3'-UTRs as a source of regulatory RNAs in bacteria

Markus Gößringer and Roland K Hartmann*

Institut für Pharmazeutische Chemie, Fachbereich Pharmazie, Philipps Universität Marburg, Marburg, Germany

*Correspondence to: roland.hartmann@staff.uni-marburg.de

The EMBO Journal (2012) 31, 3958–3960. doi:10.1038/emboj.2012.269; Published online 25 September 2012

Bacterial small non-coding RNAs, sRNAs, have up to now been identified primarily in intergenic regions. Chao *et al* reveal that the 3'-region of mRNAs is another rich reservoir of sRNAs.

Parallel to the finding of various classes of small RNAi-associated effector RNAs in eukarya, bacterial RNA biology has experienced a similar massive discovery of small non-coding RNAs (sRNAs) with regulatory function. Bacterial sRNAs, commonly between 50 and 250 nt in length, bind their targets (usually mRNAs) via short base-pairing interactions that have been termed 'seed' regions analogous to eukaryotic miRNAs. Bacterial sRNAs frequently regulate multiple mRNA targets in response to environmental cues and physiological adaptations (for details, see Chao *et al*, 2012). As an example, the sRNA GcvB (~200 nt) negatively regulates up to ~1% of all mRNAs in *Salmonella enterica* Typhimurium (Sharma *et al*, 2011), thereby promoting

changes in peptide and amino-acid uptake as well as amino-acid biosynthesis.

An important cofactor of sRNA-mediated regulation of bacterial gene expression is the Sm-like homohexameric ring protein Hfq. Two binding sites for single-stranded RNA are provided by the ring-shaped core of the Hfq hexamer. Preferentially, U-rich RNA sequences bind to the proximal face and A-rich sequences to the distal face of the hexameric ring (Link *et al* 2009). Hfq mediates short base-pairing interactions of regulatory sRNAs with their *trans*-encoded target mRNAs and often simultaneously stabilizes the sRNA (Urban and Vogel, 2007). The protein hexamer is seen as a central hub on which sRNAs and mRNAs constantly exchange at its multiple RNA-binding sites (see active cycling model by Fender *et al*, 2010), with half-lives of individual RNA-Hfq complexes not exceeding 1 min. Hfq affects sRNA function by increasing sRNA-target RNA

Figure 1 sRNA DapZ is transcribed from an internal promoter in the 3' region of the lysine biosynthesis gene *dapB* (relative lengths of *dapB*-coding region and 3'-UTR not drawn to scale). DapZ expression is dependent on the HilD transcription factor that is the major regulator of *Salmonella* invasion genes. The Hfq protein hexamer binds and stabilizes DapZ and mediates its interaction with the translation-initiation region (target) of mRNAs *oppA* and *dppA* to repress their translation. This results in inhibition of cellular peptide uptake during activation of invasion genes; sRNA DapZ binds to its target regions via a G/U-rich R1 sequence element that resembles a corresponding element in another sRNA, GcvB. DapZ and GcvB target sites in *oppA* and *dppA* overlap; *oppA* and *dppA* are the first cistrons of polycistronic operons.

association rates (Soper and Woodson, 2008) and by promoting sRNA or mRNA unfolding as a prerequisite for their complex formation (chaperone function; Geissmann and Touati, 2004). In addition, Hfq can mediate decay of sRNA-target mRNA complexes or induce polyadenylation of mRNA to trigger its 3'- to 5'-exonucleolytic degradation (Vogel and Luisi, 2011). An Hfq homologue has been identified in at least 50% of the bacteria with available genome sequence (Vogel and Luisi, 2011). Whereas Hfq is considered to be generally required for effective sRNA function in enteric Gram-negative bacteria, such as *Salmonella* and *E. coli*, the protein is apparently dispensable for the function of several sRNAs in Gram-positive bacteria, such as *Staphylococcus aureus* or *Bacillus subtilis* (Gottesman and Storz, 2011).

Jörg Vogel and coworkers have studied Hfq-dependent sRNAs in *S. typhimurium*, a strain whose virulence is dependent on the presence of Hfq (Sittka *et al.*, 2007). In their very recent work, Chao *et al.* (2012) used a *Salmonella* mutant strain where the wild-type *hfq* gene is replaced with a gene variant coding for a FLAG-tagged Hfq protein. The tagged Hfq and associated cellular RNAs were co-immunoprecipitated and RNA fractions extracted and converted to cDNA for high-throughput sequencing (RNA-seq). Out of a total of ~3500 mRNAs detected in the control wild-type strain, ~1250 were enriched at least three-fold after co-immunoprecipitation (coIP) with Hfq. Taking Hfq's rather promiscuous RNA-binding properties into account, combined with its affinity for transcription terminators (stem-loop structures flanked by single-stranded segments), the task to extract meaningful information from this huge data set looks like a tough one. A clue to mastering this task was harvest of cells at different growth stages, including exponential phase, early stationary phase when *Salmonella* invasion genes are transiently activated, and more extended stationary phase. This revealed a distinct, growth phase-specific enrichment of Hfq-associated sRNAs. The authors found that the RNA-seq data agreed very well with the expression profiles of several sRNAs studied previously. They also noticed that their Hfq-enrichment libraries abundantly represented RNAs that overlapped in sense with the 3'-UTR of mRNAs, in addition to the classical sRNAs encoded in intergenic regions. For eight of these 3'-UTR RNAs, the size and abundance inferred from northern blots was in line with the RNA-seq data. Intriguingly, for the majority of those eight 3'-UTR RNAs, northern blots revealed a longer mRNA signal as well, whose time point of highest expression deviated in several cases from that of the overlapping 3'-UTR-derived sRNA.

Chao *et al.* selected one of the eight candidates for a functional in-depth analysis. The RNA was termed DapZ owing to its overlap with the 3'-UTR of the biosynthetic *dapB* mRNA (Figure 1). A physiological function of DapZ RNA was likely based on the fact that it contributed ~12% of the cDNA reads in the Hfq-enrichment library at the transition to stationary phase. In other words, DapZ occupied a substantial fraction of Hfq under these conditions, although the protein's cellular amounts are limited. Moreover, DapZ was clearly demonstrated to be Hfq-dependent, as its copy number dropped from ~100/cell to less than one/cell and its half-life decreased from ~2.5 min to <30 s in a Δhfq -mutant strain. As DapZ expression sharply peaked in early stationary phase where *Salmonella* invasion genes are transiently

activated, the authors suspected that its expression may be governed by a transcription factor associated with invasion genes located on the *Salmonella* pathogenicity island 1 (SPI-1). Using deletion strains and inducible complementation plasmids, Chao *et al.* found the transcription factor HilD, a known master regulator of *Salmonella* invasion genes, to control DapZ RNA expression in *Salmonella* (Figure 1). RNA-seq and RACE analysis further revealed that DapZ is predominantly, if not entirely, synthesized as a primary transcript from an internal promoter overlapping the stop codon of the *dapB* mRNA-coding sequence. Sequence analysis of DapZ RNA upstream of the terminator stem-loop and alignment with corresponding sequences from other related Gram-negatives attracted the authors' attention to a conserved 10-nt G/U-rich sequence element (termed R1) almost identical to the R1 'seed' of the global sRNA regulator GcvB. A series of elegant experiments then showed that DapZ translationally represses mRNAs of the *dpp* and *opp* operons (encoding major ABC transporters), which represent a subset of mRNAs repressed by the sRNA GcvB (for an overview, see Figure 1). The target sequences of GcvB and DapZ in *oppA* and *dppA* mRNAs—encoded by the first cistrons of the *opp* and *dpp* operon, respectively—fully overlap and likely block ribosomal initiation at these mRNAs, with the GcvB RNA interaction being somewhat more extended than that for DapZ RNA.

In conclusion, Chao *et al.* (2012) have unravelled a new facet of the post-transcriptional Hfq-sRNA network: they identified DapZ sRNA, encoded in the 3'-UTR of the conserved *dapB* gene, as a translational repressor with a GcvB-like seed domain. This is the first example of a widely conserved bacterial gene whose 3'-UTR has dual function. Their study further indicates that seven other *Salmonella* 3'-UTRs are subject to the same dual use as the *dapB*/DapZ pair. Some of these sRNAs are transcribed from their own gene-internal promoter and independent of the mRNA at the same locus, others are released by processing of the respective full-length mRNA. The authors thus posit that 3' regions of mRNA transcripts will generally be a large reservoir for Hfq-dependent sRNAs in bacteria. Specifically, expression of DapZ sRNA is mechanistically unique to *Salmonella*, as the RNA is transiently expressed under control of the horizontally acquired transcription factor HilD to repress certain ABC transporters as part of a global adaptation to conditions of host cell invasion. The analysis by Chao *et al.* represents the most comprehensive analysis of Hfq-associated RNAs to date. The DapZ RNA would have escaped identification by any of the existing bioinformatic algorithms applied to the search for potential sRNAs, owing to its poorly conserved primary sequence and its overlap with the conserved *dapB* gene. Finally, the study has substantiated the 'seed' concept of bacterial sRNAs, analogous to eukaryotic miRNAs. Both DapZ and GcvB RNAs bind their target mRNAs *oppA* and *dppA* with a 5'-proximal fully complementary seed, followed by a bulge or internal loop and then a second base-pairing interaction. One difference to miRNAs is that bacterial sRNAs predominantly bind at or close to the ribosome-binding site of mRNAs, whereas miRNAs mainly target the 3'-UTRs of mRNAs.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J (2012) An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J* **31**: 4005–4019
- Fender A, Elf J, Hampel K, Zimmermann B, Wagner EG (2010) RNAs actively cycle on the Sm-like protein Hfq. *Genes Dev* **24**: 2621–2626
- Geissmann TA, Touati D (2004) Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO J* **23**: 396–405
- Gottesman S, Storz G (2011) Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol* **3**: a003798
- Link TM, Valentin-Hansen P, Brennan RG (2009) Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *Proc Natl Acad Sci USA* **106**: 19292–19297
- Sharma CM, Papenfort K, Pernitzsch SR, Mollenkopf HJ, Hinton JC, Vogel J (2011) Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol* **81**: 1144–1165
- Sittka A, Pfeiffer V, Tedin K, Vogel J (2007) The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Mol Microbiol* **63**: 193–217
- Soper TJ, Woodson SA (2008) The *rpoS* mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA* **14**: 1907–1917
- Urban JH, Vogel J (2007) Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*. *Nucleic Acids Res* **35**: 1018–1037
- Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. *Nat Rev Microbiol* **9**: 578–589