

# Cell-wall synthesis and ribosome maturation are co-regulated by an RNA switch in *Mycobacterium tuberculosis*

Stefan Schwenk<sup>1</sup>, Alexandra Moores<sup>1</sup>, Irene Nobeli<sup>2</sup>, Timothy D. McHugh<sup>3</sup> and Kristine B. Arnvig<sup>1,\*</sup>

<sup>1</sup>Institute for Structural and Molecular Biology, University College London, London WC1E 6BT, UK, <sup>2</sup>Institute for Structural and Molecular Biology, Birkbeck, London WC1E 7HX, UK and <sup>3</sup>Centre for Clinical Microbiology, Royal Free Campus, University College London, London NW3 2QG, UK

Received December 12, 2017; Revised March 11, 2018; Editorial Decision March 14, 2018; Accepted March 15, 2018

## ABSTRACT

The success of *Mycobacterium tuberculosis* relies on the ability to switch between active growth and non-replicating persistence, associated with latent TB infection. Resuscitation promoting factors (Rpf) are essential for the transition between these states. Rpf expression is tightly regulated as these enzymes are able to degrade the cell wall, and hence potentially lethal to the bacterium itself. We have identified a regulatory element in the 5' untranslated region (UTR) of *rpfB*. We demonstrate that this element is a transcriptionally regulated RNA switch/riboswitch candidate, which appears to be restricted to pathogenic mycobacteria, suggesting a role in virulence. We have used translation start site mapping to re-annotate the RpfB start codon and identified and validated a ribosome binding site that is likely to be targeted by an *rpfB* antisense RNA. Finally, we show that *rpfB* is co-transcribed with *ksgA* and *ispE* downstream. *ksgA* encodes a universally conserved methyltransferase involved in ribosome maturation and *ispE* encodes an essential kinase involved in cell wall synthesis. This arrangement implies co-regulation of resuscitation, cell wall synthesis and ribosome maturation via the RNA switch.

## INTRODUCTION

The ability to switch between actively replicating and non-replicating persistence (NRP) is at the heart of *Mycobacterium tuberculosis*' success as a pathogen. *M. tuberculosis* expresses five resuscitation promoting factors (RpfA-E) (1). These are cell wall remodelling enzymes critical for the transition of *M. tuberculosis* between dormancy and resuscitation, and for reactivation of tuberculosis (TB) in animal

models (2–4). In an *in vivo* environment, *M. tuberculosis* forms cells that can only be grown with Rpf supplementation (5).

Precise and tight control of Rpf expression is vital as these enzymes are able to degrade the bacterial cell wall posing a potentially lethal threat to *M. tuberculosis* itself. Expression of the five Rpfs is induced by different triggers, many of which are associated with the host environment (6,7). ChIP-seq data indicates that several transcription factors, including MtrA regulate these promoters (8).

RNA-based regulation (riboregulation) of bacterial gene expression has attracted increasing attention over the last decade, as the wealth of molecules and the systems they regulate become more apparent (9–14). One class of riboregulators are the RNA switches, *cis*-regulatory elements, located largely within the 5' untranslated region (UTR) of the mRNA they regulate. Upon sensing a physiological signal such as temperature, pH, metabolites, RNA or proteins, they switch between conformations that are either permissive or non-permissive for downstream gene expression; RNA switches regulated by small molecule ligands are specifically referred to as riboswitches, and these currently make up the largest class of RNA switches (13,15,16). Riboswitches are formed of distinct domains with an aptamer domain responsible for binding a specific ligand, and an expression platform that regulates transcription or translation downstream (17). Most of the riboswitches described to date are widespread and associated with biosynthetic pathways; however, there are examples of less widespread riboswitches, and it is likely that there are many more, some of which may never be identified due to their rare occurrence (17,18). Riboswitches have been highlighted as potential drug targets due to their inherent ability to interact with a variety of ligands. For example, the FMN riboswitch has been suggested as potential drug target against *M. tuberculosis* infection (19). Several riboswitches have been predicted in *M. tuberculosis* by sequence homology and covariance

\*To whom correspondence should be addressed. Tel: +44 2076792242; Fax: +44 2076792242; Email: k.arnvig@ucl.ac.uk

(20). Among these is a homologue of the cyclic-di-AMP (c-d-AMP) sensing *ydaO* riboswitch, found in the 5' UTR of the *rpfA* mRNA (21,22). The *ydaO* riboswitch is widely conserved and generally regulates genes associated with cell wall metabolism, osmotic stress and sporulation (22).

Here, we identify a transcriptional RNA switch (riboswitch candidate) located within the 5' UTR of *M. tuberculosis* *rpfB*, and seemingly restricted to a subset of pathogenic mycobacteria. Based on experimental evidence, we have re-annotated the *RpfB* start codon and identified a likely Shine-Dalgarno (SD) sequence (23) that overlaps with an asRNA transcribed opposite to *rpfB*. The genetic arrangement of *rpfB* flanked upstream by the *tatD* nuclease, and downstream by the universally conserved *ksgA* methyltransferase and the essential *ispE* kinase is conserved in a wide range of Actinobacteria (Supplementary Figure S1) (24). We show that *rpfB*, *ksgA* and *ispE* are co-transcribed indicating a tight regulatory link between resuscitation, cell wall synthesis and ribosome maturation, subject to regulation by this element.

## MATERIALS AND METHODS

### Bacterial strains and growth conditions

*Escherichia coli* DH5 $\alpha$  were grown in LB liquid media or agar (1.5%) supplemented with 50  $\mu$ g/ml kanamycin or 250  $\mu$ g/ml hygromycin B as required.

*Mycobacterium smegmatis* mc<sup>2</sup>155 (25) was grown on LB agar supplemented with 50  $\mu$ g/mL hygromycin B as required, and in liquid LB media supplemented with 0.05% Tween 80 and 50  $\mu$ g/ml hygromycin B as required.

*Mycobacterium tuberculosis* H37Rv (26) and *Mycobacterium bovis* BCG were grown on Middlebrook 7H11 agar supplemented with 10% OADC, 0.5% glycerol and 50  $\mu$ g/ml hygromycin B as required and in liquid Middlebrook 7H9 medium supplemented with 10% ADC, 0.4% glycerol and 0.05% Tween 80 in roller bottles (Cell Master, Griener Bio-One) or PETG flasks (Nalgene, Thermo Scientific), respectively. Exponential phase cultures were harvested at OD<sub>600</sub> 0.6–0.8. Stationary phase cultures for *M. tuberculosis* and *M. bovis* BCG were harvested at least 1 week after 1.0 OD<sub>600</sub>. For time-course experiments, cultures were harvested as indicated. Biofilms were formed by adding 10 ml of an exponential phase culture to 50 ml polypropylene tubes, sealing tightly and leaving for the indicated amount of time. At time of harvest the pellicle was removed and processed for RNA. Mycobacteria were transformed by electroporation.

### Plasmid construction

Plasmids used in this study are listed in Supplementary Table S1.

### Oligonucleotides

Oligonucleotides used during this study are listed in Supplementary Table S2.

### RNA isolation

Total RNA extraction was performed as previously described. Briefly, ice was added directly to the culture, which was centrifuged at 5000 rpm for 10 min at 2°C and total RNA was extracted using the FastRNA Pro Blue Kit (MP Bio) according to the manufacturer's instructions. RNA concentration and quality was determined using a Nanodrop 2000 (27,28)

### cDNA synthesis and 3' rapid amplification of cDNA ends (RACE)

cDNA was synthesised using random hexamers and Superscript III reverse transcriptase (Invitrogen), largely according to manufacturer's protocol except for an additional extension step for 30 min at 55°C.

3' RACE was performed as previously described (29). Samples were reverse transcribed and primed using oligo d(T) adapter primer (oligonucleotide 2.07). RACE targets were amplified using adapter primer and a gene specific primer (oligonucleotide 2.09 and 2.15 respectively).

Co-transcription of *rpfB*, *ksgA* and *ispE* was analysed using cDNA generated with random hexamers. cDNA was amplified with REDTaq PCR reaction using primers flanking the *rpfB/ksgA* or *rpfB/ispE* region complementary to the sequence in *M. tuberculosis*, *bovis* BCG and *smegmatis* (oligonucleotides 5.51, 8.04 and 8.62).

### Northern blotting

Northern blotting and probing was performed as described in (29).

Template oligonucleotides are listed in Supplementary Table S2. Membranes were exposed to a phosphor screen and developed using Typhoon FLA 9500 (GE). Sizing of transcripts were done using Century marker (Ambion).

### Quantitative RT-PCR

'SensiFast SYBR Hi-ROX master mix' (Bio-line) was used to amplify cDNA for quantitative RT-PCR (qRT-PCR), according to manufacturer's instructions. *M. tuberculosis* H37Rv DNA was used to create a standard curve. Wells were loaded with either 1  $\mu$ l standard in three technical replicates or 1  $\mu$ l cDNA in four technical replicates.

All reactions were carried out using a 'QuantiStudio 6 Flex Real-Time PCR System' and analysed using QuantiStudio Real-time PCR software v1.1 (Applied Biosciences).

### $\beta$ -galactosidase assay

Protein extracts were obtained from cultures of *M. smegmatis* and assayed as previously described (28).

Miller units were expressed as a percentage of the average WT value. Statistical significance was calculated using one-way ANOVA with Tukey post hoc analysis in IBM SPSS  $\pm$ 1 standard deviation. Significance thresholds specified as: 'NS' (no significant difference,  $P > 0.05$ ), '\*' ( $P \leq 0.05$ ), '\*\*' ( $P \leq 0.01$ ) and '\*\*\*' ( $P \leq 0.001$ ).

### In vitro transcription

*Escherichia coli* RNAP *in vitro* transcription assays were carried out using previously described methods for producing halted transcription elongation complexes (TECs) (30,31). Transcription templates were cloned into pGAM-rnnX (details in supplemental methods).

### Q5 site directed mutagenesis

Site-directed mutagenesis (SDM) was carried out using the 'Q5 SDM kit' (NEB) following the manufacturer's protocol. Correct constructs were sub-cloned into un-treated vector.

### Overlap extension mutagenesis

For small mutations, a pair of Phusion GC polymerase PCR reactions were carried out: reaction (A) used an upstream forward primer and a mutagenic reverse primer spanning the region to be mutated, reaction (B) used a downstream reverse primer and a mutagenic forward primer spanning the region to be mutated. The resulting amplicons contained a region of complementarity exploited in reaction (C) by combining 1 µl of each as template in another PCR reaction with the non-mutagenic primers of the original reactions.

### Alignment of the *tatD-rpfB* intergenic regions

Test alignments of the intergenic regions between the *tatD* and *rpfB* genes in a number of mycobacteria indicated that a small number of mycobacterial species aligned well whereas others had large insertions and deletions. Based on the preliminary alignments, we selected a number of species with well-conserved intergenic regions to align first (Figure 9) and subsequently added to this alignment three more species to highlight divergence in the sequences of the latter (Supplementary Figure S7). Further details can be found in supplementary methods.

### Consensus structure of *rpfB* switch

Sequences with seemingly functional P1 (based on -10 box) and terminator (based on Transterm (32)) were used to generate the consensus in LocARNA (33) using default parameters. Poly-U tails were removed for clarity.

## RESULTS

### Promoters and transcripts of the *rpfB* locus

Through interrogation of *M. tuberculosis* (d)RNA-seq (34), we found that *rpfB* is expressed from two promoters: P1, with transcription start site (TSS) at G1127876, and P2 with TSS at A1127955. For both TSS we identified canonical (TANNNT) -10 regions (Figure 1). RNA-seq also indicates the presence of an antisense RNA expressed from P<sub>as</sub> with TSS at G1128048.

Expression from these promoters was validated by cloning the region from 140 bp upstream of P1 to the annotated ATG start codon in frame to a *lacZ* reporter (Figure 1). In addition, we made three derivatives mutating

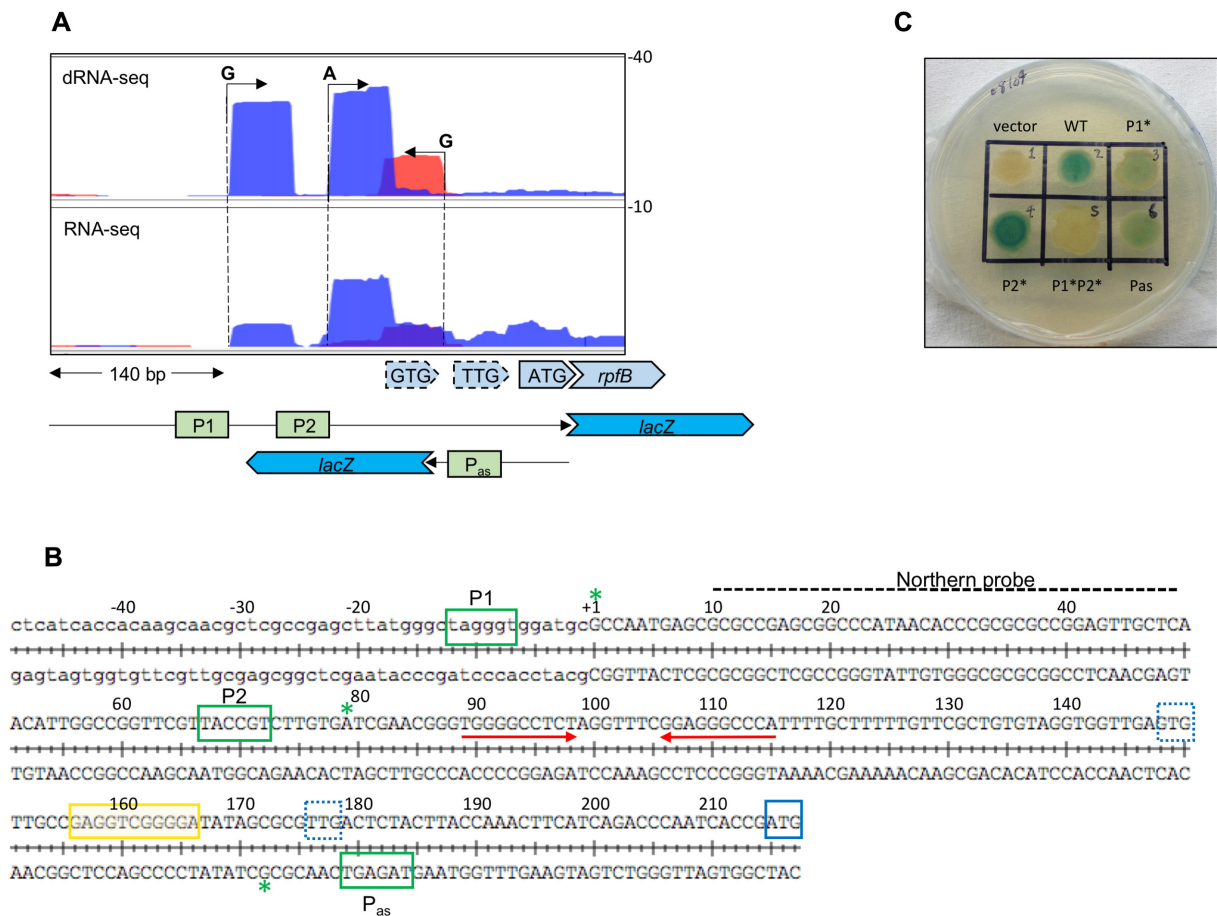
the -10 regions of either P1 or P2 separately or in both P1 and P2. Finally, we made a transcriptional fusion of P<sub>as</sub> including 100 bp upstream of the TSS. The constructs were transformed into *M. tuberculosis* and promoter activity was assessed by colony colour on X-gal plates (Figure 1C). Mutating the promoters individually suggested that P1 and P2 are both active in *M. tuberculosis*, corroborating the RNA-seq data. The lack of expression in the double mutant indicates that P1 and P2 are the only promoters driving *rpfB* expression, which is supported by the TSS mapping. Moreover, the results indicate that P<sub>as</sub> is active and may play a role in *rpfB* expression. To investigate if the asRNA had an effect on *rpfB* expression, we employed a dual expression/reporting vector designed in our lab (28). To avoid unwanted promoter effects, both asRNA and the *rpfB-lacZ* fusion (similar to the one above) were expressed from divergently transcribed heterologous promoters, meaning the asRNA was no longer *cis*-encoded. Somewhat surprisingly, the results of the reporter gene assay indicated that the asRNA had very little, if any effect on *rpfB-lacZ* expression in this context (Supplementary Figure S2).

### RpfB translation start site

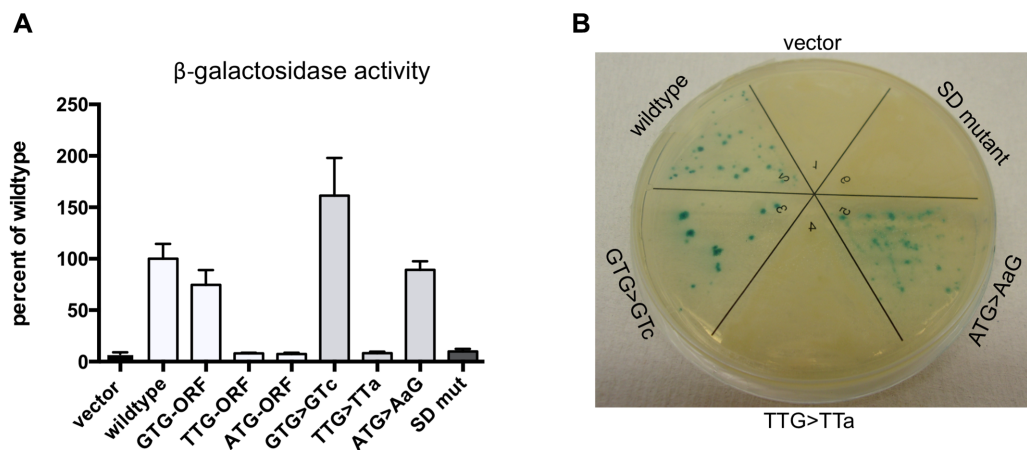
The annotated translation start site of RpfB is ATG (Figure 1). However, there is no obvious SD sequence proximal to this start codon; moreover, the start sites of the RpfB homologues in *Mycobacterium leprae* and *Mycobacterium smegmatis* have been annotated 13 codons further upstream, corresponding to the alternative TTG start, which has a likely SD sequence upstream (Figure 1). In line with previous observation, we considered that the RpfB start codon may have been mis-annotated (35,36), and we identified two potential start sites (TTG and GTG) upstream of the annotated ATG (Figure 1). To define which of the potential start sites was correct, we modified the method developed by Smollett et al. for translation start site mapping (36), using the wildtype translational *lacZ* fusion described above. Frameshift mutations were introduced separately between GTG and TTG and between TTG and ATG. If a frameshift were located within the resulting coding sequence, functional beta-galactosidase (β-gal) would not be expressed. The constructs were transformed into *M. smegmatis*, a tractable surrogate host for the expression of *M. tuberculosis* genes, and cell extracts were assayed for β-gal activity.

The results, shown in Figure 2, demonstrate that the frameshift between GTG and TTG retained ~75% of wild-type β-gal activity level, suggesting that this part of the transcript was outside the translated region. However, the frameshift between TTG and ATG reduced β-gal activity to the level of the empty vector, indicating the mutation lay within the translated region and hence that TTG was the correct start codon (Figure 2). As this result was in conflict with previously published data (37), we employed an alternative method to validate our findings. Each of the three potential start sites (GTG, TTG, ATG) was mutated to non-start codons (GTC, TTA, AAG), and β-gal activity of the resulting constructs assayed. The results (Figure 2) corroborated our findings from the frameshift experiment; changing GTG and ATG to non-start codons did not significantly





**Figure 1.** Transcription start sites (TSS) and promoter elements in the *rpfB* locus. (A) dRNA-seq and RNA-seq of the promoter region, 5' UTR and early ORF of *rpfB*. Numbers on the right indicate normalised reads. Three TSS, two sense and one antisense, were identified (data from (34)). Below are schematics of the regions covered by the reporter constructs. (B) Sequence of promoter region and 5' UTR of *rpfB*. Green boxes indicate the -10 hexamers of promoters P1, P2 and P<sub>as</sub>; green asterisks: TSS. Red arrows: inverted repeat leading to a stem-loop structure; blue box: annotated translation start site, blue dotted boxes: alternative translation start sites, yellow box: putative ribosome binding site. (C) X-gal plate with reporter constructs expressed in *M. tuberculosis*. Wildtype refers to 'full-length' construct from -140 upstream of P1 to the ATG. Mutations in the -10 regions (TANNNT to CANNNC) are indicated by asterisk (i.e. P1\* has inactivated P1, but functional P2 and *vice versa*).



**Figure 2.** Translation start site mapping. Left panel shows results of  $\beta$ -gal assays on translational reporters expressed in *Mycobacterium smegmatis*. GTG-ORF, TTG-ORF, ATG-ORF indicate the results of introducing frameshifts in the open reading frames downstream of the indicated putative start site. GTG > GTc, TTG > TTa, ATG > AaG indicate the results of changing the putative start site to a non-start codon. The last bar in the graph shows the activity of the mutated SD sequence (GAGGTCGGGGA to ctccTCccct). The values represent the mean and standard deviation of six biological replicates; \* $P \leq 0.05$ ; \*\*\* $P \leq 0.001$ . Right panel shows translation start site and SD mutants expressed in *M. tuberculosis*.



reduce  $\beta$ -gal activity, while changing the TTG to TTA reduced the expression to empty vector level, thus verifying that TTG was the correct start codon. Further supporting this notion was the fact that we could only identify a putative SD sequence -10 to -20 relative to TTG (Figure 1B). To investigate if this sequence affected *rpfB* expression, we mutated the SD purines to pyrimidines in the *lacZ* fusion. The  $\beta$ -gal activity of the resulting construct was reduced to the level of the empty vector (SD mut, Figure 2), suggesting that this was a likely ribosome binding site.

Finally, we transformed selected constructs with altered start sites into *M. tuberculosis* to ensure there were no significant differences compared to *M. smegmatis*. The results in *M. tuberculosis*, seen as blue/white colony colour (Figure 2B), were in perfect agreement with the results obtained in *M. smegmatis*, supporting the notion that the correct translation start site for *M. tuberculosis* RpfB is the relatively unusual TTG codon.

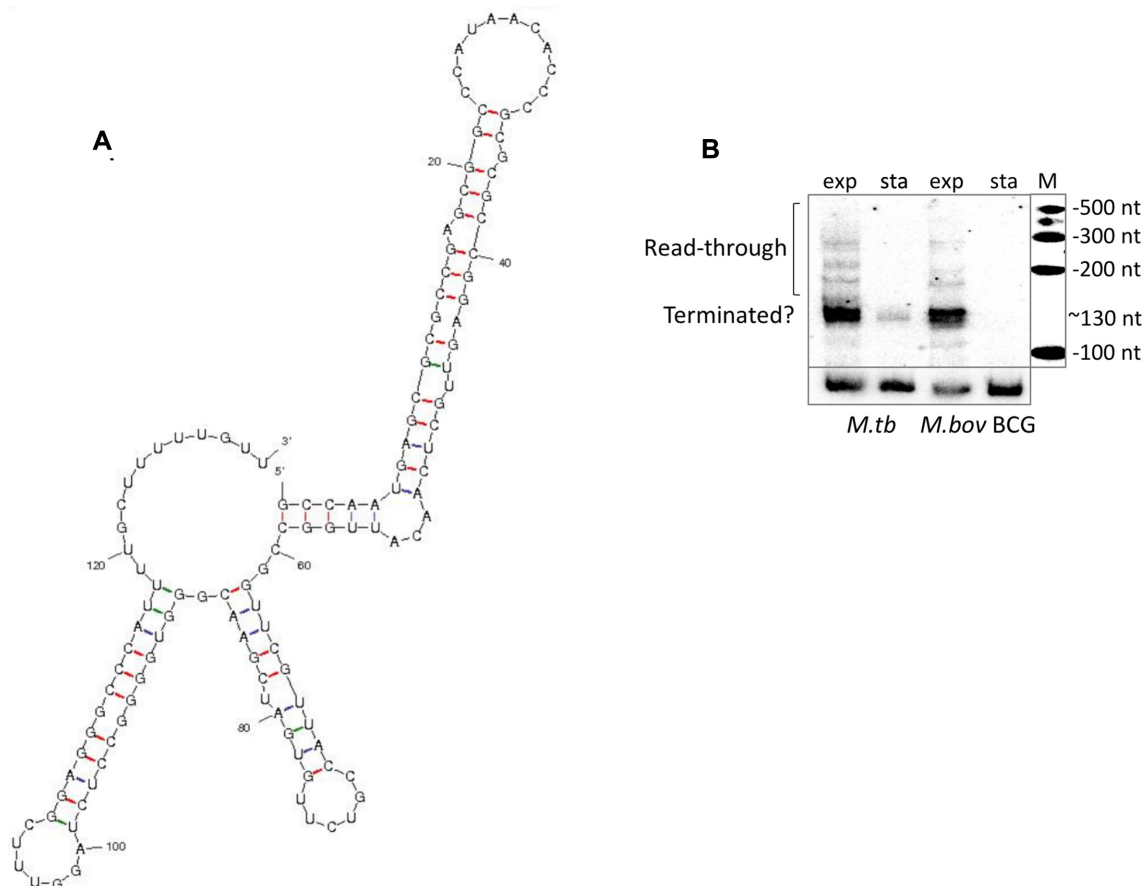
### The *rpfB* 5' UTR

The first 130 nucleotides of the *rpfB* 5' UTR expressed from P1 include an inverted repeat (red arrows, Figure 1B) followed by a poly-U tract, suggestive of a potential intrinsic terminator. Using *mfold* (38), we found that the predicted

structure of the 130 nucleotides does indeed contain a stem-loop followed by a poly-U tract (Figure 3A). To determine if this sequence might lead to premature transcription termination, we analysed RNA from exponential and stationary phase cultures of *M. tuberculosis* and the closely related *Mycobacterium bovis* BCG by Northern blotting. Figure 3B shows a Northern blot with a strong signal around 130 nucleotides in exponential phase from both species, consistent with a terminated transcript. In addition, there are several weaker signals corresponding to larger transcripts. In stationary phase, there was little or no expression in both species, in concordance with previous observations (6,27).

To identify more precisely the 3' termini associated with the *rpfB* 5' UTR, we performed 3' RACE as previously described (28). The results indicate that 12% of transcripts had a 3' end well upstream of the poly-U tract, while 42% of 3' ends fell within or proximal to the poly-U tract with U123 and U124 alone accounting for 17% (Supplementary Figure S3). Further downstream we found that 8% of the 3' ends were located at the newly annotated TTG start codon.

The fact that more than a third of all 3' termini fall within the poly-U tract, supports the notion of a functional intrinsic terminator. The results also suggest that U117 is part of the poly-U tail and not the preceding stem as the structure in Figure 3A suggests. We therefore re-



**Figure 3.** *rpfB* 5' UTR. (A) *mfold* (38) predicted structure (without constraints) of the first 130 nucleotides of the *rpfB* 5' UTR containing an intrinsic terminator structure. (B) Northern blot of RNA from exponential and stationary phase cultures of *M. tuberculosis* and *M. bovis* BCG. RNA was separated by PAGE, transferred to a nylon membrane and probed with a ribo-probe indicated in Figure 1B.

modelled the RNA with the constraints that residues downstream of A116 were unpaired. This suggested in two alternative structures. One had a slightly modified terminator and lower free energy than the original ( $\Delta G$  -50.2 versus -49.6 kcal/mol). The second was significantly different, without intrinsic terminator and with a higher free energy ( $\Delta G$  -44.7 kcal/mol); we consider the latter a potential antiterminated or readthrough conformation (Figure 4). The P2 derived transcript, which lacks the initial 79 nucleotides of the P1 derived transcript, cannot form the antiterminated structure, and is predicted to form only the terminator.

In summary, our results indicate that the P1 derived 5' UTR of *rpfB* can adopt two conformations, one of which contains an intrinsic terminator, suggesting that this element comprises an RNA switch.

### Translational reporter fusions support the notion of an RNA switch

To verify and further characterise this putative RNA switch we employed the previously described translational *lacZ* fusion. First, we compared constructs with and without the RNA switch, deleting the entire region from TSS1 to the end of the poly-U tract. This resulted in a significant increase in  $\beta$ -gal activity, suggesting that the RNA switch provides an additional layer of control by reducing RpfB expression during exponential growth (Figure 5). The two conformations of the *rpfB* 5' UTR are both likely to exist *in vivo*. We used these structures to predict single-nucleotide substitutions that could stabilise either conformation. Thus, a U6C substitution (green circles, Figure 4) would favour the antiterminated structure, while a G112C substitution (red circles, Figure 4) would favour the terminated structure. The mutations were introduced into the *lacZ*-fusions and  $\beta$ -gal activity determined. The results demonstrate that stabilising the predicted terminator leads to significantly reduced *lacZ* expression, while stabilising the antiterminator structure leads to significantly increased expression (Figure 5). To further probe the intrinsic terminator, we made a mutant in which U117 to U119 were changed to adenines. This resulted in increased expression similar to that observed for the U6C mutant. These results substantiate the presence of the two structures and the potential to switch between these.

Together our results strongly support that the *rpfB* 5' UTR comprises a transcriptional RNA switch that provides an additional layer of regulation to RpfB expression. As the terminated conformation has the lowest predicted free energy of the two, we assume it is the default conformation, and that its cognate ligand would promote readthrough.

### *rpfB*, *ksgA* and *ispE* form a tri-cistronic operon

Immediately downstream of the *rpfB* gene lies a gene encoding the highly conserved methyltransferase, KsgA that specifically methylates two adjacent adenosine residues in the 3' end of the 16S ribosomal RNA (residues 1511 and 1512 within the sequence GGAAG in *M. tuberculosis*). This process is regarded as a checkpoint for ribosome maturation (39). There are no TSS identified between the *rpfB* 5' UTR and the *ksgA* gene (Figure 5), indicating that the two genes are part of the same operon. Moreover, according to

the annotation, the ORFs for these two genes overlap, suggesting a very tight coupling in their expression. We tested if the two genes were co-transcribed using RT-PCR. The results, shown in Figure 6 suggest that *rpfB* and *ksgA* are co-transcribed in both *M. tuberculosis*, *M. bovis* BCG and in the more distantly related *M. smegmatis*. In *M. tuberculosis*, but not in *M. smegmatis*, lies the essential *ispE* downstream of *ksgA*. This gene encodes an ATP-dependent kinase involved in isoprenoid synthesis and ultimately, cell wall synthesis by providing the linker unit between arabinogalactan and peptidoglycan (40). Although there is a weak TSS 37 basepairs upstream of the annotated IspE GTG start codon as well as a consensus -10 motif (TAGTCT), we tested the possibility that *ispE* was co-transcribed with *rpfB* and *ksgA* due to the close proximity of the ORFs. The result, shown in Figure 6, indicates that this is indeed the case and hence that *rpfB*, *ksgA* and *ispE* form a tri-cistronic operon in *M. tuberculosis* with an internal promoter driving baseline expression of *ispE*.

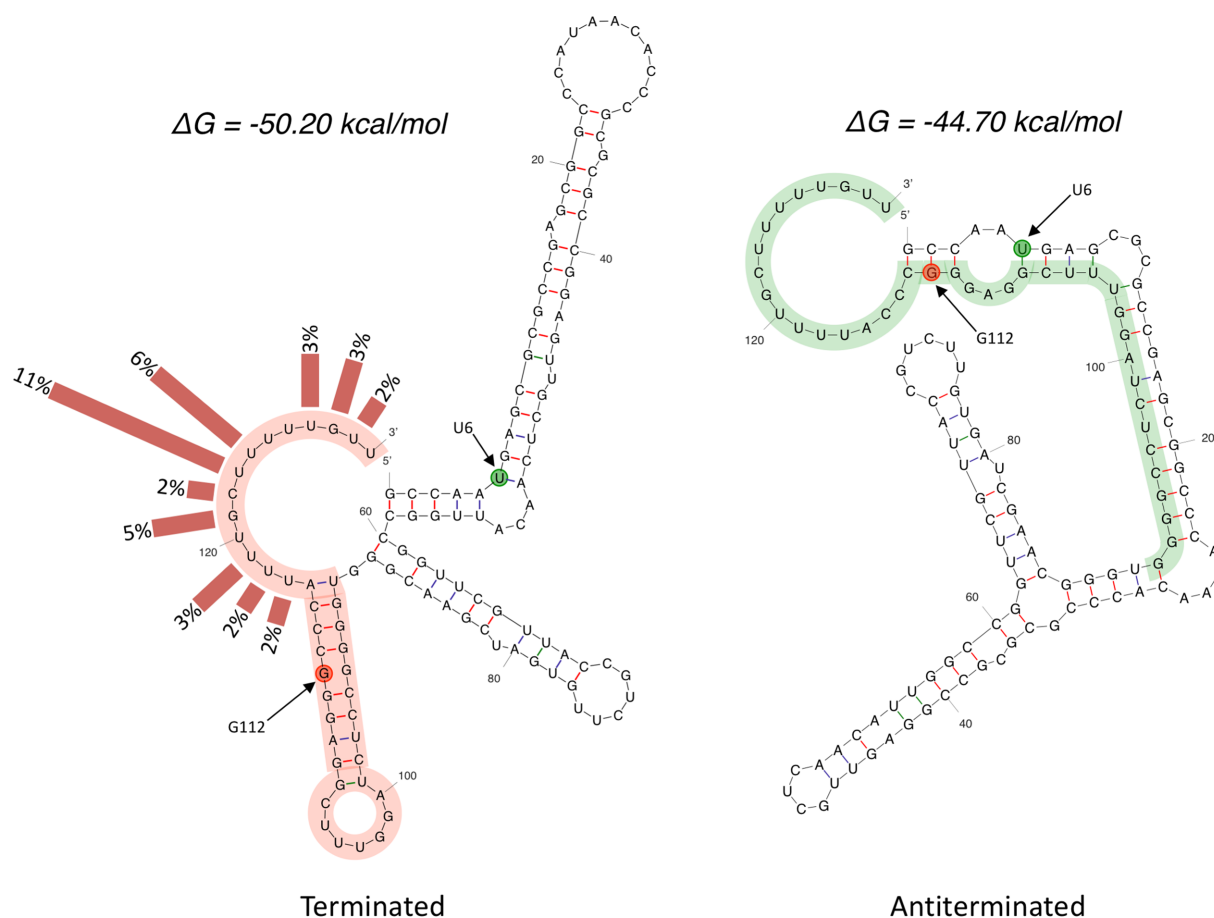
This, in turn indicates that *rpfB*, *ksgA* and *ispE* expression is regulated by the same RNA switch in *M. tuberculosis*. This arrangement provides a regulatory link between resuscitation, ribosome maturation and cell wall synthesis. It also offers the possibility that a cognate ligand could be associated with KsgA or IspE as well as with RpfB.

### Expression of *rpfB* during re-growth and nutrient starvation

To obtain a more detailed picture of termination and readthrough of the RNA switch, we investigated the expression under different growth conditions. Initially we looked at expression as cells emerge from stationary phase into log-phase. A stationary phase culture, in which *rpfB* is poorly expressed, was diluted into fresh medium followed by RNA sampling over time. Figure 7A shows a Northern blot of the time course probed for the RNA switch, which indicates robust expression of the terminated transcript after one hour in fresh medium, while expression of the longer, readthrough transcripts reached a maximum later (around 5 hours) into the time course, suggesting that the cells require more time to achieve ligand concentrations permissive of readthrough. We also investigated expression after the cells had been shifted to starvation conditions. Exponential phase culture was resuspended in PBS + Tween80 followed by RNA sampling over time. Figure 7B shows that P1-driven expression of *rpfB* ceases relatively quickly following nutrient starvation.

### Expression of *rpfB* in biofilms

The formation of mycobacterial biofilms requires significant changes in gene expression followed by substantial re-arrangements of the cell wall (41); however, changes in Rpf expression have not been reported. We investigated the expression of the RNA switch as well as *rpfB*, *as-rpfB* and *ksgA* transcripts in biofilms of *M. bovis* BCG, a close, more tractable relative of *M. tuberculosis* in which the entire *rpfB-ispE* transcript, including 5' UTR, is 100% conserved. Biofilms were allowed to form in static, non-aerated cultures for the indicated period of time after which the pellicle was removed and processed for RNA. Quantitative



**Figure 4.** Alternative structures of the *rpfB* 5' UTR (1–130). The figure shows the two structures that were predicted with *mfold* (38) with the constraints that U117 is unpaired. Frequency of each 3' end determined by RACE has been indicated with bars. Point mutations that stabilise either the left conformation or the right conformation have been indicated with red and green circles, respectively. Red highlight indicates the terminator with the same sequence shown in green in its antiterminated conformation.

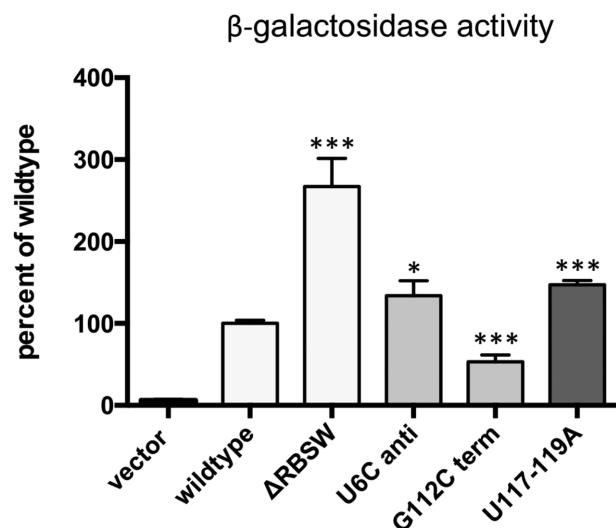
real-time PCR (qRT-PCR) was performed for the 5' UTR, P1 readthrough, *rpfB*, *asrpfB* and *ksgA*; details of these amplicons are outlined in Supplementary Figure S4, and the results are shown in Supplementary Figure S5. These indicate that the level of all measured transcripts was slightly, albeit not significantly reduced during the initial stages of biofilm formation but recovering as the biofilm matured. As *ispE* expression is driven by an additional promoter, we did not include this in our investigation.

To obtain values for transcriptional readthrough versus termination within P1 derived transcripts, we normalized the raw values as outlined in Supplementary Figure S4. The final result, shown in Figure 7C, indicates the level of P1-derived *rpfB* transcripts normalized to the values obtained for the 5' UTR. The results can therefore be used as an approximation of the proportion of transcripts that proceed through the terminator region into the *rpfB* coding region. Similarly, we normalized the values for *ksgA* transcripts to *rpfB* transcripts to obtain a measure of relative abundance of the two *cistrons*. Overall the results indicate that there are no significant changes in the relative amounts of the investigated transcripts during biofilm formation.

### Transcription of the *rpfB* attenuator *in vitro*

To demonstrate that the *rpfB* attenuator is capable of promoting termination of transcription, and to screen putative ligands of the RNA switch, we designed a single-round *in vitro* transcription assay. Since all four nucleotides are present within the first six positions of the RNA switch, we modified the 5' end marginally to obtain a template that was suitable for single-round *in vitro* transcription (see Supplementary methods). To mimic *M. tuberculosis* RNA polymerase (RNAP), the elongation rate of *E. coli* RNAP was reduced by limiting the NTP concentration to 50  $\mu$ M (Supplementary Figure S6). We first tested the wildtype RNA switch and the three mutants from the reporter constructs, expressed from a heterologous promoter. Transcription readthrough was observed either as template run-off or readthrough to the *synB* synthetic terminator (42). Halted elongation complexes were formed using *E. coli* RNAP and chased in the presence of heparin. The results demonstrated that the *rpfB* terminator is recognised by the *E. coli* RNAP resulting in approximately half of the complexes pausing/terminating at the predicted site (Figure 8A, lanes 1 and 5), while the remaining continue transcription to obtain either the run-off transcript (lane 1) or the *synB* ter-





**Figure 5.** Reporter gene assays support the presence of an RNA switch. The figure shows  $\beta$ -galactosidase activity of translational reporters expressed in *M. smegmatis*. The constructs include the promoter region, 5' UTR and 14 codons of the *rpfB* ORF, including ATG, as shown in Figure 1.  $\Delta$ RBSW: entire RNA switch, including P2 deleted from the construct; U6C anti: point mutation predicted to stabilise the antiterminated conformation; G112C term: point mutation predicted to stabilise the terminated conformation; U117-119A: change of three U residues to A residues. The values represent the mean and standard deviation of six biological replicates; \* $P \leq 0.05$ ; \*\*\* $P \leq 0.001$ .

minated transcript (lane 5). Stabilising the terminator stem led to multiple signals around the *rpfB* terminator (lanes 2 and 6), while the run-off and the *synB* terminated transcript were both replaced by aberrant signals that were ~30–40 nucleotides longer.

As this size transcript exceeded the theoretical maximum length possible using the template, we treated the samples with DNase to investigate the possibility of template labelling activity (not shown). However, this did not remove the aberrant signal, a phenomenon that we are currently unable to explain. More importantly, the two mutants, U6C and U117–119A both displayed decreased termination at the *rpfB* terminator and increased readthrough, supporting the *in vivo* findings and lending significant support to the presence of a transcriptionally regulated RNA attenuator (Figure 8A, lanes 3, 4, 7 and 8). Some transcriptionally regulated RNA attenuators require the RNAP to pause at specific sites to allow co-transcriptional folding and ligand binding (43). We investigated the pausing pattern of the RNA switch in a time-course experiment in the presence and absence of NusA, a transcription factor known to promote transcriptional pausing. As suspected, there were several pause sites within the sequence, most of which were enhanced in the presence of NusA, resulting in an overall reduced elongation rate (Figure 8B). We observed a particularly enhanced pause signal around position 41 and 43, corresponding to positions 37 and 39 in the true RNA switch transcript (indicated with +++ in Figure 8B).

These results suggest that *in vivo*, NusA may be required to allow more time for potential ligand interactions which may be necessary for antiterminator formation and transcriptional readthrough.

### The *rpfB* attenuator appears to be restricted to a subset of pathogenic mycobacteria

Many RNA switches are highly conserved, particularly between closely related species. To investigate the occurrence and conservation of the *rpfB* attenuator, we aligned sequences upstream of the *rpfB* coding region from six (pathogenic) mycobacterial species; as *Mycobacterium canettii* showed some variations within the chosen region, we used three different strains (Figure 9).

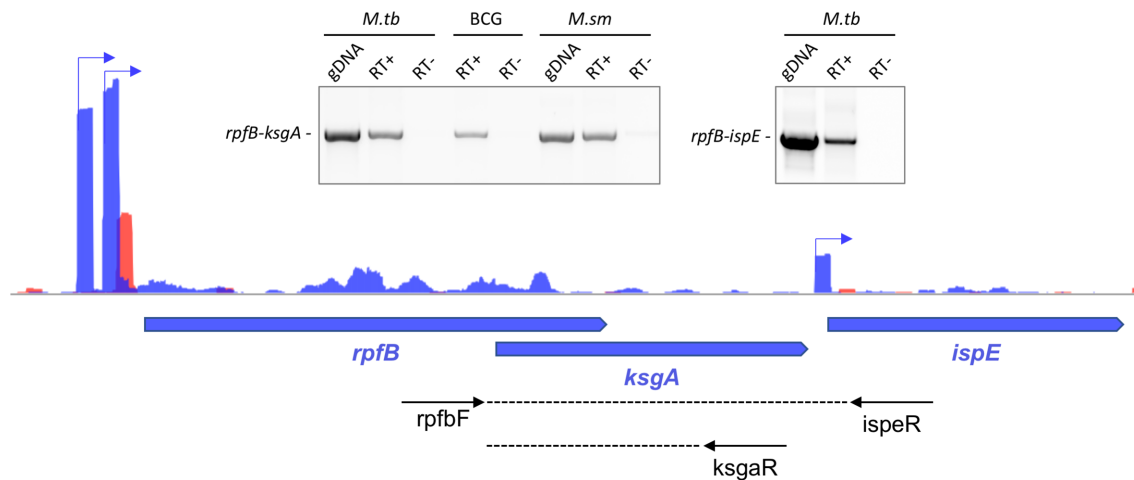
The alignment indicates that the P2–10 region is identical in all of the selected species.

P1, on the other hand, is less well-conserved, and the TANNNT -10 consensus indicative of a functional promoter is only seen in *M. tuberculosis*, *M. canettii*, *Mycobacterium marinum* and *Mycobacterium ulcerans*, suggesting that the long 5' UTR may be restricted to a subset of mycobacteria; additional species from the *M. tuberculosis* complex (MTBC), i.e. *M. bovis*, *Mycobacterium africanum* and *Mycobacterium microti* had sequences that were identical to *M. tuberculosis* (not shown). A more extensive alignment including more distantly related species is shown in Supplementary Figure S6. According to this, *Mycobacterium leprae* may also have a functional P1 promoter. Next, to identify putative hairpins/terminators within the 5' UTR of the selected species, regardless of P1, we analysed the same region using Transtern (32). Only *M. tuberculosis*, *M. canettii*, *M. marinum*, *M. ulcerans*, *Mycobacterium shinjukuense* and *Mycobacterium kansasii* were predicted to have a functional terminator (Supplementary Figure S7), meaning that species predicted to have a functional P1 promoter and a functional terminator are MTBC species as well as *M. marinum* and *M. ulcerans*. Thus, we conclude that the RNA attenuator is only present in a subset of pathogenic mycobacteria with a phylogenetic split between MBTC/*M. marinum*/*M. ulcerans* on one side and *M. leprae*, *M. avium* on the other, which is consistent with the split seen by aligning 16S ribosomal RNA sequences (44) or randomly selected genes (45). Based on our findings we used LocARNA (33), to produce a consensus structure for the species predicted to have long 5' UTRs (based on P1) and terminator, i.e. *M. tuberculosis*, *M. canettii* (three strains with different sequences), *M. marinum* and *M. ulcerans*, shown in Figure 10. This structure reveals some variation in the first stem, while the two remaining stems, including the terminator and the unpaired loop regions are more well-conserved. Overall this structure aligns well to the original *M. tuberculosis* structure without constraints (Figure 3).

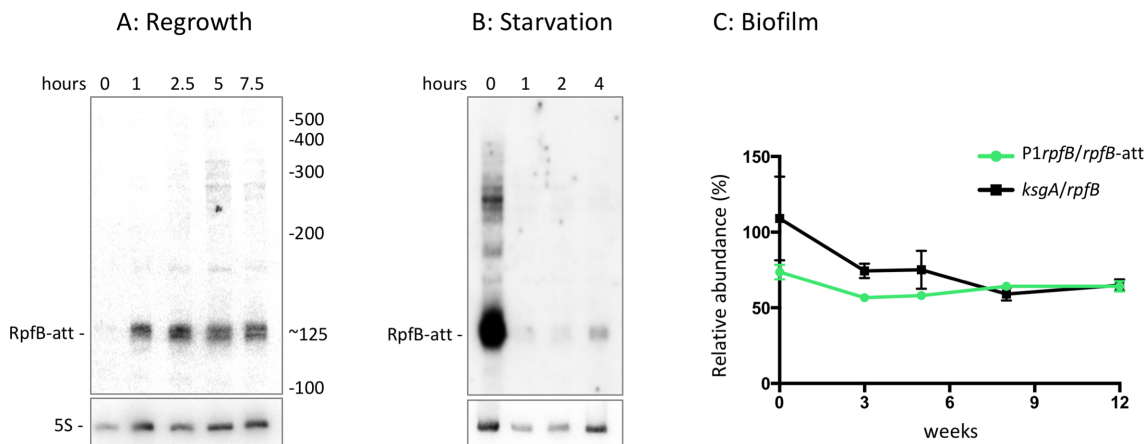
Finally, only MTBC species and *M. avium* have an antisense promoter that adheres to the consensus sequence (Figure 9), indicating that the presumably tight regulation provided by multiple promoters, RNA attenuator and asRNA is specific for species within the MTBC. The association of this element with certain pathogenic species only, offers the possibility that its function is associated with pathogenesis and adaptation to the host environment.

### DISCUSSION

Rpfs are cell wall remodelling enzymes with the potential to lyse and kill the cells that express them. Hence, their expression is under tight often multi-layered control.



**Figure 6.** Co-transcription of *rpfB* with downstream genes. Main image shows three TSS associated with the *M. tuberculosis* *rpfB* locus on the plus strand; two for *rpfB* and a minor for *ispE*, according to global TSS mapping (34). Black arrows below locus indicate primers used for RT-PCR. Inserts show RT-PCR; left: *rpfB* and *ksgA* are co-transcribed in *M. tuberculosis* (*M.tb*), *M. bovis* BCG (BCG) and *M. smegmatis* (*M.sm*); right: *rpfB*, *ksgA* and *ispE* are co-transcribed in *M. tuberculosis*.

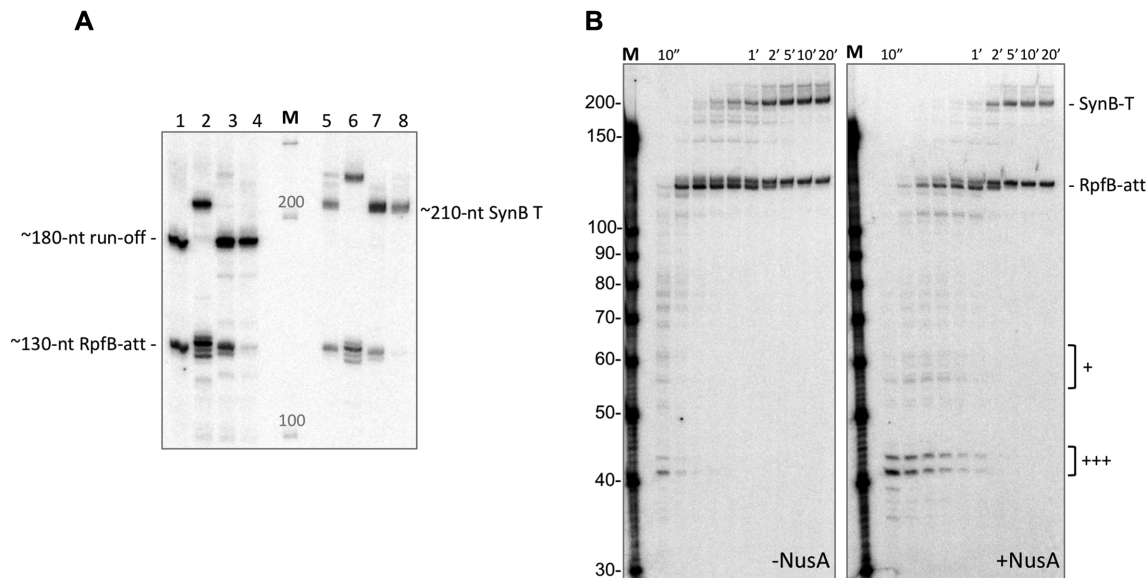


**Figure 7.** Expression and turnover of the *rpfB* 5' UTR. Northern blots of *M. tuberculosis* RNA harvested at the indicated time points; RpfB-att corresponds to terminated transcript. (A) After dilution of a stationary phase culture (1 week after OD<sub>600</sub> = 1) into fresh medium. (B) Shows that expression of RNA switch ceases quickly after cells have been shifted to PBS + 0.05% Tween80. For both, 15 µg of RNA was separated by PAGE, transferred to a nylon membrane and probed for the RNA switch (oligos 1.48 for 5S RNA and 5.22 for RNA switch). (C) Expression during biofilm formation. Normalised expression of P1 readthrough and *ksgA* transcripts. Values for P1 derived *rpfB* mRNA (Supplementary Figure S4) were normalised to values for 5' UTR RNA (P1*rpfB*/*rpfB*-att), and values for *ksgA* coding RNA were normalised to *rpfB* coding RNA. The graph illustrates the amount of P1-derived *rpfB* transcript relative to the amount of 5' UTR transcript over 12 weeks of biofilm formation. Values represent mean and SD of three biological replicates.

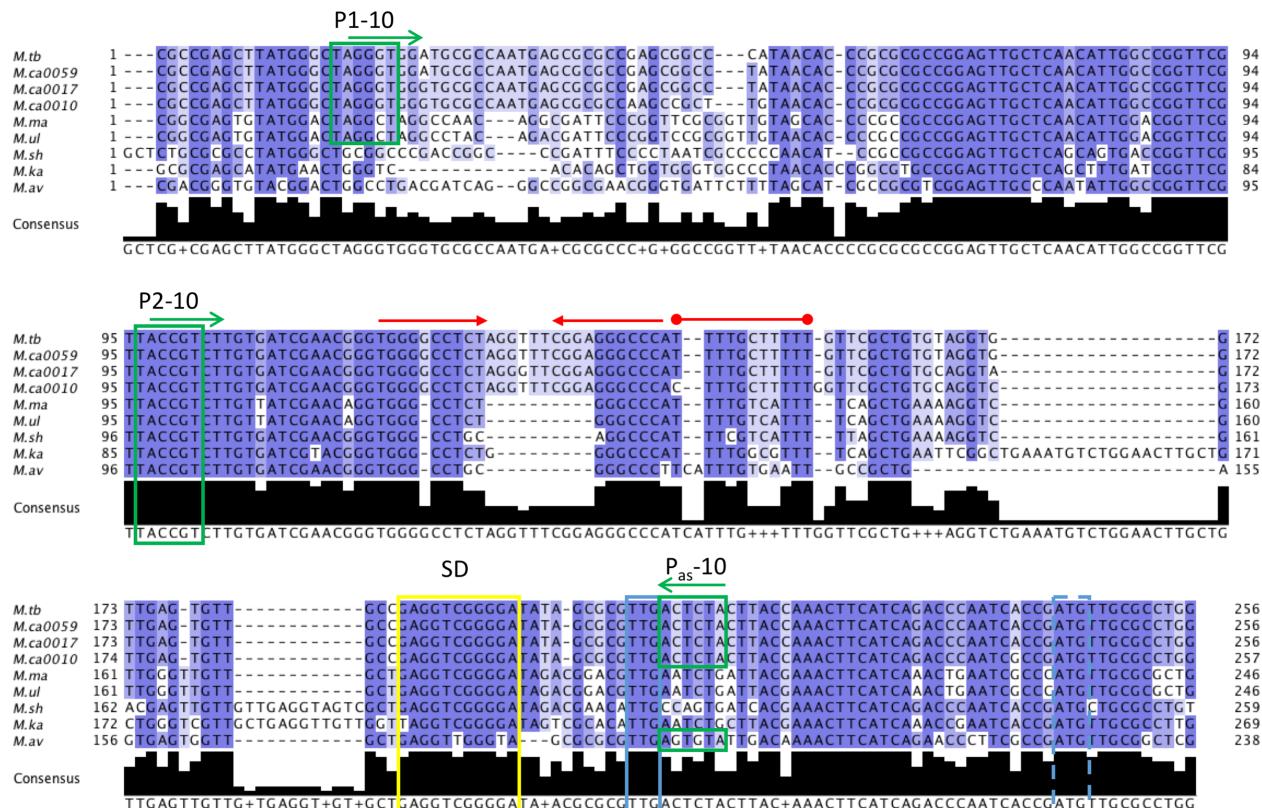
In the current study, we have shown that multipronged regulation also applies to the expression of *M. tuberculosis* RpfB. This gene is transcribed from two promoters and post-transcriptionally regulated by a pathogen-specific, transcriptionally regulated RNA switch. In this context, it is remarkable that three of the five *rpf* mRNAs have long 5' UTRs and two of these are post-transcriptionally regulated by RNA switches. Moreover, we provide evidence for a functional antisense promoter, although we did not observe changes in *rpfB* expression when the asRNA was expressed in *trans*. This does not rule out the possibility that transcription from the native antisense promoter in the correct context could have an effect on RpfB expression by RNA polymerase collision, mRNA processing or both. However, changing the antisense promoter within the *rpfB* locus would inevitably alter the codons of the *rpfB* mRNA,

and hence it would be difficult to determine the cause of any changes in expression.

We also show that the high level of multi-layered control appears to be specific for species within the MTBC, as one or more of the described elements are absent from other mycobacterial species. Moreover, we show that *rpfB*, *ksgA* and *ispE* form a tri-cistronic operon, implying that all of these regulatory mechanisms may extend to *ksgA* and *ispE* expression as well. However, *ispE* is essential and expression likely to be affected in the previously described *rpfB* deletion strain (46). Hence, we assume that the weak TSS upstream of *ispE* is sufficient for survival, while the tri-cistronic arrangement with *rpfB* and *ksgA* ensures coordinated expression of the genes directed by the RNA switch. In addition, we have re-annotated the translation start site and identi-

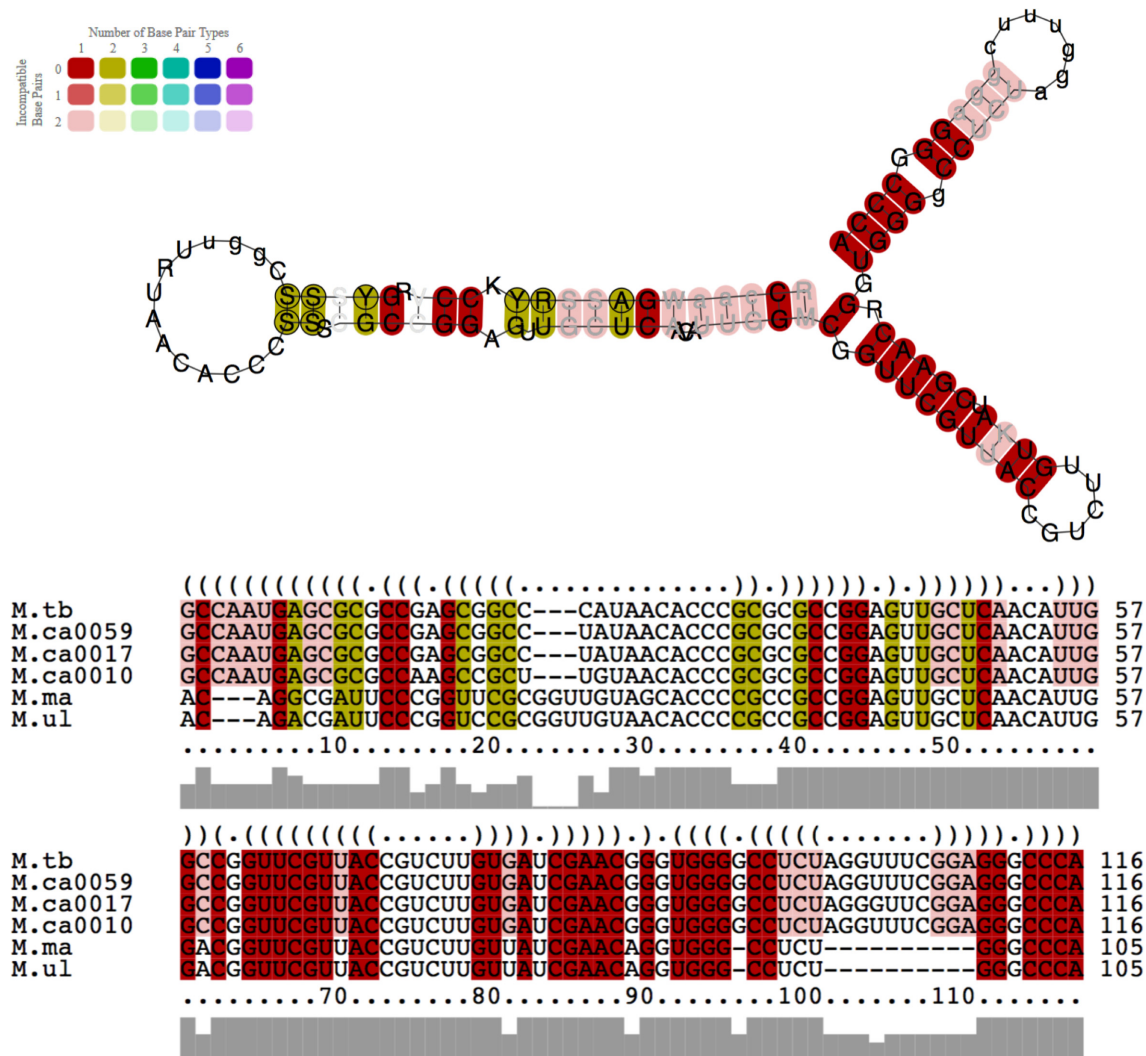


**Figure 8.** *In vitro* transcription of RNA switch. Transcription was initiated with GpU, omitting UTP from the initial reaction and labelling with  $^{32}\text{P}$ - $\alpha\text{ATP}$ . (A) Transcription of mutant templates; lanes 1–4 show reactions with run-off template; lanes 5–8 shows reaction from template with *synB*-mediated termination instead of run-off. Lanes, 1 + 5: wildtype; 2 + 6: G112C term; 3 + 7: U6C antiterm; 4 + 8: U117-119A poly(A). (B) Single-round *in vitro* transcription of the wildtype RpfB RNA switch. The *rpfB* RNA switch was transcribed *in vitro* with *E. coli* RNAP. Initiation complexes were stalled at position 11 and elongated in the presence of heparin and 50  $\mu\text{M}$  NTP at 37°C. Left image is with RNAP only and right gel image is in the presence of 5-fold molar excess NusA. + symbols indicate regions with NusA enhanced pausing.



**Figure 9.** Sequence alignment of *rpfB* promoter regions and 5' UTRs. Selected *Mycobacterium spp* are aligned: *M. tuberculosis* (*M.tb*), three strains of *M. canettii*: CIPT 140010059 (*M.ca0059*), CIPT 140070017 (*M.ca0017*) and CIPT 140070010 (*M.ca0010*), *M. marinum* (*M.ma*), *M. ulcerans* (*M.ul*), *M. shinjukuense* (*M.sh*), *M. kansasii* (*M.ka*), *M. avium* (*M.av*); an extended alignment can be seen in Supplementary Figure S5. The alignment is coloured by % sequence identity (darker blue = higher conservation). Green boxes: –10 regions; blue box: TTG start codon; blue dashed box: previously annotated ATG start; yellow box: SD sequence. Red arrows: inverted repeats, followed by red line indicating poly-U tract based on the *M. tuberculosis* sequence.





fied a likely ribosome-binding site based on several lines of experimental evidence.

fusions and previously published dRNA-seq (34) confirm that there is no promoter activity in that region. We employed two different means of determining the translation start and findings were further supported by the presence of a likely SD sequence. Therefore, we regard TTG as the correct start site. This also means that the asRNA is positioned immediately upstream of the start codon and covering the newly identified SD sequence.

diated by a small molecule, characteristic of a *bona fide* riboswitch, but rather a protein ligand, similar to ribosomal protein operons or yet another molecule/mechanism capable of stabilising one of the two conformations. The lack of widespread conservation and the fact that the genes regulated by this element are not associated with metabolic pathways further complicates the prediction of this ligand. The coordinated expression of *rpfB*, *ksgA* and *ispE* fits a model in which one or more molecular signals leading to resuscitation and cell wall remodelling/synthesis associated with growth, also lead to activation of protein synthesis by allowing the final steps in ribosome maturation. Moreover, the coordinated expression of the genes within this operon ensures that the cell maintains a carefully balanced ratio between different aspects of macromolecular synthesis, which is also apparent in operons encoding RNA polymerase subunits together with ribosomal proteins. The regulatory link described in this study means that resuscitation and ribosome maturation or rephrased, cell wall synthesis and protein synthesis, two classical antimicrobial targets could be simultaneously targeted via the *rpfB* RNA switch.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Galina Mukamolova and Jeffrey Green for helpful discussions and Finn Werner for critical reading of the manuscript.

## FUNDING

Medical Research Council New Investigator Award [MR/L018519/1 to K.B.A.]; PhD bursary (to S.S.). Funding for open access charge: University College London.

*Conflict of interest statement.* None declared.

## REFERENCES

- Rosser, A., Stover, C., Pareek, M. and Mukamolova, G.V. (2017) Resuscitation-promoting factors are important determinants of the pathophysiology in *Mycobacterium tuberculosis* infection. *Crit. Rev. Microbiol.*, **43**, 621–630.
- Chao, M.C. and Rubin, E.J. (2010) Letting sleeping dogs lie: does dormancy play a role in tuberculosis? *Annu. Rev. Microbiol.*, **64**, 293–311.
- Kana, B.D. and Mizrahi, V. (2010) Resuscitation-promoting factors as lytic enzymes for bacterial growth and signaling. *FEMS Immunol. Med. Microbiol.*, **58**, 39–50.
- Tufariello, J.M., Mi, K., Xu, J., Manabe, Y.C., Kesavan, A.K., Drumm, J., Tanaka, K., Jacobs, W.R. Jr and Chan, J. (2006) Deletion of the *Mycobacterium tuberculosis* resuscitation-promoting factor Rv1009 gene results in delayed reactivation from chronic tuberculosis. *Infect. Immun.*, **74**, 2985–2995.
- Turapov, O., Glenn, S., Kana, B., Makarov, V., Andrew, P.W. and Mukamolova, G.V. (2014) The in vivo environment accelerates generation of Resuscitation-promoting factor-dependent mycobacteria. *AJRCCM*, **190**, 1455–1457.
- Gupta, R.K., Srivastava, B.S. and Srivastava, R. (2010) Comparative expression analysis of *rpf*-like genes of *Mycobacterium tuberculosis* H37Rv under different physiological stress and growth conditions. *Microbiology*, **156**, 2714–2722.
- Tan, S., Sukumar, N., Abramovitch, R.B., Parish, T. and Russell, D.G. (2013) *Mycobacterium tuberculosis* responds to chloride and pH as synergistic cues to the immune status of its host cell. *PLoS Pathog.*, **9**, e1003282.
- Chatterjee, A., Sharma, A.K., Mahatha, A.C., Banerjee, S.K., Kumar, M., Saha, S., Basu, J. and Kundu, M. (2018) Global mapping of MtrA-binding sites links MtrA to regulation of its targets in *Mycobacterium tuberculosis*. *Microbiology*, **164**, 99–110.
- Barquist, L. and Vogel, J. (2015) Accelerating discovery and functional analysis of small RNAs with new technologies. *Annu. Rev. Genet.*, **49**, 367–394.
- Dar, D., Shamir, M., Mellin, J.R., Koutero, M., Stern-Ginossar, N., Cossart, P. and Sorek, R. (2016) Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science*, **352**, aad9822.
- Dersch, P., Khan, M.A., Muhlen, S. and Gorke, B. (2017) Roles of regulatory RNAs for antibiotic resistance in bacteria and their potential value as novel drug targets. *Front. Microbiol.*, **8**, 803.
- Kavita, K., de Mets, F. and Gottesman, S. (2017) New aspects of RNA-based regulation by Hfq and its partner sRNAs. *Curr. Opin. Microbiol.*, **42**, 53–61.
- Serganov, A. and Nudler, E. (2013) A decade of riboswitches. *Cell*, **152**, 17–24.
- Wagner, E.G. and Romby, P. (2015) Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Adv. Genet.*, **90**, 133–208.
- Breaker, R.R. (2012) Riboswitches and the RNA world. *Cold Spring Harbor Perspect. Biol.*, **4**, a003566.
- Sherwood, A.V. and Henkin, T.M. (2016) Riboswitch-mediated gene regulation: novel RNA architectures dictate gene expression responses. *Annu. Rev. Microbiol.*, **70**, 361–374.
- McCown, P.J., Corbino, K.A., Stav, S., Sherlock, M.E. and Breaker, R.R. (2017) Riboswitch diversity and distribution. *RNA*, **23**, 995–1011.
- Breaker, R.R. (2011) Prospects for riboswitch discovery and analysis. *Mol. Cell*, **43**, 867–879.
- Colameco, S. and Elliot, M.A. (2017) Non-coding RNAs as antibiotic targets. *Biochem. Pharmacol.*, **133**, 29–42.
- Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R., Floden, E.W., Gardner, P.P., Jones, T.A., Tate, J. et al. (2015) Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.*, **43**, D130–D137.
- Arnvig, K.B. and Young, D.B. (2012) Regulatory RNAs in *Mycobacterium tuberculosis*. *RNA Biol.*, **9**, 1–10.
- Nelson, J.W., Sudarsan, N., Furukawa, K., Weinberg, Z., Wang, J.X. and Breaker, R.R. (2013) Riboswitches in eubacteria sense the second messenger c-di-AMP. *Nat. Chem. Biol.*, **9**, 834–839.
- Shine, J. and Dalgarno, L. (1974) The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. U.S.A.*, **71**, 1342–1346.
- Ravagnani, A., Finan, C.L. and Young, M. (2005) A novel firmicute protein family related to the actinobacterial resuscitation-promoting factors by non-orthologous domain displacement. *BMC Genomics*, **6**, 39.
- Snapper, S.B., Melton, R.E., Mustafa, S., Kieser, T. and Jacobs, W.R. Jr (1990) Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol. Microbiol.*, **4**, 1911–1919.
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E. 3rd et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, **393**, 537–544.
- Arnvig, K.B., Comas, I., Thomson, N.R., Houghton, J., Boshoff, H.I., Croucher, N.J., Rose, G., Perkins, T.T., Parkhill, J., Dougan, G. et al. (2011) Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of *Mycobacterium tuberculosis*. *PLoS Pathog.*, **7**, e1002342.
- Moore, A., Riesco, A.B., Schwenk, S. and Arnvig, K.B. (2017) Expression, maturation and turnover of DrrS, an unusually stable, DosR regulated small RNA in *Mycobacterium tuberculosis*. *PLS One*, **12**, e0174079.
- Arnvig, K.B. and Young, D.B. (2009) Identification of small RNAs in *Mycobacterium tuberculosis*. *Mol. Microbiol.*, **73**, 397–408.

30. Grundy, F.J., Winkler, W.C. and Henkin, T.M. (2002) tRNA-mediated transcription antitermination in vitro: codon-anticodon pairing independent of the ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 11121–11126.
31. Landick, R., Wang, D. and Chan, C.L. (1996) Quantitative analysis of transcriptional pausing by *Escherichia coli* RNA polymerase: his leader pause site as paradigm. *Methods Enzymol.*, **274**, 334–353.
32. Kingsford, C.L., Ayanbule, K. and Salzberg, S.L. (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol.*, **8**, R22.
33. Will, S., Joshi, T., Hofacker, I.L., Stadler, P.F. and Backofen, R. (2012) LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA*, **18**, 900–914.
34. Cortes, T., Schubert, O.T., Rose, G., Arnvig, K.B., Comas, I., Aebersold, R. and Young, D.B. (2013) Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. *Cell Rep.*, **5**, 1121–1131.
35. Dejesus, M.A., Sacchettini, J.C. and Ioerger, T.R. (2013) Reannotation of translational start sites in the genome of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb.)*, **93**, 18–25.
36. Smollett, K.L., Fivian-Hughes, A.S., Smith, J.E., Chang, A., Rao, T. and Davis, E.O. (2009) Experimental determination of translational start sites resolves uncertainties in genomic open reading frame predictions - application to *Mycobacterium tuberculosis*. *Microbiology*, **155**, 186–197.
37. Sharma, A.K., Chatterjee, A., Gupta, S., Banerjee, R., Mandal, S., Mukhopadhyay, J., Basu, J. and Kundu, M. (2015) MtrA, an essential response regulator of the MtrAB two-component system, regulates the transcription of resuscitation-promoting factor B of *Mycobacterium tuberculosis*. *Microbiology*, **161**, 1271–1281.
38. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
39. Connolly, K., Rife, J.P. and Culver, G. (2008) Mechanistic insight into the ribosome biogenesis functions of the ancient protein KsgA. *Mol. Microbiol.*, **70**, 1062–1075.
40. Eoh, H., Brennan, P.J. and Crick, D.C. (2009) The *Mycobacterium tuberculosis* MEP (2C-methyl-d-erythritol 4-phosphate) pathway as a new drug target. *Tuberculosis (Edinb.)*, **89**, 1–11.
41. Richards, J.P. and Ojha, A.K. (2014) *Mycobacterial* biofilms. *Microbiol. Spectr.*, **2**, doi:10.1128/microbiolspec.MGM2-0004-2013.
42. Czyz, A., Mooney, R.A., Iaconi, A. and Landick, R. (2014) *Mycobacterial* RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *mBio*, **5**, e00931.
43. Steinert, H., Sochor, F., Wacker, A., Buck, J., Helmling, C., Hiller, F., Keyhani, S., Noeske, J., Grimm, S., Rudolph, M.M. *et al.* (2017) Pausing guides RNA folding to populate transiently stable RNA structures for riboswitch-based transcription regulation. *Elife*, **6**, e21297.
44. Zakhm, F., Aouane, O., Ussery, D., Benjouad, A. and Ennaji, M.M. (2012) Computational genomics-proteomics and Phylogeny analysis of twenty one mycobacterial genomes (*Tuberculosis* & non *Tuberculosis* strains). *Microb. Inform. Exp.*, **2**, 7.
45. Wang, J. and Behr, M.A. (2014) Building a better bacillus: the emergence of *Mycobacterium tuberculosis*. *Front. Microbiol.*, **5**, 139.
46. Tufariello, J.M., Jacobs, W.R. Jr and Chan, J. (2004) Individual *Mycobacterium tuberculosis* resuscitation-promoting factor homologues are dispensable for growth in vitro and in vivo. *Infect. Immun.*, **72**, 515–526.