit translation initiation, creating a feedback loop that controls the ratio of protein to rRNA. Interestingly, our in-cell SHAPE data reveal that many of these RAREs are only partially formed or adopt alternative structures in the absence of bound protein, implying that RNA dynamics are important for their regulatory function (Figure S6; Table S1). Critically, the high sensitivity of the low-SHAPE/low-entropy strategy for finding known elements strongly supports that structural data can be used to identify novel functional elements *de novo*.

Overall, we identified 58 low-SHAPE/low-entropy structures located in 51 (35%) of the 147 searched UTRs and IGRs. 49 of these motifs are uncharacterized and represent compelling novel regulatory motif candidates. We substantiated the potential functions of these motifs by three approaches. First, for non-CDS-overlapping motifs, phylogenetic analysis revealed that 82% are evolutionarily conserved, with many conserved in 100% of enterobacterial species (Figure 6B; Table S1; see also Figure S7). Second, literature searches readily revealed that 23 of these uncharacterized structures (47%) are located in genomic regions with either strong or moderate evidence of biochemical function (Figure 6B; Table S1). Finally, for three candidate RAREs newly identified as low-SHAPE/low-entropy

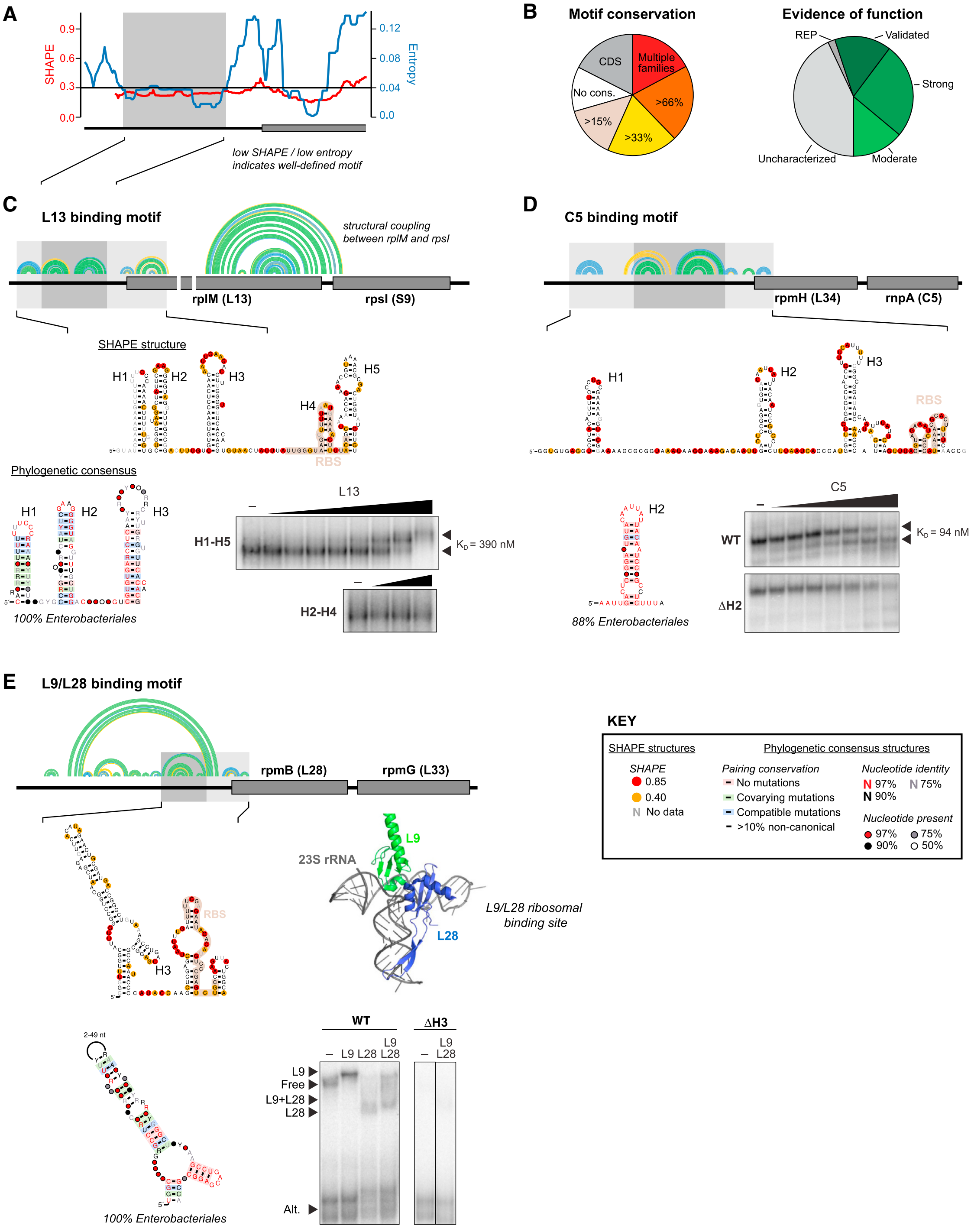


Figure 6. Structure-Based Discovery of Novel RNA-Regulatory Motifs

(A) Candidate motifs are identified in non-coding regions based on the ability to form stable, well-defined structures, as defined by low SHAPE reactivity and low structural entropy. The low-SHAPE/low-entropy region is emphasized with gray shading.

(B) Conservation of identified low-SHAPE/entropy structures in enterobacteria and evidence of function from prior literature (n = 58; Table S1).

(C–E) Identification and validation of the L13-binding motif, C5-binding motif, and L9/L28-binding motif. For each motif, the defining low-SHAPE/entropy region is highlighted in dark gray on the transcript model, with expansions to incorporate surrounding sequences shown in light gray (top). The two secondary structures shown illustrate SHAPE probing data superimposed on the structure of the 5′ UTR construct used for validation and the consensus structure labeled by percent conservation in enterobacteria. Gels show electrophoretic mobility shift assays for the designated protein-RNA interactions. In (E), the structure of the 23S rRNA

(legend continued on next page)

motifs, we validated functional RNA-protein interactions by electrophoretic mobility shift assays. We discuss these new RAREs below and provide detailed discussions of all 58 motifs in Table S1.

Our search identified a highly conserved multi-hairpin structure in the 5′ UTR of the *rplM-rpsI* transcript that encodes ribosomal proteins L13 and S9 (Figure 6C). We hypothesized that this structure constituted a novel RARE, and, indeed, a contemporaneous study found that L13 translationally represses the *rplM-rpsI* operon *in vivo* (Aseev et al., 2016). No RNA structure or mechanistic information have been reported for the putative L13-binding motif. The 5′ UTR and CDS form five well-defined hairpins under cell-free conditions; however, in cells, the H1 and H2 hairpins are moderately destabilized, and the H4 hairpin, which sequesters the start codon, is completely destabilized (data not shown). The L13 protein specifically bound to an RNA containing helices H1–H5 (Figure 6C; K_D = 390 ± 60 nM), but no binding was observed to a truncated construct containing H2–H4 (Figure 6C). Thus, L13 binds RNA containing the H1–H5 hairpins and likely inhibits *rplM* translation by stabilizing H4 and occluding access to the RBS. L13 has also been shown to negatively regulate translation of the downstream *rpsI* gene (Aseev et al., 2016). Our structure models revealed that *rpsI* is structurally linked to *rplM* (Figure 6C), indicating that co-regulation of these two genes is likely achieved via RNA structure-mediated translational coupling.

Another well-defined structure occurs in the 5′ UTR of the *rpmH-rnpA* transcript, which encodes ribosomal protein L34 and protein C5, the protein component of RNase P (Figure 6D). Helix H2 is highly conserved upstream of the *rpmH-rnpA* operon in enterobacteria, and the conserved juxtaposition of these two genes suggests that the regulatory circuits governing RNase P and ribosome biosynthesis are co-regulated. C5 binds tightly to the *rpmH-rnpA* 5′ UTR (K_D = 94 ± 9 nM) but not to a mutant lacking the H2 hairpin (Figure 6D). The increased electrophoretic mobility of the C5-bound UTR is consistent with protein binding inducing a global conformational change in the UTR structure (Ryder et al., 2008). Intriguingly, the H2 hairpin is similar to C5-binding hairpins identified by *in vitro* selection (Lee et al., 2002). Because the L34 coding sequence lies between the 5′ UTR and the coding sequence for C5, binding of C5 likely regulates the expression of both L34 and C5, with function at either the transcriptional or translational stage. To our knowledge, this is the first example of a "moonlighting" regulatory function for C5.

Finally, we identified a well-defined motif in the 5′ UTR of the *rpmB-rpmG* operon, encoding ribosomal proteins L28 and L33. Remarkably, this highly conserved three-helix junction motif shows strong structural similarity to the 23S rRNA binding sites for both L28 and ribosomal protein L9, the latter of which is encoded on a separate operon (Figure 6E). Prior studies failed to observe autoregulation of the *rpmB-rpmG* operon by L28 or L33, but the potential involvement of L9 was not explored (Aseev et al., 2016; Maguire and Wild, 1997). The *rpmB-rpmG* 5′ UTR folds into several conformations at high salt concentrations, as visualized by non-denaturing gel electrophoresis, one of which has exceptionally slow mobility suggestive of a defined tertiary structure (Figures 6E and S8). Strikingly, L9 specifically binds this low-mobility conformation (K_D ≈ 300 nM), and L9 and L28 can jointly bind the slow conformation (Figures 6E and S8). L28 and L33 also bind independently to the UTR without discriminating between the low- and high-mobility states. L9 and L33 binding are mutually exclusive, with L9 competing off L33 (Figure S8). Interactions are specific to the native UTR because deletion of helix H3 eliminates the low-mobility conformation and, consequently, L9 and L28 binding (Figure 6E). This motif, identified *de novo* by structure-informed discovery, reveals remarkable complexity and likely constitutes a novel RARE that integrates regulation of L9, L28, and L33 across multiple operons.

In sum, phylogenetic analysis, prior functional genetics studies, and our biochemical validation support clear functional roles for many of the novel RNA motifs identified by our study (Figure 6). With limited exception, the motifs identified here have remained structurally uncharacterized, and 31% of the motifs derive from fully novel loci not even suggested by large-scale bioinformatic predictions (Table S1). Thus, our analysis indicates that, with high-quality probing data, it is possible to discover novel RNA regulatory motifs *de novo* based on RNA structure information alone.

DISCUSSION

High-throughput structure probing experiments have the potential to transform our understanding of the diverse cellular functions of RNA structure. Many studies to date have emphasized rapid and large-scale data acquisition, with less emphasis placed on the quality or completeness of data or on the quality of the resulting structure models. Such strategies place fundamental limits on the ability to resolve individual RNA structures, which is essential for understanding biological mechanisms. In the present work, we took an alternative approach by performing extensive structure-probing experiments and then curating these data to focus on transcripts for which we could obtain nearly complete, quantitative, and nucleotide-resolution profiles (Figure 1). For the roughly 400 genes examined here, our structure probing data are comparable in quality to prior highly focused studies of individual RNAs. The completeness and quality of these SHAPE data make it possible to derive realistic structure models for individual RNAs, for individual motifs within these RNAs, and for per-nucleotide structure changes within individual motifs. Ultimately, we were able to discover and validate multiple new mechanisms by which RNA structure governs gene expression in *E. coli* (Figure 7).

The most fundamental result of our study is that individual mRNAs have highly idiosyncratic architectures; in essence, each mRNA has its own distinctive structural "personality." Previous studies have presented evidence that mRNAs are

binding site for ribosomal proteins L9 and L28 is also shown (PDB: 4YBB). In (C), L13 concentrations varied from 22 to 800 nM for the H1–H5 construct and 288 to 800 nM for the H2–H4 construct. In (D), C5 varied from 10 to 240 nM. In (E), L9 and L28 concentrations were 500 nM. –, no protein. Note that CDSs in transcript models are not drawn to scale.

See also Figure S8 and Table S1.

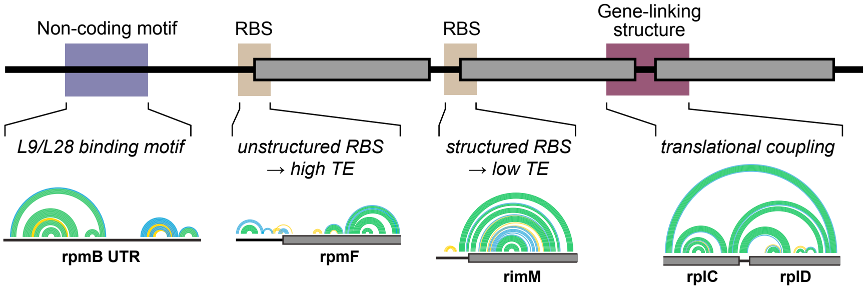


Figure 7. Mechanisms Identified in This Study through which RNA Structure Regulates Gene Expression

The function of identified novel non-coding motifs is supported by direct binding studies, evolutionary conservation, and literature cross-references. The function of RBS structure in regulating gene TE is supported by transcriptome-wide analysis and reporter gene assays. The role of gene-linking structures in mediating translational coupling is supported by transcriptome-wide analysis and literature cross-references.

frequently structured in cells but were unable to resolve this functionally important variability or distinguish the extent to which RNA structure differs between in-cell and cell-free environments (Del Campo et al., 2015; Ding et al., 2014; Lu et al., 2016; Ramani et al., 2015; Rouskin et al., 2014; Spitale et al., 2015; Wan et al., 2014; Zubradt et al., 2017). Comparisons between cell-free, in-cell, and kasugamycin-treated SHAPE datasets reveal that translation destabilizes RNA structure in highly translated genes and reduces long-range base pairing in these genes (Figure 2). Importantly, however, RNA structure is largely conserved in cells, leaving intact the potential for sequence-encoded structures to mediate gene regulation.

Significantly, our high-quality structural models allow us to address long-standing controversies regarding how translation is regulated in native endogenous genes. Studies of simplified engineered genes have shown that TE is strongly related to RBS structure (Goodman et al., 2013; Kudla et al., 2009; Salis et al., 2009), but studies of native genes have failed to recapitulate this relationship (Boël et al., 2016; Guimaraes et al., 2014; Li et al., 2014; Tuller et al., 2010b). Thus, it has remained unclear whether endogenous genes are regulated by alternative, still unknown mechanisms or, rather, that the role of RBS structure has been obscured by inaccuracies when modeling structures of native genes. Our work strongly supports the latter conclusion: TE is regulated by RBS structure in similar ways for both engineered and endogenous genes, but endogenous genes have highly diverse and much more complex structures. We explicitly validate this commonality by transplanting idiosyncratic endogenous RBS sequences in front of exogenous GFP reporters and recover a strong relationship between RBS structure and gene expression (Figure 4). This conclusion differs from that of a recent study (Burkhardt et al., 2017) that interpreted strong correlations between TE and the DMS reactivity of endogenous genes as evidence that TE is regulated by coding sequence structure. Our analysis indicates that TE is only weakly correlated with $\Delta G^{\ddagger}_{\text{unfold}}$ in coding regions (Figure 3D). In addition, given that correlations between SHAPE reactivity and TE are best explained by ribosome-mediated unfolding of the CDS (Figure 2), reduced CDS structure is most likely a consequence rather than a cause of high TE. Overall, the model that endogenous genes rely on RBS structure to tune TE explains the unique evolutionary constraint of RBS-adjacent sequences (Bentele et al., 2013; Tuller et al., 2010a) and unifies our understanding of translation regulation for synthetic and endogenous genes. Again, these broad insights into the regulation of TE require robust models of the underlying mRNA structure.

Our data also allow us to distinguish whether translation initiation depends on kinetics versus equilibrium unfolding of RBS structure. This distinction is essential for understanding the multi-step, highly regulated mechanism of translation initiation and, correspondingly, how translation is dynamically reprogrammed in response to cellular stimuli such as heat shock. The possibility of a kinetic mechanism was first proposed from a theoretical analysis showing that, at equilibrium, the lifetime of the unfolded state for a well-structured RBS is much too short to bind a 30S subunit (de Smit and van Duin, 2003). This limitation can be overcome if the 30S subunit first binds non-specifically to an mRNA and transiently "stands by" until the RBS unfolds. The importance of standby sites in translation initiation is now well supported (Espah Borujeni et al., 2014; Studer and Joseph, 2006). However, whether translation initiation depends on RNA unfolding kinetics has been essentially untestable because of the difficulty of modeling long-range RNA structures; not modeling such long-range structures effectively hides differences between equilibrium versus kinetic unfolding mechanisms (STAR Methods). The kinetic unfolding model explains roughly 40% of the observed TE variation in endogenous genes compared with only 13% explained by the equilibrium unfolding model. Necessary approximations made in our analysis leave open the possibility of contributions from an equilibrium mechanism (STAR Methods), but overall, our data imply that the kinetic mechanism predominates. When combined with accurate mRNA secondary structure models, incorporation of the kinetic mechanism into holistic biophysical models of translation is likely to yield further improvements in the ability to predict and rationally tune gene TE (Espah Borujeni et al., 2017).

Our work also reveals that large-scale RNA structure probing and modeling, when sufficiently accurate, make it possible to discover and understand complex post-transcriptional regulatory mechanisms. We found that searching for well-defined and highly structured RNA elements (low-SHAPE/low-entropy motifs) identifies 70% of previously known regulatory structures. The few known structures missed by our analysis consist of small and dynamic RNA motifs that present challenges for any detection strategy. This initial finding supports the hypothesis that many functional motifs have been evolutionary selected to have uniquely well-defined structures relative to the genetic background and that searching for such motifs will be useful for identifying novel regulatory elements. Strikingly, searching for low-SHAPE/low-entropy motifs across all non-coding regions in our dataset revealed well-structured motifs occur in

35% of UTRs and IGRs. The large majority of these motifs are well-conserved, and many overlap functional sites of protein binding, RNase processing, transcription termination, and small RNA binding, strongly implying involvement of RNA structure in diverse post-transcriptional regulatory processes (Table S1). We specifically validated the protein binding activity for three regulatory elements upstream of the ribosomal protein genes *rplM*, *rpmB*, and *rpmH*. The discovery of novel RNA motifs is particularly significant given that our analysis was limited to highly expressed housekeeping genes in *E. coli*, which represent some of the most intensively interrogated and finely parsed genetic loci in biology. Although outside the scope of our current study, the 46 other novel structures identified by our motif search represent compelling targets for future functional studies (Table S1); for example, complex motifs were found in front of essential genes *rpsT* (ribosomal protein S20), *csrA* (carbon storage regulator A, CsrA), *rho* (Rho terminator factor), *rpoB* and *rpoC* (RNA polymerase subunits β and β'), and *accA* and *accB* (subunits of acetyl-coenzyme A [CoA] carboxylase).

More important than any individual conclusion, our data collectively imply that regulation by RNA structure is much more common than previously appreciated. Indeed, either by tuning TE via RBS structure or using non-coding structure to achieve more complex differential regulation, every single gene examined here is regulated in a meaningful way by RNA structure (Figure 7). Our dataset covers roughly 8% of the *E. coli* genome, suggesting that the majority of RNA-regulatory structures and functions have yet to be discovered.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - ○ In-cell SHAPE probing
  - ○ SHAPE probing of cell-free RNA
  - ○ Reverse transcription
  - ○ Library preparation and sequencing
  - ○ Translation reporter assays
  - ○ Electrophoretic mobility shift assays
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ Read trimming and sequence alignment
  - ○ SHAPE reactivity calculation
  - ○ Calculation of gene median SHAPE
  - ○ Coding region aperiodicity
  - ○ Transcript boundary assignment
  - ○ Secondary structure modeling and analysis
  - ○ Calculation of $\Delta G^{\ddagger}_{unfold}$ and ΔG_{unfold}
  - ○ RBS–TE correlations
  - ○ CDS–TE correlations
  - ○ Translational coupling analysis
  - ○ Automated motif detection
  - ○ Motif conservation analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes eight figures and five tables and can be found with this article online at https://doi.org/10.1016/j.cell.2018.02.034.

ACKNOWLEDGMENTS

This work was supported by grants from the NIH (R35 GM122532) and the NSF (MCB-1121024) to K.M.W. and NIBR internal funding. A.M.M. is supported by a postdoctoral fellowship from the Arnold and Mabel Beckman Foundation and was supported by the UNC Lineberger NIH training fellowship (T32-CA009156). We are indebted to D. Mathews (University of Rochester) for helpful discussions; M. Smola (UNC) for sharing his ΔSHAPE analysis framework; and P. Irving (UNC), S. Gowrisankar (NIBR), A. Ho (NIBR), O. Iartchouk (NIBR), and B. Jones (NIBR) for assistance.

AUTHOR CONTRIBUTIONS

K.M.W. conceived the study. A.M.M., S.B., G.M.R., J.L.B., and K.M.W. designed the experiments with assistance from C.E.H., B.K.P., N.K., and R.N. A.M.M., S.B., and K.M.W. designed the analyses. A.M.M. and G.M.R. performed the experiments with assistance from V.M.R. A.M.M. and S.B. performed the analyses with assistance from C.E.H. and B.K.P. A.M.M. and K.M.W. wrote the paper with input from all authors.

DECLARATION OF INTERESTS

K.M.W. is an advisor to and holds equity in Ribometrix, to which SHAPE-MaP technologies have been licensed. C.E.H., B.K.P., V.M.R., N.K., R.N., and J.L.B. were employees of the Novartis Biomedical Research Institute. All other authors declare no competing interests.

Received: November 12, 2017
Revised: January 2, 2018
Accepted: February 15, 2018
Published: March 15, 2018

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Angelini et al. (2008); Aseev et al. (2008, 2016); Bardey et al. (2005); Barry et al. (1980); Burton et al. (1983); Byström et al. (1989); Chiaruttini et al. (1996); Christensen et al. (1984); Chung et al. (1993); Climie and Friesen (1988); Conway et al. (2014); Dennis (1984); Desnoyers et al. (2009); Donly and Mackie (1988); Downing and Dennis (1987); Dubey et al. (2005); Dykxhoorn et al. (1996, 1997); Fu et al. (2013); Giuliodori et al. (2010); Guillier et al. (2005); Hansen et al. (1982, 1985); He et al. (1993); Hemm et al. (2008); Higashi et al. (2008); Hollands et al. (2014); Holmqvist et al. (2016); Iben and Draper (2008); James and Cronan (2004); Kakuda et al. (1994); Kalamorz et al. (2007); Kawano et al. (2005); Keseler et al. (2013); Li et al. (2014); Livny et al. (2006, 2008); Mackie and Parsons (1983); Matsumoto et al. (1986); McDowall et al. (1994); Meades et al. (2010); Morita et al. (1999); My et al. (2015); Nagai et al. (1991); Nawrocki et al. (2015); Ott et al. (2012); Pannuri et al. (2016); Parsons and Mackie (1983); Parsons et al. (1988); Passador and Linn (1989, 1992); Peacock et al. (1985); Peng et al. (2014a, 2014b); Peters et al. (2009, 2011); Pichon et al. (2012); Podkovyrov and Larson (1995); Post et al. (1993); Powell et al. (1995); Pulvermacher et al. (2008); Rivas and Eddy (2001); Romeo et al. (2013); Saito and Nomura (1994); Salgado et al. (2013); Salim et al. (2012); Salvail et al. (2010); Schlax et al. (2001); Sevostyanova and Groisman (2015); Sharma et al. (2007); Shen et al. (1988); Smith and Cronan (2014); Steward and Linn (1992); Tran et al. (2009); Tsui and Winkler (1994); Tsui et al. (1994, 1996); Turnbough and Switzer (2008); Urban and Vogel (2008); Uzilov et al. (2006); Washietl et al. (2005); Wikström et al. (1988); Wilson and Sharp (2006); Wirth et al. (1982); Yajnik and Godson (1993); Yakhnin et al. (2011); Zengel and Lindahl (1996); Zhang et al. (2015).

REFERENCES

Angelini, S., Gerez, C., Ollagnier-de Choudens, S., Sanakis, Y., Fontecave, M., Barras, F., and Py, B. (2008). NfuA, a new factor required for maturing Fe/S proteins in Escherichia coli under oxidative stress and iron starvation conditions. J. Biol. Chem. *283*, 14084–14091.

Aseev, L.V., Bylinkina, N.S., and Boni, I.V. (2015). Regulation of the rplY gene encoding 5S rRNA binding protein L25 in Escherichia coli and related bacteria. RNA *21*, 851–861.

Aseev, L.V., Koledinskaya, L.S., and Boni, I.V. (2016). Regulation of Ribosomal Protein Operons rplM-rpsI, rpmB-rpmG, and rplU-rpmA at the Transcriptional and Translational Levels. J. Bacteriol. *198*, 2494–2502.

Aseev, L.V., Levandovskaya, A.A., Tchufistova, L.S., Scaptsova, N.V., and Boni, I.V. (2008). A new regulatory circuit in ribosomal protein operons: S2-mediated control of the rpsB-tsf expression in vivo. RNA *14*, 1882–1894.

Bardey, V., Vallet, C., Robas, N., Charpentier, B., Thouvenot, B., Mougin, A., Hajnsdorf, E., Régnier, P., Springer, M., and Branlant, C. (2005). Characterization of the molecular mechanisms involved in the differential production of erythrose-4-phosphate dehydrogenase, 3-phosphoglycerate kinase and class II fructose-1,6-bisphosphate aldolase in Escherichia coli. Mol. Microbiol. *57*, 1265–1287.

Barry, G., Squires, C., and Squires, C.L. (1980). Attenuation and processing of RNA from the rplJL–rpoBC transcription unit of Escherichia coli. Proc. Natl. Acad. Sci. U.S.A *77*, 3331–3335.

Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., and Blüthgen, N. (2013). Efficient translation initiation dictates codon usage at gene start. Mol. Syst. Biol. *9*, 675.

Boël, G., Letso, R., Neely, H., Price, W.N., Wong, K.-H., Su, M., Luff, J., Valecha, M., Everett, J.K., Acton, T.B., et al. (2016). Codon influence on protein expression in E. coli correlates with mRNA levels. Nature *529*, 358–363.

Burkhardt, D.H., Rouskin, S., Zhang, Y., Li, G.-W., Weissman, J.S., and Gross, C.A. (2017). Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. eLife *6*, 811.

Burton, Z.F., Gross, C.A., Watanabe, K.K., and Burgess, R.R. (1983). The operon that encodes the sigma subunit of RNA polymerase also encodes ribosomal protein S21 and DNA primase in E. coli K12. Cell *32*, 335–349.

Byström, A.S., Gabain, von, A., and Björk, G.R. (1989). Differentially expressed trmD ribosomal protein operon of Escherichia coli is transcribed as a single polycistronic mRNA species. J. Mol. Biol. *208*, 575–586.

Cech, T.R., and Steitz, J.A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. Cell *157*, 77–94.

Chiaruttini, C., Milet, M., and Springer, M. (1996). A long-range RNA-RNA interaction forms a pseudoknot required for translational control of the IF3-L35-L20 ribosomal protein operon in Escherichia coli. EMBO J *15*, 4402–4413.

Christensen, T., Johnsen, M., Fiil, N.P., and Friesen, J.D. (1984). RNA secondary structure and translation inhibition: analysis of mutants in the rplJ leader. EMBO J *3*, 1609–1612.

Chung, T., Resnik, E., Stueland, C., and LaPorte, D.C. (1993). Relative expression of the products of glyoxylate bypass operon: contributions of transcription and translation. J. Bacteriol *175*, 4572–4575.

Climie, S.C., and Friesen, J.D. (1988). In vivo and in vitro structural analysis of the rplJ mRNA leader of Escherichia coli. Protection by bound L10-L7/L12. J. Biol. Chem. *263*, 15166–15175.

Conway, T., Creecy, J.P., Maddox, S.M., Grissom, J.E., Conkle, T.L., Shadid, T.M., Teramoto, J., San Miguel, P., Shimada, T., Ishihama, A., et al. (2014). Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. MBio *5*, e01442–e14.

de Smit, M.H., and van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc. Natl. Acad. Sci. USA *87*, 7668–7672.

de Smit, M.H., and van Duin, J. (2003). Translational standby sites: how ribosomes may deal with the rapid folding kinetics of mRNA. J. Mol. Biol. *331*, 737–743.

Deigan, K.E., Li, T.W., Mathews, D.H., and Weeks, K.M. (2009). Accurate SHAPE-directed RNA structure determination. Proc. Natl. Acad. Sci. USA *106*, 97–102.

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.

Dennis, P.P. (1984). Site specific deletions of regulatory sequences in a ribosomal protein-RNA polymerase operon in Escherichia coli. Effects on beta and beta’ gene expression. J. Biol. Chem. *259*, 3202–3209.

Desnoyers, G., Morissette, A., Prévost, K., and Massé, E. (2009). Small RNA-induced differential degradation of the polycistronic mRNA iscRSUA. EMBO J *28*, 1551–1561.

Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., and Assmann, S.M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature *505*, 696–700.

Donly, B.C., and Mackie, G.A. (1988). Affinities of ribosomal protein S20 and C-terminal deletion mutants for 16S rRNA and S20 mRNA. Nucleic Acids Res. *16*, 997–1010.

Downing, W.L., and Dennis, P.P. (1987). Transcription products from the rplKAJL-rpoBC gene cluster. J. Mol. Biol. *194*, 609–620.

Dubey, A.K., Baker, C.S., Romeo, T., and Babitzke, P. (2005). RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. RNA *11*, 1579–1587.

Dykxhoorn, D.M., St Pierre, R., and Linn, T. (1996). Synthesis of the beta and beta’ subunits of Escherichia coli RNA polymerase is autogenously regulated in vivo by both transcriptional and translational mechanisms. Mol. Microbiol. *19*, 483–493.

Dykxhoorn, D.M., St Pierre, R., Van Ham, O., and Linn, T. (1997). An efficient protocol for linker scanning mutagenesis: analysis of the translational regulation of an Escherichia coli RNA polymerase subunit gene. Nucleic Acids Res. *25*, 4209–4218.

Eddy, S.R. (2014). Computational analysis of conserved RNA secondary structure in transcriptomes and genomes. Annu. Rev. Biophys. *43*, 433–456.

Espah Borujeni, A., Channarasappa, A.S., and Salis, H.M. (2014). Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. Nucleic Acids Res. *42*, 2646–2659.

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 Res. *45*, 5437–5448.

Fu, Y., Deiorio-Haggar, K., Anthony, J., and Meyer, M.M. (2013). Most RNAs regulating ribosomal protein biosynthesis in Escherichia coli are narrowly distributed to Gammaproteobacteria. Nucleic Acids Res. *41*, 3491–3503.

Fu, Y., Deiorio-Haggar, K., Soo, M.W., and Meyer, M.M. (2014). Bacterial RNA motif in the 5′ UTR of rpsF interacts with an S6:S18 complex. RNA *20*, 168–176.

Goodman, D.B., Church, G.M., and Kosuri, S. (2013). Causes and effects of N-terminal codon bias in bacterial genes. Science *342*, 475–479.

Giuliodori, A.M., Di Pietro, F., Marzi, S., Masquida, B., Wagner, R., Romby, P., Gualerzi, C.O., and Pon, C.L. (2010). The cspA mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. Mol. Cell *37*, 21–33.

Guillier, M., Allemand, F., Dardel, F., Royer, C.A., Springer, M., and Chiaruttini, C. (2005). Double molecular mimicry in Escherichia coli: binding of ribosomal protein L20 to its two sites in mRNA is similar to its binding to 23S rRNA. Mol. Microbiol. *56*, 1441–1456.

Guimaraes, J.C., Rocha, M., and Arkin, A.P. (2014). Transcript level and sequence determinants of protein abundance and noise in Escherichia coli. Nucleic Acids Res. *42*, 4791–4799.

Hansen, F.G., Hansen, E.B., and Atlung, T. (1982). The nucleotide sequence of the dnaA gene promoter and of the adjacent rpmH gene, coding for the ribosomal protein L34, of Escherichia coli. EMBO J *1*, 1043–1048.

Hansen, F.G., Hansen, E.B., and Atlung, T. (1985). Physical mapping and nucleotide sequence of the rnpA gene that encodes the protein component of ribonuclease P in Escherichia coli. Gene *38*, 85–93.

He, B., Choi, K.Y., and Zalkin, H. (1993). Regulation of Escherichia coli glnB, prsA, and speA by the purine repressor. J. Bacteriol *175*, 3598–3606.

Hemm, M.R., Paul, B.J., Schneider, T.D., Storz, G., and Rudd, K.E. (2008). Small membrane proteins found by comparative genomics and ribosome binding site models. Mol. Microbiol. *70*, 1487–1501.

Higashi, K., Terui, Y., Suganami, A., Tamura, Y., Nishimura, K., Kashiwagi, K., and Igarashi, K. (2008). Selective structural change by spermidine in the bulged-out region of double-stranded RNA and its effect on RNA function. J. Biol. Chem. *283*, 32989–32994.

Hollands, K., Sevostiyanova, A., and Groisman, E.A. (2014). Unusually long-lived pause required for regulation of a Rho-dependent transcription terminator. Proc. Natl. Acad. Sci. U.S.A *111*, E1999–E2007.

Holmqvist, E., Wright, P.R., Li, L., Bischler, T., Barquist, L., Reinhardt, R., Backofen, R., and Vogel, J. (2016). Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. EMBO J *35*, 991–1011.

Iben, J.R., and Draper, D.E. (2008). Specific interactions of the L10(L12)4 ribosomal protein complex with mRNA, rRNA, and L11. Biochemistry *47*, 2721–2731.

James, E.S., and Cronan, J.E. (2004). Expression of two Escherichia coli acetyl-CoA carboxylase subunits is autoregulated. J. Biol. Chem. *279*, 2520–2527.

Kakuda, H., Hosono, K., Shiroishi, K., and Ichihara, S. (1994). Identification and characterization of the ackA (acetate kinase A)-pta (phosphotransacetylase) operon and complementation analysis of acetate utilization by an ackA-pta deletion mutant of Escherichia coli. J. Biochem *116*, 916–922.

Kalamorz, F., Reichenbach, B., März, W., Rak, B., and Görke, B. (2007). Feedback control of glucosamine-6-phosphate synthase GlmS expression depends on the small RNA GlmZ and involves the novel protein YhbJ in Escherichia coli. Mol. Microbiol. *65*, 1518–1533.

Kamiyama, D., Sekine, S., Barsi-Rhyne, B., Hu, J., Chen, B., Gilbert, L.A., Ishikawa, H., Leonetti, M.D., Marshall, W.F., Weissman, J.S., and Huang, B. (2016). Versatile protein tagging in cells with split fluorescent protein. Nat. Commun. *7*, 11046.

Kawano, M., Reynolds, A.A., Miranda-Rios, J., and Storz, G. (2005). Detection of 5‘‘- and 3-’’UTR-derived small RNAs and cis-encoded antisense RNAs in Escherichia coli. Nucleic Acids Res. *33*, 1040–1050.

Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M., et al. (2013). EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Res. *41*, D605–D612.

Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene *361*, 13–37.

Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. Science *324*, 255–258.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lesage, P., Chiaruttini, C., Graffe, M., Dondon, J., Milet, M., and Springer, M. (1992). Messenger RNA secondary structure and translational coupling in the Escherichia coli operon encoding translation initiation factor IF3 and the ribosomal proteins, L35 and L20. J. Mol. Biol. *228*, 366–386.

Lee, J.H., Kim, H., Ko, J., and Lee, Y. (2002). Interaction of C5 protein with RNA aptamers selected by SELEX. Nucleic Acids Res. *30*, 5360–5368.

Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell *157*, 624–635.

Livny, J., Brencic, A., Lory, S., and Waldor, M.K. (2006). Identification of 17 Pseudomonas aeruginosa sRNAs and prediction of sRNA-encoding genes in 10 diverse pathogens using the bioinformatic tool sRNAPredict2. Nucleic Acids Res. *34*, 3484–3493.

Livny, J., Teonadi, H., Livny, M., and Waldor, M.K. (2008). High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs. PLoS ONE *3*, e3197.

Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., et al. (2016). RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. Cell *165*, 1267–1279.

Mackie, G.A., and Parsons, G.D. (1983). Tandem promoters in the gene for ribosomal protein S20. J. Biol. Chem. *258*, 7840–7846.

Maguire, B.A., and Wild, D.G. (1997). Mutations in the rpmBG operon of Escherichia coli that affect ribosome assembly. J. Bacteriol. *179*, 2486–2493.

Matelska, D., Purta, E., Panek, S., Boniecki, M.J., Bujnicki, J.M., and Dunin-Horkawicz, S. (2013). S6:S18 ribosomal protein complex interacts with a structural motif present in its own mRNA. RNA *19*, 1341–1348.

Matsumoto, Y., Shigesada, K., Hirano, M., and Imai, M. (1986). Autogenous regulation of the gene for transcription termination factor rho in Escherichia coli: localization and function of its attenuators. J. Bacteriol *166*, 945–958.

Mattheakis, L.C., and Nomura, M. (1988). Feedback regulation of the spc operon in Escherichia coli: translational coupling and mRNA processing. J. Bacteriol. *170*, 4484–4492.

Mauger, D.M., Golden, M., Yamane, D., Williford, S., Lemon, S.M., Martin, D.P., and Weeks, K.M. (2015). Functionally conserved architecture of hepatitis C virus RNA genomes. Proc. Natl. Acad. Sci. USA *112*, 3692–3697.

McDowall, K.J., Lin-Chao, S., and Cohen, S.N. (1994). A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. J. Biol. Chem. *269*, 10790–10796.

McGinnis, J.L., Liu, Q., Lavender, C.A., Devaraj, A., McClory, S.P., Fredrick, K., and Weeks, K.M. (2015). In-cell SHAPE reveals that free 30S ribosome subunits are in the inactive state. Proc. Natl. Acad. Sci. USA *112*, 2425–2430.

Meades, G., Benson, B.K., Grove, A., and Waldrop, G.L. (2010). A tale of two functions: enzymatic activity and translational repression by carboxyltransferase. Nucleic Acids Res. *38*, 1217–1227.

Morita, M., Kanemori, M., Yanagi, H., and Yura, T. (1999). Heat-induced synthesis of sigma32 in Escherichia coli: structural and functional dissection of rpoH mRNA secondary structure. J. Bacteriol *181*, 401–410.

My, L., Achkar, N.G., Viala, J.P., and Bouveret, E. (2015). Reassessment of the Genetic Regulation of Fatty Acid Synthesis in Escherichia coli: Global Positive Control by the Functional Dual Regulator FadR. J. Bacteriol *197*, 1862–1872.

Nagai, H., Yuzawa, H., and Yura, T. (1991). Interplay of two cis-acting mRNA regions in translational control of sigma 32 synthesis during the heat shock response of Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. *88*, 10515–10519.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics *29*, 2933–2935.

Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R., Floden, E.W., Gardner, P.P., Jones, T.A., Tate, J., and Finn, R.D. (2015). Rfam 12.0: updates to the RNA families database. Nucleic Acids Res. *43*, D130–D137.

Ott, A., Idali, A., Marchais, A., and Gautheret, D. (2012). NAPP: the Nucleic Acid Phylogenetic Profile Database. Nucleic Acids Res. *40*, D205–D209.

Pannuri, A., Vakulskas, C.A., Zere, T., McGibbon, L.C., Edwards, A.N., Georgellis, D., Babitzke, P., and Romeo, T. (2016). Circuitry Linking the Catabolite Repression and Csr Global Regulatory Systems of Escherichia coli. J. Bacteriol *198*, 3000–3015.

Parsons, G.D., and Mackie, G.A. (1983). Expression of the gene for ribosomal protein S20: effects of gene dosage. J. Bacteriol *154*, 152–160.

Parsons, G.D., Donly, B.C., and Mackie, G.A. (1988). Mutations in the leader sequence and initiation codon of the gene for ribosomal protein S20 (rpsT) affect both translational efficiency and autoregulation. J. Bacteriol *170*, 2485–2492.

Passador, L., and Linn, T. (1989). Autogenous regulation of the RNA polymerase beta subunit of Escherichia coli occurs at the translational level in vivo. J. Bacteriol *171*, 6234–6242.

Passador, L., and Linn, T. (1992). An internal region of rpoB is required for autogenous translational regulation of the beta subunit of Escherichia coli RNA polymerase. J. Bacteriol *174*, 7174–7179.

Peacock, S., Lupski, J.R., Godson, G.N., and Weissbach, H. (1985). In vitro stimulation of Escherichia coli RNA polymerase sigma subunit synthesis by NusA protein. Gene *33*, 227–234.

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T.C., and Waldo, G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. Nat. Biotechnol. *24*, 79–88.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread Co-translational RNA Decay Reveals Ribosome Dynamics. Cell *161*, 1400–1412.

Peng, Y., Curtis, J.E., Fang, X., and Woodson, S.A. (2014a). Structural model of an mRNA in complex with the bacterial chaperone Hfq. Proc. Natl. Acad. Sci. U.S.A. *111*, 17134–17139.

Peng, Y., Soper, T.J., and Woodson, S.A. (2014b). Positional effects of AAN motifs in rpoS regulation by sRNAs and Hfq. J. Mol. Biol. *426*, 275–285.

Peters, J.M., Mooney, R.A., Kuan, P.F., Rowland, J.L., Keles, S., and Landick, R. (2009). Rho directs widespread termination of intragenic and stable RNA transcription. Proc. Natl. Acad. Sci. U.S.A. *106*, 15406–15411.

Peters, J.M., Vangeloff, A.D., and Landick, R. (2011). Bacterial transcription terminators: the RNA 3′-end chronicles. J. Mol. Biol. *412*, 793–813.

Pichon, C., du Merle, L., Caliot, M.E., Trieu-Cuot, P., and Le Bouguénec, C. (2012). An in silico model for identification of small RNAs in whole bacterial genomes: characterization of antisense RNAs in pathogenic Escherichia coli and Streptococcus agalactiae strains. Nucleic Acids Res. *40*, 2846–2861.

Podkovyrov, S., and Larson, T.J. (1995). Lipid biosynthetic genes and a ribosomal protein gene are cotranscribed. FEBS Lett. *368*, 429–431.

Post, D.A., Hove-Jensen, B., and Switzer, R.L. (1993). Characterization of the hemA-prs region of the Escherichia coli and Salmonella typhimurium chromosomes: identification of two open reading frames and implications for prs expression. J. Gen. Microbiol. *139*, 259–266.

Powell, B.S., Court, D.L., Inada, T., Nakamura, Y., Michotey, V., Cui, X., Reizer, A., Saier, M.H., and Reizer, J. (1995). Novel proteins of the phosphotransferase system encoded within the rpoN operon of Escherichia coli. Enzyme IIANtr affects growth on organic nitrogen and the conditional lethality of an erats mutant. J. Biol. Chem. *270*, 4822–4839.

Pulvermacher, S.C., Stauffer, L.T., and Stauffer, G.V. (2008). The role of the small regulatory RNA GcvB in GcvB/mRNA posttranscriptional regulation of oppA and dppA in Escherichia coli. FEMS Microbiol. Lett. *281*, 42–50.

Ramani, V., Qiu, R., and Shendure, J. (2015). High-throughput determination of RNA structure by proximity ligation. Nat. Biotechnol. *33*, 980–984.

Reuter, J.S., and Mathews, D.H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics *11*, 129.

Rio, D.C., Ares, M., Jr., Hannon, G.J., and Nilsen, T.W. (2011). RNA: a laboratory manual (Cold Spring Harbor: Cold Spring Harbor Laboratory Press).

Rivas, E., and Eddy, S.R. (2001). Noncoding RNA gene detection using comparative sequence analysis. BMC Bioinformatics *2*, 8.

Rivas, E., Klein, R.J., Jones, T.A., and Eddy, S.R. (2001). Computational identification of noncoding RNAs in E. coli by comparative genomics. Curr. Biol. *11*, 1369–1373.

Rivera-León, R., Green, C.J., and Vold, B.S. (1995). High-level expression of soluble recombinant RNase P protein from *Escherichia coli*. J. Bacteriol. *177*, 2564–2566.

Romeo, T., Vakulskas, C.A., and Babitzke, P. (2013). Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. Environ. Microbiol. *15*, 313–324.

Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature *505*, 701–705.

Ryder, S.P., Recht, M.I., and Williamson, J.R. (2008). Quantitative Analysis of Protein-RNA Interactions by Gel Mobility Shift. In RNA-Protein Interaction Protocols, R.-J. Lin, ed. (Humana Press), pp. 99–115.

Saito, K., and Nomura, M. (1994). Post-transcriptional regulation of the str operon in Escherichia coli. Structural and mutational analysis of the target site for translational repressor S7. J. Mol. Biol. *235*, 125–139.

Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muñiz-Rascado, L., García-Sotelo, J.S., Weiss, V., Solano-Lira, H., Martínez-Flores, I., Medina-Rivera, A., et al. (2013). RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. Nucleic Acids Res. *41*, D203–D213.

Salim, N.N., Faner, M.A., Philip, J.A., and Feig, A.L. (2012). Requirement of upstream Hfq-binding (ARN)x elements in glmS and the Hfq C-terminal region for GlmS upregulation by sRNAs GlmZ and GlmY. Nucleic Acids Res. *40*, 8021–8032.

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

Salvail, H., Lanthier-Bourbonnais, P., Sobota, J.M., Caza, M., Benjamin, J.-A.M., Mendieta, M.E.S., Lépine, F., Dozois, C.M., Imlay, J., and Massé, E. (2010). A small RNA promotes siderophore production through transcriptional and metabolic remodeling. Proc. Natl. Acad. Sci. U.S.A. *107*, 15223–15228.

Schlax, P.J., Xavier, K.A., Gluick, T.C., and Draper, D.E. (2001). Translational repression of the Escherichia coli alpha operon mRNA: importance of an mRNA conformational switch and a ternary entrapment complex. J. Biol. Chem. *276*, 38494–38501.

Sevostyanova, A., and Groisman, E.A. (2015). An RNA motif advances transcription by preventing Rho-dependent termination. Proc. Natl. Acad. Sci. U.S.A *112*, E6835–E6843.

Sharma, C.M., Darfeuille, F., Plantinga, T.H., and Vogel, J. (2007). A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. Genes Dev. *21*, 2804–2817.

Shen, P., Zengel, J.M., and Lindahl, L. (1988). Secondary structure of the leader transcript from the Escherichia coli S10 ribosomal protein operon. Nucleic Acids Res. *16*, 8905–8924.

Siegfried, N.A., Busan, S., Rice, G.M., Nelson, J.A.E., and Weeks, K.M. (2014). RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). Nat. Methods *11*, 959–965.

Slinger, B.L., Deiorio-Haggar, K., Anthony, J.S., Gilligan, M.M., and Meyer, M.M. (2014). Discovery and validation of novel and distinct RNA regulators for ribosomal protein S15 in diverse bacterial phyla. BMC Genomics *15*, 657.

Smith, A.C., and Cronan, J.E. (2014). Evidence against translational repression by the carboxyltransferase component of Escherichia coli acetyl coenzyme A carboxylase. J. Bacteriol *196*, 3768–3775.

Smola, M.J., Calabrese, J.M., and Weeks, K.M. (2015a). Detection of RNA-Protein Interactions in Living Cells with SHAPE. Biochemistry *54*, 6867–6875.

Smola, M.J., Rice, G.M., Busan, S., Siegfried, N.A., and Weeks, K.M. (2015b). Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. Nat. Protoc. *10*, 1643–1669.

Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.-W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T., and Chang, H.Y. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. Nature *519*, 486–490.

Steen, K.-A., Siegfried, N.A., and Weeks, K.M. (2011). Synthesis of 1-methyl-7-nitroisatoic anhydride (1M7). Protoc. Exch. Published online November 3, 2011. https://doi.org/10.1038/protex.2011.255.

Steward, K.L., and Linn, T. (1992). Transcription frequency modulates the efficiency of an attenuator preceding the rpoBC RNA polymerase genes of Escherichia coli: possible autogenous control. Nucleic Acids Res. *20*, 4773–4779.

Studer, S.M., and Joseph, S. (2006). Unfolding of mRNA secondary structure by the bacterial translation initiation complex. Mol. Cell *22*, 105–115.

Sugimoto, Y., Vigilante, A., Darbo, E., Zirra, A., Militti, C., D'Ambrogio, A., Luscombe, N.M., and Ule, J. (2015). hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. Nature *519*, 491–494.

Takyar, S., Hickerson, R.P., and Noller, H.F. (2005). mRNA helicase activity of the ribosome. Cell *120*, 49–58.

Tatusova, T., Ciufo, S., Fedorov, B., O'Neill, K., and Tolstoy, I. (2014). RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Res. *42*, D553–D559.

Thomas, M.S., Bedwell, D.M., and Nomura, M. (1987). Regulation of α operon gene expression in Escherichia coli. A novel form of translational coupling. J. Mol. Biol. *196*, 333–345.

Tran, T.T., Zhou, F., Marshburn, S., Stead, M., Kushner, S.R., and Xu, Y. (2009). De novo computational prediction of non-coding RNA genes in prokaryotic genomes. Bioinformatics *25*, 2897–2905.

Tsui, H.C., and Winkler, M.E. (1994). Transcriptional patterns of the mutL-miaA superoperon of Escherichia coli K-12 suggest a model for posttranscriptional regulation. Biochimie *76*, 1168–1177.

Tsui, H.C., Feng, G., and Winkler, M.E. (1996). Transcription of the mutL repair, miaA tRNA modification, hfq pleiotropic regulator, and hflA region protease genes of Escherichia coli K-12 from clustered Esigma32-specific promoters during heat shock. J. Bacteriol *178*, 5719–5731.

Tsui, H.C., Leung, H.C., and Winkler, M.E. (1994). Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12. Mol. Microbiol. *13*, 35–49.

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010a). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell *141*, 344–354.

Tuller, T., Waldman, Y.Y., Kupiec, M., and Ruppin, E. (2010b). Translation efficiency is determined by both codon bias and folding energy. Proc. Natl. Acad. Sci. USA *107*, 3645–3650.

Turnbough, C.L., and Switzer, R.L. (2008). Regulation of pyrimidine biosynthetic gene expression in bacteria: repression without repressors. Microbiol. Mol. Biol. Rev. *72*, 266–300–tableofcontents.

Tyrrell, J., McGinnis, J.L., Weeks, K.M., and Pielak, G.J. (2013). The cellular environment stabilizes adenine riboswitch RNA structure. Biochemistry *52*, 8777–8785.

Urban, J.H., and Vogel, J. (2008). Two seemingly homologous noncoding RNAs act hierarchically to activate glmS mRNA translation. PLoS Biol. *6*, e64.

Uzilov, A.V., Keegan, J.M., and Mathews, D.H. (2006). Detection of non-coding RNAs on the basis of predicted secondary structure formation free energy change. BMC Bioinformatics *7*, 173.

Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E., and Chang, H.Y. (2014). Landscape and variation of RNA secondary structure across the human transcriptome. Nature *505*, 706–709.

Washietl, S., Hofacker, I.L., and Stadler, P.F. (2005). Fast and reliable prediction of noncoding RNAs. Proc. Natl. Acad. Sci. USA *102*, 2454–2459.

Watters, K.E., Abbott, T.R., and Lucks, J.B. (2016). Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. Nucleic Acids Res. *44*, e12–e12.

Weeks, K.M. (2015). Review toward all RNA structures, concisely. Biopolymers *103*, 438–448.

Weinberg, Z., and Breaker, R.R. (2011). R2R–software to speed the depiction of aesthetic consensus RNA secondary structures. BMC Bioinformatics *12*, 3.

Weinberg, Z., Kim, P.B., Chen, T.H., Li, S., Harris, K.A., Lünse, C.E., and Breaker, R.R. (2015). New classes of self-cleaving ribozymes revealed by comparative genomics analysis. Nat. Chem. Biol. *11*, 606–610.

Wikström, P.M., Byström, A.S., and Björk, G.R. (1988). Non-autogenous control of ribosomal protein synthesis from the trmD operon in Escherichia coli. J. Mol. Biol. *203*, 141–152.

Wilson, L.A., and Sharp, P.M. (2006). Enterobacterial repetitive intergenic consensus (ERIC) sequences in Escherichia coli: Evolution and implications for ERIC-PCR. Mol. Biol. Evol. *23*, 1156–1168.

Wirth, R., Littlechild, J., and Böck, A. (1982). Ribosomal protein S20 purified under mild conditions almost completely inhibits its own translation. Mol. Gen. Genet. *188*, 164–166.

Yajnik, V., and Godson, G.N. (1993). Selective decay of Escherichia coli dnaG messenger RNA is initiated by RNase E. J. Biol. Chem. *268*, 13253–13260.

Yakhnin, H., Yakhnin, A.V., Baker, C.S., Sineva, E., Berezin, I., Romeo, T., and Babitzke, P. (2011). Complex regulation of the global regulatory gene csrA: CsrA-mediated translational repression, transcription from five promoters by Eσ^{70} and Eσ(S), and indirect transcriptional activation by CsrA. Mol. Microbiol. *81*, 689–704.

Yao, Z., Barrick, J., Weinberg, Z., Neph, S., Breaker, R., Tompa, M., and Ruzzo, W.L. (2007). A computational pipeline for high- throughput discovery of cis-regulatory noncoding RNA in prokaryotes. PLoS Comput. Biol. *3*, e126.

Yates, J.L., and Nomura, M. (1980). E. coli ribosomal protein L4 is a feedback regulatory protein. Cell *21*, 517–522.

Zengel, J.M., and Lindahl, L. (1996). A hairpin structure upstream of the terminator hairpin required for ribosomal protein L4-mediated attenuation control of the S10 operon of Escherichia coli. J. Bacteriol *178*, 2383–2387.

Zhang, Y., Mandava, C.S., Cao, W., Li, X., Zhang, D., Li, N., Zhang, Y., Zhang, X., Qin, Y., Mi, K., et al. (2015). HflX is a ribosome-splitting factor rescuing stalled ribosomes under stress conditions. Nat. Struct. Mol. Biol. *22*, 906–913.

Zubradt, M., Gupta, P., Persad, S., Lambowitz, A.M., Weissman, J.S., and Rouskin, S. (2017). DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. Nat. Methods *14*, 75–82.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
E. coli K12 MG1655	Bo Li, UNC, Chapel Hill	RRID: SCR_002433
E. coli TOP10	Invitrogen	Cat# C404010
E. coli BL21-AI	Invitrogen	Cat# C607003
Chemicals, Peptides, and Recombinant Proteins		
1-methyl-7-nitroisatoic anhydride (1M7)	(Steen et al., 2011)	CAS: 73043-80-8
Kasugamycin	Sigma	CAS: 19408-46-9
SuperScript II reverse transcriptase	Invitrogen	Cat# 18064014
T7 RNA polymerase	(Rio et al., 2011)	N/A
E. coli L9 protein	This paper	N/A
E. coli L28 protein	This paper	N/A
E. coli L33 protein	This paper	N/A
E. coli C5 protein	This paper	N/A
E. coli L13 protein	This paper	N/A
rpmB WT RNA	This paper	N/A
rpmB $H1_{insA\text{-}GC}$ RNA	This paper	N/A
rpmB ΔH3 RNA	This paper	N/A
rpmB WT_{trunc} RNA	This paper	N/A
rplM H1-H5 RNA	This paper	N/A
rplM H2-H4 RNA	This paper	N/A
rpmH WT RNA	This paper	N/A
rpmH ΔH2 RNA	This paper	N/A
Critical Commercial Assays		
Bacterial Ribo-Zero rRNA removal kit	Illumina	Cat# MRZMB126
NEBNext second strand synthesis module	NEB	Cat# E6111S
NexteraXT library prep kit	Illumina	Cat# FC-131-1024
Isothermal assembly cloning kit	NEB	Cat# E5520S
Deposited Data		
Raw SHAPE-MaP sequencing reads	This paper	https://www.ebi.ac.uk/ena/data/view/PRJEB23974
E. coli K12 MG1655 reference genome, U00096.2	GenBank	https://www.ncbi.nlm.nih.gov/nuccore/U00096.2
Conway et al. transcript annotations	(Conway et al., 2014)	Table S4 at http://mbio.asm.org/content/5/4/e01442-14.full
RegulonDB transcript annotations	(Salgado et al., 2013)	http://regulondb.ccg.unam.mx
RefSeq bacterial genomes	RefSeq	ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria/assembly_summary.txt
Oligonucleotides		
Oligos used for construction of pTrc-TE plasmid, see Table S4	This paper	N/A
Oligos used for RNA transcription, see Table S5	This paper	N/A
Recombinant DNA		
Plasmid: pTrcHis A	Invitrogen	Cat# V36020
Plasmid: pTrc-TE	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmids: pTrc-TE with inserted leader sequences, see Table S3.	GenScript, this paper	N/A
Plasmid: pET-29a(+)-rplI	GenScript, this paper	N/A
Plasmid: pET-29a(+)-rplM	GenScript, this paper	N/A
Plasmid: pET-29a(+)-rpmB	GenScript, this paper	N/A
Plasmid: pET-29a(+)-rpmG	GenScript, this paper	N/A
Plasmid: pET-29a(+)-rnpA	GenScript, this paper	N/A
Software and Algorithms		
Bowtie2	(Langmead and Salzberg, 2012)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
RNAstructure (v5.8)	(Reuter and Mathews, 2010)	https://rna.urmc.rochester.edu/RNAstructure.html
SuperFold	(Siegfried et al., 2014; Smola et al., 2015b)	http://www.chem.unc.edu/rna/software.html
ShapeMapper (v1)	(Siegfried et al., 2014; Smola et al., 2015b)	http://www.chem.unc.edu/rna/software.html
Motif finder algorithm	This paper	http://www.chem.unc.edu/rna/software.html
Homolog search algorithm	This paper	http://www.chem.unc.edu/rna/software.html
FlowJo	FlowJo	https://www.flowjo.com
Infernal (v1.1.1)	(Nawrocki and Eddy, 2013)	http://eddylab.org/infernal/
R2R (v1.0.4)	(Weinberg and Breaker, 2011)	http://breaker.research.yale.edu/R2R/
Other		
List of known non-coding motifs	RFAM (Nawrocki et al., 2015); RAREs (Aseev et al., 2015; Fu et al., 2013; 2014; Matelska et al., 2013).	N/A
Processed SHAPE data and structure models	This paper	http://www.chem.unc.edu/rna/publications.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Kevin Weeks (weeks@unc.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

SHAPE probing experiments were performed on *E. coli* K12 MG1655 (gift of Bo Li, UNC, Chapel Hill), grown at 37°C in LB Broth. Translation reporter assays were done using transformants of *E. coli* TOP10 (Invitrogen), grown at 37°C in Terrific Broth. Proteins for *in vitro* binding assays were expressed in *E. coli* BL21-AI (Invitrogen), grown in Terrific Broth or ZYM-5052 auto-induction media at 37°C with shift to 18°C during induction.

METHOD DETAILS

In-cell SHAPE probing

In full biological replicates, 2 mL of overnight culture were added to 48 mL of LB. Cells were incubated with shaking until the culture reached OD_{600} ~0.5 (~30 min). To each culture, 5.55 mL of 10 mg/mL kasugamycin or LB was added, and cells were incubated with shaking for 20 minutes. Next, the media was buffered by addition of 3 mL 2 M HEPES pH 8.0 (100 mM final), and the cultures incubated by shaking for two minutes. SHAPE probing was performed in culture tubes by adding 9 mL of cells to 600 μL of 167 mM 1M7 (Steen et al., 2011) in DMSO and shaking for 2 minutes. The samples were transferred to a new tube containing 200 μL of 500 mM 1M7 in DMSO and incubated at 37°C for 2 minutes. This was repeated once more for a total of three rounds of 1M7 modification. The same procedure was performed for untreated control samples, but adding only DMSO. Cells were pelleted at 4000 *g* for 20 minutes at 4°C. Supernatant was discarded, and the cell pellet was resuspended in 200 μL of 0.5 × TE buffer (pH 8.0) with lysozyme (1 mg/mL) and incubated on ice for 5 minutes. 1 mL of Trizol reagent (Invitrogen) was added, and the reaction tubes were incubated at room temperature for 5 minutes. To each sample was added 200 μL of cold chloroform. Samples were mixed by shaking for 15 s and incubated at room temperature for 2-3 minutes. Tubes were centrifuged at 12,000 *g* for 15 min at 4°C. The aqueous upper layer was transferred to a new tube, and a 1.1 × volume of isopropanol was added. Reactions were incubated at −20°C for 30 minutes and then centrifuged at 15,000 *g* for 30 minutes at 4°C. The supernatant was discarded and pellets were washed twice with 500 μL 80% ethanol, centrifuging 5 minutes at 15,000 *g* between washes. Following the washes, the supernatant was discarded, and pellets were air-dried for 5 minutes. Samples were then treated with DNase I (Ambion) and purified on an affinity column (RNeasy Mini Kit; QIAGEN).

SHAPE probing of cell-free RNA

In full biological replicates, 2 mL of overnight culture were added to 48 mL of LB. Cells were grown to mid-log phase (OD_{600} = 0.6), and RNA was gently extracted under non-denaturing conditions as described (Deigan et al., 2009). Total RNA was exchanged using a gravity-flow Sephadex column (PD-10; GE Healthcare) into folding buffer containing 50 mM HEPES, pH 8.0, 200 mM potassium acetate, and 5 mM $MgCl_2$ and incubated at 37°C for 20 minutes. RNA was modified using three consecutive additions of 1M7 as follows: Folded RNA (360 μL) was combined with 32 μL of 167 mM 1M7 solution in anhydrous DMSO, rapidly mixed, and incubated at 37°C for 2 minutes. Subsequently, 8 μL of 500 mM 1M7 in DMSO was then added, samples were quickly vortexed and incubated for 2 minutes at 37°C, and this was repeated once. Following modification, RNA was isolated by affinity chromatography (RNeasy Mini kit; QIAGEN), followed by DNase I treatment (Ambion), and a second affinity column (RNeasy).

Reverse transcription

The integrity of each total RNA sample was evaluated using an Agilent Bioanalyzer 2100; RIN numbers were greater than 8.0 for all samples. rRNA was subsequently removed from 15 μg of total RNA (bacterial Ribo-zero kit; Illumina), yielding 50-100 ng of mRNA. All recovered RNA was input into SHAPE-MaP reverse transcription reactions, using SuperScript II (Invitrogen), 6 mM Mn^{2+}, and random nonamer primers (Siegfried et al., 2014; Smola et al., 2015b). Following reverse transcription, Mn^{2+} was removed using G-25 microspin columns (GE Healthcare). Next, second-strand synthesis was performed (NEB), and dsDNA was isolated using a spin column (PureLink micro spin column; Life Technologies).

Library preparation and sequencing

Libraries were prepared using NexteraXT (Illumina) from 1 ng of each second-strand synthesis product. Final libraries were size-selected (AmpureXP beads; Agencourt) with a 1:1 bead to sample ratio (targeting products greater than 200 bp long), and quantified using an Agilent Bioanalyzer 2100 and QuBit high-sensitivity dsDNA assay. For quality control, sequencing was initially performed on a MiSeq. Subsequently, samples were sequenced on an Illumina HiSeq 2500 using version 4 chemistry and 2 × 125 reads. 20-100 million mapped sequencing reads were obtained per experimental conditional (Table S2), with 88% of base calls above Q30.

Translation reporter assays

Gene panel selection

We selected a subset of genes covered by our in-cell SHAPE data that had constant Shine-Dalgarno sequence strength ($-7.5 \leq \Delta G_{hyb} \leq -5.5$ kcal/mol, ΔG_{hyb} calculated as described in RBS–TE correlations) and reasonably well-defined structures around the RBS. Regions consisting of 34-191 nts upstream (mean 90 nts) and 45-231 nts downstream (mean 120 nts) relative to the start codon were then identified for each gene that could be excised with minimum perturbation of the observed endogenous RBS structure (Table S3). These sequences were synthesized with flanking BamHI and HindIII restriction sites and cloned into the pTrc-TE plasmid containing sfCherry and sfGFP under the control of independent Trc promoters (described below), with BamHI and HindIII sites allowing in-frame insertion in front of sfGFP. Gene synthesis and cloning was performed by Genscript.

Plasmid construction

The pTrc-TE plasmid was constructed as follows. A pTrcHis A (Invitrogen V36020) was linearized by PCR using pTrcHis_rev and pTrcHis_for primers (Table S4). Following PCR, plasmid template was digested with DpnI (NEB) and purified using a PCR cleanup kit. Sequences for sfCherry (Kamiyama et al., 2016), a double terminator stem (iGEM part BBa_B0015) with restriction sites, and sfGFP (Pédelacq et al., 2006) were designed with ∼40 nucleotides of overlapping sequence and ordered as geneBlocks (IDT) (Table S4). Linearized backbone and gene blocks were assembled using isothermal assembly (NEB E5520S).

Measurement of GFP expression

Final cloned plasmids containing inserted endogenous leaders were transformed into TOP10 *E. coli* (Invitrogen C404010). Overnight cultures were mixed 1:1 with 50% glycerol, aliquoted in 40 μL volumes to 96-well deep-well plates, and stored at −80°C. TE experiments were initiated by adding 360 μL Terrific Broth supplemented with 50 μg/mL carbenicillin (TB+carb) to thawed plates and growing overnight. These overnight cultures were diluted 1:700 into TB+carb and outgrown for 2 hours, followed by induction of GFP/RFP-expression by addition of 0.2 mM IPTG for 2 hours. After the induction period, aliquots were removed to measure OD_{600} via plate reader and the remaining culture was pelleted by centrifugation at 2000*g* for 10 min at 4°C, resuspended in ice-cold PBS, and immediately forwarded to fluorescence measurement. A Beckman Coulter CytoFLEX flow-cytometer was used to measure at least 10,000 cells, exciting at 488 nM and monitoring at 510 nM (525/40 filter) for GFP, and exciting at 561 nM and monitoring at 610 nM (610/20 filter) for RFP. Data were analyzed in FlowJo, using forward/side-scatter gates to mask debris and

FSC-A/FSC-Width gates to isolate singlet cells. The median RFP and GFP fluorescence was then calculated from the population of RFP positive cells, with normalized GFP (nGFP) computed as the ratio of GFP to RFP. Results represent the average of three biological replicates performed on separate days.

Exclusion of atypical transformants

In total, we made GFP-fusion expression transformants for 53 different genes. However, 21 transformants exhibited severe slow-growth or low RFP fluorescence phenotypes, indicative of cellular toxicity caused by the endogenous 5′ UTR or CDS leader of the fusion gene. We excluded these transformants from further analysis due to the unpredictable effects that toxicity can have on translation. In particular, we excluded transformants where the median fraction of RFP positive cells across three biological replicates was < 0.6, or where the median post-induction OD_{600} was < 0.001. In addition, we excluded three strains that exhibited > 5-fold variability in RFP or GFP fluoresence across replicates. This left the 29 transformants shown in Figure 4.

Electrophoretic mobility shift assays

Protein expression and purification

His_6-tagged genes for the five *E. coli* proteins *rplI* (L9), *rplM* (L13), *rpmB* (L28), *rpmG* (L33), and *rnpA* (C5) were synthesized and cloned into pET-29a(+) vectors using NdeI and XhoI restriction sites (GenScript). C5 contained an N-terminal $MRGSH_6GS$ sequence tag (Rivera-León et al., 1995), while all other proteins contained C-terminal GSH_6 tags. Vectors were transformed into BL21- Arabinose-inducible *E. coli* cells (Invitrogen). For L9, L13, and L28, overnight cultures were used to inoculate Terrific Broth and grown to $OD_{600} = 0.6$ at 37°C, followed by induction for ~16 hours at 18°C with L-arabinose at 0.02% (w/v) and IPTG at 0.1 mM final concentrations. For C5 and L33, ZYM-5052 autoinduction media was inoculated and grown to $OD_{600} = 2.5$, followed by addition of L-arabinose to 0.02% (w/v) and shift to 18°C for ~16 hours. Cultures were collected by centrifugation, resuspended in A_1 Ni-binding buffer (50 mM $NaPO_4$ pH 7.4, 0.5 M NaCl, 40 mM imidazole), lysed by sonication, and clarified by centrifugation at 10,000*g* for 30 minutes at 4°C. Supernatant was mixed and incubated with Nickel-NTA Sepharose-FF beads (GE Healthcare), collected by centrifugation, and washed twice with A_1 binding buffer, twice with A_2 wash buffer (1 × DPBS (GIBCO), 860 mM NaCl, 40 mM imidazole), and twice with A_3 wash buffer (1 × DPBS, 360 mM NaCl, 40 mM imidazole). Washed beads were resuspended in elution buffer (1 × DPBS, 110 mM NaCl, 250 mM Imidazole), followed by centrifugation and removal of the supernatant containing the eluted protein. Millipore Amicon Ultra 0.5 mL 3000 Da filters were used to concentrate and buffer exchange proteins; L9, L13, and L28 were exchanged into 20 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA; and C5 and L33 were exchanged into 20 mM Tris pH 7.5, 500 mM KCl. Concentrations of C5, L13, and L28 were determined by A_{280} with extinction coefficients estimated by Expasy, and the concentrations of L9, and L33 were determined via Bradford assay (ThermoFisher, calibrated using BSA standard). SDS-PAGE indicated purities of > 95% for C5, L9, L13, and L28, and ~80% purity for L33. L13 was stored at 4°C in the final exchange buffer noted above. C5, L9, L28, and L33 were stored at 4°C in the above-noted buffer for several weeks before being diluted into glycerol (50% v/v final glycerol concentration) and stored −20°C.

RNA transcription

DNA oligos for *in vitro* RNA synthesis (IDT; single-stranded oligos or double-stranded gBlocks) were PCR amplified using Q5 hot-start DNA polymerase (NEB) (sequences listed in Table S5). ^{32}P-body-labeled RNAs were synthesized using T7 RNA polymerase (Rio et al., 2011) and $\alpha^{32}P$-ATP, purified by 6% denaturing PAGE, eluted overnight using the crush and soak method, and precipitated with ethanol. RNA concentrations were determined using the Qubit RNA HS assay (Invitrogen).

Binding assays

For binding reactions, RNAs were denatured at 95°C for 2 minutes, cooled on ice for 2 minutes, and then mixed with protein and binding buffer and incubated at 25°C for 40 minutes. Final reaction concentrations were 5 nM ^{32}P-RNA, protein (variable concentrations), 12 mM Tris-HCl (pH 7.5), 0.1 mg/μL yeast tRNA, 0.1 mg/mL BSA, 5 mM DTT, 1 unit/μL recombinant RNasin (Promega), and KCl and $MgCl_2$ optimized for each system. Final salt concentrations were as follows (mM KCl, mM $MgCl_2$): *rplM* RNAs (80, 1); *rpmH* RNAs (80, 1); and *rpmB* RNAs (250, 10). Protein dilutions from glycerol stocks were performed to maintain constant final glycerol concentrations of 4% (v/v) for all binding reactions. For L13, which was not stored in glycerol, the binding buffer was supplemented with 2.5% final (v/v) glycerol. Following equilibration, samples were mixed with glycerol loading dye to 10% final glycerol concentration and immediately loaded onto running native polyacrylamide gels (0.5 × TBE; 0.4-mm × 28.5-cm × 30-cm). 8% (37.5:1 acrylamide: bisacrylamide) gels were used for *rpmB* and *rplM* RNAs and 6% (29:1 acrylamide:bisacrylamide) gels were used for *rpmH* RNAs. Gels were run for 4 hours in a cold room at 720 V, which maintained the gel temperature < 15°C, with at least 1 hour of prerun.

Gel imaging and quantification

Gels were imaged using a GE Healthcare Typhoon Trio phosphoimager, and bands quantified using ImageQuant. K_d values were obtained from fitting to the equation:

$$f = b + \left[\frac{m-b}{1+(K_d/P_t)^n}\right]$$

where *b* and *m* are the upper and lower asymptotes of the fraction of RNA bound, respectively, P_t is the concentration of protein, and *n* is the Hill coefficient. Non-linear least square fits were obtained using the curve_fit module of SciPy in Python. *n* ranged from 1.1-2.6. Reported K_d values represent the average and standard deviation of at least two independent datasets.

QUANTIFICATION AND STATISTICAL ANALYSIS

Read trimming and sequence alignment

Forward and reverse reads were quality trimmed by computing 5-nt averages of the Phred score, trimming at the first 5-nt window with an average Phred score below 20. Reads shorter than 25 nts after trimming were excluded. Trimmed reads were aligned using *Bowtie2* (Langmead and Salzberg, 2012) to the *E. coli* strain MG1655 genome (GenBank: U00096.2). *Bowtie2* alignment was performed in paired-end mode with the following arguments:–local -D 20 -R 3 -N 1 -L 15 -i S,1,0.50 –score-min G,20,8 –ma 2 –mp 6,2 –rdg 5,1 –rfg 5,1 –dpad 100 –maxins 700. Reads not mapping to *E. coli* or with *Bowtie*-reported mapping quality scores below 30 were excluded from analysis.

SHAPE reactivity calculation

Data processing and quality control

Data were processed using the ShapeMapper software (Siegfried et al., 2014). Apparent mutation rates were calculated at each genomic position by summing the number of mismatches and deletions and dividing by the number of reads overlapping the position. Sequence insertions and ambiguously aligned deletions were excluded. Mutations spanning multiple adjacent nucleotides were treated as single mutations at the 3′-most position (Siegfried et al., 2014). Nucleotides with apparent mutation rates above 0.02 in any untreated sample were excluded from analysis. In some genomic regions, we observed clusters of elevated mutation rates that appeared to correspond to local self-complementarity artifacts, possibly arising from PCR. These artifacts were identified as regions of at least 10 nucleotides in which three or more of the 10 nucleotides showed mutation rates above 0.03 in the absence of 1M7 treatment, or modified mutation rates above 0.1 in any condition, and were also excluded from analysis. Except where noted elsewhere, SHAPE reactivities were only computed for nucleotides possessing sequencing depths above 1000 in both modified and untreated samples; nucleotides not passing this filter were treated as ‘‘no data’’ and excluded from analysis. Genes were required to have SHAPE data across 80% of the coding sequence for a given condition to be included in gene-specific analyses.

SHAPE reactivity normalization

SHAPE reactivities were calculated as the difference in mutation rates between 1M7-modified and untreated samples. Reactivities were normalized within each probing condition to the mean of the 92-98th percentile reactivities of nucleotides from the ncRNAs RNase P, tmRNA, and 6S RNA, as these ncRNAs were sequenced to high depths and showed few changes across experimental probing conditions.

Calculation of gene median SHAPE

In Figure 1B, medians were computed over all coding sequence nucleotides with defined SHAPE reactivities. In Figure 2A, medians were computed over the region 30 nucleotides 5′ of the start codon to 30 nucleotides 5′ of the stop codon; this captures potential SHAPE reactivity changes associated with translation initiation at the considered gene while excluding changes attributable to translation initiation at neighboring genes.

Coding region aperiodicity

Previous transcriptome-wide structure-probing experiments in *E. coli*, yeast, and mammalian cells have been interpreted to indicate that mRNA coding regions exhibit periodic reactivity profiles (Del Campo et al., 2015; Ding et al., 2014; Spitale et al., 2015; Wan et al., 2014).

To provide the best comparison to these prior studies, we collectively averaged over all internal coding region 99-nt windows with at least 60% SHAPE data coverage, aligning to preserve a common reading frame. This meta-gene analysis revealed that coding regions have aperiodic SHAPE reactivity profiles in both cell-free and in-cell conditions (Figure S2). There are several potential explanations for this discrepancy. First, prior studies of *E. coli* relied on enzymatic reagents with known sequence biases (Del Campo et al., 2015). Given that coding regions have inherently periodic sequences, periodic structure-probing signal may be a consequence of sequence bias. The 1M7 SHAPE reagent by contrast has minimal sequence bias. Second, prior studies relied on detecting truncated RNA fragments via a ligation-based library preparation strategy. Such strategies introduce sequence biases that are avoided by the SHAPE-MaP strategy (Smola et al., 2015a; 2015b; Weeks, 2015). Third, in truncation based detection strategies, any cellular or experimental process that generates truncated or degraded RNA fragments will give artifactual signals. For example, cotranslational decay in yeast yields intermediates consistent with the periodicity observed in structure-probing experiments (Pelechano et al., 2015). Since SHAPE-MaP detects 1M7 modifications as mutations within continuous RNA sequences, such artifacts are avoided. Fourth, previous *E. coli* probing experiments were performed on *in vitro* refolded RNAs, compared to the natively extracted cell-free and in-cell conditions used here. Thus, differences in experimental conditions may contribute to these discordant observations.

Transcript boundary assignment

General strategy

Our SHAPE data represent averages over all transcript isoforms and thus primarily report on the structure of the most highly expressed isoforms. We used manual analysis of the sequencing depths observed in the in-cell dataset to determine the transcript isoform most consistent with the expression observed at each genomic locus. Hallmark signs of consistent read-depth across a transcript with drop-offs near the transcript boundaries were cross-referenced to *E. coli* transcript annotations compiled from high-throughput end-mapping experiments (Conway et al., 2014). For transcripts without clear dominant transcription start sites, we assigned the transcription start site to the most distal, significantly expressed transcript. Transcripts with internal terminators were modeled as the read-through product if the read-depth of downstream genes was sufficient for accurate calculation of SHAPE reactivities.

Treatment of dominant internal promoters

Several operons exhibited expression profiles consistent with dominant expression of a ''short'' transcript from an internal start site and lesser, although significant, expression of a ''long'' transcript from a start site in front of upstream genes. In such cases, the genes downstream of the internal start site were assigned to the short transcript isoform, and the long isoform was truncated to include only the upstream genes. To prevent structure modeling errors associated with using an artificial 3′ boundary, the long isoform was modeled with a 3′ UTR that extended 600 nts past the stop codon of the last gene, or if closer, to the natural transcript termini.

Treatment of annotation-inconsistent transcripts

Approximately 20% of loci had expression profiles inconsistent with any annotated transcription unit (Conway et al., 2014). We searched regulonDB (Salgado et al., 2013) for alternative start sites and/or terminator annotations that better fit the observed expression and found annotations for the majority of such loci. Visual analysis was used to estimate transcript boundaries for the remaining 8% of loci with unannotated transcription start or termination sites.

Secondary structure modeling and analysis

Modeling methodology

While nucleotides possessing < 1000 read-depth were otherwise excluded from SHAPE analyses, for the purposes of structure modeling we included SHAPE reactivities for all nucleotides possessing sequencing depths of > 350 in both the modified and untreated samples. This choice was made to minimize regions with no data near transcript boundaries and is justified by prior studies showing that SHAPE reactivities computed from as few as 200 reads provide useful information for guiding secondary structure prediction (Siegfried et al., 2014). Minimum free energy secondary structures and base pairing probabilities were generated for each mRNA transcript using the SuperFold algorithm and SHAPE reactivities as restraints (Smola et al., 2015b). SuperFold uses a windowing approach to fold large RNAs. First, partition function calculations are performed for overlapping windows and are merged, yielding transcript-wide base-pairing probabilities and base-pairing (Shannon) entropies. The minimum free energy structure is then predicted in sliding windows, constrained by highly probable pairs observed in the merged partition function. Partition function and minimum free energy calculations were performed using RNAstructure (v 5.8) (Reuter and Mathews, 2010). SuperFold parameters were as follows: SHAPEslope = 1.8, SHAPEintercept = −0.6, trimInterior = 300, partitionWindowSize = 1500, partitionStepSize = 100, foldWindowSize = 3000, foldStepSize = 300, maxPairingDist = 500. ''No-data'' models were generated using the same SuperFold parameters, but setting SHAPE reactivities to −999 (equivalent to NaN).

Cross-condition comparisons of MFE structures

Analysis for a given condition was limited to transcripts possessing > 80% SHAPE data coverage (defined using > 1000 read-depth threshold); due to varying read-depths in different samples, the number of transcripts passing this threshold varied from 59 to 157 (194 transcripts have at least one coding sequence with 80% data coverage in one sample). Comparisons between minimum free energy (MFE) structures indicated that in-cell, cell-free, and kasugamycin transcript models share on average ∼60% of base pairs (Figure S3A). A larger fraction of in-cell pairs are shared with cell-free structures than vice versa. This asymmetry arises from the higher number of base pairs in cell-free models (Figure S3B); in cells, translation likely disrupts weak base pairs. Supporting this interpretation, structures in kasugamycin-treated cells have more base pairs than in-cell structures but fewer than cell-free structures (Figure S3B). Note that the apparent increased similarity in Figure S3A between kasugamycin and cell-free structures, and kasugamycin and in-cell structures is misleading, and arises from the smaller number of transcripts with SHAPE data in the kasugamycin condition. In-cell, kasugamycin, and cell-free structures shared comparable fractions of base pairs when analysis was limited to the same subset of transcripts.

Structural variation in dynamic regions

RNAs with poorly-defined dynamic structures can form multiple structures with similar free energies as the MFE structure, which can cause structure modeling to be artificially sensitive to insignificant differences in SHAPE data. We therefore repeated our analysis considering only well-defined base pairs (pairing probability > 0.9). Shown in Figure S3D, 25%–30% of nucleotides participate in well-defined base pairs, representing ∼50% of the base pairs in each MFE structure. Again consistent with translation destabilizing RNA structure, there are fewer high probability pairs in in-cell than in cell-free models. Notably, high-probability pairs are more likely than MFE pairs to be shared between conditions (> 70% of in-cell p > 0.9 pairs are also observed in cell-free models; Figure S3C). As a complementary analysis, we also analyzed how MFE structure similarity varies as a function of base-pairing entropy (a measure of how well-defined a structure is). Similarity between models is strongly anticorrelated with base-pairing entropy (Figures S3E and S3F). Together, these analyses indicate that differences between in-cell and cell-free structures are primarily localized to poorly defined regions. Some of these differences are caused by ribosome-induced unfolding in cells, which reduces the overall number of base pairs observed in cells. However, in-cell and kasugamycin models differ from each other to similar degrees as they differ with respect to cell-free models, implying that the cellular environment does not induce large-scale changes in RNA structure.

Calculation of $\Delta G^{\ddagger}_{unfold}$ and ΔG_{unfold}

General strategy

We tested four different models of the RBS unfolding process that occurs during mRNA accommodation into the 30S subunit (Figure S4A). Equilibrium versus non-equilibrium unfolding allows versus disallows the mRNA molecule to refold to a new minimum free energy structure after unfolding of the RBS. Local versus complete unfolding allows versus disallows base pairs spanning the unfolded RBS window. For all four models, ΔG_{unfold} was computed as

$$\Delta G_{unfold} = \Delta G_{cons} - \Delta G_{ref}$$

ΔG_{ref} is the free energy of the reference SHAPE-directed transcript structure. ΔG_{cons} is the free energy of the "constrained" transcript structure with the RBS window constrained as single-stranded. For complete unfolding, the constrained structure was also prevented from having base pairs spanning the RBS window. All ΔG calculations were performed using the *efn2* command of RNAstructure (excluding SHAPE pseudo-energies) (Reuter and Mathews, 2010).

Calculation of $\Delta G^{\ddagger}_{unfold}$

The non-equilibrium ΔG_{unfold} is assumed to correspond to the unfolding transition state free energy, referred to as $\Delta G^{\ddagger}_{unfold}$ throughout the text. For these calculations, the SuperFold minimum free energy transcript structure was used as the reference structure. The constrained structure was obtained by deleting base pairs involving the RBS window from the reference. In the case of complete unfolding, all base pairs spanning the RBS window were also deleted.

Calculation of ΔG_{unfold}

Equilibrium ΔG_{unfold} calculations required computing new sets of structure models. The reference structure for each gene was obtained by folding up to a 1500-nt subsequence centered on the start codon. For genes with start codons < 750 nts from either the 5′ or 3′ transcript boundary, the subsequence extended from the proximal boundary up to 1500 nts, or to the distal boundary. For the local unfolding scenario, the constrained structure was obtained by refolding the same subsequence with the RBS constrained as single stranded. For the complete unfolding scenario, the subsequence was refolded in two segments (5′ and 3′ to the RBS window) to prevent RBS-spanning pairs; ΔG_{cons} was then obtained by summing the ΔG computed for each segment. These folding calculations were performed using the *Fold* command of RNAstructure with parameters –mfe –md 500 –si –0.6 –sm 1.8 and the same SHAPE restraints as used with SuperFold.

RBS–TE correlations

Gene inclusion criteria

Downstream genes in polycistronic transcripts with TE (Li et al., 2014) within 2-fold of the TE of the immediate upstream gene were classified as potentially translationally coupled and excluded (Figure 3A). Analysis was restricted to genes possessing SHAPE data for > 80% of nucleotides in the 200-nt window centered around the start codon. If the gene start codon was less than 100 nts from the transcript boundary this window extended from the boundary to 100 nts upstream of the start codon. Genes with non-canonical start codons (not AUG or GUG) or lacking Shine-Dalgarno sequences were excluded. Shine-Dalgarno sequences were assessed by computing the hybridization free energy ΔG_{hyb} between the 16S rRNA anti-Shine-Dalgarno sequence CACCUCCU and the gene subsequence from -16 to -3 relative to the start codon. Genes with valid Shine-Dalgarno sequences were defined as having $\Delta G_{hyb} \leq 0$, with the terminal Shine-Dalgarno/anti-Shine-Dalgarno base pair located within the interval [-10, -4] relative to the gene start. ΔG_{hyb} calculations were performed using RNAstructure (Bifold –i).

Investigation of different RBS unfolding models

Correlations were computed between TE and the local and complete equilibrium energy of unfolding (ΔG_{unfold}), and local and complete non-equilibrium energy of unfolding ($\Delta G^{\ddagger}_{unfold}$) for varied sized windows around the RBS. As shown in Figure S4B, local $\Delta G^{\ddagger}_{unfold}$ was strongly correlated with TE for RBS windows between 30-nt to 50-nt in size (r < –0.6). Supporting that this strong correlation reflects a true physical unfolding process, the correlation between $\Delta G^{\ddagger}_{unfold}$ and TE markedly decreased once the RBS window was expanded beyond the anticipated physical unfolding window (beyond -25 or +25 of the gene start). By contrast, the correlations between TE and complete $\Delta G^{\ddagger}_{unfold}$, local ΔG_{unfold}, and complete ΔG_{unfold} are significantly weaker for all analyzed windows ($r \leq -0.4$). For other analyses, RBS $\Delta G^{\ddagger}_{unfold}$ and RBS ΔG_{unfold} were taken to be the local $\Delta G^{\ddagger}_{unfold}$ and ΔG_{unfold}, respectively, computed for the window [-25, +25] around the start codon. All linear regressions were computed using the stats.linregress function of ScipPy in Python.

Analysis limitations

In general, SHAPE-directed structure models are more accurate in modeling short-range than long-range base pairs. Thus, the reduced correlation between complete $\Delta G^{\ddagger}_{unfold}$ and TE compared to local $\Delta G^{\ddagger}_{unfold}$ may be a consequence of including lower-accuracy long-range base pairs in the complete unfolding calculation. Additionally, our equilibrium ΔG_{unfold} calculations are compromised by the necessary assumption that our SHAPE data can be used to model the RBS-unfolded structure.

Comparison to prior studies of synthetic genes

Studies of overexpressed synthetic genes have observed comparable ($r \approx -0.6$) correlations between TE and complete equilibrium RBS unfolding (complete ΔG_{unfold}) as we observe between TE and $\Delta G^{\ddagger}_{unfold}$ for native genes (Espah Borujeni et al., 2014; Goodman et al., 2013; Kudla et al., 2009; Salis et al., 2009). The similar observed correlations suggest a common mechanism despite the important conceptual difference between $\Delta G_{\mathrm{unfold}}$ and $\Delta G^{\ddagger}_{\mathrm{unfold}}$. We note that several features made these prior experiments insensitive to differences between kinetic and equilibrium mechanisms. First, previously studied mRNAs were engineered to have well-defined, modular secondary structures. Consequently, RBS unfolding is unlikely to promote refolding of adjacent mRNA sequences, rendering equilibrium $\Delta G_{\mathrm{unfold}}$ and non-equilibrium $\Delta G^{\ddagger}_{\mathrm{unfold}}$ equivalent. Second, the studied mRNAs had only short-range pairing interactions, and thus were insensitive to differences between complete versus local unfolding. Finally, the kinetic mechanism is based on the assumption that individual mRNA species comprise a small fraction of the total cellular mRNA. This assumption may be violated for highly overexpressed mRNAs, which are more likely to be at or near equilibrium with the pool of free 30S subunits.

CDS–TE correlations

The local $\Delta G^{\ddagger}_{\mathrm{unfold}}$ (or $\Delta G_{\mathrm{unfold}}$) was computed for 50-nt windows across the CDS, relative to the start or stop codon of each gene, using the same methodology described above for the RBS. For windows relative to the start codon (5′ CDS), the regression of ln(TE) on $\Delta G^{\ddagger}_{\mathrm{unfold}}$ (or $\Delta G_{\mathrm{unfold}}$) was computed for the same genes as used for RBS regressions, with the additional restriction that genes must be > 200 nts long (in-cell N = 150; kasugamycin N = 102; cell-free N = 120). For windows relative to the stop codon (3′ CDS), regressions were computed for genes passing the same start codon, Shine-Dalgarno strength, translational coupling, and > 200 nt length filters, while requiring SHAPE data for > 80% of nucleotides in the 200-nt window centered around the stop codon (in-cell N = 155; kasugamycin N = 92; cell-free N = 173). Linear regressions and significance were computed using the stats.linregress function of ScipPy in Python.

Translational coupling analysis

The number of gene-linking pairs (*LP*) between a given gene and its upstream neighbor was computed from the base pairing partition function as:

$$LP = \sum_{i=1}^{t_u} \sum_{j>i}^{t_t} p(i,j) \cdot I_A\left(j \geq s_g - w\right)$$

where $p(i, j)$ is the pairing probability between positions i and j, I_A is the indicator function, s_g is the position of the gene start, and t_u and t_t are the termini of the upstream gene and the transcript, respectively. The w parameter specifies the size of the included RBS window (for example, pairs linking the Shine-Dalgarno sequence to the upstream gene are included in *LP*). We used $w = 25$, matching the RBS window size used elsewhere in the text. A similar trend of decreasing TE variation with *LP* was observed for different w values. Mechanistic considerations distinct from potential structural coupling make translational coupling unlikely between genes separated by very long intergenic regions, or between genes with significantly overlapping coding sequences. Therefore, to prevent such genes from skewing analyses, we limited our analysis to genes where $-5 < s_g\text{-}t_u < 100$, but comparable results were obtained when the analysis was applied to all genes. Analysis was restricted to genes that had SHAPE data for > 80% of nts in the 200-nt window centered around the gene start.

Automated motif detection

Algorithm description

We built on a previously described strategy for identifying well-structured motifs in large RNA molecules (Figure 6A) (Siegfried et al., 2014; Smola et al., 2015b). Local median SHAPE reactivity and entropy were computed over centered, sliding 51-nt windows using the cell-free dataset. At boundaries, local medians were computed using all nucleotides within ± 26 nts of the considered position (for example, for a window centered on nucleotide 10, the median was computed using nucleotides [1, 36]). At least 26 nts were required to have SHAPE data in order to compute a valid local median. Well-structured regions were identified as regions where the local median SHAPE fell below 0.3 and median entropy fell below 0.04 for more than 25 contiguous nucleotides. These regions were then expanded by up to 50 nts on either side to incorporate nested structures with pairing probability (pp) > 0.9. To confirm identified structures also existed in cells, > 95% of cell-free pp > 0.9 base pairs were required to have pp > 0.3 in-cell. If this 95% cutoff was not satisfied, the region was trimmed to the maximal sub-region meeting this requirement. Finally, all nucleotides with pp < 0.5 were trimmed from the 5′ and 3′ ends. Final trimmed consensus regions that were shorter than 25 nts or possessed < 80% cell-free or in-cell SHAPE data coverage were rejected. Following automated identification, each motif was visually inspected and in some cases manually adjusted to include (or exclude) adjacent structures that were judged to be part of (or distinct from) the algorithmically identified structure.

Our use of fixed-value SHAPE and base-pairing entropy cutoffs differs modestly from our previously described algorithm, where regions were identified from comparisons to the global medians of SHAPE and entropy (Siegfried et al., 2014; Smola et al., 2015b). Fixed-value cutoffs are required for analyzing RNAs that are potentially poorly structured overall (such as the mRNAs analyzed here), or, conversely, those that are highly structured overall (such as structured ncRNAs). The 0.3 SHAPE cutoff corresponds to the maximum median reactivity expected of paired nucleotides, and the 0.04 base-pairing entropy cutoff corresponds to a pp ≈0.95.

UTR/IGR data coverage criteria

We limited our search to UTRs and IGRs > 25 nts long, and which contained at least one 25-nt stretch with 75% SHAPE coverage in both the in-cell and cell-free datasets. Annotated REP (Keseler et al., 2013) and ERIC (Wilson and Sharp, 2006) motifs were masked out.

Sensitivity of detecting known motifs

We compiled a list of all *E. coli* RFAM (Nawrocki et al., 2015) motifs and known RAREs (Aseev et al., 2015; Fu et al., 2013; 2014; Matelska et al., 2013). Structures identified from our *de novo* structure models were considered "true positives" if they recapitulated any portion of the known structure. Thirteen of these known motifs fall within UTR/IGRs passing our length and data coverage filters, and of these thirteen, we positively identified nine, corresponding to a sensitivity of 69%. The four known motifs we failed to identify were the *rplK*, *rpsO*, *rpsF*, and *rplY* RAREs. We note that our motif search also identified the so-called Pseudomonas sRNA P26 motif listed in RFAM (named intergenic *rplL-rpoB* motif in Table S1). Despite its entry in RFAM, we determined that this motif is better described as "functionally uncharacterized" due to a lack of validation (see Table S1), and therefore excluded this motif from our sensitivity calculations. If we include this motif in our sensitivity calculations, we detect 10 out of 14 (71%) of known RFAM and RARE motifs.

Comparisons with prior comparative genomics predictions

We compared the UTR/IGR motifs identified here against prior comparative genomics and bioinformatics predictions of functional RNAs (Livny et al., 2008; Ott et al., 2012; Pichon et al., 2012; Rivas et al., 2001; Tran et al., 2009; Uzilov et al., 2006). Several of these algorithms were optimized to predict small RNA genes rather than functional UTR/IGR motifs, but were nonetheless included for completeness. The study by Uzilov et al. includes predictions made using three algorithms: Dynalign (Uzilov et al., 2006), QRNA (Rivas and Eddy, 2001), and RNAz (Washietl et al., 2005); comparisons were performed to all three sets of predictions (requiring $p > 0.9$ for Dynalign and RNAz). Motifs were considered "previously predicted" if they overlapped a predicted functional loci by at least 50 nts and were located on the same strand (if specified).

Motif conservation analysis

Algorithm for identifying homologs

We constructed an automated pipeline to search for motif homologs in other bacterial genomes (Figure S7). Similar to other comparative genomics pipelines (Slinger et al., 2014; Yao et al., 2007), we use iterative Infernal (v1.1.1) (Nawrocki and Eddy, 2013) searches to train a covariation model (CM) constructed from a single input *E. coli* structure. The initial CM was built and calibrated from a Stockholm file containing the *E. coli* sequence and base pairs (cmbuild–F; cmcalibrate). cmsearch was performed against a non-redundant bacterial genome database using a lenient e-value cutoff of 1.0 (cmsearch –incE 1.0 –mid –cpu 8). The genetic context of each identified homolog was cross-referenced to *E. coli*, filtering out homologs found in different contexts or at unannotated loci. The filtered homologs were then aligned (cmalign –cpu 8 –noprob) and used to construct a new CM. This process was repeated a total of three times, yielding a "trained" CM. The trained CM was then used to perform a final search against the bacterial database using a e-value cutoff of 0.01.

Homolog genetic context filtering

Genetic context filtering was performed using RefSeq annotations (Tatusova et al., 2014). The "transcript" of each homolog was inferred by first identifying adjacent same-strand genes within 400 nts. The "transcript" was then extended from both directions to incorporate additional same-strand genes, allowing a maximum intergenic distance of 400 nts. These genes were then cross-referenced against the genes of the parent *E. coli* transcript, defining shared context as at least one common gene between the two transcripts. Cross-referencing was performed using both gene names and products: names were cross-referenced using gene and gene_synonym fields; products were cross-referenced using manually specified keywords.

Bacterial genetic database details

The genomic database was constructed by downloading the RefSeq (Tatusova et al., 2014) bacterial genome assembly summary from ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria/assembly_summary.txt (on October 19, 2015). Genomes that were not "latest," "Complete Genome," "reference," or "representative" were discarded. From the remaining genomes, a single genome was chosen for each species and downloaded from the NCBI genome ftp (also on October 19, 2015). Reference genomes were prioritized over representative genomes. For species with multiple reference genomes, or multiple representative but no reference, the last listed genome was used.

Consensus motif analysis

The homologs returned from our algorithmic search were manually assessed for context specificity and phylogenetic diversity. For the large majority of motifs, the search procedure returned homologs with 100% context-specificity and reasonable structure conservation. Our homolog searches for identified ribosomal protein autoregulatory motifs provide strong positive controls, yielding consensus structures and phylogenetic diversities comparable to prior studies (Table S1) (Fu et al., 2013). However, in some cases, searches using the trained CM returned homologs with poor context or secondary structure conservation. This was attributable to either progressive loss of CM specificity during the refinement stage, or for small motifs, low information content of the original motif. These cases are noted in Table S1, and were excluded from downstream conservation and consensus structure analysis. R2R (Weinberg and Breaker, 2011) was used to draw consensus structure diagrams and assess secondary structure conservation (–GSC-weighted-consensus 3 0.97 0.9 0.75 4 0.97 0.9 0.75 0.5 0.1).

Conservation calculation

Conservation within enterobacteria was computed as the number of enterobacterial homologs identified divided by 32, the total number of enterobacteria in our database. The endosymbionts *Wigglesworthia glossinidia* and *Buchnera aphidicola* were excluded from conservation calculations.

DATA AND SOFTWARE AVAILABILITY

The accession number for the raw sequencing reads from SHAPE experiments reported in this paper is ENA: PRJEB23974 (https://www.ebi.ac.uk/ena/data/view/PRJEB23974).

Processed SHAPE data, RNA structure models, Python code used to perform automated low-SHAPE/low-entropy motif detection, and Python code to perform automated homology searches are freely available at the Lead Contact's webpage, http://www.chem.unc.edu/rna/.

Supplemental Figures

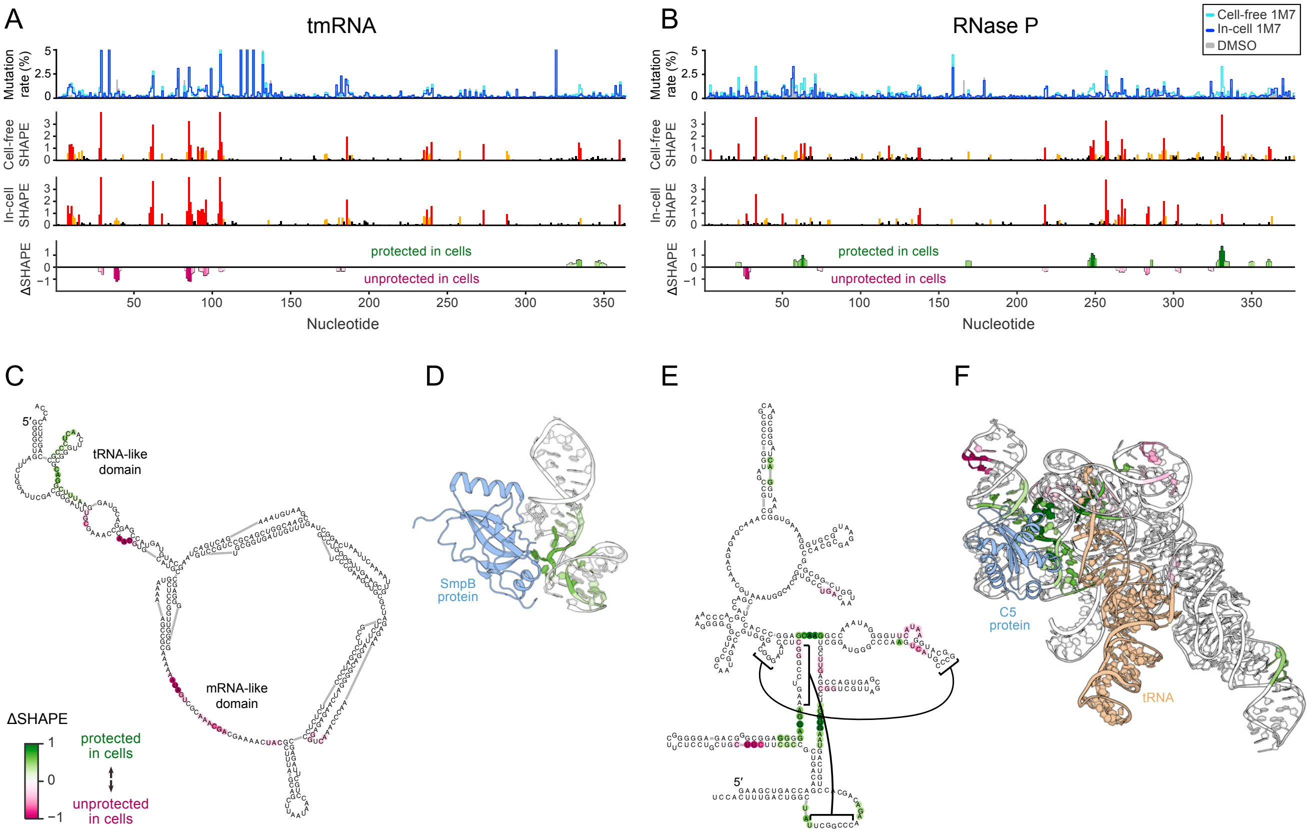


Figure S1. In-Cell SHAPE Resolves Protein Binding Sites in Non-coding RNAs, Related to Figure 1

(A and B) Raw modification profiles and resultant SHAPE profiles of tmRNA and RNase P probed under in-cell and cell-free conditions. DMSO no-reagent control samples were collected for both cell-free and in-cell conditions, but for simplicity, only a single DMSO profile is shown. Smoothed SHAPE reactivity differences were calculated using the ΔSHAPE framework (Smola et al., 2015a).

(C) SHAPE reactivity changes mapped on the *E. coli* tmRNA secondary structure.

(D) SHAPE reactivity changes mapped on the crystal structure of the tRNA-like domain of *A. aeolicus* tmRNA (PDB 1P6V). In-cell SHAPE reactivity protections (green) correspond closely with the SmpB binding site.

(E) SHAPE reactivity changes mapped on the *E. coli* RNase P RNA secondary structure.

(F) SHAPE reactivity changes mapped on the crystal structure of *T. maritima* RNase P (PDB 3QIQ). In-cell SHAPE reactivity protections (green) correspond closely with C5 protein and tRNA binding sites.

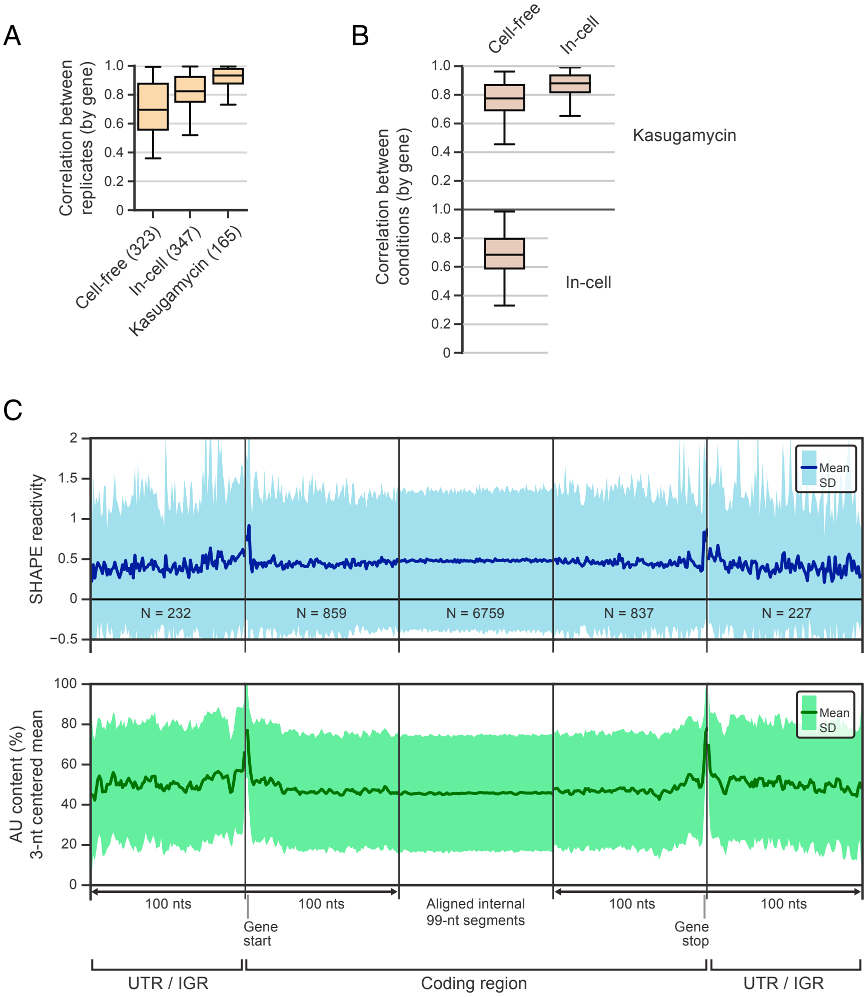


Figure S2. Reproducibility and Meta-gene Analysis of SHAPE Reactivity, Related to Figure 1

(A) Per-gene Pearson correlation between SHAPE profiles across biological replicates. Medians are denoted by black bisecting lines, boxes indicate the interquartile range (IQR), and whiskers indicate data within 1.5 × IQR of the top and bottom quartiles.

(B) Per-gene Pearson correlation between SHAPE profiles across experimental conditions.

(C) Meta-gene analysis of cell-free SHAPE reactivity provides little information on the structure of individual mRNAs, but indicates that coding regions do not have periodic structures (top; see also STAR Methods). Note that changes in average SHAPE reactivity are much smaller than the per-nucleotide standard deviation. Note also that the increased SHAPE reactivity observed at the meta-gene start and stop codons mirror AU-sequence biases (bottom). Averaging was performed transcriptome-wide, including all 100-nt windows with at least 60% cell-free SHAPE data coverage irrespective of whether the parent transcript had sufficient full-length SHAPE coverage for other analyses. Hence, this analysis reflects a larger pool of genes, and is comparable in makeup to other transcriptome-wide studies. The number of windows used for each average is denoted.

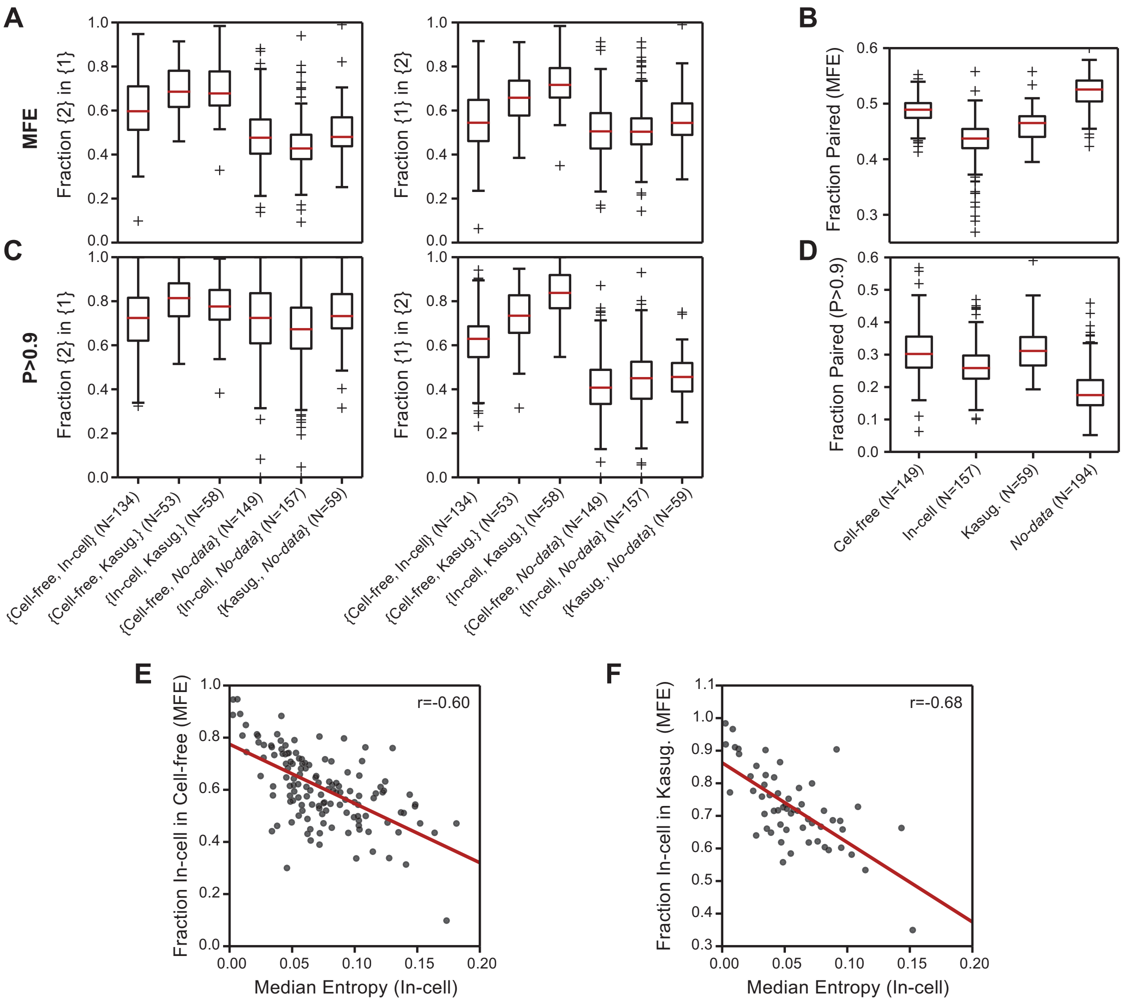


Figure S3. Comparison between SHAPE-Directed and No-Data Structure Models, Related to Figure 2

(A) Similarity between MFE structure models for each transcript. Comparisons were performed by computing the fraction of base pairs shared between the first and second structures and vice versa (first and second correspond to order listed on x axis). These fractions correspond to positive predictive value (ppv) and sensitivity, respectively, which are conventionally used when comparing structure models to known references.

(B) Fraction of nucleotides that are base paired in MFE structures for different conditions.

(C) Similarity between the set of highly probable (p > 0.9) base pairs for each condition. Comparisons were performed as described in panel A.

(D) Fraction of nucleotides paired with p > 0.9 under different conditions. In panels A-D, medians are denoted by red bisecting lines, boxes indicate the IQR, whiskers indicate data within 1.5 × IQR of the top and bottom quartiles, and outliers are indicated by crosses.

(E) Correlation between base-pairing entropy and the fraction of MFE pairs shared between in-cell and cell-free models. High entropy indicates structures are poorly defined.

(F) Correlation between base-pairing entropy and the fraction of MFE pairs shared between in-cell and kasugamycin models.

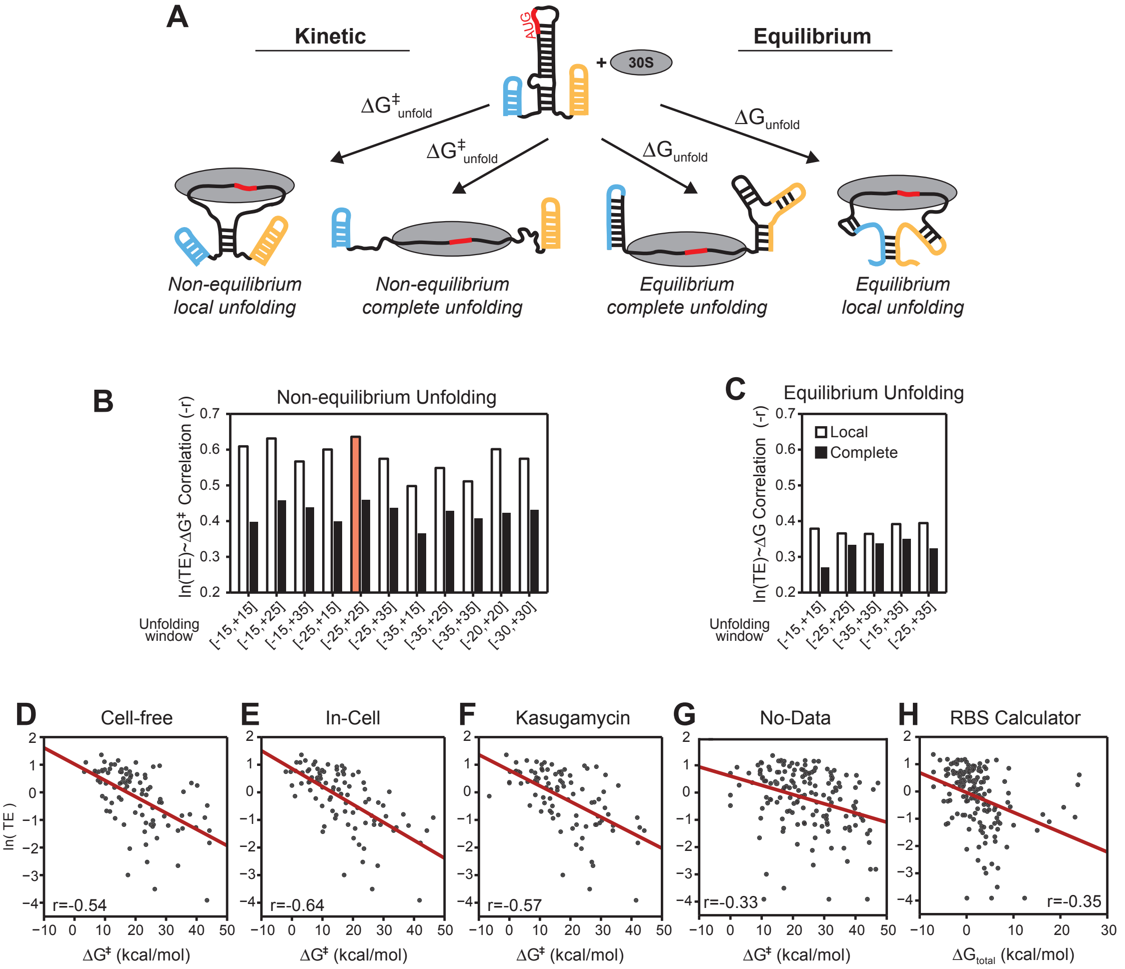


Figure S4. Correlation between TE and ΔG_{unfold} and $\Delta G^{\ddagger}_{unfold}$, Related to Figure 3

(A) Scheme illustrating different models of mRNA accommodation into the 30S subunit. For equilibrium calculations, the mRNA molecule is allowed to refold to a new minimum free energy structure after unfolding the RBS, but not in non-equilibrium (kinetic) calculations. Local versus complete unfolding allows versus disallows base pairs across the RBS window. Non-equilibrium unfolding energies are assumed to correspond to $\Delta G^{\ddagger}_{unfold}$, the free energy of the unfolding transition state (see STAR Methods).

(B and C) Correlation coefficients computed using different sized windows for local (filled bars) and complete (open bars) RBS unfolding models. Correlations were computed using in-cell structures, excluding potential translationally coupled genes (n = 157). In panel B, red shading indicates the model used for all remaining analyses.

(D–F) Correlation between TE and local $\Delta G^{\ddagger}_{unfold}$ for the three probing conditions. To facilitate direct comparison, we only show genes that possess sufficient data coverage in all three SHAPE probing conditions (n = 92).

(G) Correlation between TE and local $\Delta G^{\ddagger}_{unfold}$ computed from “no-data” structure models.

(H) Correlation between TE and ΔG_{total} predicted by the RBS calculator (v1.0), a representative thermodynamics-based TE calculator (Salis et al., 2009). Analyses in panels G and H were performed on genes possessing in-cell SHAPE data (n = 157) and thus can be directly compared to Figure 3C.

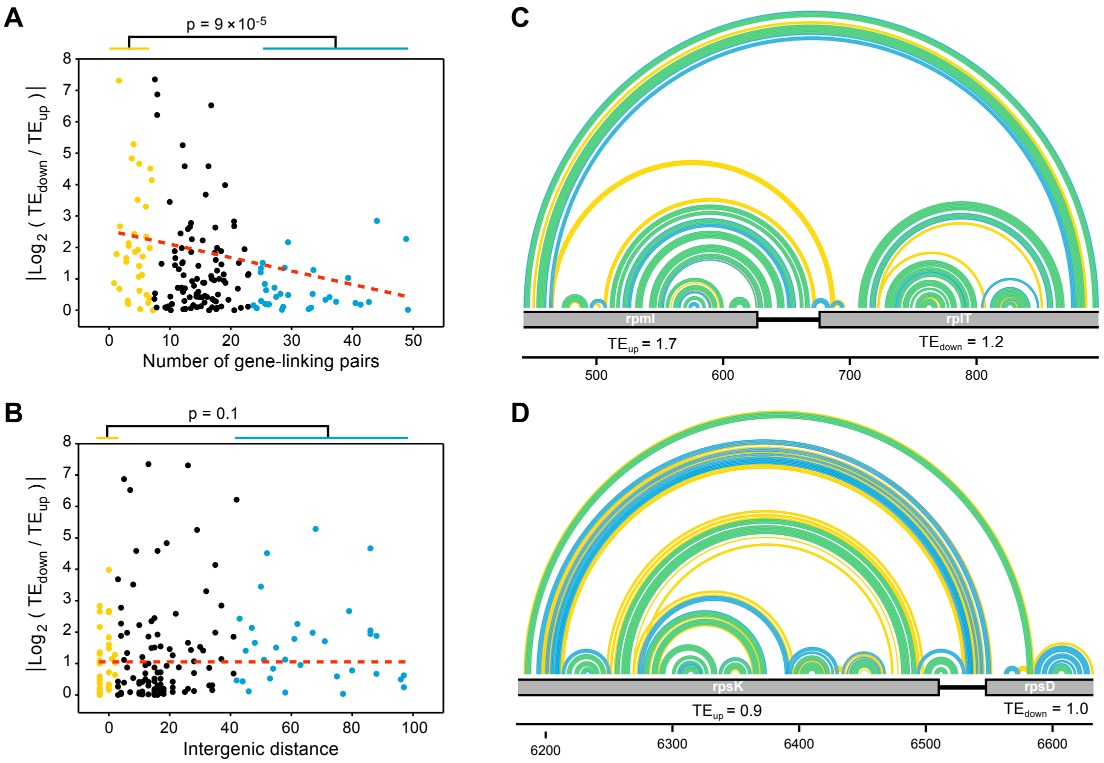


Figure S5. RNA Structure Couples Translation of Adjacent Genes, Related to Figure 5

(A) Relationship between the TE ratio of adjacent genes as a function of the number base pairs linking the genes. Bottom and top quintiles are shown in yellow and blue, respectively; these quintiles correspond to the "few" and "many" linking-pairs categories in Figure 5. The red dashed line highlights the consistent decrease in TE variability as genes are linked by more base pairs.

(B) Relationship between TE of adjacent genes as a function of the length of the intervening intergenic region. This analysis shows clearly that translational coupling is not a simple function of intergenic distance. Top and bottom quintiles are shown as in (A). Statistical significance between the top and bottom quintiles is indicated above (A) and (B) and was tested using two-tailed Mann-Whitney U-tests.

(C and D) Examples of structure-mediated translational coupling over long intergenic regions. The *rpmI-rplT* IGR is 53-nt long, and the *rpsK-rpsD* IGR is 34-nt long. Structures are shown as pairing probability arcs (key shown in Figure 3).

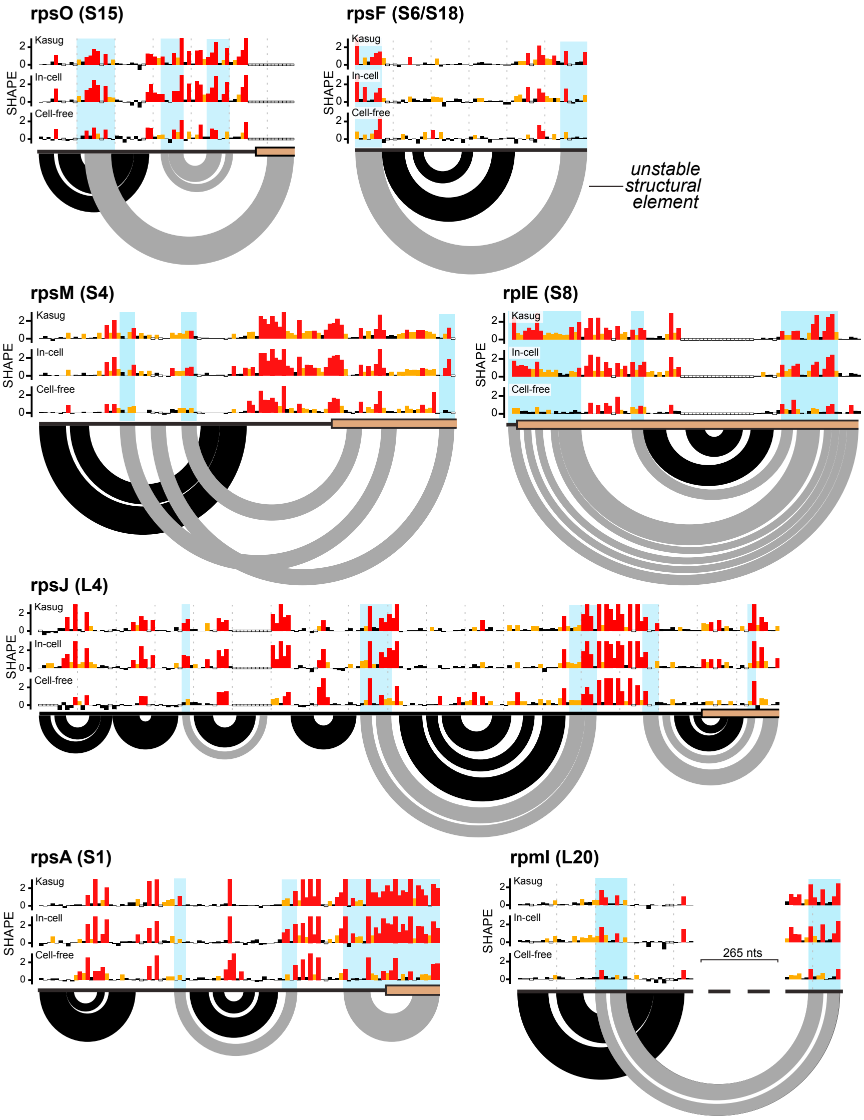


Figure S6. SHAPE Data Reveal that Many Known RARE Structures Are Unstable in the Absence of Bound Protein, Related to Figure 6

Motifs are labeled by downstream gene, with the ribosomal protein ligand listed in parentheses. Accepted functional structures and SHAPE data are shown for each motif (Aseev et al., 2015; Fu et al., 2013; 2014; Matelska et al., 2013). Regions where SHAPE data are inconsistent with an accepted structure are highlighted with light blue shading, and corresponding unstable structural elements are shown using gray arcs. Brown boxes indicate coding sequences. Note that the *rplE* motif is located entirely within the *rplE* coding region, and thus was not included in our automated motif search.

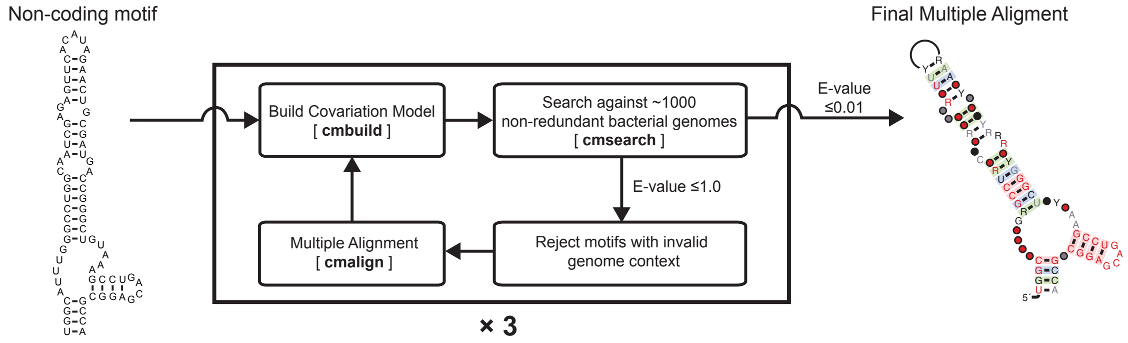


Figure S7. Outline of the Homolog Search Strategy, Related to STAR Methods

Each *E. coli* structure was used to build an Infernal (Nawrocki and Eddy, 2013) covariation model. The initial model was refined three times by incorporating additional homologs identified in similar genetic contexts. The trained covariation model was then used to perform a final search, with returned homologs used to construct consensus structures using R2R (Weinberg and Breaker, 2011).

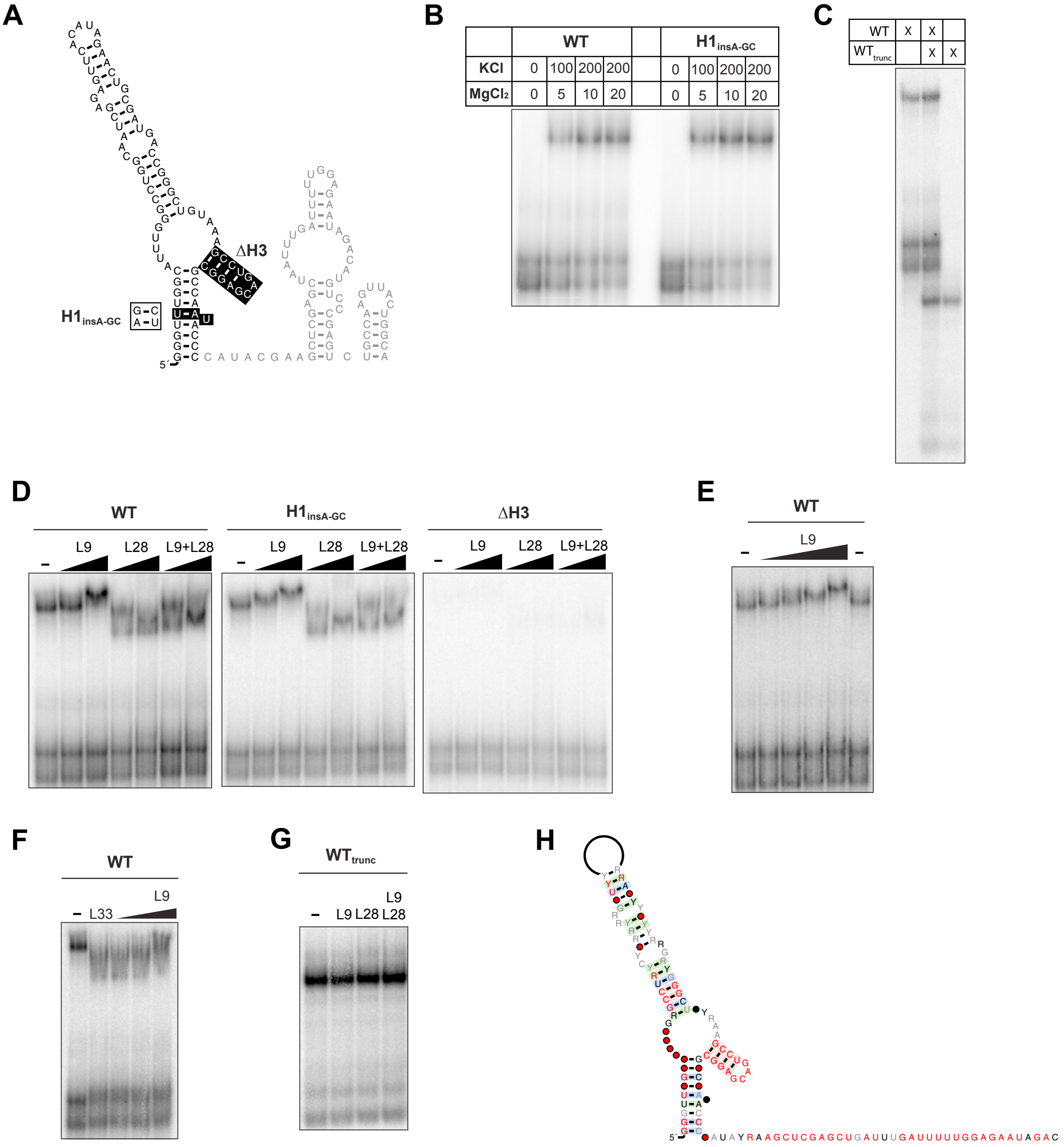


Figure S8. The *rpmB* 5′ UTR Binds L9 and L28, Related to Figure 6

(A) Constructs used. Alterations made in variant constructs are highlighted in black. In the stabilizing H1$_{insA-GC}$ construct, an A is inserted to pair with the bulged U, and the neighboring AU pair is changed to a GC. The WT$_{trunc}$ construct contains only the three-way junction (truncated nucleotides are drawn in gray).

(B) The low-mobility conformation of the *rpmB* 5′ UTR is salt-dependent, and is stabilized by the H1$_{insA-GC}$ mutation. Quantification indicates that 64% of H1$_{insA-GC}$ RNA is in the low-mobility conformation at 200 mM KCl and 20 mM MgCl$_2$, compared to 48% for WT RNA. 10 nM RNA was folded as described in Methods in 10 mM Tris-HCl (pH 7.5), 0.1 mg/mL yeast tRNA, and varying KCl and MgCl$_2$. Concentrations are in mM.

(C) Co-incubation experiment indicates that the WT and WT$_{trunc}$ constructs do not interact, confirming that the slow conformation is not a dimer. 2.5 nM isolated or mixed RNAs were denatured and folded as described in Methods in L9-binding buffer.

(D) Binding of L9 and L28 to different constructs. L28 appears to bind both high- and low-mobility states, as evidenced by the appearance of new bands in both regions of the gel (see also Figure 6E). L9 and L28 concentrations are 250 and 500 nM. The no protein and 500 nM L9+L28 lanes in the ΔH3 panel are identical to the ΔH3 panel in Figure 6E.

(E) Concentration-dependent binding of L9 (concentrations vary from 178 to 600 nM). Estimate, $K_D \approx 300$ nM.

(F) L33 binds the *rpmB* 5′ UTR but is competed by L9 (L33 = 500 nM; L9 varies from 125 to 500 nM).

(G) The WT$_{trunc}$ construct does not bind L9 or L28 (500 nM concentrations).

(H) Consensus 5′ UTR across all enterobacterial species indicates that sequences downstream of the three-way junction are highly conserved. Key for the consensus is located in main text Figure 6.