RNA-binding proteins in bacteria

Erik Holmqvist and Jörg Vogel

Abstract | RNA-binding proteins (RBPs) are central to most if not all cellular processes, dictating the fate of virtually all RNA molecules in the cell. Starting with pioneering work on ribosomal proteins, studies of bacterial RBPs have paved the way for molecular studies of RNA–protein interactions. Work over the years has identified major RBPs that act on cellular transcripts at the various stages of bacterial gene expression and that enable their integration into post-transcriptional networks that also comprise small non-coding RNAs. Bacterial RBP research has now entered a new era in which RNA sequencing-based methods permit mapping of RBP activity in a truly global manner in vivo. Moreover, the soaring interest in understudied members of host-associated microbiota and environmental communities is likely to unveil new RBPs and to greatly expand our knowledge of RNA–protein interactions in bacteria.

Ribosomal proteins
(r-proteins). All proteins that, together with the ribosomal RNA, form the two subunits of the ribosome, the molecular machine that synthesizes polypeptides from an mRNA template.

Small non-coding RNAs
(sRNAs). Bacterial, in most cases non-coding, RNAs in the size range of 50–300 nucleotides that carry out regulatory functions, either by base pairing to complementary sequences in other RNAs or by functioning as binding partners for proteins.

RNA-binding proteins (RBPs) are found in all living organisms. The most conserved RBPs are a core set of ribosomal proteins (r-proteins) that were already present in the last universal common ancestor. Bacteria use RBPs both as structural components of larger complexes, such as the ribosome, and as regulators of many cellular processes, including the synthesis, modification, translation, processing and degradation of RNA (FIG. 1a).

RBPs interact with their ligands through RNA-binding domains (RBDs) that recognize short RNA sequences. Classic RBDs are widespread among bacterial RBPs. These include the S1 domain and the cold-shock domain (CSD) of the oligonucleotide/oligosaccharide binding (OB) superfamily, the Sm and Sm-like domains, the RNA recognition motif (RRM), the K homology (KH) domain, the double-stranded RNA-binding domain (dsRBD) and the PAZ and PIWI domains (FIG. 1b).

Our present knowledge of the functions of RBPs largely stems from work in eukaryotes, in which the specificity and versatility of RNA–protein interactions often arise from cooperative binding of several RBPs to one transcript or by combining multiple RBDs — different ones or repeats of the same — in a single protein. A modular architecture enables RBPs to recognize different RNA motifs in the same RNA ligand or to simultaneously interact with different RNA ligands. Bacterial RBPs tend to be more streamlined, often possessing only a few or even a single RBD per protein (FIG. 1b,c).

The fact that r-proteins make up one-third of the ~180 annotated RBPs in a typical bacterium but only one-tenth of the annotated 1,500 RBPs in the human genome suggests a lesser role of bacterial RBPs in functions other than protein synthesis. Furthermore, bacteria were traditionally thought to control their genes almost exclusively at the level of transcription, whereas extensive post-transcriptional processes, such as mRNA splicing involving dozens, if not hundreds, of different RBPs, would be seen as a hallmark of eukaryotic gene expression.

Things have changed. Research over the past 2 decades has revealed extensive post-transcriptional control in bacteria, which includes large regulatory networks comprising RBPs and small non-coding RNAs (sRNAs)14. We have also achieved a much better understanding of how RBPs act on nascent bacterial transcripts, modulate the translation rate of mRNAs or selectively alter the decay rates of specific sRNAs (FIG. 1c). Novel profiling methods using RNA sequencing (RNA-seq) are now providing evidence that bacterial RBPs can function as truly global factors that directly associate with and influence the fate of several hundreds of transcripts from across the genome (Supplementary Box 1). New methods that rely on mass spectrometry enable the unbiased identification of RBPs and RNA–protein complexes (Supplementary Box 2), and molecular studies of CRISPR–Cas systems have uncovered bacterial ribonucleoprotein particles (RNPs) of unexpected structural and functional complexity17. Similarly, a recent census of cytosolic complexes suggests that many bacterial RBPs are part of larger molecular assemblies (FIG. 1e). In light of these developments, it is timely to review the molecular mechanisms and functions of major RBPs in bacterial gene expression.

In this Review, we explore the roles of specific RBPs in transcription, protein synthesis and RNA decay, with an emphasis on their molecular properties. We focus on recent findings regarding the molecular interactions between proteins and RNA. Although ribonucleases and RNA-modification proteins interact with RNA, they are not commonly referred to as RBPs. Readers interested in these types of enzymes are referred to a number of recent specialized reviews18–20. Similarly, the molecular properties of Cas proteins have recently been discussed elsewhere16.
**RNA sequencing**

[RNA-seq]. A method for the parallel determination of sequence and abundance of RNA molecules in a biological sample.

**CRISPR–Cas systems**

Bacterial adaptive immune systems consisting of short RNAs that guide Cas proteins to target invading nucleic acids for destruction.

**Ribonuclease particles (RNPs):** Macromolecular complexes that consist of RNA and RNA-binding proteins.

**Ribonucleases**

Enzymes that catalyse the cleavage of RNA.

**Endoribonucleases** cleave RNA internally, whereas

**Exoribonucleases** degrade RNA from either the 5′ end or the 3′ end.

**RNA-modification proteins**

Proteins that introduce chemical modifications in RNA.

**Intrinsic termination**

A mechanism used by bacteria to stop transcription elongation and release the newly synthesized RNA. RNA polymerase is released from the nascent transcript as it encounters a palindromic GC-rich DNA sequence followed by a stretch of T residues. Also called Rho-independent termination.

**Rho-dependent termination**

A bacterial transcription termination mechanism whereby the release of the transcribing RNA polymerase from the DNA template is mediated by the Rho protein.

**Rho**

A homohexameric protein that promotes termination of transcription elongation by means of ATP hydrolysis. Rho is the target of the antibiotic bicyclomycin.

### Regulation of transcription

During RNA synthesis, numerous proteins — in addition to RNA polymerase (RNAP) itself — make contact with the nascent transcript. Most of the proteins modulate the termination phase of transcription, either by promoting the release of RNAP or by preventing premature termination (Fig. 2). For both induction and prevention of termination, RBPs may either act globally or regulate only specific genes.

**Termination transcription.** Bacteria use two general mechanisms for transcription termination:

- **Intrinsic termination and Rho-dependent termination** (Fig. 2A). Rho is among the best-studied RBPs in bacteria. It was first described ~50 years ago as a factor that promotes transcription termination in phage λ. Following early observations that Rho also acts on host genes, this protein was linked to termination in ~25% of all operons in *Escherichia coli*. In addition to terminating full-length mRNAs, Rho acts as an attenuator at many 5′ UTRs, a process that can be inhibited by base pairing sRNAs. Rho is an essential protein in many bacterial species, including *E. coli*, owing to diverse roles in the silencing of horizontally acquired genes, inhibition of R-loop formation, and suppression of replication-transcription conflicts. Rho forms large (~280 kDa) homohexameric rings (Fig. 2D), presenting two RNA-binding sites in each of the monomers. Termination initiation involves the oligosaccharide-binding (OB) fold of the primary RNA-binding site recognizing cytosine-rich Rho utilization (rut) sites in a cellular transcript. The full molecular mechanism of how Rho translocates along the RNA and finally halts transcription elongation remains to be determined. Possibly, contacts by the secondary RNA-binding sites induce the closure of the Rho ring and stimulate ATP-dependent 5′ → 3′ movement along the nascent RNA with Rho remaining tethered to the rut sequences. Alternatively, Rho may associate with RNAP already at the start of the transcription cycle. In any case, termination occurs as Rho encounters the RNA exit channel of the transcribing RNAP and releases the nascent transcript by unwinding the RNA–DNA duplex.

Intrinsic termination is controlled by Nus factors, which induce termination globally. Originally discovered in *E. coli* as host factors of phage λ (see Box 1), Nus factor-like proteins have since been found in many other bacteria. *E. coli* transcription termination/antitermination protein NusA is an essential protein with several different roles in termination: providing antitermination activity for phage transcription and enhancing RNA hairpin-dependent pausing and termination in chromosomal transcription. In *Bacillus subtilis*, 25% of all termination events are ascribed to NusA, especially at suboptimal intrinsic terminators. Not only does this protein affect intrinsic termination but it may also enhance Rho-dependent silencing of horizontally transferred DNA and globally antagonize Rho function (Fig. 2C).

A protein of many functions, NusA exemplifies a bacterial RBP with multiple functional domains, as beautifully illustrated through cryo-electron microscopy (cryo-EM) reconstructions of NusA bound to a transcriptionally paused RNAP or as part of the λN-based antitermination complex. In the former, the amino (N)-terminal domain, the carboxy (C)-terminal domain and a region between the two KH domains of NusA all contact the RNAP. Furthermore, the N-domain and the C-terminal domain provide a positively charged cavity around an RNA hairpin structure. This may stimulate both the formation and stability of RNA hairpins and explain how NusA promotes hairpin-dependent pausing and intrinsic termination. By contrast, in the context of the λN-based antitermination complex, NusA contacts nascent RNA via the two KH domains (Fig. 2D).

Contrasting the global effects of Rho, numerous RBPs induce termination at specific loci (Fig. 2B). A classic example is the trp RNA-binding attenuation protein (TRAP) in *B. subtilis*, a negative regulator of the tryptophan biosynthesis operon. The functional TRAP protein comprises a ring of 11 subunits and binds tryptophan in hydrophobic pockets between its monomers when this amino acid is plentiful. The activated TRAP then recognizes (G/U)AG repeats in the *trp* leader and terminates transcription upstream of the *trp* genes. The RBP PyrR regulates the pyrimidine operon (*pyr*) in a similar fashion, and it also provides an excellent example of how an RBP structurally remodels an RNA leader sequence to abort transcription.

The *pyr* operon contains three segments that, when transcribed, can each adopt alternative and mutually exclusive termination or antitermination structures. By binding to and stabilizing an RNA structure called the antiantiterminator, PyrR promotes the formation of the terminator, leading to premature termination and reduced expression of the *pyr* genes (Fig. 2B).

TRAP and PyrR both modulate intrinsic termination, but others act by modulating Rho-dependent termination. For instance, binding of carbon storage regulator A (CsrA) to leader sequences can expose rut sites that then mediate premature termination of transcription via the recruitment of Rho (Fig. 2B).

**Antitermination.** RNA structure strongly influences the rate of transcription elongation. RNA hairpins that invade the RNAP exit channel induce pausing, which ultimately can lead to transcription termination. Antiterminator RBPs either directly prevent the formation of terminator hairpins or favour the formation of alternative structures (Fig. 2C) to generally suppress aberrant premature termination events or to regulate gene expression as part of a physiological response.

Cold-shock proteins (CSPs) have been of much interest as a potentially major source of global antitermination activity. These are small (~7 kDa) proteins of a ubiquitous family that also includes eukaryotic Y-box proteins and which interact with single-stranded nucleic acids through their shared CSD. Purified CSPs decrease termination and pausing in vitro, and their overexpression in vivo increases the expression of genes preceded by several intrinsic terminators.

The structure of the founding member of this protein family, the cold-induced CspA of *E. coli*, revealed five antiparallel β-strands with an overall negatively
Fig. 1 | Overview of bacterial RBPs. a| Pie chart of all Escherichia coli proteins classified as ‘RNA-binding’ according to the Gene Ontology divided into the indicated functional categories. b| Shown are representative examples of well-characterized bacterial RNA-binding proteins (RBPs) to demonstrate their variability in length and number of RNA-binding domains (RBDs). RBDs are represented as rounded boxes that are colour-coded as indicated in the box. ‘Miscellaneous RBDs’ indicate RBDs that do not belong to a major RBD family. c| Size distribution of all E. coli proteins with Gene Ontology termed RNA-binding. d| Cellular processes commonly regulated by bacterial RBPs. Bacterial RBPs function on the levels of transcription, translation and RNA decay. Well-characterized RBPs that regulate each process are indicated. e| Distribution of RBPs as part of cellular complexes. Heat map showing relative abundance of Salmonella RBPs in fractionated bacterial lysates. For each protein (rows), the fraction in which the protein was most abundant was set to unity. For the RBPs indicated, the heat map colour was changed to red for better visualization. AlaS, alanine–tRNA ligase; BglG, cryptic β-glucoside bgl operon antiterminator; CSD, cold-shock domain; CSPs, cold-shock proteins; CsrA, carbon storage regulator A; CspA, cold-shock protein CspA; CspC, cold-shock protein CspC; HutP, hut operon positive regulatory protein; KH, K homology; IF-3, translation initiation factor 3; MpAgo, Marinitoga piezophila Argonaute; NusA, transcription termination/antitermination protein NusA; ProQ, RNA chaperone proQ; PyrR, RNA binding protein PyrR; RapZ, RNase adapter protein RapZ; RhlB, ATP-dependent RNA helicase RhlB; RNAP, RNA polymerase; RpsA, 30S ribosomal protein S1; RRM, RNA recognition motif; rRNA, ribosomal RNA; ThrRS, threonine–tRNA synthetase; TRAP, trp RNA-binding attenuation protein; YxiN, ATP-dependent RNA helicase DbpA. Data for part e from REF. 10.
Intrinsic termination

Rho-dependent termination

RBP-mediated induction of termination

RBP-mediated antitermination

Rho
NusA
TRAP or PyrR
CsrA
BglG or EutV
NusA
CspB

Rho
NusA
TRAP
CspB

Rho
NusA
TRAP
CspB
**Fig. 2** | RBP-mediated regulation of transcription termination. a | Bacteria use two different processes to terminate transcription: intrinsic termination and Rho-dependent termination. Intrinsic termination occurs after transcription of stem–loop structures followed by a poly(U) tract and is often stimulated by transcription termination/antitermination protein NusA. Rho-dependent termination requires the binding of Rho to Rho utilization (rut) sites in single-stranded RNA followed by Rho-mediated release of the RNA polymerase (RNAP) form the nascent RNA. b | Regulatory mechanisms by which RNA-binding proteins (RBPs) stimulate termination. Upper panel: mRNA leaders can adopt mutually exclusive structures that either promote transcription readthrough (red–blue stem–loop) or promote termination (red–red stem–loop). Binding of RBPs such as trp RNA-binding attenuation protein (TRAP) and RBP PyrR promotes the formation of terminator structures to prevent expression of downstream genes. Lower panel: to initiate transcription termination, Rho recognizes single-stranded rut sites. The pgaA mRNA leader contains several rut sites sequestered in a stem–loop. Binding of the Rho carbon storage regulator A (CsrA) renders the rut sites single-stranded, allowing Rho to bind and terminate transcription. c | Regulatory mechanisms of RBP-dependent antitermination. Upper panel: mRNA leaders that can adopt mutually exclusive structures that either promote (red–red stem–loop) or prevent (blue–red stem–loop) premature transcription termination. RBPs such as cryptic β-glucoside bgl operon antiterminator (BglG) and HutP, stabilize the formation of antitermination structures to induce the expression of downstream genes. Lower panel: competition for RNA binding between transcription termination/antitermination protein NusA and Rho prevents transcription termination when binding sites for each of these RBPs overlap in nascent RNA transcripts. d | Structures of four major RBPs that control transcription termination: Escherichia coli Rho bound to UUUUUU RNA (PDB ID: 5JJI), E. coli NusA bound to nut RNA in context of the XN-based transcription antitermination complex (PDB ID: 2ASB), Geobacillus stearothermophilus TRAP bound to a 53-nucleotide-long single-stranded RNA (PDB ID: 1C9S) and Bacillus subtilis cold-shock protein CspB bound to UUUUUU RNA (PDB ID: 3PF5). BglG, transcription antiterminator BglG.

charged surface.

Box 1 | RBPs expressed from plasmids and phages

Plasmids and bacteriophages often use RNA-binding proteins (RBPs) to regulate gene expression, and many early discoveries of RBP-mediated regulation were made in these systems. RBP-based positive regulation was identified in phages λ and Mu,

whereas examples of negative regulation came from bacteriophages MS2 and T4 (REFS [165,166]). Protein N from phage λ was the first antiterminator protein to be discovered,

which together with Nus factors (transcription termination/antitermination proteins) NusA, NusB, NusC, and NusG, and RNAP forms a transcription antitermination complex (TAC).

The structural basis for this form of antitermination was recently described.

As the elongating TAC encounters an intrinsic terminator, the S1 domain of NusA prevents terminator formation by sequestering the upstream arm of the terminator hairpin. The RNA-binding coat protein of mS2 has a dual role in genome encapsidation and in translation repression. The high affinity of this RBP for its RNA ligand has been used for RNA–protein capture techniques in bacteria.

Plasmids use RBPs to regulate processes such as DNA replication and conjugation. On F plasmids, the antisense RNA FinP controls the synthesis of the positive regulator of conjugation protein. This regulation requires the RBP fertility inhibition protein (FinO), which both stabilizes FinP and aids its annealing to the tral RNA. FinO-like proteins have recently garnered much attention because of the unexpected functions of family members from bacterial chromosomes (see the main text).

Other plasmid-expressed RBPs may have chromosomal homologues with a broader physiological function than currently thought. The Rop (also known as Rom) proteins, which facilitate antisense regulation on plasmid ColE1, seem good candidates to start with.

charged amino acid residues of the conserved RNA-binding motifs RNP-1 and RNP-2 mediate RNA binding, as shown for CspB, a CspA homologue in B. subtilis.[26] Interestingly, the structural similarity and functional redundancy between the CSPs and S1 domains hint at a common ancestry from an ancient RBP.[26]

 Whereas E. coli CspA globally unfolds mRNA structure to promote translation during acclimation after cold shock,[26] CspG and CspE are now established as truly globally acting RBPs at regular growth temperature.[26,27] Their combined activities affect the expression of 20% of all genes in Salmonella enterica (henceforth Salmonella) and are essential for the virulence of this pathogen.[26] In Staphylococcus aureus, CspA binds more than 500 different transcripts and has global effects on gene expression.[26] However, it is unclear how many of these effects involve antitermination, as CSPs may also modulate translation initiation and RNA stability.[26,28–30,32] (see below).

Many mechanistic aspects of the CSPs are unclear, including the molecular mechanism of antitermination itself, except for their melting activity on RNA secondary structures.[26] Strikingly, the affinity for RNA targets is typically in the micromolar range,[26,33] whereas many other RBPs described here have binding affinities in the nanomolar range. In vitro selection has identified putative consensus sequences for several E. coli CSPs,[26] but how RNA targets are recognized in vivo remains unclear. It will be important to elucidate how these very similar RBPs, of which some enteric bacteria have no fewer than 11, select targets in a seemingly non-redundant manner in vivo.[26]

Whereas CSPs represent global factors, other RBPs inhibit termination only at specific loci, often as part of a specific physiological response.[26] The hexameric antiterminator hut operon positive regulatory protein (HutP) acts on the histidine utilization genes in B. subtilis. Its RNA-binding activity is activated by a conformational change that is triggered by sensing the intracellular concentration of L-histidine.[26] By binding to six NAG triplets (with N representing any nucleotide) between its own gene and the downstream hut operon, HutP directly prevents the formation of the terminator hairpin to promote transcription elongation.[26] By contrast, RBPs of the BglG (cryptic β-glucoside bgl operon antiterminator)/SacY (levansucrase and sucrase synthase operon antiterminator) family bind inverted repeat sequences in mRNA leaders to induce RNA structures that are mutually exclusive with terminator formation[26] (Fig. 2c). Their target genes encode carbohydrate-metabolizing enzymes, and their RNA-binding activity is regulated through sugar availability-dependent phosphorylation.[26]

The recently discovered Nus-like protein LoaP promotes readthrough of intrinsic terminators in antibiotic gene clusters in Firmicutes.[26] Furthermore, two
studies highlighted how bacteria integrate RBPs with other regulatory molecules to control antitermination. In *Enterococcus faecalis* and *Listeria monocytogenes*, the phosphorylated form of the RBP EutV binds to hairpin structures in the *eut* RNAs, thereby causing antitermination and thus expression of the ethanolamine utilization (*eut*) locus. Active EutV can be out-titrated by an sRNA that in turn is controlled by a riboswitch responding to adenosylcobalamin, a cofactor for ethanolamine catabolism.

---

**Fig. 3 | RBP-based regulation of translation.**

**a** | Major mechanisms by which bacterial RNA-binding proteins (RBPs) mediate translation repression. Competition with the 30S ribosomal subunit for binding to the ribosome binding site (RBS) as shown for carbon storage regulator A (CsrA), Hfq–small non-coding RNA (sRNA) complexes and protein S1 and entrapment of the 30S ribosomal subunit mediated by S1.

**b** | RBP-dependent mechanisms of translation activation. Inhibitory RNA structures occluding 30S binding to the RBS can be relieved by binding of S1 or Hfq–sRNA complexes.

**c** | Molecular structures and RNA recognition for Csr and regulator of secondary metabolism (Rsm) (NMR structure of *Pseudomonas fluorescens* RsmE in complex with the SD sequence of *hcnA* mRNA, PDB ID: 2JPP), Hfq (crystal structure of *Salmonella enterica* subsp. enterica serovar Typhimurium Hfq in complex with RydC sRNA, PDB ID: 4V2S) and RNA chaperone ProQ (docking of NMR-derived structures of *E. coli* ProQ amino-terminal and carboxy-terminal domains, as well as a model of sRNA SraB, into a SAXS-derived envelope). Structure of ProQ in part c adapted with permission from REF. The RNA Society.

---

studies highlighted how bacteria integrate RBPs with other regulatory molecules to control antitermination. In *Enterococcus faecalis* and *Listeria monocytogenes*, the phosphorylated form of the RBP EutV binds to hairpin structures in the *eut* RNAs, thereby causing antitermination and thus expression of the ethanolamine utilization (*eut*) locus. Active EutV can be out-titrated by an sRNA that in turn is controlled by a riboswitch responding to adenosylcobalamin, a cofactor for ethanolamine catabolism.
Ribosome binding site (RBS). An mRNA sequence that recruits the 30S ribosomal subunit to initiate translation. The Shine–Dalgarno sequence of the RBS is complementary to the 16S ribosomal RNA (rRNA) and enables 30S–mRNA interaction.

Although the above-mentioned examples illustrate how intrinsic termination can be prevented by RBP binding, RBPs can antagonize Rho-dependent termination as well. A recent report showed that overlapping binding sites of NusA and Rho can lead to NusA-dependent prevention of Rho loading genome-wide24 (Fig. 2c).

In conclusion, RBPs engage in regulating transcription termination by promoting or inhibiting the formation of intrinsic terminators or by affecting the availability of nascent transcripts for termination factor Rho.

**Regulation of translation**

Mechanistically, most RBPs that affect translation function at the level of translation initiation (Fig. 5a,b); that is, the association of the 30S ribosomal subunit with the mRNA at the ribosome binding site (RBS). RBPs may directly compete with the 30S subunit for binding to an mRNA (for example, Csr and regulator of secondary metabolism (Rsm) proteins) or induce a structural change in the mRNA that alters the accessibility of the RBS. Another major mechanism whereby bacteria control mRNA translation involves RBP-mediated recruitment of sRNAs, the primary example of which is the regulatory network formed by the Hfq protein. Moreover, CSPs and fertility inhibition protein (FinO)-like RBPs also regulate mRNAs globally.

There are many translation-related RBPs that control their own mRNA, such as the r-protein S1 that both autoregulates its own synthesis and affects global translation (see below). Finally, ribosome biogenesis and function rely upon several additional well-conserved RBPs (Box 2).

**Csr/Rsm proteins.** Csr/Rsm proteins are highly conserved RBPs in the ∼7 kDa range; predicted homologues of E. coli CsrA — the founding member of this class — are found in almost all bacterial phyla25. Their physiological importance reveals itself by the usually strong pleiotropic phenotypes, if not lethality, upon genetic inactivation. For example, CsrA was discovered through its major impact on glycogen production and carbon storage in E. coli26; the former results from CsrA directly inhibiting translation of the first gene in a glycolysis biosynthesis operon27; a dimer of CsrA recognizes two single-stranded GGA triplets in the glgC 5′ UTR to obstruct binding of the 30S subunit45 (Fig. 5a).

Csr/Rsm proteins operate truly globally, targeting several hundreds of mRNAs in some Gram-negative model species25,67. In addition to their canonical mechanism whereby Csr/Rsm binds in 5′ UTRs to inhibit translation68,69, regulatory mechanisms affecting transcription or RNA decay have been described70.

Crosslinking immunoprecipitation coupled with high-throughput sequencing (CLIP-seq) studies capturing hundreds of CsrA–RNA interactions inside live E. coli and Salmonella71,72 have helped to refine the recognition motif as AUGGA, preferably located in the apical loops of RNA hairpins (Fig. 5c). This motif is not only found in mRNAs but also abounds in a particular set of sRNAs that act as decoys of Csr/Rsm proteins45,71. For example, the 369-nucleotide CarB sRNA contains no fewer than 22 GGA sequences, of which 18 are high-affinity CsrA binding sites72. The structural analysis of one such 'sponge', the RsmZ sRNA of Pseudomonas fluorescens, revealed how the RBP RsmE binds RNA in an ordered, sequential and cooperative fashion72. Specifically, sequential binding of RsmE to RsmZ induces allosteric changes in the RNA that strongly enhance the affinity for subsequent RsmE binding events. The resulting highly stable RNP of five RsmE dimers not only sequesters the protein but also protects the RsmZ sRNA from nucleolytic degradation72. It is easy to envisage the same process in the 5′ UTRs of target mRNAs leading to stable complexes that fully inhibit translation initiation.

Although studies in Gram-negative bacteria revealed intricate regulatory circuits composed of Csr/Rsm proteins and antagonistic sRNAs72, these proteins can also be regulated by protein–protein interactions. For example, in enteropathogenic E. coli, the protein CesT not only functions as a cytosolic chaperone of secreted virulence effectors but also moonlights as a direct inhibitor of CsrA to lift the CsrA-dependent repression of effector mRNAs73. Likewise, the interaction of CsrA with other proteins and the hagA mRNA precisely times the expression and secretion of the flagellin Hag in B. subtilis74–77.

The recognition mode of Csr/Rsm proteins seems highly conserved because homologues from distant species can often complement each other's function in vivo74. However, some bacteria express multiple Csr/Rsm homologues with partly non-redundant activities75. One open question is whether there may be auxiliary proteins that modulate the affinity of these RBPs for certain mRNAs, either directly or by affecting the subcellular localization of a target72. The growing atlases of in vivo binding sites45,69 should assist in determining whether the Csr/Rsm proteins act on certain targets in a combinatorial fashion with any of the other major RBPs described below.

---

**Box 2 | Non-ribosomal RBPs that function in translation**

Proper ribosome biogenesis and function rely on many well-conserved RNA-binding proteins (RBPs). Of several ribosome maturation factors, the GTPase RsgA uses its oligosaccharide binding (OB)-fold to interact with the 16S ribosomal RNA (rRNA) to suppress the formation of kinetically trapped 30S intermediates189. Turning to ribosome function, the RBP Ffh together with 4.5S RNA constitutes the bacterial signal recognition particle (SRP), the RNP that delivers translating ribosomes to the membrane. Structural analyses of SRP have revealed an unexpected arrangement in which a methionine-rich domain of Ffh binds to one end of 4.5S RNA, whereas another domain contacts the opposite end189. Among non-ribosomal bacterial RBPs, the 18 kDa small protein B (SmpB) of the trans-translation ribosome rescue system has been characterized best. SmpB, together with transfer-messenger RNA (tmRNA), promotes the release and recycling of ribosomes that have reached the very 3′ end of an mRNA without undergoing translation termination and tags the released truncated polypeptide for degradation190. SmpB specifically interacts with both tmRNA and the mRNA channel of the ribosome, the latter explaining how this rescue system specifically distinguishes stalled from actively translating ribosomes191. Considering their structural robustness, protein synthesis-related RBPs will continue to provide intricate details of RNA–protein interactions in particular, as cryogenic electron microscopy (cryo-EM) has begun to elucidate higher-order complexes of bacterial gene expression191.
Hfq. If bacteria needed a ‘poster child’ for RBPs that facilitate global RNA networks of eukaryotic-like complexity, the choice would be Hfq. The history of this protein, from its discovery as a host factor that unwinds phage RNA for efficient replication to its increasing implication in endogenous gene control, has been reviewed\(^\text{112}\). To date, the primary role of Hfq is seen in helping sRNAs find and regulate trans-encoded mRNAs through short, imperfect base pairing interactions\(^\text{81}\). In \textit{E. coli} and \textit{Salmonella}, up to 100 sRNAs may recognize more than 25% of all mRNAs via Hfq, comprising thousands of potential RNA interactions\(^\text{61,82-84}\). In the current standard model, Hfq simultaneously binds sRNAs and mRNAs, and, if these possess sufficient complementarity, an sRNA–mRNA complex forms (FIG. 3a). The resulting RNA duplex usually sequesters the RBS, silencing translation of the mRNA; however, other modes of inhibition and various mechanisms of target activation are known\(^\text{114,115}\) (FIG. 3b).

How does Hfq work as a global RNA ‘matchmaker’? Hfq possesses a conserved Sm domain\(^\text{45,48} –\) an N-terminal core that harbours two sequence motifs named Sm1 and Sm2 — and a variable C terminus. To be functional, six Hfq protomers assemble into the typical ring-like architecture of the Hfq–Sm–LSm family found in archaea, bacteria, and eukaryotes\(^\text{49}\). The homohexameric Hfq ring binds single-stranded RNA at four different sites: the proximal and the distal faces, the rim and the C-terminal tail. Although a high-resolution structure of full-length Hfq in complex with two paired RNAs is still lacking, the results of different in vitro and in vivo analyses have led to a model of how regulation is brought about (FIG. 3c).

The proximal face of Hfq preferentially interacts with single-stranded U-rich sequences at sRNA 3′ ends\(^\text{90,93}\); additional rim interactions help to position the sRNA such that its seed region is poised for contacting mRNA targets\(^\text{84}\). By contrast, the distal face interacts with single-stranded A-rich sequences present in mRNA 5′ UTRs\(^\text{90-93}\), bringing potential mRNA target sites close to the sRNA-binding proximal face\(^\text{48}\). When Hfq is loaded with both sRNA and cognate mRNA, the rim contacts their UA-rich sequences to promote RNA pairing\(^\text{92}\). Although this is the canonical view of how Hfq contacts sRNAs, the enormous sequence diversity of RNA partners predicts case-to-case variations. Indeed, mutational studies suggest two major classes of Hfq-associated sRNAs that differ by their dependencies for face or rim contacts\(^\text{86}\).

The function of the fourth major site on Hfq, the poorly conserved and intrinsically disordered C terminus, has been a matter of controversy\(^\text{97,98}\). However, recent data suggest that it helps to displace bound transcripts, which may not only help to rid Hfq of nonspecific RNA binders\(^\text{85}\) but also promote a rapid cycling of cellular RNA whose total concentration exceeds that of Hfq\(^\text{100,101}\). One can imagine how the six flexible C termini protruding from the ring brush RNA off the core. In addition, their acidic tip may transiently bind the basic Sm core residues necessary for RNA annealing, thereby autoregulating RNA binding to the Sm ring\(^\text{102}\).

Following the pioneering structure analysis of \textit{S. aureus} Hfq\(^\text{92}\), much of the above model is based on work on the \textit{E. coli} and \textit{Salmonella} proteins, the hexamers of which are ~70 kDa. However, the Hfq family is diverse and contains proteins with predicted hexamers of up to 140 kDa\(^\text{103}\). There are more reasons to suspect that we have yet to understand the true functional diversity of these proteins. On the one hand, heterologous expression of diverse Hfq proteins in \textit{Salmonella} showed that members of this family have an intrinsic propensity to bind sRNAs\(^\text{84}\). On the other hand, the collective results of Hfq studies in Gram-positive bacteria show little evidence for a prominent sRNA–mRNA matchmaking function that is a hallmark of Hfq in Gram-negative bacteria\(^\text{103}\).

Even in Gram-negative species, Hfq can modulate mRNA translation by sRNA-independent mechanisms. For example, during the mismatch repair pathway in \textit{E. coli}, the distal face of Hfq binds to the leader region of \textit{mutS} mRNA to inhibit translation\(^\text{104}\). Moreover, studies of how Hfq of \textit{Pseudomonas aeruginosa} directly represses translation of mRNAs under conditions of catabolite repression have suggested a direct interaction with the catabolite repression control (Crc) protein\(^\text{107}\). Crc was initially reported to inhibit translation by binding A-rich sequences in mRNAs required for growth on non-preferred carbon sources; this activity would be counteracted by the CrcY and CrcZ sRNAs\(^\text{108}\). However, more recent work suggests that the translation inhibitor is Hfq, whereas Crc stimulates the mRNA-binding activity of Hfq\(^\text{107,109}\). Crc and Hfq form a complex when RNA is bound to the distal side of Hfq, which results in a prolonged lifetime of Hfq–RNA interactions\(^\text{107}\). How Crc turns Hfq into a better mRNA repressor is not yet understood, but the finding itself raises the possibility that other proteins exist that may guide Hfq to mRNA targets in an sRNA-independent fashion.

\textbf{ProQ and other RBPs of the FinO family.} Whereas Hfq and CsrA have well-established roles in translational control, the functions of RNA chaperone ProQ and other FinO domain-containing RBPs are just emerging. The founding member of this class, FinO, has been studied for its role as a RNA chaperone in antisense regulation of F plasmid conjugation in \textit{E. coli}\(^\text{110}\).

The ~25 kDa ProQ protein is a chromosomal homologue of FinO. However, whereas FinO has only two known RNA targets, ProQ stably associates with several hundreds of cellular transcripts in \textit{Salmonella} and \textit{E. coli}\(^\text{110,111}\). These targets of ProQ include >70 sRNAs, including the sRNA that forms base pairs with the RBS of the mRNA of histone-like protein HU\(^\text{112}\). Not only does ProQ stabilize this sRNA, it also strengthens RNA duplex formation\(^\text{112}\), a general property that had already been seen with artificial substrates\(^\text{113}\). RocC, a FinO-like protein expressed from the chromosome of \textit{Legionella pneumophila}, also mediates translational repression, helping the RocR sRNA to bind multiple mRNAs of the DNA uptake system\(^\text{114}\).

Although these examples seem to reiterate the functions of Hfq, what happens after target recognition may differ: whereas Hfq, being at limiting concentration,
quickly releases a matched sRNA–mRNA couple, ProQ seems to hold on to sRNA–mRNA duplexes for a longer time as if assuring that 30S is always excluded\textsuperscript{112}. Cellular abundance would favour this scenario: there are 5–10 times more ProQ monomers (the active form) than Hfq hexamers\textsuperscript{110}, albeit the number of RNA targets is similar\textsuperscript{20,111}.

Potential FinO homologues are found on the chromosomes and on plasmids of alphaproteobacteria, betaproteobacteria, gammaproteobacteria and acidithiobacilli\textsuperscript{110,114}, promising a rich diversity of physiological functions and molecular mechanisms. Moreover, results from different bacterial species\textsuperscript{20,111} have consistently shown that the cellular RNA target suites of CsrA, Hfq and ProQ are distinct. Intriguingly, whereas both Hfq and CsrA target defined sequence motifs in single-stranded RNA, RBPs harbouring a FinO domain may preferentially recognize targets by RNA structure. For example, CLIP-Seq analysis in *E. coli* and *Salmonella* predicts ProQ to typically recognize a stable RNA hairpin, often a transcription terminator\textsuperscript{111}. Similarly, FinO and RocC bind to the terminator hairpins of the FinP and RocK sRNAs, respectively\textsuperscript{110,114}. In other words, ProQ seems to govern its global network of sRNAs and mRNAs by reading RNA structure rather than sequence\textsuperscript{111}, while at the same time, it successfully discriminates against the abundant and hairpin-rich ribosomal RNAs (rRNAs) and tRNAs.

Structural analysis of FinO-like RBPs is in its infancy, but it has great potential for elucidating their structural recognition code. Such analyses may also help to explain why some members, such as FinO and RocC, have only a few cellular RNA targets, whereas ProQ binds hundreds, although all these RBPs carry the same FinO domain. Is their selectivity determined by their distinct N-terminal and/or C-terminal extensions\textsuperscript{110–117}? A structural analysis of *E. coli* ProQ indicates that in addition to the FinO domain, other regions contribute to RNA binding and thus may modulate target selectivity\textsuperscript{118} (Fig. 5c). For *E. coli* ProQ, there is another open question: this protein was stumbled upon as being required for full expression of the proline transporter ProP (hence its name)\textsuperscript{119}, but the underlying mechanism remains unsolved.

**The rich world of r-proteins.** The largest functional class of bacterial RBPs is those involved in protein synthesis, such as r-proteins, ribosome-associated proteins, tRNA synthetases and enzymes that modify tRNA and/or rRNA. There are 57 bacterial r-proteins, 34 of which are conserved in all domains of life\textsuperscript{1}. Both the biogenesis of ribosomes and the ribosome-mediated process of translation are highly complex and dynamic processes that involve countless transient and stable interactions of r-proteins with RNA. As such, the rapidly growing information on ribosome assembly pathways\textsuperscript{119} and molecular structures\textsuperscript{120} should provide a ‘treasure trove’ for the understanding of RBP functions. Upon binding, an r-protein may induce a local structural change in the rRNA that enables the association with other proteins\textsuperscript{130}. Interestingly, however, r-proteins also help to move the mRNA within the ribosome and secure translational accuracy through selection of the correct aminoacyl-tRNAs\textsuperscript{132}.

One interesting aspect of r-proteins involves their general use of positively charged residues, rather than of defined amino acids, to specifically recognize the shape and charge of the RNA backbone\textsuperscript{133}. Many of these proteins possess globular domains with classic RBPs that are extended by elongated tails and internal loops to enable contacts with distinct, and sometimes multiple, rRNA regions\textsuperscript{132}. However, another common RNA target has helped to better understand the specificity of r-proteins: their own mRNA\textsuperscript{134}.

As a substantial proportion of cellular resources is devoted to produce ribosomes, the synthesis of each r-protein is controlled to match the stoichiometry of the other ribosome constituents. In most cases, this is achieved through autoinhibition at the level of translation initiation. Autoinhibition by r-proteins usually involves molecular mimicry; the mRNA presents a binding motif akin to the cognate site in the rRNA\textsuperscript{135}. A difference in the affinity of the two sites ensures that r-protein-mediated auto-repression occurs only when rRNA synthesis is reduced, for example, during starvation. Mechanistically, the molecular mimicry can entrap the ribosome\textsuperscript{136} or outcompete the 30S subunit\textsuperscript{137} (Fig. 3a).

The r-protein S1 generally promotes protein synthesis by unfolding structured mRNAs to enable docking of the 30S subunit and optimal positioning of the start codon\textsuperscript{138} (Fig. 3b). For autoregulation, it uses a mechanism distinct from the above: its C-terminal region anchors S1 on the ribosome via protein–protein interactions, while its N-terminal region binds AU-rich RNA sequences. When S1 is in molar excess over ribosomes, free S1 binds its own mRNA to inhibit translation\textsuperscript{139}.

Autoinhibition via RNA binding has become a recurrent theme for bacterial RBPs, well beyond r-proteins. Other well-studied examples are the translation initiation factor IF-3, which suppresses translation initiation at the non-canonical AUU start codon of its own mRNA\textsuperscript{140}, and threonine–tRNA synthetase (ThrRS), which translationally represses its synthesis by binding to its mRNA at a 5’ UTR–located RNA motif that structurally mimics its substrate, tRNA\textsubscript{Thr} (REFS\textsuperscript{131–133}). Moreover, several nucleases cleave their own mRNA\textsuperscript{134,135}.

**RNA turnover**

Bacterial RNA turnover is generally sequential: transcripts are first attacked by single-strand–specific endoribonucleases, such as RNase E or RNase Y, or the double-strand–specific endoribonuclease RNase III before decay is completed by exoribonucleases and oligoribonucleases\textsuperscript{141}. These activities can be combined in multi-protein RNA degradation complexes called degradosomes, which are found in many bacterial species\textsuperscript{142}.

For RBPs that regulate the stability of cellular transcripts, major mechanisms include the direct competition with RNases for the same site, the RBP-mediated recruitment of RNases to specific RNA ligands and the RBP-dependent positioning of RNase cleavage sites\textsuperscript{137} (Fig. 6). Regarding CsrA, Hfq and ProQ, we focus...
**Endoribonucleolytic RNA decay**

| Activation |
|------------|
| sRNA → mRNA |

**Inactivation**

| RapZ | RNase |
|------|--|
| sRNA | Hfq |

**Exoribonucleolytic RNA decay**

| Activation |
|------------|
| RNA helicase |

**Inactivation**

| Hfq or ProQ |
|-------------|

**Fig. 4 | RBP-dependent regulation of RNA decay.** Mechnisms by which RNA-binding proteins (RBPs) activate or inhibit RNA degradation by endoribonucleases (part a) or exoribonucleases (part b), respectively. a | Activation: RBPs such as RNase adapter protein RapZ and Hfq can recruit the major endoribonuclease RNase E to specific RNAs for degradation. Inactivation: binding of RBPs can directly compete for endoribonuclease cleavage sites to stabilize RNA transcripts. b | Activation: RNA helicases unwind RNA secondary structures to permit degradation of structured RNA by exoribonucleases. Inactivation: RBPs, such as Hfq and RNA chaperone ProQ, prevent exoribonucleolytic decay by binding to RNA 3′ ends. CspC, cold-shock protein CspC; CspE, cold-shock protein CspE; CsrA, carbon storage regulator A; sRNA, small non-coding RNA.

The helicase in the E. coli degradosome is RhlB. Stimulated by direct contact with RNase E, RhlB unwinds RNA secondary structures to enable full decay by the exoribonuclease PNPase, which is another component of the degradosome. CshA is the corresponding DEAD-box protein in the degradosomes of Gram-positive species. Interestingly, work in S. aureus suggests a more specific function of CshA in accelerating the degradation of mRNAs from the agr quorum-sensing system to promote biofilm formation.

Another way for RBPs to promote RNA decay is the active recruitment of nucleases. For example, it was proposed that Hfq-associated sRNAs, as they base pair with mRNAs to inhibit translation, also actively recruit RNase E to render silencing irreversible. Recruitment would involve the formation of an sRNA–Hfq–RNase E complex that is distinct from the degradosome (Fig. 4a). Such a triple complex in which the RBP Hfq recruits RNase E is yet to be proved structurally; in fact, others have argued that the sRNA rather than the RBP recruits RNase E. Regardless of the nature of this elusive complex, there is evidence that Hfq and sRNAs can guide RNase E to cleave mRNAs, even in the absence of translational inhibition.

A well-characterized example of RBP-mediated nuclease recruitment is RNase adapter protein RapZ (previously known as YhbJ) (Fig. 4a). Originally identified through a phenotype in amino sugar metabolism, the 32 kDa RapZ protein of E. coli was shown to selectively bind and present the GlmZ sRNA to RNase E for inactivation. Structural analysis has indicated a molecular origin of this RBP from metabolic enzymes. The Csr pathway contributes another putative adaptor protein for specific RNA decay: the ~73 kDa CsrD protein stimulates RNase E if it recruits RNase E143. Regardless of the nature of this elusive complex, there is evidence that Hfq and sRNAs can guide RNase E to cleave mRNAs, even in the absence of translational inhibition.

**RNA stabilization and processing.** Two major principles of how RBPs positively affect transcript stability have emerged: direct competition with an endoribonuclease for an internal cleavage site and competition with an exonuclease activity at the transcript termini (Fig. 4a,b). The former has been well studied for Hfq-associated RNAs, which are stabilized by Hfq occupying an internal RNase E recognition site. A comparison of recent global maps of Hfq occupancy and RNase E binding sites in Salmonella indicates Hfq-mediated protection to be common among sRNAs. Moreover, by suppressing certain RNase E sites in precursor transcripts, Hfq guides RNase E to the correct processing site in some sRNAs. This role of Hfq in processing seems to be particularly important for sRNAs that are derived from mRNA 3′ UTRs. By contrast, Hfq-mediated protection of mRNAs against RNase E largely depends on associated on direct mechanisms where changes in RNA stability are more than a consequence of altered translation rates.

**RNA decay.** Although the composition of degradosomes varies among bacteria, all contain members of the DEAD-box family of RNA helicases. RNA helicases are RBPs that unwind RNA duplexes powered by ATP hydrolysis. Because most RNases act on single-stranded RNA, the activity of an RNA helicase usually promotes RNA degradation (Fig. 4a).
not only in its decoy, the CsrB sRNA148, but also in the protein of S. aureus negatively autoregulates its own mRNA of the E. coli binant CspC or CspE proteins in vitro48. The CspA resists RNase E attack when pre-incubated with recom-

Complementing in vivo evidence, the ecnB mRNA fully end152. 12 RNase E target sites that all cluster in its 5′ end152,184. In regard to CSPs, one promising example is the short than internal entry18). In regard to CSPs, one promising example is the short ccsA mRNA70, raising the pos-

flhDC end of the flhDC mRNA14, raising the possibility that CsrA blocks RNase E access at the earliest possible stage (for RNase E, 5′-initiated decay is faster than internal entry185). In regard to CSPs, one promising example is the short ccsA transcript, which possesses 12 RNase E target sites that all cluster in its 5′ end152. Complementing in vivo evidence, the ccsA mRNA fully resists RNase E attack when pre-incubated with recombinant CspC or CspE proteins in vitro46. The CspA protein of S. aureus negatively auto- regulates its own expression by counteracting RNase III processing in the cspA 5′ UTR46.

Although endonucleolytic cleavage is the rate-limiting step in bacterial RNA turnover, exoribonucleolytic activity can influence RNA decay rates too. Consequently, RBPs can exert stabilizing effects at transcript termini. Hfq does so by binding to RNA 3′ ends, where it antagonizes exonuclease activity or stimulates polyadenylation154. Examples include sRNAs whose half-lives decrease upon loss of Hfq but are restored when PNPase is simultaneously inactivated119. CspC and CspE may also have a general effect on RNA degradation as they can counteract PNPase activity118. Adding to this list, a fine-mapping of ProQ sites in E. coli107 predicted no protein partners of Hfq and only three for CsrA, RNA chaperone ProQ was associated with an astonishing 40 non-ribosomal proteins. Albeit, this number should be treated with caution; the presence of higher-order ProQ-containing complexes is supported by the aforementioned gradient profiling10.

Box 3 | Emerging aspects of bacterial RBPs

**Bifunctional RBPs**

Global screens in eukaryotes have revealed not only many RNA-binding proteins (RBPs) of unknown function but also many proteins with established functions (especially metabolic enzymes that seem to moonlight as RBPs)13. Although analogous RBPs are lacking in bacteria, bifunctional RBPs are also known in bacteria. Aconitase, an iron-containing tricarboxylic acid (TCA) cycle protein, also functions as a post-transcriptional regulator. In Bacillus subtilis, this non-metabolic activity regulates physiological processes such as sporulation147. Both bacterial and eukaryotic aconitases autoinhibit their expression via RNA-binding at low intracellular iron concentration; in Escherichia coli, aconitase protects its own mRNA against RNase E174. Other bifunctional RBPs include cytoskeletal protein RodZ, which promotes virulence mRNA decay in Shigella sonnei155; translocator protein YopD, which inhibits translation of Yersinia virulence factors; RNA-modification enzyme TruB, which moonlights as a tRNA chaperone176; and (p)ppGpp synthetase RelQ of Enterococcus faecalis, the enzymatic activity of which is inhibited by single-stranded RNA177. Global DNA-binding proteins HU, HN-S and StpA also bind RNA in vivo: HU binds non-coding RNA to promote DNA condensation179, and HN-S and StpA alter the half-lives of some small non-coding RNAs (sRNAs)180,181. Similarly, the DNA-binding virulence transcriptional regulator SarA in Staphylococcus aureus was reported to be a global RNA binder181. Determining the full set of RNA ligands for these proteins with RNA sequencing (RNA-seq)175 may broaden our knowledge of how enzymatic and structural functions are combined with RNA-binding activities.

**Cellular localization of RBPs**

Bacterial cell biology has shown that even these simple organisms exhibit ample spatial and temporal organization of both their proteome and transcriptome182. For example, a membrane-localized ribonucleoprotein particle (RNP) that confers resistance to alcohol-induced and cold-induced stress in Firmicutes contains the non-coding RNA OLE, which is anchored to the inner membrane through the trans-membrane domain-containing RBP OLA183. Also implicated in RNA localization are the aforementioned RodZ protein in S. sonnei184 and carbon storage regulator A (CsrA) in Campylobacter jejuni, which influences the localization of flagellin mRNA185.

These recent findings notwithstanding, intracellular localization remains a greatly understudied aspect of bacterial RBP biology. Among the available resources, the ASKA collection of GFP-tagged E. coli ORFs lends itself for a bird’s eye view of the cellular localization of an RBP of interest185. However, the results should be interpreted with caution because of mislocalization issues with fluorescent tags, as demonstrated for Hfq186.

**RNA–protein complexes**

Eukaryotic RBPs often participate in dynamic higher-order complexes, such as the spliceosome or the RNA-induced silencing complexes187. Except for the ribosome and SRP, these complexes are not conserved in bacteria. Conversely, bacteria contain several complexes that are absent from eukaryocytes, including the degradosome188, several CRISPR–Cas-related RNPs and complexes of toxin–antitoxin systems189.

How many bacterial RNPs may be out there? A proteome analysis of Salmonella gradient fractions190 reveals many candidate proteins with RNA-related functions that form complexes greater than their own molecular mass (see FIG. 10). In addition, global protein–protein interaction data that have become available for different bacteria over the past 2 decades may guide the discovery of additional complexes. Although the latest screen in E. coli107 predicted no protein partners of Hfq and only three for CsrA, RNA chaperone ProQ was associated with an astonishing 40 non-ribosomal proteins. Albeit, this number should be treated with caution; the presence of higher-order ProQ-containing complexes is supported by the aforementioned gradient profiling10.
RNase E accumulates at the inner membrane — one may speculate about additional mechanisms of modulating RNA decay, such as an RBP-mediated relocalization of specific transcripts into or out of areas with high nucleolase activity.

Outlook
This Review aimed to provide a high-level view of how and where bacterial RBPs function at the levels of transcription, translation and RNA decay. There are several other exciting aspects of bacterial RBPs that could not be fully covered here owing to space constraints. These include bifunctional metabolic enzymes that moonlight as RBPs and unexpected RNA–protein complexes, as well as growing appreciation of particular subcellular localization of RBPs (Box 3).

As to the true number of RBPs in bacteria, technical obstacles have prevented a transfer of game-changing in vivo crosslinking and purification technology for genome-wide cataloguing of eukaryotic RBPs (Supplementary Box 2). Although putative bacterial RBPs have already been identified bioinformatically, it may be only a matter of time before bacteriologists devise global experimental screens of similar success, for example, by using selective purification of RBPs with altered solvent preferences as the result of ultraviolet crosslinking to RNA or metabolic labelling of cellular RNA followed by affinity purification of RNA-bound proteins. The results of those eukaryotic screens have already revealed many previously unknown RBPs lacking classic RBDs, the counterpart of which await to be discovered in bacteria as well. Among the many exciting unanswered questions, such screens are expected to shed light on whether RBPs have only cytosolic functions or are even secreted to modulate RNA metabolism in the host cell of a pathogenic bacterium.

New RBPs may carry out fully unexpected functions, as illustrated by those of CRISPR–Cas systems that few bacteriologists had on their radar. Once understood, they can inform the design of artificial RBPs for synthetic types of gene regulation. Moreover, our general approach taken here for the sake of brevity naturally neglected that the bacteria are in fact a very large and diverse group of organisms. Inasmuch as Jacques Monod’s quote that “What is true for E. coli is true for the elephant” has been proved to be too simple, we have seen ample evidence suggesting that the major RBP functions described here on the basis of work in very few model species, predominantly gammaproteobacteria, may not be representative in the many poorly characterized, often unculturable bacteria from the human microbiota and environmental communities.

The growing information about which RBPs are expressed when and about their cellular RNA ligands now enables a systems-level understanding of how these regulatory proteins function. Integrating such data with large-scale functional and phenotypic screens will certainly be a fruitful approach to further understand RBP functionality.

Published online 11 July 2018
This study describes the crystal structure of TRAP in complex with a CAG repeat-containing RNA target. This is the first structural basis for TRAP-mediated regulation of transcription and translation.

Babitzke, R., Stults, T. J., Shire, S. J. & Yanofsky, C. TRAP, the trp RNA-binding attenuation protein of Bacillus subtilis, is a multisubunit complex that appears to recognize GUGA repeats in the trpE/D/CFBA and trpG transcripts. J. Bacteriol. 166, 31651–31656 (2001).

This study describes the crystal structure of TRAP in complex with a CAG repeat-containing RNA target. This is the first structural basis for TRAP-mediated regulation of transcription and translation.

Babitzke, R., Stults, T. J., Shire, S. J. & Yanofsky, C. TRAP, the trp RNA-binding attenuation protein of Bacillus subtilis, is a multisubunit complex that appears to recognize GUGA repeats in the trpE/D/CFBA and trpG transcripts. J. Bacteriol. 166, 31651–31656 (2001).

This study describes the crystal structure of TRAP in complex with a CAG repeat-containing RNA target. This is the first structural basis for TRAP-mediated regulation of transcription and translation.

Babitzke, R., Stults, T. J., Shire, S. J. & Yanofsky, C. TRAP, the trp RNA-binding attenuation protein of Bacillus subtilis, is a multisubunit complex that appears to recognize GUGA repeats in the trpE/D/CFBA and trpG transcripts. J. Bacteriol. 166, 31651–31656 (2001).
Holmqvist, E., Li, L., Bischler, T., Barquist, L. & Attaiech, L., Glover, J. N. & Charpentier, X. RNA. Davis, J. H. & Williamson, J. R. Structure and dynamics by an RNA chaperone and a multitarget small binding in Salmonella and E. coli suggest a target in this study, ultraviolet CLIP-seq maps of ProQ microorganisms. Nucleic Acids Res. 46, 1470–1485 (2011).

This study establishes that the Crc protein of Pseudomonas is an auxiliary protein. Sittka, A., Sharma, C. M., Rolle, K. & Vogel, J. Deep RNA degradosome. Nature 537, 684–685 (2015).

Alt-Bara, S. & Carposius, A. J. RNA degradosomes in bacteria and chloroplasts: classification, distribution and evolution of RNase E homologs. Mol. Microbiol. 97, 1021–1135 (2015). This is an excellent review describing the biology and evolution of RNase E.

Laalami, S., Zig, L. & Putzer, H. Initiation of mRNA decay in bacteria. Cell. Mol. Life Sci. 71, 1799–1829 (2014).

Jain, C. & Belasco, J. G. Autoregulation of RNase E a regulator of gene expression. Mol. Microbiol. 61, 1382–1395 (2007).

Caillet, J. et al. The modular structure of Escherichia coli threonyl-tRNA synthetase expression in vivo. J. Biol. Chem. 281, 10562–10571 (2006).

Dittrich, C., Sharan, M., Eulalio, A. & Vogel, J. CORPSE: an integrated computational pipeline for the sequence-based identification and characterization of RNA-binding proteins. Nucleic Acids Res. 45, 696 (2017).

Said, N. et al. In vivo expression and purification of aptamer-tagged small RNA regulators. Nucleic Acids Res. 37, e133 (2009).
