RNA-binding proteins in bacteria

Erik Holmqvist1 and Jörg Vogel2,3,*

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.

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 ancestor1. 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) superfamily2,3, the Sm and Sm-like domains4, the RNA recognition motif (RRM)5, the K homology (KH) domain6, the double-stranded RNA-binding domain (dsRBD)7 and the PAZ and PIWI domains8 (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 ligands9. 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 bacterium10 but only one-tenth of the annotated 1,500 RBPs in the human genome11,12 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 expression13.

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 sRNAs15 (FIG. 1d). 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 genome16 (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 assemblies18 (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 elsewhere21.

1Department of Cell and Molecular Biology, Biomedical Center, Uppsala University, Uppsala, Sweden.
2Helmholtz Institute for RNA-based Infection Research (HIRI), Würzburg, Germany.
3Institute of Molecular Infection Biology, University of Würzburg, Würzburg, Germany.
*E-mail: Joerg.Vogel@uni-wuerzburg.de

https://doi.org/10.1038/s41579-018-0049-5

NATURE REVIEWS | MICROBIOLOGY

VOLUME 16 | OCTOBER 2018 | 601
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.

Transcription termination. Bacteria use two general mechanisms for transcription termination\(^{11}\): 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 \(\lambda\).\(^{22}\) Following early observations that Rho also acted on host genes, this protein was linked to termination in >25% of all operons in \(\text{Escherichia coli}\).\(^{23}\) 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\(^{24}\). Rho is an essential protein in many bacterial species, including \(E.\ coli\), owing to diverse roles in the silencing of horizontally acquired genes\(^{25}\), inhibition of R-loop formation\(^{26}\) and suppression of replication–transcription conflicts\(^{27,28}\). Rho forms large (~280 kDa) homohexameric rings (Fig. 2d), presenting two RNA-binding sites in each of the monomers. Termination initiates as the oligosaccharide-binding (OB) fold of the primary RNA-binding site recognizes 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\(^{29}\). 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\(^{30}\). Alternatively, Rho may associate with RNAP already at the start of the transcription cycle\(^{30}\). 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\(^{39}\).

Intrinsic termination is controlled by Nus factors, which induce termination globally. Originally discovered in \(E.\ coli\) as host factors of phage \(\lambda\) (see BOX 1), Nus factor-like proteins have since been found in many other bacteria\(^{31}\). \(E.\ coli\) transcription termination/antitermination protein NusA is an essential protein with several different roles in termination: providing antitermination activity for phage transcription\(^{41}\) and enhancing RNA hairpin-dependent pausing and termination in chromosomal transcription\(^{32}\) (Fig. 2a). In \(\text{Bacillus subtilis}\), 25% of all termination events are ascribed to NusA, especially at suboptimal intrinsic terminators\(^{33}\). Not only does this protein affect intrinsic termination but it may also enhance Rho-dependent silencing of horizontally transferred DNA\(^{33}\) and globally antagonize Rho function\(^{41}\) (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 \(\lambda\)-based antitermination complex\(^{41}\). 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 \(\lambda\) domain and the N-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 \(\lambda\)-based antitermination complex, NusA contacts nascent RNA via the two KH domains\(^{38}\) (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 \(\text{B. subtilis}\), a negative regulator of the tryptophan biosynthesis operon\(^{39}\). The functional TRAP protein comprises a ring of 11 subunits\(^{40}\) (Fig. 2d) 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\(^{39}\). 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\(^{41}\).

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\(^{38}\) (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\(^{42}\). Purified CSPs decrease termination and pausing in vitro, and their overexpression in vivo increases the expression of genes preceded by several intrinsic terminators\(^{42}\).

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. Well-characterized RBPs that regulate each process are indicated. 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.

For example, Hfq, CsrA, ProQ, CspC, RapZ and RhlB

For example, Hfq, CsrA, ProO, Csd, Sm-like, FinO-like, PIWI, RRM, Miscellaneous RBDs

**Regulation of transcription termination**

**Regulation of translation**

**Regulation of RNA decay**

For example, Hfq, CsrA, ProQ, e-proteins, IF-3 and ThrRS

For example, CFH5, CspA, HutP, BglG and EutV

For example, NusA, TRAP and PyrR

**Termination and/or antitermination**

Chaperones, helicases, etc.

tRNA modification

tRNA synthetases

rRNA modification

Ribosome-associated

Translation factors

Ribonucleases

RNAP

RNase E

RNase II

S9

L11

30S subunit

50S subunit

Small complexes

Large complexes

Data for part e from REF. 10.
charged surface\textsuperscript{35,44}. Positively 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 \textit{B. subtilis}\textsuperscript{36} (Fig. 2d). Interestingly, the structural similarity and functional redundancy between the CSPs and S1 domains hint at a common ancestry from an ancient RBP\textsuperscript{36}.

Whereas \textit{E. coli} CspA globally unfolds mRNA structure to promote translation during acclimation after cold shock\textsuperscript{37}, CspG and CspE are now established as truly globally acting RBPs at regular growth temperature\textsuperscript{38,39}. Their combined activities affect the expression of 20% of all genes in \textit{Salmonella enterica} (henceforth \textit{Salmonella}) and are essential for the virulence of this pathogen\textsuperscript{40}. In \textit{Staphylococcus aureus}, CspA binds more than 500 different transcripts and has global effects on gene expression\textsuperscript{41}. However, it is unclear how many of these effects involve antitermination, as CSPs can also modulate translation initiation and RNA stability\textsuperscript{42,43–45} (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\textsuperscript{46}. Strikingly, the affinity for RNA targets is typically in the micromolar range\textsuperscript{47,48}, whereas many other RBPs described here have binding affinities in the nanomolar range. In vitro selection has identified putative consensus sequences for several \textit{E. coli} CSPs\textsuperscript{34}, 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\textsuperscript{49}.

Whereas CSPs represent global factors, other RBPs inhibit termination only at specific loci, often as part of a specific physiological response\textsuperscript{50}. The hexameric antiterminator \textit{hut} operon positive regulatory protein (HutP) acts on the histidine utilization genes in \textit{B. subtilis}. Its RNA-binding activity is activated by a conformational change that is triggered by sensing the intracellular concentration of \textit{L}-histidine\textsuperscript{51}. By binding to six NAG triplets (with N representing any nucleotide) between its own gene and the downstream \textit{hut} operon, HutP directly prevents the formation of the terminator hairpin to promote transcription elongation\textsuperscript{52}. By contrast, RBPs of the BglG (cryptic \textit{β}-glucoside bgl operon antiterminator (BglG) and EutV, 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: \textit{Escherichia coli} Rho bound to UUUUUU RNA (PDB ID: 5JJj), \textit{E. coli} NusA bound to nut RNA in context of the XN-based transcription antitermination complex (PDB ID: 2ASB), \textit{Geobacillus stearothermophilus} TRAP bound to a 53-nucleotide-long single-stranded RNA (PDB ID: 1C9S) and \textit{Bacillus subtilis} cold-shock protein CspB bound to UUUUUU RNA (PDB ID: 3PF5). BglG, transcription antiterminator BglG.

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 \textit{λ} and \textit{Mu}\textsuperscript{22,24}, whereas examples of negative regulation came from bacteriophages MS2 and T4 (REFS\textsuperscript{49,50}). Protein N from phage \textit{λ} was the first antiterminator protein to be discovered\textsuperscript{23}, which together with Nus factors (transcription termination/antitermination proteins) NusA, NusB, NusC and NusG, and RNAP forms a transcription antitermination complex (TAC)\textsuperscript{51}. The structural basis for this form of antitermination was recently described\textsuperscript{52}. 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\textsuperscript{53–55}.

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 FinO. This regulator requires the RBP fertility inhibition protein (FinO), which both stabilizes FinP and aids its annealing to the tral mRNA\textsuperscript{56}. FinO-like proteins have recently garnered much attention because of the unexpected functions of family members from bacterial chromosomes (see the main text)\textsuperscript{57,58}. 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\textsuperscript{58}. 
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** | Inhibition of translation.

**b** | Activation of translation.

**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. 116, The RNA Society.
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-wide (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. 3a,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 phyla. 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. coli*12; the former results from CsrA directly inhibiting translation of the first gene in a glycogen biosynthesis operon13; a dimer of CsrA recognizes two single-stranded GGA triplets in the glcC 5′ UTR to obstruct binding of the 30S subunit43 (Fig. 3a).

Csr/Rsm proteins operate truly globally, targeting several hundreds of mRNAs in some Gram-negative model species45–47. In addition to their canonical mechanism whereby Csr/Rsm binds in 5′ UTRs to inhibit translation, regulatory mechanisms affecting transcription or RNA decay have been described65,66.

Crosslinking immunoprecipitation coupled with high-throughput sequencing (CLIP-seq) studies capturing hundreds of CsrA–RNA interactions inside live *E. coli* and *Salmonella*65,66 have helped to refine the recognition motif as AUGGA, preferably located in the apical loops of RNA hairpins (Fig. 3c). This motif is not only found in mRNAs but also abounds in a particular set of sRNAs that act as decoys of Csr/Rsm proteins65,71. For example, the 369-nucleotide CarB sRNA contains no fewer than 22 GGA sequences, of which 18 are high-affinity CsrA binding sites.61 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 fashion14. 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 degradation. It is easy to envision 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 sRNAs14, 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 mRNAs54. Likewise, the interaction of CsrA with other proteins and the hagA mRNA precisely times the expression and secretion of the flagellin Hag in *B. subtilis*55–57.

The recognition mode of Csr/Rsm proteins seems highly conserved because homologues from distant species can often complement each other’s function in vivo. However, some bacteria express multiple Csr/Rsm homologues with partly non-redundant activities54. 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 target15. The growing atlases of in vivo binding sites16–18 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 intermediates18. 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 end19. 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 degradation20. 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 ribosomes.21. 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 expression22.
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\(^{66}\). To date, the primary role of Hfq is seen in helping sRNAs find and regulate trans-encoded mRNAs through short, imperfect base pairing interactions\(^{81}\). In *E. coli* and *Salmonella*, up to 100 sRNAs may recognize more than 25% of all mRNAs via Hfq, comprising thousands of potential RNA interactions\(^{43,44}\). 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\(^{14,45,86}\) (Fig. 3b).

How does Hfq work as a global RNA ‘matchmaker’? Hfq possesses a conserved Sm domain\(^{67,68}\) — 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\(^{66}\). 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\(^{80,86}\); additional rim interactions help to position the sRNA such that its seed region is poised for contacting mRNA targets\(^{65}\). By contrast, the distal face interacts with single-stranded A-rich sequences present in mRNA 5′ UTRs\(^{80,89}\), bringing potential mRNA target sites close to the sRNA-binding proximal face\(^{66}\). When Hfq is loaded with both sRNA and cognate mRNA, the rim contacts their UA-rich sequences to promote RNA pairing\(^{82}\). Although this is the canonical view of how Hfq contacts sRNAs and mRNAs, 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\(^{86}\).

The function of the fourth major site on Hfq, the poorly conserved and intrinsically disordered C terminus, has been a matter of controversy\(^{97,98}\). However, recent data suggest that it helps to displace bound transcripts, which may not only help to rid Hfq of nonspecific RNA binders\(^{89}\) but also promote a rapid cycling of cellular RNA whose total concentration exceeds that of Hfq\(^{100,101}\). 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\(^{99}\).

Following the pioneering structure analysis of *S. aureus* Hfq\(^{66}\), much of the above model is based on work on the *E. coli* and *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\(^{103}\). 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 *Salmonella* showed that members of this family have an intrinsic propensity to bind sRNAs\(^{104}\). 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\(^{105}\).

Even in Gram-negative species, Hfq can modulate mRNA translation by sRNA-independent mechanisms. For example, during the mismatch repair pathway in *E. coli*, the distal face of Hfq binds to the leader region of *mutS* mRNA to inhibit translation\(^{106}\). Moreover, studies of how Hfq of *Pseudomonas aeruginosa* directly represses translation of mRNAs under conditions of catabolite repression have suggested a direct interaction with the catabolite repression control (Crc) protein\(^{107}\). 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\(^{108}\). However, more recent work suggests that the translation inhibitor is Hfq, whereas Crc stimulates the mRNA-binding activity of Hfq\(^{107,109}\). 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\(^{107}\). 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.

**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 an RNA chaperone in antisense regulation of F plasmid conjugation in *E. coli*\(^{110}\).

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 *Salmonella* and *E. coli*\(^{110,111}\). 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\(^{112}\). Not only does ProQ stabilize this sRNA, it also strengthens RNA duplex formation\(^{111}\), a general property that had already been seen with artificial substrates\(^{113}\). RocC, a FinO-like protein expressed from the chromosome of *Legionella pneumophila*, also mediates translational repression, helping the RocR sRNA to bind multiple mRNAs of the DNA uptake system\(^{114}\).

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\(^\text{112}\). Cellular abundance would favour this scenario: there are 5–10 times more ProQ monomers (the active form) than Hfq hexamers\(^\text{10}\), albeit the number of RNA targets is similar\(^\text{111}\).

Potential FinO homologues are found on the chromosomes and on plasmids of alphaproteobacteria, betaproteobacteria, gammaproteobacteria and aciditiothiobacilli\(^\text{116,117}\), promising a rich diversity of physiological functions and molecular mechanisms. Moreover, results from different bacterial species\(^\text{85,111}\) 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 \textit{E. coli} and \textit{Salmonella} predicts ProQ to typically recognize a stable RNA hairpin, often a transcription terminator\(^\text{111}\). Similarly, FinO and RocC bind to the terminator hairpins of the FinP and RocR sRNAs, respectively\(^\text{10,111}\). In other words, ProQ seems to govern its global network of sRNAs and mRNAs by reading RNA structure rather than sequence\(^\text{111}\), while at the same time, it successfully discriminates against the abundant 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\(^\text{116,117}\)? A structural analysis of \textit{E. coli} ProQ indicates that in addition to the FinO domain, other regions contribute to RNA binding and thus may modulate target selectivity\(^\text{118}\) (Fig. 5c). For \textit{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)\(^\text{118}\), 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\(^\text{1}\). 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\(^\text{119}\) and molecular structures\(^\text{120}\) 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\(^\text{121}\). Interestingly, however, r-proteins also help to move the mRNA within the ribosome and secure translational accuracy through selection of the correct aminoacyl-tRNAs\(^\text{122}\).

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 rRNA backbone\(^\text{123}\). 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\(^\text{123}\). However, another common RNA target has helped to better understand the specificity of r-proteins: their own mRNA\(^\text{124}\).

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\(^\text{125}\). A difference in the affinity of the two sites ensures that r-protein-mediated autoinhibition occurs only when rRNA synthesis is reduced, for example, during starvation. Mechanistically, the molecular mimicry can entrap the ribosome\(^\text{125}\) or outcompete the 30S subunit\(^\text{127}\) (Fig. 3b).

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\(^\text{128}\) (Fig. 3a). For autorogulation, 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\(^\text{129}\).

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\(^\text{130}\), 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\(^\text{131-133}\). Moreover, several nucleases cleave their own mRNA\(^\text{134,135}\).

**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\(^\text{136}\). These activities can be combined in multi-protein RNA degradation complexes called degradosomes, which are found in many bacterial species\(^\text{136}\).

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\(^\text{137}\). Regarding CsrA, Hfq and ProQ, we focus...
a  Endoribonucleolytic RNA decay

<table>
<thead>
<tr>
<th>Activation</th>
<th>Inactivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>sRNA</td>
<td>mRNA</td>
</tr>
<tr>
<td>RapZ, RNase</td>
<td>Hfq</td>
</tr>
<tr>
<td>CsrA, CspC and CspE</td>
<td>CsrA, CspC and CspE</td>
</tr>
<tr>
<td>Hfq</td>
<td>Hfq</td>
</tr>
</tbody>
</table>

b  Exoribonucleolytic RNA decay

<table>
<thead>
<tr>
<th>Activation</th>
<th>Inactivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA helicase</td>
<td>mRNA</td>
</tr>
<tr>
<td>Hfq or ProQ</td>
<td>Hfq</td>
</tr>
</tbody>
</table>

Fig. 4 | RBP-dependent regulation of RNA decay. Mechanisms 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<sup>139</sup>. 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<sup>140</sup>. 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<sup>141</sup>.

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<sup>42</sup> [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<sup>143</sup>. 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<sup>144,145</sup>.

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<sup>146</sup>. Structural analysis has indicated a molecular origin of this RBP from metabolic enzymes<sup>147</sup>. The Csr pathway contributes another putative adaptor protein for specific RNA decay: the ~73 kDa CsrD protein stimulates CsrB decay by counteracting CsrA binding to an RNase E site<sup>148,149</sup>. Acting on a more global level, a bacterial homologue of Ro (a well-studied eukaryotic RBP) was recently shown to promote the degradation of highly structured RNAs in *Deinococcus radiodurans*. This RBP is tethered by the non-coding Y RNAs to PNPase, yielding a much larger RNP that displays different substrate specificity than PNPase alone<sup>150</sup>. Together, these examples provide molecular models for a better understanding of how RBPs contribute to shaping the selectivity of RNA decay.

**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<sup>151</sup>. A comparison of recent global maps of Hfq occupancy<sup>152</sup> and RNase E binding sites<sup>153</sup> 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<sup>132</sup>. This role of Hfq in processing seems to be particularly important for sRNAs that are derived from mRNA 3’ UTRs<sup>152,153</sup>. By contrast, Hfq-mediated protection of mRNAs against RNase E largely depends on associated
sRNAs that sequester crucial cleavage sites in the target or a higher ribosome density due to sRNA-induced translational activation\textsuperscript{146,144}. YbEY, an RNAse that carries a putative MID domain of Argonaute proteins, has also been predicted to have links with Hfq-dependent and/or sRNA-dependent RNA decay\textsuperscript{150}. That major RBPs directly protect mRNAs against endonuclease activity was demonstrated with CsrA and some CSPs (Fig. 4a). CsrA shields RNase E cleavage sites not only in its decoy, the CsrB sRNA\textsuperscript{148}, but also in the mRNA of the \textit{E. coli} master activator of motility genes, FlhDC\textsuperscript{38}. Interestingly, one of the two CsrB sites lies at the very 5’ end of the \textit{flhDC} mRNA\textsuperscript{39}, raising the possibility that CsrA blocks RNase E access at the earliest possible stage (for RNase E, 5’-initiated decay is faster than internal entry\textsuperscript{40}). In regard to CSPs, one promising example is the short \textit{ecnB} transcript, which possesses 12 RNase E target sites that all cluster in its 5’ end\textsuperscript{152}. Complementing in vivo evidence, the \textit{ecnB} mRNA fully resists RNase E attack when pre-incubated with recombinant CspC or CspE proteins in vitro\textsuperscript{45}. The CspA protein of \textit{S. aureus} negatively autoregulates its own expression by counteracting RNase III processing in the \textit{cspA} 5’ UTR\textsuperscript{40}.

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 polyadenylation\textsuperscript{154}. Examples include sRNAs whose half-lives decrease upon loss of Hfq but are restored when PNPase is simultaneously inactivated\textsuperscript{156}. CspC and CspE may also have a general effect on RNA degradation as they can counteract PNPase activity\textsuperscript{151}. Adding to this list, a fine-mapping of ProQ sites in \textit{E. coli} and \textit{Salmonella} revealed substantial binding of this RBP to mRNA 3’ ends. Genetic evidence suggests that ProQ stabilizes at least one of these targets, the \textit{cspE} mRNA, against 3’ → 5’ degradation by the major exoribonuclease RNase II\textsuperscript{111}.

Collectively, these examples highlight how RBPs may directly interfere with RNases at both internal and terminal transcript positions. Considering that nuclease may not be evenly distributed in a bacterial cell — for example,
RNase E accumulates at the inner membrane— one may speculate about additional mechanisms of modulating RNA decay, such as an RBP-mediated relocation 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 how 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 cytotoxic 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 a high-throughput method for analysing global RNA–protein interactions and establishes PrsQ as a global bacterial RBP.**


**This is an excellent overview of the findings from recent screens for eukaryotic RBPs.**

25. Leela, J. K., Syeda, A. H., Arupama, K. & Govrishankar, J. Rho-dependent transcription termination is essential to prevent excessive

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


Phadtare, S. & Severinov, K. *Structural* *analysis* *of* *Escherichia* *coli* *global* *RNA* *expression* *profiles* *in* *response* *to* *overexpression* *of* *the* *global* *transcriptional* *regulator* *gene* *CspA*. *J. Bacteriol.* 188, 2521–2527 (2006).


This paper describes the pioneering work of ultraviolet crosslinking and RNA sequencing in bacteria to globally map Hfq binding sites in vivo.


In this study, the Ro protein is shown to work with Hfq and RNA-dependent small RNA degradation in bacteria.
173. Benjamin, J. A. & Massé, E. The iron-sensing
174. Mitobe, J. et al. RodZ regulates the post-
170. Ataide, S. F. et al. The crystal structure of the signal
169. NADH oxidase RiesC of Bacillus subtilis
172. Neubauer, C., Gillet, R., Kelley, A. C. & Ramakrishnan,
168. Lopez-Alonso, J. P. et al. RsgA couples the maturation
167. Beljantseva, J. et al. Negative allosteric regulation of
176. Qian, Z., Zhurkin, V. B. & Adhya, S. DNA-RNA
178. Rajagopala, S. V. et al. The binary protein-protein
180. Deighan, P., Free, A. & Dorman, C. J. A role for the
186. Deighan, P., Free, A. & Dorman, C. J. A role for the
181. Rajagopala, S. V. et al. The binary protein-protein
175. Mitobe, J. et al. RodZ regulates the post-
174. Benjamin, J. A. & Massé, E. The iron-sensing
171. Huter, P., Muller, C., Arezu, S., Beckert, B. &
170. Naik, S. et al. The crystal structure of the signal
169. NADH oxidase RiesC of Bacillus subtilis
168. Lopez-Alonso, J. P. et al. RsgA couples the maturation
167. Beljantseva, J. et al. Negative allosteric regulation of
176. Qian, Z., Zhurkin, V. B. & Adhya, S. DNA-RNA
178. Rajagopala, S. V. et al. The binary protein-protein
180. Deighan, P., Free, A. & Dorman, C. J. A role for the
181. Rajagopala, S. V. et al. The binary protein-protein
177. Beljantseva, J. et al. Negative allosteric regulation of
176. Qian, Z., Zhurkin, V. B. & Adhya, S. DNA-RNA
178. Rajagopala, S. V. et al. The binary protein-protein