Regulatory RNA in *Mycobacterium tuberculosis*, back to basics

Stefan Schwenk and Kristine B. Arnvig*

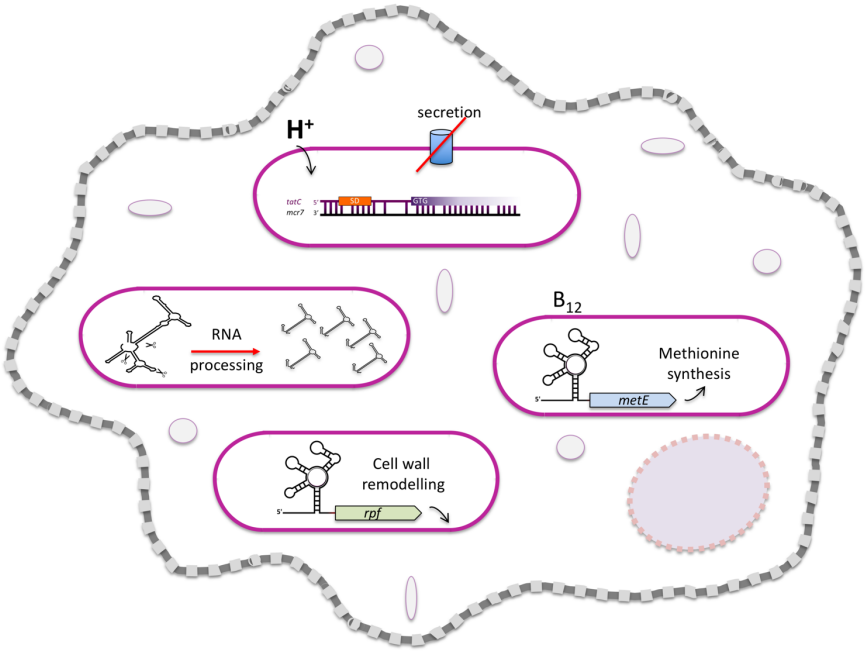
Institute for Structural and Molecular Biology, University College London WC1E 6BT, UK

Keywords: *Mycobacterium tuberculosis*, post-transcriptional regulation, regulatory RNA, small RNA, riboswitch, riboregulation

*Corresponding author: Institute for Structural and Molecular Biology, Darwin Building, University College London WC1E 6BT, UK; tel. +44 2076792242; k.arnvig@ucl.ac.uk

Abstract

Since the turn of the millenium, RNA-based control of gene expression has added an extra dimension to the central dogma of molecular biology. Still, the roles of *Mycobacterium tuberculosis* regulatory RNAs, and the proteins that facilitate their functions remain elusive, although there can be no doubt that RNA biology plays a central role in the baterium's adaptation to its many host environments. In this review we have presented examples from model organisms and from *M. tuberculosis* to showcase the abundance and versatility of regulatory RNA, in order to emphasize the importance of these 'fine-tuners' of gene expression.



Graphical abstract: aspects of *M. tuberculosis* regulatory RNA discussed in this review.

Introduction

Bacterial gene expression consists of two tightly coupled processes, transcription and translation. A detailed, systematic and molecular to global characterisation of how these processes are regulated in pathogens is critical for development and improvement of disease interventions. While the regulation of transcription initiation is chiefly protein-based

and orchestrated by sigma- and transcription factors, post-transcriptional regulation is to a large extent influenced by regulatory RNA species or 'riboregulators'.

Until the turn of the millenium, riboregulators were few and far between; their discovery had been serendipituous, and in particular small regulatory RNAs (sRNAs) were mostly associated with plasmid copy number control, phages and transposable elements, reviewed in (Wagner *et al.*, 2002). Two publications from 2001 signalled a change in attitude towards regulatory RNA with systematic searches for intergenic sRNAs based on cDNA cloning, tiling arrays and bioinformatics (Argaman *et al.*, 2001, Wassarman *et al.*, 2001). This was soon followed by the identification and characterisation of a number of highly conserved *cis*-regulatory elements, e.g. (Grundy *et al.*, 2002, Mironov *et al.*, 2002, Nahvi *et al.*, 2002, Winkler *et al.*, 2002). The trend continued at an accelerated pace with the advent of Deep sequencing methods, including variations such as dRNA-seq, ribo-seq, term-seq and grad-seq providing unbiased and comprehensive mapping of global transcriptomes e.g (Ingolia *et al.*, 2009, Sharma *et al.*, 2010, Dar *et al.*, 2016, Smirnov *et al.*, 2016). Currently, a plethora of riboregulators with steadily increasing complexity and functionality has been revealed in a substantial number of prokaryotic species, reviewed in (Wagner & Romby, 2015), and by now, the number of riboregulators is likely to equal or exceed the number of transcription factors in many bacteria (Ishihama, 2010, Rau *et al.*, 2015, Holmqvist & Wagner, 2017, Smirnov *et al.*, 2017). The more we uncover, the clearer it becomes that just like the processes of transcription and translation are tightly coupled, the functions of their regulators are interwoven and part of each other's regulons (Arnvig & Young, 2010, Beisel & Storz, 2010, Lee & Gottesman, 2016).

RNA-based control of gene expression

The question 'why regulatory RNA?' keeps emerging in scientific discussions. There are several answers to this question that to some extent dependent on the type of element. Judging by their evolutionary conservation, some regulatory RNAs, including certain riboswitches are very old and possibly remnants from the RNA world and thus predate

proteinaceous regulators (McCown *et al.*, 2017). This is not the case for sRNAs, which have evolved later than other non-coding RNAs (ncRNAs) (Peer & Margalit, 2014). Many sRNAs are induced by stress and their synthesis is faster and less costly than transcription and translation of a proteinaceous transcription factor, two qualities that may be of significance during stress. In addition, one might argue from first principles that RNA is the obvious interaction partner for RNA - because basepairing is less complicated than evolving RNA-protein interactions. New sRNAs are continuously emerging as a result of single nucleotide polymorphisms (SNPs) leading to spurious promoters e.g. (Rose *et al.*, 2013), and probably continuously evolving, as sRNA-mRNA interactions can be easily modulated by a few SNPs adjusting their basepairing (Updegrove *et al.*, 2015). Whether these SNPs become fixed depends on the resulting fitness gain or -loss.

Several reports indicate that sRNA regulators have different kinetic properties than protein regulators and that (sRNAs as well as asRNAs) may play a role in suppressing transcriptional noise (Levine & Hwa, 2008, Lasa *et al.*, 2011, Holmqvist & Wagner, 2017).

Finally, it is important to keep in mind that regulatory RNAs operate not instead of, but in addition to, conventional protein-based regulation thereby significantly expanding the number of available components and the complexity of cellular regulatory network.

M. tuberculosis differs from model organisms

M. tuberculosis is subject to multiple stresses and environments through the course of transmission, infection, immune response attack, dormancy and resuscitation, and surviving these changes requires extensive rewiring of its gene expression programme. Regrettably, protein-centric gene expression analysis still provides the main point of reference for *M. tuberculosis*. Our knowledge about mycobacterial riboregulation to a large extent still remains ‘dark matter’, although there is currently no doubt that this is an important aspect of *M. tuberculosis*’s intracellular life. Most of our knowledge about bacterial riboregulation originates from studies of the Gram-negative *Escherichia coli* and *Salmonella typhimurium*,

and the Gram-positive *Listeria monocytogenes, Bacillus subtilis* and *Staphylococcus aureus*. These model organisms have provided extensive and novel insights into the structure, function and mechanisms of different types of regulatory RNA. In addition to integrating 'ribo-knowledge' into our understanding of how *M. tuberculosis* gene expression is controlled, we may ask whether it has anything to offer to the already vast knowledgebase of RNA biology or whether it is all about finding a cure? Should we simply rely on extrapolating findings from significantly more tractable model organisms instead of sweating in the Cat 3 lab playing catch up? The answer is a resounding 'No!' and the reasons will hopefully be clearer after reading this review. In here we provide a brief overview of different types of riboregulation, showcasing pertinent examples from model organisms as well as *M. tuberculosis*. Due to the rapidly expanding field of regulatory RNA, we are not able to provide a comprehensive overview, and we apologize in advance for the ommission of our colleagues' work that has not been included due to space constraints. We refer to more extensive reviews on riboswitches e.g. (Serganov & Nudler, 2013, Sherwood & Henkin, 2016, Quereda & Cossart, 2017) and on sRNAs e.g. (Updegrove *et al.*, 2015, Wagner & Romby, 2015, Smirnov *et al.*, 2017). Our key aim is to highlight and further the appreciation of riboregulators as important operators in the regulation of gene expression in *M. tuberculosis.*

Cis- and *trans*-, non-coding and regulatory RNAs

In the early days of the era of riboregulation, classifications were relatively simple. However, as we learn more about the origins and the functions of the different types of regulatory RNA, the boundaries between *cis*-regulatory elements, *trans*-acting RNA, and *cis*- and *trans*-encoded regulatory RNAs and have become increasingly blurred. In BOX 1 we offer a short list of definitions, bearing in mind that there may be overlaps between these.

BOX1

Non-coding RNA (ncRNA) is a transcript or part of a transcript that does not encode protein or peptide; it may or may not be regulatory. Conversely, there are several examples of regulatory RNAs that are coding.

***Cis*-regulatory elements** are part of the RNAs they regulate, and they include RNA leaders (5' of both coding and ncRNA) and 3' untranslated regions (UTRs), both of which may be a source of sRNAs.

Trans-acting RNAs, as opposed to ***cis*-regulatory elements**, refer to RNAs that are not part of the transcripts they regulate; they include antisense RNA (asRNA) and small regulatory RNA (sRNA).

sRNAs comprise a small number of protein binding RNAs, and a large number of basepairing RNAs, which again are divided into *cis*- or *trans*-encoded.

asRNAs are encoded opposite their targets and sizes vary significantly.

Dual function sRNAs are transcripts that act both as riboregulators and as templates for protein synthesis.

RNA chaperones and ribonucleases

As riboregulators adopt more prominent roles, the proteins that regulate their expression, stability and/or degradation have attracted more attention as well. Two classes of proteins important for riboregulator function are the RNA chaperones and Ribonucleases (RNases).

RNA chaperones can have a profound influence on the effect that riboregulators can exert, both as molecular matchmakers, but also due to their role in modulating the stability of these regulators, reviewed in (Vogel & Luisi, 2011). For several years, the widely conserved RNA chaperone, Hfq was considered *the* facilitator of sRNA-mRNA interactions, but different species rely on Hfq to different extents. For example, several sRNAs in the otherwise highly Hfq-dependent *S. typhimurium* and *E. coli* do not bind Hfq; there are conflicting reports on the requirement for Hfq in Gram-positive bacteria, and certain species including *Helicobacter*

pylori and *M. tuberculosis* do not encode an Hfq homologue at all (Jousselin *et al.*, 2009, Arnvig & Young, 2012, Oliva *et al.*, 2015).

Recently, the new global RNA chaperone ProQ was discovered by a combination of density gradient fractionation with mass spectrometry- (of proteins) and and RNA-seq analysis of each fraction (Grad-seq) (Smirnov *et al.*, 2016). Similar to Hfq, there is no ProQ homologue in *M. tuberculosis*, but its identification does rekindle the question as to whether *M. tuberculosis* sRNAs require a matchmaker and if so, how does it work?

A third contender as a general RNA chaperone with a wide range of targets is Cold Shock Protein A (CspA), which belongs to a large family of Csp's, and this protein does have a highly expressed homologue in *M. tuberculosis* (Arnvig *et al.*, 2011, Caballero *et al.*, 2018). CspA is an RNA binding protein that facilitates the melting of secondary RNA structures, which are otherwised stabilised at low temperatures (Jiang *et al.*, 1997). Although well-expressed at all times, the *E. coli* cspA mRNA is itself regulated by a thermoswitch that leads to increased translation at low temperatures (Giuliodori *et al.*, 2010). Whether *M. tuberculosis* CspA ia regulated in a similar manner remains to be seen, but certainly an RNA chaperone that specialises in melting highly structured RNA would seem appropriate for *M. tuberculosis*.

Degrading RNA

The abundance of a transcript is a carefully controlled balance between synthesis and degadation, and hence, ribonucleases (RNases) are crucial players in riboregulation.

In addition to encoding an unprecedented number of type II toxins (Ramage *et al.*, 2009, Sala *et al.*, 2014), *M. tuberculosis* (and other mycobacteria) encode a curious complement of RNases, with elements from both Gram negative (*E. coli*) and Gram positive (*B. subtilis*) species. A comprehensive review of the similarities and differences can be found in (Durand *et al.*, 2015), and we will only mention a few important corner stones. Mycobacteria contain functional homologues of both RNase E and RNase J, while RNase E is absent from *B.*

subtilis and RNase J is absent from *E. coli* (Even *et al.*, 2005, Taverniti *et al.*, 2011, Durand *et al.*, 2015).

The endonucleolytic activity of *E. coli* RNase E as well as the exonucleolytic activity of *B. subtilis* RNase J are both sensitive to the phosphorylation state of the 5' nucleotide of their substrates, i.e. they both have a strong preference for mono-phosphorylated transcripts as substrates, and the same 'rules' are likely to apply to the *M. tuberculosis* enzymes (Mackie, 1998, Koslover *et al.*, 2008, Li de la Sierra-Gallay *et al.*, 2008, Mathy *et al.*, 2010, Taverniti *et al.*, 2011). RNA 5' monophosphates can be generated either by endonucleolytic cleavage of a transcript or by the removal of a pyrophosphate group from the 5' nucleotide by the action of the Nudix hydrolase, RppH (RNA pyrophosphohydrolase) (Deana *et al.*, 2008). So far, there are no reports of RppH homologues in mycobacteria, although there are some indications that such an enzyme exists (Moores *et al.*, 2017). Regardless of the enzymes involved and the pathways employed, the degradation of RNA is as important as its synthesis for managing resources and ensuring appropriate execution of the cell's gene expression programme.

RNA 5' leaders

RNA leaders serve as hubs for post-transcriptional regulation, which in many cases involves some means of controlling ribosome entry. In its most basic form, the RNA leader is short and simply provides a Shine-Dalgarno (SD) sequence for binding of the 30S ribosomal subunit (Shine & Dalgarno, 1974). Translation efficiency can be modulated by changing the SD sequence to be more or less complimentary to the 16S sequence or by altering the spacing between the SD and the start codon (Vellanoweth & Rabinowitz, 1992). Certain (longer) RNA leaders have the ability to switch between mutually exclusive conformations that are either permissive or non-permissive for downstream gene expression (transcription or translation), and these are referred to as RNA switches. The switching between alternative RNA structures can be triggered by changes in temperature, pH, metal ions and

metabolites, but also by RNA and proteins (Babitzke *et al.*, 2009, Nechooshtan *et al.*, 2009, Ferre-D'Amare & Winkler, 2011, Kortmann & Narberhaus, 2012, Sherwood & Henkin, 2016, McCown *et al.*, 2017). Moreover, the leader is often targeted by sRNAs with a variety of outcomes.

Protein binding 5' leaders

Many 5' leaders regulate downstream gene expression by binding specific proteins reviewed in (Babitzke *et al.*, 2009). This is particularly evident in the case of ribosome biosynthesis, which represents a major drain on cellular resources, and therefore has to be tightly regulated. The expression of many ribosomal protein (r-protein) operons is regulated by direct binding of one or more r-proteins encoded in these operons to their cognate mRNA leaders, usually proximal to the translation initiation region (TIR) typically blocking ribosome entry. This has been extensively characterised in *E. coli* and examples include r-proteins S4, S8, L20, and a complex of L10/L12 (Babitzke *et al.*, 2009). The highly conserved organisation of many of these genes between *E. coli* and other bacteria including *M. tuberculosis* suggests that these feedback mechanisms are also conserved (Arnvig *et al.*, 2011).

In parallel, the transcription of ribosomal RNA (rRNA), is likewise heavily regulated by the 5' leader, which is removed by nucleolytic cleavage from the nascent RNA to generate the mature transcript (Deutscher, 2009). Studies in *E. coli* have elucidated how binding of the antitermination factor NusB and the r-protein S10 to the rRNA leader nucleates a conformational change in the transcription elongation complex that leads to an increase in elongation rate and processivity, i.e. antitermination e.g. (Greive *et al.*, 2005). Similar mechanisms are likely to occur in *M. tuberculosis*, although a specific role for the NusB/E heterodimer has not been demonstrated. However, *M. tuberculosis* has contributed to the antitermination story via its NusA protein and one of the earliest investigations on mycobacterial regulatory RNA. The *M. tuberculosis* NusA lacks the C-terminal domain that masks part of the RNA-binding domain in its *E. coli* counterpart (Gopal *et al.*, 2001). This in

turn facilitated the identification of a highly specific interaction between the KH domains of NusA and the antitermination site of the *M. tuberculosis* rRNA leader, an interaction that was also shown to affect RNAP processivity (Arnvig *et al.*, 2004, Beuth *et al.*, 2005). It remains to be seen how NusA in this context changes from a pausing/termination factor to an antitermination factor.

Riboswitches, leaders sensing metabolites

The term 'riboswitch' refers to a subset of RNA switches that sense changes in the concentration of metabolites or ions without the aid of accessory proteins, and these currently make up the largest class of RNA switches (Serganov & Nudler, 2013, Sherwood & Henkin, 2016, McCown *et al.*, 2017). Riboswitch aptamers sense and interact with a variety of small molecules including nucleotides, amino acids and enzyme co-factors with high affinity and specificity, making them potentially ideal drug targets (Sherwood & Henkin, 2016, Dersch *et al.*, 2017). In many cases these ligands are synthesized by enzymes encoded by the genes that are controlled by the riboswitches itself, thereby implementing feedback regulation (Nudler & Mironov, 2004). The expression platform executes the regulatory output triggered by the presence or absence of ligand binding. If ligand binding leads to reduced expression, the switch is classified as an 'Off' switch (Fig. 1); conversely, if ligand binding leads to increased expression, the switch is classified as an 'On' switch.

Similar to other leader-based regulators, many riboswitches function by blocking/unblocking ribosome entry. Whether the default (i.e. ligand free) conformation is 'On' or 'Off' depends on the individual switch, but in all cases ligand binding to the aptamer domain induces the alternative conformation within the expression platform.

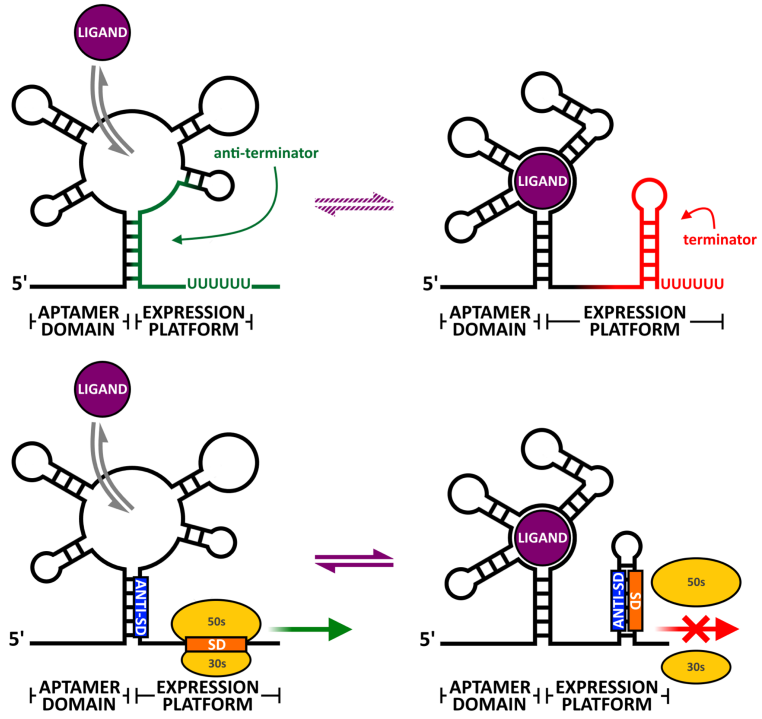


Fig. 1: Riboswitch architecture. Top panel illustrates how a ligand induces transcriptional termination in a transcriptionally controlled 'Off' switch; panel below illustrates a translational 'Off' switch.

Other riboswitches are based on intrinsic (i.e. factor-independent) termination/antitermination, which sets them apart from other leader-based modes of regulation as access to the SD is not involved. Moreover, while the two conformers of a translationally regulated riboswitch may exist in a dynamic equilibrium a transcriptionally regulated riboswitch cannot, since both transcription termination and readthrough are irreversible events. This adds additional kinetic requirements to transcriptional riboswitches, as the decision between one or the other conformer has to be made after transcribing the aptamer domain and sensing of a cognate ligand, but before reaching the end of the expression platform. This may require the RNAP to pause at specific and functionally critical positions within the riboswitch to allow for correct co-transcriptional folding of the RNA (Steinert *et al.*, 2017).

A curious characteristic of some riboswitches is that conserved aptamer domains, recognising identical ligands, may be associated with different expression platforms in

different species, and in some cases even within the same species. For example, the *B. subtilis* RFN element, which senses Flavin Mononucleotide (FMN) uses SD sequestration within the ypaA riboflavin transporter mRNA, but transcriptional termination in the ribDEAHT mRNA, encoding a series of FMN biosynthetic enzymes (Winkler *et al.*, 2002). It is still unclear exactly where and when a translational expression platform is more or less advantageous than a transcriptional expression platform. While the latter requires ongoing RNA synthesis, the former could in theory regulate the translation of extant transcripts provided that these are relatively stable.

Riboswitches in *M. tuberculosis*

The vitamin B_{12}-sensing riboswitch

Currently only a single metabolite-sensing riboswitch has been experimentally validated in *M. tuberculosis*, although several have been predicted by sequence homology and covariance analyses (Warner *et al.*, 2007, Nawrocki *et al.*, 2015). This is the cobalamine or B_{12}-sensing riboswitch upstream of the *metE* gene, encoding a B_{12}-independent methionine synthase. In the presence of B_{12}, expression of the MetE enzyme, which catalyses the conversion of homocysteine to methionine, is repressed, making this an 'Off' switch. In *M. tuberculosis* H37Rv, this reaction is instead carried out by the B_{12}-dependent isozyme, encoded by *metH*. This means that in the absence of B_{12}, *metE* is required, while in the presence of B_{12}, *metH* is required, due to riboswitch-mediated repression of *metE*. However, this gene has been partially disrupted in *M. tuberculosis* CDC1551 with the result that this strain of *M. tuberculosis* has a severe growth defect in the presence of vitamin B_{12} (Warner *et al.*, 2007).

A second B_{12}-sensing rioswitch is located in the 5' leader of the *PPE2-cobQ1-cobU* operon. PPE2 (Rv0256c) belongs to the family of proteins sharing proline-proline-glutamate (PPE) N-terminal motifs that were identified in the *M. tuberculosis* genome sequence, many of which are found on the cell surface of *M. tuberculosis*. PPE2 was originally predicted to be a

277 vitamin B_{12} transporter (Rodionov *et al.*, 2003, Vitreschak *et al.*, 2003). However, a more
278 recent study demonstrated that Rv1819c, an ABC transporter is the 'sole corrinoid
279 transporter' responsible for vitamin B_{12} uptake in *M. tuberculosis* under standard *in vitro*
280 growth conditions (Gopinath *et al.*, 2013). Expression of *Rv1819c* is not controlled by a B_{12}
281 riboswitch, and the exact function of PPE2 remains obscure, although the presence of the
282 riboswitch and the *cobQ1-cobU* genes does suggest a role in B_{12} uptake/metabolism.

283 ***A Cyclic-di-AMP sensing riboswitch regulates rpfA expression***

284 *M. tuberculosis* encodes five so-called resuscitation promoting factors (RpfA-E). These are
285 cell wall remodelling enzymes critical for the transition between dormancy and resuscitation
286 (Chao & Rubin, 2010, Kana & Mizrahi, 2010, Mukamolova *et al.*, 2010, Turapov *et al.*, 2014).
287 Precise control of Rpf expression is vital as these enzymes are potentially lethal for *M.*
288 *tuberculosis* itself, and multiple, at times shared signals converge in the control of *rpf*
289 transcription (Fig. 2). The rpfA 5' leader is 272 nucleotides in length and harbours a
290 homologue of the ydaO aptamer domain (Block *et al.*, 2010, Arnvig & Young, 2012).
291 Identified almost a decade before its cognate ligand, cyclic di-AMP (c-di-AMP), the ydaO
292 riboswitch regulates genes associated with cell wall metabolism and osmotic stress in a wide
293 range of bacteria (Barrick *et al.*, 2004, Nelson *et al.*, 2013). The prolonged pursuit for the
294 correct ligand illustrates the difficulty of identifying some riboswitch ligands, even after an
295 element has been characterised. A similar element has been identified in the 5' leader of
296 *Streptomyces coelicolor* rpfA mRNA, where it has been shown to control expression of RpfA
297 in a c-di-AMP-dependent manner (St-Onge *et al.*, 2015, St-Onge & Elliot, 2017). Due to the
298 close relationship between *S. coelicolor* and *M. tuberculosis*, we expect the *M. tuberculosis*
299 ydaO homologue may also respond to c-di-AMP. Curiously, unlike the *B. subtilis* element,
300 there are no apparent intrinsic terminators, i.e. a stable stem-loop followed by a poly-U tail,
301 associated with neither the *Streptomyces* nor the *M. tuberculosis* riboswitch, suggesting a
302 different expression platform (Nelson *et al.*, 2013, St-Onge & Elliot, 2017) (J. Green and G.
303 Mukamolova personal communication).

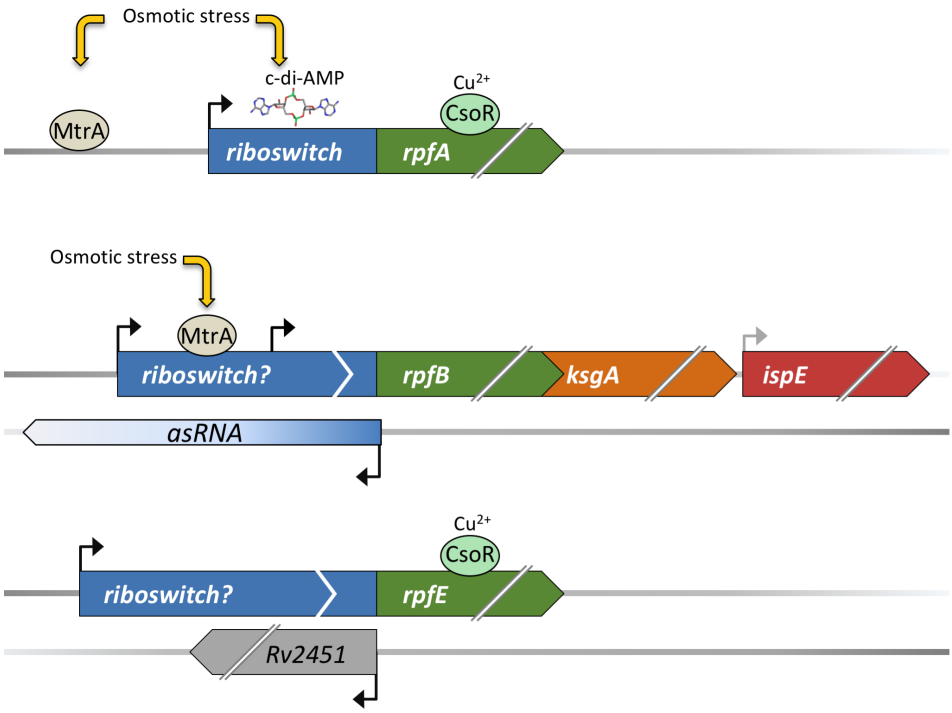


Fig. 2: Control of *rpf* expression in *M. tuberculosis*. The figure illustrates how different, sometimes shared, transcriptional regulators contribute to *rpf* regulation in addition to long 5' leaders, which in the case of *rpfA* harbours a riboswitch with a known ligand (c-d-AMP), in *rpfB* a riboswitch candidate, with unknown ligand and in *rpfE*, a so far entirely uncharacterised element.

It remains to be seen how this element affects *M. tuberculosis* pathogenesis, but adaptation to changing osmolarity does play an important role in *M. tuberculosis*'s lifestyle as well as in phenotypic drug tolerance (Larrouy-Maumus *et al.*, 2016). Moreover, while the rpfA CDS is highly polymorphic in *M. bovis* (Amadio *et al.*, 2005), the ydao element is 100% conserved between *M. tuberculosis* and *Mycobacterium bovis*, suggesting an important role for this riboswitch.

A novel riboswitch candidate regulating expression of rpfB

Remarkably, three of the five rpf mRNAs (encoding RpfA, B and E), have extensive 5' leaders of more than 100 nucleotides in length (Arnvig *et al.*, 2011, Cortes *et al.*, 2013), suggestive of post-transcriptional regulation; the Rpfs encoded by the same three genes are critical players for Rpf-mediated phenomena such as resuscitation of dormant mycobacteria, growth on solid medium and resistance to detergents (Kana & Mizrahi, 2010).

An RNA switch without a known ligand may be considered a riboswitch candidate (Meyer *et al.*, 2011). Similar to rpfA, the 176-nucleotide 5' leader of the rpfB mRNA harbours an RNA switch (or riboswitch candidate), and like rpfA, identification of the rpfB element precedes identification of its ligand. Unlike ydaO however, the rpfB switch has a recognisable intrinsic terminator structure, and also unlike ydaO, the rpfB switch appears to be restricted to a small subset of pathogenic mycobacteria (Schwenk *et al.*, 2018). By extensive genetic and biochemical analysis, this switch has been shown to control *rpfB* transcription via an intrinsic terminator located immediately upstream of the TTG start codon, which was experimentally re-annotated in the same study. The rpfB switch regulates a *tri-cistronic* operon, which also encodes the methyltransferase KsgA, crucial for ribosome biogenesis and IspE, essential for early steps in *M. tuberculosis* cell wall synthesis (Connolly *et al*., 2008, Schwenk *et al.*, 2018).

This arrangement provides an intriguing, regulatory link between riboswitch co-ordinated resuscitation from dormancy, ribosome maturation and cell wall synthesis. Moreover, as the operon represents two classical drug targets, i.e. cell wall synthesis and ribosome function under one regulatory roof, it is tempting to speculate that this riboswitch candidate may represent a new target for anti-tuberculosis drug development. Identification of the cognate ligand will undoubtedly provide novel insights into coordinated regulation of macromolecular synthesis as well as post-transcriptional regulation of gene expression in *M. tuberculosis*.

A potential RNA switch regulating rpfE

Little is known about the regulation of *rpfE* expression other than it is induced by chloride, and it is not yet clear if this effect is transcriptional or post-transcriptional (Tan *et al.*, 2013). TSS mapping indicates that the rpfE 5' leader is at least 251 nucleotides in length, and overlaps the divergently transcribed Rv2451 (of unknown function). Similar to the rpfB leader, the rpfE leader harbours the potential to form a stem-loop followed by a poly-U tail close to the TIR. However, the rpfE poly-U tail is short with only three uridine residues, which

may be insufficient to confer intrinsic termination without the support of additional factors/ligands.

Across bacterial species it is clear that only a fraction of riboswitches has been identified to date, and rare (i.e. not broadly conserved) riboswitches are unlikely to be identified by genome alignments. Novel, more experimental approaches are required to tackle this conservation bias. One such approach is Term-seq, which provides a genome-wide display of RNA 3' ends facilitating the identification of conditional terminators and potential novel riboswitches (Dar *et al.*, 2016). Finally, it is worth mentioning in this context that transcriptionally terminated riboswitches can act in *trans* as sRNAs, thus blurring the boundaries *cis*-regulatory elements and *trans*-acting RNA (Loh *et al.*, 2009).

Trans-acting RNAs

In contrast to *cis*-regulatory elements, asRNAs and sRNAs are not part of the transcript they regulate and may therefore be considered '*trans*-acting' (Lease & Belfort, 2000, Loh *et al.*, 2009). This class of transcripts include a small number of protein binding RNAs, and a very large number of basepairing RNAs, which again are divided into *cis*- or *trans*-encoded. Due to space constraints we will focus on basepairing RNAs in this review.

***Cis*-encoded RNAs** are transcripts encoded opposite their target mRNAs (i.e.'true' asRNAs). These transcripts have perfect complementarity to their mRNA targets, suggesting that the resulting hybrids are ideal RNase III substrates. asRNAs can be of varying sizes from <100 nucleotides to several kb, and they are likely to have different modes of action depending on their size and location.

Many, smaller asRNAs are encoded opposite the TIR of their mRNA targets, where they function in a manner similar to *trans*-encoded sRNAs, by blocking ribosome entry and translation. An important class of such small asRNAs are those associated with type I toxin-antitoxin (TA) systems (Brantl & Jahn, 2015). Curiously however, while there is an

abundance of Type II/protein based TA systems in *M. tuberculosis*, so far no Type I systems have been identified.

Longer asRNAs can be several hundred nucleotides long and in a few cases even several kb (Arnvig *et al.*, 2011, Lasa *et al.*, 2011, Sesto *et al.*, 2013). An example is the asRNA covering *Rv2817-2816c*, encoding Cas1 and Cas2, respectively in the *M. tuberculosis* CRISPR locus, and while this transcript is relatively abundant, there is very little expression of the coding strand under standard *in vitro* growth conditions, suggesting an inverse correlation in abundance between sense and antisense (Arnvig *et al.*, 2011). The function of these asRNAs is still debated, but pervasive antisense transcription may suppress sense transcriptional noise via transcriptional (RNAP) interferences and/or RNase III mediated cleavage of hybridised sense-antisense transcripts (Lasa *et al.*, 2011).

The 5' leaders of divergently transcribed genes or 3' UTRs of convergently transcribed genes, can also act as asRNA on mRNAs transcribed from the opposing strand, once more blurring the boundaries between cis-regulatory elements and *trans*-acting RNA. This phenomenon was first observed in *L. monocytogenes* (Toledo-Arana *et al.*, 2009) where it has since led to the to the 'Excludon' concept, coined by Pascale Cossart's group. The excludon specifically refers to 'an unusually long asRNA that spans divergent genes or operons with related or opposing functions (Sesto *et al.*, 2013).

In *M. tuberculosis*, converging 3' UTRs make a significant contribution to the overall antisense transcriptome, and these show a striking enrichment of genes associated with cell wall functions (Arnvig *et al.*, 2011, Cortes *et al.*, 2013). Future studies on gene function and expression should reveal if an excludon mechanism is employed in *M. tuberculosis*.

Finally, some *cis*-encoded sRNAs also have the potential to act as *trans*-encoded sRNAs on mRNA targets with similar sequences as the primary targets e.g. (Arnvig & Young, 2009, Jager *et al.*, 2012).

***Trans*-encoded sRNAs** are encoded in different genomic locations to their targets. The majority of these transcripts are induced by stress and therefore often associated with pathogen adaptation to hostile host environments. In the early days of sRNA identification, searches for sRNAs focused on intergenic regions, and hence this class of regulators were perceived to originate primarily from disinct promoters within these regions, e.g. (Argaman *et al.*, 2001, Wassarman *et al.*, 2001, Arnvig & Young, 2009, Dichiara *et al.*, 2010). However, with the accumulation of data from RNA-seq based methods, it has become evident that many sRNAs are in fact derived from mRNAs. As already mentioned, transcriptionally attenuated leaders can act as sRNAs (Loh *et al.*, 2009); and mRNA 3' UTRs are avid sRNA generators either from processing or from internal promoters (Chao *et al.*, 2012, Chao *et al.*, 2017).

Mode of action

Unlike the interaction between *cis*-encoded (as)RNAs and their targets, the interaction between *trans*-encoded sRNAs and their targets proceeds via limited basepairing apart from a short 'seed sequence', which means that in many cases, *trans*-encoded sRNAs depend on an RNA chaperone to facilitate the interaction with their targets (Vogel & Luisi, 2011). In addtion to the seed sequence, most sRNAs contain another characteristic feature, which is an intrinsic terminator critical for the interaction with the RNA chaperone Hfq (Otaka *et al.*, 2011, Morita *et al.*, 2017). The limited complementarity also means that prediction of targets can be challenging, and several algorithms have been developed to facilitate this, e.g. TargetRNA2, (Kery *et al.*, 2014) and CopraRNA (Wright *et al.*, 2013). Moreover, a number of experimental approaches have been developed, e.g. RIL-seq, which exploits the proximity of sRNAs to mRNA targets on Hfq (Melamed *et al.*, 2016) or MAPS (pull-downs with MS2-tagged sRNAs), which does not require a protein (Lalaouna *et al.*, 2017). Both predictive and experimental approaches require further validation, in particular in an organism such as *M. tuberculosis*, where little remains known about sRNA targets. Individual sRNAs can both

repress and increase expression of genes in their regulons, depending on the location of the target region.

Repressing interactions

The most commonly known mode of action for *trans*-encoded sRNAs is repression of translation by blocking the TIR, often followed by mRNA degradation (Fig. 3), reviewed in (Wagner & Romby, 2015).

If the TIR is located early within a multi-cistronic operon, this block may also lead to Rho-dependent termination of transcription further downstream (i.e. polarity), (Bossi *et al.*, 2012). The interaction can also take place downstream of the TIR, several codons into the coding region of the mRNA (Pfeiffer *et al.*, 2009). This may be a means of regulating the many leaderless transcripts in *M. tuberculosis* (Cortes *et al.*, 2013).

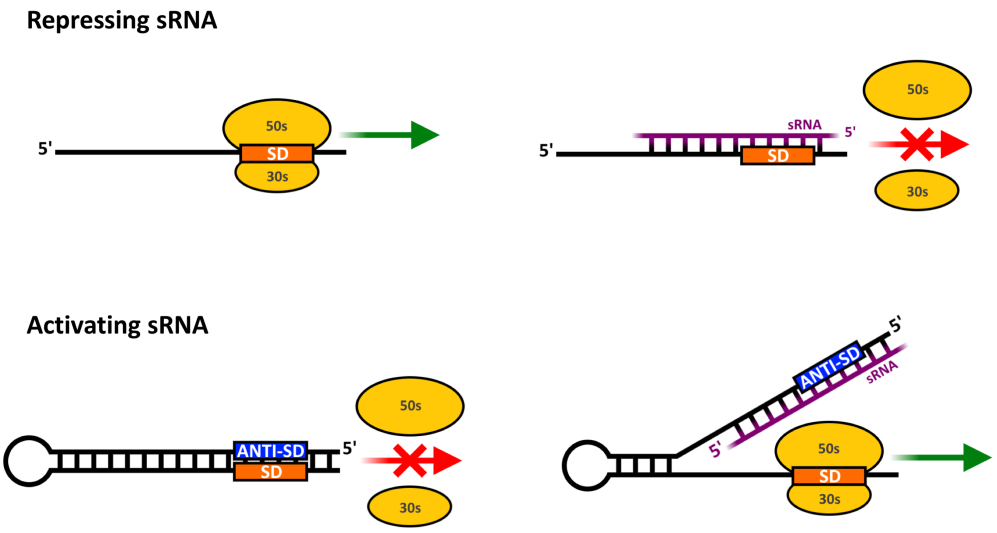


Fig. 3: Basic sRNA modes of action. Top half illustrates how an sRNA (*cis*- or *trans*-encoded) can block ribosome entry and translation. Bottom panel illustrates how an sRNA can activate translation by an anti-antisense mechanism; in this situation the mRNA leader itself blocks translation, by masking the TIR, but an sRNA can interact with the leader to unmask the TIR.

Activating interactions

sRNA-mRNA interaction can also lead to increased translation either by direct stabilisation of the mRNA, by unmasking of the TIR and/or by interfering with Rho-dependent termination, reviewed in (Papenfort & Vanderpool, 2015).

An example of direct stabilisation has been observed in *Salmonella*, where the RydC sRNA blocks an RNase E cleavage site in the cfa1 mRNA. This interaction leads to stabilisation of the mRNA even in the absence of translation (Frohlich *et al.*, 2013).

A somewhat more sophisticated means of activation involves a so-called 'anti-antisense' mechanism (Majdalani *et al.*, 1998). In this situation, the leader of the target mRNA contains an auto-inhibitory secondary structure that masks the TIR, and which can be unmasked sRNA binding. A well-characterised example is the *E. coli* rpoS mRNA, which encodes the stationary phase sigma factor, Sigma38 (Battesti *et al.*, 2011). The rpoS mRNA harbours a 567-nucleotide 5' leader, which blocks its own TIR (Majdalani *et al.*, 1998, Peng *et al.*, 2014). Upon binding of one of three sRNAs (DsrA, RprA, ArcZ) to the inhibitory region, the SD sequence and start codon are unmasked via the anti-antisense mechanism to permit translation (Battesti *et al.*, 2011).

Recently, it was shown that the same three sRNAs in addition to unmasking the rpoS mRNA TIR, could also inhibit Rho-dependent termination of *rpoS* transcription in *E. coli* by masking one or more Rho binding sites in the rpoS leader, thus making the sRNA activating effect two-pronged. The authors argued that this novel sRNA-regulated antitermination is likely to be widespread in long leaders (Sedlyarova *et al.*, 2016).

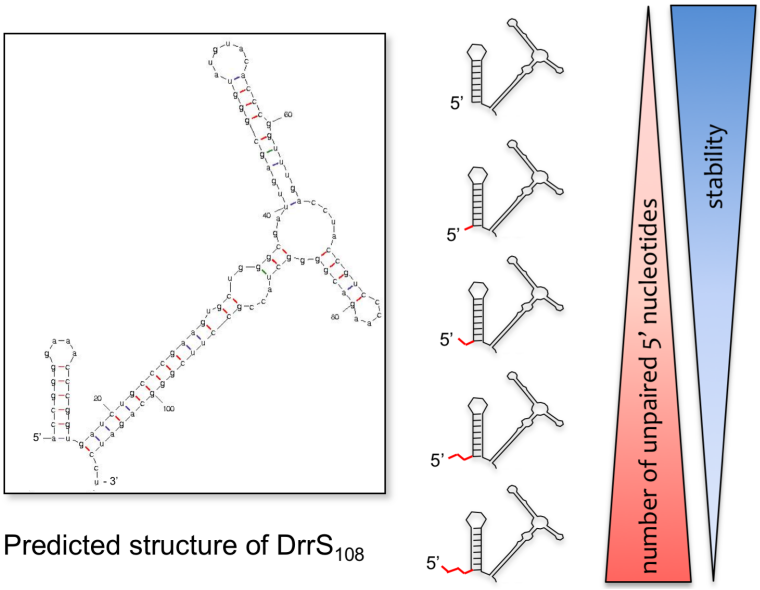
To summarise, sRNAs can both repress or promote translation initiation, and repress or promote Rho-dependent termination of transcription. Moreover, the effect of an sRNA can be greatly enhanced if the mRNA target encodes a regulator such as a sigma or a transcription factor. An overview of different regulatory networks, and their evolution can be found in (Beisel & Storz, 2010, Peer & Margalit, 2014).

M. tuberculosis sRNAs

In spite of several *M. tuberculosis* sRNAs being identified and mapped, and their expression patterns investigated e.g. (Arnvig & Young, 2009, Dichiara *et al.*, 2010, Arnvig *et al.*, 2011, Miotto *et al.*, 2012), only few, including MTS2823, ncRv12659, DrrS and Mcr7, have been functionally characterised in any greater detail (Arnvig *et al.*, 2011, Houghton *et al.*, 2013, Solans *et al.*, 2014, Moores *et al.*, 2017). Like their counterparts in model organisms, *M. tuberculosis* sRNAs are often stress induced and some are highly abundant during infection. The evolutionary conservation of *M. tuberculosis* small RNAs is subject to considerable variation. Some sRNAs, such as ncRv12659 are specific for a subset of *M. tuberculosis* strains (Houghton *et al.*, 2013), some are found throughout species of the *M. tuberculosis* complex, some a little further afield including non-tuberculous, pathogenic mycobacteria and a few *M. tuberculosis* sRNAs are conserved in *Mycobacterium smegmatis* and other Actinomycetes e.g. (Arnvig & Young, 2009, Dichiara *et al.*, 2010, Haning *et al.*, 2014). Many *M. tuberculosis* sRNAs are highly structured, in part due to the high GC content of the bacterium. Furthermore, by comparing results from 5' and 3' RACE, RNA-seq, northern blotting and RNA structure prediction, it is evident that many *M. tuberculosis* sRNAs do not contain conventional intrinsic terminator structures. For some time this lack of conventional terminators was attributed to the presence of so-called I-shaped terminators, i.e. stem-loop structures without a poly-U tail (Mitra *et al.*, 2008). However, more recently, RNA-seq and *in vitro* transcription experiments using *M. bovis* RNA polymerase, have demonstrated that in most cases this type of structure is not sufficient for termination of transcription *in vivo* or *in vitro* (Arnvig *et al.*, 2011, Czyz *et al.*, 2014). This in turn suggests that many sRNA 3' termini may be generated by processing in *M. tuberculosis*, setting them apart from the well-known Hfq-dependent sRNAs that require a poly-U tail to function (Otaka *et al.*, 2011). The predicted processing also suggests that some sRNAs may exist as different isoforms, as is the case for the DosR regulated sRNA, DrrS (Moores *et al.*, 2017).

493 The 108-nucleotide DrrS was first identified by RNA-seq and shown to accumulate to high
494 levels during chronic mouse infection (Arnvig *et al.*, 2011). Recently it was shown that DrrS
495 expression is induced by DosR, but it is a combination of DosR-dependent induction and the
496 unrivalled stability of DrrS that determines the overall levels (Moores *et al.*, 2017).

497 DrrS has a half-life in the order of several hours due to a stable stem-loop structure at its 5'
498 end. The addition of two or more unpaired nucleotides 5' of this stem-loop, reduces stability
499 significantly, suggesting the involvement of a mycobacterial RppH homologue (Fig. 4).
500 Moreover, this structure increases expression of a *lacZ* reporter when added to the 5' end of
501 its mRNA, suggesting that it represents a general stabilising feature (Moores *et al.*, 2017). In
502 addition to elucidating how RNA stabiity may be modulated in *M. tuberculosis*, DrrS provides
503 insights into sRNA processing. DrrS is transcribed as a longer (>300 nucleotide) precursor,
504 ($DrrS^{+}$) that is rapidly (in *M. tuberculosis* terms) processed to the shorter, stable 108-
505 nucleotide sRNA ($DrrS_{108}$).



Predicted structure of $DrrS_{108}$

506

507 Fig. 4: Stability of DrrS. Large image shows the predicted structure of $DrrS_{108}$, while the schematic
508 representation illustrates how the number of unpaired nucleotides 5' are inversely correlated to
509 transcript stability.

510

While $DrrS^{+}$ levels peak in early stationary phase, $DrrS_{108}$ accumulates continuously for at least three weeks into stationary phase (Moores *et al.*, 2017). The substantial difference in size and maximum expression between DrrS and $DrrS^{+}$ implies that the longer isoform may play a different role than the the shorter isoform. Apart from shedding light on RNA processing and stability, the DrrS example also highlights the importance of thoroughly characterising multiple aspects of an sRNA before defining its regulon. The application of Term-seq to define *M. tuberculosis* 3' ends on a global scale (Dar *et al.*, 2016), is likely to be hugely informative at this stage.

The best characterised *M. tuberculosis* sRNA in terms of biological role is Mcr7. This sRNA was first identified as a 350-400 nucleotide transcript by cloning and sequencing of *M. bovis* BCG cDNA, and in the same study predicted by sequence homology to be conserved throughout the *M. tuberculosis* complex (Dichiara *et al.*, 2010). RNA-seq later confirmed high expression in *M. tuberculosis* H37Rv (Arnvig *et al.*, 2011). Mcr7 is encoded downstream of Rv2395 and according to TSS mapping, a single promoter drives transcription in the region downstream of Rv2395 and into PE_PGRS41 (Cortes *et al.*, 2013), suggesting that Mcr7 is (part of) the 5' leader of the latter. However, there is more to this locus than a PE_PGRS protein with a long 5' leader.

In 2011 David Russell's group reported the characterisation of the PhoPR-dependent *aprABC* (Acid and Phagosome Regulated) locus encoding the conserved hypothetical proteins, AprA and AprB, as well as PE_PGRS41 (AprC) (Abramovitch *et al.*, 2011). The aprA coding region lies entirely within the boundaries of Mcr7 (62 basepairs downstream of the annotated TSS), with AprB and AprC encoded downstream of Mcr7 (Fig.5).

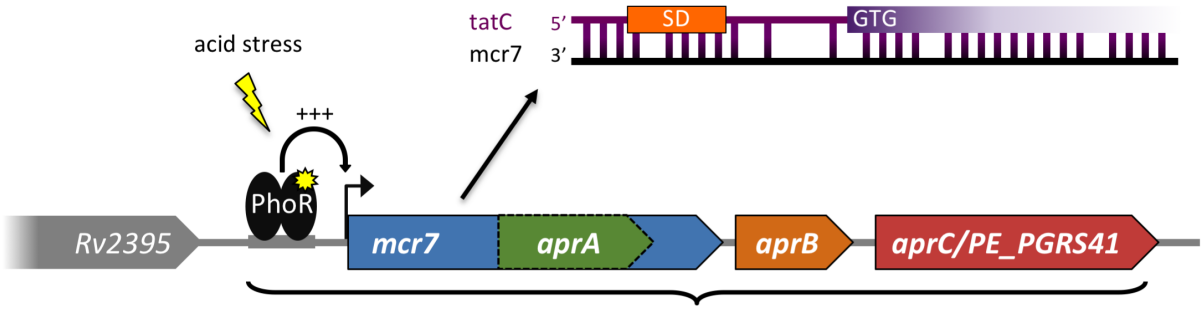


Fig. 5: The *mcr7/aprABC* locus in *M. tuberculosis*. The figure illustrates the elements associated with the PhoP/R regulated operon with the ncRNA Mcr7, which contains an open reading frame encoding the acid inducible AprA, and the proposed interaction between Mcr7 and the tatC mRNA.

The proteins have not yet been experimentally validated in *M. tuberculosis*, but *aprA* does have a likely ribosome binding site upstream of its start codon. Moreover, the recombinant protein has been expressed and purified in *E. coli*, suggesting this is a *bona fide*, stable protein (Abramovitch *et al.*, 2011). Was this then an indication that Mcr7 had been wrongly annotated as an sRNA?

In 2014 the *mcr7*/*aprABC* promoter was identified as one of the major targets of PhoR (Solans *et al.*, 2014). Based on this finding and the assumption that Mcr7 was a post-transcriptional regulator of gene expression, the authors used *in silico* prediction to identify putative targets of Mcr7, one of which was the tatC mRNA. More specifically positions -16 to +19 relative to the annotated GTG start of the tatC mRNA are targeted by the central portion (nucleotides 119 to 151) of Mcr7, i.e. well within the coding region of *aprA*. The prediction suggests that PhoP/R dependent expression of Mcr7 represses the translation of TatC resulting in reduced secretion of TAT-dependent proteins, which was supported by proteomics on culture supernatants on *M. tuberculosis* wildtype and *phoP* mutant. This study therefore strongly supports the notion of Mcr7 being an sRNA that represses translation of TatC, thereby changing the secretome and modifying the host-pathogen interface (Solans *et al.*, 2014).

So, although AprA has not yet been identified in *M. tuberculosis* and a direct interaction between Mcr7 and tatC mRNA has not been experimentally validated, it appears that this sRNA is a prime candidate for a dual function sRNA in *M. tuberculosis*. As there are no additional TSS in this operon, it also suggests that the 5' end of Mcr7 may regulate *aprA* expression via an as yet uncharacterised post-transcriptional mechanism. If all these elements really represents their annotated functions, this operon represents a complex arrangement of a 5' leader that acts as a *trans*-encoded dual function sRNA.

Concluding remarks

Pathogen survival depends on constant monitoring of, and adaptation to, a range of host environments, an adaptation that sometimes requires rapid and drastic changes in gene expression. This is most efficiently achieved by multi-pronged approaches combining several layers of control, such as transcriptional, post-transcriptional and post-translational regulation. A comprehensive insight into all of these mechanisms is necessary to fully understand how a pathogen interacts with its host, and more importantly, how we might exploit this to our own advantage. Whether the aim is drug discovery or vaccine development, a thorough understanding of the basic molecular mechanisms of the pathogen in question is fundamental.

In this review we have illustrated (i) how riboregulators work, (ii) argued why riboregulation should be considered by the *M. tuberculosis* community, and (iii) why *M. tuberculosis* should be considered by the RNA community. Although some general rules may apply, riboregulation is still full of surprises, and *M. tuberculosis* is different; with its high GC content (>65%), abundance of leader-less mRNA, distinct complement of RNases and lack of Hfq and ProQ chaperones. In summary, *M. tuberculosis* has the potential to greatly advance our knowledge of RNA based control of gene expression.

Funding

This work was supported by The Medical Research Council New Investigator Award for KBA [grant number MR/L018519/1].

Acknowledgements

We thank Galina Mukamolova and Finn Werner for critical reading of the manuscript.

References

Abramovitch RB, Rohde KH, Hsu FF & Russell DG (2011) aprABC: a Mycobacterium tuberculosis complex-specific locus that modulates pH-driven adaptation to the macrophage phagosome. *Molecular microbiology* **80**: 678-694.

Amadio A, Romano MI, Bigi F, Etchechoury I, Kubica T, Niemann S, Cataldi A & Caimi K (2005) Identification and characterization of genomic variations between Mycobacterium bovis and M. tuberculosis H37Rv. *Journal of clinical microbiology* **43**: 2481-2484.

Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H & Altuvia S (2001) Novel small RNA-encoding genes in the intergenic regions of Escherichia coli. *Curr Biol* **11**: 941-950.

Arnvig KB & Young DB (2009) Identification of small RNAs in Mycobacterium tuberculosis. *Molecular microbiology* **73**: 397-408.

Arnvig KB & Young DB (2010) Regulation of pathogen metabolism by small RNA. *Drug Discov Today Dis Mech* **7**: e19-e24.

Arnvig KB & Young DB (2012) Regulatory RNAs in Mycobacterium tuberculosis. *RNA biology*.

Arnvig KB, Pennell S, Gopal B & Colston MJ (2004) A high-affinity interaction between NusA and the rrn nut site in Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 8325-8330.

Arnvig KB, Comas I, Thomson NR*, et al.* (2011) Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis. *PLoS Pathog* **7**: e1002342.

Babitzke P, Baker CS & Romeo T (2009) Regulation of translation initiation by RNA binding proteins. *Annual review of microbiology* **63**: 27-44.

Barrick JE, Corbino KA, Winkler WC*, et al.* (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 6421-6426.

Battesti A, Majdalani N & Gottesman S (2011) The RpoS-mediated general stress response in Escherichia coli. *Annual review of microbiology* **65**: 189-213.

Beisel CL & Storz G (2010) Base pairing small RNAs and their roles in global regulatory networks. *FEMS microbiology reviews* **34**: 866-882.

Beuth B, Pennell S, Arnvig KB, Martin SR & Taylor IA (2005) Structure of a Mycobacterium tuberculosis NusA-RNA complex. *EMBO J* **24**: 3576-3587.

Block KF, Hammond MC & Breaker RR (2010) Evidence for widespread gene control function by the ydaO riboswitch candidate. *Journal of bacteriology* **192**: 3983-3989.

Bossi L, Schwartz A, Guillemardet B, Boudvillain M & Figueroa-Bossi N (2012) A role for Rho-dependent polarity in gene regulation by a noncoding small RNA. *Genes & development* **26**: 1864-1873.

Brantl S & Jahn N (2015) sRNAs in bacterial type I and type III toxin-antitoxin systems. *FEMS microbiology reviews* **39**: 413-427.

624 Caballero CJ, Menendez-Gil P, Catalan-Moreno A*, et al.* (2018) The regulon of the RNA
625 chaperone CspA and its auto-regulation in Staphylococcus aureus. *Nucleic acids research*.
626 Chao MC & Rubin EJ (2010) Letting sleeping dos lie: does dormancy play a role in
627 tuberculosis? *Annual review of microbiology* **64**: 293-311.
628 Chao Y, Papenfort K, Reinhardt R, Sharma CM & Vogel J (2012) An atlas of Hfq-bound
629 transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J* **31**:
630 4005-4019.
631 Chao Y, Li L, Girodat D*, et al.* (2017) In Vivo Cleavage Map Illuminates the Central Role of
632 RNase E in Coding and Non-coding RNA Pathways. *Molecular cell* **65**: 39-51.
633 Cortes T, Schubert OT, Rose G, Arnvig KB, Comas I, Aebersold R & Young DB (2013) Genome-
634 wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in
635 Mycobacterium tuberculosis. *Cell Rep* **5**: 1121-1131.
636 Czyz A, Mooney RA, Iaconi A & Landick R (2014) Mycobacterial RNA polymerase requires a
637 U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *mBio* **5**:
638 e00931.
639 Dar D, Shamir M, Mellin JR, Koutero M, Stern-Ginossar N, Cossart P & Sorek R (2016) Term-
640 seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science* **352**:
641 aad9822.
642 Deana A, Celesnik H & Belasco JG (2008) The bacterial enzyme RppH triggers messenger
643 RNA degradation by 5' pyrophosphate removal. *Nature* **451**: 355-358.
644 Dersch P, Khan MA, Muhlen S & Gorke B (2017) Roles of Regulatory RNAs for Antibiotic
645 Resistance in Bacteria and Their Potential Value as Novel Drug Targets. *Frontiers in*
646 *microbiology* **8**: 803.

Deutscher MP (2009) Maturation and degradation of ribosomal RNA in bacteria. *Prog Mol Biol Transl Sci* **85**: 369-391.

Dichiara JM, Contreras-Martinez LM, Livny J, Smith D, McDonough KA & Belfort M (2010) Multiple small RNAs identified in Mycobacterium bovis BCG are also expressed in Mycobacterium tuberculosis and Mycobacterium smegmatis. *Nucleic acids research*.

Durand S, Tomasini A, Braun F, Condon C & Romby P (2015) sRNA and mRNA turnover in Gram-positive bacteria. *FEMS microbiology reviews* **39**: 316-330.

Even S, Pellegrini O, Zig L, Labas V, Vinh J, Brechemmier-Baey D & Putzer H (2005) Ribonucleases J1 and J2: two novel endoribonucleases in B.subtilis with functional homology to E.coli RNase E. *Nucleic acids research* **33**: 2141-2152.

Ferre-D'Amare AR & Winkler WC (2011) The roles of metal ions in regulation by riboswitches. *Met Ions Life Sci* **9**: 141-173.

Frohlich KS, Papenfort K, Fekete A & Vogel J (2013) A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *EMBO J* **32**: 2963-2979.

Giuliodori AM, Di Pietro F, Marzi S, Masquida B, Wagner R, Romby P, Gualerzi CO & Pon CL (2010) The cspA mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. *Molecular cell* **37**: 21-33.

Gopal B, Haire LF, Gamblin SJ, Dodson EJ, Lane AN, Papavinasasundaram KG, Colston MJ & Dodson G (2001) Crystal structure of the transcription elongation/anti-termination factor NusA from Mycobacterium tuberculosis at 1.7 A resolution. *Journal of molecular biology* **314**: 1087-1095.

Gopinath K, Venclovas C, Ioerger TR, Sacchettini JC, McKinney JD, Mizrahi V & Warner DF (2013) A vitamin B12 transporter in Mycobacterium tuberculosis. *Open biology* **3**: 120175.

Greive SJ, Lins AF & von Hippel PH (2005) Assembly of an RNA-protein complex. Binding of NusB and NusE (S10) proteins to boxA RNA nucleates the formation of the antitermination complex involved in controlling rRNA transcription in Escherichia coli. *The Journal of biological chemistry* **280**: 36397-36408.

Grundy FJ, Winkler WC & Henkin TM (2002) tRNA-mediated transcription antitermination in vitro: codon-anticodon pairing independent of the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 11121-11126.

Haning K, Cho SH & Contreras LM (2014) Small RNAs in mycobacteria: an unfolding story. *Front Cell Infect Microbiol* **4**: 96.

Holmqvist E & Wagner EGH (2017) Impact of bacterial sRNAs in stress responses. *Biochemical Society transactions* **45**: 1203-1212.

Houghton J, Cortes T, Schubert OT, Rose G, Rodgers A, De Ste Croix M, Aebersold R, Young DB & Arnvig KB (2013) A small RNA encoded in the Rv2660c locus of Mycobacterium tuberculosis is induced during starvation and infection. *PloS one* **8**: e80047.

Ingolia NT, Ghaemmaghami S, Newman JR & Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**: 218-223.

Ishihama A (2010) Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. *FEMS microbiology reviews* **34**: 628-645.

Jager D, Pernitzsch SR, Richter AS, Backofen R, Sharma CM & Schmitz RA (2012) An archaeal sRNA targeting cis- and trans-encoded mRNAs via two distinct domains. *Nucleic acids research* **40**: 10964-10979.

692 Jiang W, Hou Y & Inouye M (1997) CspA, the major cold-shock protein of Escherichia coli, is
693 an RNA chaperone. *The Journal of biological chemistry* **272**: 196-202.
694 Jousselin A, Metzinger L & Felden B (2009) On the facultative requirement of the bacterial
695 RNA chaperone, Hfq. *Trends in microbiology* **17**: 399-405.
696 Kana BD & Mizrahi V (2010) Resuscitation-promoting factors as lytic enzymes for bacterial
697 growth and signaling. *FEMS immunology and medical microbiology* **58**: 39-50.
698 Kery MB, Feldman M, Livny J & Tjaden B (2014) TargetRNA2: identifying targets of small
699 regulatory RNAs in bacteria. *Nucleic acids research* **42**: W124-129.
700 Kortmann J & Narberhaus F (2012) Bacterial RNA thermometers: molecular zippers and
701 switches. *Nature reviews Microbiology* **10**: 255-265.
702 Koslover DJ, Callaghan AJ, Marcaida MJ, Garman EF, Martick M, Scott WG & Luisi BF (2008)
703 The crystal structure of the Escherichia coli RNase E apoprotein and a mechanism for RNA
704 degradation. *Structure* **16**: 1238-1244.
705 Lalaouna D, Prevost K, Eyraud A & Masse E (2017) Identification of unknown RNA partners
706 using MAPS. *Methods* **117**: 28-34.
707 Larrouy-Maumus G, Marino LB, Madduri AV, Ragan TJ, Hunt DM, Bassano L, Gutierrez MG,
708 Moody DB, Pavan FR & de Carvalho LP (2016) Cell-Envelope Remodeling as a Determinant of
709 Phenotypic Antibacterial Tolerance in Mycobacterium tuberculosis. *ACS Infect Dis* **2**: 352-
710 360.
711 Lasa I, Toledo-Arana A, Dobin A*, et al.* (2011) Genome-wide antisense transcription drives
712 mRNA processing in bacteria. *Proceedings of the National Academy of Sciences of the United*
713 *States of America*.

714 Lease RA & Belfort M (2000) A trans-acting RNA as a control switch in Escherichia coli: DsrA
715 modulates function by forming alternative structures. *Proceedings of the National Academy*
716 *of Sciences of the United States of America* **97**: 9919-9924.
717 Lee HJ & Gottesman S (2016) sRNA roles in regulating transcriptional regulators: Lrp and
718 SoxS regulation by sRNAs. *Nucleic acids research* **44**: 6907-6923.
719 Levine E & Hwa T (2008) Small RNAs establish gene expression thresholds. *Current opinion in*
720 *microbiology* **11**: 574-579.
721 Li de la Sierra-Gallay I, Zig L, Jamalli A & Putzer H (2008) Structural insights into the dual
722 activity of RNase J. *Nat Struct Mol Biol* **15**: 206-212.
723 Loh E, Dussurget O, Gripenland J, Vaitkevicius K, Tiensuu T, Mandin P, Repoila F, Buchrieser
724 C, Cossart P & Johansson J (2009) A trans-acting riboswitch controls expression of the
725 virulence regulator PrfA in Listeria monocytogenes. *Cell* **139**: 770-779.
726 Mackie GA (1998) Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* **395**: 720-
727 723.
728 Majdalani N, Cunning C, Sledjeski D, Elliott T & Gottesman S (1998) DsrA RNA regulates
729 translation of RpoS message by an anti-antisense mechanism, independent of its action as
730 an antisilencer of transcription. *Proceedings of the National Academy of Sciences of the*
731 *United States of America* **95**: 12462-12467.
732 Mathy N, Hebert A, Mervelet P, Benard L, Dorleans A, Li de la Sierra-Gallay I, Noirot P,
733 Putzer H & Condon C (2010) Bacillus subtilis ribonucleases J1 and J2 form a complex with
734 altered enzyme behaviour. *Molecular microbiology* **75**: 489-498.
735 McCown PJ, Corbino KA, Stav S, Sherlock ME & Breaker RR (2017) Riboswitch Diversity and
736 Distribution. *RNA*.

737 Melamed S, Peer A, Faigenbaum-Romm R, Gatt YE, Reiss N, Bar A, Altuvia Y, Argaman L &
738 Margalit H (2016) Global Mapping of Small RNA-Target Interactions in Bacteria. *Molecular*
739 *cell* **63**: 884-897.
740 Meyer MM, Hammond MC, Salinas Y, Roth A, Sudarsan N & Breaker RR (2011) Challenges of
741 ligand identification for riboswitch candidates. *RNA biology* **8**: 5-10.
742 Miotto P, Forti F, Ambrosi A, Pellin D, Veiga DF, Balazsi G, Gennaro ML, Di Serio C, Ghisotti D
743 & Cirillo DM (2012) Genome-wide discovery of small RNAs in Mycobacterium tuberculosis.
744 *PloS one* **7**: e51950.
745 Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, Perumov DA & Nudler E
746 (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in
747 bacteria. *Cell* **111**: 747-756.
748 Mitra A, Angamuthu K & Nagaraja V (2008) Genome-wide analysis of the intrinsic
749 terminators of transcription across the genus Mycobacterium. *Tuberculosis (Edinb)* **88**: 566-
750 575.
751 Moores A, Riesco AB, Schwenk S & Arnvig KB (2017) Expression, maturation and turnover of
752 DrrS, an unusually stable, DosR regulated small RNA in Mycobacterium tuberculosis. *PloS*
753 *one* **12**: e0174079.
754 Morita T, Nishino R & Aiba H (2017) Role of the terminator hairpin in the biogenesis of
755 functional Hfq-binding sRNAs. *RNA* **23**: 1419-1431.
756 Mukamolova GV, Turapov O, Malkin J, Woltmann G & Barer MR (2010) Resuscitation-
757 promoting factors reveal an occult population of tubercle Bacilli in Sputum. *Am J Respir Crit*
758 *Care Med* **181**: 174-180.
759 Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL & Breaker RR (2002) Genetic control by a
760 metabolite binding mRNA. *Chemistry & biology* **9**: 1043.

761 Nawrocki EP, Burge SW, Bateman A*, et al.* (2015) Rfam 12.0: updates to the RNA families
762 database. *Nucleic acids research* **43**: D130-137.
763 Nechooshtan G, Elgrably-Weiss M, Sheaffer A, Westhof E & Altuvia S (2009) A pH-responsive
764 riboregulator. *Genes & development* **23**: 2650-2662.
765 Nelson JW, Sudarsan N, Furukawa K, Weinberg Z, Wang JX & Breaker RR (2013)
766 Riboswitches in eubacteria sense the second messenger c-di-AMP. *Nature chemical biology*.
767 Nudler E & Mironov AS (2004) The riboswitch control of bacterial metabolism. *Trends*
768 *Biochem Sci* **29**: 11-17.
769 Oliva G, Sahr T & Buchrieser C (2015) Small RNAs, 5' UTR elements and RNA-binding
770 proteins in intracellular bacteria: impact on metabolism and virulence. *FEMS microbiology*
771 *reviews* **39**: 331-349.
772 Otaka H, Ishikawa H, Morita T & Aiba H (2011) PolyU tail of rho-independent terminator of
773 bacterial small RNAs is essential for Hfq action. *Proceedings of the National Academy of*
774 *Sciences of the United States of America* **108**: 13059-13064.
775 Papenfort K & Vanderpool CK (2015) Target activation by regulatory RNAs in bacteria. *FEMS*
776 *microbiology reviews* **39**: 362-378.
777 Peer A & Margalit H (2014) Evolutionary patterns of Escherichia coli small RNAs and their
778 regulatory interactions. *RNA* **20**: 994-1003.
779 Peng Y, Soper TJ & Woodson SA (2014) Positional effects of AAN motifs in rpoS regulation by
780 sRNAs and Hfq. *Journal of molecular biology* **426**: 275-285.
781 Pfeiffer V, Papenfort K, Lucchini S, Hinton JC & Vogel J (2009) Coding sequence targeting by
782 MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat*
783 *Struct Mol Biol* **16**: 840-846.

784 Quereda JJ & Cossart P (2017) Regulating Bacterial Virulence with RNA. *Annual review of*
785 *microbiology* **71**: 263-280.
786 Ramage HR, Connolly LE & Cox JS (2009) Comprehensive functional analysis of
787 Mycobacterium tuberculosis toxin-antitoxin systems: implications for pathogenesis, stress
788 responses, and evolution. *PLoS genetics* **5**: e1000767.
789 Rau MH, Bojanovic K, Nielsen AT & Long KS (2015) Differential expression of small RNAs
790 under chemical stress and fed-batch fermentation in E. coli. *BMC genomics* **16**: 1051.
791 Rodionov DA, Vitreschak AG, Mironov AA & Gelfand MS (2003) Comparative genomics of
792 the vitamin B12 metabolism and regulation in prokaryotes. *The Journal of biological*
793 *chemistry* **278**: 41148-41159.
794 Rose G, Cortes T, Comas I, Coscolla M, Gagneux S & Young DB (2013) Mapping of genotype-
795 phenotype diversity amongst clinical isolates of Mycobacterium tuberculosis by sequence-
796 based transcriptional profiling. *Genome Biology and Evolution*.
797 Sala A, Bordes P & Genevaux P (2014) Multiple toxin-antitoxin systems in Mycobacterium
798 tuberculosis. *Toxins (Basel)* **6**: 1002-1020.
799 Schwenk S, Moores A, Nobeli I, McHugh TD & Arnvig KB (2018) Cell-wall synthesis and
800 ribosome maturation are co-regulated by an RNA switch in Mycobacterium tuberculosis
801 *Nucleic acids research* **in press**.
802 Sedlyarova N, Shamovsky I, Bharati BK, Epshtein V, Chen J, Gottesman S, Schroeder R &
803 Nudler E (2016) sRNA-Mediated Control of Transcription Termination in E. coli. *Cell* **167**:
804 111-121 e113.
805 Serganov A & Nudler E (2013) A decade of riboswitches. *Cell* **152**: 17-24.

Sesto N, Wurtzel O, Archambaud C, Sorek R & Cossart P (2013) The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. *Nature reviews Microbiology* **11**: 75-82.

Sharma CM, Hoffmann S, Darfeuille F*, et al.* (2010) The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* **464**: 250-255.

Sherwood AV & Henkin TM (2016) Riboswitch-Mediated Gene Regulation: Novel RNA Architectures Dictate Gene Expression Responses. *Annual review of microbiology* **70**: 361-374.

Shine J & Dalgarno L (1974) The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences of the United States of America* **71**: 1342-1346.

Smirnov A, Schneider C, Hor J & Vogel J (2017) Discovery of new RNA classes and global RNA-binding proteins. *Current opinion in microbiology* **39**: 152-160.

Smirnov A, Forstner KU, Holmqvist E, Otto A, Gunster R, Becher D, Reinhardt R & Vogel J (2016) Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proceedings of the National Academy of Sciences of the United States of America* **113**: 11591-11596.

Solans L, Gonzalo-Asensio J, Sala C, Benjak A, Uplekar S, Rougemont J, Guilhot C, Malaga W, Martin C & Cole ST (2014) The PhoP-dependent ncRNA Mcr7 modulates the TAT secretion system in Mycobacterium tuberculosis. *PLoS Pathog* **10**: e1004183.

St-Onge RJ & Elliot MA (2017) Regulation of a muralytic enzyme-encoding gene by two non-coding RNAs. *RNA biology* 0.

St-Onge RJ, Haiser HJ, Yousef MR, Sherwood E, Tschowri N, Al-Bassam M & Elliot MA (2015) Nucleotide second messenger-mediated regulation of a muralytic enzyme in Streptomyces. *Molecular microbiology* **96**: 779-795.

Steinert H, Sochor F, Wacker A*, et al.* (2017) Pausing guides RNA folding to populate transiently stable RNA structures for riboswitch-based transcription regulation. *Elife* **6**.

Tan S, Sukumar N, Abramovitch RB, Parish T & Russell DG (2013) Mycobacterium tuberculosis responds to chloride and pH as synergistic cues to the immune status of its host cell. *PLoS Pathog* **9**: e1003282.

Taverniti V, Forti F, Ghisotti D & Putzer H (2011) Mycobacterium smegmatis RNase J is a 5'-3' exo-/endoribonuclease and both RNase J and RNase E are involved in ribosomal RNA maturation. *Molecular microbiology* **82**: 1260-1276.

Toledo-Arana A, Dussurget O, Nikitas G*, et al.* (2009) The Listeria transcriptional landscape from saprophytism to virulence. *Nature* **459**: 950-956.

Turapov O, Glenn S, Kana B, Makarov V, Andrew PW & Mukamolova GV (2014) The in vivo environment accelerates generation of Resuscitation-promoting factor-dependent mycobacteria. *AJRCCM*.

Updegrove TB, Shabalina SA & Storz G (2015) How do base-pairing small RNAs evolve? *FEMS microbiology reviews* **39**: 379-391.

Vellanoweth RL & Rabinowitz JC (1992) The influence of ribosome-binding-site elements on translational efficiency in Bacillus subtilis and Escherichia coli in vivo. *Molecular microbiology* **6**: 1105-1114.

Vitreschak AG, Rodionov DA, Mironov AA & Gelfand MS (2003) Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element. *RNA* **9**: 1084-1097.

852 Vogel J & Luisi BF (2011) Hfq and its constellation of RNA. *Nature reviews Microbiology* **9**:
853 578-589.
854 Wagner EG & Romby P (2015) Small RNAs in bacteria and archaea: who they are, what they
855 do, and how they do it. *Adv Genet* **90**: 133-208.
856 Wagner EG, Altuvia S & Romby P (2002) Antisense RNAs in bacteria and their genetic
857 elements. *Adv Genet* **46**: 361-398.
858 Warner DF, Savvi S, Mizrahi V & Dawes SS (2007) A riboswitch regulates expression of the
859 coenzyme B12-independent methionine synthase in Mycobacterium tuberculosis:
860 implications for differential methionine synthase function in strains H37Rv and CDC1551.
861 *Journal of bacteriology* **189**: 3655-3659.
862 Wassarman KM, Repoila F, Rosenow C, Storz G & Gottesman S (2001) Identification of novel
863 small RNAs using comparative genomics and microarrays. *Genes & development* **15**: 1637-
864 1651.
865 Winkler W, Nahvi A & Breaker RR (2002) Thiamine derivatives bind messenger RNAs directly
866 to regulate bacterial gene expression. *Nature* **419**: 952-956.
867 Winkler WC, Cohen-Chalamish S & Breaker RR (2002) An mRNA structure that controls gene
868 expression by binding FMN. *Proceedings of the National Academy of Sciences of the United*
869 *States of America* **99**: 15908-15913.
870 Wright PR, Richter AS, Papenfort K, Mann M, Vogel J, Hess WR, Backofen R & Georg J (2013)
871 Comparative genomics boosts target prediction for bacterial small RNAs. *Proceedings of the*
872 *National Academy of Sciences of the United States of America* **110**: E3487-3496.

873

874