Bacterial RNA motif in the 5' UTR of *rpsF* interacts with an S6:S18 complex

YANG FU, KAILA DEIORIO-HAGGAR, MARK W. SOO, and MICHELLE M. MEYER¹

Department of Biology, Boston College, Chestnut Hill, Massachusetts 02467, USA

ABSTRACT

Approximately half the transcripts encoding ribosomal proteins in *Escherichia coli* include a structured RNA motif that interacts with a specific ribosomal protein to inhibit gene expression, thus allowing stoichiometric production of ribosome components. However, many of these RNA structures are not widely distributed across bacterial phyla. It is increasingly common for RNA motifs associated with ribosomal protein genes to be identified using comparative genomic methods, yet these are rarely experimentally validated. In this work, we characterize one such motif that precedes operons containing *rpsF* and *rpsR*, which encode ribosomal proteins S6 and S18. This RNA structure is widely distributed across many phyla of bacteria despite differences within the downstream operon, and examples are present in both *E. coli* and *Bacillus subtilis*. We demonstrate a direct interaction between an example of the RNA from *B. subtilis* and an S6:S18 complex using in vitro binding assays, verify our predicted secondary structure, and identify a putative protein-binding site. The proposed binding site bears a strong resemblance to the S18 binding site within the 16S rRNA, suggesting molecular mimicry. This interaction is a valuable addition to the canon of ribosomal protein mRNA interactions. This work shows how experimental verification translates computational results into concrete knowledge of biological systems.

Keywords: ribosome assembly; autogenous regulation; rRNA mimic; comparative genomics; ribosomal protein

INTRODUCTION

In the past few years, increasingly sophisticated comparative genomic approaches have predicted many putative RNA structures within bacterial genomes (Barrick et al. 2004; Weinberg et al. 2007, 2010; Yao et al. 2007; Meyer et al. 2009; Xu et al. 2009). Many of these RNAs have been experimentally characterized as riboswitches (for review, see Serganov and Nudler 2013), or other functional RNAs (Irnov and Winkler 2010). However, most have no experimental validation, and in many cases determining a biological function can be challenging (Meyer et al. 2011). As the volume of sequence data available continues to grow, experimental follow-up is increasingly necessary to translate findings from computational studies into biological knowledge.

Several conserved RNA motifs are associated with ribosomal protein (r-protein) operons (Weinberg et al. 2007; Yao et al. 2007; Meyer et al. 2009; Xu et al. 2009). In *Escherichia coli*, three decades of research have identified a series of mRNA structures that collectively regulate approximately half the transcriptional units encoding r-proteins (Zengel and Lindahl 1994). These mRNA structures allow autogenous

regulation, wherein an r-protein interacts with a structured portion of the transcript to negatively regulate transcription or translation (Dean et al. 1981; Zengel and Lindahl 1992). Many of the r-proteins responsible for this regulatory activity are primary ribosomal RNA (rRNA) binding proteins that interact with the rRNA early during ribosome assembly (e.g., S8, S4, S15, L1, L10, L20) (Dean and Nomura 1980). However, autogenous regulatory activity has been reported for tertiary rRNA binding proteins (e.g., S1, S2) (Skouf et al. 1990; Aseev et al. 2008). Several of the mRNA structures form secondary or tertiary structures that appear to mimic the rRNA binding sites of their protein interaction partners. For example, the mRNA structures interacting with L1, L10(L12)₄ complex, S8, and L20 each bear striking resemblance to the rRNA binding sites for these proteins (Merianos et al. 2004; Guillier et al. 2005; Nevskaya et al. 2005; Iben and Draper 2008). In other cases there is no clear resemblance between the mRNA and rRNA binding sites (Tang and Draper 1989; Mathy et al. 2004). These regulatory RNAs maintain stoichiometric levels of r-proteins (Sykes et al. 2010) and overexpression of their respective binding-proteins can result

¹Corresponding author

E-mail m.meyer@bc.edu

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.041285.113>.

© 2014 Fu et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported), as described at <http://creativecommons.org/licenses/by-nc/3.0/>.

in severe growth defects (Dean et al. 1981; Wower et al. 1992). Despite their important role in regulating an essential process, only three of more than ten RNA structures known in *E. coli* are widely distributed to many bacterial phyla. The majority are narrowly distributed to only a few orders of γ -proteobacteria (Fu et al. 2013). There are a number of cases in which multiple RNA structures have arisen in different bacterial phyla to perform similar functions (Nomura et al. 1980; Grundy and Henkin 1992; Scott and Williamson 2001; Guillier et al. 2002; Serganov et al. 2003; Choonee et al. 2007).

In this work, we demonstrate an RNA motif preceding *rpsF* and *rpsR*, encoding r-proteins S6 and S18 interacts with an S6:S18 dimer. Both S6 and S18 are secondary rRNA binding proteins that form a heterodimer prior to assembly with the S15-rRNA complex (Held et al. 1974; Recht and Williamson 2001). Although this RNA structure was originally identified in Firmicutes (Yao et al. 2007), our subsequent analysis indicates that it is widely distributed to many bacterial phyla. We verify the secondary structure predicted from comparative genomics using SHAPE (Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension) (Wilkinson et al. 2006) and identify a putative binding site that resembles the S18 rRNA binding site using nuclease protection assays (Ehresmann et al. 1987).

RESULTS

An RNA structure preceding *rpsF* is present in many bacterial species

A computational search for structured RNAs in Firmicute genomes identified an RNA motif preceding *rpsF* (Yao et al. 2007), and additional putative RNA structures preceding this gene have been identified by similar searches in other bacterial phyla (Weinberg et al. 2007, 2010). Examination of the motifs obtained from these searches revealed a common structural element that includes a stem-loop with a bulge containing a pair of conserved cytosines. We combined the shared regions of these motifs into a common alignment and subsequently used this to identify additional homologs in completed microbial genomes.

We identified more than 1300 sequences matching this RNA motif across many bacterial phyla, including Proteobacteria, Actinobacteria, Cyanobacteria, Spirochaetes, and Firmicutes (alignment in Supplemental File 1; secondary structures from selected examples drawn in

Supplemental Fig. 1). A consensus representation of this alignment is in Figure 1A. The double cytosine in the bulge is strikingly conserved, and the H1 stem sequesters a potential Shine-Dalgarno sequence in many examples, suggesting possible regulatory activity. However, the mere proximity of the conserved motif to the translational start site is potentially sufficient to allow regulation of gene expression upon protein binding through steric inhibition of ribosome binding. H2 is variable in length (13–107 nucleotides, typically paired), but the region proximal to the bulge is fairly conserved. Apart from this shared motif, the organization of the operon is not well conserved. The RNA invariably precedes *rpsF*, but the subsequent genomic context can vary, although it contains both *rpsF* and *rpsR* the vast majority of the time (Supplemental Table 1). In most Deltaproteobacteria and some γ -proteobacteria *rpsF* and *rpsR* occur together with no intervening gene. However, in β -proteobacteria and the remaining γ -proteobacteria they are separated by *priB*, and in Firmicutes and Actinobacteria they are separated by *ssbA*. In *E. coli* the operon contains *rpII* (encoding L9) (Isono and Kitakawa 1978), which occurs in isolation at a different

FIGURE 1. (A) Consensus secondary structure of the RNA motif preceding *rpsF* derived from more than 1300 sequences. This RNA motif is widely distributed across many bacterial phyla. (B) Example of the RNA motif from *B. subtilis* 168. The transcript start site, an alternative pairing element, and potential regulatory features are indicated. (C) Consensus secondary structure for the more than 500 RNA sequences primarily identified from Firmicutes and Actinobacteria that contain a potential conserved H3 helix. (D) rRNA binding site for the S6:S18 complex. Sequence is derived from *B. subtilis*, black nucleotides are conserved >90%, and red nucleotides conserved >98% across 4214 bacterial 16S rRNA sequences (Cannone et al. 2002). Bases in contact with S18 in *T. thermophilus* structure are boxed (Agalarov et al. 2000).

locus in Firmicute genomes (Akanuma et al. 2012). The only bacteria where we find the RNA without both *rpsF* and *rpsR* are some species of Cyanobacteria, where *rpsF* occurs in isolation in the genome and *rpsR* is found near an *rpmG* (encoding L33) homolog.

rpsF, *ssbA*, and *rpsR* are cotranscribed from S6 operon in *Bacillus subtilis*

Because the conserved RNA motif was originally identified in Firmicutes, we chose to study an example of the RNA occurring in the model gram-positive bacterium *Bacillus subtilis*. To ensure that the RNA motif is cotranscribed with both the *rpsF* and *rpsR* genes, we performed 5'-RLM RACE using primers in the S6 coding region. From this analysis, we identified a transcription start site that completely contains the predicted RNA structure (Fig. 1B) and is preceded in *B. subtilis* by a canonical promoter sequence (Supplemental Fig. 2). We also performed 3'-RACE to confirm the 3' end of the transcript, which lies 22 nucleotides after the stop codon for *rpsR*. Identification of the transcription start site reveals a potential alternative structure (H3) in the *B. subtilis* transcript that may disrupt the Shine-Dalgarno sequestration in H1 (Fig. 1B). Although this alternative pairing is not universally conserved throughout our alignment, we identified it in many Firmicute and Actinobacteria species (Fig. 1C).

Predicted RNA structure interacts with S6:S18 protein dimer in vitro

Based on the genomic organization of the operon, we hypothesized that either r-protein S6 or S18 was likely to interact with the RNA structure. In addition, previous studies of ribosome assembly indicate that S18 is insoluble and un-

structured prior to interaction with S6 (Held et al. 1974) and that S6 and S18 dimerize prior to association with the 16S rRNA-S15 complex (Recht and Williamson 2001), leading us to include the S6:S18 dimer as a potential binding partner. To determine the binding partner of the *B. subtilis* RNA structure, we conducted electrophoretic mobility shift assays (EMSA) using S6, S18, and the S6:S18 dimer (S6 and S18 mixed at a 1:1 ratio). Due to difficulties overexpressing *B. subtilis* S18 in *E. coli*, we utilized S6 and S18 proteins from a close thermophilic relative, *Geobacillus kaustophilus*. The RNA motifs from *G. kaustophilus* and *B. subtilis* are identical with the exception of minor differences in the sequence and length of the variable stem H2 (see Supplemental Fig. 1 for the *G. kaustophilus* RNA sequence and predicted secondary structure). In addition, the proteins share 75% (S6) and 89% (S18) sequence identity, and the vast majority of residues making contacts to the rRNA are conserved between *B. subtilis* and *G. kaustophilus* (see Supplemental Fig. 3). Although species-specific interactions between mRNA regulatory RNAs and ribosomal proteins are reported (Scott and Williamson 2005), these typically occur when the regulatory RNAs are nonhomologous. In this case the RNAs are homologous; thus, this substitution is unlikely to negatively impact potential binding interactions. The RNA construct derived from *B. subtilis* used for most of our in vitro assays (*rpsF*_5-69) was truncated at the 5' end. However, similar results were observed for a construct beginning at the native transcription start site (*rpsF*_1-69). Both of these constructs encompass the entire conserved RNA structure and continue through the translational start site and are thus likely to include any relevant features required for protein-binding.

We observed a specific RNA-protein interaction between the *rpsF*_5-69 RNA and the S6:S18 dimer. S18 alone binds to *rpsF*_5-69 RNA at high protein concentrations ($K_D > 500$ nM), and S6 does not bind at all (Fig. 2A–C). The predicted RNA structure has nM affinity for the protein dimer (155 ± 17 nM) (Fig. 2D), and we observe a high Hill coefficient (4.1 ± 0.5) for this interaction. Several different experimental artifacts can potentially cause the exceedingly steep binding curve observed. First, many r-proteins are “sticky” and can be absorbed onto surfaces of tubes, resulting in no observed binding until the surfaces are saturated. Second, due to the insolubility of S18, aggregation of S18 or the S6:S18 dimer could be occurring. Finally, the proportion of correctly folded heterodimer in solution could be limiting the extent to which RNA binding occurs. Our binding buffer contains 0.8% BSA, which should largely alleviate protein absorption onto tube surfaces. To examine the potential for aggregation, we repeated our experiments in the presence of detergent (0.02% IGEPAL CA630) (Recht and Williamson 2001) and found no significant change in either the K_D or the Hill coefficient.

To further assess the causes of our steep binding curve, we also individually titrated S6 and S18, where the other component of the heterodimer was held constant at high concentration (Fig. 3A–C). Under these conditions, the majority of

FIGURE 2. (A) Gel-shift of *rpsF*_5-69 with increasing concentration of S18 (0 nM, 19–600 nM). (B) Gel-shift of *rpsF*_5-69 with increasing concentration of S6 (0 nM, 9.7–600 nM). (C) Gel-shift of *rpsF*_5-69 with increasing S6 and S18 (mixed 1:1, 0 nM, 9.7–600 nM). Arrows indicate quantified bands. (D) Quantification of the fraction of *rpsF*_5-69 bound by the protein with S6, S18, and S6:S18. Error bars represent standard deviation on three to six independent experiments. Some error bars fall within the boundaries of the marker. Curve represents a standard binding curve with a K_D of 155 nM, and a Hill coefficient of 4.

FIGURE 3. (A) Gel-shift of rpsF₅₋₆₉ with 500 nM S18 and increasing concentration of S6 (no protein, 0 nM S6, 4.6–500 nM S6). (B) Gel-shift of rpsF₅₋₆₉ with 600 nM S6 and increasing concentration of S18 (0 nM, 18.5–600 nM S18, no protein). Arrows on A and B indicate quantified bands. (C) Quantification of the fraction of rpsF₅₋₆₉ bound by the S6:S18 complex. Solid thick line corresponds to the titration of S6:S18 shown in Figure 2D (K_D of 155 nM, Hill coefficient of 4, maximum fraction bound 100%). Dashed line is the curve fit of the S6 titration (K_D of 37 nM, Hill coefficient 1.45, maximum fraction bound 93.5%). Thin line is the curve fit for the S18 titration (K_D of 60 nM, Hill coefficient 1.5, maximum fraction bound of 100%).

the titrated protein should form the heterodimer allowing for more accurate assessment of heterodimer:RNA interactions. These experiments show that RNA binding can occur at approximately two- to threefold lower concentrations (K_D for S6 titration [S18 at 500 nM] is 37 ± 7 nM, K_D for S18 titration [S6 at 600 nM] is 60 ± 3 nM). As anticipated, the steep transition previously observed is largely alleviated under these conditions (Hill coefficients are 1.45 ± 0.40 and 1.5 ± 0.23 for S6 and S18 titrations, respectively). These experiments suggest that the steep transition observed for the S6:S18 titration is most likely the result of inadequate dimer formation when S6 and S18 are both titrated rather than aggregation or absorption of the proteins onto tube surfaces.

To ensure the specificity of the observed interaction, we constructed several mutated RNAs. A series of mutations that disrupted various portions of the secondary structure (M1–M3) and the conserved pair of cytosines (M4) were analyzed with the S6:S18 protein dimer using EMSAs (Fig. 4A). The RNA–protein interactions were either abolished (M2) or significantly impaired by these mutations (M1, M3, M4) ($K_D > 500$ nM). To further confirm our secondary structure predictions, we introduced compensatory mutations (M5 and M6). The M6 compensatory mutation to the H1 stem fully restores protein binding (Fig. 4B) ($K_D = 125$ nM $\pm$ 13), but the M5 mutation within helix H2 does not ($K_D > 500$ nM). This indicates that either the primary sequence changes have altered the global structure in some unanticipated way, or the primary sequence in this area is important for a productive interaction.

Structural probing confirms predicted secondary structure

To confirm our secondary structure predicted from comparative genomics, we conducted SHAPE (Wilkinson et al. 2006)

in the presence and absence of the S6:S18 dimer. For this experiment, an additional hairpin was appended to the 3' end of the rpsF₁₋₆₉ RNA to allow for primer binding. Limitations in resolving the 5' end of RNA on the SHAPE gel (Supplemental Fig. 4; quantified nucleotide reactivity in Supplemental Table 2) prevented quantification of reverse transcription for nucleotides 1–6, and a strong pause site at the base of the H2 stem prevents quantitative analysis of nucleotides 35–39. From analysis of the SHAPE data we find that the 3' end of the RNA is largely flexible in both the protein-bound and protein-unbound states. The protein-bound and -unbound states show moderately similar, but not identical, patterns suggesting that there are changes in flexibility; but there do not appear to be any

large whole-scale changes to the secondary structure between the two states (Fig. 5A,B). Reactive SHAPE nucleotides are predominantly located in the predicted loop regions of H1 and H2, within the conserved bulge, and at the 3' termini of the RNA that is not predicted to be structured. Somewhat surprisingly, we also consistently observe significant flexibility in the 3' portion of the predicted H1 stem. However, this may partially be due to over estimation of cleavage at the 3' end of the molecule due to the signal decay that occurs during primer extension.

To further assess the secondary structure, we also conducted nuclease probing with RNase V1 and RNase A (Fig. 5C). RNase V1 nonspecifically cleaves double-stranded RNA, and significant cleavage is observed for bases predicted to be paired in H2 (28–29 and 35–38) and for the bases resolved on the gel that are predicted to be paired in H1 (52–56, 20–25) (Fig. 5C,E). RNase A cleaves single-stranded C and U, and we can see that U32, C39, U43, and C48 are cleaved, in

FIGURE 4. (A) The RNA construct rpsF₅₋₆₉ used for in vitro binding assays with mutations indicated. (B) Interactions of S6:S18 with mutants displayed in A. Solid black curve represents the wild-type interaction from Figure 2D. A binding curve is plotted for M6 ($K_D = 115$ nM, Hill coefficient 3). For the remaining mutants RNA binding was not saturated, indicating a $K_D > 500$ nM and no curves are drawn.

FIGURE 5. (A) Highly reactive (black circles) and moderately reactive (gray circles) nucleotides determined from SHAPE analysis of RNA in isolation mapped to the rpsF_1-69 structure. Positions not resolved are in gray. (B) Reactive nucleotides in SHAPE analysis of RNA bound to the S6:S18 complex. (C) Nuclease probing data mapped to the secondary structure of the rpsF_5-69 RNA structure. Starred bases are cleaved by V1 nuclease and arrows indicate bases cleaved by RNase A. Bases protected from cleavage by the addition of protein are circled. Bases not resolved on the gel are in gray. Numbering starts from the transcription start site (A). (D,E) Nuclease probing gels of rpsF_5-69 (D) and rpsF_5-69 M4 (E) using RNase A (left), which cleaves single-stranded uridine and cytosine, and RNase V1 (right), which cleaves double-stranded RNA. RNA was incubated with increasing concentrations of S6 and S18 mixed at a 1:1 ratio. (F) V1 nuclease probing gel of rpsF_5-69 in the presence of S18 alone. OH⁻ and T1 lanes indicate partial alkaline hydrolysis and RNase T1 digest under denaturing conditions used to map the RNA sequence. On the right, bases cleaved by RNase V1 are indicated. On the left, nucleotides cleaved by RNase T1 (G) and RNase A (C or U, bold) are indicated.

accordance with the SHAPE data, suggesting that these regions are flexible (Fig. 5A,B). Although we predict that the conserved bulge is not double-stranded, the RNase A cleavage patterns suggest that portions of this bulge have potential tertiary structure. In particular, we do not observe cleavage at several positions (e.g., C41, U46 and C47) predicted to be single-stranded within this bulge.

Nuclease protection analysis suggests a protein-binding region

To assess which portions of the RNA are necessary for interaction with the S6:S18 dimer, we performed nuclease protection assays using RNase V1 and RNase A with increasing concentrations of the S6:S18 dimer (Fig. 5C–E). Although such experiments can only approximate sites of protein binding,

they do give insight into general regions of the RNA that are impacted by protein binding. In the presence of the S6:S18 dimer, several regions of the rpsF_5-69 RNA are protected from RNase cleavage, indicating that the protein dimer most likely has some contact with the bulge and portions of H1 and H2 proximal to the bulge (Fig. 5C,D). Upon increasing protein concentrations, significant RNase V1 protection is seen for bases 28–29 and 35–38 in the H2 stem. This direct interaction between the protein complex and this portion of H2 suggests that our mutations to H2 (M3 and M5) likely disrupt not only secondary structure, but also potential sequence specific contacts. RNase V1 protection was also observed for H1 bases 52–53, which includes a putative Shine-Dalgarno sequence. From the RNase A protection, it is apparent that U43 and C48 within the bulge are significantly protected in the presence of the S6:S18 dimer.

This protection within the bulge further supports our hypothesis that this region is important for protein binding. The loop region of H2 is unaffected by the presence of protein. This finding is in accordance with the multiple sequence alignment, which indicates that H2 can vary significantly in length. Thus beyond some minimum length, H2 is not likely to be important for binding. Despite an overall very similar cleavage pattern in the absence of the S6:S18 dimer, no protection was observed with the M4 mutant RNA. This indicates that changes in nuclease cleavage are attributable to specific protein–RNA interactions (Fig. 5D).

We also conducted nuclease protection assays with just S18 rather than the S6:S18 complex to further examine the potential S18–mRNA interaction observed at high S18 concentrations during gel-shift assays. In these experiments, we did not observe any specific changes to protection upon increasing protein concentration up to 600 nM. This suggests that the results from the gel-shift assays, indicating that S18 alone can interact with the RNA, may reflect nonspecific or weak interactions that cannot prevent nuclease cleavage.

DISCUSSION

In this work, we show that a putative RNA structure preceding the *B. subtilis* operon containing *rpsF* and *rpsR* specifically interacts with an S6:S18 dimer. Although weak interactions with S18 are detected, it is likely that the biologically active molecule is the S6:S18 dimer. S18 is insoluble and only partially folded in the absence of S6 (Held et al. 1974), and the previously reported dissociation constant for this interaction is significantly below that we observe for S6:S18 interaction with the RNA structure. Furthermore, S18 interactions alone are not sufficient to prevent nuclease cleavage, suggesting that the potential interaction observed is either not specific, or very weak. Like most nonribosomal RNA binding sites for r-proteins (Zengel and Lindahl 1994), the dissociation constant for the *rpsF*_5-69 and S6:S18 dimer complex is significantly higher than that reported for the S6:S18 dimer with the 16S rRNA–S15 complex (*A. aeolicus* ~6 nM under equilibrium conditions) (Recht and Williamson 2001).

We do observe an exceedingly steep transition between the unbound and protein-bound states in our gel-shift assays. We conclude that this steep transition is most likely a result of inadequate or inefficient heterodimer formation. Previous studies of *Aquifex aeolicus* S6:S18 dimer formation and interactions with the assembling ribosome (S15–rRNA complex) report a dissociation constant for the S6:S18 heterodimer of 8 nM and a dissociation constant between this complex and the S15–RNA complex of 6 nM. The proximity of these two values results in a similar steep transition curve measured under equilibrium conditions due to difficulties maintaining appropriate concentrations of the heterodimer at low protein concentrations. Our protein concentrations are significantly higher, but the overall effect is similar, suggesting that either the dissociation constant between *G. kaustophilus* S6 and S18

is somewhat higher than that reported for *A. aeolicus*, or more likely that we have poor reconstitution of the dimer protein in our assays due to inadequate refolding of S18.

One characteristic feature of the S6:S18 interacting RNA structure is the conserved pair of cytosines present in the bulge (C47 and C48). Our mutagenesis results (M4) indicate that the cytosines are important for binding, and although coarse-grained, our nuclease probing assays indicate that C48 is significantly protected from cleavage by its interaction with the S6:S18 dimer, suggesting that this region is proximal to the protein binding region. S18 makes extensive contacts with a region of the 16S rRNA that also contains a pair of conserved cytosines (C719 and C720) located in the three-helix junction between rRNA helices 22, 23a, and 23b (Fig. 1D; Agalarov et al. 2000). Furthermore, these cytosines are also conserved in >90% of bacterial 16S rRNA sequences (Fig. 1D). Based on our observed interaction between S18 and the *rpsF*_5-69 structure and the similarities between the rRNA and mRNA primary sequences, we speculate the mRNA structure we observe is mimicking a portion of the rRNA binding site for S18.

The interaction of S6:S18 with its mRNA and rRNA binding sites is reminiscent of the interaction of the L10 (L12)₄ complex with its own mRNA and rRNA binding sites. In both cases, the protein making the majority of contacts to the RNA is largely insoluble (S18 or L10) (Iben and Draper 2008), but may bind the mRNA independently of its protein partner (S6 or L12) (Brot et al. 1980). However, the protein complex, S6:S18 or L10(L12)₄, has an improved interaction with the mRNA due to increased solubility and folding of the dominant protein partner (S18 or L10).

The RNA we have characterized from *B. subtilis* joins the collection of RNA structures that interact with r-proteins. However, its wide distribution across many bacterial phyla is the exception rather than the rule for such RNAs. Of the r-protein interacting RNA structures known in *E. coli*, only those interacting with L1, L10, and S2 show a similar broad distribution to many bacterial phyla (Fu et al. 2013). Like the example here, the RNAs interacting with both L1 and L10 are believed to act through mimicry of the rRNA binding sites (Nevskaya et al. 2005; Iben and Draper 2008). Although there are narrowly distributed r-protein mRNA binding sites that are obvious mimics of the rRNA (e.g., that interacting S8) (Merianos et al. 2004), most r-protein binding motifs showing obvious mimicry are broadly distributed.

MATERIALS AND METHODS

Homology searches

Covariance models for each RNA were constructed and calibrated using Infernal 1.1 (cmbuild, cmcalibrate), and homologs identified for each RNA (cmsearch) (Nawrocki and Eddy 2013). Cmsearch was performed against a custom sequence database as described previously (Deiorio-Haggard et al. 2013) using a lenient *E*-value cutoff.

Potential homologs were assessed on the basis of genomic context, using a custom visualization tool (GenomeChart) (Miller et al. 2013) and for fit to the existing alignment. Alignments were manually adjusted as necessary when sequences with variable-length helices and/or loops were added. The search process was repeated approximately six to seven times, to expand sequence diversity. Phylogenetic diversity was determined from the number of completed genomes within refseq58 based on the final alignments utilizing queries to our custom database. Consensus secondary structure diagrams were created from the alignments using GSC-weighting in R2R (Weinberg and Breaker 2011).

RACE

Bacillus subtilis total RNA was extracted from a log phase culture and 5' RACE performed using the Invitrogen GeneRacer kit. First strands were synthesized with a gene-specific primer (*rpsF*, 5'-ACCCCAATCCTTTGTTCAGTG; *rpsR*, 5'-GATCTACATCTTTGTAGTCGATGTGC). PCRs were conducted using the first strand cDNA as the template and a second gene-specific primer (*rpsF*, 5'-GACTCTTCGTCAATGTTTGGGCGG; *rpsR*, 5'-CGCACGACCGCCTCTGCGTCCTC). 3' RACE began with adenylation of extracted total RNA followed by reverse transcription with a dT primer (5'-GCGGTCACGCTTACTTAGCCCTCACTGAAATTTTTTTTTTTT TTTT) to generate the cDNA. PCR was conducted using this cDNA as the template, a gene-specific primer (5'-GATCAAACGC GCACGTCAAATGG) and an Anchor primer (5'-GCGGTCACGC TTACTTAGCCCTCACTGAA). All PCR products were cloned using Invitrogen TOPO-cloning kit and sequenced to identify the transcription initiation site and termination site of the *S6* operon and the cotranscription of related genes.

Preparation of RNA transcripts

The 65-nucleotide DNA fragment preceding *rpsF* was PCR amplified from *Bacillus subtilis* genomic DNA with a T7 promoter and was transcribed into mRNA in vitro with T7 RNA polymerase (Milligan et al. 1987). Purified mRNA was 5'-labeled with $\gamma^{32}\text{P}$ -ATP and purified again on a denaturing acrylamide gel. Transcripts carrying mutations were produced in the same way from a DNA template carrying mutations that had been introduced through QuickChange mutagenesis.

Preparation of *Geobacillus kaustophilus* S6 and S18 proteins

The genes coding for *Geobacillus kaustophilus* S6 and S18 were cloned into pET-HT (Block et al. 2011). Clones were transformed into BL21(DE3) cells and proteins were over-expressed from the T7 promoter. After over-expression, proteins were extracted and purified similarly to that previously described (Culver and Noller 1999). Cells were lysed in Buffer E (20 mM HEPES [pH 7.6], 20 mM KCl, and 6 mM BME). S6 was soluble and was purified with Buffer C (20 mM NaOAc [pH 5.0], 20 mM KCl, and 6 mM BME) at 4°C by nondenaturing FPLC cation exchange chromatography with a linear gradient (20 mM–1 M KCl). S18 was recovered from inclusion bodies with Buffer B (20 mM Tris-HCl [pH 7.6], 20 mM KCl, 6 M urea, and 6 mM BME) and was purified at 4°C by denaturing FPLC cation exchange chromatography with a linear

gradient (20 mM–1 M KCl). Protein-containing fractions were examined by SDS-PAGE, collected, and dialyzed against Buffer D (S18 with 4 M urea). After dialysis proteins were examined by SDS-PAGE again, the concentration was determined by Bradford assays. Proteins were stored in buffer D (S18 with 4 M urea) at 4°C.

Electrophoretic mobility shift assays

Protein–RNA interaction was examined by electrophoretic mobility shift assays (EMSA). A fixed amount of radio-labeled RNA (1000 cpm, ~ 1 nM) was incubated with protein in serial dilutions (0–600 nM) in a total volume of 10 μL for 30 min at 25°C (2 $\times$ binding buffer: 100 mM Tris-HCl pH7.6, 40 mM MgCl_2 , 300 mM KCl, 2 mM BME, 0.08% BSA). Five microliters 50% glycerol was added to each reaction before loading onto a native 10% acrylamide gel which had been prerun for 60 min at 160V under 25°C. The loaded gel was run for 4 h at 350V at 4°C and subsequently dried and imaged using GE Healthcare STORM 820 phosphorimager and bands quantified using ImageQuant. Quantified bands were fit to a standard binding curve of the form $(F_{\text{max}} \cdot y^n)/(x^n + K_D^n)$ where n is the Hill coefficient, F_{max} is the maximum fraction bound, x is the concentration of protein in the reaction, and y is the measured fraction bound, using the Solver function in Microsoft Excel to minimize the summed residuals subject to the constraint $F_{\text{max}} \leq 1.0$. K_D and Hill coefficient values are reported $\pm$ the standard deviation of three or more data sets.

For individual component titrations, 600 nM S6 was incubated with a S18 serial dilution series (0 nM, 18.75–600 nM, twofold dilutions) in 1 $\times$ binding buffer for 10 min before a fixed amount of radio-labeled RNA (*rpsF*_5-69, 1000 cpm, ~ 1 nM) was added. The reactions were incubated at 25°C for an additional 20 min. RNA–protein interaction was examined by EMSA as described above. 500 nM S18 was incubated with a dilution series of S6 (0 nM, 4.6–300 nM in twofold serial dilutions, and 500 nM). Titrations with S18 at 600 nM displayed significant aggregation and did not yield interpretable results.

SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) analysis

RNA composed of the predicted RNA structure (*rpsF*_1-69) with a 3' primer binding site appended (5'-TCGATCCGGTTCGCCGGATCCAAATCGGGCTTCGGTCCGGTTC) was prepared as described above using T7 transcription (Milligan et al. 1987). To form acylated products, 3 μL NMIA (130 mM) was added into RNA–protein binding reactions (with and without protein) as described above and incubated for 15 min at 25°C. Modified RNAs were purified by phenol-chloroform extraction followed by ethanol precipitation and redissolved in 10 μL TE buffer (0.5 $\times$). Primer extension was performed by incubating the redissolved RNA with a 5' radiolabeled primer (5'-GAACCGGACCGAAGCCCCG) for 30 min at 55°C in the presence of Superscript III (Invitrogen). Reverse transcription was stopped by addition of 1 μL NaOH (4 M) and 29 μL acid stop mix (Wilkinson et al. 2006). cDNAs were loaded on a 10% denaturing Acrylamide/Bis gel and run at 55W at 25°C. The gel was dried and examined using a GE Healthcare STORM phosphorimager and SAFA footprinting software (Das et al. 2005; Laederach et al. 2008). Data was normalized as described (Wilkinson et al. 2006; Deigan et al. 2009).

Nuclease protection analysis

For S6:S18 nuclease protection assays, the RNA–protein binding reactions identical to those used for EMSA assays above (50 μ L reaction volume with S6:S18 dimer concentrations of 0, 50, 100, 300, and 600 nM) were assembled. For S18 nuclease protection assays binding reactions with S18 concentrations of 0, 25, 50, 100, 300, and 600 nM were similarly assembled. After protein–RNA incubation, each reaction was treated with 1 μ L RNase A and V1 (purchased from Ambion: 20 pg and 0.002 units, respectively) for 10 min and 5 min, respectively, at 25°C. The nuclease was inactivated with inactivation/precipitation buffer (Ambion), and RNA fragments were recovered by ethanol precipitation. Precipitated RNAs were resuspended in 10 μ L water and 10 μ L urea loading solution (10 M urea, 1.5 mM EDTA). Ten microliters of each reaction was loaded on a 12% denaturing Acrylamide/Bis gel and run at 35W at 25°C. Partial alkaline hydrolysis reactions were generated by incubating RNA in 50 mM Na₂CO₃ pH 9.0, 1 mM EDTA for 10 min at 95°C. Denaturing T1 reaction was conducted by incubating 1U of RNase T1 (Roche) in 25 mM Na Citrate pH 5.0, 5.5 M urea for 15 min at 55°C. The gel was dried and examined using a GE Healthcare STORM 820 phosphorimager and ImageQuant.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

The authors thank Arianne M. Babina for her work cloning the S6 RNA and proteins from *B. subtilis*. This work was generously supported by a bioinformatics starter grant from the PhRMA Foundation, the Alfred P. Sloan Foundation, and Boston College. M.W.S. is supported by an American Society of Microbiology Undergraduate Research Fellowship.

Received July 12, 2013; accepted October 29, 2013.

REFERENCES

- Agalarov SC, Sridhar Prasad G, Funke PM, Stout CD, Williamson JR. 2000. Structure of the S15,S6,S18-rRNA complex: Assembly of the 30S ribosome central domain. *Science* **288**: 107–113.
- Akanuma G, Nanamiya H, Natori Y, Yano K, Suzuki S, Omata S, Ishizuka M, Sekine Y, Kawamura F. 2012. Inactivation of ribosomal protein genes in *Bacillus subtilis* reveals importance of each ribosomal protein for cell proliferation and cell differentiation. *J Bacteriol* **194**: 6282–6291.
- Aseev LV, Levandovskaya AA, Tchufistova LS, Scaptsova NV, Boni IV. 2008. A new regulatory circuit in ribosomal protein operons: S2-mediated control of the *rpsB-ts* expression in vivo. *RNA* **14**: 1882–1894.
- Barrick J, Corbino K, Winkler W, Nahvi A, Mandal M, Collins J, Lee M, Roth A, Sudarsan N, Jona I. 2004. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci* **101**: 6421–6426.
- Block KF, Puerta-Fernandez E, Wallace JG, Breaker RR. 2011. Association of OLE RNA with bacterial membranes via an RNA–protein interaction. *Mol Microbiol* **79**: 21–34.
- Brot N, Caldwell P, Weissbach H. 1980. Autogenous control of *Escherichia coli* ribosomal protein L10 synthesis *in vitro*. *Proc Natl Acad Sci* **77**: 2592–2595.
- Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, et al. 2002. The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3**: 2.
- Choonee N, Even S, Zig L, Putzer H. 2007. Ribosomal protein L20 controls expression of the *Bacillus subtilis* *infC* operon via a transcription attenuation mechanism. *Nucleic Acids Res* **35**: 1578–1588.
- Culver G, Noller H. 1999. Efficient reconstitution of functional *Escherichia coli* 30S ribosomal subunits from a complete set of recombinant small subunit ribosomal proteins. *RNA* **5**: 832–843.
- Das R, Laederach A, Pearlman SM, Herschlag D, Altman RB. 2005. SAFA: Semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA* **11**: 344–354.
- Dean D, Nomura M. 1980. Feedback regulation of ribosomal protein gene expression in *Escherichia coli*. *Proc Natl Acad Sci* **77**: 3590–3594.
- Dean D, Yates J, Nomura M. 1981. *Escherichia coli* ribosomal protein S8 feedback regulates part of *spc* operon. *Nature* **289**: 89–91.
- Deigan KE, Li TW, Mathews DH, Weeks KM. 2009. Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci* **106**: 97–102.
- Deiorio-Haggar K, Anthony J, Meyer M. 2013. RNA structures regulating ribosomal protein biosynthesis in bacilli. *RNA Biol* **10**: 1180–1184.
- Ehresmann C, Baudin F, Mougél M, Romby P, Ebel JP, Ehresmann B. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res* **15**: 9109–9128.
- Fu Y, Deiorio-Haggar K, Anthony J, Meyer MM. 2013. Most RNAs regulating ribosomal protein biosynthesis in *Escherichia coli* are narrowly distributed to γ -proteobacteria. *Nucleic Acids Res* **41**: 3491–3503.
- Grundy F, Henkin T. 1992. Characterization of the *Bacillus subtilis* *rpsD* regulatory target site. *J Bacteriol* **174**: 6763–6770.
- Guillier M, Allemand F, Raibaud S, Dardel F, Springer M, Chiaruttini C. 2002. Translational feedback regulation of the gene for L35 in *Escherichia coli* requires binding of ribosomal protein L20 to two sites in its leader mRNA: A possible case of ribosomal RNA–messenger RNA molecular mimicry. *RNA* **8**: 878–889.
- Guillier M, Allemand F, Dardel F, Royer C, Springer M, 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**: 1141–1456.
- Held W, Ballou B, Mizushima S, Nomura M. 1974. Assembly mapping of 30 S ribosomal proteins from *Escherichia coli*: Further studies. *J Biol Chem* **249**: 3103–3111.
- Iben J, Draper D. 2008. Specific Interactions of the L10(L12)₄ ribosomal protein complex with mRNA, rRNA, and L11. *Biochemistry* **47**: 2721–2731.
- Irnov I, Winkler WC. 2010. A regulatory RNA required for antitermination of biofilm and capsular polysaccharide operons in Bacillales. *Mol Microbiol* **76**: 559–575.
- Isono K, Kitakawa M. 1978. Cluster of ribosomal protein genes in *Escherichia coli* containing genes for proteins S6, S18, and L9. *Proc Natl Acad Sci* **75**: 6163–6167.
- Laederach A, Das R, Vicens Q, Pearlman SM, Brenowitz M, Herschlag D, Altman RB. 2008. Semiautomated and rapid quantification of nucleic acid footprinting and structure mapping experiments. *Nat Protoc* **3**: 1395–1401.
- Mathy N, Pellegrini O, Serganov A, Patel DJ, Ehresmann C, Portier C. 2004. Specific recognition of *rpsO* mRNA and 16S rRNA by *Escherichia coli* ribosomal protein S15 relies on both mimicry and site differentiation. *Mol Microbiol* **52**: 661–675.
- Merianos H, Wang J, Moore P. 2004. The structure of a ribosomal protein S8/*spc* operon mRNA complex. *RNA* **10**: 954–964.
- Meyer M, Ames T, Smith D, Weinberg Z, Schwabach M, Giovannoni S, Breaker R. 2009. Identification of candidate structured RNAs in the marine organism '*Candidatus Pelagibacter ubique*'. *BMC Genomics* **10**: 268.

- Meyer MM, Hammond MC, Salinas Y, Roth A, Sudarsan N, Breaker RR. 2011. Challenges of ligand identification for riboswitch candidates. *RNA Biol* **8**: 5–10.
- Miller CA, Anthony J, Meyer MM, Marth G. 2013. Scribl: An HTML5 Canvas-based graphics library for visualizing genomic data over the web. *Bioinformatics* **29**: 381–383.
- Milligan J, Groebe D, Witherell G, Uhlenbeck O. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* **15**: 8783–8798.
- Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**: 2933–2935.
- Nevskaya N, Tishchenko S, Gabdoulkhakov A, Nikonova E, Nikonov O, Nikulin A, Platonova O, Garner M, Nikonov S, Piendl W. 2005. Ribosomal protein L1 recognizes the same specific structural motif in its target sites on the autoregulatory mRNA and 23S rRNA. *Nucleic Acids Res* **33**: 478–485.
- Nomura M, Yates J, Dean D, Post L. 1980. Feedback regulation of ribosomal protein gene expression in *Escherichia coli*: Structural homology of ribosomal RNA and ribosomal protein mRNA. *Proc Natl Acad Sci* **77**: 7084–7088.
- Recht MI, Williamson JR. 2001. Central domain assembly: Thermodynamics and kinetics of S6 and S18 binding to an S15-RNA complex. *J Mol Biol* **313**: 35–48.
- Scott LG, Williamson JR. 2001. Interaction of the *Bacillus stearothermophilus* ribosomal protein S15 with its 5'-translational operator mRNA. *J Mol Biol* **314**: 413–422.
- Scott L, Williamson J. 2005. The binding interface between *Bacillus stearothermophilus* ribosomal protein S15 and its 5'-translational operator mRNA. *J Mol Biol* **351**: 280–290.
- Serganov A, Nudler E. 2013. A decade of riboswitches. *Cell* **152**: 17–24.
- Serganov A, Polonskaia A, Ehresmann B, Ehresmann C, Patel D. 2003. Ribosomal protein S15 represses its own translation via adaptation of an rRNA-like fold within its mRNA. *EMBO J* **22**: 1898–1908.
- Skouv J, Schnier J, Rasmussen M, Subramanian A, Pedersen S. 1990. Ribosomal protein S1 of *Escherichia coli* is the effector for the regulation of its own synthesis. *J Biol Chem* **265**: 17044–17049.
- Sykes M, Sperling E, Chen S, Williamson J. 2010. Quantitation of the ribosomal protein autoregulatory network using mass spectrometry. *Anal Chem* **82**: 5038–5045.
- Tang C, Draper D. 1989. Unusual mRNA pseudoknot structure is recognized by a protein translational repressor. *Cell* **57**: 531–536.
- Weinberg Z, Breaker RR. 2011. R2R—software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics* **12**: 3.
- Weinberg Z, Barrick JE, Yao Z, Roth A, Kim JN, Gore J, Wang JX, Lee ER, Block KF, Sudarsan N, et al. 2007. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res* **35**: 4809–4819.
- Weinberg Z, Wang J, Bogue J, Yang J, Corbino K, Moy R, Breaker R. 2010. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol* **11**: R31.
- Wilkinson K, Merino E, Weeks K. 2006. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): Quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1**: 1610–1616.
- Wower I, Kowaleski MP, Sears LE, Zimmermann RA. 1992. Mutagenesis of ribosomal protein S8 from *Escherichia coli*: Defects in regulation of the *spc* operon. *J Bacteriol* **174**: 1213–1221.
- Xu X, Ji Y, Stormo GD. 2009. Discovering *cis*-regulatory RNAs in *Shewanella* genomes by Support Vector Machines. *PLoS Comput Biol* **5**: e1000338.
- Yao Z, Barrick J, Weinberg Z, Neph S, Breaker R, Tompa M, Ruzzo WL. 2007. A computational pipeline for high-throughput discovery of *cis*-regulatory noncoding RNA in prokaryotes. *PLoS Comput Biol* **3**: e126.
- Zengel JM, Lindahl L. 1992. Ribosomal protein L4 and transcription factor NusA have separable roles in mediating terminating of transcription within the leader of the S10 operon of *Escherichia coli*. *Genes Dev* **6**: 2655–2662.
- Zengel JM, Lindahl L. 1994. Diverse mechanisms for regulating ribosomal protein synthesis in *Escherichia coli*. *Prog Nucleic Acid Res Mol Biol* **47**: 331–370.

RNA

A PUBLICATION OF THE RNA SOCIETY

Bacterial RNA motif in the 5' UTR of *rpsF* interacts with an S6:S18 complex

Yang Fu, Kaila Deiorio-Haggar, Mark W. Soo, et al.

RNA 2014 20: 168-176 originally published online December 5, 2013

Access the most recent version at doi:[10.1261/rna.041285.113](https://doi.org/10.1261/rna.041285.113)

Supplemental Material

<http://rnajournal.cshlp.org/content/suppl/2013/11/21/rna.041285.113.DC1>

References

This article cites 50 articles, 19 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/20/2/168.full.html#ref-list-1>

Creative Commons License

This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported), as described at <http://creativecommons.org/licenses/by-nc/3.0/>.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:

<http://rnajournal.cshlp.org/subscriptions>