A *Mycobacterium tuberculosis* Mbox controls a conserved, small upstream ORF via a translational expression platform and Rho-dependent termination of transcription

Alexandre D'Halluin[1, 2], Terry Kipkorir[1, 3], Catherine Hubert[1], Declan Barker[1], Kristine Arnvig[1 *]

[1] Structural and Molecular Biology, University College London, London WC1E 6BT U.K.

[2] Curent location: EGM CNRS, Université Paris-Cité, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France

[3] Current location: London School of Hygiene and Tropical Diseases, Department of Infection Biology, TB Centre, London WC1E 7HT U.K.

* Corresponding author: k.arnvig@ucl.ac.uk

ABSTRACT

Magnesium is vital for bacterial survival, and its homeostasis is tightly regulated. Intracellular pathogens like *Mycobacterium tuberculosis* (Mtb) often face host-mediated magnesium limitation, which can be counteracted by upregulating the expression of Mg^{2+} transporters. This upregulation may be via Mg^{2+}-sensing regulatory RNA such as the *Bacillus subtilis ykoK* Mbox riboswitch, which acts as a transcriptional "OFF-switch" under high Mg^{2+} conditions. Mtb encodes two Mbox elements with strong similarity to the *ykoK* Mbox.

In the current study, we characterize the Mbox encoded upstream of the Mtb *pe20* operon, which is required for growth in low Mg^{2+}/low pH. We show that this switch operates via a translational expression platform and Rho-dependent transcription termination, which is the first such case reported for an Mbox. Moreover, we show that the switch directly controls a small ORF encoded upstream of *pe20*. We have annotated this highly conserved uORF *rv1805A*, but its role remains unclear. Interestingly, a homologous gene exists outside the Mbox-regulated context, suggesting functional importance beyond magnesium stress.

Overall, this study uncovers a dual mechanism of riboswitch-regulation in Mtb, combining translational control with Rho-mediated transcription termination. These findings expand our understanding of RNA-based gene regulation in mycobacteria, with implications for pathogenesis and stress adaptation.

INTRODUCTION

Magnesium is required for a wide range of cellular functions in all domains of life and the most abundant divalent cation in living cells (Smith et al. 1998). In bacteria, these functions include cell wall integrity, biofilm formation, macromolecular metabolism and -function, making magnesium homeostasis essential (Thomas and Rice 2014; Subramani et al. 2016; Yamagami et al. 2021; Chatterjee et al. 2024). This represents an extra challenge for intracellular pathogens as host immune responses include mechanisms for restricting access to magnesium in certain cellular compartments such as the phagosome (Forbes and Gros 2001; Pokorzynski and Groisman 2023). To counteract these defence mechanisms, pathogens express an array of transporters to ensure adequate Mg^{2+} uptake. At least four types of magnesium channels and transporters regulate and maintain essential Mg^{2+} levels in prokaryotes: CorA, CorB/C, MgtA/B and MgtE (Franken et al. 2022). *Mycobacterium*

tuberculosis (Mtb) encodes CorA (Rv1239) and MgtE (Rv0362) however, Mtb does not encode homologues of MgtA/B, making the function of MgtC elusive (Alix and Blanc-Potard 2007).

Riboregulated, i.e. RNA-based, stress-responses are widespread in bacteria, with small RNAs and riboswitches being the most prominent elements. Riboswitches are located in the 5' leader regions of mRNAs regulating gene expression in *cis*; they are composed of a highly structured ligand binding aptamer domain and an expression platform. The latter exerts gene expression control by either modulating premature termination of RNA Polymerase (RNAP) and/or restricting access of the ribosome to the Ribosome Binding Site (RBS) of the mRNA (Salvail and Breaker 2023). Nonpermissive control mechanisms may involve the formation of intrinsic terminators, unmasking of Rho-binding sites or occlusion of Shine-Dalgarno (SD) sequence of the downstream Open Reading Frame (ORF). The latter may in addition be associated with Rho-dependent termination of transcription within the ORF (Salvail and Breaker 2023). Binding of the specific ligand can either allow ("ON-switch") or inhibit ("OFF-switch") expression of the downstream gene (Breaker 2018; Kavita and Breaker 2023; Schwenk and Arnvig 2018). The genes regulated by riboswitches are often, but not always, involved in the metabolism or transport of the cognate ligand (Kavita and Breaker 2023; Roth and Breaker 2009; Sherlock and Breaker 2020). Riboswitch ligands range from sugars, amino acids, nucleotides and cofactors to metal ions including Mg^{2+} (Breaker 2022; Barrick et al. 2004; Dann et al. 2007; Mccown et al. 2017). A magnesium-sensing riboswitch, referred to as Mbox, was first discovered in the *Bacillus subtilis ykoK* gene encoding a MgtE-type magnesium transporter (Barrick et al. 2004; Ramesh and Winkler 2010; Townsend et al. 1995). The Mbox is a transcriptional 'OFF-switch'; magnesium binding to the aptamer leads to conformational changes of the RNA and the formation of an intrinsic terminator preventing *ykoK* expression. At low Mg^{2+} concentrations, the absence of the terminator is permissive to *ykoK* expression which facilitates increased Mg^{2+} uptake (Ramesh and Winkler 2010).

Successful infection by Mtb requires its sensing of, and adaptation to, multiple micro-environments including different types of macrophages and their subcellular compartments such as phagosomes(Chandra et al. 2022; Samuels et al. 2022; Sholeye et al. 2022). Mtb has evolved mechanisms to either escape this organelle or to endure the hostile environment within by a range of adaptive responses (Ehrt et al. 2018; Ernst 2012; Huang et al. 2019). Riboswitches are likely to play a role in this adaptation by directly sensing host environments via specific metabolites. Several Mtb riboswitches have been predicted (Rfam RF00380) and

their expression validated by RNA-seq, Term-seq, inline probing and functional assays (Nawrocki et al. 2015; Arnvig et al. 2011; D'Halluin et al. 2023; Kipkorir et al. 2024a; Kolbe et al. 2020; Kipkorir et al. 2024b). These include two predicted Mbox aptamers upstream of Rv1535 and Rv1806 (*pe20 locus*), respectively. Both *loci* are upregulated in low magnesium (Walters et al. 2006), and in a recent study Kolbe et al. demonstrated that magnesium-dependent control resides within the leader, not the promoter of *pe20* (Kolbe et al. 2020). Moreover, the two Mtb aptamers show a high level of structural similarity to the *Bacillus subtilis ykoK* aptamer, although some results suggest that these interact differently with divalent cations including Mg^{2+} (Bahoua et al. 2021). The *pe20 locus* (encoding PE20, PPE31, PPE32, PPE33 Rv1810 and MgtC) is associated with magnesium homeostasis and acid stress, and PE20-PPE31 have been shown to be necessary for maintaining growth in a combination of low Mg^{2+} and low pH, conditions that mimic the phagosomal environment (Walters et al. 2006; Wang et al. 2020). The function of Rv1535 remains unknown.

We recently mapped premature termination of transcription in Mtb at genome-scale and identified hundreds of RNA leaders with an abundance of potential new riboswitches and translated small upstream ORFs (uORFs) (D'Halluin et al. 2023). We validated predicted riboswitches and demonstrated that both Mtb Mbox leaders were associated with premature, Rho-dependent termination of transcription upstream of the annotated ORFs (D'Halluin et al. 2023).

Here, we show that *pe-ppe* associated Mboxes are widely conserved across *Mycobacterium* and are co-transcribed with *mgtC* in Mtb. The Mtb Mbox upstream of *pe20* is unusual as it combines a translational expression platform with a Rho-dependent transcription terminator. This is to our knowledge the first translational Mbox to be described. Using translational reporter fusion constructs we show that a conserved uORF located between the Mbox and *pe20* is highly expressed. This peptide is highly conserved in the context of Mboxes across the *Mycobacterium* genus. While its function remains opaque, a paralogue of this ORF is expressed from an additional, Mbox-independent *locus* in Mtb, supporting the biological and regulatory importance of this peptide.

RESULTS

Conservation and genomic context of *M. tuberculosis* Mboxes

Two Mbox aptamers have been identified within the Mtb H37Rv genome (Rfam RF00380). We used these sequences to predict their structures and compare these to the Mbox consensus structure from Rfam. The results, shown in Figure 1A, indicate a high degree of similarity between the two Mtb aptamers and the *B. subtilis ykoK* element, suggesting these are functional Mg^{2+}-sensing elements as reported in the case of *rv1535* by (Bahoua et al. 2021).

Next, we investigated the conservation of the element and its context across the *Mycobacterium* genus. Both elements have been shown to be associated with multiple uORFs, and at least one of these is translated (u2, D'Halluin *et al*., 2023). Based on a phylogenetic analysis of the two Mtb elements and Mboxes from other species, we identified four classes of mycobacterial Mboxes, represented by *pe20*-type, manganese-type (Mn^{2+}-type), *rv1535*-type and *mgtE*-type elements, respectively; these classes are further supported by a conserved gene synteny (Figure 1B).

The *mgtE*-type is the only Mbox found in the non-pathogenic *M. smegmatis*, and its genomic neighbourhood shows that the genes immediately downstream encode predicted metal transporters (MT) and/or associated proteins (e.g. MgtE), or proteins of unknown function. The other three branches are seen across fast- and slow-growing pathogenic mycobacteria.

The first branch, the *pe20*-type is almost exclusively found upstream of multiple *pe-ppe* genes, which in Mtb*, M. ulcerans, M. marinum* and *M. kansasii* are followed by genes of unknown function (GUF) and MT (as MgtC was originally annotated as a magnesium transporter). The second Mn^{2+}-type branch includes a cluster of manganese transporter-type downstream of the riboswitch, a constellation that is not seen in Mtb. The third *rv1535*-type branch is found upstream of GUF like *rv1535* but followed by a cluster of T-box/*ileS* elements or various transferases. These results suggest the regulation of metal transporters by both Mboxes is well-conserved across mycobacteria, while the the *pe-ppe* clusters and mosaic appearance suggest insertion events may have taken place in pathogenic / slow growing species.

Rho-dependent premature termination of transcription within Mbox leaders

TSS mapping and RNA-seq suggest that the *rv1535* mRNA is monocistronic, while the *pe20* mRNA is polycistronic spanning *pe20* to *mgtC* (Arnvig et al. 2011; Cortes et al. 2013; D'Halluin et al. 2023) (Supplementary figure 1). Importantly, the entire polycistronic *pe20* operon is upregulated during growth in low magnesium controlled by the Mbox (Walters et al., 2006).

We recently mapped premature transcription termination (TTS) in Mtb genome-wide and identified two dominant TTS associated with the Mbox leaders (D'Halluin et al. 2023). TTS1062 is located ~210 nucleotides downstream of the *rv1535* TSS and 40 nucleotides downstream of the aptamer. TTS1209 is located ~185 nucleotides downstream of the *pe20* TSS and 10 nucleotides downstream of the aptamer (Figure 2A).

In both *loci,* multiple smaller peaks are flanking the TTS, suggesting a degree of flexibility in the TTS. Both TTS were located a significant distance (>200 nucleotides) upstream of their annotated ORFs, revealing the premature termination of transcription within the two leader regions, and neither were associated with canonical intrinsic terminator structures.

In our Mtb TTS mapping we predicted and validated Rho-dependent termination using RhoTermPredict (Di Salvo et al. 2019) and depletion of Rho using the Rho-DUC strain (Botella et al. 2017; D'Halluin et al. 2023). Two Rho-binding (*rut*) sites were predicted in each Mbox leader; one in each aptamer (T5468 and T6425) and one between aptamers and annotated ORFs (T5469 and T6426), while the mapped TTS1209 and TTS1062 are located between these (Table 1; Figure 2A). The calculated readthrough (RT) scores for the mapped TTS after Anhydrous Tetracyclin (ATc) induced depletion of Rho validated that transcription termination was in fact due to Rho (D'Halluin et al. 2023). To further confirm Rho-dependent, premature termination of transcription, we performed Northern Blotting on RNA from H37Rv and from Rho-depleted cultures probing for both Mboxes.

The homology between the Mbox aptamers from *rv1535* and *pe20* made it impossible to design a 5' probe that could distinguish between the two transcripts. To ensure that the signals were specific for either *pe20* or *rv1535*, we used a probe that was located 180 nucleotides into the transcripts beyond the homologous regions (supplementary figure 2) and as a result, transcripts shorter than this could not be detected.

Several strong signals between 200 and 300 nucleotides roughly corresponding to the TTS mapping suggesting multiple points of premature termination of transcription within both

leaders (Figure 2B). In H37Rv and in Rho-DUC time 0, we observed limited readthrough beyond 300 nucleotides for *pe20*, while *rv1535* displayed multiple larger signals primarily around 400 nucleotides consistent with the TTS pattern. Depletion of Rho led to an increase in larger transcripts for *pe20* suggesting increased readthrough i.e. reduced termination. In contrast, the *rv1535* termination pattern changed only marginally over the Rho depletion time course, suggesting that Rho plays a greater role for *pe20* regulation as compared to *rv1535*.

The regions downstream of the Mbox aptamers harbour multiple uORFs

Ribosome profiling demonstrates that the regions between the Mbox aptamers and the two annotated open reading frames (ORF)s (*rv1535* and *pe20*, respectively) are bound by ribosomes in agreement with on-going translation upstream of the annotated genes (D'Halluin et al. 2023; Smith et al. 2022; Sawyer and Cortes 2022). Sequence alignment of *rv1535* and *pe20* leaders with other *pe20*-type leaders indicated several regions of conservation including near-identical SD sequences located at the end of the aptamer (SD1) and a second, highly conserved SD (SD2) further downstream. The ORF downstream of SD1 (upstream ORF1/uORF1) shows poor conservation, while the uORF downstream of SD2 (uORF2) is highly conserved (Supplementary figure 2). Moreover, we have previously shown that uORF2 from both loci is expressed (D'Halluin et al. 2023).

To characterize the relationship between the *pe20* Mbox and the two uORFs, we first investigated expression using translational *lacZ*-fusions. All constructs included the 5' leader from the TSS and were gradually extended downstream to the end of uORF1 (Mbox-*uORF1-lacZ*); the end of uORF2 (Mbox-*uORF2-lacZ*), or the start codon of *pe20* ORF (Mbox-*pe20::lacZ*), respectively. All were fused in-frame to *lacZ*, expressed from a heterologous, constitutive promoter and integrated into the *M. smegmatis* genome in single copy (Figure 3). Next, we performed β-galactosidase (β-gal) assays, which showed that Mbox-*uORF2-lacZ* expression was >10 fold higher than Mbox-*pe20::lacZ* (~650 Miller Units compared to 60 Miller Units), while Mbox-*uORF1-lacZ* was only slightly higher than the background (Figure 3B). To validate the start codons of the two uORFs, we mutated each to non-start codons (GTG to GTC and ATG to ACG for *uORF1* and *uORF2*, respectively. This reduced β-gal activity significantly in both constructs, suggesting reduced expression in support of the annotated translation start sites, although Mbox-*uORF2-lacZ* expression was still higher than Mbox-*pe20::lacZ* expression (Supplementary figure 3).

As the translation initiation region (TIR) for uORF1 and uORF2 (i.e. SD1 and SD2 and their distances to the start codons) were almost identical, and we had not observed any premature TTS in the region, we reasoned that the coding region of uORF1 was responsible for the lower β-gal activity. To explore this possibility, we deleted the majority of uORF1 from the Mbox-*uORF1-lacZ* construct except the first two codons (Mbox- uORF1Δ-*lacZ*) and measured *lacZ* expression.

The results, shown in Figure 3B indicate that expression of this truncated uORF1 was 2.5-fold higher than that of the full-length uORF1, suggesting that the uORF1 coding region did indeed suppress β-gal activity. As uORF1 contains several rare (≤5/1000) codons, i.e. TGC, CCT, TGC, TGT, TGT, AGG, AGG (Figure 3C), we assume that this was due to poor translation elongation, but alternative explanations such as the uORF1 peptide interfering with β-gal activity cannot be ruled out at this stage.

The *pe20* Mbox operates via a translational expression platform

In conjunction, the Rho-dependent premature termination of transcription, the highly conserved SDs at the end of the aptamer, located upstream of a well-expressed conserved uORF made us speculate that the *pe20* Mbox operates via a translational expression platform.

A functional translational expression platform requires the potential for the SD to be masked, e.g. by a pyrimidine-rich region (an αSD) that in turn can be sequestered by an ααSD under different conditions.

We identified such a region approximately halfway between SD1 and SD2. This αSD and its flanking regions have the potential to pair with the entire translation initiation region (TIR) of uORF2 (shown in blue in Figure 4) or alternatively, with the aptamer-associated SD1 and its flanks (yellow in Figure 4).

To explore this hypothesis further, we measured uORF2 expression after introducing mutations that could interfere with the proposed interactions. One was the abolishing the uORF1 start codon, the rationale being that this would partially unmask SD1 thereby favouring the SD1-αSD interaction, leading to an increase in uORF2-*lacZ* expression

Similarly, deleting the αSD should also lead to higher expression of uORF2, as SD2 would no longer be sequestered. The results, shown in Figure 5A indicate a moderate (~1.3-fold), but significant increase in expression, when the start codon of uORF1 was changed (Mbox-*uORF2*^{G184C}-*lacZ*) and a larger (~2-fold) increase in expression, when the αSD was deleted

(Mbox-ΔαSD-*uORF2-lacZ*). Combining the two mutations did not result in an additive effect, suggesting they involved the same mechanism (Figure 5A).

These results support a model in which SD1 (ααSD), αSD and SD2 interact to control the expression of uORF2-*lacZ*.

To further validate this model, we assessed the contribution of each element by gradually extending the region between uORF2 and the Mbox in uORF2-*lacZ* fusions (Figure 5B). The SD2-uORF2 construct displayed β-gal expression levels of ~500 Miller Units, and the addition of the αSD motif reduced the β-gal expression by ~35%. However, a further extension including the ααSD motif led to a substantial increase in uORF2 expression. This is likely due to the unmasking of SD2 and corroborates our model of a translational expression platform controlling uORF2 expression.

Our results support a model in which uORF2 is controlled by a translational Mbox riboswitch combined with Rho-dependent termination of transcription. Based on sequence homology, we propose that the *rv1535* Mbox likewise operates via a translational expression platform. To the best of our knowledge, these are the first examples of an Mbox translational expression platform and Rho-dependent termination of transcription.

An Mbox-independent homologue of uORF2 encoded in a separate Mtb *locus*

Considering the high conservation between uORF2 in the *rv1535* and *pe20 loci*, we carried out deeper sequence searches and identified a third homologue of the uORF2 region including its SD downstream of the *gca-gmhA-gmhB-hddA* operon. This *locus* has been acquired by horizontal gene transfer (Becq et al. 2007) and the uORF2 homologue annotated as Rv0115A (Figure 6A).

We identified two TSS and associated promoter motifs within this *locus*. The first drives the transcription of the *gca-hddA* operon, which terminates downstream of *hddA* (D'Halluin et al. 2023). The second drives the transcription of *rv0115A*, and potentially also a second ORF, *rv0115B*. The *gca* and *rv0115A* promoters have similar unusual motifs in the form of an AANCAT -10 hexamer, an extended -10 motif (TGN), a perfect -35 hexamer and in the case of *cga*, a Cytidine TSS (Figure 6B).

A further alignment of the promoter regions from -120 to a few basepairs downstream of the mapped TSS, had a remarkable similarity more than 100 basepairs upstream of the TSS suggestive of a gene duplication event (Figure 6B). There are no Mbox elements associated

with the leaders of these genes, but according to TBDB (http://tbdb.bu.edu/tbdb_sysbio/MultiHome.html) both promoters include a binding site for PhoP, linking expression to pH stress (Abramovitch et al. 2011).

Alignment of uORF2 peptide homologues including Rv0115A across mycobacterial species reveals a well-conserved N-terminal region, including a universally conserved Proline residue (Figure 7). This peptide is specific for mycobacteria, which indicates that uORF2 peptides and their homologues have functions uniquely associated with this genus. Based on this finding, we suggest renaming uORF2 from the *pe20* operon Rv1805A.

No evident role for Rv1805A in biofilm formation during magnesium stress

Realising the ubiquitous presence of Rv1805A homologues, we sought to find a role for this peptide. PE20 and PPE31 are necessary for Mtb growth in conditions of low Mg^{2+} combined with low pH (Wang et al. 2020). To probe a potential role of uORF2 in this process, we exploited the fact that magnesium is required for biofilm formation in mycobacteria (Chatterjee et al. 2024) and leveraged the trick that *Mycobacterium smegmatis*, a closely related species, has no homolog of *pe20* locus.

In agreement with literature, the growth and biofilm formation of *M. smegmatis* were compromised in low Mg^{2+}, and that this phenotype was exacerbated at acidic pH values (Figure 8). We tested whether the expression of *pe20-ppe31* or *rv1805A-pe20-ppe31* might rescue this phenotype by transforming *M. smegmatis* with plasmids expressing the cognate genes. The results in figure 8 indicate no visible difference between strains expressing *pe20-ppe31* with or without *rv1805A* or *rv0115A*; further investigations are required to identify a role of this peptide and its homologues in mycobacterial biology.

DISCUSSION

In the current study we have revealed a novel complex riboregulatory system which controls *pe20* gene expression in Mtb. Our results show that premature termination occurring in the 5' leader of *pe20* (and *rv1535*) relies on Rho-dependent termination of transcription (D'Halluin et al. 2023). Moreover, the *pe20* Mbox contributes a translational expression platform, where the translation initiation region including the SD of the first gene in the operon can be sequestered by an αSD motif. This is also, to the best of our knowledge, the first example of a translationally controlled Mbox. This type of control is consistent with the scarcity of intrinsic

terminators in Mtb, and it echoes the finding that a mycobacterial T-box is the only known T-box with a translational expression platform (Sherwood et al. 2018) Finally, we identified a highly conserved uORF (*rv1805A*), which is the primary regulated ORF within the *pe20* operon and, based on homology, likely also in the *rv1535* operon.

Expression of the *pe20* operon is suppressed *via* its leader by high Mg^{2+} concentrations similar to the Mbox controlled *ykoK* gene in *B. subtilis* (Walters et al. 2006; Ramesh and Winkler 2010; Kolbe et al. 2020). *pe20* and *ppe31* are critical for magnesium uptake in low-pH/low-magnesium conditions suggesting that the gene products form (part of) a magnesium transporter (Feng et al. 2021; Wang et al. 2020). We propose that the Mbox–*rv1805A* module acts as the key regulatory gate, enabling expression of the magnesium-responsive PE/PPE transporter complex only under specific environmental conditions, such as low Mg^{2+} and acidic pH.

The structure of the *pe20* operon, including the presence of *mgtC* raises questions about its ancestry. Given what is known about *pe-ppe* gene expansion (Fishbein et al. 2015) and what we have observed in other riboswitch-controlled *pe/ppe loci (i.e. the Cbl-ppe2-cobQ locus*, and the PE-containing uORF recently identified downstream of the Mtb glycine riboswitch (D'Halluin et al. 2023; Kipkorir et al. 2024a), it is tempting to speculate that an early *pe*(*-ppe*) element invaded the current *pe20 locus* and subsequently expanded whereby Rv1805A became the first gene in this operon.

A recent study suggests that the *rv1535* Mbox, and by extension likely also the *pe20* Mbox associates with other divalent cations in addition to Mg^{2+} (Bahoua et al. 2021), while Kolbe et al. have demonstrated strong Mg^{2+}-dependent control of *pe20* expression via its leader (Kolbe et al. 2020). Regardless of the identity of the cognate ligand, our results suggest an ability to alternate between two structures: a non-permissive (ligand-bound) structure that sequesters SD1, allowing αSD/αTIR to pair with SD2/TIR thereby preventing translation of uORF2/Rv1805A. This could in turn lead Rho-dependent termination of transcription, which will affect the entire operon (Hao et al. 2021; Molodtsov et al. 2023). We note, however, that according to Term-seq results, the primary TTS is located upstream of *rv1805A*, suggesting that Rho-dependent termination does not depend on translation of this ORF. An alternative explanation of our results could therefore be that the pyrimidine-rich region that we have annotated as αSD, might act as a Rho-binding (*rut*) site that would be masked by translation of uORF1. Deleting this region increased expression 2.5-fold, likely due to reduced termination

of transcription or by unmasking of SD2 or both. The marginal increase in the expression of uORF2 in the context of an untranslated uORF1 (Mbox-*rv1805A*^{G184C}-*lacZ*, Figure 5) and the conservation of the ααSD-αSD interaction, suggests a functional interaction. The two models are not mutually exclusive, and further experiments will elucidate the structural and mechanistic basis underlying the regulation. Along the same lines, we note that expression of uORF1 might affect the activity of LacZ, although this is unlikely to affect the overall conclusions.

What is the function of Rv1805A and its homologues? Given its conservation and position upstream of *pe20*, we hypothesize that Rv1805A may act as a regulatory peptide modulating the activity or assembly of the PE20–PPE31 complex. Alternatively, it may serve as a structural component of a magnesium-responsive transporter. Conservation between Rv1805A, the Rv1535 uORF2 (Rv1535A) and Rv0115A, and their associations with magnesium and pH stress suggests important roles for these peptides during infection. Future work will focus on identifying interaction partners of Rv1805A and assessing its role in magnesium uptake and stress responses.

In conclusion, our findings reveal a previously unrecognized mode of riboswitch control in Mtb, where a translational Mbox integrates with Rho-dependent termination to regulate a conserved uORF. This multilayered modus operandi underscores the sophistication of RNA-based regulation in Mtb stress adaptation.

MATERIAL AND METHODS

Strains and cultures

Strains used in this study are listed in Supplementary Table 1. *M. tuberculosis* H37Rv and *M. smegmatis* MC^2 155 were cultured on solid media Middlebrook agar 7H11 supplemented with 10% OADC (Sigma), 0.5% Glycerol and 50 µg/ml hygromycin if appropriate. Liquid cultures were done in Middlebrook 7H9 supplemented with 10% ADC (Sigma), 0.5% Glycerol, 0.05% Tween 80 and 50 µg/ml hygromycin where appropriate. Cultures were harvested at an OD_{600nm} ~0.6 for mid-log phase.

Mtb RhoDUC strain, a gift obtained from Professor Dirk Schnappinger, was grown as previously described with 50 µg/ml hygromycin, 20 µg/ml Kanamycin and 50 µg/ml zeocin (Botella et al. 2017; D'Halluin et al. 2023). When the cultures reached an OD_{600nm}~0.6,

depletion of Rho was induced using 500 ng/ml of anhydrotetracycline. Cells were harvested after 0, 1.5, 3 and 4.5 hours.

Escherichia coli DH5α was used for cloning the *lacZ* fusion reporters and were cultured on solid LB 1.5% agar supplemented with 50 µM of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) or in liquid LB supplemented with 250 µg/ml Hygromycin.

Plasmids constructions and primers

Plasmids and primers used in this study are listed in Supplementary Table 1 and 2. pIRATE plasmids, describe in D'Halluin *et al*., 2023, were used for *lacZ* translational fusion reporters and for Beta-galactosidase assay. Reporters were constructed using Gibson assemblies with oligos (Sigma) or geneBlocks (IDT) listed in Table 3 between HindIII and NcoI sites. Point mutations and deletions were generated using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs). Plasmids were cloned in *E. coli* DH5α, extracted and sequenced by Sanger sequencing. Plasmids were transformed into *M. smegmatis* by electroporation and selected on Middlebrook 7H11 agar plates containing 50 µg/ml Hygromycin.

RNA extraction and Northern Blotting

M. tuberculosis H37Rv were stopped using 37.5% of cold ice and centrifuge 10min at 5000 rpm 4°C. Total RNA was extracted as previously described using the FastRNA Pro Blue kit (MP Biomedicals) according to the manufacturer's protocol (D'Halluin et al. 2023; Arnvig et al. 2011). RNA concentration and purity was assessed using the Nanodrop 2000 (ThermoFisher), residual genomic DNA removed using Turbo DNase (ThermoFisher) and RNA integrity assessed with 2100 Bioanalyzer (Agilent). 10 µg of total RNA were separated on a denaturing 8% acrylamide:bis-acrylamide (19:1) gel and transfer to a nylon membrane. An RNA probe was synthetized using the mirVana miRNA probe synthesis kit (Ambion) to reveal the *pe20* and *rv1535* Mbox transcripts and labelled with 3µM final concentration of ^{32}P α-UTP (3000Ci/mmol; Hartmann AnalyticGmbH). Northern blots were revealed using radiosensitive screens and visualized on a Typhoon FLA 9500 phosphoimager (GEHealthcare).

Beta-galactosidase activity

M. smegmatis carrying the *lacZ* reporter fusions were cultured at OD_{600nm} ~0.6 and centrifuge 10min 5000 rpm. Pellets were washed four times in Z-buffer composed of 60mM

Na_2HPO_4, 40mM NaH_2PO_4, 10mM KCl, 1mM $MgSO_4$ and lysed using beads with the FastPrep bio-pulveriser (MP Biomedicals). The supernatant was kept after centrifugation and the protein level assessed using a Bradford yield with the BCA kit (ThermoFisher) following the manufacturer's recommendations. Beta-galactosidase were done using the Beta-galactosidase assay kit (ThermoFisher) following manufacturer's protocol. Proteins were pre-incubated for 5min at 28°C before addition of ONPG.

Biofilm formation

M. smegmatis expressing *pe20-ppe31, rv1805A-pe20-ppe31* or *rv0115A+pe20-ppe31* was grown to mid-log phase, washed in Mg^{2+}-free medium, resuspended in 1 mL of the indicated medium at OD 0.01 and seeded in 24-well plates. Plates were sealed in plastic bags and left for static incubation at 37℃ for a week. Biofilm formation was monitored every day for a week.

Folding, sequence conservation and distribution across mycobacteria

Representative genomes of several mycobacteria were selected for sequences conservation: *Mycobacterium tuberculosis* H37Rv (NB_000962), *Mycobacterium leprae* TN (AL450380), *Mycobacterium avium* K10 (NZ_CP106873), *Mycobacterium kansasii* Kuro I (AP023343), *Mycobacterium ulcerans* ATCC33728 (NZ_AP017624), *Mycobacterium marinum* M (CP000854), *Mycobacterium abscessus* ATCC19977 (NC_010397), *Mycobacterium haemophilum* DSM 44634 (CP011883) and *Mycobacterium smegmatis* MC^2155 (NZ_CP009494).

The aptamer sequences of the Mboxes were extracted from RFam database (Rfam RF00380) (Nawrocki et al. 2015) and extended to the next annotated ORF. DNA and peptidic sequences were aligned using ClustalW (Thompson et al. 1994), and alignment strengthen using T-coffee (Notredame et al. 2000). The conservation of uORF2 across mycobacteria was determined using Blast (Altschul et al. 1990) and amino acid sequences aligned using ClustalW (Thompson et al. 1994) and Chimera (Meng et al. 2006). The phylogenetic tree was generated by Clustal Omega using the sequences from the aptamer sequence to the start codon of the next in frame annotated ORF (Sievers et al. 2011). Aptamer secondary structures were predicted using the RNAstructure Web Server for RNA Secondary Structure Prediction (Reuter

and Mathews 2010). The resulting Connectivity Table (CT) file was then uploaded to RNAcanvas (Johnson and Simon 2023) for visualization and structural editing

FUNDING

KBA was funded by The UK Medical Research Council grants MR/S009647/1 and MR/X009211/1. TK was funded by The Newton International Fellowship grants (NIF\R1\180833 & NIF\R5A\0035), the Wellcome Institutional Strategic Support Fund grant (204841/Z/16/Z), and the Wellcome Early Career Award (225605/Z/22/Z). The views expressed are those of the author(s) and not necessarily those of the funders. The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication.

AUTHOR CONTRIBUTIONS

AD, TK, KBA designed the study. AD, TK, CH, KBA performed experiments. AD, TK, KBA performed data analysis and wrote the manuscript.

ACKNOWLEDGEMENTS

We are grateful for Finn Werner's comments on the manuscript

REFERENCES

Abramovitch, R. B., Rohde, K. H., Hsu, F. F., & Russell, D. G. (2011). aprABC: a Mycobacterium tuberculosis complex-specific locus that modulates pH-driven adaptation to the macrophage phagosome. *Molecular microbiology*, *80*(3), 678–694. https://doi.org/10.1111/J.1365-2958.2011.07601.X

Alix, E., & Blanc-Potard, A. B. (2007). MgtC: a key player in intramacrophage survival. *Trends in Microbiology*, *15*(6), 252–256. https://doi.org/10.1016/j.tim.2007.03.007

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

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., & Young, D. B. (2011). Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis. *PLoS pathogens*, *7*(11). https://doi.org/10.1371/JOURNAL.PPAT.1002342

Bahoua, B., Sevdalis, S. E., & Soto, A. M. (2021). Effect of Sequence on the Interactions of Divalent Cations with M-Box Riboswitches from Mycobacterium tuberculosis and Bacillus subtilis. *Biochemistry*, *60*(37), 2781–2794. https://doi.org/10.1021/ACS.BIOCHEM.1C00371,

Barrick, J. E., Corbino, K. A., Winkler, W. C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., Wickiser, J. K., & Breaker, R. R. (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*(17), 6421. https://doi.org/10.1073/PNAS.0308014101

Becq, J., Gutierrez, M. C., Rosas-Magallanes, V., Rauzier, J., Gicquel, B., Neyrolles, O., & Deschavanne, P. (2007). Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli. *Molecular Biology and Evolution*, *24*(8), 1861–1871. https://doi.org/10.1093/MOLBEV/MSM111,

Botella, L., Vaubourgeix, J., Livny, J., & Schnappinger, D. (2017). Depleting Mycobacterium tuberculosis of the transcription termination factor Rho causes pervasive transcription and rapid death. *Nature Communications*. https://doi.org/10.1038/ncomms14731

Breaker, R. R. (2018). Riboswitches and Translation Control. *Cold Spring Harbor Perspectives in Biology*, *10*(11). https://doi.org/10.1101/CSHPERSPECT.A032797

Breaker, R. R. (2022). The Biochemical Landscape of Riboswitch Ligands. *Biochemistry*, *61*(3), 137–149. https://doi.org/10.1021/ACS.BIOCHEM.1C00765,

Chandra, P., Coullon, H., Agarwal, M., Goss, C. W., & Philips, J. A. (2022). Macrophage global metabolomics identifies cholestenone as host/pathogen cometabolite present in human Mycobacterium tuberculosis infection. *Journal of Clinical Investigation*, *132*(3). https://doi.org/10.1172/JCI152509,

Chatterjee, D., Daya Manasi, A. R., Rastogi, S. K., Panda, A. P., Biju, B., Bhattacharyya, D., & Ghosh, A. S. (2024). Involvement of CorA of Mycobacterium smegmatis in exerting intrinsic resistance towards structurally unrelated antibiotics. *Journal of Applied Microbiology*, *135*(12). https://doi.org/10.1093/JAMBIO/LXAE298,

Cortes, T., Schubert, O. T., Rose, G., Arnvig, K. B., Comas, I., Aebersold, R., & Young, D. B. (2013). Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in Mycobacterium tuberculosis. *Cell reports*, *5*(4), 1121–1131. https://doi.org/10.1016/j.celrep.2013.10.031

Dann, C. E., Wakeman, C. A., Sieling, C. L., Baker, S. C., Irnov, I., & Winkler, W. C. (2007). Structure and mechanism of a metal-sensing regulatory RNA. *Cell*, *130*(5), 878–892. https://doi.org/10.1016/J.CELL.2007.06.051

D'Halluin, A., Polgar, P., Kipkorir, T., Patel, Z., Cortes, T., & Arnvig, K. B. (2023). Premature termination of transcription is shaped by Rho and translated uORFS in Mycobacterium tuberculosis. *iScience*, *26*(4). https://doi.org/10.1016/J.ISCI.2023.106465

Di Salvo, M., Puccio, S., Peano, C., Lacour, S., & Alifano, P. (2019). RhoTermPredict: An algorithm for predicting Rho-dependent transcription terminators based on Escherichia coli, Bacillus subtilis and Salmonella enterica databases. *BMC Bioinformatics*, *20*(1), 1–11. https://doi.org/10.1186/S12859-019-2704-X/TABLES/5

Ehrt, S., Schnappinger, D., & Rhee, K. Y. (2018). Metabolic principles of persistence and pathogenicity in Mycobacterium tuberculosis. *Nature Reviews Microbiology*, *16*(8), 496–507. https://doi.org/10.1038/S41579-018-0013-4,

Ernst, J. D. (2012). The immunological life cycle of tuberculosis. *Nature Reviews Immunology*, *12*(8), 581–591. https://doi.org/10.1038/NRI3259,

Feng, S., Hong, Z., Zhang, G., Li, J., Tian, G. B., Zhou, H., & Huang, X. (2021). Mycobacterium PPE31 Contributes to Host Cell Death. *Frontiers in Cellular and Infection Microbiology*, *11*. https://doi.org/10.3389/FCIMB.2021.629836/FULL

Fishbein, S., van Wyk, N., Warren, R. M., & Sampson, S. L. (2015). Phylogeny to function: PE/PPE protein evolution and impact on Mycobacterium tuberculosis pathogenicity. *Molecular Microbiology*, *96*(5), 901–916. https://doi.org/10.1111/MMI.12981,

Forbes, J. R., & Gros, P. (2001). Divalent-metal transport by NRAMP proteins at the interface of host-pathogen interactions. *Trends in Microbiology*, *9*(8), 397–403. https://doi.org/10.1016/S0966-842X(01)02098-4

Franken, G. A. C., Huynen, M. A., Martínez-Cruz, L. A., Bindels, R. J. M., & de Baaij, J. H. F. (2022). Structural and functional comparison of magnesium transporters throughout evolution. *Cellular and Molecular Life Sciences*, *79*(8), 1–17. https://doi.org/10.1007/S00018-022-04442-8/FIGURES/7

Hao, Z., Svetlov, V., & Nudler, E. (2021). Rho-dependent transcription termination: a revisionist view. *Transcription*, *12*(4), 171–181. https://doi.org/10.1080/21541264.2021.1991773

Huang, L., Nazarova, E. V., & Russell, D. G. (2019). Mycobacterium tuberculosis : Bacterial Fitness within the Host Macrophage . *Microbiology Spectrum*, *7*(2). https://doi.org/10.1128/MICROBIOLSPEC.BAI-0001-2019,

Johnson, P. Z., & Simon, A. E. (2023). RNAcanvas: Interactive drawing and exploration of nucleic acid structures. *Nucleic Acids Research*, *51*(W1), W501–W508. https://doi.org/10.1093/NAR/GKAD302,

Kavita, K., & Breaker, R. R. (2023). Discovering riboswitches: the past and the future. *Trends in Biochemical Sciences*, *48*(2), 119–141. https://doi.org/10.1016/j.tibs.2022.08.009

Kipkorir, T., Polgar, P., Barker, D., D'Halluin, A., Patel, Z., & Arnvig, K. B. (2024). A novel regulatory interplay between atypical B12 riboswitches and uORF translation in Mycobacterium tuberculosis. *Nucleic Acids Research*, *52*(13), 7876. https://doi.org/10.1093/NAR/GKAE338

Kipkorir, T., Polgar, P., D'Halluin, A., Gap-Gaupool, B., Makarov, V. A., Mukamolova, G. V., & Arnvig, K. B. (2024). The RpfB switch is a novel B12-sensing riboswitch regulating (non-replicating) persistence in Mycobacterium tuberculosis. *bioRxiv*, 2024.07.19.603033. https://doi.org/10.1101/2024.07.19.603033

Kolbe, K., Bell, A. C., Prosser, G. A., Assmann, M., Yang, H. J., Forbes, H. E., Gallucci, S., Mayer-Barber, K. D., Boshoff, H. I., & Barry, C. E. (2020). Development and Optimization of Chromosomally-Integrated Fluorescent Mycobacterium tuberculosis Reporter Constructs. *Frontiers in Microbiology*, *11*, 591866. https://doi.org/10.3389/FMICB.2020.591866/BIBTEX

Mccown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E., & Breaker, R. R. (2017). Riboswitch diversity and distribution. *RNA*, *23*(7), 995–1011. https://doi.org/10.1261/RNA.061234.117/-/DC1

Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C., & Ferrin, T. E. (2006). Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics*, *7*. https://doi.org/10.1186/1471-2105-7-339,

Molodtsov, V., Wang, C., Firlar, E., Kaelber, J. T., & Ebright, R. H. (2023). Structural basis of Rho-dependent transcription termination. *Nature*, *614*(7947), 367–374. https://doi.org/10.1038/S41586-022-05658-1

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., & Finn, R. D. (2015). Rfam 12.0: Updates to the RNA families database. *Nucleic Acids Research*, *43*(D1), D130–D137. https://doi.org/10.1093/NAR/GKU1063,

Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, *302*(1), 205–217. https://doi.org/10.1006/jmbi.2000.4042

Pokorzynski, N. D., & Groisman, E. A. (2023). How Bacterial Pathogens Coordinate Appetite with Virulence. *Microbiology and Molecular Biology Reviews*, *87*(3). https://doi.org/10.1128/MMBR.00198-22,

Ramesh, A., & Winkler, W. C. (2010). Magnesium-sensing riboswitches in bacteria. *RNA Biology*, *7*(1), 77–83. https://doi.org/10.4161/RNA.7.1.10490

Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, *11*. https://doi.org/10.1186/1471-2105-11-129,

Roth, A., & Breaker, R. R. (2009). The structural and functional diversity of metabolite-binding riboswitches. *Annual review of biochemistry*, *78*, 305–334. https://doi.org/10.1146/ANNUREV.BIOCHEM.78.070507.135656

Salvail, H., & Breaker, R. R. (2023). Riboswitches. *Current Biology*, *33*(9), R343–R348. https://doi.org/10.1016/j.cub.2023.03.069

Samuels, A. N., Wang, E. R., Harrison, G. A., Valenta, J. C., & Stallings, C. L. (2022). Understanding the contribution of metabolism to Mycobacterium tuberculosis drug tolerance. *Frontiers in Cellular and Infection Microbiology*, *12*. https://doi.org/10.3389/FCIMB.2022.958555/PDF

Sawyer, E. B., & Cortes, T. (2022). Ribosome profiling enhances understanding of mycobacterial translation. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/FMICB.2022.976550,

Schwenk, S., & Arnvig, K. B. (2018). Regulatory RNA in Mycobacterium tuberculosis, back to basics. *Pathogens and Disease*, *76*(4). https://doi.org/10.1093/femspd/fty035

Sherlock, M. E., & Breaker, R. R. (2020). Former orphan riboswitches reveal unexplored areas of bacterial metabolism, signaling, and gene control processes. *RNA*, *26*(6), 675–693. https://doi.org/10.1261/RNA.074997.120,

Sherwood, A. V., Frandsen, J. K., Grundy, F. J., & Henkin, T. M. (2018). New tRNA contacts facilitate ligand binding in a Mycobacterium smegmatis T box riboswitch. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(15), 3894–3899. https://doi.org/10.1073/PNAS.1721254115,

Sholeye, A. R., Williams, A. A., Loots, D. T., Tutu van Furth, A. M., van der Kuip, M., & Mason, S. (2022). Tuberculous Granuloma: Emerging Insights From Proteomics and Metabolomics. *Frontiers in Neurology*, *13*, 804838. https://doi.org/10.3389/FNEUR.2022.804838

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, *7*. https://doi.org/10.1038/MSB.2011.75

Smith, C., Canestrari, J. G., Wang, A. J., Champion, M. M., Derbyshire, K. M., Gray, T. A., & Wade, J. T. (2022). Pervasive translation in Mycobacterium tuberculosis. *eLife*, *11*. https://doi.org/10.7554/ELIFE.73980

Smith, R. L., Kaczmarek, M. T., Kucharski, L. M., & Maguire, M. E. (1998). Magnesium transport in Salmonella typhimurium: Regulation of mgtA and mgtCB during invasion of epithelial and macrophage cells. *Microbiology*, *144*(7), 1835–1843. https://doi.org/10.1099/00221287-144-7-1835/CITE/REFWORKS

Subramani, S., Perdreau-Dahl, H., & Morth, J. P. (2016). The magnesium transporter A is activated by cardiolipin and is highly sensitive to free magnesium in vitro. *eLife*, *5*. https://doi.org/10.7554/ELIFE.11407

Thomas, K. J., & Rice, C. V. (2014). Revised model of calcium and magnesium binding to the bacterial cell wall. *BioMetals*, *27*(6), 1361–1370. https://doi.org/10.1007/S10534-014-9797-5/METRICS

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680. https://doi.org/10.1093/NAR/22.22.4673,

Townsend, D. E., Esenwine, A. J., George, J., Bross, D., Maguire, M. E., & Smith, R. L. (1995). Cloning of the mgtE Mg2+ transporter from Providencia stuartii and the distribution of mgtE in gram-negative and gram-positive bacteria. *Journal of Bacteriology*, *177*(18), 5350–5354. https://doi.org/10.1128/JB.177.18.5350-5354.1995,

Walters, S. B., Dubnau, E., Kolesnikova, I., Laval, F., Daffe, M., & Smith, I. (2006). The Mycobacterium tuberculosis PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. *Molecular Microbiology*, *60*(2), 312–330. https://doi.org/10.1111/J.1365-2958.2006.05102.X

Wang, Q., Boshoff, H. I. M., Harrison, J. R., Ray, P. C., Green, S. R., Wyatt, P. G., & Barry, C. E. (2020). PE/PPE proteins mediate nutrient transport across the outer membrane of Mycobacterium tuberculosis. *Science*, *367*(6482), 1147–1151. https://doi.org/10.1126/SCIENCE.AAX3072

Yamagami, R., Sieg, J. P., & Bevilacqua, P. C. (2021). Functional Roles of Chelated Magnesium Ions in RNA Folding and Function. *Biochemistry*, *60*(31), 2374–2386. https://doi.org/10.1021/ACS.BIOCHEM.1C00012,

TABLES

M-Box gene	Mapped TTS	RTP number*	*rut* site start	*rut* site end	Distance from TSS***	Distance to annotated ORF***
rv1535	1735718	T5468	1735526	1735604	17	-372
		T5469	1735816	1735894	307	-82
pe20	2047779	T6425	2047633	2047711	38	-361
		T6426	2047887	2047965	292	-107

Table 1: Mapped transcription termination sites (TTS), predicted Rho-dependent terminators (RTP) and their locations according to (D'Halluin et al. 2023).

FIGURE LEGENDS

Figure 1: Conservation of Mbox elements. A) Mbox aptamer structures from *Bacillus subtilis ykoK* and the two *M. tuberculosis* aptamers from *rv1535* and *pe20* (*rv1806*). Structures were predicted using RNAstructure Web Server (Reuteur & Mathews, 2010) and drawn by extracting the bracket-dot plot to RNAcanvas (Johnson & Simons, 2023). B) Distribution of Mboxes and

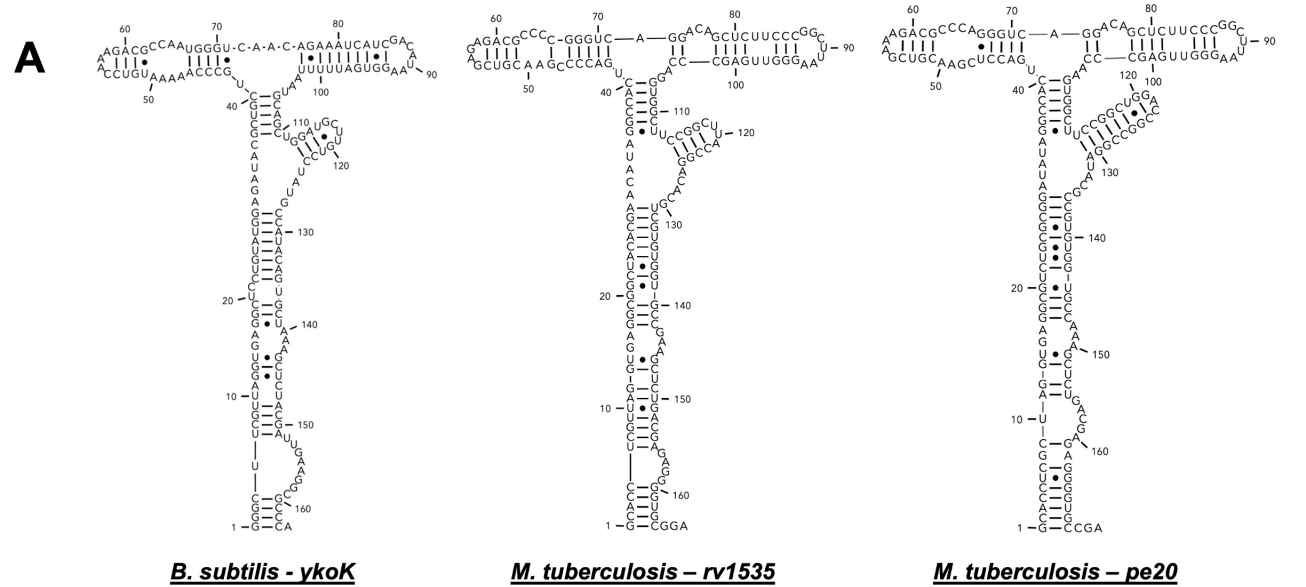
642 their associated genes in *Mycobacteria*. These can be split into the four types indicated, based
643 on the aptamer and their downstream sequences. Notably, the two *M. tuberculosis* elements
644 fall into different groups.
645
646 **Figure 2: Premature termination of transcription within Mtb Mbox *loci*.** A) The two Mbox-
647 associated genes, *rv1535* and *pe20* are shown with their respective leaders. TSS from (Cortes
648 *et al.,* 2013), Term-seq data and Transcription termination sites (TTS) from (D'Halluin *et al.,*
649 2023). Distances from TSS to dominant TTS peaks and further to the start codons of have been
650 indicated. B) Northern blot with log-phase total RNA from Mtb H37Rv and from RhoDUC
651 (Botella *et al*., 2017) following depletion of Rho. Total RNA was separated on an 8% acrylamide
652 gel, electroblotted and probed for leader sequences distinct for the two genes, approximately
653 180 nucleotides downstream of the TSS. The 5S RNA was probed as a loading control.
654
655 **Figure 3: Expression of *pe20* uORFs.** To ascertain expression of uORF1 and uORF2 from the
656 *pe20* operon, we made translational *lacZ*-fusions and measured ***β***-galactosidase (***β***gal) activity
657 of the different constructs. Experiments were done in triplicates and differences of expression
658 tested with a t-test (p-val<0.01). A) Genomic context of *pe20* and the uORFs associated (green)
659 including SD1 (yellow box) and SD2 (blue box); B) Schematic showing each reporter constructs
660 (left) and their expression in Miller units (right). uORF1D refers to a truncated version of uORF1
661 encoding only its first two codons. C) uORF1 sequence with amino acids and their codons. Start
662 codon is shown in green and rare codons (<5/1000 frequency) are shown in red.
663
664 **Figure 4: Model for a translational expression platform.** The figure shows how the translation
665 initiation region (TIR, blue) can be sequestered by base-pairing with the αTIR (orange), which
666 in turn can base-pair with the ααTIR (yellow), depending on the conformation of the aptamer.
667 Structure of the aptamer is shown on the left with part of the ααTIR shown in yellow.
668
669 **Figure 5: Testing the model for a translational expression platform.** A) Reporter constructs
670 assessing the effect of uORF1 changes on uORF2 expression; changing the start codon of
671 uORF1 to a no-start (G184C), deleting the proposed αSD, which is part of uORF1 or a
672 combination of the two. B) Effect of gradual extension of region upstream of uORF2.
673 Expression decreases, when αSD is included and increases again, when SD1 (ααSD) is included.

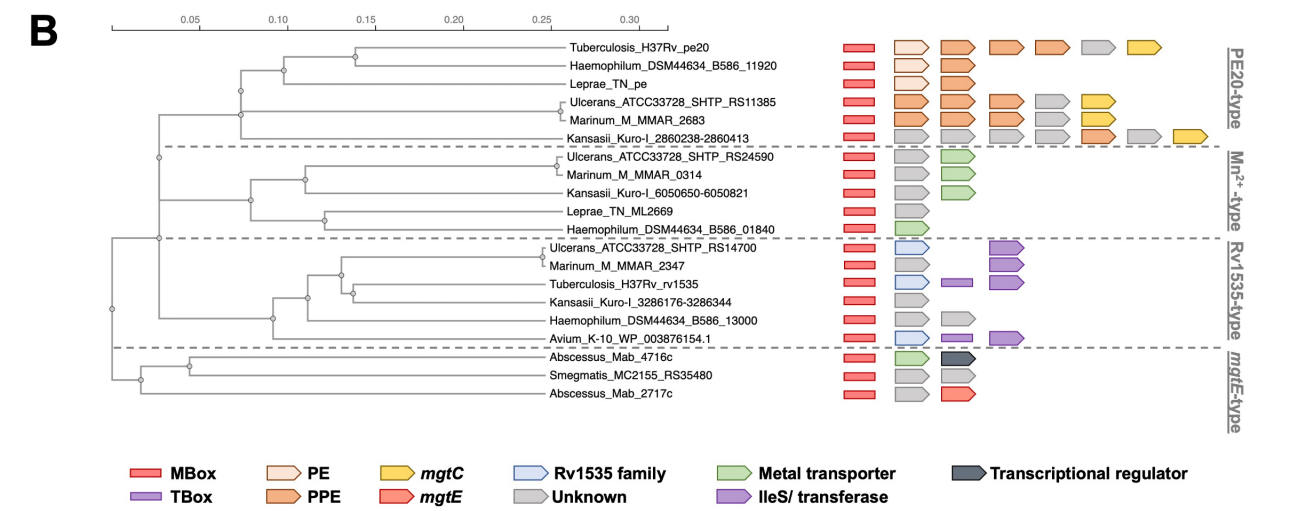
C) Structures indicating how the reporter constructs relate to the model proposed in Figure 4. Experiments were done in triplicates and differences of expression tested with a t-test (p-val<0.01).

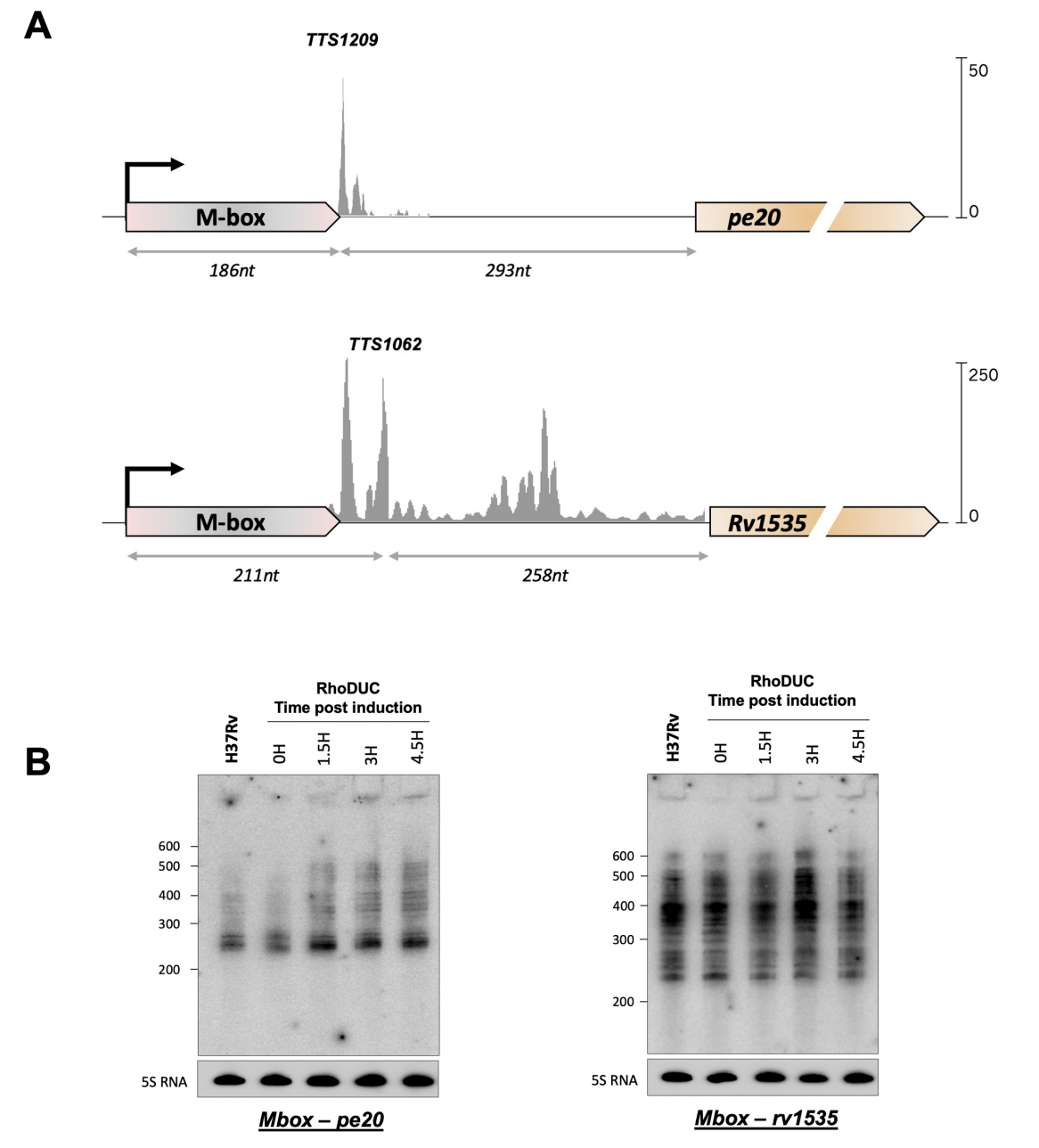
Figure 6: The Rv0115A locus. BLAST identified Rv0115A to be a homologue of Rv1805A. A) Rv0115A (green) is encoded downstream of the *gca* operon (golden) but transcribed from its own promoter. B) Alignment of the promoter regions of *gca* and *rv0115A* show very high degree of similarity, suggesting a duplication event. The blue arrow indicates *hddA* coding sequence upstream of *rv0115A*. The promoter elements, -35, extended -10 and -10 are highlighted in grey. PhoP binding regions, according to TBDB, are shown in orange and purple with their respective centres boxed in same colour.

Figure 7: Conservation of uORF2 within Mycobacteria. Alignment of Mbox associated uORF2 extracted from Figure 1B and Mtb Rv0115A peptides showed high conservation of several residues mainly at the N-terminal sequence, including 100% conservation of a proline at position 7 in most peptides. Consensus sequence and amino acid conservation were assessed using Chimera (Meng *et al.,* 2006).

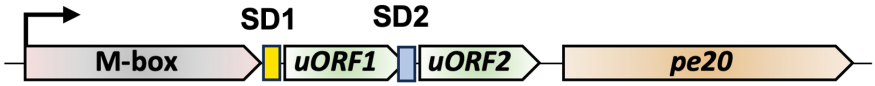
Figure 8: Biofilm formation in *M. smegmatis* during Mg2+-depletion and acid stress. Cultures of *Mycobacterium smegmatis* were grown to mid-log phase, washed in Mg2+-free medium, resuspended in 1 mL of indicated medium at OD 0.01. Plates were sealed in plastic bags and left for static incubation at 37˚C for a week. Plates shown are representative of three independent experiments.


A
B. subtilis - ykoK
M. tuberculosis – rv1535
M. tuberculosis – pe20

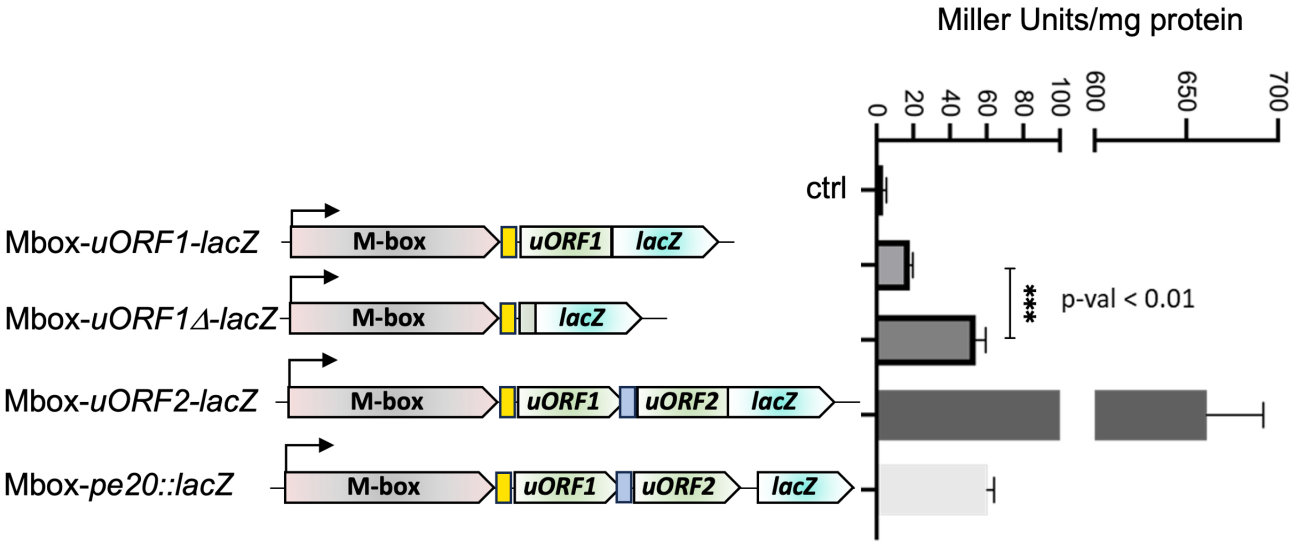

B
0.05
0.10
0.15
0.20
0.25
0.30
Tuberculosis_H37Rv_pe20
Haemophilum_DSM44634_B586_11920
Leprae_TN_pe
Ulcerans_ATCC33728_SHTP_RS11385
Marinum_M_MMAR_2683
Kansasii_Kuro-I_2860238-2860413
Ulcerans_ATCC33728_SHTP_RS24590
Marinum_M_MMAR_0314
Kansasii_Kuro-I_6050650-6050821
Leprae_TN_ML2669
Haemophilum_DSM44634_B586_01840
Ulcerans_ATCC33728_SHTP_RS14700
Marinum_M_MMAR_2347
Tuberculosis_H37Rv_rv1535
Kansasii_Kuro-I_3286176-3286344
Haemophilum_DSM44634_B586_13000
Avium_K-10_WP_003876154.1
Abscessus_Mab_4716c
Smegmatis_MC2155_RS35480
Abscessus_Mab_2717c
PE20-type
Mn2+-type
Rv1535-type
mgtE-type
MBox
TBox
PE
PPE
mgtC
mgtE
Rv1535 family
Unknown
Metal transporter
IleS/ transferase
Transcriptional regulator


A
TTS1209
50
0
M-box
pe20
186nt
293nt
TTS1062
250
0
M-box
Rv1535
211nt
258nt
B
RhoDUC
Time post induction
H37Rv
0H
1.5H
3H
4.5H
600
500
400
300
200
5S RNA
Mbox – pe20
RhoDUC
Time post induction
H37Rv
0H
1.5H
3H
4.5H
600
500
400
300
200
5S RNA
Mbox – rv1535

A Genomic context

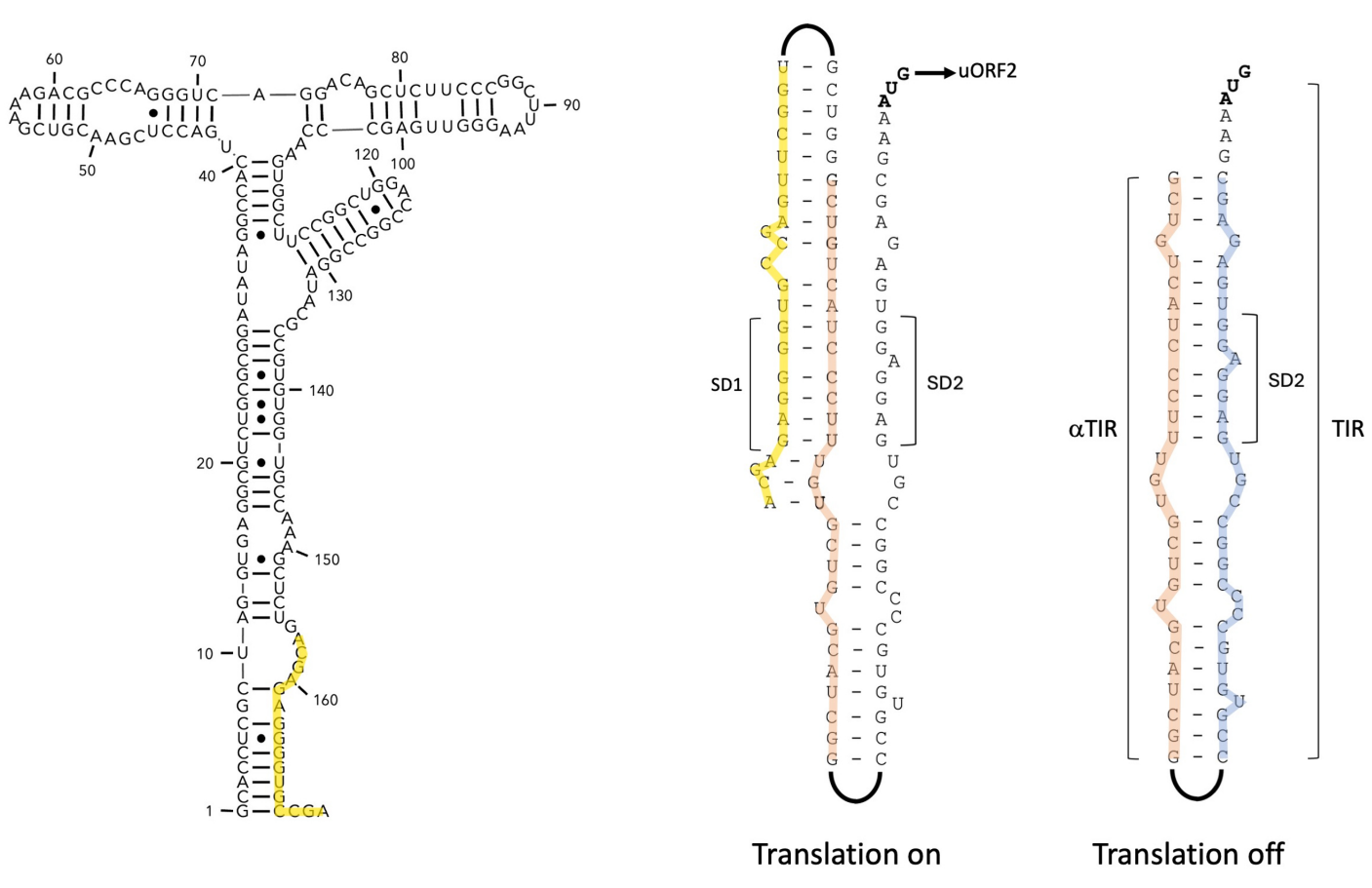


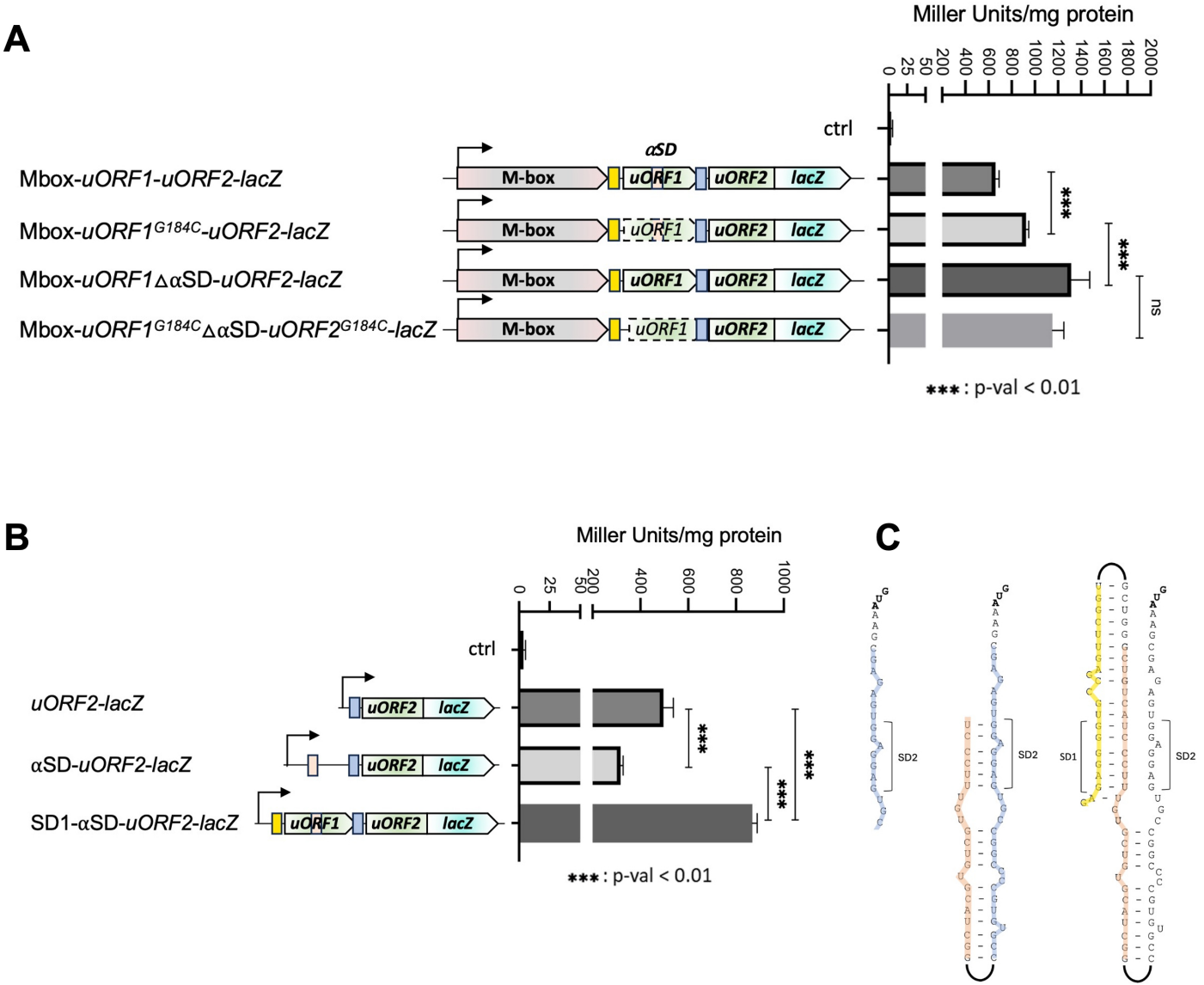
B Expression

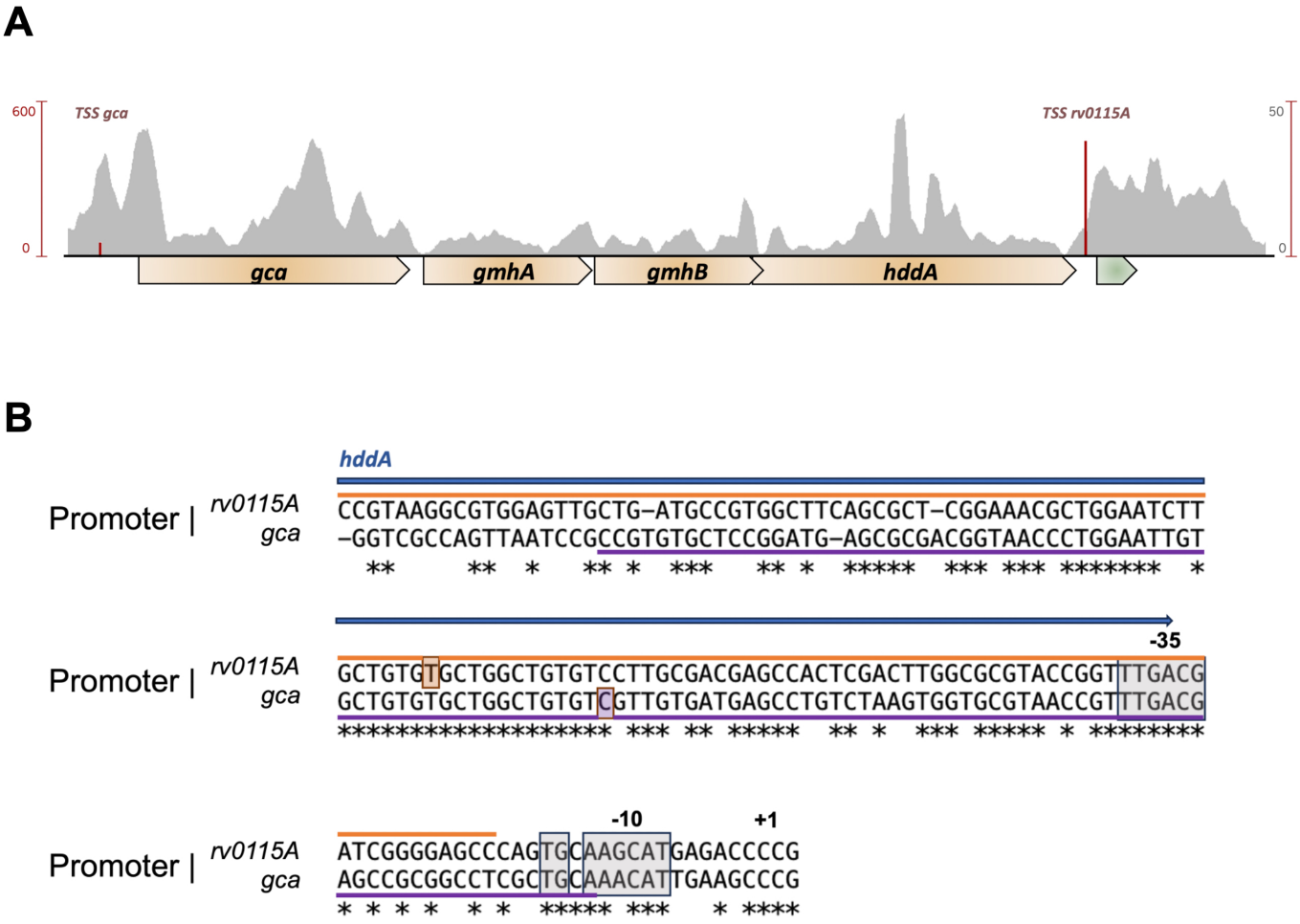


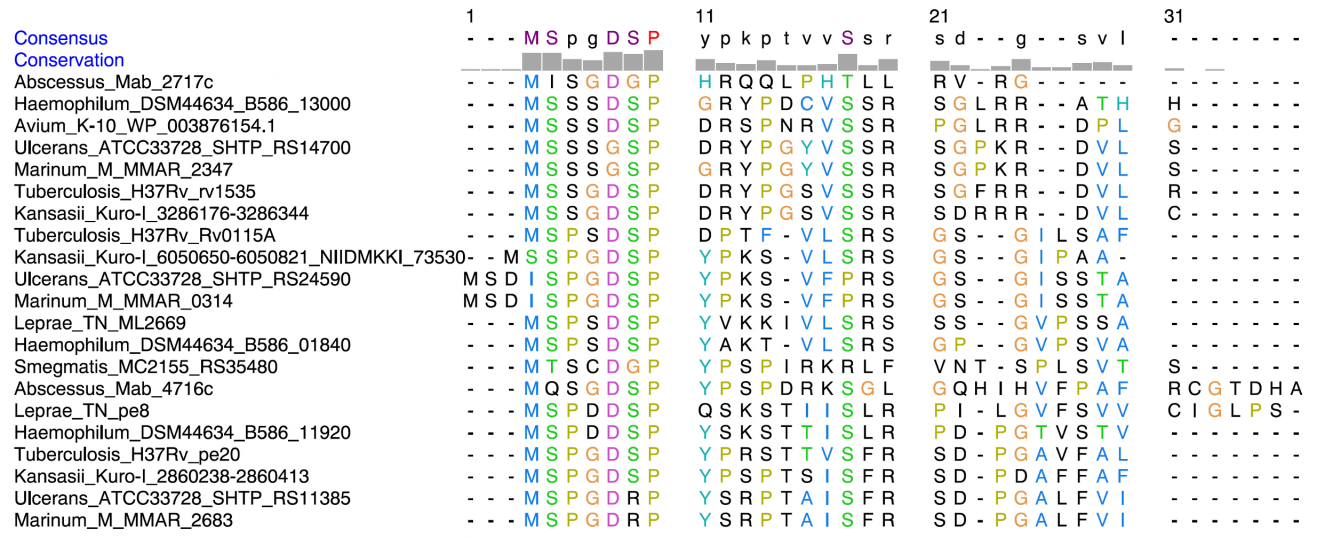
C uORF1 sequence

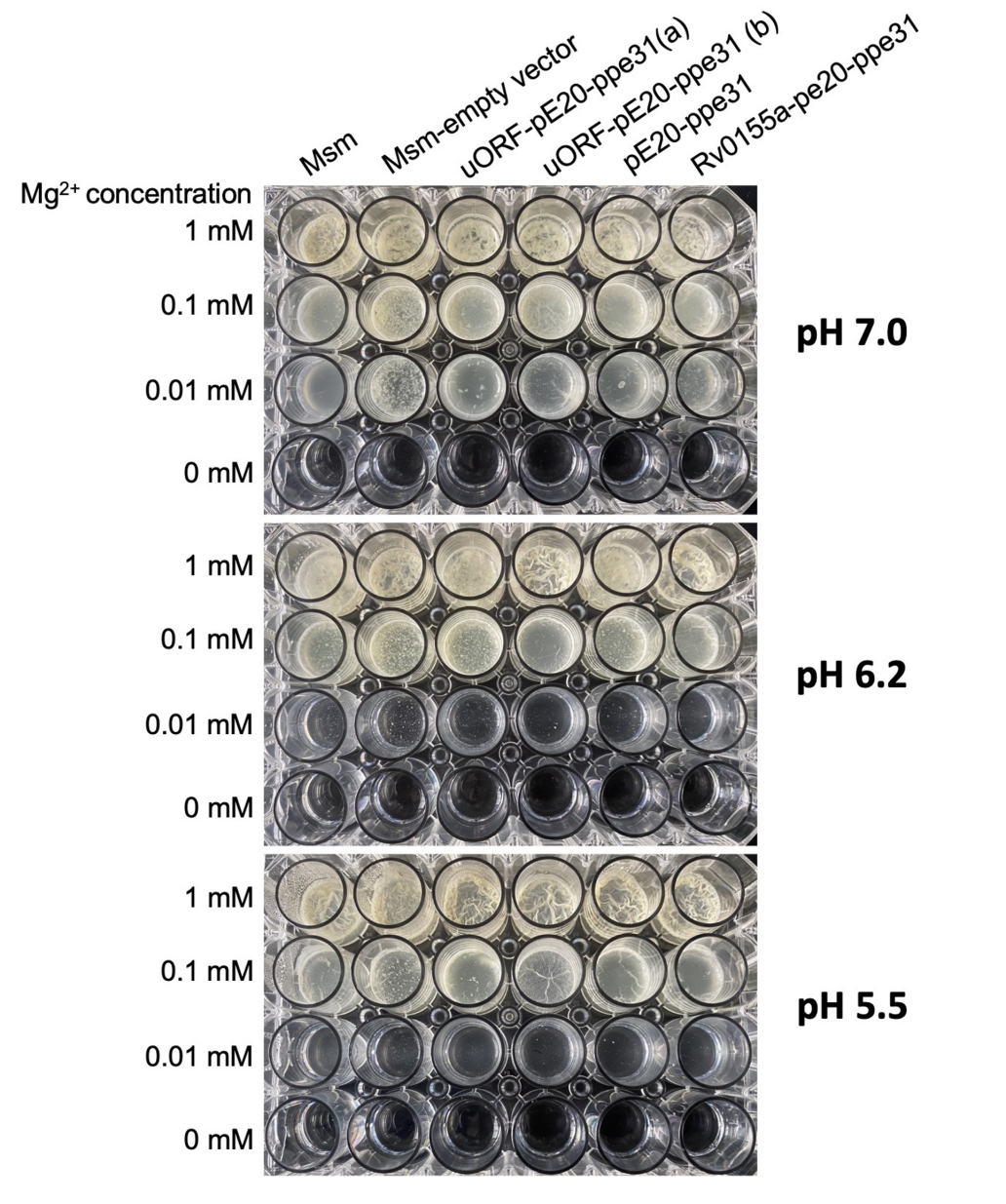
Amino Acid	M	V	C	W	A	V	I	P	L	C	C	A	C	A	S	P	C	A	P	A	V	R	R
Codon	GTG	GTC	TGC	TGG	GCT	GTC	ATC	CCT	TTG	TGC	TGT	GCA	TCG	GCA	TCC	CCG	TGT	GCC	CCG	GCC	GTG	AGG	AGG


uORF2
SD1
SD2
αTIR
SD2
TIR
Translation on
Translation off


A
Miller Units/mg protein
0
25
50
200
400
600
800
1000
1200
1400
1600
1800
2000
ctrl
αSD
Mbox-uORF1-uORF2-lacZ
Mbox-uORF1G184C-uORF2-lacZ
Mbox-uORF1ΔαSD-uORF2-lacZ
Mbox-uORF1G184CΔαSD-uORF2G184C-lacZ
M-box
uORF1
uORF2
lacZ
***
***
ns
***: p-val < 0.01
B
Miller Units/mg protein
0
25
50
200
400
600
800
1000
ctrl
uORF2-lacZ
αSD-uORF2-lacZ
SD1-αSD-uORF2-lacZ
uORF2
lacZ
uORF1
***
***
***
***: p-val < 0.01
C
SD1
SD2


A
600
0
TSS gca
TSS rv0115A
50
0
gca
gmhA
gmhB
hddA
B
hddA
Promoter | rv0115A
gca
CCGTAAGGCGTGGAGTTGCTG–ATGCCGTGGCTTCAGCGCT–CGGAAACGCTGGAATCTT
–GGTCGCCAGTTAATCCGCCGTGTGCTCCGGATG–AGCGCGACGGTAACCCTGGAATTGT
-35
GCTGTGTGCTGGCTGTGTCCTTGCGACGAGCCACTCGACTTGGCGCGTACCGGTTTGACG
GCTGTGTGCTGGCTGTGTCGTTGTGATGAGCCTGTCTAAGTGGTGCGTAACCGTTTGACG
-10
+1
ATCGGGGAGCCCAGTGCAAGCATGAGACCCCG
AGCCGCGGCCTCGCTGCAAACATTGAAGCCCG


1
11
21
31
Consensus
Conservation
Abscessus_Mab_2717c
Haemophilum_DSM44634_B586_13000
Avium_K-10_WP_003876154.1
Ulcerans_ATCC33728_SHTP_RS14700
Marinum_M_MMAR_2347
Tuberculosis_H37Rv_rv1535
Kansasii_Kuro-I_3286176-3286344
Tuberculosis_H37Rv_Rv0115A
Kansasii_Kuro-I_6050650-6050821_NIIDMKKI_73530
Ulcerans_ATCC33728_SHTP_RS24590
Marinum_M_MMAR_0314
Leprae_TN_ML2669
Haemophilum_DSM44634_B586_01840
Smegmatis_MC2155_RS35480
Abscessus_Mab_4716c
Leprae_TN_pe8
Haemophilum_DSM44634_B586_11920
Tuberculosis_H37Rv_pe20
Kansasii_Kuro-I_2860238-2860413
Ulcerans_ATCC33728_SHTP_RS11385
Marinum_M_MMAR_2683


Msm
Msm-empty vector
uORF-pE20-ppe31(a)
uORF-pE20-ppe31 (b)
pE20-ppe31
Rv0155a-pe20-ppe31
Mg2+ concentration
1 mM
0.1 mM
0.01 mM
0 mM
pH 7.0
1 mM
0.1 mM
0.01 mM
0 mM
pH 6.2
1 mM
0.1 mM
0.01 mM
0 mM
pH 5.5

A *Mycobacterium tuberculosis* Mbox controls a conserved, small upstream ORF via a translational expression platform and rho-dependent termination of transcription

Alexandre D'Halluin, Terry Kipkorir, Catherine Hubert, et al.

RNA published online November 21, 2025