

# S6:S18 ribosomal protein complex interacts with a structural motif present in its own mRNA

DOROTA MATELSKA,<sup>1</sup> ELZBIETA PURTA,<sup>1</sup> SYLWIA PANEK,<sup>1</sup> MICHAL J. BONIECKI,<sup>1</sup> JANUSZ M. BUJNICKI,<sup>1,2,3</sup> and STANISLAW DUNIN-HORKAWICZ<sup>1,3</sup>

<sup>1</sup>Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, Warsaw, 02-109, Poland

<sup>2</sup>Bioinformatics Laboratory, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznań, 61-614, Poland

## ABSTRACT

Prokaryotic ribosomal protein genes are typically grouped within highly conserved operons. In many cases, one or more of the encoded proteins not only bind to a specific site in the ribosomal RNA, but also to a motif localized within their own mRNA, and thereby regulate expression of the operon. In this study, we computationally predicted an RNA motif present in many bacterial phyla within the 5' untranslated region of operons encoding ribosomal proteins S6 and S18. We demonstrated that the S6:S18 complex binds to this motif, which we hereafter refer to as the S6:S18 complex-binding motif (S6S18CBM). This motif is a conserved CCG sequence presented in a bulge flanked by a stem and a hairpin structure. A similar structure containing a CCG trinucleotide forms the S6:S18 complex binding site in 16S ribosomal RNA. We have constructed a 3D structural model of a S6:S18 complex with S6S18CBM, which suggests that the CCG trinucleotide in a specific structural context may be specifically recognized by the S18 protein. This prediction was supported by site-directed mutagenesis of both RNA and protein components. These results provide a molecular basis for understanding protein-RNA recognition and suggest that the S6S18CBM is involved in an auto-regulatory mechanism.

**Keywords:** ribosomal protein S6; feedback regulation; RNA secondary structure; RNA structure prediction; ribosome

## INTRODUCTION

In prokaryotes, most ribosomal protein (r-protein) genes are organized within highly conserved operons (Roberts et al. 2008). In a number of cases their expression is regulated through a feedback mechanism, in which one or more protein products of a given operon acts as a regulator (Babitzke et al. 2009). Upon saturation of the binding site on a ribosomal RNA (rRNA), a regulator-protein binds to its own mRNA, frequently within the 5' untranslated region (5' UTR) (Nomura et al. 1980; Freedman et al. 1987; Thomas and Nomura 1987; Portier et al. 1990; Philippe et al. 1993; Guillier et al. 2002). This interaction inhibits translation initiation by occlusion of a ribosome binding site or ribosome entrapment, permitting maintenance of the balance of r-protein and rRNA levels. In *Escherichia coli*, there are 12 known distinct mRNA regulatory elements that control 11 r-protein operons, i.e.,  $\alpha$ , spc, S10, str, *rpsT*, *rpsB-tsF*, *rspA*, *rpmI-rplT*, *rpsO*, *rplK-rplA*, and *rplJ-rplL* (Fu et al. 2013).

According to the mimicry hypothesis proposed more than three decades ago by Nomura et al. (1980), regulatory binding sites on r-protein mRNAs have evolved to resemble native r-protein binding sites on rRNA. Numerous studies have shown that the similarity between these two types of binding sites is detectable at the structural level (Serganov et al. 2003; Merianos et al. 2004; Nevskaya et al. 2005), and sometimes they also share common sequence patterns (Olins and Nomura 1981; Iben and Draper 2008). The regulatory sites can be conserved in closely related bacteria (Robert and Brakier-Gingras 2001; Aseev et al. 2008), between different bacterial phyla (Fu et al. 2013), or even between domains of life (Kraft et al. 1999), but in many cases they appear to have originated independently in the individual taxonomic groups (Allen et al. 1999; Springer and Portier 2003).

S6 and S18 are bacteria-specific r-proteins with undefined functions (Melnikov et al. 2012). They bind cooperatively as a heterodimer to a specific 16S rRNA region previously stabilized by another ribosomal protein, S15 (Held et al. 1974; Agalarov et al. 2000; Recht and Williamson 2001). Since the structure of S18 is irregular, it has been suggested that it is capable of folding only in the context of S6 and/or the 16S rRNA-S15 complex (Agalarov et al. 2000). In most bacteria, the *rpsF* and *rpsR* genes—coding for S6 and S18, respectively—are located within one operon. However, no mechanism regulating

<sup>3</sup>Corresponding authors

E-mail [sdh@genesilico.pl](mailto:sdh@genesilico.pl)

E-mail [iamb@genesilico.pl](mailto:iamb@genesilico.pl)

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.038794.113>.

their expression has been proposed so far. In a high-throughput bioinformatics screening study, the Breaker group identified a conserved RNA motif associated with *rpsF* (Yao et al. 2007), suggesting presence of a regulatory mechanism. However, this study focused only on Firmicutes bacteria and did not elaborate on the structure or the potential role of this motif.

Using a comparative genomics approach, we have defined an RNA motif present in most bacterial phyla within operons encoding the S6 and S18 proteins. This motif adopts a hairpin-like structure that contains a predicted r-protein binding site, and it is characterized by the presence of a three-nucleotide CCG pattern. We found that this pattern is present in an analogous structural context in the S6:S18 binding site on the 16S rRNA, suggesting that the motif identified in the *rpsF* 5' UTR may also constitute a binding site for the S6:S18 complex. Using electrophoretic mobility shift and nitrocellulose filter binding assays, we demonstrated that the S6:S18 complex indeed binds specifically to the motif, which we hereafter refer to as the S6:S18 complex-binding motif (S6S18CBM). Moreover, the replacement of the CCG trinucleotide with AAA abolishes the interaction. A possible mechanism for the regulation of the *rpsF* operon by the S6S18CBM is discussed.

## RESULTS

### The 5'-UTR region of the *rpsF* operon contains a conserved RNA structural motif

Experimental data on prokaryotic operon architecture is too sparse for comparative genomics analyses; therefore, we predicted operons computationally for all fully sequenced genomes, using an intergenic distance criterion (see Materials and Methods). For further analysis, we considered only operons containing at least one copy of *rpsF*. Although gene order in the predicted operons differs across phyla, they almost always start with *rpsF* and usually contain *rpsR* coding for r-protein S18 (Table 1). Coexpression of these two proteins is favorable, as S6 and S18 form a native complex, which binds cooperatively to the 16S rRNA with bound S15 (Agalarov et al. 2000). Besides *rpsF* and *rpsR*, the predicted operons, depending on a phylum (Table 1), often contain the following genes: *priB* (coding for primosomal replication protein b), *rplI* (r-protein L9), *ssb* (single-stranded DNA binding protein), *rnb* (ribonuclease II), and *rpmG* (r-protein L33).

Analysis of the 5'-UTR sequences of the operons encoding S6 r-proteins led us to the discovery of the highly conserved S6S18CBM RNA motif. The S6S18CBM consists of two parts: an irregular hairpin of 7–25 bp (15 bp on average) and a bulged loop of 8–23 nt (Fig. 1). Despite divergence (i.e., deletions and insertions) across bacterial phyla, S6S18CBM shows a clear pattern of correlated substitutions, implying the relevance of the consensus structure (Figs. 1, 2). The base-paired region contains the consensus Shine-Dalgarno

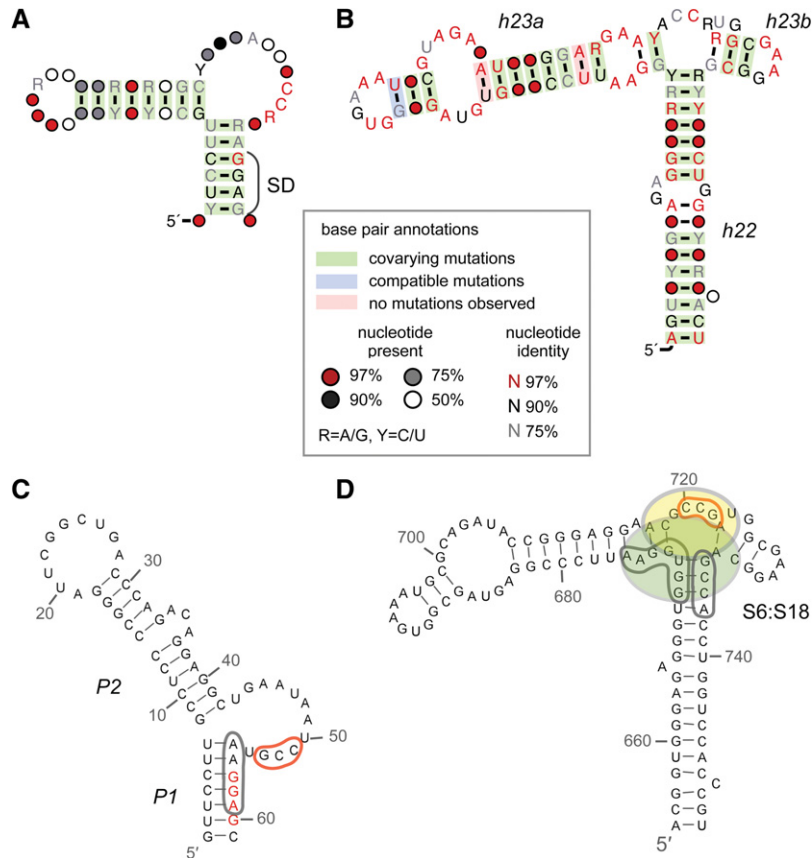
**TABLE 1.** Distribution of the S6S18CBM and most common *rpsF* operon structures across prokaryotic phyla

Taxonomic group	Occurrence of S6S18CBM	Most common operon architecture
$\alpha$ -Proteobacteria	–	<i>rpsF-rpsR-rplI</i>
$\beta$ -Proteobacteria	++	<i>rpsF-priB-rpsR-rplI</i>
$\gamma$ -Proteobacteria	+++	<i>rpsF-priB-rpsR-rplI</i>
$\delta/\epsilon$ -Proteobacteria	+++	<i>rpsF-ssb-rpsR</i> and <i>rspF-rpsR</i>
Actinobacteria	++	<i>rpsF-ssb-rpsR-rplI</i>
Bacteroidetes/Chlorobi	–	<i>rpsF-rpsR-rplI</i>
Cyanobacteria	+	<i>rpsF</i>
Firmicutes	+++	<i>rpsF-ssb-rpsR</i>
Spirochaetes	+	<i>rpsF-ssb-rpsR</i>
Tenericutes	++	<i>rpsF-ssb-rpsR</i>
Thermotogae	+	<i>rpsF-ssb-rpsR</i>
Archaea	–	<i>rps6e</i>

(–) The motif was found in <5% of the analyzed genomes of a given phylum; (+) in 5%–33%; (++) in 34%–66%; and (+++) in 67%–100%.

(SD) sequence at its 3' end, whereas the bulged loop includes a highly conserved short sequence pattern CCR (R = A/G) located 2–3 nt upstream of the SD sequence.

The newly discovered S6S18CBM motif is broadly distributed among bacteria. To investigate its taxonomic distribution, we represented the initially discovered motif instances as a covariance model and used it to scan all fully sequenced prokaryotic genomes (see Materials and Methods for details; see Supplemental Data 1 for the results). Results showed that despite its simplicity, the motif is specific for *rpsF* operon 5' UTRs, i.e., does not occur in other genomic locations. The distribution of the S6S18CBM instances across different phyla is presented in Table 1. It was found in most  $\gamma$ -Proteobacteria, Firmicutes, and Tenericutes; however, it is absent in Chlorobi,  $\alpha$ -Proteobacteria, and Archaea groups. Examination of DNA sequences upstream of chlorobial *rpsF* gene clusters revealed that the predicted transcriptional units do not contain SD sequences, suggesting an alternative mechanism of regulation. Although the S6S18CBM was not found in  $\alpha$ -Proteobacteria, the *rpsF* operon of *Sinorhizobium meliloti* contains a ~200-nt long 5' UTR (Ulve et al. 2007). However, its predicted secondary structure differs substantially from the S6S18CBM, suggesting that  $\alpha$ -Proteobacteria may employ a distinct regulatory system. Interestingly, the two bacterial phyla that do not contain the S6S18CBM motif share the same dominant operon organization, in which *rpsR* gene directly follows *rpsF* (Table 1). We analyzed all predicted bacterial *rpsF* operons and found that this organization strongly correlates ( $\chi^2$  *P*-value of  $2.92 \times 10^{-38}$ ) with the absence of a S6S18CBM motif in the 5' UTR, implying the presence of a general rule. The archaeal counterparts of *rpsF* and *rpsR* genes—*rps6e* and *rps26e* (Malygin and Karpova 2010)—are located separately in the genome, and in *Sulfolobus solfataricus*, both are expressed with short leaders (Wurtzel et al. 2010), which nonetheless do not resemble S6S18CBM.



**FIGURE 1.** Secondary structures of the S6S18CBM in *rpsF* operon mRNAs and the structurally corresponding junction in 16S rRNA. (A) Predicted RNA secondary structure of the S6S18CBM in *rpsF* operons. (B) Three-way junction of 16S rRNA—a binding region of the S6:S18 protein complex in the ribosome crystal structure. Conservation of individual nucleotides is calculated based on rRNA sequences from organisms with predicted S6S18CBM instances. (C) Proposed secondary structure of the S6S18CBM in *Escherichia coli*, with a depicted CCR trinucleotide motif. SD sequence is colored red; nucleotides forming the predicted S6:S18 binding interface are outlined in red and gray. (D) Three-way junction of the *Thermus thermophilus* 16S rRNA. Depicted is a S6:S18 binding interface; nucleotide residues contacting the complex in a base-specific manner are marked in red. Green and yellow transparent ovals indicate S6 and S18 proteins, respectively.

According to the recent study on RNA motifs regulating r-protein biosynthesis (Fu et al. 2013), such universal occurrence puts the S6S18CBM motif among most widespread *cis*-regulatory motifs in bacterial r-protein operon leaders.

### The CCR sequence pattern is conserved in 16S rRNA

During the assembly of the central domain of the ribosomal small subunit, the S6:S18 complex binds to the rRNA across the junction of helix (h) 22, h23a, and h23b (Fig. 1; Agalarov et al. 2000). The complex cannot bind (Held et al. 1974) or binds very weakly (Recht and Williamson 2001) to the junction in the absence of the S15 protein. S15 stabilizes the 16S rRNA tertiary structure and organizes the binding site for the S6:S18 complex (Agalarov et al. 2000), but does not interact with it directly. In the structure of *T. thermophilus* small ribosomal subunit, bulged-out nucleotides C720, C721,

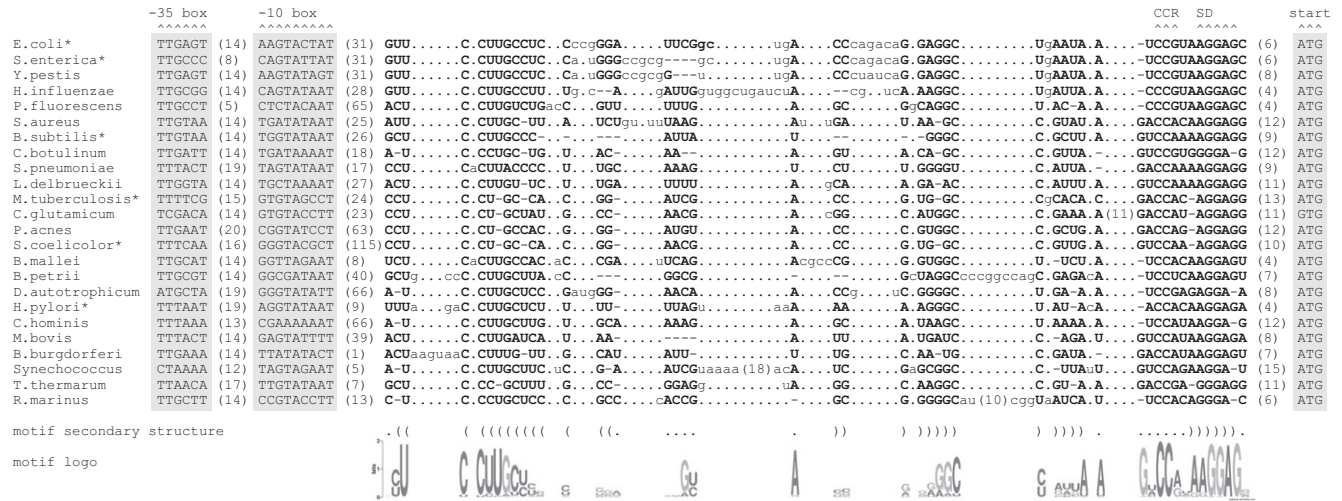
and G722 make four base-specific hydrogen bonds to residues Lys71 and Arg74 of r-protein S18 (Figs. 1, 3). These are the only base-specific contacts in the S6/S18-RNA interface; and importantly, the CCR (R = A/G) sequence is conserved in all genomes where the S6 RNA motif is found (Fig. 1). Furthermore, both amino acid residues involved in this interaction are highly conserved in S18 proteins (data not shown). The aforementioned evidence, in addition to the fact that the CCR sequence pattern is localized in equivalent structural positions in the S6S18CBM and 16S rRNA, suggests that it is the key element for the binding of the S6:S18 complex to both rRNA and to the S6S18CBM in mRNA.

It is difficult to assess the significance of the structural similarity between the S6S18CBM and the S6:S18 binding site in rRNA based solely on the comparison of secondary structure, because the protein-bound RNA conformation of S6S18CBM most likely mimics the conformation of the S6:S18 binding site in rRNA altered by the interaction with S15. To this end, and to better understand the potential role of the CCR pattern, we built a tertiary structure model of the *E. coli* S6S18CBM bound by the S6:S18 complex (Figs. 1, 3). The model shows conservation of the binding interface (Fig. 1C, D). In the protein-S6S18CBM model, S6, which contacts the minor groove of h22 and h23b in 16S rRNA (Fig. 1D, residues shown in gray), binds to the RNA

backbone of helix P1 (Fig. 1C, residues shown in gray) with the purine-rich tract containing the SD sequence (nucleotide residues 55–58 colored red). S18 contacts phosphate groups in the helix P1 and, analogously to the ribosome structure, makes base-specific contacts to the single-stranded CCR sequence (residues 51–53) within a bulged-out loop. This model suggests that despite differences in sequence and structure between rRNA and mRNA, their recognition by the S6:S18 complex may occur in a similar way.

### In *Escherichia coli* S6S18CBM is specifically recognized by the S6:S18 protein complex

Reconstitution studies of the *E. coli* 30S ribosomal subunit have shown that S6 r-protein binding depends on the presence of the S18 r-protein (Mizushima and Nomura 1970). Central domain assembly studies using the *Aquifex aeolicus* system revealed that S6 and S18 r-proteins form a tightly associated



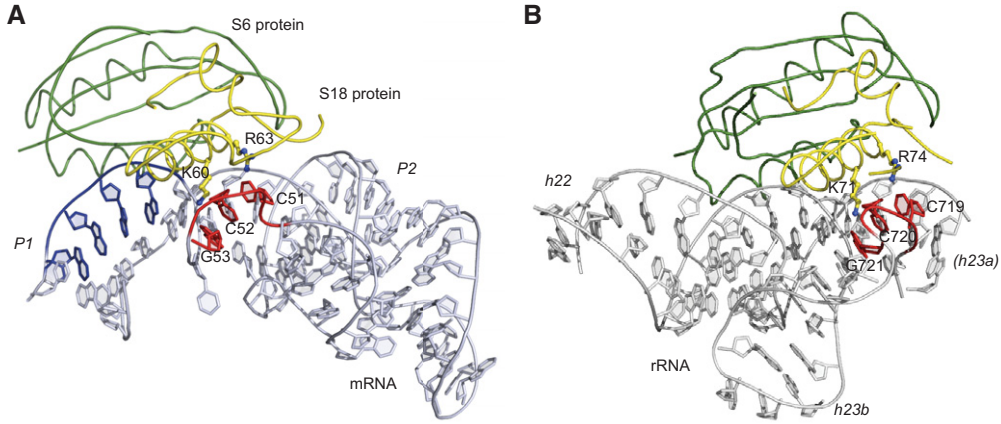
**FIGURE 2.** Constrained sequence alignment of representative sequences of the S6S18CBM. Sequences experimentally proved to be transcribed (Lindner et al. 2004; Sharma et al. 2010; Arnvig et al. 2011; Maciag et al. 2011; Vockenhuber et al. 2011; Ramachandran et al. 2012) are marked with an asterisk, and insertions are shown as lowercase letters and dots. Possible promoter -35 and -10 boxes and start codons of downstream rpsF are highlighted; unpaired conserved motif CCR and Shine-Dalgarno sequence are indicated.

heterodimer prior to binding to ribosomal RNA and showed no evidence for binding of either protein alone (Recht and Williamson 2001). Due to the fact that the structure of S18 is quite irregular, it is unlikely that S18 is folded alone, and it is thought to fold only upon binding to S6. Therefore, we decided to coexpress and copurify the *E. coli* S6:S18 complex. The hypothesis that the S6:S18 protein complex binds to the S6S18CBM RNA was tested experimentally using in vitro transcribed RNA and purified S6 and S18 proteins. Titration of the S6:S18 complex with the 5' UTR of the *E. coli rpsR* gene allowed for monitoring of the S6:S18-RNA complex formation by an electrophoretic mobility shift assay (EMSA). Addition of increasing amounts of S6:S18 causes a shift of the RNA band

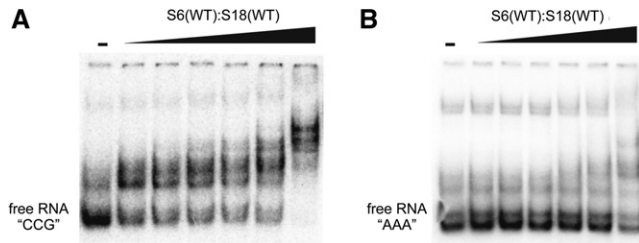
in the gel (Fig. 4A), indicating an interaction. The supershifted bands visible at S6:S18 concentrations >5 μM are presumably the result of unspecific binding of protein aggregates.

**The CCG sequence pattern in the S6S18CBM of RNA is required for recognition by the S6:S18 protein complex**

Because the unpaired CCR trinucleotide is present in both 16S rRNA and in the S6S18CBM, we asked whether it is essential for mRNA binding by the S6:S18 complex. To this end, we prepared a variant of the *E. coli* 5'-UTR sequence in which the CCG trinucleotide was replaced by AAA. This replacement



**FIGURE 3.** Interfaces of interactions between the S6:S18 complex and RNA. (A) Modeled interaction interface between S6:S18 and the S6S18CBM of the *rpsF* operon mRNA in *E. coli*. Residues proposed as recognition determinants (K60, R63 of S18 protein and C51, C52, G53 of S6S18CBM) are labeled. Nucleotides corresponding to the Shine-Dalgarno (SD) sequence are colored blue. Numbering of S6S18CBM residues follows that in Figure 1C. (B) Interaction interface between the S6:S18 r-protein heterodimer and 16S rRNA in the S15,S6,S18-rRNA complex of *Thermus thermophilus* (PDB code: 1G1X). Amino acid and nucleotide residues making specific contacts (K71, R74 of S18 protein and C719, C720, G721 of 16S rRNA) are labeled.



**FIGURE 4.** Binding of the S6:S18 r-protein complex to a 5'-UTR RNA [(A) wild-type, (B) mutant with "CCG" substituted by "AAA"] from the *E. coli rpsF* gene containing the S6S18CBM, monitored by EMSA. In vitro-transcribed  $^{32}\text{P}$ -labeled 5'-UTR RNA was incubated with an equimolar mixture of the wild-type S6:S18 complex (0.3, 0.7, 1.3, 2.7, 4.0, and 5.5  $\mu\text{M}$ ; indicated by wedges above the figure).

was predicted to have no significant effect on the secondary structure of the S6S18CBM by the RNA mutants program (Waldispühl et al. 2009). The mutant RNA was unable to bind the S6:S18 protein dimer according to both EMSA (Fig. 4) and filter-binding assay (data not shown), validating our hypothesis that the CCG trinucleotide is important for protein binding. This result is in agreement with the structural model that predicted that bases of the CCG trinucleotide are directly contacted by the S6:S18 protein complex. On the other hand, an RNA molecule with length and predicted secondary structure content similar to that of *rpsF* mRNA, was not bound by the S6:S18 protein complex (data not shown).

To analyze the importance of amino acid residues forming the base-specific contacts with RNA, two variants of the S6:S18 complex were created, carrying substitutions K60 or R63 to alanine in the S18 protein. The filter binding assay was used to measure the binding affinity of the two S6:S18 mutant complexes relative to the wild-type complex. Under the conditions tested, substitutions of Lys60 and Arg63 resulted in a decrease in the binding affinity, confirming the importance of the selected residues for RNA recognition. The wild-type S6:S18 complex binds to the 5'-UTR mRNA with a dissociation constant ( $K_d$ ) of 0.3  $\mu\text{M}$  ( $\pm 0.05$ ), whereas the variant with the K60A substitution binds with a  $K_d$  of 1.9  $\mu\text{M}$  ( $\pm 0.38$ ) (Supplemental Fig. 1). The saturation curve generated for the R63A variant could not be fitted using a one-site-specific binding model; and although it indicates very weak binding, no  $K_d$  value could be reliably calculated.

## DISCUSSION

### A newly discovered RNA motif, S6S18CBM, presumably participates in the translational regulation of the *rpsF* operon

In this study, we identified a conserved RNA motif, termed S6S18CBM, present in most bacterial phyla within 5' UTRs of *rpsF* operons encoding r-protein S6. The motif, which is characterized by a conserved CCG sequence presented in a bulge flanked by a stem and a hairpin structure, is located immediately upstream of the start codon of *rpsF* gene and in-

cludes the ribosome-binding site. Comparative sequence analysis and structural modeling led us to hypothesize that the S6S18CBM mimics the native S6:S18 complex binding site in 16S rRNA, and thus is capable of recognizing the S6:S18 complex. This hypothesis was confirmed experimentally using in vitro binding assays and the determinants of the interaction were identified.

We propose that the S6S18CBM competes with the ribosome for binding to the S6:S18 complex, and S6:S18 complex binding stabilizes the structure of the S6S18CBM, thereby leading to a conformation in which the SD sequence is sequestered (Fig. 1). This sequestration, in turn, may provide the *rpsF* operon with a regulatory feedback mechanism that ensures the balanced and coordinated synthesis of the proteins encoded. In analogous systems, associated with other r-protein operons, the regulatory effect is either entirely coupled (e.g., affects equally translation of all genes) or weakens in downstream genes (Nomura et al. 1980; Baughman and Nomura 1984). As most of the bacterial *rpsF* operons encode both components of the S6:S18 complex, the coupled mechanism of regulation appears to be more appealing. However, considering the fact that S18 folding depends on S6, it is also possible that only S6 translation is regulated, and the concentration of the functional form of S18 is controlled on the level of folding. Further in vivo experiments are required to assess the influence of S6:S18 binding to the 5' UTR on the expression levels of the individual *rpsF* operon proteins.

### RNA recognition by the S6:S18 complex

Despite no sequence resemblance, the predicted structure of the S6S18CBM is similar to the structure of the S6:S18 binding site in 16S rRNA. Furthermore, the S6:S18 complex appears to bind both RNAs using the same interface involving positively charged amino acid residues in S18 (Fig. 3). To gain more insight into the structural features governing the RNA recognition by the S6:S18 complex, we compared the S6S18CBM-S6:S18 model with the crystal structure of *T. thermophilus* small ribosomal subunit (Fig. 3). In the ribosome, the 16S rRNA-S6:S18 protein complex interface contains 19 electrostatic and hydrogen-bonding contacts; however, only four of them are base-specific (Agalarov et al. 2000). Interestingly, these positions involve nucleotide residues that are preserved in both the 16S rRNA and in the structurally corresponding region of the S6S18CBM (Figs. 1, 3). We therefore hypothesized that they may be essential for the recognition of the S6:S18 complex. Indeed, mutation of the CCG pattern in the *E. coli* S6S18CBM considerably reduced S6:S18 complex binding, indicating that these residues are essential for the interaction and may be involved in base-specific contacts, as predicted by our structural model. These findings demonstrate that structural similarity to a S6:S18 binding site in 16S rRNA is not sufficient for recognition, and base-specific contacts appear to play a key role in defining the binding specificity of the S6:S18 complex. This result is also in line with studies

showing that single-base bulges modulate the strength of interactions between mRNA and r-proteins S6 (Wu et al. 1994) and S15 (Serganov et al. 2003).

S6 and S18, as secondary r-proteins in ribosome assembly, bind to 16S rRNA only if it is structurally modified by a previous interaction with S15 (Agalarov et al. 2000). We showed that the S6:S18 complex alone is capable of binding to the S6S18CBM, suggesting that the S6S18CBM may mimic the rRNA region already altered by S15. This, in turn, indicates that in S15-depleted conditions, the expression of the *rpsF* operon should be down-regulated and suggests that the S6S18CBM may be involved in a more complex regulatory network.

## MATERIALS AND METHODS

### Sequence analysis

DNA sequences of 1306 fully sequenced prokaryotic genomes with annotated ncRNA and protein-coding genes were obtained from the NCBI database (Sayers et al. 2012). A sequence of genes transcribed from the same strand, with intergenic regions not longer than 100 nt and containing an *rpsF*, was considered as a gene cluster belonging to an *rpsF* operon (throughout the study, we define an operon as a polycistronic mRNA and its 5' UTR). Definition of *rpsF* genes was based on the KEGG Orthology (K02990 cluster) (Kanehisa et al. 2012). A nucleotide sequence between the first gene upstream of an *rpsF* operon gene cluster and the first gene of the cluster was considered as a 5' UTR. Sequences with a 5' UTR shorter than 40 nt were removed, and those longer than 400 nt were limited to 400 residues at the 5' end. The 5'-UTR sequences were searched for conserved RNA motifs with CMfinder (Yao et al. 2007), a local RNA sequence alignment algorithm based on covariance models. The program was started using two parameter sets: (1) returning up to five candidate motifs of length 30–100 nt and containing one stem-loop; and (2) returning up to five motifs of length 40–100 nt and containing two stem-loops. Overlapping motifs were subsequently combined using the CombMotif.pl script from the CMfinder package. To favor thermostable motifs present in many genomes, with conserved sequence and secondary structure, for each of the motifs a score was assigned according to the heuristic function proposed by Yao et al. (2007). The best-scored motif was used to calculate a covariance model, subsequently used to scan the entire genome sequences with cmsearch (Nawrocki et al. 2009) and e-value threshold of 0.01. Only three of 713 significant matches were present in another genomic context than the *rpsF* operon, implying high specificity of the obtained covariance model. Matches from *rpsF* operons were aligned with cmlign from the Infernal package, and their taxonomic distribution was analyzed. Full sequence alignment is available as Supplemental File 1.  $\sigma^{70}$  promoters associated with the representative motifs (Fig. 2) were predicted with BPROM (<http://linux1.softberry.com/berry.phtml>). To estimate the sequence conservation of the S6:S18 complex binding site, 16S rRNA sequences from the genomes containing the S6S18CBM were extracted and aligned using SSU-ALIGN (Nawrocki 2009). Consensus secondary structure diagrams were created based on alignments with R2R using GSC-weighting (Weinberg and Breaker 2011).

### Modeling the tertiary structure of the S6S18CBM-S6:S18 complex interaction

The 3D structure of the *E. coli* S6S18CBM (Fig. 1) in complex with the S6 and S18 proteins was modeled using methodology previously validated in the course of the RNA-puzzles experiment (Cruz et al. 2012). A part of an RNA helix and a loop were modeled using the comparative approach with ModeRNA (Rother et al. 2011). As the template, we used helix 22 and the junction region of the rRNA three-way junction from the crystal structure of the S15,S6,S18-rRNA complex from *T. thermophilus* (PDB code: 1G1X). Insertions were modeled de novo with MC-SYM (Parisien and Major 2008), with constraints on secondary structure and with a fragment-based approach using the RNABricks database of RNA structure fragments (G Chojnowski, unpubl.). Structures of *E. coli* S6 and S18 r-proteins were taken from the crystal structure of the 70S *E. coli* ribosome (PDB code: 3I1M). RNA and protein components of the modeled complex were evaluated with Model Quality Assessment Tools, i.e., RASP (Capriotti et al. 2011) and MolProbity (Davis et al. 2007). Poorly scored regions in RNA (in particular in regions modeled de novo and at the insertion/deletion sites) were refined with SimRNA, a method for RNA folding simulations that uses a coarse-grained representation and a statistical potential (Rother et al. 2012). The final full-atom model was obtained following local refinement using HyperChemPro 7.51 (Hypercube, Inc.) with the AMBER force field (Case et al. 2005), followed by global refinement using PHENIX (Adams et al. 2010). Atom coordinates of the model are available from <ftp://genesilico.pl/iamb/models/S6S18CBM/S6S18CBM&RNA.pdb>.

### RNA transcription

A DNA fragment coding for the *Escherichia coli* K12 5' UTR of the *rpsF* operon (93 nt) was amplified by PCR and cloned into the pTZ19R vector as an NcoI-XhoI fragment. Site-directed mutagenesis of the CCG motif was performed by a PCR-based technique. All constructs were linearized with XhoI, gel purified, and used as templates (1  $\mu$ g) for in vitro transcription (T7-Flashscribe Transcription Kit, CellScript) in the presence of 10  $\mu$ Ci of [ $\alpha^{32}$ P] guanosine-5'-triphosphate (GTP), performed according to the manufacturers' protocol. Following incubation for 2 h at 37°C, the template DNA was digested with 1 unit of RNase-free DNaseI for 15 min at 37°C. The labeled RNA molecules were phenol/chloroform extracted and passed through mini Quick Spin oligo columns (Roche) to remove free radionucleotides. RNA transcripts were resolved on 8% polyacrylamide/8 M urea gel and visualized by autoradiography. Full-length products were excised and extracted from the gel.

### Cloning, expression, and purification of S6 and S18 r-proteins

The *E. coli rpsF* and *rpsR* genes have been cloned previously with noncleavable N-terminal His6 tags into the recombinant expression plasmid pCA24N by Saka et al. (2005). The *rpsF* gene with T5 promoter originating from pCA24N was amplified in a PCR reaction and cloned into pET28a as a BglII-XbaI fragment, resulting in the pETrpsF. Subsequently, *rpsR* was cloned into pETrpsF as NcoI-XhoI fragment, resulting in a pETrpsFrpsR construct expressing full-length S6 with an N-terminal His6 tag and S18 without a

tag. Mutant variants of the *rpsR* gene corresponding to r-protein variants with K60A and R63A substitutions were constructed by site-directed mutagenesis in PCR reactions by amplifying the whole pCA24NrpsR plasmid. Mutant genes were sequenced and found to contain only the desired changes. The *E. coli* BL21(DE3) strain (Novagen) was used for protein overexpression. Cells were grown in LB medium containing 50 µg/mL kanamycin at 37°C. Protein expression was induced with 1 mM IPTG for 4 h at 37°C. Cells were harvested by centrifugation (4000g for 15 min, 4°C) and pelleted. The cell pellet was first washed with STE buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl and 1 mM EDTA), resuspended, and lysed by sonication in lysis buffer (50 mM Hepes, pH 8.0, 300 mM NaCl, 1 mM PMSF, 10 mM β-mercaptoethanol, 10% glycerol, 10 mM imidazole). The lysate was clarified by centrifugation at 40,000g for 30 min at 4°C, and the soluble fraction was incubated with His-Select Nickel Affinity Gel (Sigma) for 1 h at 4°C. Subsequently, the resin was washed with 10 bed volumes of wash buffer 1 (lysis buffer with 2 M NaCl), 5 bed volumes of wash buffer 2 (lysis buffer with 20 mM imidazole), and the proteins were eluted with elution buffer (lysis buffer with 250 mM imidazole). Samples of eluates were resolved in 20% SDS-PAGE, and protein concentration was determined using ImageQuant TL image analysis software (GE Healthcare). Proteins were expressed in a molar ratio close to 1:1.

### Electrophoretic mobility shift assay

A constant amount of internally <sup>32</sup>P-labeled RNA (0.1 µM) was incubated with increasing concentrations of equimolar mixture of S6/S18 proteins (0.3–5.5 µM). The RNA transcript was heated to 90°C for 2 min and slowly cooled down. The RNA:protein complexes were assembled in 20 mM Hepes-KOH (pH 7.6 at 25°C), 330 mM KCl, 10 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 2 µg/µL poly(dI-dC), 5 µg/mL heparin, and 0.01% nondenaturing detergent Igepal CA630. Reactions of 20 µL were incubated for 1 h at 25°C prior to addition of 4 µL of glycerol (50% w/v). The complexes were resolved by native PAGE. Gels were vacuum dried, exposed to a Storage Phosphor Screen (GE Healthcare), and visualized by PhosphorImaging (Typhoon Trio, GE Healthcare).

### Nitrocellulose filter binding assay

Binding reactions were performed with internally <sup>32</sup>P-labeled RNA (0.01 µM) and equimolar mixture of S6/S18 proteins (0.05–10 µM) in the presence of 20 mM Hepes-KOH (pH 7.6 at 25°C), 330 mM KCl, 10 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 2 µg/µL poly(dI-dC), 10 µg/mL bovine serum albumin (BSA), 5 µg/mL heparin, and 0.01% Igepal CA630. The nitrocellulose membrane was soaked in the wash buffer (20 mM Hepes-KOH [pH 7.6 at 25°C], 330 mM KCl, 10 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 5 µg/mL heparin, and 0.01% Igepal CA630) for at least 30 min, with agitation. Following 1-h incubation at 25°C, the reactions of 100 µL were spotted onto nitrocellulose membrane with Bio-Dot Microfiltration Apparatus (Bio-Rad). The membrane was washed four times with 100 µL of wash buffer, dried, and subjected to autoradiography. The obtained data were analyzed using the GraphPad Prism Version 6.01 for Windows software (GraphPad Software; <http://www.graphpad.com>). K<sub>d</sub> values were calculated by nonlinear curve fitting using a one-site specific binding model.

## SUPPLEMENTAL MATERIAL

Supplemental material is available for this article and contains the sequence alignment of the S6/S18CBM instances in the Stockholm format (Supplemental File 1.stk) and a plot showing relative binding affinity of S6/S18 variants measured using filter binding assay (Supplemental Fig. 1.pdf).

## ACKNOWLEDGMENTS

We thank Grzegorz Chojnowski for critical reading of the manuscript and for help with the RNABricks database, and Albert Bogdanowicz and Marcin Magnus for access to RNA Model Quality Assessment Tools. We thank Allana Schooley for language corrections. This work was supported by the Foundation of Polish Science (FNP) TEAM/2009-4/2 to J.M.B.; S.D.-H. has been supported by the 7th Framework Programme of the E.U. (contract number 286556 to J. M.B.); S.P. has been supported by the European Research Council (ERC) StG grant 261351 to J.M.B.; M.J.B. has been supported by Deutsche Forschungsgemeinschaft (grant number SE 1195/12-2, subcontract to J.M.B.). S.D.-H. also acknowledges a support from FNP (the START fellowship) and Polish Ministry of Science and Higher Education (fellowship for outstanding young scientists).

Received March 12, 2013; accepted July 5, 2013.

## REFERENCES

- Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**: 213–221.
- 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.
- Allen T, Shen P, Samsel L, Liu R, Lindahl L, Zengel JM. 1999. Phylogenetic analysis of L4-mediated autogenous control of the S10 ribosomal protein operon. *J Bacteriol* **181**: 6124–6132.
- Arnvig KB, Comas I, Thomson NR, Houghton J, Boshoff HI, Croucher NJ, Rose G, Perkins TT, Parkhill J, Dougan G, et al. 2011. Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of *Mycobacterium tuberculosis*. *PLoS Pathog* **7**: e1002342.
- Aseev LV, Levandovskaya AA, Tchufistova LS, Saptsova NV, Boni IV. 2008. A new regulatory circuit in ribosomal protein operons: S2-mediated control of the *rpsB-tsF* expression in vivo. *RNA* **14**: 1882–1894.
- Babitzke P, Baker CS, Romeo T. 2009. Regulation of translation initiation by RNA binding proteins. *Annu Rev Microbiol* **63**: 27–44.
- Baughman G, Nomura M. 1984. Translational regulation of the L11 ribosomal protein operon of *Escherichia coli*: Analysis of the mRNA target site using oligonucleotide-directed mutagenesis. *Proc Natl Acad Sci* **81**: 5389–5393.
- Capriotti E, Norambuena T, Marti-Renom MA, Melo F. 2011. All-atom knowledge-based potential for RNA structure prediction and assessment. *Bioinformatics* **27**: 1086–1093.
- Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. 2005. The Amber biomolecular simulation programs. *J Comput Chem* **26**: 1668–1688.
- Cruz JA, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cao S, Das R, Ding F, Dokholyan NV, Flores SC, et al. 2012. RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction. *RNA* **18**: 610–625.
- Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB III, Snoeyink J, Richardson JS, et al.

2007. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**: W375–W383.
- Freedman LP, Zengel JM, Archer RH, Lindahl L. 1987. Autogenous control of the S10 ribosomal protein operon of *Escherichia coli*: Genetic dissection of transcriptional and posttranscriptional regulation. *Proc Natl Acad Sci* **84**: 6516–6520.
- Fu Y, Deiorio-Hagggar K, Anthony J, Meyer MM. 2013. Most RNAs regulating ribosomal protein biosynthesis in *Escherichia coli* are narrowly distributed to Gammaproteobacteria. *Nucleic Acids Res* **41**: 3491–3503.
- 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.
- Held WA, 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 JR, Draper DE. 2008. Specific interactions of the L10(L12)<sub>4</sub> ribosomal protein complex with mRNA, rRNA, and L11. *Biochemistry* **47**: 2721–2731.
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40**: D109–D114.
- Kraft A, Lutz C, Lingenhel A, Gröbner P, Piendl W. 1999. Control of ribosomal protein L1 synthesis in mesophilic and thermophilic archaea. *Genetics* **152**: 1363–1372.
- Lindner C, Nijland R, van Hartskamp M, Bron S, Hamoen LW, Kuipers OP. 2004. Differential expression of two paralogous genes of *Bacillus subtilis* encoding single-stranded DNA binding protein. *J Bacteriol* **186**: 1097–1105.
- Maciag A, Peano C, Pietrelli A, Egli T, De Bellis G, Landini P. 2011. *In vitro* transcription profiling of the  $\sigma^5$  subunit of bacterial RNA polymerase: Re-definition of the  $\sigma^5$  regulon and identification of  $\sigma^5$ -specific promoter sequence elements. *Nucleic Acids Res* **39**: 5338–5355.
- Malygin AA, Karpova GG. 2010. Structural motifs of the bacterial ribosomal proteins S20, S18 and S16 that contact rRNA present in the eukaryotic ribosomal proteins S25, S26 and S27A, respectively. *Nucleic Acids Res* **38**: 2089–2098.
- Melnikov S, Ben-Shem A, Garreau de Loubresse N, Jenner L, Yusupova G, Yusupov M. 2012. One core, two shells: Bacterial and eukaryotic ribosomes. *Nat Struct Mol Biol* **19**: 560–567.
- Merianos HJ, Wang J, Moore PB. 2004. The structure of a ribosomal protein S8/spc operon mRNA complex. *RNA* **10**: 954–964.
- Mizushima S, Nomura M. 1970. Assembly mapping of 30S ribosomal proteins from *E. coli*. *Nature* **226**: 1214–1218.
- Nawrocki EP. 2009. “Structural RNA homology search and alignment using covariance models.” PhD thesis, Washington University in Saint Louis, Saint Louis, MO.
- Nawrocki EP, Kolbe DL, Eddy SR. 2009. Infernal 1.0: Inference of RNA alignments. *Bioinformatics* **25**: 1335–1337.
- Nevskaya N, Tishchenko S, Gabdoulkhakov A, Nikonova E, Nikonov O, Nikulin A, Platonova O, Garber 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 JL, Dean D, Post LE. 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.
- Olins PO, Nomura M. 1981. Regulation of the S10 ribosomal protein operon in *E. coli*: Nucleotide sequence at the start of the operon. *Cell* **26**: 205–211.
- Parisien M, Major F. 2008. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**: 51–55.
- Philippe C, Eyermann F, Benard L, Portier C, Ehresmann B, Ehresmann C. 1993. Ribosomal protein S15 from *Escherichia coli* modulates its own translation by trapping the ribosome on the mRNA initiation loading site. *Proc Natl Acad Sci* **90**: 4394–4398.
- Portier C, Dondon L, Grunberg-Manago M. 1990. Translational auto-control of the *Escherichia coli* ribosomal protein S15. *J Mol Biol* **211**: 407–414.
- Ramachandran VK, Shearer N, Jacob JJ, Sharma CM, Thompson A. 2012. The architecture and ppGpp-dependent expression of the primary transcriptome of *Salmonella* Typhimurium during invasion gene expression. *BMC Genomics* **13**: 25.
- 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.
- Robert F, Brakier-Gingras L. 2001. Ribosomal protein S7 from *Escherichia coli* uses the same determinants to bind 16S ribosomal RNA and its messenger RNA. *Nucleic Acids Res* **29**: 677–682.
- Roberts E, Sethi A, Montoya J, Woese CR, Luthe-Schulten Z. 2008. Molecular signatures of ribosomal evolution. *Proc Natl Acad Sci* **105**: 13953–13958.
- Rother M, Milanowska K, Putoń T, Jeleniewicz J, Rother K, Bujnicki JM. 2011. ModeRNA server: An online tool for modeling RNA 3D structures. *Bioinformatics* **27**: 2441–2442.
- Rother K, Rother M, Boniecki M, Putoń T, Tomala K, Lukasz P, Bujnicki JM. 2012. Template-based and template-free modeling of RNA 3D structure: Inspirations from protein structure modeling. In *RNA 3D structure analysis and prediction* (ed. Leontis NB, Westhof E), pp. 67–90. Springer-Verlag, Berlin.
- Saka K, Tadenuma M, Nakade S, Tanaka N, Sugawara H, Nishikawa K, Ichiyoshi N, Kitagawa M, Mori H, Ogasawara N, et al. 2005. A complete set of *Escherichia coli* open reading frames in mobile plasmids facilitating genetic studies. *DNA Res* **12**: 63–68.
- Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvermin V, Church DM, Dicuccio M, Federhen S, et al. 2012. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **40**: D13–D25.
- Serganov A, Polonskaia A, Ehresmann B, Ehresmann C, Patel DJ. 2003. Ribosomal protein S15 represses its own translation via adaptation of an rRNA-like fold within its mRNA. *EMBO J* **22**: 1898–1908.
- Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, et al. 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**: 250–255.
- Springer M, Portier C. 2003. More than one way to skin a cat: Translational autoregulation by ribosomal protein S15. *Nat Struct Biol* **10**: 420–422.
- Thomas MS, Nomura M. 1987. Translational regulation of the L11 ribosomal protein operon of *Escherichia coli*: Mutations that define the target site for repression by L1. *Nucleic Acids Res* **15**: 3085–3096.
- Ulvé VM, Sevin EW, Chéron A, Barloy-Hubler F. 2007. Identification of chromosomal  $\alpha$ -proteobacterial small RNAs by comparative genome analysis and detection in *Sinorhizobium meliloti* strain 1021. *BMC Genomics* **8**: 467.
- Vockenhuber MP, Sharma CM, Statt MG, Schmidt D, Xu Z, Dietrich S, Liesegang H, Mathews DH, Suess B. 2011. Deep sequencing-based identification of small non-coding RNAs in *Streptomyces coelicolor*. *RNA Biol* **8**: 468–477.
- Waldispuhl J, Devadas S, Berger B, Clote P. 2009. RNAmutants: A web server to explore the mutational landscape of RNA secondary structures. *Nucleic Acids Res* **37**: W281–W286.
- Weinberg Z, Breaker RR. 2011. R2R—software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics* **12**: 3.
- Wu H, Jiang L, Zimmermann RA. 1994. The binding site for ribosomal protein S8 in 16S rRNA and *spc* mRNA from *Escherichia coli*: Minimum structural requirements and the effects of single bulged bases on S8–RNA interaction. *Nucleic Acids Res* **22**: 1687–1695.
- Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R. 2010. A single-base resolution map of an archaeal transcriptome. *Genome Res* **20**: 133–141.
- 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.