Ribosomal RNA Mimicry in RNA Regulation of Gene Expression

Michelle M. Meyer

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

Email: m.meyer@bc.edu

ABSTRACT

The ribosomal RNA (rRNA) is the largest and most abundant RNA in bacterial and archaeal cells. It is also one of the best-characterized RNAs in terms of its structural motifs and sequence variation. Production of ribosome components including over 50 ribosomal proteins (r-proteins) consumes significant cellular resources. Thus RNA cis-regulatory structures that interact with r-proteins to repress further r-protein synthesis play an important role in maintaining appropriate stoichiometry between r-proteins and rRNA. Classically, such mRNA structures were thought to directly mimic the rRNA. However, over 30 years of research has demonstrated that a variety of different recognition and regulatory paradigms are present. This chapter will demonstrate how structural mimicry between the rRNA and mRNA cis-regulatory structures may take many different forms. The collection of mRNA structures that interact with r-proteins to regulate r-protein operons are best characterized in *E. coli*, but are increasingly found within species from nearly all phyla of bacteria and several archaea. Furthermore, they represent a unique opportunity to assess the plasticity of RNA structure in the context of RNA-protein interactions. The binding determinants imposed by r-proteins to allow regulation can be fulfilled in many ways. Some r-protein interacting mRNAs are immediately obvious as rRNA mimics from primary sequence similarity, others are identifiable only after secondary or tertiary structure determination, and some show no obvious similarity. In addition, across different bacterial species a host of different mechanisms of action have been characterized showing that there is no simple one size fits all solution.

Despite the many roles for RNA as a regulator in eukaryotes, archaea, and bacteria, the ribosomal RNA (rRNA) is the most abundant cellular RNA and the size of the rRNA outstrips nearly all other functional RNAs. Furthermore, the ribosome is also composed of over 50 ribosomal proteins (r-proteins); the majority of which directly contact the rRNA forming specific interactions with RNA (1). Since most regulatory RNAs in bacteria appear to be relatively recent inventions (2-5), they most certainly have evolved in the context of abundant rRNA and r-proteins, and thus have been shaped by them. Many regulatory RNA structures contain portions that bear strong resemblance to motifs within the rRNA. Some of this similarity is due to the role that rRNA plays in our understanding of RNA structure, and in other cases it is due to interaction with an r-protein. This review will first illustrate the role of the ribosome in our understanding of RNA structures generally and subsequently examine how r-proteins may interact with RNA outside the ribosome to act in a regulatory capacity.

THE RIBOSOMAL RNA AS A SOURCE OF RNA STRUCTURAL MOTIFS

The rRNA plays an outsized role in our general understanding of RNA structure. Despite over a decade since publication of the initial high-resolution ribosome structures and significant growth in the number and diversity of RNA structures in the Protein Data Bank, the rRNA still represents a significant proportion of the three-dimensional structure information available for RNA and RNA-protein complexes. Of the 3692 structures containing RNA, 1082 contain segments derived from the rRNA or otherwise associated with the ribosome. The ribosome has also significantly influenced the development of RNA structure descriptions (6). Many recurring RNA structure motifs such as kink-turns, loop-E, and loop-C motifs (7-11) were first recognized in the context of the ribosome and our knowledge of the sequences that may fold into many such

features is heavily influenced by rRNA alignments (12-14). These structural motifs form the basis of not only the rRNA but many other structured RNAs including riboswitches (15), T-boxes (16), as well as other catalytic RNAs such as the group I and II introns (17, 18). Several reviews specifically addressing the roles such motifs play in RNA structure are available (19-21).

RIBOSOMAL PROTEINS AS AUTOGENOUS REGULATORS

Many ribosomal proteins (r-proteins) have secondary functions (22, 23) as negative regulators of their own synthesis. R-proteins and other protein components necessary for translation can account for up to 40% of cellular proteins (24) and 41% of active translation in actively growing cells in rich medium (25). Thus maintaining stoichiometry among the over 60 ribosome components is essential for efficient resource utilization, and the mRNA structures responsible for implementing regulation are only one of several regulatory layers. In *E. coli* over half of the r-protein operons are regulated by autogenous regulatory mechanisms where an individual r-protein will bind to a portion of its own transcript to inhibit transcription or translation. Often the mRNA will take a structure that bears significant similarity to the rRNA, however, there are several different paradigms for RNA-protein recognition that are embodied by the mRNA structures that mediate r-protein autogenous regulation.

Discovery of R-Protein Autogenous Regulatory mRNA Structures

The mRNA structures enabling regulation of r-protein synthesis in *E. coli* were among the first mRNA regulatory sites discovered. Many distinct *E. coli* examples were described based on similar observations and using the same experimental approaches. Initial studies demonstrated that over-expression of specific r-proteins resulted in inhibited synthesis of entire r-protein operons (26, 27), and that these effects were operon specific (28). Using *in vitro*

transcription/translation systems as well as reporter gene assays, the inhibitory properties for several r-proteins including L1, L4, S4, S7, S8 and the L10(L12)$_4$ complex were uncovered (29-33). Most mechanisms involve inhibition of translation (30, 34-37), however, alterations to the mRNA decay rate (27, 38-40) and attenuation (premature transcription termination) mechanisms also occur in conjunction with translational inhibition (41, 42).

In many cases mimicry between the mRNA regulatory sites responding to an r-protein and its rRNA binding site was proposed as soon as a DNA sequence became available (e.g. S4, S7, S8, L1, L4, and L10 (35, 43-46)). However, demonstration of direct RNA-protein contacts that such similarity would imply lagged the speculation considerably (47-49). In several cases proposed similarities were merely the result of sequence gazing and it has become apparent that the rRNA and mRNA binding sites do not have structural similarity (e.g. S4, L4)(50, 51). In other cases, the initially observed similarity between the mRNA and rRNA was verified when three-dimensional structural data became available (e.g. S8, L1)(52, 53).

Since the initial discoveries of r-protein autogenous mRNA structures in *E. coli*, an additional nine mRNA structures responding to r-proteins (S1, S2, S15, S20, L19, L20 (2 sites), L25) (54-60) or r-protein complexes (S6:S18) (61-63) have been described in *E. coli* and today there are a total of 15 r-protein interacting mRNA structures described in *E. coli* (Figure 1A, Table 1). Many have been extensively characterized, but for others the mechanisms of action, or even whether a direct RNA-protein interaction occurs, remain undetermined. With some exceptions, the complement of r-proteins and organization of r-protein operons is largely conserved across bacterial species (64, 65). However, many of the structures allowing regulation in *E. coli* are not widely distributed to organisms outside of a few orders of gammaproteobacteria (66-71). Furthermore, most enterobacterial endosymbionts appear to have lost these structures

during the course of genome reduction (71, 72). The only organism with significant study of r-protein regulation other than *E. coli* is the gram-positive model bacterium *Bacillus subtilis*. This organism shares the mRNA-binding sites that interact with r-proteins L1, L10, S2, and S6 with *E. coli* (Figure 1, Table 1), but the other 11 structures known in *E. coli* are not apparent in *B. subtilis* or its relatives. Alternative regulatory structures that respond to S4, L20, and S15 have been described (73-75) (Figure 1B, Table 1).

With the growing number and diversity of sequenced bacterial genomes, comparative genomics has also proved to be a powerful approach for discovery. The combination of RNA-specific homology search tools (76) and the availability of RNA structural families corresponding to most known r-protein responsive structures (71, 75, 77) enables accurate annotation of these structures in bacterial genomes. In addition to characterized mRNA structures, hundreds of novel putative cis-regulatory mRNA motifs have been identified in bacterial genomes, many of which are associated with r-proteins or bear resemblance to the rRNA (78-82). The low-cost of sequencing has also enabled the direct discovery of regulatory RNAs through comparative transcriptomics (4), 5'-end sequencing (83), and RNA-protein immunoprecipitations (84). However, relatively few such motifs have been experimentally validated.

The S8-Interacting mRNA Structure: A Prototype R-Protein Cis-Regulatory RNA

The mRNA segment bound by r-protein S8 to regulate the *spc* operon is the prototype mRNA binding motif that embodies all the properties initially hypothesized for all mRNA structures bound by r-proteins. S8 is a primary rRNA binding protein that interacts with the rRNA early during ribosome assembly. The interaction site for S8 on the mRNA is within the

intergenic region between *rplX* and *rplE* (encoding L24 and L5) and the coding region of *rplE* (Figure 1A). S8 inhibits translation of several proteins following the protein binding-site (L5, S14, S8, L6, L18, S5, L20, L15) and there is evidence that the two genes upstream of the S8 binding region (*rplN* and *rplX*, encoding L14 and L24) are also down-regulated in response to S8 due to increased mRNA degradation (40, 85).

The initial observed sequence similarity between the rRNA and mRNA binding sites for S8 extends to shared secondary structure (35, 86-88) and three-dimensional structure (52, 86) (Figure 2). The S8 binding site consists of an internal loop. The motif centers on two internal Watson-Crick base-pairs that are separated from the rest of the pairing element by bulged bases on either side, although many of the base identities are not strongly conserved in the case of the rRNA (Figure 2B). S8 itself directly contacts the minor groove of the internal loop. Structures of the mRNA and rRNA are directly superimposable (Figure 2C) (52). The major difference between the rRNA and mRNA binding sites is an additional bulged base in the mRNA structure a few nucleotides away from the S8 recognition sites (orange). While this base decreases binding affinity by about 10-fold, it does not directly interact with S8. Despite a highly conserved rRNA-S8 interface across all bacteria (88-92) and archaea (93), the S8 responsive regulatory RNA structure observed in *E. coli* is narrowly distributed a few orders of gammaproteobacteria (71). What if any regulation occurs in other organisms has not yet been characterized and the causes of the narrow distribution are unclear. The phylogenetic distribution of the S8-interacting mRNA structure is similar to those of many r-protein mRNA regulators identified in *E. coli*, suggesting that similar selective constraints influenced the evolution of all the regulatory structures. The preponderance of known structures in *E. coli* is likely due to a significant discovery bias. Ribosome assembly and stoichiometry is by far the best studied in *E. coli*. Similar regulators

may be present, but as of yet unidentified in other bacteria. The narrow distribution displayed by most of the *E. coli* structures makes it more difficult to utilize comparative genomic approaches for discovery, and it is likely that several of the characterized motifs in *E. coli* would not be easily re-discovered using state of the art comparative genomic tools.

The L10(L12)$_4$-Interacting Regulatory Structure: Homologous Binding Sites, Different Mechanisms of Action

The L10(L12)$_4$ interacting mRNA structure also represents a mimic of the rRNA (47, 48, 94-96) and participates in the regulation of translation initiation in *E. coli* (34, 97) directly impacting only *rplJ* and *rplL* (Figure 1A). Sequence similarity between the mRNA and rRNA binding sites has been described (46, 68), but the L10(L12)$_4$ complex is typically not resolved in ribosome crystal structures and three-dimensional data for an mRNA-L10(L12)$_4$ complex is not available. The L10(L12)$_4$ binding site consists of a kink-turn motif that is four base-pairs away from an internal loop containing a pair of adenosines. In the rRNA the internal loop is a multi-stem junction (Figure 3), while in the mRNA the structure it is often a bulge, but may be a multi-stem junction (71). In both the rRNA, and the mRNA the adenosines are highly conserved and mutating them reduces binding affinity substantially (68).

In contrast to the S8-interacting mRNA structure, the RNA structure responsible for interacting with L10 in *E. coli* is widely conserved throughout many bacterial species (68, 71, 80). However, the mechanism of action is not the same across all species. In many gram-positive species the L10-interacting structure is followed by an intrinsic transcription terminator (80) and the mechanism of regulation in *Bacillus subtilis* is regulation of transcription termination (98).

Thus, r-protein binding structures are similar to riboswitches where homologous sensor domains may utilize different mechanisms of action in diverse species (99).

The L1-Interacting mRNA Structure: Convergence on the Same Binding Determinants

Like the L10- and S8- responsive mRNA structures, the L1-interacting mRNA structure shows obvious similarity to the rRNA (45, 53, 100) and examples of the L1 recognition site are found across nearly all bacterial phyla (71) as well as archaea (101-103). The binding determinants for L1 are often accommodated in a short hairpin of < 30 nucleotides and consist of a base-paired region containing an asymmetric internal loop closed by a non-canonical A•G pairing (Figure 4A). In three dimensions this corresponds to two canonical helixes, one of which is capped by the non-canonical A•G pair, that are separated by a sharp turn (53). Diverse L1 homologs are able to interact with an example of the mRNA binding site from *Methanococcus vannielii* (103), and structural data show that the rRNA and mRNA sites are nearly superimposable (53).

In *E. coli* the binding site is within the 5'-UTR of the transcript encoding both *rplK* and *rplA* and L1 regulates translation initiation of L11 and L1 (45) (Figure 4). Surprisingly in the archaea *M. vannielii, M. jannaschii,* and *M. thermolithotrophicus* the L1 binding site appears approximately 30 nucleotides inside the coding region for L1 and regulates translation of L1, L10, and P1 (homolog of L12 (104)) and the gene encoding L11 occurs elsewhere in the genome (101, 103). In *Sulfolobus solfataricus* the L1 binding site is found within the L11 coding region, which directly precedes genes encoding L1, L10, and P1 (Figure 4B) (103).

In addition to examples that have been explicitly examined, a systematic homology search for L1 binding sites in bacterial genomes identified the site within transcripts encoding L1

and L11 in many bacterial species (71). However like the examples identified in Archaea, the location of the binding site relative to the coding regions is not consistent. In Cyanobacteria, Actinobacteria, and Chloroflexi, the L1 binding site precedes *rplA*, typically between *rplK* and *rplA*. In Proteobacteria, Spirochaetes, Thermotogoa and Tenericutes the binding site precedes *rplK,* presumably to regulate both *rplA* and *rplK*. Furthermore, in many species of Firmicutes, L1 binding sites appear preceding both *rplA* and *rplK*. In *Geobacillus kaustophilus* both sites are capable of binding L1 *in vitro* (71). Interestingly, there is evidence of loss for each individual binding site within species scattered throughout Firmicutes (Figure 4B). The combination of the wide distribution and changing position of the L1 binding site relative to the regulated genes suggest that the site may have evolved convergently in many species.

L20-Interacting mRNA Regulatory Structures: Diverse Scaffolds Support the Same Binding Determinants

In addition to cases where there is a single mRNA binding site that mimics the rRNA, there are also cases where homologous r-proteins interact with distinct mRNA secondary structures in different bacterial species. Three L20 interacting mRNA structures are known, two in *E. coli* and one in *B. subtilis*. Each structure mimics the rRNA, but uses a different arrangement of secondary structure to support the necessary bases in the correct geometry required for recognition (Figure 5). In *E. coli,* two L20-responsive mRNA structures control the IF3 operon (*infC, rpmI,* and *rplT*, encoding IF3, L35 and L20). One structure is found within the intergenic region between *infC* and *rpmI* (70)(Figure 5A), and consists of a relatively straightforward bulged stem loop where the binding site includes a pair of adenosines within the bulge and a set of consecutive G-C base-pairs just after the closing base-pair of the loop (Figure

5C, mRNA-II). This arrangement is the closest mimic of the rRNA. The second structure is comprised of a pseudoknot formed by long-range interactions between a sequence within *infC* and sequence adjacent to the start of the *rpmI* coding region (54, 70) (Figure 5C, mRNA-I). In this structure the pair of adenosines is found in the single-stranded region just prior to the 3'-most portion of the pseudoknot. Both of these structures are required for full translational repression of the operon *in vivo* and L20 binds independently to each (70, 105). A high quality alignment and phylogenetic distribution is only available for the mRNA structure preceding *rpmI*. The pseudoknotted binding-site is challenging to identify using RNA-specific homology search tools (106) due to its significant overlap with coding sequence, long-range interactions, and pseudoknotted structure. However, the structure preceding *rpmI* is narrowly distributed to gammaproteobacteria.

In addition to the two L20-responsive structures in *E. coli*, L20 binds to a regulatory structure in *B. subtilis* that precedes *infC* (Figure 5B,C). While this structure shares many features with the L20-interacting structures identified in *E. coli*, the binding features present near the multi-stem junction are supported by a different arrangement of secondary structure, and the ordering of the elements with respect to one another in a linear sequence is distinct. A potential intrinsic transcription terminator follows this mRNA structure, and the mechanism of action is L20 induced structural change resulting in early transcription termination (74). In this structure the conserved adenosines are in a single stranded region just 5' of the first hairpin and the G-C pairs in the second hairpin. This structure is found in most Firmicutes (75, 80), although more frequently in the class *Bacilli* than in *Clostridia* and the transcription attenuation mechanism appears conserved in these organisms. While *infC* is part of this operon, in *B. subtilis* IF3 levels are decoupled from those of L20 and L35 through the presence of a second upstream promoter.

The transcript produced from this promoter is cleaved by RNAseY and only allows translation of L20 and L35 translation (107).

The three L20 sites all present the same effective binding geometry for L20 recognition (Figure 5C). Both *E. coli* structures are capable of interacting with an L20 homolog from *Aquifex aeolicus* to repress gene expression from *rpmI'-'lacZ* reporter constructs (108), and the *E. coli* homolog is able to stimulate premature transcription termination during *in vitro* assays with the *B. subtilis* mRNA structure (74). Thus, the L20 regulators serve as an example of how the same three-dimensional geometry may be supported by different arrangements of secondary structure elements. This example also illuminates how challenging identification of common binding sites may be. Despite similar binding determinants, the distinct arrangements of the necessary recognition elements make automatic detection difficult or impossible.

S15-Interacting Regulatory Structures: Diverse Binding Determinants Produce Diverse Structures

R-protein S15 also regulates gene expression using multiple distinct mRNA binding sites in diverse bacterial species. To date four different S15-interacting mRNA structures spanning several bacterial phyla have been experimentally characterized (109-112) and several additional putative structures identified (112) (Figure 6). Each of these structures directly precedes and controls expression of *rpsO*, the gene encoding S15. In *E. coli* the mechanism of action is through entrapment of the translation initiation complex (110), but in other species the mechanism has not explicitly been characterized.

The structures share very little in the way of a single recognizable sequence or structural motif. This is partially due to the bidentate nature of the S15 binding site on the rRNA. S15

recognizes two portions of the 16S rRNA: a multi-stem junction, and a stem containing a slight defect characterized by a G•U/G-C set of base pairs directly adjacent to the junction (Figure 6). The mRNA regulatory structures that interact with S15 often only partially mimic this binding site. For example, the mRNA from *Thermus thermophilus* includes a three-stem junction formed by the bases of three adjacent pairing elements. However, the pairing elements themselves show no evidence for the G•U/G-C defect recognized by S15 in the rRNA. In contrast mRNA structures described from *E. coli*, *Rhizobium radiobacter* (formerly *Agrobacterium radiobacter*), and *Geobacillus stearothermophilus* (previously *Bacillus stearothermophilus*) include more obvious mimics of the G•U/G-C elements, and require this element for interaction. In several cases the mRNAs have additional recognition elements that are necessary, but do not directly mimic the rRNA (Figure 6)(109, 113, 114).

The differences between the mRNAs are sufficiently large such that specificity of interaction has been reported (113, 115). For example the S15 homolog from *G. kaustohphilus* does not interact with the mRNA structure originating from *E. coli* and the S15 homolog from *T. thermophilus* does not intact with several of the mRNA structures containing only the G•U/G-C motif and no mimic of the three-stem junction (113, 115). Mutagenesis studies indicate that the same face of S15 appears to be used for interaction (115, 116). However, in *E. coli* different S15 amino acids are implicated in rRNA and mRNA binding (114, 117). Furthermore, selective recognition of the *Geobacillus* and *E. coli* mRNA structures may be traced to specific amino acids that are differentially conserved in S15 homologs originating from organisms containing RNAs of each type (113). Thus, the diversity of S15-interacting structures is not only due to the bidentate recognition site that may allow a larger set of potential interaction partners, but also to differences in the protein homologs that change recognition. These findings suggest that despite

very similar rRNA recognition sites across all bacteria, the r-proteins and their mRNA binding sites are influencing each other's evolution.

BEYOND AUTOGENOUS REGULATION: L7Ae

Archaeal r-protein L7Ae participates processes in well beyond of its role in the ribosome. L7Ae interacts with kink-turn (k-turn) and k-loop motifs as a component of the ribosome (1), RNase P (118), the C/D box and H/ACA box snoRNPs responsible for site-selective 2'-O-methylation (119, 120), and in mammals a L7Ae homolog binds to the U4 snRNP of the spliceosome (121). There is no r-protein that directly corresponds to L7Ae in prokaryotes. Two L7Ae homologs in *B. subtilis* exist, and both bind to kink-turns (122), but their biological function is unknown. The role of the k-turn as a fundamental RNA structural building block has already been discussed. L7Ae specifically recognizes this motif, and therefore has a role in many RNA complexes, primarily to stabilize RNA structure.

A recent RIP-seq study of L7Ae in *Sulfolobus acidocaldarius* identified several mRNA fragments in addition to the expected interaction partners (84). Many of these mRNA fragments contained sequences corresponding to the consensus sequences for a k-turn suggesting a biologically relevant interaction. Among the mRNAs identified are those encoding L7Ae, Nop5 and fibrillarin (other components of the snoRNP), a hypothetical DNA binding protein, and a hypothetical glycosyl transferase. Subsequent reporter gene assays and phylogenetic analysis showed that L7Ae negatively regulates the transcript encoding L7Ae in *S. acidocaldarius* and several other diverse archaea species. The presence of k-turn motifs preceding several genes and L7Ae interaction with these motifs suggests that L7Ae may regulate not only it's own synthesis, but also synthesis of its interaction partners in snoRNPs, Nop5 and Fibrillarin (84).

ENGINEERED R-PROTEIN RESPONSIVE REGULATORY RNA SYSTEMS

R-protein binding motifs have also been used to create synthetic regulatory systems. Repressing systems designed for eukaryotic cells were created by placing the L7Ae binding site near the translational start site (123), allowing L7Ae to prevent translation initiation. Systems with the L7Ae binding site within the coding region proved more effective than those where the binding site was placed in the 5'-UTR, both in an *in vitro* translation system and within HELA cells. Activating systems, where L7Ae binding removes a trans-acting RNA to prevent translation, also proved effective *in vitro*. In addition, L7Ae-mediated activation was achieved in Hela cells by adding an L7Ae binding site to a synthetic shRNA (short-hairpin RNA); thus L7Ae binding prevented shRNA-mediated mRNA degradation (124). These examples demonstrate how the L7Ae protein-binding site may be easily transferred to an alternative context and harnessed for gene expression in a modular manner.

Indeed, creation of synthetic regulatory systems responding to r-proteins within cells appears to be facile in comparison to the creation of many types of synthetic regulators where the transition from *in vitro* to *in vivo* can be challenging (125). Several synthetic regulatory systems responding to r-protein S15 have also been created (126). Unlike the L7Ae examples, these regulators were created through *in vitro* selection of RNA aptamers interacting with r-protein S15 from *Geobacillus kaustophilus* rather than transplantation of a known binding site. One striking observation from this work is that even without explicit selection for regulation, a high proportion of aptamers enable regulation when positioned correctly relative to the start codon. A second finding is that r-protein S15 can interact with a wide variety of different binding sites (127). This observation is echoed in previous work where *in vitro* selection to r-protein S8

yielded both aptamers similar to the natural RNA binding partners, as well as those showing substantial differences (128).

CONCLUDING REMARKS

Regulatory RNA structures displaying motifs found in the rRNA are commonly identified. While, in some cases similarity may be due to shared RNA tertiary structure motifs, in other cases structural similarity can imply a shared r-protein binding partner. Many r-proteins have a secondary role as negative regulators of their own synthesis, and while it was postulated that all such regulatory structures would resemble the rRNA, this has proved true only in some cases. This review illustrates a range of different regulatory mRNA structures that display similarity to the rRNA, but it is by no means exhaustive. While the mRNA structures controlling r-protein synthesis in *E. coli* remain the best characterized, r-protein responsive mRNA structures hail from nearly all species of bacteria and several archaea. From these examples it is apparent that r-protein responsive mRNA structures can be direct and obvious mimics of the rRNA, but they do not have to be. Many r-protein interacting mRNA structures display no similarity to their cognate rRNA sites (e.g. *E. coli* S4 regulator), while others share only partial similarity. Second, very similar binding sites can appear in diverse organisms, but may use alternative mechanisms regulate gene expression, or display different positioning relative to regulated genes. Third, due to the structural plasticity of RNA, a geometrically similar binding site may be displayed in several very different manners. Finally, from the diversity of natural regulatory mRNA structures, *in vitro* selection of aptamers, and the design of r-protein responsive regulatory mechanisms, it is clear that the sequence space that allows for r-protein binding and subsequent gene regulation may be quite large. This conclusion combined with the

lack of knowledge of r-protein regulation outside of *E. coli* suggests that many r-protein responsive mRNA structures, including those not directly associated with r-protein operons, remain undiscovered or unverified.

ACKNOWLEDGEMENTS

This material is based on work partially supported by NSF grants MCB:1411970 and 1715440 to M. Meyer. Special thanks to Arianne Babina and Betty Slinger for their hard work and insightful discussions of this topic.

REFERENCES

1. **Ban N**, **Nissen P**, **Hansen J**, **Moore PB**, **Steitz TA**. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science **289**:905–920.

2. **Peer A**, **Margalit H**. 2014. Evolutionary patterns of *Escherichia coli* small RNAs and their regulatory interactions. RNA **20**:994–1003.

3. **Hoeppner MP**, **Gardner PP**, **Poole AM**. 2012. Comparative Analysis of RNA Families Reveals Distinct Repertoires for Each Domain of Life. PLoS Comput Biol **8**:e1002752.

4. **Lindgreen S**, **Umu SU**, **Lai AS-W**, **Eldai H**, **Liu W**, **McGimpsey S**, **Wheeler NE**, **Biggs PJ**, **Thomson NR**, **Barquist L**, **Poole AM**, **Gardner PP**. 2014. Robust Identification of Noncoding RNA from Transcriptomes Requires Phylogenetically-Informed Sampling. PLoS Comput Biol **10**:e1003907.

5. **Kacharia FR**, **Millar JA**, **Raghavan R**. 2017. Emergence of New sRNAs in Enteric Bacteria is Associated with Low Expression and Rapid Evolution. Journal of Molecular Evolution **84**:204–213.

6. **Leontis NB**, **Westhof E**. 2002. The Annotation of RNA Motifs. Comp Funct Genom **3**:518–524.

7. **Klein D**, **Schmeing T**, **Moore P**, **Steitz T**. 2001. The kink-turn: a new RNA secondary structure motif. The EMBO Journal **20**:4214–4221.

8. **Lescoute A**. 2005. Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. Nucleic Acids Res **33**:2395–2409.

9. **Fox GE**, **Woese CR**. 1975. 5S RNA secondary structure. Nature **256**:505–507.

10. **Leontis NB**, **Westhof E**. 1998. A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. J Mol Biol **283**:571–583.

11. **Correll CC**, **Freeborn B**, **Moore PB**, **Steitz TA**. 1997. Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. Cell **91**:705–712.

12. **Woese CR**, **Winker S**, **Gutell RR**. 1990. Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". Proc Natl Acad Sci USA **87**:8467–8471.

13. **Gutell RR**, **Larsen N**, **Woese CR**. 1994. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. Microbiol Rev **58**:10–26.

14. **Wu JC**, **Gardner DP**, **Ozer S**, **Gutell RR**, **Ren P**. 2009. Correlation of RNA secondary structure statistics with thermodynamic stability and applications to folding. J Mol Biol **391**:769–783.

15. **Serganov A**, **Huang L**, **Patel DJ**. 2008. Structural insights into amino acid binding and

gene control by a lysine riboswitch. Nature **455**:1263–1267.

16. **Wang J**, **Nikonowicz EP**. 2011. Solution Structure of the K-Turn and Specifier Loop Domains from the *Bacillus subtilis* tyrS T-Box Leader RNA. J Mol Biol **408**:99–117.

17. **Strobel SA**. 2004. RNA kink turns to the left and to the right. RNA **10**:1852–1854.

18. **Keating KS**, **Toor N**, **Perlman PS**, **Pyle AM**. 2010. A structural analysis of the group II intron active site and implications for the spliceosome. RNA **16**:1–9.

19. **Garst AD**, **Edwards AL**, **Batey RT**. 2011. Riboswitches: Structures and Mechanisms. Cold Spring Harbor Perspectives in Biology **3**:a003533–a003533.

20. **Huang L**, **Lilley DMJ**. 2016. The Kink Turn, a Key Architectural Element in RNA Structure. J Mol Biol **428**:790–801.

21. **Chan CW**, **Chetnani B**, **Mondragón A**. 2013. Structure and function of the T-loop structural motif in noncoding RNAs. WIREs RNA **4**:507–522.

22. **Aseev LV**, **Boni IV**. 2011. Extraribosomal functions of bacterial ribosomal proteins. Mol Biol (Mosk) **45**:805–816.

23. **Warner JR**, **McIntosh KB**. 2009. How Common Are Extraribosomal Functions of Ribosomal Proteins? Molecular Cell **34**:3–11.

24. **Schmidt A**, **Kochanowski K**, **Vedelaar S**, **Ahrné E**, **Volkmer B**, **Callipo L**, **Knoops K**, **Bauer M**, **Aebersold R**, **Heinemann M**. 2015. The quantitative and condition-dependent *Escherichia coli* proteome. Nat Biotech **34**:104–110.

25. **Li G-W**, **Burkhardt D**, **Gross C**, **Weissman JS**. 2014. Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. Cell **157**:624–635.

26. **Fallon AM**, **Jinks CS**, **Yamamoto M**, **Nomura M**. 1979. Expression of ribosomal protein genes cloned in a hybrid plasmid in *Escherichia coli*: gene dosage effects on synthesis of ribosomal proteins and ribosomal protein messenger ribonucleic acid. J Bacteriol **138**:383–396.

27. **Fallon AM**, **Jinks CS**, **Strycharz GD**, **Nomura M**. 1979. Regulation of ribosomal protein synthesis in *Escherichia coli* by selective mRNA inactivation. Proc Natl Acad Sci USA **76**:3411–3415.

28. **Lindahl L**, **Zengel JM**. 1979. Operon-specific regulation of ribosomal protein synthesis in *Escherichia coli*. Proc Natl Acad Sci USA **76**:6542–6546.

29. **Dean D**, **Yates J**, **Nomura M**. 1981. *Escherichia coli* ribosomal protein S8 feedback regulates part of spc operon. Nature **289**:89–91.

30. **Yates J**, **Arfsten A**, **Nomura M**. 1980. In vitro expression of *Escherichia coli* ribosomal protein genes: autogenous inhibition of translation. Proc Natl Acad Sci USA **77**:1837–1841.

31. **Dean D**, **Nomura M**. 1980. Feedback regulation of ribosomal protein gene expression in *Escherichia coli*. Proc Natl Acad Sci USA **77**:3590–3594.

32. **Brot N**, **Caldwell P**, **Weissbach H**. 1980. Autogenous control of *Escherichia coli* ribosomal protein L10 synthesis in vitro. Proc Natl Acad Sci USA **77**:2592.

33. **Holowachuk E**, **Friesen J**, **Fiil N**. 1980. Bacteriophage lambda vehicle for the direct cloning of *Escherichia coli* promoter DNA sequences: feedback regulation of the *rplJL-rpoBC* operon. Proc Natl Acad Sci USA **77**:2124.

34. **Robakis N**, **Meza-Basso L**, **Brot N**, **Weissbach H**. 1981. Translational control of ribosomal protein L10 synthesis occurs prior to formation of first peptide bond. Proc Natl Acad Sci USA **78**:4261.

35. **Olins P**, **Nomura M**. 1981. Translational regulation by ribosomal protein S8 in *Escherichia coli*: structural homology between rRNA binding site and feedback target on mRNA. Nucleic Acids Res **9**:1757.

36. **Dean D**, **Yates JL**, **Nomura M**. 1981. Identification of ribosomal protein S7 as a repressor of translation within the str operon of *E. coli*. Cell **24**:413–419.

37. **Jinks-Robertson S**, **Nomura M**. 1982. Ribosomal protein S4 acts in trans as a translational repressor to regulate expression of the *alpha* operon in *Escherichia coli*. J Bacteriol **151**:193–202.

38. **Singer P**, **Nomura M**. 1985. Stability of ribosomal protein mRNA and translational feedback regulation in *Escherichia coli.* Mol Gen Genet **199**:543–546.

39. **Cole JR**, **Nomura M**. 1986. Changes in the half-life of ribosomal protein messenger RNA caused by translational repression. J Mol Biol **188**:383–392.

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

41. **Friedman DI**, **Schauer AT**, **Baumann MR**, **Baron LS**, **Adhya SL**. 1981. Evidence that ribosomal protein S10 participates in control of transcription termination. Proc Natl Acad Sci USA **78**:1115–1118.

42. **Zengel J**, **Lindahl L**. 1990. *Escherichia coli* ribosomal protein L4 stimulates transcription termination at a specific site in the leader of the S10 operon independent of L4-mediated inhibition of translation. J Mol Biol **213**:67–78.

43. **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 USA **77**:7084–7088.

44. **Olins P**, **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.

45. **Baughman G**, **Nomura M**. 1983. Localization of the target site for translational regulation of the L11 operon and direct evidence for translational coupling in *Escherichia coli*. Cell **34**:979–988.

46. **Lindahl L**, **Zengel JM**. 1986. Ribosomal genes in *Escherichia coli*. Annu Rev Genet **20**:297–326.

47. **Johnsen M**, **Christensen T**, **Dennis P**, **Fiil N**. 1982. Autogenous control: ribosomal protein L10-L12 complex binds to the leader sequence of its mRNA. The EMBO Journal **1**:999.

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

49. **Deckman I**, **Draper D**. 1985. Specific interaction between ribosomal protein S4 and the. *alpha* operon messenger RNA. Biochemistry **24**:7860–7865.

50. **Tang CK**, **Draper DE**. 1989. Unusual mRNA pseudoknot structure is recognized by a protein translational repressor. Cell **57**:531–536.

51. **Shen P**, **Zengel J**, **Lindahl L**. 1988. Secondary structure of the leader transcript from the *Escherichia coli* S10 ribosomal protein operon. Nucleic Acids Res **16**:8905.

52. **Merianos H**, **wang J**, **Moore P**. 2004. The structure of a ribosomal protein S8/spc operon mRNA complex. RNA **10**:954.

53. **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.

54. **Lesage P**, **Truong H**, **Graffe M**, **Dondon J**, **Springer M**. 1990. Translated translational operator in *Escherichia col*i auto-regulation in the *infC-rpmI-rplT* operon. J Mol Biol **213**:465–475.

55. **Portier C**, **Dondon L**, **Grunberg-Manago M**. 1990. Translational autocontrol of the *Escherichia coli* ribosomal protein S15. J Mol Biol **211**:407.

56. **Parsons G**, **Donly B**, **Mackie G**. 1988. Mutations in the leader sequence and initiation codon of the gene for ribosomal protein S20 (*rpsT*) affect both translational efficiency and autoregulation. J Bacteriol **170**:2485–2992.

57. **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-tsf* expression *in vivo*. RNA **14**:1882–1894.

58. **Aseev LV**, **Bylinkina NS**, **Boni IV**. 2015. Regulation of the *rplY* gene encoding 5S rRNA binding protein L25 in *Escherichia coli* and related bacteria. RNA **21**:851–861.

59. **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. Journal of Biological Chemistry **265**:17044.

60. **Aseev LV**, **Koledinskaya LS**, **Boni IV**. 2016. Regulation of Ribosomal Protein Operons *rplM-rpsI, rpmB-rpmG,* and *rplU-rpmA* at the Transcriptional and Translational Levels. J Bacteriol **198**:2494–2502.

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

62. **Fu Y**, **Deiorio-Haggar K**, **Soo MW**, **Meyer MM**. 2014. Bacterial RNA motif in the 5' UTR of *rpsF* interacts with an S6:S18 complex. RNA **20**:168–176.

63. **Babina AM**, **Soo MW**, **Fu Y**, **Meyer MM**. 2015. An S6:S18 complex inhibits translation of *E. coli rpsF*. RNA **21**:2039–2046.

64. **Fujita K**, **Baba T**, **Isono K**. 1998. Genomic Analysis of the Genes Encoding Ribosomal Proteins in Eight Eubacterial Species and *Saccharomyces cerevisiae.* Genome Inform Ser Workshop Genome Inform **9**:3–12.

65. **Coenye T**, **Vandamme P**. 2005. Organisation of the S10, *spc* and alpha ribosomal protein gene clusters in prokaryotic genomes. FEMS Microbiology Letters **242**:117–126.

66. **Allen T**, **Shen P**, **Samsel L**, **Liu R**, **Lindahl L**, **Zengel J**. 1999. Phylogenetic analysis of L4-mediated autogenous control of the S10 ribosomal protein operon. J Bacteriol **181**:6124.

67. **Allen T**, **Watkins T**, **Lindahl L**, **Zengel**. 2004. Regulation of ribosomal protein synthesis in *Vibrio cholerae*. J Bacteriol **186**:5933–5937.

68. **Iben J**, **Draper D**. 2008. Specific Interactions of the L10 $(L12)_4$ Ribosomal Protein Complex with mRNA, rRNA, and L11. Biochemistry **47**:2721–2731.

69. **Aseev LV**, **Levandovskaya AA**, **Skaptsova NV**, **Boni IV**. 2009. Conservation of regulatory elements controlling the expression of the *rpsB-tsf* operon in γ-proteobacteria. Mol Biol **43**:101–107.

70. **Guillier M**, **Allemand F**, **Raibaud S**, **Dardel F**, **Springer M**, **Chairuttini 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.

71. **Fu Y**, **Deiorio-Haggar 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.

72. **Matelska D**, **Kurkowska M**, **Purta E**, **Bujnicki JM**, **Dunin-Horkawicz S**. 2016. Loss of Conserved Noncoding RNAs in Genomes of Bacterial Endosymbionts. Genome Biology and Evolution **8**:426–438.

73. **Grundy F**, **Henkin TM**. 1991. The *rpsD* gene, encoding ribosomal protein S4, is autogenously regulated in *Bacillus subtilis*. J Bacteriol **173**:4595–4602.

74. **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.

75. **Deiorio-Haggar K**, **Anthony J**, **Meyer MM**. 2013. RNA structures regulating ribosomal protein biosynthesis in bacilli. RNA Biology **10**:1180–1184.

76. **Nawrocki E**, **Kolbe D**, **Eddy S**. 2009. Infernal 1.0: inference of RNA alignments. Bioinformatics **25**:1335–1337.

77. **Griffiths-Jones S**, **Bateman A**, **Marshall M**, **Khanna A**, **Eddy SR**. 2003. Rfam: an RNA family database. Nucleic Acids Res **31**:439–441.

78. **Weinberg Z**, **Wang JX**, **Bogue J**, **Yang J**, **Corbino K**, **Moy RH**, **Breaker RR**. 2010. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. Genome Biol **11**:1–17.

79. **Weinberg Z**, **Barrick JE**, **Yao Z**, **Roth A**, **Kim JN**, **Gore J**, **Wang JX**, **Lee ER**, **Block KF**, **Sudarsan N**, **Neph S**, **Tompa M**, **Ruzzo WL**, **Breaker RR**. 2007. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. Nucleic Acids Res **35**:4809–4819.

80. **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.

81. **Tseng H-H**, **Weinberg Z**, **Gore J**, **Breaker RR**, **Ruzzo WL**. 2009. Finding non-coding RNAs through genome-scale clustering. J Bioinform Comput Biol **7**:373–388.

82. **Weinberg Z**, **Lunse CE**, **Corbino KA**, **Ames TD**, **Nelson JW**, **Roth A**, **Perkins KR**, **Sherlock ME**, **Breaker RR**. 2017. Detection of 224 candidate structured RNAs by comparative analysis of specific subsets of intergenic regions. Nucleic Acids Res 1–13.

83. **Dar D**, **Shamir M**, **Mellin JR**, **Koutero M**, **Stern-Ginossar N**, **Cossart P**, **Sorek R**.

2016. Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. Science **352**:187–187.

84. **Daume M**, **Uhl M**, **Backofen R**, **Randau L**. 2017. RIP-Seq Suggests Translational Regulation by L7Ae in Archaea. mBio **8**:e00730–17–14.

85. **Mattheakis L**, **Vu L**, **Sor F**, **Nomura M**. 1989. Retroregulation of the synthesis of ribosomal proteins L14 and L24 by feedback repressor S8 in *Escherichia coli*. Proc Natl Acad Sci USA **86**:448–452.

86. **Gregory RJ**, **Cahill PB**, **Thurlow DL**, **Zimmermann RA**. 1988. Interaction of Escherichia coli ribosomal protein S8 with its binding sites in ribosomal RNA and messenger RNA. J Mol Biol **204**:295–307.

87. **Cerretti D**, **Mattheakis L**, **Kearney K**, **Vu L**, **Nomura M**. 1988. Translational regulation of the *spc* operon in *Escherichia coli*: Identification and structural analysis of the target site for S8 repressor protein. J Mol Biol **204**:309–329.

88. **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.

89. **Davies C**, **Ramakrishnan V**, **White SW**. 1996. Structural evidence for specific S8-RNA and S8-protein interactions within the 30S ribosomal subunit: ribosomal protein S8 from Bacillus stearothermophilus at 1.9 A resolution. Structure **4**:1093–1104.

90. **Vysotskaya V**, **Tischenko S**, **Garber M**, **Kern D**, **Mougel M**, **Ehresmann C**, **Ehresmann B**. 1994. The ribosomal protein S8 from *Thermus thermophilus* VK1. Eur J Biochem **223**:437–445.

91. **Kalurachchi K**, **Uma K**, **Zimmermann RA**, **Nikonowicz EP**. 1997. Structural features of the binding site for ribosomal protein S8 in *Escherichia coli* 16S rRNA defined using NMR spectroscopy. Proc Natl Acad Sci USA **94**:2139–2144.

92. **Nevskaya N**, **Tishchenko S**, **Nikulin A**, **Al-Karadaghi S**, **Liljas A**, **Ehresmann B**, **Ehresmann C**, **Garber M**, **Nikonov S**. 1998. Crystal structure of ribosomal protein S8 from *Thermus thermophilus* reveals a high degree of structural conservation of a specific RNA binding site. J Mol Biol **279**:233–244.

93. **Tishchenko S**, **Nikulin A**, **Fomenkova N**, **Nevskaya N**, **Nikonov O**, **Dumas P**, **Moine H**, **Ehresmann B**, **Ehresmann C**, **Piendl W**, **Lamzin V**, **Garber M**, **Nikonov S**. 2001. Detailed analysis of RNA-protein interactions within the ribosomal protein S8-rRNA complex from the archaeon *Methanococcus jannaschii*. J Mol Biol **311**:311–324.

94. **Friesen J**, **Tropak M**, **An G**. 1983. Mutations in the *rplJ* leader of *Escherichia coli* that abolish feedback regulation. Cell **32**:361–369.

95. **Climie S**, **Friesen J**. 1987. Feedback regulation of the *rplJL-rpoBC* ribosomal protein operon of *Escherichia coli* requires a region of mRNA secondary structure. J Mol Biol **198**:371–381.

96. **Climie S**, **Friesen J**. 1988. In vivo and in vitro structural analysis of the *rplJ* mRNA leader of *Escherichia coli*. Protection by bound L10-L7/L12. Journal of Biological Chemistry **263**:15166–15175.

97. **Yates J**, **Dean D**, **Strycharz W**, **Nomura M**. 1981. *E. coli* ribosomal protein L10 inhibits translation of L10 and L7/L12 mRNAs by acting at a single site. Nature **294**:190–192.

98. **Yakhnin H**, **Yakhnin AV**, **Babitzke P**. 2015. Ribosomal protein $L10(L12)_4$ autoregulates expression of the *Bacillus subtilis rplJL* operon by a transcription attenuation mechanism. Nucleic Acids Res **43**:7032–7043.

99. **Barrick JE**, **Breaker RR**. 2007. The distributions, mechanisms, and structures of metabolite-binding riboswitches. Genome Biol **8**:R239.

100. **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 USA **81**:5389.

101. **Hanner M**, **Mayer C**, **Kohrer C**, **Golderer G**, **Grobner P**, **Piendl W**. 1994. Autogenous translational regulation of the ribosomal MvaL1 operon in the archaebacterium *Methanococcus vannielii.* J Bacteriol **176**:409–418.

102. **Kraft A**, **Lutz C**, **Lingenhel A**, **Grobner P**, **Piendl W**. 1999. Control of ribosomal protein L1 synthesis in mesophilic and thermophilic archaea. Genetics **152**:1363.

103. **Köhrer C**, **Mayer C**, **Neumair O**, **Gröbner P**, **Piendl W**. 1998. Interaction of ribosomal L1 proteins from mesophilic and thermophilic Archaea and Bacteria with specific L1-binding sites on 23S rRNA and mRNA. European Journal of Biochemistry **256**:97–105.

104. **Shimmin L**, **Ramirez C**, **Matheson A**, **Dennis P**. 1989. Sequence alignment and evolutionary comparison of the L10 equivalent and L12 equivalent ribosomal proteins from archaebacteria, eubacteria, and eucaryotes. Journal of Molecular Evolution **29**:448–462.

105. **Haentjens-Sitri J**, **Allemand F**, **Springer M**, **Chiaruttini C**. 2008. A Competition Mechanism Regulates the Translation of the *Escherichia coli* Operon Encoding Ribosomal Proteins L35 and L20. J Mol Biol **375**:612–625.

106. **Nawrocki EP**, **Eddy SR**. 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics **29**:2933–2935.

107. **Bruscella P**, **Shahbabian K**, **Laalami S**, **Putzer H**. 2011. RNase Y is responsible for

uncoupling the expression of translation factor IF3 from that of the ribosomal proteins L35 and L20 in *Bacillus subtilis*. Molecular Microbiology **81**:1526–1541.

108. **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. Molecular Microbiology **56**:1141–1456.

109. **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.

110. **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 USA **90**:4394–4398.

111. **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. The EMBO Journal **22**:1898–1908.

112. **Slinger BL**, **Deiorio-Haggar K**, **Anthony JS**, **Gilligan MM**, **Meyer MM**. 2014. Discovery and validation of novel and distinct RNA regulators for ribosomal protein S15 in diverse bacterial phyla. BMC Genomics **15**:657.

113. **Slinger BL**, **Newman H**, **Lee Y**, **Pei S**, **Meyer MM**. 2015. Co-evolution of Bacterial Ribosomal Protein S15 with Diverse mRNA Regulatory Structures. PLoS Genet **11**:e1005720.

114. **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. Molecular Microbiology **52**:661–675.

115. **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.

116. **Serganov A**, **Ennifar E**, **Portier C**, **Ehresmann B**, **Ehresmann C**. 2002. Do mRNA and rRNA binding sites of *E. coli* ribosomal protein S15 share common structural determinants? J Mol Biol **320**:963–978.

117. **Ehresmann C**, **Ehresmann B**, **Ennifar E**, **Dumas P**, **Garber M**, **Mathy N**, **Nikulin A**, **Portier C**, **Patel D**, **Serganov A**. 2004. Molecular mimicry in translational regulation: the case of ribosomal protein S15. RNA Biology **1**:66–73.

118. **Cho I-M**, **Lai LB**, **Susanti D**, **Mukhopadhyay B**, **Gopalan V**. 2010. Ribosomal protein L7Ae is a subunit of archaeal RNase P. Proc Natl Acad Sci USA **107**:14573–14578.

119. **Kuhn JF**, **Tran EJ**, **Maxwell ES**. 2002. Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. Nucleic Acids Res **30**:931–941.

120. **Rozhdestvensky TS**, **Tang TH**, **Tchirkova IV**, **Brosius J**, **Bachellerie J-P**, **Hüttenhofer A**. 2003. Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. Nucleic Acids Res **31**:869–877.

121. **Nottrott S**, **Hartmuth K**, **Fabrizio P**, **Urlaub H**, **Vidovic I**, **Ficner R**, **Lührmann R**. 1999. Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. The EMBO Journal **18**:6119–6133.

122. **Baird NJ**, **Zhang J**, **Hamma T**, **FerrÉ-D'AmarÉ AR**. 2012. YbxF and YlxQ are bacterial homologs of L7Ae and bind K-turns but not K-loops. RNA **18**:759–770.

123. **Saito H**, **Kobayashi T**, **Hara T**, **Fujita Y**, **Hayashi K**, **Furushima R**, **Inoue T**. 2009. Synthetic translational regulation by an L7Ae–kink-turn RNP switch. Nature chemical biology **6**:71–78.

124. **Saito H**, **Fujita Y**, **Kashida S**, **Hayashi K**, **Inoue T**. 2011. synthetic human cell fate regulation by protein-driven RNA switches. Nature Communications **2**:160–8.

125. **Berens C**, **Suess B**. 2015. Riboswitch engineering — making the all-important second and third steps. Current Opinion in Biotechnology **31**:10–15.

126. **Slinger BL**, **Meyer MM**. 2016. RNA regulators responding to ribosomal protein S15 are frequent in sequence space. Nucleic Acids Res **44**:9331–9341.

127. **Pei S**, **Slinger BL**, **Meyer MM**. 2017. Recognizing RNA structural motifs in HT-SELEX data for ribosomal protein S15. BMC Bioinformatics **18**:298.

128. **Moine H**, **Cachia C**, **Westhof E**, **Ehresmann B**, **Ehresmann C**. 1997. The RNA binding site of S8 ribosomal protein of *Escherichia coli*: Selex and hydroxyl radical probing studies. RNA **3**:255–268.

129. **Cannone JJ**, **Subramanian S**, **Schnare MN**, **Collett JR**, **D'Souza LM**, **Du Y**, **Feng B**, **Lin N**, **Madabusi LV**, **Müller KM**, **Pande N**, **Shang Z**, **Yu N**, **Gutell RR**. 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.

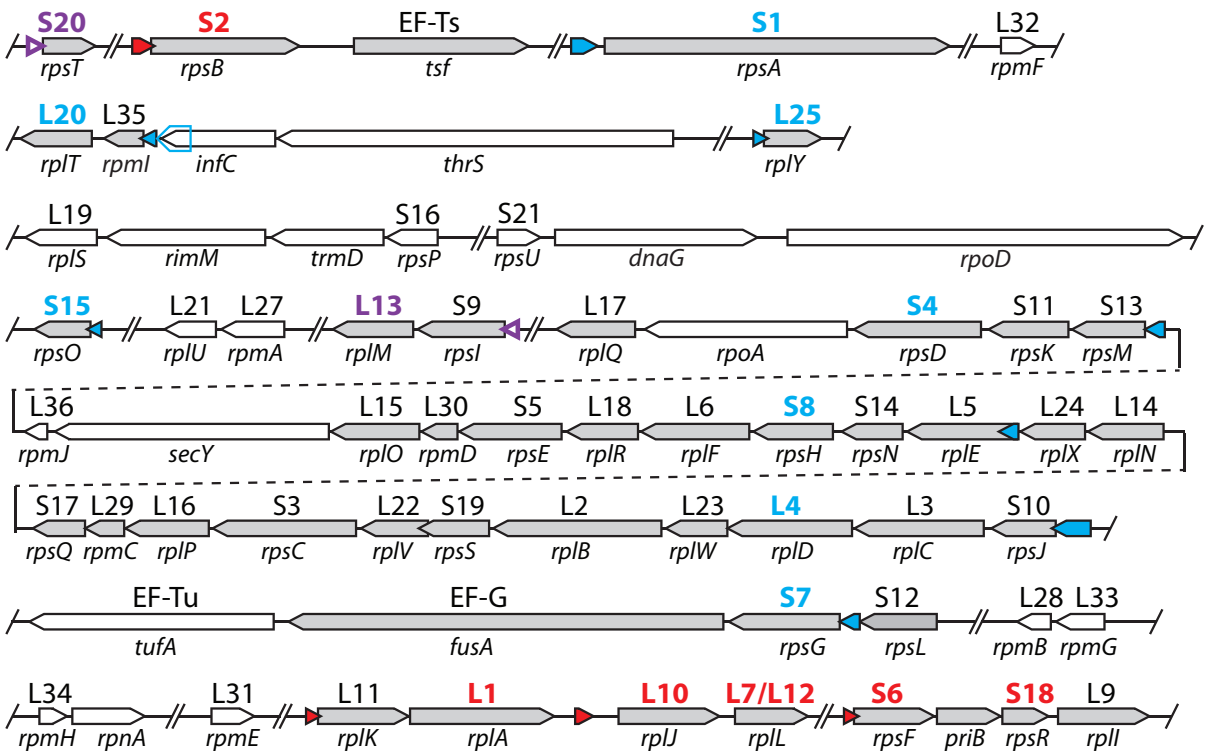
130. **Dunkle JA**, **Wang L**, **Feldman MB**, **Pulk A**, **Chen VB**, **Kapral GJ**, **Noeske J**, **Richardson JS**, **Blanchard SC**, **Cate JHD**. 2011. Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. Science **332**:981–984.

710

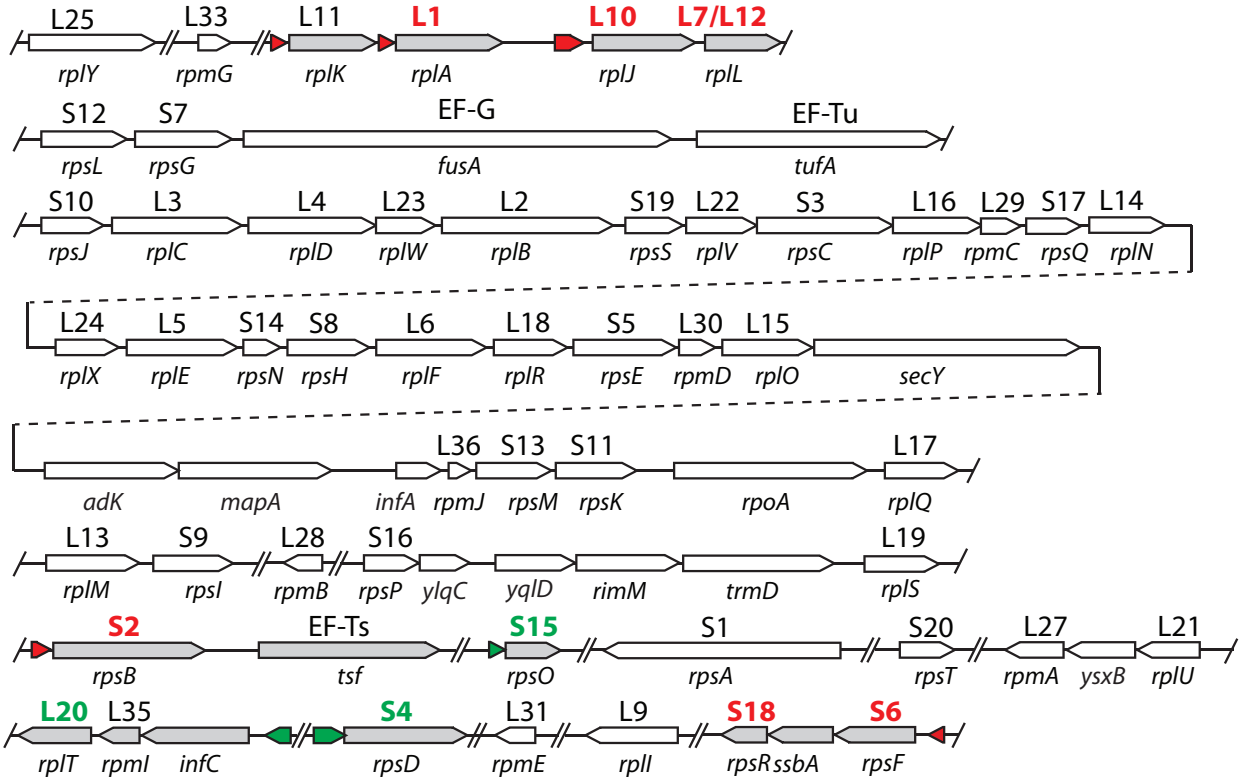
Binding Partner	Regulated Genes	Position	Species Distribution
L1[a,b,c,d]	*rplA, rplK*, rplP0**	varied	Archaea/Bacteria
L4[a,b,c,d]	*rpsJ, rplC, rplD, rplW, rplB, rpsS, rplV, rpsC, rplP, rpmC, rpsQ*	*rpsJ* 5'-UTR	Gammaproteobacteria
L10/ L10(L12)$_4$[a,b,c,d]	*rplJ, rplL*	*rplJ* 5'-UTR	Bacteria
L13[a]	*rplM, rpsI*	*rplM* 5'-UTR	*Escherichia coli*
L20[a,b,c,d]	*rpmI, rplT*	*infC* 5'-UTR	Firmicutes
L20[a,b,c,d]	*rpmI, rplT*	*infC-rpmI* intergenic	Gammaproteobacteria
L20[a,b,c,d]	*rpmI, rplT*	*infC* coding/ *infC-rpmI* intergenic	*Escherichia coli*
L25[a,c]	*rplY*	*rplY* 5'-UTR	Gammaproteobacteria
S1[a,b,c,d]	*rpsA*	*rpsA* 5'-UTR	Gammaproteobacteria
S2[a,c]	*rpsB*	*rpsB* 5'-UTR	Bacteria
S4[a,b,c,d]	*rpsM, rpsK, rpsD, rplQ*	*rpsM* 5'-UTR	Gammaproteobacteria
S4[a,c]	*rpsD*	*rpsD* 5'-UTR	Firmicutes
S6:S18[a,b,c]	*rpsF, rpsR, rplL**	*rpsF* 5'-UTR	Bacteria
S7[a,b,c]	*rpsL, rpsG, fusA*	*rpsL-rpsG* intergenic	Gammaproteobacteria
S8[a,b,c,d]	*rplN, rplX, rplE, rpsN, rpsH, rplF, rplR, rpsE, rpmD, rplO, secY, rpmJ*	*rplX-rplE* intergenic	Gammaproteobacteria
S15[a,b,c,d]	*rpsO*	*rpsO* 5'-UTR	Gammaproteobacteria
S15[b,c]	*rpsO*	*rpsO* 5'-UTR	Firmicutes
S15[b,c]	*rpsO*	*rpsO* 5'-UTR	*Thermus thermophilus*
S15[a,b,c]	*rpsO*	*rpsO* 5'-UTR	Alphaproteobacteria
S20[a,b]	*rpsT*	*rpsT* 5'-UTR	*Escherichia coli*

Table 1: Summary of ribosomal protein interacting mRNAs that allow regulation of r-protein genes. [a] regulation demonstrated using *in vitro* transcription/translation system or reporter gene assays; [b] direct RNA-protein interaction demonstrated *in vitro*; [c] structure of mRNA binding site characterized; [d] mechanism of action known, *may only be regulated in some species. Where a single species is listed for distribution either no structure is available, or no comparative genomic work has been conducted for the RNA and only the species of characterization is given.

A.


S20 S2 EF-Ts S1 L32
rpsT rpsB tsf rpsA rpmF
L20 L35 L25
rplT rpmI infC thrS rplY
L19 S16 S21
rplS rimM trmD rpsP rpsU dnaG rpoD
S15 L21 L27 L13 S9 L17 S4 S11 S13
rpsO rplU rpmA rplM rpsI rplQ rpoA rpsD rpsK rpsM
L36 L15 L30 S5 L18 L6 S8 S14 L5 L24 L14
rpmJ secY rplO rpmD rpsE rplR rplF rpsH rpsN rplE rplX rplN
S17 L29 L16 S3 L22 S19 L2 L23 L4 L3 S10
rpsQ rpmC rplP rpsC rplV rpsS rplB rplW rplD rplC rpsJ
EF-Tu EF-G S7 S12 L28 L33
tufA fusA rpsG rpsL rpmB rpmG
L34 L31 L11 L1 L10 L7/L12 S6 S18 L9
rpmH rpnA rpmE rplK rplA rplJ rplL rpsF priB rpsR rplI

B.


L25 L33 L11 L1 L10 L7/L12
rplY rpmG rplK rplA rplJ rplL
S12 S7 EF-G EF-Tu
rpsL rpsG fusA tufA
S10 L3 L4 L23 L2 S19 L22 S3 L16 L29 S17 L14
rpsJ rplC rplD rplW rplB rpsS rplV rpsC rplP rpmC rpsQ rplN
L24 L5 S14 S8 L6 L18 S5 L30 L15
rplX rplE rpsN rpsH rplF rplR rpsE rpmD rplO secY
L36 S13 S11 L17
adK mapA infA rpmJ rpsM rpsK rpoA rplQ
L13 S9 L28 S16 L19
rplM rpsI rpmB rpsP ylqC yqlD rimM trmD rplS
S2 EF-Ts S15 S1 S20 L27 L21
rpsB tsf rpsO rpsA rpsT rpmA ysxB rplU
L20 L35 S4 L31 L9 S18 S6
rplT rpmI infC rpsD rpmE rplI rpsR ssbA rpsF

720

Figure 1: Diagrams of r-protein operons from *E. coli* (A) and *B. subtilis* (B). Genes are shown in the order in which they appear in the genome and to scale. Gray genes are subject to r-protein autogenous regulation; white genes have no described autogenous regulation. Colored arrows represent r-protein RNA binding structures. Red arrows indicate structures that are widely distributed to many bacterial phyla, blue arrows indicate RNA structures that are confided to Gammaproteobacteria, green arrows indicate RNA structures confined to Firmicutes, and purple arrows indicate presumed r-protein binding sites where no explicit RNA secondary structure has been described. For each operon the effector protein is colored to match the RNA site with which it interacts.

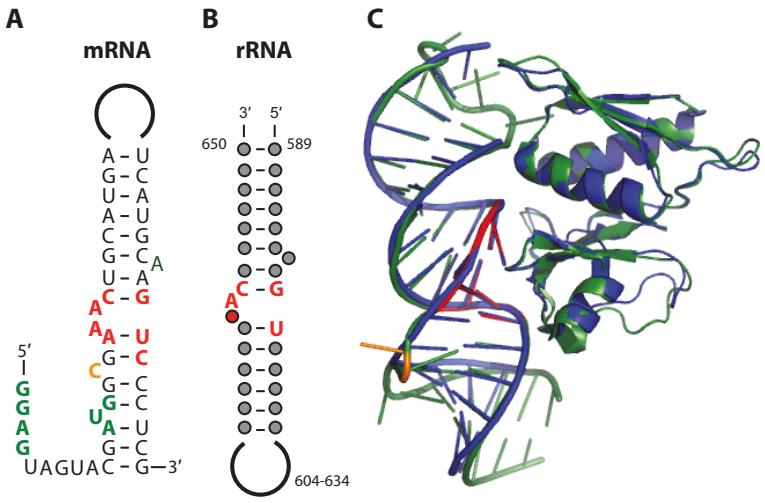


Figure 2 S8 mRNA binding site in *E. coli* mRNA (A) and rRNA (B) consensus structure. Green nucleotides indicate Shine-Dalgarno sequence and translational start, red nucleotides directly contact S8 in the three-dimensional structure (52). rRNA nucleotides conserved <90% are shown as filled circles, nucleotides conserved ≥90% are indicated by letters. Numbering corresponds to bacterial consensus sequence (129). (C) Aligning structural data for each site based on the S8 protein backbone shows that the two binding sites are superimposable. The structure of the S8 with its mRNA binding site are shown in green (1s03.cif,(52)), and the structure of S8 interacting with the rRNA is shown in blue (4v9d.cif, (130)). Bases of the mRNA directly contacting S8 are colored red, a bulged base in the mRNA that differentiates the rRNA and mRNA binding sites is colored orange.

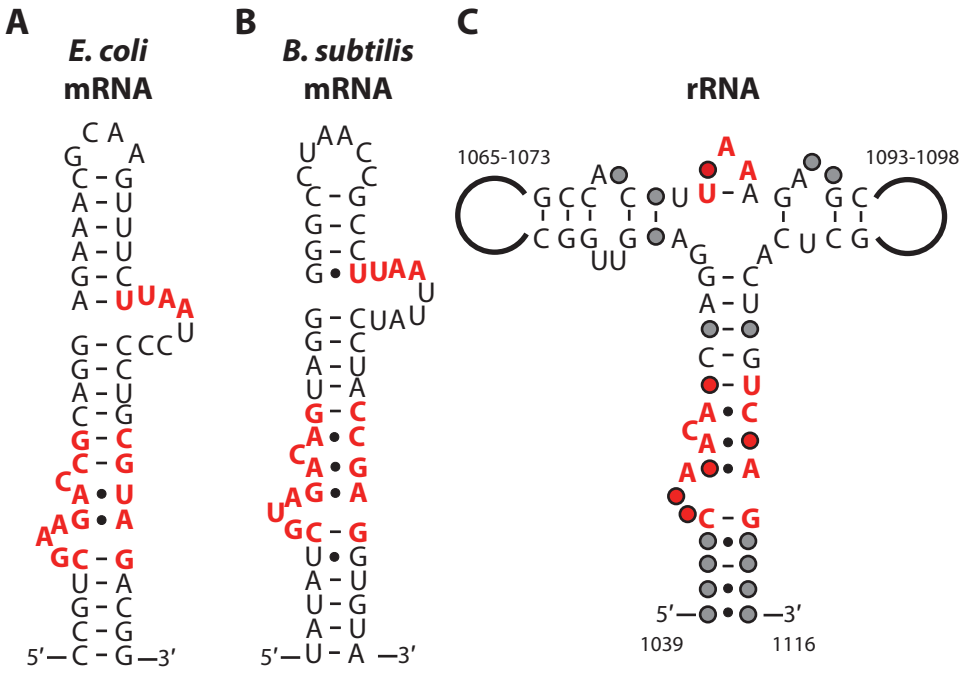


Figure 3: L10(L12)$_4$ mRNA binding sites from *E. coli* (A) and *B. subtilis* (B) and the rRNA consensus structure (C). Red nucleotides are implicated in binding, rRNA nucleotides conserved <90% are shown as filled circles, nucleotides conserved ≥90% are indicated by letters. Numbering corresponds to bacterial consensus sequence (129).

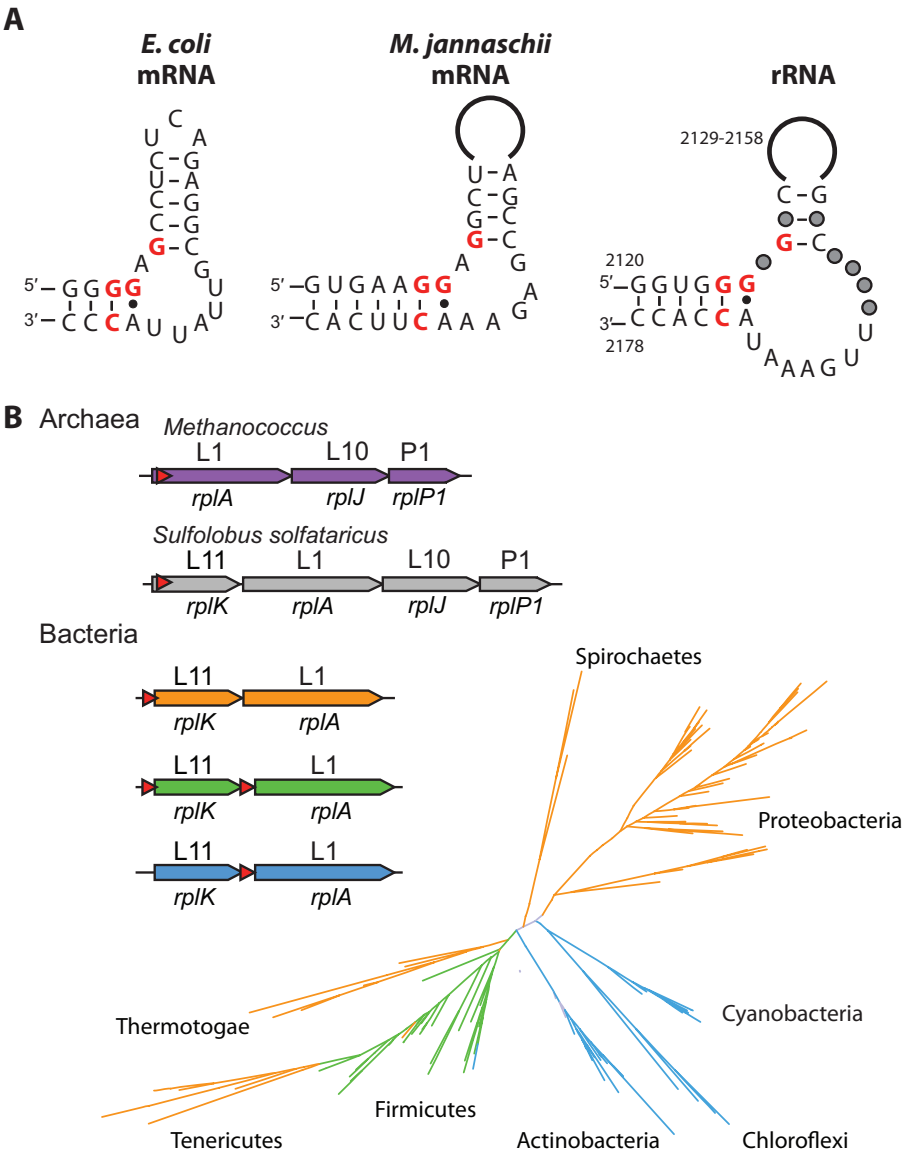


Figure 4: (A) L1-interacting mRNA structures from *E. coli* and *M. jannaschii*, and the L1 rRNA binding site (bacterial consensus). Red nucleotides directly contact L1 in the three-dimensional structure (53). rRNA nucleotides conserved <90% are shown as filled circles, nucleotides conserved ≥90% are indicated by letters. Numbering corresponds to bacterial consensus sequence (129). (B) Diagrams indicating the genomic position of L1 mRNA binding sites in two archaea clades (several *Methanococcus* species and *Sulfolobus solfataricus*) and in bacteria. Bacterial genomic position of L1 binding site are mapped to a 16S rRNA tree.

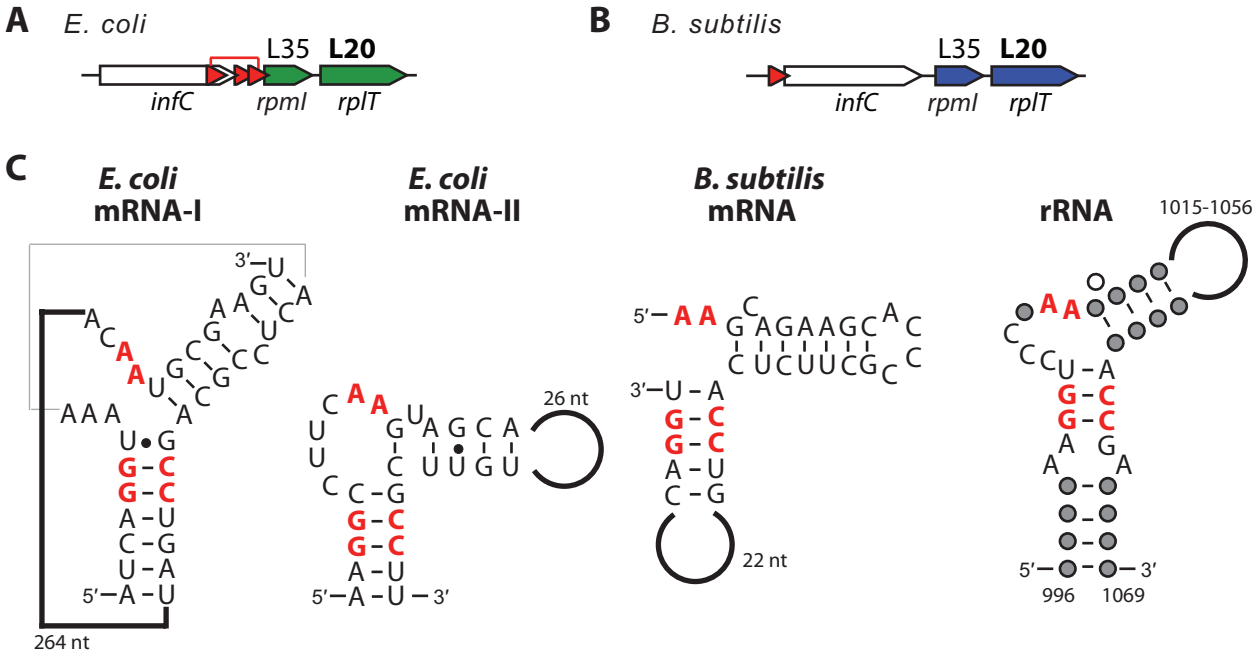
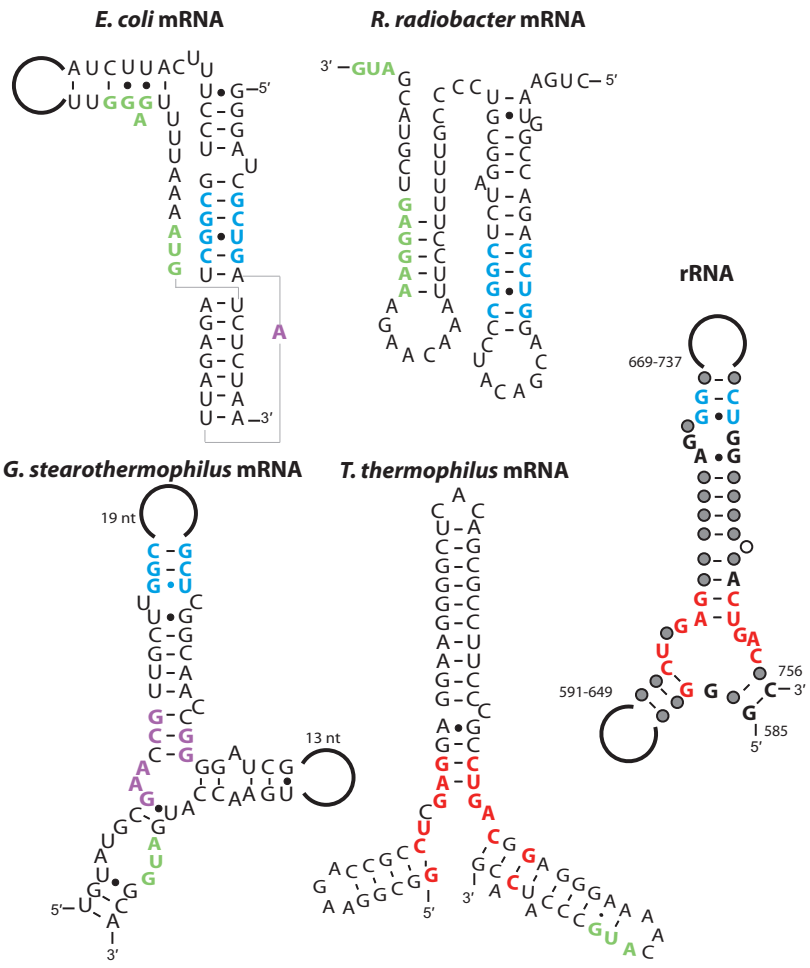


Figure 5: Diagram of *infC* operons showing genomic positions of L20-interacting mRNA structures (red arrows) in *E. coli* (A) and *B. subtilis* (B). Genes regulated by the RNA structure are colored. (C) L20-interacting mRNA structures from *E. coli* (mRNA-I and mRNA-II) and *B. subtilis* and the consensus rRNA L20 binding site. Red nucleotides are important for L20 interaction. rRNA nucleotides conserved <90% are shown as filled circles, nucleotides conserved >90% are indicated by letters. Numbering corresponds to bacterial consensus sequence (129).



765

766 **Figure 6**: S15-interacting mRNA structures in different bacterial phyla and the consensus S15
767 rRNA binding site. Red nucleotides correspond to the rRNA 3-stem junction and its direct
768 mimics. Blue nucleotides correspond to G•U/G-C helix imperfection in rRNA binding site and
769 its mimics in mRNA structures. Purple nucleotides are important for S15 recognition, but do not
770 directly correspond to any rRNA motif. Green nucleotides correspond to Shine-Dalgarno or
771 translational start sequences. rRNA nucleotides conserved <90% are shown as filled circles,
772 nucleotides conserved ≥90% are indicated by letters. Numbering corresponds to bacterial
773 consensus sequence (129)

774