Identification of small RNAs in *Mycobacterium smegmatis* using heterologous Hfq

SAI-KAM LI,^{1,4} PATRICK KWOK-SHING NG,^{2,4,5} HAO QIN,³ JEFFREY KWAN-YIU LAU,¹ JONATHAN PAK-YUEN LAU,¹ STEPHEN KWOK-WING TSUI,² TING-FUNG CHAN,^{3,6} and TERRENCE CHI-KONG LAU^{1,6}

¹Department of Biology and Chemistry, City University of Hong Kong, Hong Kong SAR

²School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong SAR

³School of Life Sciences, The Chinese University of Hong Kong, Hong Kong SAR

ABSTRACT

Gene regulation by small RNAs (sRNAs) has been extensively studied in various bacteria. However, the presence and roles of sRNAs in mycobacteria remain largely unclear. Immunoprecipitation of RNA chaperone Hfq to enrich for sRNAs is one of the effective methods to isolate sRNAs. However, the lack of an identified mycobacterial *hfq* restricts the feasibility of this approach. We developed a novel method that takes advantage of the conserved inherent sRNAs-binding capability of heterologous Hfq from *Escherichia coli* to enrich sRNAs from *Mycobacterium smegmatis*, a model organism for studying *Mycobacterium tuberculosis*. We validated 12 *trans*-encoded and 12 *cis*-encoded novel sRNAs in *M. smegmatis*. Many of these sRNAs are differentially expressed at exponential phase compared with stationary phase, suggesting that sRNAs are involved in the growth of mycobacteria. Intriguingly, five of the *cis*-encoded novel sRNAs target known transposases. Phylogenetic conservation analysis shows that these sRNAs are pathogenicity dependent. We believe that our findings will serve as an important reference for future analysis of sRNAs regulation in mycobacteria and will contribute significantly to the development of sRNAs prediction programs. Moreover, this novel method of using heterologous Hfq for sRNAs enrichment can be of general use for the discovery of bacterial sRNAs in which no endogenous Hfq is identified.

Keywords: mycobacteria; *Mycobacterium smegmatis*; small RNA; deep sequencing

INTRODUCTION

Regulatory RNAs in bacteria are mostly not translated and are small in size and thus termed small RNAs (sRNAs). Most sRNAs regulate gene expression and provide quick responses to the changes in environmental conditions such as nutrient deprivation, stress, or virulence conditions (Vogel and Papenfort 2006; Repoila and Darfeuille 2009; Waters and Storz 2009; Papenfort and Vogel 2010). Recent evidence shows that sRNAs act as signal transducers of environmental cues by participating in the precise coordination of gene expression in multiple infectious processes and play key roles in microbial pathogenesis (Papenfort and Vogel 2010).

Regulatory RNAs can be grouped into different classes. One of the classes comprises the *cis*-acting regulatory sRNAs, which are part of the mRNAs they regulate and are usually located within the 5' untranslated region (5' UTR). The other class is the *trans*-acting sRNAs. Many of these sRNAs act by base pairing with target mRNAs and modulate target translational efficiency as well as stability. The base-pairing sRNAs usually target the 5' UTR of genes and compete with ribosome binding to repress protein translation. However, some base-pairing sRNAs can hybridize to the coding regions (CDS) of genes to generate duplex molecules and activate the mRNA degradation pathway by recruiting RNase III or RNase E (Vogel et al. 2004; Morita et al. 2005; Gottesman et al. 2006; Pfeiffer et al. 2009; Waters and Storz 2009). In many cases of sRNAs–mRNA interactions, base pairing is imperfect and contains mismatches and unpaired nucleotides (nt). In order to overcome such interrupted pairings, RNA chaperones such as Hfq are usually required in some bacterial species (Franze de Fernandez et al. 1968; Schumacher et al. 2002; Valentin-Hansen et al. 2004; Aiba 2007; Nielsen et al.

⁴These authors contributed equally to this work.

⁵Present address: Department of Systems Biology, The University of Texas, MD Anderson Cancer Center, Houston, TX 77030, USA

⁶Corresponding authors
E-mail chiklau@cityu.edu.hk
E-mail tf.chan@cuhk.edu.hk

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.034116.112>.

2009). Hfq facilitates sRNAs–mRNA annealing and enhances the rate of duplex formation (Arluison et al. 2007). Hfq protein belongs to the Sm protein family and contains an N-terminal-helical domain followed by an antiparallel five-stranded sheet (Schumacher et al. 2002). Hfq monomers form a doughnut-shaped homohexameric ring structure that contains at least two separate RNA-binding sites: one located on the proximal side that binds AU-rich tracts on sRNAs as well as mRNA, and the other located on the distal side that binds poly(A) on mRNAs (Folichon et al. 2003).

Several lines of evidence have shown that Hfq is a sRNAs-binding protein that recognizes various RNA species, in particular sRNAs (Brescia et al. 2003; Folichon et al. 2003; Sittka et al. 2008). The sRNAs-binding capability of Hfq has been further confirmed and characterized when heterologous Hfq was used (Sittka et al. 2009). Intriguingly, however, Hfq was also shown to be dispensable in some organisms (Bohn et al. 2007). It is believed that those bacteria utilize an RNA chaperone either related to or distinct from Hfq, or use a different mechanism to facilitate sRNAs–mRNA interactions (Chao and Vogel 2010). Finally, the sRNAs-binding properties of Hfq have been exploited in RNA immunoprecipitation to enrich for sRNAs for high-throughput sequencing for the identification and analysis of the bacterial sRNAs repertoire (Vogel and Papenfort 2006).

Mycobacteria are gram-positive, acid-fast, and GC-rich (62%–70%) organisms (Edson 1951; Ramakrishnan et al. 1972; Ratledge and Stanford 1982). They are aerobic, rod shaped, and characterized by a unique cell wall coated with mycolic acids (Cook et al. 2009; Kaur et al. 2009). The genus *Mycobacterium* is comprised of organisms that are true pathogens, opportunistic pathogens, or saprophytes. *Mycobacterium tuberculosis* (*M. tuberculosis*, MTB) is one of the true pathogens that cause significant morbidity and mortality worldwide.

The success of MTB as a human pathogen can be attributed to its extraordinary stealth and capacity to adapt to environmental changes throughout the course of infection (Cook et al. 2009). sRNAs-mediated regulation is one of the emerging research areas in the study of gene regulation in mycobacteria. However, no Hfq orthologs have been found in the *Mycobacterium* species (Sun et al. 2002). Recently, numerous sRNAs were identified by either direct analysis of low-molecular-weight RNAs isolated from mycobacteria (cloning based) (Arnvig and Young 2009), bioinformatics prediction based on consensus transcription start and termination sites (computation-based), or sRNAs homology analysis in a sequenced transcriptome (high-throughput sequencing) (DiChiara et al. 2010; Arnvig et al. 2011). However, the limited number of novel sRNAs identified by cloning-based methods, and the inaccuracy of computational predictions due to the lack of sRNAs sequence homology among species hinder understanding of sRNAs networking and sRNAs-mediated gene regulation in mycobacteria. Here, we used a novel method to isolate

sRNAs in mycobacteria in which no Hfq ortholog has been reported so far. We take advantage of the conserved inherent sRNAs-binding capability of heterologous Hfq from other species to enrich for sRNAs, and identified novel sRNAs by high-throughput sequencing. In order to test our hypothesis, we transformed the FLAG-tagged heterologous Hfq from *Escherichia coli* (*E. coli*) into *Mycobacterium smegmatis* (*M. smegmatis*), and performed RNA immunoprecipitation, followed by RNA sequencing. This report presents 24 novel sRNAs identified with this approach. Our results illustrate the usefulness of heterologous Hfq in genome-wide screening of sRNAs in *M. tuberculosis* as well as in other Hfq-negative bacteria.

RESULTS

Expression and immunoprecipitation of heterologous Hfq in *M. smegmatis*

The sRNAs-binding capability of Hfq has been utilized to capture sRNAs in bacteria. Since Hfq has not been identified in mycobacteria, we performed RNA immunoprecipitation using heterologous and well-characterized Hfq from *E. coli*. We cloned the *E. coli* *hfq* into a mycobacterial expression vector (Fan et al. 2009) and expressed the heterologous Hfq with the FLAG-tag in *M. smegmatis* (Fig. 1A). The empty vector was used as a control. We then performed RNA immunoprecipitation using anti-FLAG agarose to capture the Hfq-bound RNA.

Deep sequencing of the *E. coli* Hfq-bound sRNAs in *M. smegmatis*

In order to increase the number of reads from sRNAs and reduce the masking effect of ribosomal RNA (rRNA) during sequencing, rRNA depletion was performed using MICROBExpress Bacterial mRNA Enrichment Kit from Ambion, resulting in the removal of ~80% of the rRNA from the extracted RNA. The enriched sRNAs was then reverse-transcribed and a library was constructed for deep sequencing. Sequencing was performed using the Illumina

FIGURE 1. Deep sequencing reveals sRNAs in *M. smegmatis*. (A) The FLAG epitope-tagged Hfq protein was expressed in *M. smegmatis* in parallel with the vector control. (B) Pie diagrams showing the relative proportions of the different RNA species.

high-throughput sequencing protocol for sRNAs. A total of 13 million reads constituting 168 million base pairs (168 Mbp), ranging from 17 to 32 bp (after base-quality trimming) were generated. Reads were mapped to the *M. smegmatis* genome using BWA and allowing only one mismatch (Li and Durbin 2010). RNAs were classified based on their functions: intergenic, mRNA, rRNA, and tRNA as shown in Figure 1B. When compared with the control coIP, enrichment of RNAs in intergenic regions (23% vs. 6%) and mRNA (39% vs. 34%) was observed in the Hfq coIP sample, suggesting that heterologous *E. coli* Hfq functions as a specific RNA-binding protein in mycobacteria.

sRNAs in intergenic and antisense regions of mycobacterial genome

We analyzed the Hfq-bound sRNAs in both intergenic and coding regions based on orientation and strand specificity. The *trans*-encoded intergenic sRNAs were named with prefix IGR whereas the *cis*-encoded antisense sRNAs were given the prefix AS. As shown in Figure 2A, half of the Hfq-bound sRNAs are located at intergenic regions, whereas one-fourth of the sRNAs are antisense to the coding regions. These antisense sRNAs were further categorized according to KEGG classification (Fig. 2B; Kanehisa et al.

2010). Most of the antisense sRNAs target the coding regions of genes in metabolic pathways, suggesting the regulatory roles of sRNAs in mycobacteria. Two previously reported sRNAs in *M. tuberculosis*, C8 and B11, were also found in our sequencing (1) (Fig. 2C). This indicates that our methodology of using heterologous Hfq to immunoprecipitate and enrich for sRNAs is a feasible approach to discovering sRNAs in bacterial species without known endogenous Hfq.

Validation of sRNAs in *M. smegmatis* using Northern blot analysis

To rule out the possibility of the nonphysiological effects of expressing heterologous Hfq on sRNAs as well as to verify the sequencing results, sRNAs were detected in the total RNA sample of wild-type *M. smegmatis* strain using Northern blot analysis. As shown in Figure 3A, 24 sRNAs including 12 *trans*-encoded IGR-sRNAs and 12 *cis*-encoded AS-sRNAs, were confirmed by Northern blot analysis. All of the sRNAs showed differential expression at exponential (mid-log, M) and stationary phases (S), and their sizes ranged from 28 to 400 nt. Multiple forms of some sRNAs (IGR-6, IGR-7, AS-1, AS-5, AS-6, AS-8, AS-10, and AS-11) were present at both exponential and stationary phase. In

contrast, some forms of the sRNAs were detected at either exponential phase or stationary phase, such as IGR-5 and IGR-12. Although we cannot rule out the possibility that the multiple bands observed on the Northern blots are due to the sRNAs degradation, these results suggest that sRNAs biogenesis is highly dynamic and sRNAs changes throughout different stages of mycobacterial growth.

Identification of transcription start sites (TSS) by 5' RACE

We performed 5' RACE to determine the transcription start sites (TSS) for the newly identified sRNAs to further confirm the sequencing results. The coordinates of these sRNAs are shown in Supplemental Tables S1 and S2 (Supplemental Material). TSS determined by 5' RACE are mostly consistent with the 5' distal end found by the high-throughput sequencing of the sRNAs. The distance of IGR sRNAs from the upstream genes was calculated based on the TSS found by 5' RACE or by the mapped results. However, we cannot rule out the possibility that some of the IGR

FIGURE 2. Statistics of the Hfq-bound sRNAs. (A) Pie diagram showing the relative proportions of different classes of sRNAs mapped to *M. smegmatis* genome. (B) KEGG classification of target genes of antisense sRNAs. (C) Visualization of high-throughput sequencing data with Integrated Genome Viewer of known sRNAs (C8 and B11) in *Mycobacterium tuberculosis*.

FIGURE 3. sRNAs of *M. smegmatis* validated on Northern blots. (A) Expression of sRNAs was detected at exponential mid-log (M) and stationary (S) phases. The intergenic sRNAs were represented as IGR, whereas the antisense sRNAs as AS. The bands of the sRNAs were indicated with arrows, while the approximate size of sRNAs detected by Northern blot was stated at the top. (B) Genomic positions of sRNAs in *M. smegmatis* were indicated. The approximate size of sRNAs determined by Northern blot was shown. The distance between sRNAs and the neighboring genes was calculated based on the 5' RACE results and mapping results visualized in IGV. (#) The sRNAs with a size in consistency with the mapping results in IGV; (*) the distance determined by 5' RACE.

sRNAs are part of the UTRs of neighboring genes. The locations and orientations of all validated sRNAs are shown in Figure 3B. Five of the AS-sRNAs are located on the com-

plementary strands of transposase genes (AS-4, AS-8, AS-9, AS-10, and AS-11), indicating that these sRNAs regulate the expression of transposases. One of these AS-sRNAs targets on

an unknown protein. All of the remaining six AS-sRNAs were complementary to genes encoding enzymes; an amidohydrolase (AS-1), transferase (AS-2), sulfatase (AS-5), helicase (AS-6), synthetase (AS-7), and ligase (AS-12).

Prediction of secondary structures and putative sigma factor binding sites

We further predicted the secondary structures of the sRNAs using RNAfold (Hofacker 2003) and CentroidFold (Sato et al. 2009). As illustrated in Figure 4 and Supplemental Figure S1, most of the sRNAs contained a typical I-shaped terminator or a stem-loop for mycobacteria RNA without a long poly(U) stretch (Gardner et al. 2011). Notably, IGR-2 contained a short poly(A) sequence at the 3' end (Fig. 4). Based on the TSS information, potential sigma factor binding sites were determined by searching for the consensus sites in the upstream 20 nt. Due to the lack of prediction programs for finding the consensus binding sites for *M. smegmatis*, the known consensus sites were identified manually (Gerhart et al. 2008). A putative *SigA* consensus was found for IGR-2, IGR-3, AS-5, and AS-11 (Supplemental Fig. S2). A putative *SigB* consensus was found for nine sRNAs (IGR-1, IGR-4, IGR-9, IGR-11, IGR-12, AS-3, AS-6, AS-7, and AS-9). A putative *SigF* consensus was only found for IGR-4.

FIGURE 4. 5' RACE, mapping spectrum and secondary structure of *M. smegmatis* sRNAs. PCR results of 5' RACE was shown. The band found in TAP and adapter-treated RNA was cloned and sequenced to determine the TSS. The mapping spectrum in IGV was included. Secondary structure was predicted using the TSS found in 5' RACE and mapped results in IGV. Representatives of intergenic sRNAs IGR-2 and IGR-3 (A,B) and antisense sRNAs AS-5 and AS-7 (C,D) were shown.

Phylogenetic conservation of sRNAs among mycobacteria

Phylogenetic conservation of the detected sRNAs was determined in nucleotide BLAST on NCBI (Fig. 5A). Of the genes encoding the 12 detected IGR-sRNAs and 12 AS-sRNAs, 13 were exclusively present in *M. smegmatis* (IGR-2, IGR-3, IGR-4, IGR-6, IGR-9, AS-1, AS-2, AS-3, AS-6, AS-7, AS-9, AS-11, and AS-12). Although *M. smegmatis* shares over 2000 homologous genes with the more virulent *M. tuberculosis*, only two sRNAs (IGR-1, AS-5) detected in *M. smegmatis* were found in *M. tuberculosis* after excluding the previously reported C8 and B11. The novel sRNAs seemed to be even more broadly conserved among *M. smegmatis* and *M. gilvum* PYR-GCK, *M. sp. JLS*, *M. sp. KMS*, *M. sp. MCS*, and *M. vanbaalenii* PYR-1. Intriguingly, *M. smegmatis* is more closely related to these species, usually regarded as rapid growers and mostly non-pathogenic as compared with *M. tuberculosis* (Fig. 5B).

Regulatory functions of sRNAs in mycobacteria—down-regulation of p450 by IGR-4

In order to further understand the roles of sRNAs in gene regulation, we expressed one of the sRNAs candidates, IGR-4, in *M. smegmatis*. The predicted targets of IGR-4 sRNAs, MSMEG_6548 Riese iron-sulfur protein (ISP) and MSMEG_4823 cytochrome p450 (*p450*), were found using IntaRNA (Busch et al. 2008). A modified pMFAJ vector that only transcribes the sRNAs was utilized for this study. pMFAJ-IGR4, which expresses IGR-4 sRNAs, was transformed into *M. smegmatis* mc² 155, and the sRNAs expression level was detected by Northern blot. The empty pMFAJ vector of pMFAJ was used as a control. Expression of IGR4 did not result in a distinctive change to the growth morphology or growth rate of cells (data not shown). However, IGR-4 dramatically decreased the mRNA level of p450 and ISP at both the exponential mid-log (M) and stationary (S) phases (Fig. 6).

DISCUSSION

Using heterologous Hfq to isolate sRNAs in bacteria that lack Hfq

Instead of using traditional gel purification of small RNAs, we used Hfq to enrich for sRNAs. Unlike size-independent selection of sRNAs, a common method of sRNAs enrichment, immu-

FIGURE 6. Overexpression of IGR-4 in *M. smegmatis* down-regulated MSMEG_6548 and MSMEG_4823. The relative mRNA expression levels of two of its predicted targets: MSMEG_6548 Rieske iron-sulfur protein (ISP) and MSMEG_4823 cytochrome p450 (*p450*) ISP and *p450* were determined by real-time PCR. ISP and *p450* showed a decreased expression level upon overexpression of IGR-4 at both mid-exponential (M) and stationary (S) phases when compared with the vector control.

noprecipitation of sRNAs by Hfq usually results in a better recovery. In addition, this method may reduce the chance of sequencing degraded mRNAs. Hfq is a sRNAs-binding protein and plays a significant role in bacterial gene regulation. RNA immunoprecipitation with endogenous Hfq has been utilized to enrich the pool of sRNAs for the discovery of novel bacterial sRNAs. The sRNAs bound by Hfq are more likely to be functional. However, no Hfq ortholog has been found in some bacteria such as mycobacteria. Therefore, we developed a novel method that takes advantage of the sRNAs-binding capabilities of heterologous Hfq to enrich for sRNAs in mycobacteria and then identified the sRNAs using high-throughput sequencing. We have shown that *E. coli* Hfq possesses sRNAs-binding capabilities when overexpressed in *M. smegmatis*. Twelve novel trans-encoded IGR-sRNAs and 12 novel cis-encoded AS-sRNAs were found in *M. smegmatis*. The identified sRNAs include two previously found sRNAs, C8 and B11 (Arnvig and Young 2009) in MTB, which serves as a positive control for the precision and reliability of our methodology. This approach to discover sRNAs could be applicable to other Hfq-negative bacterial strains. In this project, we found two additional sRNAs conserved between *M. smegmatis* and MTB (IGR-1 and AS-5); further analysis of these sRNAs may enhance our understanding of sRNAs-mediated gene regulation in this lethal pathogen.

Putative sigma factor-binding consensus were found based on TSS

Based on the identified TSS and sigma factor-binding consensus sequences, we have noted the putative -10 boxes

and -35 boxes for some of the novel sRNAs. Most of the sRNAs genes are preceded by either *sigA* or *sigB* consensus sequences, while the IGR-4 protein contained a *sigF* consensus sequence. The *sigF* sequence was thought to be absent in *M. smegmatis* until Chowdhury et al. (2007) showed that it was involved in the regulation of *dps* gene, which is overexpressed under stress. This suggested that IGR-4 could be a stress-related sRNAs in *M. smegmatis*.

Phylogenetic analysis reveals the conservation of sRNAs in fast-growing mycobacteria

The roles of sRNAs in pathogenicity have recently been described in a few gram-positive bacteria, yet little is known for mycobacteria (Toledo-Arana et al. 2007). We evaluated the phylogenetic conservation of the newly identified sRNAs and observed that sRNAs found in *M. smegmatis* are also conserved in other rapid growers such as *M. gilvum*, *M. sp. JLS*, *M. sp. KMS*, and *M. sp. MCS*. They were less conserved in the slow growing MTB and were absent in *M. leprae*. This suggests that sRNAs are involved in growth rate determination and genome evolution.

Involvement of sRNAs in mycobacterial growth, metabolism, and gene transposition

All of the IGR-sRNAs showed differential expression at exponential and stationary phase, suggesting that sRNAs are involved in the regulation of growth and adaptation of *M. smegmatis* to changes in nutrient availability. We observed multiple bands for some sRNAs, which are likely to be the precursor and mature sRNAs, although we cannot rule out the possibility of sRNAs degradation. This suggests the existence of highly regulated sRNAs biogenesis during bacterial growth and in response to environmental changes (Vogel and Papenfort 2006; Repoila and Darfeuille 2009; Waters and Storz 2009; Papenfort and Vogel 2010; Romby and Charpentier 2010). For example, IGR-5 was detected as a band of 130 nt at both exponential and stationary phases. An extra band of 70 nt was only detected at the stationary phase, suggesting further processing of the sRNAs into a shorter sRNAs in response to the nutrient-depleted environment. Based on the presence of putative binding sites, expression of the sRNAs is likely to be regulated by different sigma factors under a variety of stress conditions including nutrient deprivation at stationary phase (Sachdeva et al. 2010).

We noticed that a number of AS-sRNAs target transposase genes that might restrict transposon movement. This suggests the potential involvement of sRNAs in inactivating transposition in bacteria (Landt et al. 2008). The role of sRNAs regulation of transposition in *E. coli* was illustrated in an elegant study by Ross et al. (2010). Hfq can be an indirect negative regulator of DsrA-mediated transposition. By disrupting Hfq, which facilitates

the base pairing of the DsrA sRNAs with the mRNA encoding the transposition-regulating H-NS protein, the transposition frequency of IS10 was greatly increased. In this study, we showed the presence of AS-sRNAs, which may target transposases more efficiently with better complementarity. Thus, by regulating the expression of these sRNAs, the frequency of transposition can be controlled. It will be interesting to determine how the sRNAs regulate expression of bacterial genes, in particular, genes related to virulence and antibiotic resistance. All other AS-sRNAs, except those related to transposases, target metabolic enzymes, suggesting extensive and highly regulated sRNAs-mediated control of enzyme expression in mycobacteria.

To investigate the functional roles of the novel mycobacterial sRNAs, we examined the transcription level of some of the predicted targets of one of the identified sRNAs. Our results suggested that sRNAs IGR-4 of *M. smegmatis* down-regulates some of the genes involved in metabolic pathways, such as cytochrome p450 and Rieske iron-sulfur protein. The roles of sRNAs in growth and metabolism control have been extensively studied in *E. coli*, as exemplified by SsrS (6S) ncRNA, which is involved in limiting the number of genes expressed at the stationary phase when the dNTP levels are low. At the outgrowth from stationary phase, 6S can be used to transcribe a 14–20 nt long pRNA, which releases the 6S RNA from inhibiting RNA polymerase (Wassarman and Saecker 2006). sRNAs are also involved in regulation of metabolism such as glucose utilization. Also in *E. coli*, SgrS (previously known as RyaA) negatively regulates the translation of PtsG, which generates toxic glucose-6-phosphate from glucose during sugar utilization. SgrS pairs with PtsG mRNA in an Hfq-dependent and endoribonuclease RNase E-dependent manner, resulting in the translation inhibition of PtsG levels, and therefore reduced glucose phosphate stress (Kimata et al. 2001; Vanderpool and Gottesman 2004). The roles of sRNAs in gram-positive bacteria have been widely studied in the *Bacillus* species, *Staphylococcus* species, and *Listeria monocytogenes* but are less well studied in mycobacteria (Romby and Charpentier 2010). Therefore, it is worthwhile to further investigate their roles in *M. smegmatis*.

Concluding remarks

This work presents the whole-genome screening of *M. smegmatis* sRNAs by a novel method that utilizes heterologous Hfq to enrich for a pool of sRNAs for high-throughput sequencing. The identification of large numbers of sRNAs in *M. smegmatis* indicates potential important roles of these small molecules in the physiological and genetic network. The discovery of these new sRNAs in *M. smegmatis* and other mycobacteria will bring new insights into the control of biological functions in these bacteria. Our findings will serve as an important reference for the future analysis of sRNAs-mediated regulation in

mycobacteria and will contribute to the development of sRNAs prediction programs. This approach to identify sRNAs in bacteria using heterologous Hfq for sRNAs enrichment will be of general applicability for species lacking endogenous Hfq.

MATERIALS AND METHODS

Bacterial strains and culture

M. smegmatis of strain MC² 155 were used in this study. They were grown in *Mycobacterium* Middlebrook 7H9 medium (BBL) supplemented with 0.5% glycerol, 10% oleic acid-albumin-dextrose-catalase (OADC) (BBL), in a conical flask. The cultures were grown at 37°C with shaking of 100 rpm under subdued light. Exponential phase cultures were harvested at optical density 600 nm (OD₆₀₀) 0.8–1.5, whereas stationary phase cultures were harvested at OD₆₀₀ = 2.0. *M. smegmatis* with pFVV16 vector or pMFA41J vector were grown in similar conditions with the addition of 20 µg/mL of kanamycin. Colonies of *M. smegmatis* were grown on Middlebrook 7H11 medium agar, supplemented with 20 µg/mL of kanamycin where necessary. *E. coli* DH5α was cultured in lysogeny broth in a conical flask. The cultures were grown at 37°C with shaking at 250 rpm.

Plasmids

The RNA chaperone gene *hfq* of *E. coli* DH5α was amplified from genomic DNA. The amplified Hfq was cloned into pFVV16 vector that was modified from pVTV16 by adding a FLAG epitope sequence at 5' terminal of the multiclonal site (MCS). Vector pMFA41 was provided by G.P. Zhao, Department of Microbiology, Fudan University, Shanghai, China (Fan et al. 2009). It was further modified to remove the 5'- and 3'-untranslated sequences such that transcription would start at +1 nucleotide of the cloned sRNAs and end at the transcription terminator at the 3' end sRNAs gene. The modified pMFA41 was named as pMFA41J. sRNAs IGR-4 was amplified and cloned into pMFA41J. The sRNAs construct was then transformed into *M. smegmatis* using electroporation at 2.5 kV, 1000Ω, 25 µF. Transformed cells were grown and harvested at the exponential and stationary phases.

Coimmunoprecipitation with Hfq and ribosomal RNA depletion

M. smegmatis was transformed with pFVV16 and pFVV16-hfq and grown in 7H9 medium to stationary phase. Coimmunoprecipitation (Co-IP) was carried out to isolate the FLAG-bound and the FLAG-hfq-bound RNA. Cells were washed twice with 1× PBS and lysed in 1× PBS with 0.5% Triton X-100. The lysate was centrifuged at 12,000g at 4°C for 15 min. The clear lysate was incubated with 50 µL of anti-FLAG antibody agarose affinity gel (Sigma-Aldrich), followed by incubation at 4°C for 2 h on a rotator. The anti-FLAG agarose was then washed in washing buffer (10 mM Tris/pH 7.4, 100 mM NaCl, 0.01% NP-40) three times. The immunoprecipitated RNA-protein complex was digested with 80 µg of proteinase K at 42°C for 30 min. RNA was purified using acidic phenol:chloroform (5:1, pH 4.5, Ambion), followed by isopropanol precipitation. The ribosomal RNA in

the immunoprecipitated RNA was depleted using MICROBExpress Bacterial mRNA Enrichment Kit (Ambion) according to manufacturer's instructions. A total of 10 µg of extracted total RNA was separated on a 6% polyacrylamide gel containing 8 M urea. The RNA fraction of size smaller than 500 bp was cut out and purified.

Next-generation sequencing and data processing

The purified RNA was sequenced using the Illumina/Solexa small RNA-sequencing protocol on an Illumina GAII platform. Base-trimming was performed to remove nucleotides of low quality at the distal ends of the reads. Processed reads were mapped onto *M. smegmatis* mc² 155 genome (GenBank accession: CP000480.1) using BWA (v0.5.7) with default parameters (Li and Durbin 2010). The mapped sequencing reads were then visualized in Integrated Genome Viewer (IGV) 1.5.18 (Robinson et al. 2011).

Total RNA preparation

M. smegmatis were grown to exponential and stationary phases in a 100-mL culture. Cells were collected from a 50-mL culture. The cell pellets were lysed in 500 µL of 20 mg/mL lysozyme (Sigma-Aldrich) at 37°C for 30 min. The cells suspension was further lysed in 10 mL TRIzol reagent (Invitrogen). The total RNA was then precipitated by isopropanol. The isolated RNA was subjected to a second-round purification using acidic phenol-chloroform (5:1, pH 4.5, Ambion). The purified RNA was reconstituted with 50 µL of nuclease-free water. The RNA quantity and quality was analyzed on a NanoDrop ND-1000 spectrophotometer (Thermo) and TBE agarose gel, respectively.

Northern blot analysis

Northern blot was used to verify the presence and size of sRNAs identified in RNA sequencing. *M. smegmatis* RNA from exponential phase and stationary phase were denatured at 70°C for 5 min in Gel Loading Buffer II (Ambion) and then separated on 6% polyacrylamide gels containing 8 M urea. For each lane, 10 µg of total RNA was used. The separated RNAs were electroblotted onto a Hybond-XL membrane (Amersham). The RNAs were then cross-linked under an ultraviolet (UV) light of 120 mJ/cm² for 2 min. DNA oligonucleotide probes specific for each candidate sRNAs were radioactively labeled using 20 pmoles of oligonucleotide in 20 µL of kinase reaction containing 20 µCi [γ -P³²]ATP and 20 units of T4 polynucleotide kinase (New England Biolabs) at 37°C for 1 h. The labeling reaction was then purified to remove the free [γ -P³²]ATP using Centri Spin column-20 (Princeton Separations) according to the manufacturer's instructions. The membranes were prehybridized with UltraHyb buffer (Ambion) at 42°C for 1 h, and then hybridized with the specific probes at 42°C for 16 h. The membranes were then washed with 1×SSC, 0.1% SDS extensively. Hybridization signals were exposed to a phosphor screen overnight and visualized with a PhosphorImager (Typhoon TRIO, Amersham Biosciences). Alternatively, the signals were determined by autoradiography on hyperfilms (Amersham).

sRNAs sequence analysis

The candidate sRNAs were computationally analyzed to predict their transcription start sites, secondary structures, regulatory

targets, and phylogenetic conservation among the bacteria genus *Mycobacterium*. The transcription start sites of the sRNAs were predicted by searching for consensus binding sites for sigma factor A, B, or F using information provided by Sachdeva et al. (2010). The sequences of detected sRNAs were proposed according to the predicted transcription start sites and the size detected on Northern blots. The secondary structure of sRNAs and the termination stem-loop were predicted using both RNAfold with default settings on the Vienna RNA secondary structure server (<http://rna.tbi.univie.ac.at>) (Hofacker 2003) using CentroidFold (Sato et al. 2009). The targets of the sRNAs were predicted using IntaRNA v1.2.4 (Busch et al. 2008). The phylogenetic conservation of the sRNAs was determined using NCBI nucleotide BLAST research with the default settings. The database nucleotide collection (nt/nr) limited to *Mycobacterium* (taxid: 1763) was used. Results with E-values of <1 were taken as positive. E-values of <1 × 10⁻⁴ and 1 × 10⁻⁸ were further classified as moderately and highly conserved. The phylogenetic tree was constructed with the neighbor-joining method (Saitou and Nei 1987) in Molecular Evolutionary Genetics Analysis (MEGA) version 5 (Tamura et al. 2011). The species distance was calculated using bacterial 16S RNA sequences obtained from Ribosomal Database Project release 10 (Cole et al. 2007, 2009).

Quantitative real-time PCR

Total RNA extracted from *M. smegmatis* was reverse-transcribed using QuantiTect Reverse Transcription Kit (Qiagen). Quantitative real-time PCR was performed on ABI Fast 7500 real-time PCR machine (Applied Biosystems, ABI). All real-time PCR assays were run in a total reaction of 10 µL consisting of 1× SYBR Green Master Mix (ABI) and 200 nM primers. Cycling conditions were 2 min at 50°C, then 10 min at 95°C, followed by 40 cycles of amplification (15 sec at 95°C, 1 min at 60°C). The primers used in real-time PCR were as follows: ISP- forward 5'-TCTCGTCGA AGTGCACACAC-3' and reverse 5'-GGTTCCGTCGAGGTTGA ACT-3'; p450- forward 5'-GACTACCGGACGTCCATCTC-3' and reverse 5'-TGATCGCTCAGGTGTGAATG-3'. The results were normalized to the amount of Ribonuclease P (Rnp A) RNA measured using rnpA forward 5'-AACGTTTCGTATCCGGTC TTG-3' and reverse 5'-ACCCAACGTGCGTTCCAAAC-3'.

Rapid amplification of cDNA ends (RACE)

5' RACE was carried out using FirstChoice RML-RACE kit (Ambion) with a modified protocol. Total RNA of *M. smegmatis* was treated with Tobacco acid pyrophosphatase (TAP). 5' RACE RNA adapter was ligated to the TAP-treated RNA, followed by reverse transcription using random primers. Controls without TAP treatment and/or without adding adapter were included. PCR was performed using a forward primer specific to the 5' RACE adapter and reverse primers specific to target sRNAs. The amplified PCR fragment was cloned into pGEM-T easy system (Promega) and then sequenced. TSS of sRNAs was found by sequencing the junction between the adapter and the sRNAs.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank G.P. Zhao (Department of Microbiology, Fudan University, Shanghai, China) for contributing the pMFA41 expression vector, and W.C. Yam (Department of Microbiology, The University of Hong Kong, Hong Kong SAR, China) for his advice. This work was supported by CityU startup grant (7200262) and Strategic Research Grant (7002753) funded to T.C.L. by the City University of Hong Kong, and a General Research Fund (GRF461809) from the Research Grant Committee, Hong Kong SAR Government granted to T.F.C.

Received April 29, 2012; accepted October 22, 2012.

REFERENCES

- Aiba H. 2007. Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr Opin Microbiol* **10**: 134–139.
- Arлуison V, Hohng S, Roy R, Pellegrini O, Regnier P, Ha T. 2007. Spectroscopic observation of RNA chaperone activities of Hfq in post-transcriptional regulation by a small non-coding RNA. *Nucleic Acids Res* **35**: 999–1006.
- Arnvig KB, Young DB. 2009. Identification of small RNAs in *Mycobacterium tuberculosis*. *Mol Microbiol* **73**: 397–408.
- Arnvig KB, Comas I, Thomson NR, Houghton J, Boshoff HI, Croucher NJ, Rose G, Perkins TT, Parkhill J, Dougan G, et al. 2011. Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of *Mycobacterium tuberculosis*. *PLoS Pathog* **7**: e1002342. doi: 10.1371/journal.pat.1002342.
- Bohn C, Rigoulay C, Bouloc P. 2007. No detectable effect of RNA-binding protein Hfq absence in *Staphylococcus aureus*. *BMC Microbiol* **7**: 10. doi: 10.1186/1471-2180-7-10.
- Brescia CC, Mikulecky PJ, Feig AL, Sledjeski DD. 2003. Identification of the Hfq-binding site on DsrA RNA: Hfq binds without altering DsrA secondary structure. *RNA* **9**: 33–43.
- Busch A, Richter AS, Backofen R. 2008. IntaRNA: Efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* **24**: 2849–2856.
- Chao Y, Vogel J. 2010. The role of Hfq in bacterial pathogens. *Curr Opin Microbiol* **13**: 24–33.
- Chowdhury RP, Gupta S, Chatterji D. 2007. Identification and characterization of the dps promoter of *Mycobacterium smegmatis*: Promoter recognition by stress-specific extracytoplasmic function sigma factors σ^H and σ^F . *J Bacteriol* **189**: 8973–8981.
- Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM, Tiedje JM. 2007. The ribosomal database project (RDP-II): Introducing myRDP space and quality controlled public data. *Nucleic Acids Res* **35**: D169–D172.
- Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, et al. 2009. The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**: D141–D145.
- Cook GM, Berney M, Gebhard S, Heinemann M, Cox RA, Danilchanka O, Niederweis M. 2009. Physiology of mycobacteria. *Adv Microb Physiol* **55**: 81–182, 318–319.
- DiChiara JM, Contreras-Martinez LM, Livny J, Smith D, McDonough KA, Belfort M. 2010. Multiple small RNAs identified in *Mycobacterium bovis* BCG are also expressed in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. *Nucleic Acids Res* **38**: 4067–4078.
- Edson NL. 1951. The intermediary metabolism of the mycobacteria. *Bacteriol Rev* **15**: 147–182.
- Fan XY, Ma H, Guo J, Li ZM, Cheng ZH, Guo SQ, Zhao GP. 2009. A novel differential expression system for gene modulation in Mycobacteria. *Plasmid* **61**: 39–46.
- Folichon M, Arлуison V, Pellegrini O, Huntzinger E, Regnier P, Hajsndorf E. 2003. The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. *Nucleic Acids Res* **31**: 7302–7310.
- Franze de Fernandez MT, Eoyang L, August JT. 1968. Factor fraction required for the synthesis of bacteriophage Q β -RNA. *Nature* **219**: 588–590.
- Gardner PP, Barquist L, Bateman A, Nawrocki EP, Weinberg Z. 2011. RNIE: Genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Res* **39**: 5845–5852.
- Gerhart E, Wagner H, Vogel J. 2008. Approaches to identify novel non-messenger RNAs in bacteria and to investigate their biological functions: Functional analysis of identified non-mRNAs. In *Handbook of RNA biochemistry* (ed. RK Hartmann et al.). Wiley-VCH Verlag GmbH, Weinheim, Germany. doi: 10.1002/9783527619504.ch37.
- Gottesman S, McCullen CA, Guillier M, Vanderpool CK, Majdalani N, Benhammou J, Thompson KM, FitzGerald PC, Sowa NA, FitzGerald DJ. 2006. Small RNA regulators and the bacterial response to stress. *Cold Spring Harb Symp Quant Biol* **71**: 1–11.
- Hofacker IL. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res* **31**: 3429–3431.
- Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* **38**: D355–D360.
- Kaur D, Guerin ME, Skovierova H, Brennan PJ, Jackson M. 2009. Chapter 2: Biogenesis of the cell wall and other glycoconjugates of *Mycobacterium tuberculosis*. *Adv Appl Microbiol* **69**: 23–78.
- Kimata K, Tanaka Y, Inada T, Aiba H. 2001. Expression of the glucose transporter gene, ptsG, is regulated at the mRNA degradation step in response to glycolytic flux in *Escherichia coli*. *EMBO J* **20**: 3587–3595.
- Landt SG, Abeliuk E, McGrath PT, Lesley JA, McAdams HH, Shapiro L. 2008. Small non-coding RNAs in *Caulobacter crescentus*. *Mol Microbiol* **68**: 600–614.
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**: 589–595.
- Morita T, Maki K, Aiba H. 2005. RNase E-based ribonucleoprotein complexes: Mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev* **19**: 2176–2186.
- Nielsen JS, Lei LK, Ebersbach T, Olsen AS, Klitgaard JK, Valentin-Hansen P, Kallipolitis BH. 2009. Defining a role for Hfq in Gram-positive bacteria: Evidence for Hfq-dependent antisense regulation in *Listeria monocytogenes*. *Nucleic Acids Res* **38**: 907–919.
- Papenfert K, Vogel J. 2010. Regulatory RNA in bacterial pathogens. *Cell Host Microbe* **8**: 116–127.
- Pfeiffer V, Papenfert K, Lucchini S, Hinton JC, Vogel J. 2009. Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat Struct Mol Biol* **16**: 840–846.
- Ramakrishnan T, Murthy PS, Gopinathan KP. 1972. Intermediary metabolism of mycobacteria. *Bacteriol Rev* **36**: 65–108.
- Ratledge C, Stanford J, ed. 1982. *The biology of the mycobacteria*. Academic Press, Ltd., London, UK.
- Repoila F, Darfeuille F. 2009. Small regulatory non-coding RNAs in bacteria: Physiology and mechanistic aspects. *Biol Cell* **101**: 117–131.
- Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26.
- Romby P, Charpentier E. 2010. An overview of RNAs with regulatory functions in gram-positive bacteria. *Cell Mol Life Sci* **67**: 217–237.
- Ross JA, Wardle SJ, Haniford DB. 2010. Tn10/IS10 transposition is downregulated at the level of transposase expression by the RNA-binding protein Hfq. *Mol Microbiol* **78**: 607–621.
- Sachdeva P, Misra R, Tyagi AK, Singh Y. 2010. The sigma factors of *Mycobacterium tuberculosis*: Regulation of the regulators. *FEBS J* **277**: 605–626.

- Saitou N, Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.
- Sato K, Hamada M, Asai K, Mituyama T. 2009. CENTROIDFOLD: A web server for RNA secondary structure prediction. *Nucleic Acids Res* **37**: (Web Server issue) W277–W280.
- Schumacher MA, Pearson RF, Moller T, Valentin-Hansen P, Brennan RG. 2002. Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: A bacterial Sm-like protein. *EMBO J* **21**: 3546–3556.
- Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JC, Vogel J. 2008. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* **4**: e1000163. doi: 10.1371/journal.pgen.1000163.
- Sittka A, Sharma CM, Rolle K, Vogel J. 2009. Deep sequencing of Salmonella RNA associated with heterologous Hfq proteins in vivo reveals small RNAs as a major target class and identifies RNA processing phenotypes. *RNA Biol* **6**: 266–275.
- Sun X, Zhulin I, Wartell RM. 2002. Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res* **30**: 3662–3671.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2685–2686.
- Toledo-Arana A, Repoila F, Cossart P. 2007. Small noncoding RNAs controlling pathogenesis. *Curr Opin Microbiol* **10**: 182–188.
- Valentin-Hansen P, Eriksen M, Udesen C. 2004. The bacterial Sm-like protein Hfq: A key player in RNA transactions. *Mol Microbiol* **51**: 1525–1533.
- Vanderpool CK, Gottesman S. 2004. Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol Microbiol* **54**: 1076–1089.
- Vogel J, Papenfort K. 2006. Small non-coding RNAs and the bacterial outer membrane. *Curr Opin Microbiol* **9**: 605–611.
- Vogel J, Argaman L, Wagner EG, Altuvia S. 2004. The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide. *Curr Biol* **14**: 2271–2276.
- Wassarman KM, Saecker RM. 2006. Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science* **314**: 1601–1603.
- Waters LS, Storz G. 2009. Regulatory RNAs in bacteria. *Cell* **136**: 615–628.

RNA

A PUBLICATION OF THE RNA SOCIETY

Identification of small RNAs in *Mycobacterium smegmatis* using heterologous Hfq

Sai-Kam Li, Patrick Kwok-Shing Ng, Hao Qin, et al.

RNA 2013 19: 74-84 originally published online November 20, 2012

Access the most recent version at doi:[10.1261/rna.034116.112](https://doi.org/10.1261/rna.034116.112)

Supplemental Material

<http://rnajournal.cshlp.org/content/suppl/2012/11/08/rna.034116.112.DC1>

References

This article cites 49 articles, 9 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/19/1/74.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to RNA go to:

<http://rnajournal.cshlp.org/subscriptions>