4292–4307 *Nucleic Acids Research, 2019, Vol. 47, No. 8* *Published online 1 March 2019*
doi: 10.1093/nar/gkz149

Mycobacterium tuberculosis 6C sRNA binds multiple mRNA targets via C-rich loops independent of RNA chaperones

Juntao Mai[1], Chitong Rao[1], Jacqueline Watt[1], Xian Sun[2], Chen Lin[2], Lu Zhang[2,3,*] and Jun Liu[1,*]

[1]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada, [2]State Key Laboratory of Genetic Engineering, Institute of Genetics, School of Life Science, Fudan University, Shanghai, China and [3]Shanghai Engineering Research Center of Industrial Microorganisms, Shanghai, China

Received August 30, 2018; Revised February 19, 2019; Editorial Decision February 20, 2019; Accepted February 22, 2019

ABSTRACT

Bacterial small regulatory RNAs (sRNAs) are the most abundant class of post-transcriptional regulators and have been well studied in Gram-negative bacteria. Little is known about the functions and mechanisms of sRNAs in high GC Gram-positive bacteria including *Mycobacterium* and *Streptomyces*. Here, we performed an in-depth study of 6C sRNA of *Mycobacterium tuberculosis*, which is conserved among high GC Gram-positive bacteria. Forty-seven genes were identified as possible direct targets of 6C sRNA and 15 of them were validated using an *in vivo* translational *lacZ* fusion system. We found that 6C sRNA plays a pleotropic role and regulates genes involved in various cellular processes, including DNA replication and protein secretion. Mapping the interactions of 6C sRNA with mRNA targets *panD* and *dnaB* revealed that the C-rich loops of 6C sRNA act as direct binding sites to mRNA targets. Unlike in Gram-negative bacteria where RNA binding proteins Hfq and ProQ are required, the interactions of 6C sRNA with mRNAs appear to be independent of RNA chaperones. Our findings suggest that the multiple G–C pairings between single stranded regions are sufficient to establish stable interactions between 6C sRNA and mRNA targets, providing a mechanism for sRNAs in high GC Gram-positive bacteria.

INTRODUCTION

Trans-encoded small regulatory RNAs (sRNAs) that act as posttranscriptional regulators have been identified in a variety of bacteria and play key roles in cellular processes and bacterial adaptations to changing environments (1,2). Most sRNAs bind mRNA targets via base-pairing in the 5′ untranslated region (UTR) near the ribosome binding site (RBS), leading to translational repression and/or mRNA destabilization (3). In Gram-negative bacteria, the base-pairing of many sRNAs with mRNAs is facilitated by the RNA chaperone Hfq (4). Hfq forms a toroid and employs four solvent exposed surfaces to interact with the polyU Rho-independent terminators in sRNAs and AU-rich sequences in mRNAs, which brings them into proximity and promotes base-pairing (5). The matchmaking function of Hfq has greatly facilitated the identification of sRNAs and their mRNA targets. More than 200 sRNAs have been identified in *Escherichia coli* and *Salmonella* (6). Recently, co-immunoprecipitation of cellular RNAs with Hfq followed by deep sequencing have provided a global sRNA interactome in *E. coli*, which showed that at least half of all mRNAs are subject to sRNA-mediated regulation (7). Hfq is also present in many low GC Gram-positive bacteria. Although the role of Hfq in these organisms has remained elusive and somewhat controversial (8), structural studies have shown that Hfq from *Staphylococcus aureus*, *Listeria monocytogenes* and *Bacillus subtilis* associates with RNAs *in vitro* (9–11). sRNAs in these organisms have been shown to fine-tune metabolic processes and regulate stress adaption or virulence (8).

More recently, another major RNA binding protein ProQ was identified in Gram-negative bacteria (12,13). ProQ contains a FinO domain that is critical for RNA binding. The ProQ/FinO family proteins are not as ubiquitous as Hfq, but they are present in many proteobacteria that also have Hfq (14). All bacterial families without Hfq also lack ProQ. It was suggested that ProQ may play a complementary function to Hfq (14).

In contrast, little is known about the functions and mechanisms of sRNAs in high GC Gram-positive bacteria, which include the medically important *Mycobacterium*

*To whom correspondence should be addressed. Tel: +1 416 946 5067; Fax: +1 416 978 6885; Email: jun.liu@utoronto.ca
Correspondence may also be addressed to Lu Zhang. Tel: +86 21 51630571; Email: zhanglu407@fudan.edu.cn

and *Streptomyces*. Despite the increasing number of sRNAs identified in these organisms, very few of these putative sRNAs have been characterized, either in regards to their target mRNAs or mechanism of action (15,16). In *Mycobacterium tuberculosis* (*M. tb*), the causative agent of tuberculosis and a leading bacterial pathogen, targets have been identified for only two sRNAs (17,18). Mcr7, the first sRNA whose function has been established, regulates *tatC* mRNA and thereby impacts the activity of the Twin Arginine Translocation (Tat) protein secretion apparatus (17). Very recently, direct interaction of a mycobacterial sRNA with a mRNA target was demonstrated in the case of MsrI (18). One challenge has been the lack of obvious homologs of Hfq and ProQ in high GC Gram-positive bacteria including mycobacteria (14,15), which prevents approaches such as Hfq co-immunoprecipitation or ProQ co-sedimentation for sRNA discovery and target identification (7,12).

6C sRNA was initially predicted in a comparative genomic study of structured RNA in bacteria, and this region was found to be conserved within high GC Gram-positive bacteria including *Mycobacterium*, *Streptomyces* and *Corynebacterium* (19). It was termed the '6C' RNA due to the presence of six conserved cytosine residues in the stem-loops (19). The presence of 6C sRNA was subsequently experimentally verified in *Streptomyces coelicolor* (20) and *M. tb* (21). In *S. coelicolor*, the expression of 6C sRNA (also called *scr3558*) was increased under conditions that induce aerial hyphae formation and sporulation (20). In *M. tb*, 6C sRNA, also called B11 or ncRv13660c with the new nomenclature (22), was mapped to 93 nucleotide (nt) in size followed by polyU (21). Constructs overexpressing 6C sRNA of *M. tb* were lethal to *M. tb* and caused defective growth in *M. smegmatis* (21). A defective growth phenotype was also observed in *Corynebacterium glutamicum* overexpressing 6C sRNA (23). These observations suggest that 6C sRNA may play a role in regulation of cell division. However, the underlying mechanism remains unknown.

In this study, we performed a detailed characterization of 6C sRNA of *M. tb* to understand its function and mechanism of action. We used *M. smegmatis* as a model organism and performed RNA-seq to identify potential targets of 6C sRNA. We constructed an *in vivo* translational fusion reporter system to validate the targets. Our results showed that 6C sRNA is a regulator of genes involved in various cellular functions, including DNA replication and protein secretion. 6C sRNA binds multiple mRNA targets through the C-rich loops. The loop-loop or loop-unpaired region interactions are sufficient for efficient repression *in vivo* and appear to be independent of RNA binding proteins. Taken together, our findings provide a target recognition mechanism for sRNAs in high GC Gram-positive bacteria.

MATERIALS AND METHODS

Bacterial strains, media, and growth conditions

Mycobacteria were grown in Middlebrook 7H9 broth containing 0.2% glycerol, 10% OADC (oleic acid, bovine serum albumin, dextrose, and catalase; Difco) and 0.05% Tween 80 or 7H11 agar supplemented with 0.2% glycerol and 10% OADC at 37°C. *Escherichia coli* strain DH5α was used for routine cloning. Where indicated, antibiotics or small molecules were used at the following concentrations: kanamycin (25 μg/ml), hygromycin (75 μg/ml), anhydrotetracycline (Atc) (100 ng/ml) and pantothenate (24 μg/ml).

Molecular cloning and plasmid construction

Supplementary Data 1 lists the primers used for cloning. The *rrnB* promoter was amplified from *M. smegmatis*, and the region between –35 and –10 of P_{rrnb} was replaced with the Tet operator (5′-TCCCTATCAGTGATAGA-3′) to generate the promoter $P_{rrnb\text{-}tetO}$. The PCR product of $P_{rrnb\text{-}tetO}$ was digested with *Xba*I and *Bam*HI and cloned into pMV261 (Supplementary Figure S1A). A synthetic transcriptional terminator (5′-AAGCTTCCCCGCGAAAGCGGGGTTTTTTTTTTT-3′) was inserted at the *Hind*III sites according to a previous study (21). The *tetR* gene was amplified by PCR from pLJR962 and cloned into the vector at the *Eco*RI sites, which generates the plasmid pSI (Supplementary Figure S1A). The expression of *tetR* was driven by a strong promoter Psmyc. 6C sRNA and its mutants were cloned into pSI at *Bam*HI and *Hind*III sites.

To construct the *lacZ* reporter plasmid pUI (Supplementary Figure S1B), the kanamycin resistance gene (Kan^R) of the integration vector pMV306 was first replaced by a hygromycin resistance gene (Hyg^R) using *Spe*I and *Nhe*I digestions. The promoter $P_{rrnb\text{-}tetO}$ was cloned into the vector by *Xba*I and *Nco*I digestions (Supplementary Figure S1B). The *tetR* gene was inserted at the *Eco*RI sites, and the expression of *tetR* was driven by a strong promoter P_{smyc}. The *lacZ* gene lacking the start codon was amplified from pTZ110, digested with *Nco*I and *Hind*III, and cloned downstream of the promoter $P_{rrnb\text{-}tetO}$. Two *Sap*I-based Golden gate restriction sites were designed at the 5′ region immediately upstream of *lacZ* (Supplementary Figure S1B). The 5′ UTR sequences were based on a previous study which performed the ribosome profiling seq and RNA-Seq to determine the transcriptional start site and the ribosome binding site (24). The complementary oligos of each 5′ UTR were synthesized, annealed and ligated into the *Sap*I-digested *lacZ* reporter plasmid (Supplementary Figure S1B). Sequences of the synthesized 5′ UTRs are shown in Supplementary Data 1. All constructs were verified by DNA sequencing.

RNA-Seq and data analyses

Total RNA was isolated as previously described (25). Briefly, *M. smegmatis* strains carrying the inducible expression vector of 6C sRNA or the empty vector were grown to OD_{600} of ∼0.6. ATc (100 ng/ml) was then added to induce the expression of the sRNA. RNA was extracted at 1- and 3- h after induction using the Qiagen RNeasy Mini kit and treated with the Turbo DNase (Life). After removal of the ribosomal RNA using the RiboZero (Bacteria) kit from Illumina, the purified RNA was used to construct cDNA library according to the TruSeq Stranded RNA LT Guide from Illumina. High-throughput sequencing was carried out on an Illumina HiSeq 2500 system according to the manufacturer's instructions and 150-bp paired-end reads were obtained. The raw reads were filtered by Seqtk and then mapped to the *M. smegmatis* mc^2-155 strain reference

sequence (GenBank NC_008596) using Bowtie2 (version: 2–2.0.5) (26). Counting of reads per gene was performed using HTSeq followed by TMM (trimmed mean of M-values) normalization (27,28). Differentially expressed genes were defined as those with a false discovery rate $Q < 0.05$ and fold-change ≥2 using the edgeR software (29). For each strain, RNA-seq experiment was done twice, one at each time point (1 and 3 h) after Atc induction.

Real-time PCR (RT-PCR) analysis

RNA was treated with Turbo-DNA free kit (Ambion) to remove DNA contamination. RNA (1 μg) was reverse-transcribed to cDNA using iScript cDNA kit (Bio-RAD). For RT-PCR experiments, the standard curve was generated using serial dilutions of the genomic DNA and the *sigA* gene was used as an internal reference. The primers for amplification of selected genes are listed in Supplementary Data 1. The RT-PCR reactions were performed in 10 μl reaction mixtures including 5 μl of 2× SsoFast EvaGreen (Bio-Rad), 500 nM of forward and reverse primers, and 1 μl of cDNA template (1:50 diluted). The RT-PCR was performed using the CFX 384 Real-time PCR detection system (Bio-RAD). PCR was carried out at 95°C for 5 s and 55°C for 5 s, for a total of 40 cycles. The data were analyzed by Bio-Rad CFX Manager. The Cq-value of *dnaB* and *sigA* was converted to the absolute mRNA level based on the standard curve, and the mRNA fold change of *dnaB* was calculated by normalizing against the *sigA* mRNA level before and after the Atc treatment. Three independent experiments were performed and the results were plotted as mean ± S.D. ($n = 3$).

Computational prediction of targets of 6C sRNA

The interactions of 6C sRNA with mRNAs were predicted using IntaRNA (30). The L3 sequence of 6C sRNA (5′-UCCUCCCCCCCU-3′) was used as sRNA input and sequences of 5′ regions (–50 to +25 relative to the start codon) of the 162 downregulated genes were used as target RNA inputs. The seed parameters used were the default setting (at least 7-nt seed length). Only interactions with a *P* value of ≤0.05 were considered and chosen for further analyses.

Quantification of fluorescence intensity

M. smegmatis carrying *mCherry* driven by $P_{rrnB\text{-}tetO}$ in pMV261 was grown to OD_{600} of ∼0.6. ATc (100ng/ml) was then added and incubated for 12 h. The culture was collected and resuspended in water. Bacterial culture (OD_{600} = 1.0) was aliquoted (0.2 ml) and added to 96-well plates, and the fluorescence intensity was measured by TECAN Infinite M200 plate reader. The specific excitation and emission wavelengths for detection of the *mCherry* reporter were 587 and 610 nm, respectively.

Assay of β-galactosidase activity

Assays to measure β-galactosidase were performed as previously described (31). Briefly, *M. smegmatis* strains were grown to OD_{600} of 0.6 and Atc (100 ng/ml) was added and incubated for 3 h. The cultures were collected and washed twice with 1 ml Z buffer (0.06 M Na_2HPO_4, 0.04 M NaH_2PO_4, 0.01 M KCl, 1 mM $MgSO_4$, and 0.05 M β-mercaptoethanol). Cells were then lysed using a bead-beater (three 30 s pulses with 1-min interval between pulses). Cell lysates (20–100 μl) were added to 1 ml Z buffer containing 1 mg/ml of *o*-nitrophenyl-D-galactoside (ONPG). The reaction was incubated at 30°C for 1 h. Activity of ß-galactosidase was measured at OD_{420}. To calculate the Miller units, the following formula was used: (OD_{420})/(reaction time [min] × aliquot volume [μl] × OD_{600}). The results (mean ± S.D.) were from three biological replicates.

Northern blot analysis

Total RNA (0.5–4.0 μg) was separated on denaturing 8% acrylamide gels and transferred onto Hybond-N^+ membranes by electroblotting. The membranes were subjected to UV cross-linking and then incubated in the pre-hybridization solution (50% formamide, 0.1% SDS, 5 × saline sodium citrate, 2 × Denhardt's solution, 0.1 mg/ml salmon sperm DNA) for 2 h at 42°C in a rotating hybridization oven. The biotinylated DNA probes for 6C sRNA and mutants were synthesized at Integrated DNA Technologies (Toronto) and are listed in Supplementary Data 1. This probe was designed to hybridize with the 6C sRNA of *M. tb* but not the endogenous 6C sRNA of *M. smegmatis*, which was confirmed by experiments. The biotinylated DNA probe for 5S rRNA, which was used as an internal control, is also listed in Supplementary Data 1. The biotinylated DNA probes (0.1 nmol) were added to the membranes in the hybridization solution (50% formamide, 0.1% SDS, 5 × saline sodium citrate, 0.1 mg/ml salmon sperm DNA) and incubated at 42°C overnight. After the hybridization, the membranes were washed once with 1 × saline sodium citrate and 0.1% SDS at room temperature for 30 min. The membranes were then washed twice, each for 20 min with 0.2× saline sodium citrate and 0.1% SDS at 60°C. For detection, the membranes were incubated in the blocking buffer (1 × PBS, 0.5% SDS, 0.1% BSA) for 30 min. The biotinylated probe was then detected by adding horseradish peroxidase-conjugated streptavidin and analyzed by the ChemiDoc MP imaging system (Bio-Rad). Two independent experiments were performed and intensity of the bands was analyzed using the ImageJ program.

Synthesis of RNA by *in vitro* transcription

For *in vitro* transcription, template DNA was generated by PCR using gene specific oligonucleotides with a T7 promoter sequence at the 5′ end of the forward primer. The primers used to amplify 6C sRNA, its mutants and the 5′ UTR of *panD* are listed in Supplementary Data 1. The DNA fragments were separated by agarose gel electrophoresis and then purified. Transcription of the DNA templates was performed using T7 RNA polymerase (NEB) following the manufacturer's instructions. The reactions were carried out at 37°C overnight. The resulting RNA was

separated by 8% polyacrylamide native gel electrophoresis and then purified.

Electrophoretic mobility shift assay (EMSA)

The *in vitro* transcribed *panD* mRNA (0.25 μg) was dephosphorylated using the shrimp alkaline phosphatase rSAP (NEB) at 37°C for 1 hr. The RNA was then end-labeled with ^{32}P using T4 polynucleotide kinase (NEB) and γ-^{32}P ATP following the manufacturer's instructions. ^{32}P-labeled *panD* mRNA (0.1 pmol) was mixed with indicated concentrations of 6C WT and mutants in 10 μL of the RNA binding buffer (20 mM Tris–HCl pH 8.0, 1 mM $MgCl_2$, 20 mM KCl, 10 mM Na_2HPO_4–NaH_2PO_4 pH 8.0) and then incubated at 95°C for 3 min. The mixture was incubated at 37°C for 1 h and then loaded onto 5% polyacrylamide native gels. The gels were run at 20 mA in Tris-borate-EDTA (TBE) buffer (0.5×) containing 5% glycerol for 0.5 h at 4°C. The gels were then dried, exposed on the phosphor storage screen, and images acquired by a Typhoon FLA 9500 laser scanner (GE). The intensities of bands were quantified using ImageJ. The data were fitted using a single-site binding model (Prism) and the K_D value was generated. These experiments were performed twice.

Silencing of *dnaB* with CRISPR interference (CRISPRi)

The methods of CRISPRi mediated gene silencing in mycobacteria were previously described (32). Briefly, the CRISPRi backbone vector pLJR962 was digested with *Bsm*BI. sgRNA oligos targeting *dnaB* (listed in Supplementary Data 1) were annealed and ligated into the digested pLJR962 to construct pLJR962::*dnaB*. To induce the CRISPRi silencing of *dnaB*, *M. smegmatis* carrying pLJR962::*dnaB* was grown to mid-log phase (OD_{600} of ∼1.0) and then 100 ng/ml of ATc was added. Target knockdown was allowed to proceed for 12 h and the bacterial growth was measured by OD_{600} and compared with *M. smegmatis* carrying the pLJR962 control vector. To quantify the effect of CRISPRi, 10 OD_{600} equivalents of bacteria with or without induction were harvested and subjected to RT-PCR analysis as described above.

Generation and screening of *M. smegmatis* MycoMar insertion library

The kanamycin resistance gene in the plasmid pSI containing 6C sRNA was replaced with a hygromycin resistance gene by *Spe*I and *Nhe*I digestions. Propagation of the MycoMar transposon phage (Kan^R) was prepared as previously described (33). For phage infection, *M. smegmatis* carrying the inducible expression vector of 6C sRNA were grown to OD_{600} ∼0.8, washed twice and re-suspended in 5 ml of MP buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10 mM $MgSO_4$, 2 mM $CaCl_2$). Phage were then added at a MOI of 10:1 and incubated at 37°C for 3 hr. Bacteria were washed, re-suspended in fresh 7H10 media containing Atc, and incubated at 37°C for 6 h. Cultures were then plated on 7H10 agar containing kanamycin (25 μg/ml), hygromycin (75 μg/ml) and Atc (100 ng/ml) and incubated at 37°C for 4–5 days. Colonies that appeared were collected and subjected to further analysis.

Fluorescence microscopy

To stain the DNA of *M. smegmatis*, bacterial cells (200 μl) were washed twice with 1× PBS and then incubated with 2.5 μg/ml of 4′,6-diamino-2-phenylindole (DAPI) for 10 min. After DAPI staining, the cells were washed with 1× PBS three times, and 20 μl of the cells were transferred to slides. The fluorescence images were acquired by the Nikon Eclipse 80i microscope.

RESULTS

Construction of a tetracycline inducible expression system of 6C sRNA

Previously, Arnvig and Young constructed a 6C sRNA overexpression vector in which 6C sRNA of *M. tb* was cloned in plasmid pMV261 under the control of the strong *rrnB* promoter of *M. smegmatis* (21). They found that this construct was lethal in *M. tb* and caused defective growth in *M. smegmatis* (21). We modified this construct by placing the Tet operator sequence (*tetO*) between the –35 and –10 of the *rrnB* promoter and adding the Tet repressor *tetR* under the control of another strong promoter P_{smyc} (Figure 1A). This construct allowed us to induce the expression of 6C sRNA by addition of tetracycline. To validate this system, we replaced 6C sRNA with the *mCherry* reporter gene, and transformed the resulting construct into *M. smegmatis*. As expected, in the absence of tetracycline, the fluorescence intensity of the culture was very low. However, addition of anhydrotetracycline (Atc) (100 ng/ml) resulted in a 19.7-fold increase of the fluorescence intensity (Figure 1B).

The inducible expression construct of 6C sRNA was transformed into *M. smegmatis* and the effect of 6C sRNA overexpression on bacterial growth was examined. *M. smegmatis* containing the plasmid without the 6C sRNA insert was used as the control. In the absence of Atc, both strains grew equally well in 7H9 medium at 37°C. The cultures were grown to $OD_{600} = 0.3$ and then Atc (100 ng/ml) was added. Compared to the control strain, *M. smegmatis* containing the 6C sRNA overexpression vector began to show reduced growth ∼9 h after the addition of Atc, and ceased growing after 12 hr, which resulted in a much lower cell density (Figure 2A). Microscopic examination revealed that these cells were significantly elongated compared to the control cells (Figure 2B). These results are consistent with the previous finding that overexpression of 6C sRNA in *M. smegmatis* caused growth arrest (21).

The C-rich loops of 6C sRNA are required for growth inhibition

The sequence of 6C sRNA is highly conserved among mycobacterial species, with 91% identity (Supplementary Figure S2A). Secondary structure prediction of 6C sRNA of *M. tb* using the RNAfold program reveals three single-stranded loops, designated L1 to L3 (Figure 2C and Supplementary Figure S3) (34). The sequences of L2 and L3 (but not L1) are also highly conserved within high GC Gram-positive bacteria, including *Mycobacterium*, *Streptomyces*, *Corynebacterium* and *Rhodococcus* (Supplementary Figure S2B). To test whether L2 and L3 are required for function, we constructed loop-deletion mutants of 6C sRNA

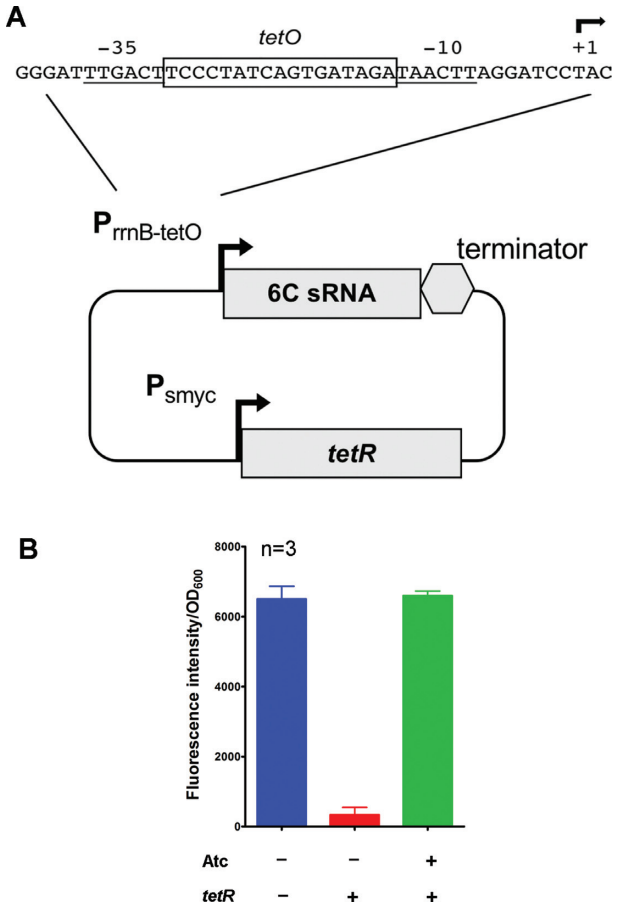

Figure 1. Construction of tetracycline inducible overexpression vector of 6C sRNA. (**A**) The region between -35 and -10 of the *rrnB* promoter was replaced with the Tet operator sequence (*tetO*) to generate $P_{rrnB\text{-}tetO}$ and 6C sRNA was cloned under the control of $P_{rrnB\text{-}tetO}$. This fragment was cloned into pMV261. The *tetR* gene under the control of another strong promoter P_{smyc} was also cloned into the same vector. (**B**) Validation of the inducible overexpression system. The 6C sRNA gene in (A) was replaced by the *mCherry* gene and the fluorescence intensity of the cultures in the presence or absence of Atc was measured. Vector that does not contain the *tetR* gene was also included for studies. The fluorescence intensity was determined by TECAN Infinite M200 plate reader. The specific excitation and emission wavelengths for detection of the *mCherry* reporter were 587 and 610 nm, respectively. The results (mean ± S.D.) were from three biological replicates ($n = 3$).

using the inducible expression construct illustrated in Figure 1A as the template. Three deletion mutants were constructed, deletion of L2 (6C ΔL2), L3 (6C ΔL3), or both (6C ΔL2ΔL3). These constructs were transformed into *M. smegmatis* and their effects on bacterial growth were determined. Results showed that overexpression of 6C ΔL2 in *M. smegmatis*, as induced by addition of Atc (100 ng/ml), inhibited cell growth and resulted in elongated cells (Figure 2A and B). These phenotypes were similar to that caused by overexpression of wild type (WT) 6C sRNA, suggesting that the deletion of L2 had minimal effect on 6C sRNA function. However, the ability of 6C sRNA to inhibit cell growth was only partially abrogated by the deletion of L3, whereas it was completely abolished by the deletion of both L2 and L3 (Figure 2A and B). Taken together, these results suggest that L3 plays a major role in 6C sRNA mediated growth inhibition. However, in the absence of L3, L2 can also exert some effects, which is evidenced by the partial growth inhibition of 6C ΔL3.

The deletion of L2 and L3 does not decrease the stability of 6C sRNA

Another possible explanation for the changes in 6C sRNA mediated growth inhibition when 6C loop deletion mutants are expressed could be changes in the stability of the 6C sRNA constructs. To address this question, cultures of *M. smegmatis* overexpressing 6C WT and deletion mutants were treated with rifampicin (200 μg/ml) following a published protocol (35), and the turnover of 6C and mutants at different time points were examined by Northern blot analysis. Results showed that the deletion of L2 and/or L3 did not decrease the stability of 6C sRNA, and in fact the three deletion mutants appeared to be even more stable than 6C WT (Figure 2D). The estimated half-life time of 6C WT was ∼40 min, whereas it was ∼2.5 h for 6C ΔL2ΔL3, and ∼5 h for both 6C ΔL2 and 6C ΔL3 (Figure 2E). There are no major changes in the secondary structure of 6C deletion mutants predicted by RNAfold (Supplementary Figure S3).

Mutations of cytosine residues in L3 affect the function of 6C sRNA

Since L3 plays a major role in 6C sRNA mediated growth inhibition, we focused on identifying residues of L3 critical for its function by site directed mutagenesis using the inducible expression vector of 6C ΔL2 as the template. The L3 is highly rich in cytosine, including 7 contiguous C in position 70–76 (Figure 2C and F). The first five C (position 70–74) are also conserved in other members of high GC Gram-positive bacteria (Supplementary Figure S2B). To test if these conserved residues are critical for function, we introduced single, double or triple point mutations in this region (Figure 2F). The resulting constructs were transformed into *M. smegmatis* and their effects on bacterial growth were determined. Results showed that a single C→G mutation (6C ΔL2M1) did not affect the ability of 6C ΔL2 to inhibit cell growth (Figure 2G). However, changing at least two C to G in this region (6C ΔL2M2, 6C ΔL2M3) abolished the growth inhibition (Figure 2G). These results indicate that cytosine residues in L3 are critical for 6C sRNA function.

Transcriptome analysis of *M. smegmatis* overexpressing 6C sRNA

To identify the potential targets of 6C sRNA, we performed RNA-seq analysis of *M. smegmatis* upon induction of 6C sRNA overexpression. For this experiment, *M. smegmatis* harboring the inducible expression vector of 6C sRNA (Figure 1A) were grown to mid-log phase ($OD_{600} = 0.6$) and then Atc (100 ng/ml) was added for 1 and 3 h. Cultures that were grown for the same time period without Atc were used as controls. We found that the expression level of 6C sRNA in cultures was low but detectable after the induction with Atc for ∼10 min, and it continued to increase overtime (Supplementary Figure S4A and B). Substantial levels of 6C were detected at 1 and 3 h after Atc induction, yet there was

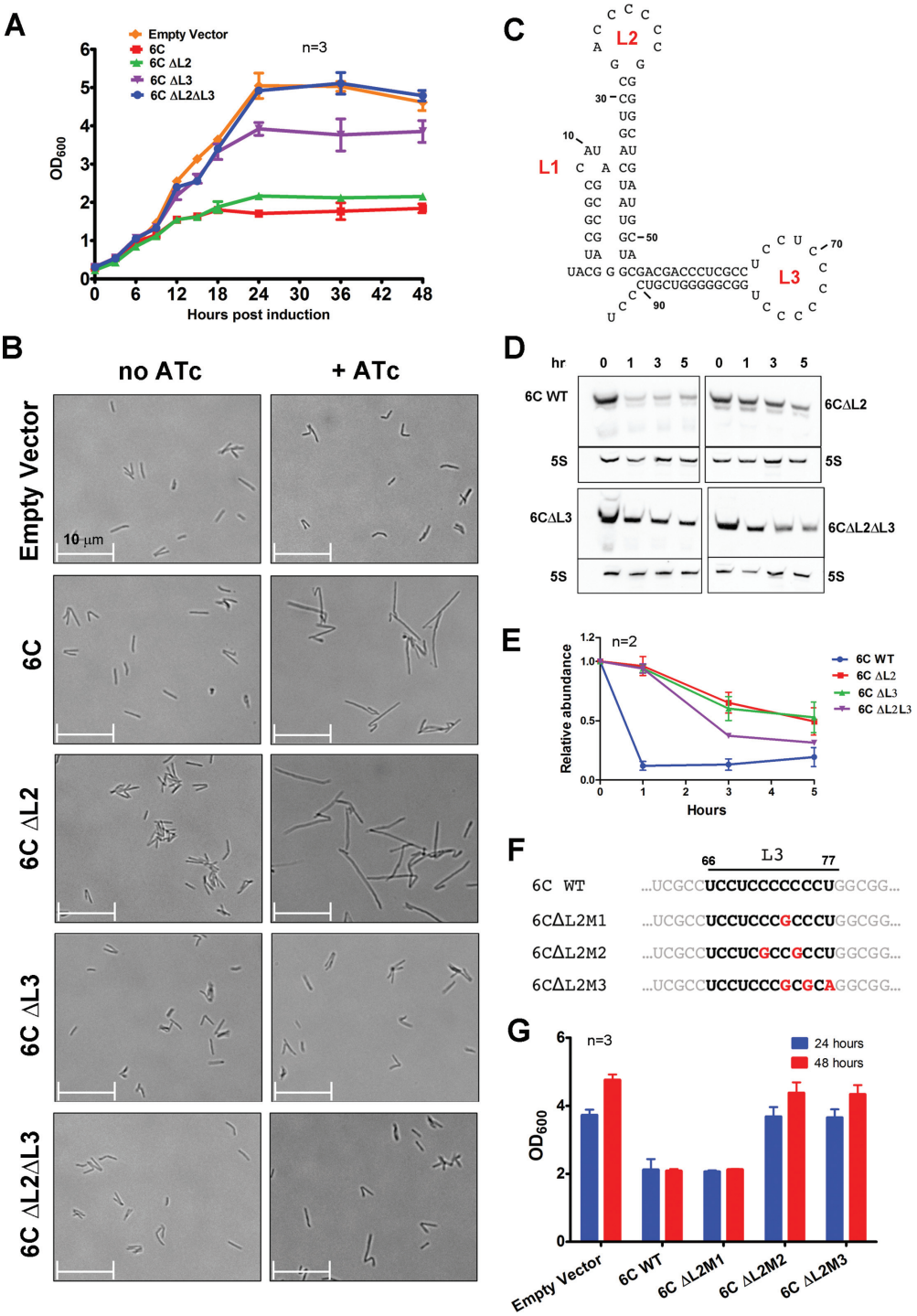

Figure 2. The C-rich loops of 6C sRNA are required for growth inhibition. (**A**) Growth curves of *M. smegmatis* strains carrying inducible overexpression vector of 6C sRNA or 6C sRNA loop deletion mutants. Cells were grown to $OD_{600} = 0.3$ and then added with 100 ng/ml ATc. The effect of Atc induction on cell growth was monitored hereafter. The results (mean ± S.D.) were from three biological replicates ($n = 3$). (**B**) Cell morphology of *M. smegmatis* strains carrying inducible overexpression vector of 6C sRNA or 6C sRNA loop deletion mutants. Cultures were grown as in (A) and examined by the Nikon Eclipse 80i microscope 24 hr after Atc induction. The scale bar is 10 μm. (**C**) Secondary structure of 6C sRNA of *M. tb* based on prediction by RNAfold. L1 to L3 indicate the three loops. (**D**) Northern blot analysis of the turnover of 6C and deletion mutants. *M. smegmatis* strains carrying inducible overexpression vector of 6C or loop deletion mutants were treated with rifampicin (200 μg/ml). At the indicated time hereafter, RNA was isolated and subjected to Norther blot analysis. The 5S rRNA was used as an internal reference. (**E**) The relative abundance of 6C and mutants over time after rifampicin treatment was plotted. The levels of 6C and mutants were normalized against that of 5S rRNA and were then compared at different time points and plotted. The data (mean ± S.D.) were from two independent experiments ($n = 2$). (**F**) A schematic of the site-directed mutagenesis in L3 of 6C ΔL2 sRNA, the locations of the mutations were indicated in red. (**G**) Growth of *M. smegmatis* strains carrying inducible overexpression vector of 6C sRNA or 6C sRNA ΔL2 point mutants. All of the strains were grown to $OD_{600} = 0.6$, and ATc (100 ng/ml) was then added. OD_{600} was measured at 24- and 48-h post induction. The results (mean ± S.D.) were from three biological replicates (n=3).

little effect on the growth of the bacteria at these time points (Supplementary Figure S4C). There appeared to be a slight decrease in colony forming units (CFU) of the cultures 3 h after induction, but the difference was not statistically significant (Supplementary Figure S4C). Therefore, we chose the 1 and 3 h time points for RNA-seq analysis to assess the effect of 6C sRNA overexpression and identify potential targets that are responsible for the growth inhibition phenotype. Under these conditions, the secondary effects on gene expression caused by differential growth would be minimized.

A total of 655 differentially expressed genes (DEGs, ≥2-fold, $Q<0.05$) were identified upon induction of 6C sRNA overexpression for 1 h, including 515 downregulated and 140 upregulated genes. More DEGs (total 1,227, 787 downregulated and 440 upregulated) were detected after 3 h induction, which might be associated with the slight decrease of bacterial growth at this time point (Supplementary Figure S4C). 162 genes were consistently downregulated after 1 and 3-h induction, and the overlap is highly significant (the hypergeometric *P*-value is 4.5e–12) (Dataset S1). Only 28 upregulated genes overlapped between these two time points ($P = 0.001$) (Dataset S2). We considered these genes, which were consistently affected at both time points, as the putative targets of 6C sRNA. It should be noted that these genes were used as an initial guide for further experiments, and only experimentally validated genes were considered as the genuine targets. Nonetheless, the majority of these genes were downregulated upon 6C sRNA overexpression, suggesting that 6C sRNA may act mainly as a negative regulator.

Development of an *in vivo* translation-reporter system to assay the interactions of 6C and target mRNAs

To validate the 6C sRNA targets, we developed a system for *in vivo* studies of the direct regulation of putative mRNA targets by 6C sRNA. This system is based on the co-expression of a translational fusion of candidate mRNA and *lacZ* that is integrated into the chromosome and 6C sRNA on the inducible expression vector (Figure 3A). We chose *panD* as an example to test this system. The *panD* gene was downregulated >10 fold upon induction of 6C sRNA overexpression and its RBS has been experimentally mapped (24) (Dataset S1). To investigate whether the regulation of *panD* mRNA by 6C sRNA was direct, a 39-nt fragment containing the 5′ UTR (24) and the first 12 nt of the coding sequence (Supplementary Data 1) was cloned as N-terminal translational fusion to *lacZ* (named as *panD*′-′*lacZ*). The *panD*′-′*lacZ* fusion was cloned in an integration vector pMV306 under the control of $P_{rrnB\text{-}tetO}$, the same promoter used to drive the inducible expression of 6C sRNA (Figure 1A). A copy of *tetR* under the control of P_{smyc} was also cloned in the same vector. For negative controls, we used the same approach and constructed 5′ UTR mRNA-*lacZ* fusions of two genes, *lsr2* and *ftsZ*, which were not targets of 6C sRNA based on RNA-seq analysis. Each of the fusion constructs (*panD*′-′*lacZ*, *lsr2*′-′*lacZ* and *ftsZ*′-′*lacZ*) was transformed into *M. smegmatis* harboring the inducible expression vector of 6C sRNA (Figure 3A) or the empty vector control (Figure 3B). The resulting recombinant strains, in which the translational *lacZ* fusions had been integrated into the chromosome, were grown and β-galactosidase activities were measured.

In the absence of Atc induction, the β-galactosidase activities in all strains were very low and close to the background level, suggesting that TetR was sufficient to inhibit the transcription of *lacZ* fusions by binding to *tetO*.

To induce the expression of 6C sRNA and derepress the TetR mediated transcriptional inhibition, *M. smegmatis* strains were grown to OD_{600} of 0.6 and Atc (100 ng/ml) was added and incubated for 3 h. The cultures were then collected and subjected to β-galactosidase activity assay. Under these conditions, 6C sRNA will decrease the *lacZ* fusion translation activity (β-galactosidase activity) only if it binds the 5′ UTR of a candidate gene.

In cells containing the *ftsZ*′-′*lacZ* or *lsr2*′-′*lacZ* fusion, overexpression of 6C sRNA did not affect the fusion activities (Figure 3C), which is consistent with our finding that *ftsZ* and *lsr2* were not the targets of 6C sRNA. In contrast, the *panD*′-′*lacZ* fusion activity was greatly reduced in cells harboring the 6C sRNA overexpression vector upon induction, showing ∼95% reduction compared to cells containing the empty vector (Figure 3C). These results indicate that *panD* is directly regulated by 6C sRNA at the posttranscriptional level.

Genetic demonstration of 6C sRNA base-pairing with *panD* mRNA

Many sRNAs exert their regulatory function by base-pairing with mRNA targets in the vicinity of RBS and inhibiting translation. The RBS of *panD* has been experimentally mapped to a G-rich sequence (AGGGGGG) (24), which is in close proximity (–12 to –6 nt) to its start codon (Figure 4A). Both the C-rich loops (L2 and L3) of 6C sRNA could potentially form multiple base pairs with the RBS of *panD* and inhibit translation (Figure 4A). To determine which loop is involved in the interaction, we employed the earlier generated constructs of 6C sRNA loop deletion mutants (6C ΔL2, ΔL3 and ΔL2ΔL3) to test their effects on the *panD*′-′*lacZ* fusion activity. Results showed that deletion of both loops abolished the translational repression, whereas deletion of L3 led to a partial abrogation and deletion of L2 had minimal effect (Figure 4B). These results mirror the effect of 6C sRNA on growth inhibition (Figure 2A) and suggest that the posttranscriptional regulation of *panD* is mainly dependent on L3.

To further map the interaction sites, we used the *panD*′-′*lacZ* fusion construct as the template and made two *panD* mutants, each containing two single nucleotide mutations at the RBS of *panD* (Figure 4C). The effect of these mutations on the 6C sRNA mediated translational repression was determined by co-transforming with the 6C sRNA overexpression vector. Results showed that both mutants were not responding to 6C sRNA overexpression (Figure 4D). The *panD* M2 mutation did not affect the intrinsic translational strength, although there was a large increase of the translational efficiency caused by the G→A mutations in M1 (Figure 4D). The fact that these point mutations completely abolished the regulation by 6C suggests that the RBS of *panD* is capable of interacting with both loops of 6C sRNA.

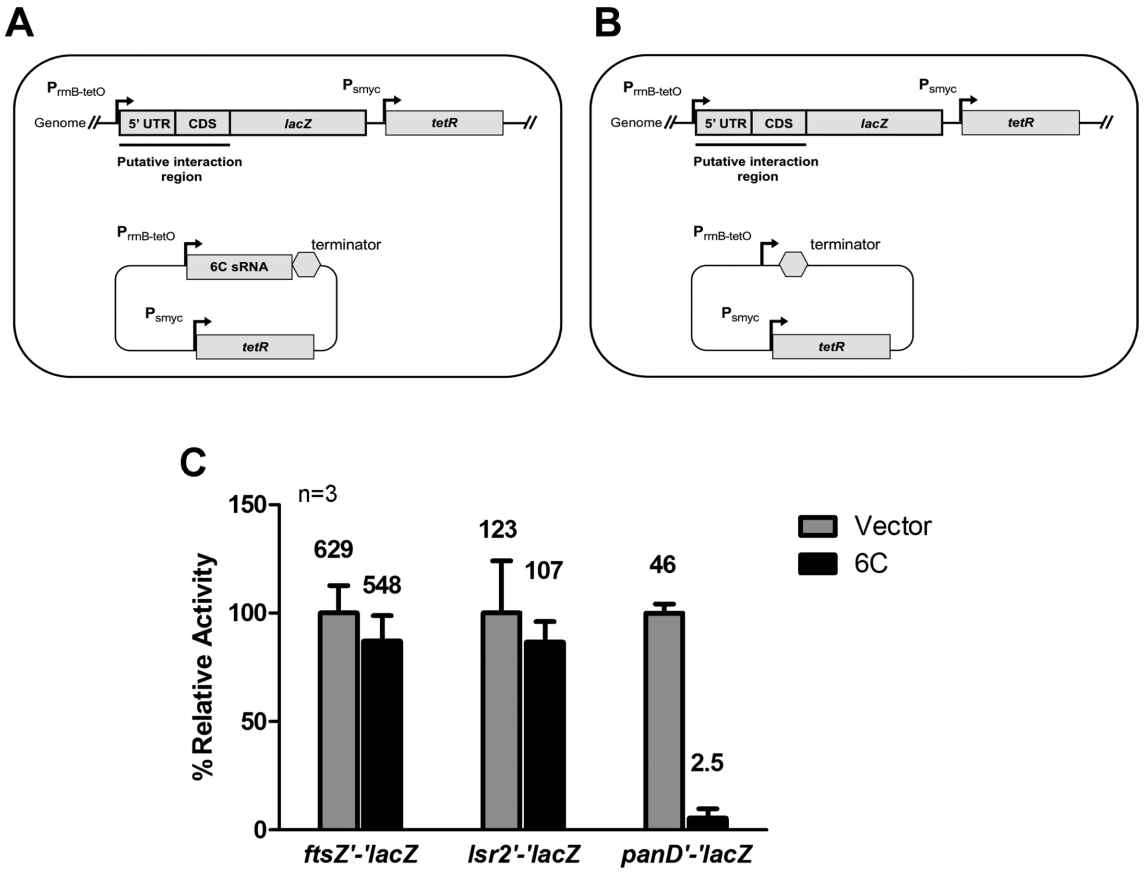

Figure 3. An *in vivo* translation-reporter system to assay the interactions of 6C and target mRNAs. (A and B) A schematic of an *in vivo* translational fusion reporter system for validation of direct targets of 6C sRNA. The 5′ UTR and the first five codons of candidate genes were fused to the *lacZ* gene lacking the start codon, and the in-frame fusion was placed under the control of the ATc inducible promoter $P_{rrnB\text{-}tetO}$. The *tetR* gene under the control of P_{smyc} was also cloned into the same integration vector (pMV306). The resulting constructs were transformed into *M. smegmatis* harboring the inducible expression vector of 6C sRNA (**A**) or the empty vector control (**B**). In the absence of Atc, TetR will repress the expression of 6C sRNA and the *lacZ* fusion by binding to *tetO*, resulting in low β-galactosidase activities. In the presence of Atc, 6C sRNA will be derepressed and will decrease the *lacZ* fusion translation activity (β-galactosidase activity) only if 6C sRNA binds the 5′ UTR of the candidate genes. (**C**) β-Galactosidase activities of *M. smegmatis* strains carrying 6C sRNA overexpression vector (**A**) or the empty vector (**B**). *M. smegmatis* strains were grown to log phase and 100 ng/ml of ATc were added. Samples were harvested 3 h after Atc addition and assayed for β-galactosidase activity. Specific activities in all samples were normalized to the levels of the corresponding strain containing the empty vector to yield percentage relative activity (reported in the graph). Specific activity values in Miller Units are reported above each bar. The results (mean ± S.D.) were from three biological replicates ($n = 3$).

Thus, in the absence of L3, L2 could still interact with the RBS of *panD*, leading to partial inhibition as demonstrated by 6C ΔL3 (Figure 4B).

To test if the repression of the *panD-M2′-′lacZ* fusion activity could be restored by reciprocal mutations in 6C sRNA, we transformed the 6C ΔL2M2 construct generated earlier (Figure 2D) into cells containing the *panD-M2′-′lacZ* fusion, which would restore perfect base-pairing between the L3 of 6C sRNA and the RBS of *panD* (Figure 4E). Indeed, overexpression of 6C ΔL2M2 resulted in >60% reduction of the *panD-M2′-′lacZ* fusion activity, whereas overexpression of WT 6C sRNA had no effect (Figure 4F).

Taken together, the above data provides genetic evidence for base-pairing of 6C sRNA with *panD* mRNA.

Biochemical demonstration of 6C sRNA binding to *panD* mRNA

To determine whether 6C sRNA can directly bind *panD* mRNA and whether the deletion of L2 and L3 affects the binding, we performed an electrophoretic mobility shift assay (EMSA) of RNA-RNA complexes. 6C sRNA and its deletion mutants, as well as the 5′ UTR of *panD* mRNA were synthesized by *in vitro* transcription. The *panD* mRNA was labeled with ^{32}P and mixed with increasing concentrations of 6C WT and deletion mutants. Analysis by polyacrylamide native gel revealed that 6C WT bound *panD* mRNA with a high affinity and yielded a saturation binding curve (Figure 5A and B). The curve was fitted well with a single-site binding model ($R^2 = 0.97$), which generated a K_D of 0.38 μM (Figure 5B). 6C ΔL2 also bound *panD* mRNA efficiently, though with a lower affinity ($K_D = 2.5$ μM, single-site fitting, $R^2 = 0.96$) compared to 6C WT (Figure 5C and D). In contrast, the binding affinity of 6C ΔL3 was greatly reduced and only ∼20% of *panD* mRNA was bound even at the highest 6C ΔL3:*panD* mRNA ratio (1300:1) (Figure 5E). As expected, the binding activity of 6C ΔL2ΔL3 was completely abolished (Figure 5F). The *in vitro panD* mRNA binding activity of 6C WT and mutants correlated

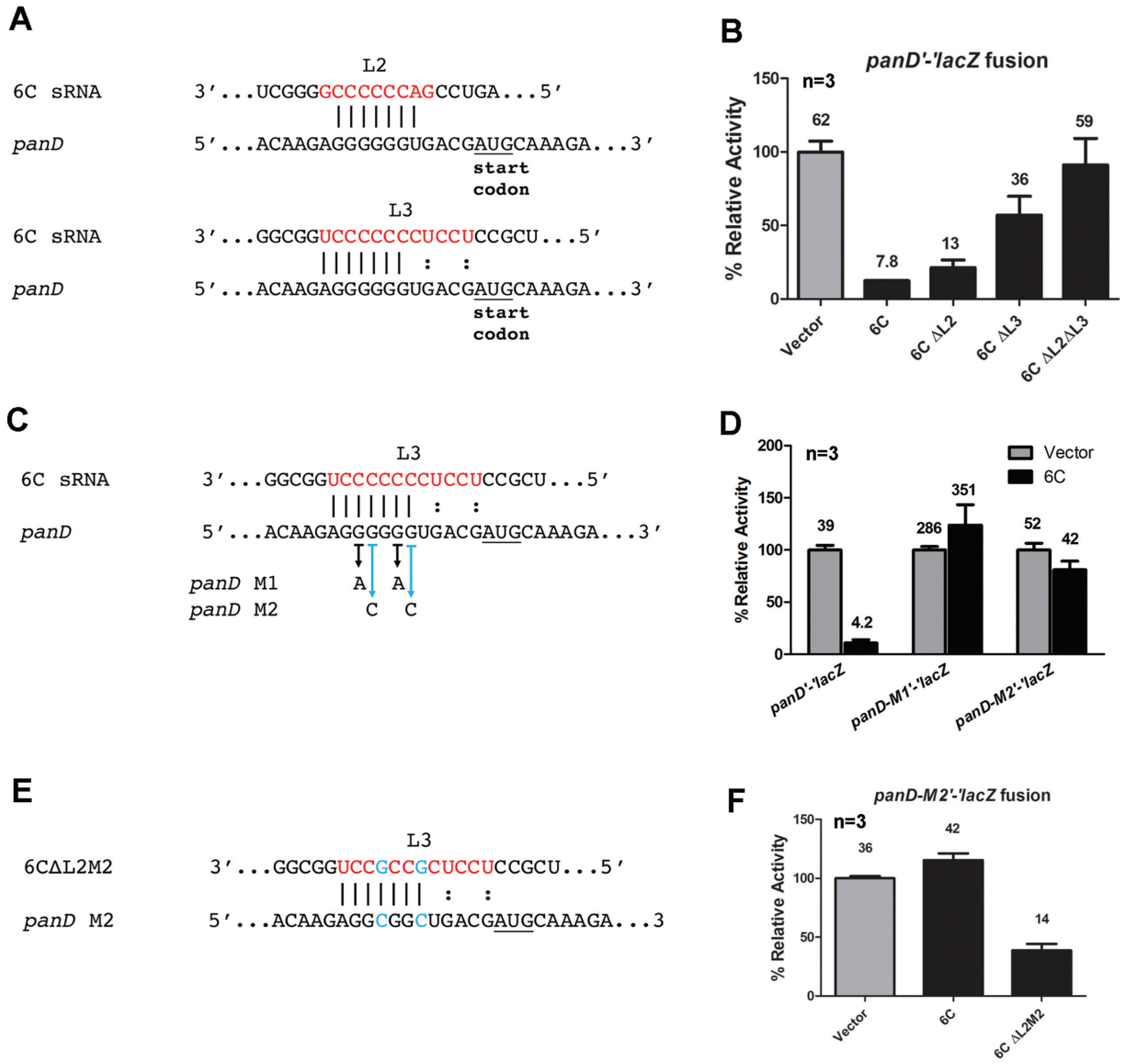

Figure 4. Genetic demonstration of 6C sRNA base-pairing with *panD* mRNA. (**A**) Predicted interactions of L2 and L3 (marked in red) of 6C sRNA with the 5′ UTR of *panD*. (**B**) L3 is required for 6C sRNA interactions with *panD*. The β-galactosidase activity was measured in *M. smegmatis* strains harboring the chromosomal *panD*′-′*lacZ* fusion and plasmids carrying 6C sRNA or the loop deletion mutants (6C ΔL2, ΔL3 and ΔL2ΔL3). Specific activities were normalized to the levels of the corresponding strain containing the empty vector. Specific activity values in Miller Units are reported above each bar. (C and D) Mutations in the G-rich RBS of *panD* that disrupt the base-pairing with the C-rich L3 abolished the interaction of 6C sRNA with *panD*. Two double mutations in the 5′ UTR of *panD* were generated (**C**) and the effect on interactions with 6C sRNA were assayed (**D**). (E and F) Reciprocal mutations in L3 of 6C sRNA restored the interaction with *panD* mutants. Two mutations in the C-rich L3 of 6C sRNA predicted to restore the base-pairing with the *panD* M2 mutant was constructed (**E**), and the effect on the fusion activity of *panD-M2*′-′*lacZ* were assayed (**F**). The results (mean ± S.D.) in (B), (D) and (F) were from three biological replicates (n=3).

with their *in vivo* inhibition of the *panD*′-′*lacZ* fusion activity (Figure 4B), thus providing strong evidence for direct interactions between 6C sRNA and *panD* mRNA.

In vivo validation of multiple mRNA targets of 6C sRNA

Having established that the C-rich L3 of 6C sRNA is critical for interactions with mRNA targets, we searched the 162 downregulated genes upon 6C sRNA overexpression to identify the potential direct targets of 6C sRNA. The –50 to +25 sequences (relative to the start codon) of these downregulated genes were extracted and their base-pairing with the L3 of 6C sRNA were analyzed by IntaRNA, a program for prediction of interactions between two RNA molecules (30). This resulted in a list of 47 candidates, including the validated *panD* (highlighted in Dataset S1).

We randomly selected 15 of the 47 genes for validation using the 5′ UTR mRNA-*lacZ* fusions. Similar to the experiments described above for *panD*, 5′ UTRs of the 15 genes (Supplementary Data 1) were translationally fused with *lacZ* and the ability of 6C sRNA to inhibit the fusion activities was determined. Remarkably, overexpression

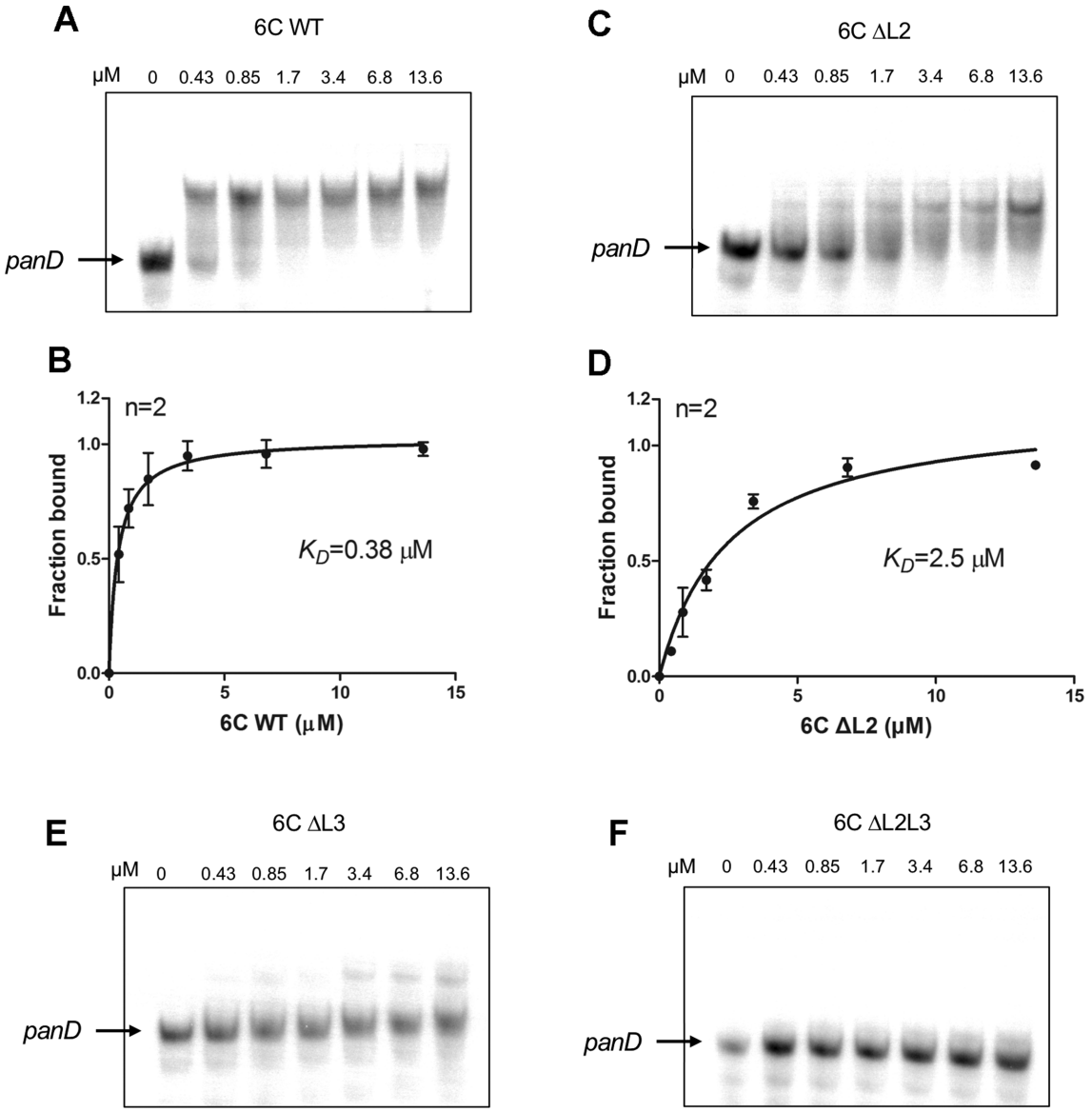

Figure 5. Biochemical demonstration of 6C sRNA binding to *panD* mRNA. The 5′ UTR of *panD* mRNA was transcribed *in vitro*, purified and end-labeled with ^{32}P. The ^{32}P-labeled *panD* mRNA (0.1 pmol) was mixed with the indicated concentrations of 6C WT (**A**) or loop deletion mutants (**C**, **E** and **F**) in 10 μL reaction mixtures. The mixtures were then separated on 5% polyacrylamide native gels. The gels were dried, exposed on the phosphor storage screen, and images acquired by a Typhoon FLA 9500 laser scanner. Two independent experiments were performed ($n = 2$). For 6C WT and 6C ΔL2, the intensities of bands were quantified using ImageJ. The fraction of bound mRNA was calculated and fitted using a single-site binding model (Prism). The resulting K_D value for 6C WT (**B**) or 6C ΔL2 (**D**) binding to *panD* mRNA was shown. Data fitting was not performed for 6C ΔL3 and ΔL2ΔL3 since there was little or no binding to *panD* mRNA.

of 6C sRNA reduced the activities of all 15 gene fusions, ranging from 40% to 90% reduction (Figure 6A).

Interestingly, secondary structure prediction by Mfold (36) revealed that 14 of the 16 validated gene targets (except *MSMEG_0408* and *mce1F*) have single-stranded G-rich sequences in their 5′ UTRs. Many of them are in loops and several examples are shown in Figure 6B and the rest are shown in Supplementary Figure S5. Presumably, these unpaired G-rich sequences could be involved in interactions with 6C sRNA.

Mechanisms of 6C sRNA mediated growth inhibition

To identify gene targets of 6C sRNA that might explain the growth inhibition phenotype, we focused on two genes, *panD* and *dnaB*. The *panD* gene encodes aspartate decarboxylase that converts L-aspartate to β-alanine, a precursor for pantothenate. Since a double deletion mutant of *M. tb* in *panC* and *panD* is an auxotrophy for pantothenate (37), repression of *panD* by 6C sRNA might have the same effect and result in growth arrest. However, supplementation of pantothenate (24 μg/ml) in 7H9 media did not restore the growth (Supplementary Figure S6).

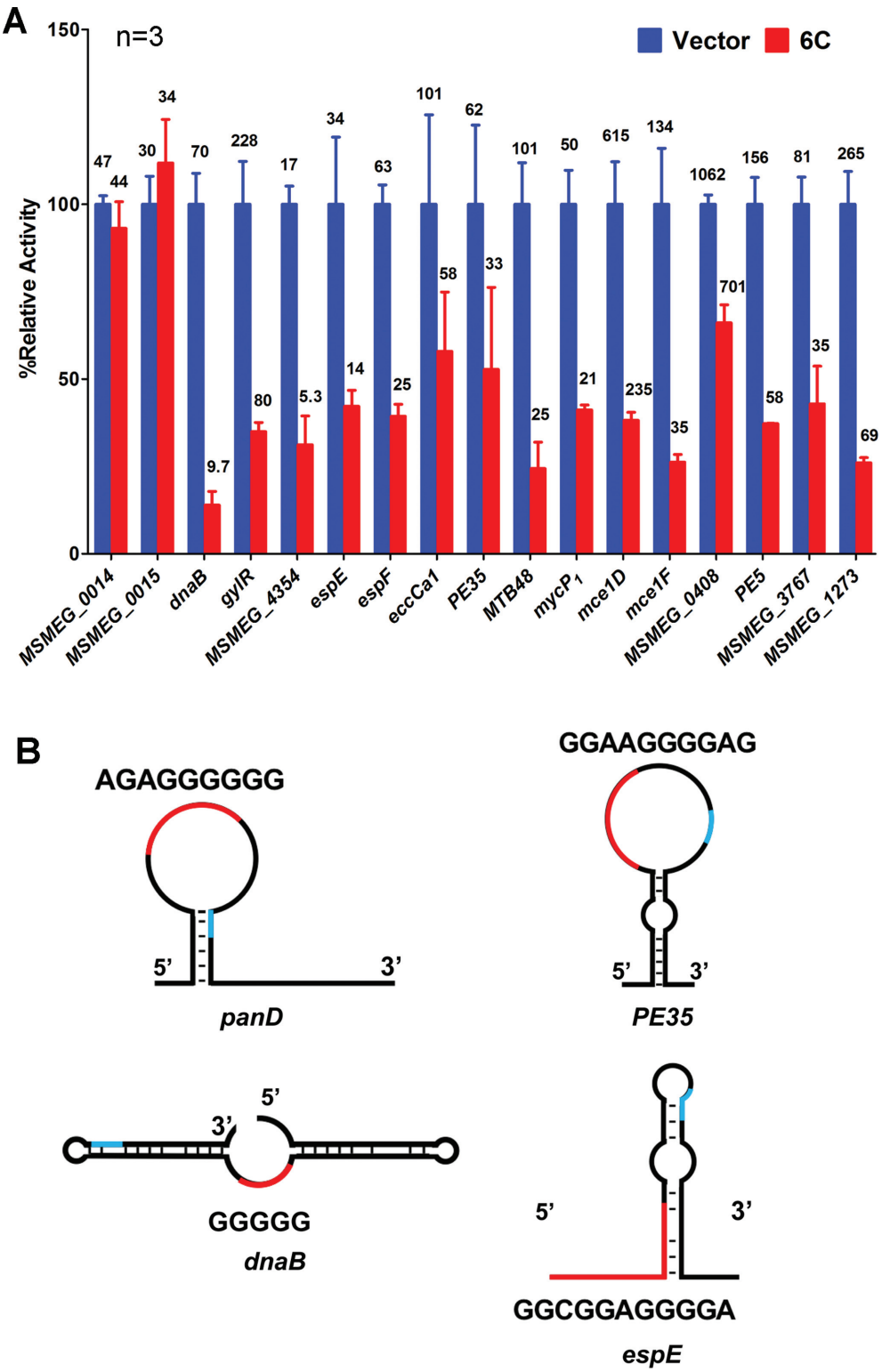

Figure 6. Validation of multiple targets of 6C sRNA. (**A**) Validation of 15 candidate target genes. 5′ UTRs of 15 candidate target genes and two negative control genes (*MSMEG_0014* and *MSMEG_0015*) were fused with *lacZ* as described in Figure 3A and B, and their interactions with 6C sRNA were assayed by measuring β-galactosidase activities. Specific activities in all samples were normalized to the levels of the corresponding strain containing the empty vector to yield percentage relative activity (reported in the graph). Specific activity values in Miller Units are reported above each bar. The results (mean ± S.D.) were from three biological replicates ($n = 3$). (**B**) Secondary structures of the 5′ UTRs of five validated 6C sRNA mRNA targets based on prediction by Mfold. G-rich stretches in unpaired regions (loops or single stranded regions) are highlighted in red, and the sequences were shown. The translation start sites (AUG or GUG) were indicated in blue.

The *dnaB* gene encodes the replicative DNA helicase, which is an essential protein that unwinds the DNA duplex ahead of the replication fork during bacterial DNA replication (38). We showed above that *dnaB* is directly regulated by 6C sRNA at the posttranscriptional level (Figure 6A). The 5′ UTR of *dnaB* contains 5 contiguous G that are unpaired (Figure 6B), which could be the direct binding site of 6C sRNA. Consistently, we found that the C-rich L3 of 6C sRNA is required for interactions with the 5′ UTR of *dnaB* (Figure 7A).

To map the interaction sites, we used the *dnaB′-′lacZ* fusion construct as the template and made three *dnaB* mutants, including single-, double-, or triple-nucleotide mutations at the RBS of *dnaB* (Figure 7B). The effect of these mutations on the 6C sRNA mediated translational repression was determined by co-transforming with the 6C sRNA overexpression vector. For *dnaB* M1, the intrinsic translational strength was greatly reduced by the G→A mutation, thus the result was inconclusive (Figure 7C). However, in *dnaB* M2 and M3, where the mutations had only partial or little effect on the intrinsic translational efficiency, the 6C mediated translational inhibition was completely abolished in *dnaB* M3, and partially abrogated in *dnaB* M2 (Figure 7C).

Since *dnaB* is an essential gene, we used the tetracycline inducible CRISPR interference (CRISPRi) system (32) to conditionally knock down *dnaB*. RT-PCR analysis showed that this resulted in a 4.6-fold reduction in mRNA transcript of *dnaB*, similar to the effect of 6C sRNA overexpression; the *dnaB* transcript was downregulated 4.2-fold after the induction of 6C sRNA overexpression (Figure 7D). Interestingly, CRISPRi treatment resulted in elongated cells that were similar in size as those overexpressing 6C sRNA. WT *M. smegmatis* contained only one or two nucleoids per cell. In contrast, cells overexpressing 6C sRNA or treated with CRISPRi both contained many more and smaller nucleoids along the cell filament (Figure 7E). These phenotypes are reminiscent of the growth defect observed in the *dnaA* and *dnaG* conditional deletion mutants of *M. smegmatis* (39,40), in which chromosomal replication is also disrupted at the initiation steps.

Taken together, this data suggests that repression of *dnaB* by 6C sRNA likely explains the growth inhibition phenotype.

6C sRNA mediated growth inhibition is independent of co-factors

We next sought to address whether co-factors (e.g. RNA binding proteins) are required for interactions of 6C sRNA and mRNA targets. We took advantage of the growth inhibition upon 6C sRNA overexpression and screened for mutations that could derepress this phenotype, with the assumption that such co-factors, if required, are not essential genes. *M. smegmatis* harboring the inducible overexpression vector of 6C sRNA (Figure 1A) were subjected to transposon mutagenesis. Approximately 10,000 transposon insertion mutants were generated and plated on 7H11 agar plates containing Atc (100 ng/ml) and incubated for growth at 37°C for 4 days. Three colonies appeared and were selected and grown in 7H11 medium. Sequencing analysis showed that all three colonies had mutations in the 6C sRNA sequence in the overexpression vector and thus produced false positive results: two of the mutations deleted or truncated the 6C sRNA coding sequence and one mutation affected the $P_{rrnB\text{-}tetO}$ (Figure 1A) used to drive 6C sRNA overexpression. We repeated this experiment with the modification of plating more colonies (∼20 000). No colony appeared on 7H11 agar plates containing Atc (100 ng/ml) after incubation. These results support the notion that base-pairing between the C-rich loop of 6C sRNA and the G-rich sequence of mRNA target is sufficient for their interactions and 6C sRNA function.

DISCUSSION

An intriguing feature of sRNAs in high GC Gram-positive bacteria including *Mycobacterium*, *Streptomyces* and *Corynebacterium* is the prevalence of unpaired C-rich stretches in their structures (19,21,41–43). Initially described in *Streptomyces*, C-rich stretches in loops or single-stranded regions are present in 9 of 25 sRNAs of *M. tb* that have been experimentally verified by Northern Blot (21) (Supplementary Figure S7). The function of the C-rich stretches in any sRNA has not been studied in high GC Gram-positive bacteria. One hypothesis is that C-rich stretch stem-loops act as alternative termination structures to known Rho-independent termination (41). Here, we demonstrated that the C-rich loops of 6C sRNA act as the direct binding site to mRNA targets through base-pairing. Our finding has strong implications for target recognition and mechanism of action of a large group of sRNAs in high GC Gram-positive bacteria including the important human pathogen *M. tb*. In addition, our study provides a framework to systematically identify and validate targets of any other sRNA.

Our findings also provide insights into an important question - can sRNAs in high GC Gram-positive bacteria interact with their targets unaided or require a chaperone? In Gram-negative bacteria, the RNA chaperone Hfq is required to facilitate the short and imperfect base pairing interactions of many *trans*-encoded sRNAs with mRNA targets (4,44,45). Hfq is a global regulator and disruption of the *hfq* gene led to a pleotropic phenotype in various bacteria, including increased sensitivity to stress conditions and reduced virulence (46). Orthologs of *hfq* are found in about half of the sequenced bacteria, including low GC Gram-positive bacteria such as *S. aureus*, *L. monocytogenes* and *B. subtilis* (47). However, Hfq and the recently identified ProQ are completely absent in high GC Gram-positive bacteria including *Mycobacterium* and *Streptomyces* (14,47). It has been postulated that alternative, hitherto unidentified chaperones may be present in these organisms, since their high GC content limit the frequency of AU-rich stretches involved in interactions with conventional Hfq (48). Alternatively, sRNAs may interact with mRNA targets via G-C pairing and do not require chaperones (16,48). G-C base pairs have significant base-stacking interactions (especially repeated G-C pairs) and can form three hydrogen bonds between each other, resulting in a stable RNA duplex. Our results favor the latter hypothesis. We showed that 6C sRNA interacted with *panD* mRNA through perfect G-C pair-

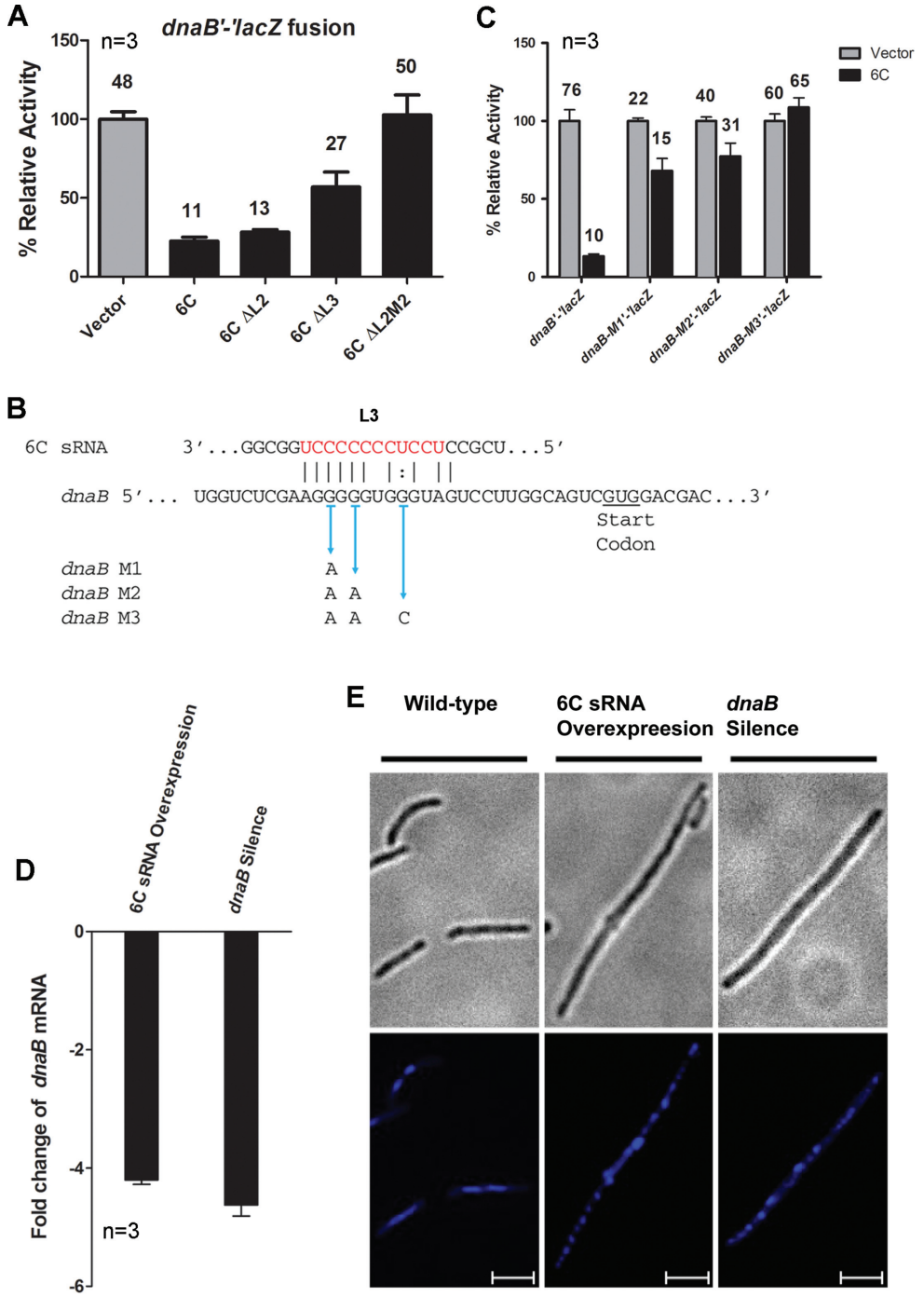

Figure 7. 6C sRNA represses *dnaB* and leads to growth inhibition. (**A**) The C-rich L3 of 6C sRNA is required for interactions with the 5′ UTR of *dnaB*. The β-galactosidase activity was measured in *M. smegmatis* strains harboring the chromosomal *dnaB′-′lacZ* fusion and plasmids carrying 6C sRNA, the loop deletion mutants (6C ΔL2, ΔL3), or the L3 point mutations of 6C ΔL2 (6C ΔL2M2). Specific activities were normalized to the levels of the corresponding strain containing the empty vector. Specific activity values in Miller Units are reported above each bar. The results (mean ± S.D.) were from three biological replicates ($n = 3$). (**B**) Mutations in the G-rich RBS of *dnaB* that disrupt the base-pairing with the C-rich L3. Three mutants were generated and indicated. (**C**) Interactions of *dnaB* mutants with 6C sRNA were assayed. Specific activities were normalized to the levels of the corresponding strain containing the empty vector. Specific activity values in Miller Units are reported above each bar. The results (mean ± S.D.) were from three biological replicates ($n = 3$). (**D**) Quantitative RT-PCR analysis of the effect of 6C sRNA overexpression or CRISPRi silencing on the *panD* mRNA level. *M. smegmatis* carrying the Atc inducible 6C sRNA overexpression vector were added with Atc (100 ng/ml) for 3 hr and the *panD* mRNA was measured and compared before and after the addition of Atc. *M. smegmatis* carrying the Atc inducible *dnaB* CRISPRi knockdown vector pLJR962::*dnaB* were added with Atc (100 ng/ml) for 12 hr, and the *panD* mRNA was measured and compared before and after the addition of Atc. Three independent experiments were performed, and the data were plotted as mean ± S.D. ($n = 3$). (**E**) Overexpression of 6C sRNA phenocopied the effect of CRISPRi silencing of *dnaB*. *M. smegmatis* strains carrying the Atc inducible 6C sRNA overexpression vector (Figure 1A) or the Atc inducible *dnaB* CRISPRi knockdown vector pLJR962::*dnaB* were grown to log phase and then added with 100 ng/ml of ATc. Cell morphology was examined 24 hr after addition of Atc. DNA was stained in blue by DAPI. The fluorescence images were acquired by the Nikon Eclipse 80i microscope. The scale bar is 2.5 μm.

ing between RNA loops. A minimum of 5 G-C pairs appeared to be required since point mutations that disrupt the base-pairing abolished their interactions (Figure 4). The direct interaction of 6C sRNA and *panD* mRNA was supported by biochemical evidence that 6C sRNA bound *panD* mRNA with a high affinity and that deletion of L3 greatly reduced the binding affinity (Figure 5). Similarly, we suggested that the C-rich L3 of 6C sRNA interacted with the single stranded G-stretches in *dnaB* (Figure 7), and that the repression of *dnaB* by 6C sRNA is likely responsible for the growth inhibition. Saturated mutagenesis of the whole genome did not yield mutations that could derepress the growth inhibition by 6C sRNA, which argued against the requirement of non-essential chaperones for 6C sRNA and *dnaB* interactions. This is reminiscent of a small RNA in low GC Gram-positive bacterium *S. aureus* called RNAIII (49). The interactions of RNAIII with mRNA targets were mediated through multiple G-C pairs involving loop-loop interactions, and deletion of *hfq* did not affect these interactions (49). However, unlike in RNAIII, where two loop-loop interactions were required for its function, one loop-loop or loop-unpaired region interaction (mediated by L3) in 6C sRNA appeared to play a dominant role for repression of *panD* and *dnaB in vivo* (Figures 4A and 7A), and for interactions with *panD* mRNA *in vitro* (Figure 5). While L2 of 6C sRNA could also form multiple base pairs with mRNA targets (Figure 4A), it consistently showed a less critical role than L3 in all the functional assays we performed. One possible explanation is that L2 might be buried in the tertiary fold of 6C sRNA thus not accessible for binding. Deletion of L3 might have changed the tertiary structure and partially exposed L2, which would explain the observation that growth inhibition mediated by 6C overexpression was only partially abrogated by the deletion of L3, whereas it was completely abolished by the deletion of both L2 and L3 (Figure 2A).

Taken together, our results suggest that the C-rich loops of 6C sRNA mediate the binding to mRNAs and confer the target specificity, and that the interactions of 6C sRNA with mRNA targets is independent of RNA chaperones. On the other hand, in two of the validated targets of 6C sRNA (*MSMEG_0408* and *mce1F*), the G-rich stretches are not predicted in loops or single-stranded regions in their structures (Supplementary Figure S7). It remains a possibility that specific RNA chaperones are required for structural changes of these mRNA targets, allowing the G-C pairing to occur.

We searched the genome of *M. smegmatis* for the occurrence of 6 contiguous G, which would represent the potential binding sites of 6C sRNA. A total of 433 sites were found, which is far below the theoretical number of 6G stretches calculated (∼10,000) based on the GC content (67.4%) and the genome size of *M. smegmatis* (6.9 Mb). For comparison, 133 sites of 6A stretches were found in the genome, which is close to the predicted number of 157. If we allowed one mismatch at any of the six positions, the total number of 5Gs found in the genome was 41,047, which was still about five times fewer than the expected number (182,238) (Supplementary Table S1). In contrast, the number of 5As found in the genome was close to and even slightly higher than the expected value (Supplementary Table S1). This suggests that the mycobacterial genome might have evolved to limit the frequency of G-rich stretches. Some of the G-rich stretches located at the 5′ UTRs are likely the RBS. While the Shine-Dalgarno RBS is not well defined in mycobacteria (50), several studies reveal an enrichment of purines displayed on the upstream region of mRNAs in *M. tb* and *M. smegmatis* (24,51). Consistently, we found that mutations in the G-rich 5′ UTR sequences strongly affected the translational efficiency of *panD* and *dnaB* (e.g. *panD* M1 in Figure 4D and *dnaB* M1 in Figure 7C). Collectively, our data suggest that 6C sRNA binds the RBS of *panD* and *dnaB* mRNAs, leading to translational repression and/or mRNA degradation.

The physiological role of 6C sRNA is less clear since we were unable to construct knock out or knock down mutants despite multiple attempts. The majority of genes identified upon 6C sRNA overexpression are unlikely direct targets. On the other hand, the knowledge on the 6C sRNA binding mechanism allowed us to identify 47 genes as possible direct targets. Of these, 15 genes randomly selected for validation all turned out to be positive (Figure 6). This result suggests that 6C sRNA may play a pleotropic role in mycobacteria, and some of these functions might be conserved in high GC Gram-positive bacteria. The G-rich region of 5′ UTR of *dnaB* mRNA is conserved among mycobacteria and a consensus sequence can be generated (Supplementary Figure S8A). We found that this consensus sequence was also bound by 6C sRNA using the *in vivo* translational *lacZ* fusion system (Supplementary Figure S8B). The 5′ UTRs of *dnaB* mRNA from *S. coelicolor* and *C. glutamicum* also contain a G-rich sequence, which could potentially bind their cognate 6C sRNA through the C-rich loops (L2 and L3) (Supplementary Figure S8C and D). Targeting *dnaB* may provide an explanation for the previous observations that overexpression of 6C sRNA caused a defective growth in *C. glutamicum* (23) and was associated with hyphae formation and sporulation in *S. coelicolor* (20). Future studies to test this hypothesis are warranted.

Besides being implicated in DNA replication and cell division through direct regulation of *dnaB*, 6C sRNA also plays a role in regulation of genes associated with mycobacterial virulence and host cell interactions. Mycobacteria contain a specialized protein secretion system called Type VII secretion system (52). Pathogenic mycobacteria such as *M. tb* have up to five of these secretion systems (ESX-1 to -5), with ESX-1 being a major virulence factor (53–55). The absence of ESX-1 is responsible for the attenuation of the live-vaccine strain *Mycobacterium bovis* BCG (56,57). Given its central role in virulence, the ESX-1 system of *M. tb* has been the focus of extensive study. Nonpathogenic mycobacterium *M. smegmatis* contains three ESX systems (ESX-1, -3 and -4), where ESX-1 is required for DNA conjugation (58,59). Remarkably, every gene in the *esx-1* locus of *M. smegmatis* (*MSMEG_0055* to *_0070*, *MSMEG_0076* to *_0084*) was downregulated upon 6C sRNA overexpression (Dataset S1), and we have validated several of them as direct targets including *MSMEG_0055* (*espE*), *MSMEG_0056* (*espF*), *MSMEG_0061* (*eccA1*), *MSMEG_0063* (*PE35*), *MSMEG_0076* (*MTB84*) and *MSMEG_0083* (*mycP1*) (Figure 6). Since gene organization of the ESX-1 system is

highly conserved between *M. smegmatis* and *M. tb*(58,59), it is reasonable to assume that 6C sRNA would have a similar effect on *esx-1* in *M. tb*. Several regulatory proteins controlling the expression of the *esx-1* system have previously been identified, including PhoP (60), WhiB6 (61), EspR (62,63), MprA (64) and Lsr2 (25). The fact that all genes in the *esx-1* locus were regulated by 6C sRNA suggests that it is a major regulator of this secretion system. To our knowledge, this is the first time a small RNA has been implicated in regulation of the *esx-1* system. 6C sRNA is also accumulated upon H_2O_2 stress (21). An emerging picture is that the ESX-1 secretion system is tightly controlled by multiple layers of regulation adapted to different conditions. Since 6C sRNA and the type VII secretion systems are both present in other high GC Gram-positive bacteria including *Streptomyces* and *Corynebacterium*, it is reasonable to assume that similar regulation mechanisms would also occur in those bacteria. Future studies will confirm this hypothesis.

DATA AVAILABILITY

The RNA-seq data are presented as Dataset S1 and Dataset S2 in the supplementary materials. The RNA-seq data has been deposited in GEO, accession number is GSE119364. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE119364

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

We thank Jeremy Rock and Sarah Fortune for providing the CRISPRi system for silencing of mycobacterial genes.
Author Contributions: J.L. conceptualized the study. J.M., C.R., L.Z. and J.L. designed experiments. J.M., C.R., J.W., X.S. and C.L. performed experiments. L.Z. and J.L. supervised experiments. J.M. and J.L. wrote the manuscript.

FUNDING

Canadian Institutes of Health Research [MOP-15107, PJT-156261 to J.L.]. Funding for open access charge: Canadian Institute of Health Research.
Conflict of interest statement. None declared.

REFERENCES

1. Wassarman,K.M. (2002) Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell*, **109**, 141–144.
2. Storz,G., Vogel,J. and Wassarman,K.M. (2011) Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell*, **43**, 880–891.
3. Desnoyers,G., Bouchard,M.P. and Masse,E. (2013) New insights into small RNA-dependent translational regulation in prokaryotes. *Trends Genet.*, **29**, 92–98.
4. Vogel,J. and Luisi,B.F. (2011) Hfq and its constellation of RNA. *Nat. Rev. Microbiol.*, **9**, 578–589.
5. Updegrove,T.B., Zhang,A. and Storz,G. (2016) Hfq: the flexible RNA matchmaker. *Curr. Opin. Microbiol.*, **30**, 133–138.
6. Hor,J., Gorski,S.A. and Vogel,J. (2018) Bacterial RNA biology on a genome scale. *Mol. Cell*, **70**, 785–799.
7. Melamed,S., Peer,A., Faigenbaum-Romm,R., Gatt,Y.E., Reiss,N., Bar,A., Altuvia,Y., Argaman,L. and Margalit,H. (2016) Global mapping of small RNA-Target interactions in bacteria. *Mol. Cell*, **63**, 884–897.
8. Brantl,S. and Bruckner,R. (2014) Small regulatory RNAs from low-GC Gram-positive bacteria. *RNA Biol.*, **11**, 443–456.
9. Schumacher,M.A., Pearson,R.F., Moller,T., Valentin-Hansen,P. and Brennan,R.G. (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J.*, **21**, 3546–3556.
10. Someya,T., Baba,S., Fujimoto,M., Kawai,G., Kumasaka,T. and Nakamura,K. (2012) Crystal structure of Hfq from Bacillus subtilis in complex with SELEX-derived RNA aptamer: insight into RNA-binding properties of bacterial Hfq. *Nucleic Acids Res.*, **40**, 1856–1867.
11. Kovach,A.R., Hoff,K.E., Canty,J.T., Orans,J. and Brennan,R.G. (2014) Recognition of U-rich RNA by Hfq from the Gram-positive pathogen Listeria monocytogenes. *RNA*, **20**, 1548–1559.
12. Smirnov,A., Forstner,K.U., Holmqvist,E., Otto,A., Gunster,R., Becher,D., Reinhardt,R. and Vogel,J. (2016) Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 11591–11596.
13. Attaiech,L., Boughammoura,A., Brochier-Armanet,C., Allatif,O., Peillard-Fiorente,F., Edwards,R.A., Omar,A.R., MacMillan,A.M., Glover,M. and Charpentier,X. (2016) Silencing of natural transformation by an RNA chaperone and a multitarget small RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 8813–8818.
14. Olejniczak,M. and Storz,G. (2017) ProQ/FinO-domain proteins: another ubiquitous family of RNA matchmakers? *Mol. Microbiol.*, **104**, 905–915.
15. Haning,K., Cho,S.H. and Contreras,L.M. (2014) Small RNAs in mycobacteria: an unfolding story. *Front. Cell Infect. Microbiol.*, **4**, 96.
16. Heueis,N., Vockenhuber,M.P. and Suess,B. (2014) Small non-coding RNAs in streptomycetes. *RNA Biol.*, **11**, 464–469.
17. Solans,L., Gonzalo-Asensio,J., Sala,C., Benjak,A., Uplekar,S., Rougemont,J., Guilhot,C., Malaga,W., Martin,C. and Cole,S.T. (2014) The PhoP-dependent ncRNA Mcr7 modulates the TAT secretion system in Mycobacterium tuberculosis. *PLoS Pathog.*, **10**, e1004183.
18. Gerrick,E.R., Barbier,T., Chase,M.R., Xu,R., Francois,J., Lin,V.H., Szucs,M.J., Rock,J.M., Ahmad,R., Tjaden,B. *et al.* (2018) Small RNA profiling in Mycobacterium tuberculosis identifies MrsI as necessary for an anticipatory iron sparing response. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 6464–6469.
19. Weinberg,Z., Barrick,J.E., Yao,Z., Roth,A., Kim,J.N., Gore,J., Wang,J.X., Lee,E.R., Block,K.F., Sudarsan,N. *et al.* (2007) Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res.*, **35**, 4809–4819.
20. Swiercz,J.P., Hindra, Bobek,J., Bobek,J., Haiser,H.J., Di Berardo,C., Tjaden,B. and Elliot,M.A. (2008) Small non-coding RNAs in Streptomyces coelicolor. *Nucleic Acids Res.*, **36**, 7240–7251.
21. Arnvig,K.B. and Young,D.B. (2009) Identification of small RNAs in Mycobacterium tuberculosis. *Mol. Microbiol.*, **73**, 397–408.
22. Lamichhane,G., Arnvig,K.B. and McDonough,K.A. (2013) Definition and annotation of (myco)bacterial non-coding RNA. *Tuberculosis (Edinb.)*, **93**, 26–29.
23. Pahlke,J., Dostalova,H., Holatko,J., Degner,U., Bott,M., Patek,M. and Polen,T. (2016) The small 6C RNA of Corynebacterium glutamicum is involved in the SOS response. *RNA Biol.*, **13**, 848–860.
24. Shell,S.S., Wang,J., Lapierre,P., Mir,M., Chase,M.R., Pyle,M.M., Gawande,R., Ahmad,R., Sarracino,D.A., Ioerger,T.R. *et al.* (2015) Leaderless transcripts and small proteins are common features of the mycobacterial translational landscape. *PLos Genet.*, **11**, e1005641.
25. Gordon,B.R., Li,Y., Wang,L., Sintsova,A., van Bakel,H., Tian,S., Navarre,W.W., Xia,B. and Liu,J. (2010) Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 5154–5159.
26. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.

27. Anders,S., Pyl,P.T. and Huber,W. (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
28. Robinson,M.D. and Oshlack,A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.*, **11**, R25.
29. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
30. Mann,M., Wright,P.R. and Backofen,R. (2017) IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic Acids Res.*, **45**, W435–W439.
31. Gonzalo-Asensio,J., Soto,C.Y., Arbues,A., Sancho,J., del Carmen Menendez,M., Garcia,M.J., Gicquel,B. and Martin,C. (2008) The Mycobacterium tuberculosis phoPR operon is positively autoregulated in the virulent strain H37Rv. *J. Bacteriol.*, **190**, 7068–7078.
32. Rock,J.M., Hopkins,F.F., Chavez,A., Diallo,M., Chase,M.R., Gerrick,E.R., Pritchard,J.R., Church,G.M., Rubin,E.J., Sassetti,C.M. *et al.* (2017) Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nat. Microbiol.*, **2**, 16274.
33. Sassetti,C.M., Boyd,D.H. and Rubin,E.J. (2001) Comprehensive identification of conditionally essential genes in mycobacteria. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 12712–12717.
34. Lorenz,R., Bernhart,S.H., Honer Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
35. Moores,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. *PLoS One*, **12**, e0174079.
36. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
37. Sambandamurthy,V.K., Wang,X., Chen,B., Russell,R.G., Derrick,S., Collins,F.M., Morris,S.L. and Jacobs,W.R. Jr (2002) A pantothenate auxotroph of Mycobacterium tuberculosis is highly attenuated and protects mice against tuberculosis. *Nat. Med.*, **8**, 1171–1174.
38. Zhang,H., Zhang,Z., Yang,J. and He,Z.G. (2014) Functional characterization of DnaB helicase and its modulation by single-stranded DNA binding protein in Mycobacterium tuberculosis. *FEBS J.*, **281**, 1256–1266.
39. Klann,A.G., Belanger,A.E., Abanes-De Mello,A., Lee,J.Y. and Hatfull,G.F. (1998) Characterization of the dnaG locus in Mycobacterium smegmatis reveals linkage of DNA replication and cell division. *J. Bacteriol.*, **180**, 65–72.
40. Greendyke,R., Rajagopalan,M., Parish,T. and Madiraju,M.V. (2002) Conditional expression of Mycobacterium smegmatis dnaA, an essential DNA replication gene. *Microbiology*, **148**, 3887–3900.
41. Panek,J., Bobek,J., Mikulik,K., Basler,M. and Vohradsky,J. (2008) Biocomputational prediction of small non-coding RNAs in Streptomyces. *BMC Genomics*, **9**, 217.
42. Vockenhuber,M.P., Sharma,C.M., Statt,M.G., Schmidt,D., Xu,Z., Dietrich,S., Liesegang,H., Mathews,D.H. and Suess,B. (2011) Deep sequencing-based identification of small non-coding RNAs in Streptomyces coelicolor. *RNA Biol.*, **8**, 468–477.
43. Moody,M.J., Young,R.A., Jones,S.E. and Elliot,M.A. (2013) Comparative analysis of non-coding RNAs in the antibiotic-producing Streptomyces bacteria. *BMC Genomics*, **14**, 558.
44. Brennan,R.G. and Link,T.M. (2007) Hfq structure, function and ligand binding. *Curr. Opin. Microbiol.*, **10**, 125–133.
45. Valentin-Hansen,P., Eriksen,M. and Udesen,C. (2004) The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol. Microbiol.*, **51**, 1525–1533.
46. Sobrero,P. and Valverde,C. (2012) The bacterial protein Hfq: much more than a mere RNA-binding factor. *Crit. Rev. Microbiol.*, **38**, 276–299.
47. Sun,X., Zhulin,I. and Wartell,R.M. (2002) Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res.*, **30**, 3662–3671.
48. Arnvig,K. and Young,D. (2012) Non-coding RNA and its potential role in Mycobacterium tuberculosis pathogenesis. *RNA Biol.*, **9**, 427–436.
49. Geissmann,T., Chevalier,C., Cros,M.J., Boisset,S., Fechter,P., Noirot,C., Schrenzel,J., Francois,P., Vandenesch,F., Gaspin,C. *et al.* (2009) A search for small noncoding RNAs in Staphylococcus aureus reveals a conserved sequence motif for regulation. *Nucleic Acids Res.*, **37**, 7239–7257.
50. 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.
51. 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.
52. Abdallah,A.M., Gey van Pittius,N.C., Champion,P.A., Cox,J., Luirink,J., Vandenbroucke-Grauls,C.M., Appelmelk,B.J. and Bitter,W. (2007) Type VII secretion–mycobacteria show the way. *Nat. Rev. Microbiol.*, **5**, 883–891.
53. Hsu,T., Hingley-Wilson,S.M., Chen,B., Chen,M., Dai,A.Z., Morin,P.M., Marks,C.B., Padiyar,J., Goulding,C., Gingery,M. *et al.* (2003) The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 12420–12425.
54. Stanley,S.A., Raghavan,S., Hwang,W.W. and Cox,J.S. (2003) Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 13001–13006.
55. Tan,T., Lee,W.L., Alexander,D.C., Grinstein,S. and Liu,J. (2006) The ESAT-6/CFP-10 secretion system of Mycobacterium marinum modulates phagosome maturation. *Cell. Microbiol.*, **8**, 1417–1429.
56. Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K. (1996) Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. *J. Bacteriol.*, **178**, 1274–1282.
57. Behr,M.A., Wilson,M.A., Gill,W.P., Salamon,H., Schoolnik,G.K., Rane,S. and Small,P.M. (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science*, **284**, 1520–1523.
58. Coros,A., Callahan,B., Battaglioli,E. and Derbyshire,K.M. (2008) The specialized secretory apparatus ESX-1 is essential for DNA transfer in Mycobacterium smegmatis. *Mol. Microbiol.*, **69**, 794–808.
59. Gey Van Pittius,N.C., Gamieldien,J., Hide,W., Brown,G.D., Siezen,R.J. and Beyers,A.D. (2001) The ESAT-6 gene cluster of Mycobacterium tuberculosis and other high G+C Gram-positive bacteria. *Genome Biol.*, **2**, RESEARCH0044.
60. Frigui,W., Bottai,D., Majlessi,L., Monot,M., Josselin,E., Brodin,P., Garnier,T., Gicquel,B., Martin,C., Leclerc,C. *et al.* (2008) Control of M. tuberculosis ESAT-6 secretion and specific T cell recognition by PhoP. *PLoS Pathog.*, **4**, e33.
61. Solans,L., Aguilo,N., Samper,S., Pawlik,A., Frigui,W., Martin,C., Brosch,R. and Gonzalo-Asensio,J. (2014) A specific polymorphism in Mycobacterium tuberculosis H37Rv causes differential ESAT-6 expression and identifies WhiB6 as a novel ESX-1 component. *Infect. Immun.*, **82**, 3446–3456.
62. Raghavan,S., Manzanillo,P., Chan,K., Dovey,C. and Cox,J.S. (2008) Secreted transcription factor controls Mycobacterium tuberculosis virulence. *Nature*, **454**, 717–721.
63. Blasco,B., Chen,J.M., Hartkoorn,R., Sala,C., Uplekar,S., Rougemont,J., Pojer,F. and Cole,S.T. (2012) Virulence regulator EspR of Mycobacterium tuberculosis is a nucleoid-associated protein. *PLoS Pathog.*, **8**, e1002621.
64. Pang,X., Samten,B., Cao,G., Wang,X., Tvinnereim,A.R., Chen,X.L. and Howard,S.T. (2013) MprAB regulates the espA operon in Mycobacterium tuberculosis and modulates ESX-1 function and host cytokine response. *J. Bacteriol.*, **195**, 66–75.