

## BIROn - Birkbeck Institutional Research Online

Evangelopoulos, Dimitrios and Gupta, Antima and Lack, N. and Maitra, Arundhati and ten Bokum, A.M.C. and Kendall, Sharon L. and Sim, E. and Bhakta, Sanjib (2014) Characterisation of a putative AraC transcriptional regulator from Mycobacterium smegmatis. Tuberculosis 94 (6), pp. 664-671. ISSN 1472-9792.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/10911/>

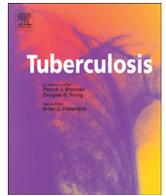
*Usage Guidelines:*

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html> or alternatively contact [lib-eprints@bbk.ac.uk](mailto:lib-eprints@bbk.ac.uk).



Contents lists available at ScienceDirect

## Tuberculosis

journal homepage: <http://intl.elsevierhealth.com/journals/tube>

## NON-TUBERCULOUS MYCOBACTERIA: GENERAL

Characterisation of a putative AraC transcriptional regulator from *Mycobacterium smegmatis*Dimitrios Evangelopoulos<sup>a, b, 1</sup>, Antima Gupta<sup>a, 2</sup>, Nathan A. Lack<sup>b, 3</sup>, Arundhati Maitra<sup>a</sup>, Annemieke M.C. ten Bokum<sup>c, 4</sup>, Sharon Kendall<sup>c</sup>, Edith Sim<sup>b, 5</sup>, Sanjib Bhakta<sup>a, \*</sup><sup>a</sup> *Mycobacteria Research Laboratory, Institute of Structural and Molecular Biology, Department of Biological Sciences, Birkbeck, University of London, Malet Street, London WC1E 7HX, UK*<sup>b</sup> *Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, UK*<sup>c</sup> *Department of Pathology and Infectious Diseases, The Royal Veterinary College, Royal College Street, London NW1 0TU, UK*

## S U M M A R Y

## Keywords:

AraC  
Transcriptional regulator  
Mycobacteria  
Mycobacterial two-hybrid system  
*nat* operon  
Protein–protein interaction

*MSMEG\_0307* is annotated as a transcriptional regulator belonging to the AraC protein family and is located adjacent to the arylamine *N*-acetyltransferase (*nat*) gene in *Mycobacterium smegmatis*, in a gene cluster, conserved in most environmental mycobacterial species. In order to elucidate the function of the AraC protein from the *nat* operon in *M. smegmatis*, two conserved palindromic DNA motifs were identified using bioinformatics and tested for protein binding using electrophoretic mobility shift assays with a recombinant form of the AraC protein. We identified the formation of a DNA:AraC protein complex with one of the motifs as well as the presence of this motif in 20 loci across the whole genome of *M. smegmatis*, supporting the existence of an AraC controlled regulon. To characterise the effects of AraC in the regulation of the *nat* operon genes, as well as to gain further insight into its function, we generated a  $\Delta$ *araC* mutant strain where the *araC* gene was replaced by a hygromycin resistance marker. The level of expression of the *nat* and *MSMEG\_0308* genes was down-regulated in the  $\Delta$ *araC* strain when compared to the wild type strain indicating an activator effect of the AraC protein on the expression of the *nat* operon genes.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

## 1. Introduction

Transcriptional factors modulate gene expression through binding to a specific DNA sequence usually found upstream of the gene or the genomic area that they control. They are important proteins that can help cells acclimatise to challenging environments based on the changing external stimuli. The AraC/XylS protein family of transcriptional regulators, present in bacterial species

is involved in a variety of cellular processes from carbon metabolism to stress responses and the regulation of virulence [1]. Common characteristics of the AraC proteins is the presence of a conserved region of 100 residues in the C-terminal region of the protein that form a helix-turn-helix structure responsible for DNA binding, a second region in the N-terminal region of the protein contains a ligand binding domain and a peptide-linker region connecting the two functional domains. The proteins that belong to the AraC/XylS family usually recognise palindromic DNA sequences and bind to them by forming dimers using the helix-turn-helix domain [2].

*Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB) can survive within macrophages as well as in the extreme environment found in granulomas during infection in the human body. For this reason, the genome of *M. tuberculosis* contains an exceptionally large number of transcriptional factors, including 13 sigma factors, 5 anti-sigma factors and 7 anti-anti-sigma factors [3] which assist its adaptation to different environments and stresses. Six of these 190 transcription factors belong to the AraC/XylS family (Rv1317, Rv1395, Rv1931c, Rv3082c, Rv3736 and Rv3833). Most of the *M. tuberculosis* AraC proteins characterised,

\* Corresponding author. Tel.: +44 (0)20 7631 6355 (office), +44 (0)20 7079 0799 (lab); fax: +44 (0)20 7631 6246.

E-mail addresses: [s.bhakta@bbk.ac.uk](mailto:s.bhakta@bbk.ac.uk), [sanjib.bhakta@ucl.ac.uk](mailto:sanjib.bhakta@ucl.ac.uk) (S. Bhakta).

<sup>1</sup> Present address: Centre for Clinical Microbiology, University College London, Royal Free Campus, Rowland Hill Street, London NW3 2PF, UK.

<sup>2</sup> Present address: Centre for Systems and Synthetic Biology (CSSB), Heinz Wolf Building, Brunel University, Uxbridge, London UB8 3PH, UK.

<sup>3</sup> Present address: School of Medicine, Koç University, Sariyer, Istanbul 34450, Turkey.

<sup>4</sup> Present address: Therapeutic Immunology Group, Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK.

<sup>5</sup> Present address: Faculty of Science, Engineering and Computing Kingston University, Penrhyn Road, Kingston upon Thames, Surrey KT1 2EE, UK.

<http://dx.doi.org/10.1016/j.tube.2014.08.007>

1472-9792/© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

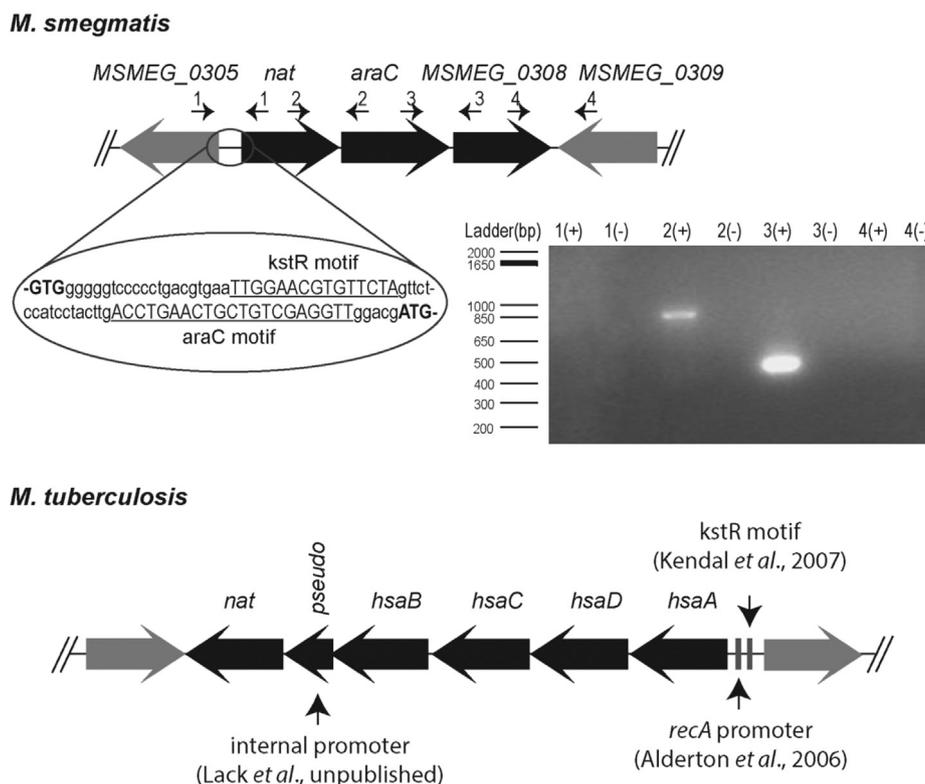
such as the AraC proteins encoded by Rv1395, Rv1931c and Rv3082c genes are linked with virulence as their genetic alteration generates an attenuated phenotype either *in vitro* in macrophage infection model or *in vivo* in mice [4–8]. Until now there has been no functional information regarding the role of the Rv3736 and Rv3833 encoded AraC proteins in *M. tuberculosis*. Furthermore, *Mycobacterium smegmatis*, the saprophytic environmental species of the genus *Mycobacterium* and a common laboratory surrogate for molecular genetic studies of *M. tuberculosis*, contains 16 different AraC proteins encoded in its genome, indicating the need of this organism to adapt to multiple niches. One of these AraC proteins is encoded by the MSMEG\_0307, gene which is located between the arylamine *N*-acetyltransferase (*nat*, MSMEG\_0306) gene and a novel oxidoreductase (MSMEG\_0308) that is believed to be involved in riboflavin biosynthesis [9].

The *nat* operon in *M. tuberculosis* has been validated as a likely therapeutic target due to its important endogenous roles in *M. tuberculosis*, related to cholesterol degradation, cell wall biogenesis, intracellular growth and altered drug susceptibility [10–12]. In *M. tuberculosis*, the *hsaA* (Rv3570c), *hsaB* (Rv3567c), *hsaC* (Rv3568c) and *hsaD* (Rv3569c) genes are co-transcribed with the *nat* (Rv3566c) gene and their corresponding proteins have been shown to be directly involved in the cholesterol metabolism pathway [13]. The NAT protein utilises acyl coenzyme A (CoA) catabolites, including acetyl CoA and *n*-propionyl CoA. These intermediates play a central role in metabolic support of cell wall biosynthesis [14,15]. Cholesterol is considered to be a vital energy source for *M. tuberculosis* cells growing within

macrophages [16]. In addition, this *nat* gene cluster in *M. tuberculosis* is under the control of the *kstR* transcription factor (Rv3574) that controls a regulon of genes involved in lipid and cholesterol metabolism [17].

In contrast, the gene organisation around *nat* differs in *M. smegmatis* and the *hsaA–D* genes (MSMEG\_6035–MSMEG\_6038) are clustered together in a region about 57 kb downstream from the *nat* locus. Nevertheless, *kstR* regulatory DNA sequences are present between the *nat* and the MSMEG\_0305 genes (Figure 1) and upstream of the *hsaA–D* gene cluster in the genome of *M. smegmatis* indicating possible co-regulation. However, it has been shown previously that the *nat* gene cluster in *M. smegmatis* is not directly controlled by the *kstR* transcriptional regulator but rather the MSMEG\_0305 is under the effect of the KstR regulator [17,18].

In the view of the presence of a transcription factor that belongs to the AraC protein family in the *nat* gene cluster in *M. smegmatis* it is important to assess whether it influences the expression of *nat* and the adjacent gene for the novel reductase (MSMEG\_0308). The question of whether the AraC transcriptional factor has a role in the regulation of genes that are involved in lipid and cholesterol metabolism in *M. smegmatis* also needs to be addressed. We report here, the characterisation of the AraC-family transcriptional regulator MSMEG\_0307 from *M. smegmatis* using biochemical assays, the characterisation of its regulon and DNA binding sites, as well as its influence on the regulation of gene expression of the *nat* gene cluster in *M. smegmatis*. In addition, we have characterised a protein–protein interaction network that is formed by the *nat* operon gene products.



**Figure 1. The *nat* operon in *M. smegmatis*.** A comparison between the *nat* operons in *M. smegmatis* and *M. tuberculosis*. In the *nat* operon of *M. smegmatis* the intergenic region between *nat* and MSMEG\_0305 genes is shown in loop with the location of the *kstR* and *araC* motifs highlighted in the sequence. In the *M. tuberculosis* *nat* operon, the identified promoter and *kstR* sequences are shown with the arrows. RT-PCR amplification of the overlapping regions of the *nat* operon genes using specific primers, shown as numbered arrows on top of the genes. Lanes with (+) on the agarose gel correspond to cDNA whereas lanes with (-) corresponds to negative control, i.e. cDNA that produced without the addition of reverse transcriptase (RT) in order to distinguish possible genomic DNA contamination. Primer set 1: MSMEG\_0305-*nat*, primer set 2: *nat*-MSMEG\_0307, primer set 3: MSMEG\_0307-MSMEG\_0308 and primer set 4: MSMEG\_0308-MSMEG\_0309.

## 2. Materials and methods

### 2.1. Bacterial strains, growth conditions and plasmids

The bacterial strains and plasmids used in this study are listed in Table S4. The oligonucleotide primers are listed in the Supplementary Table S1. *Escherichia coli* strains JM109 and BL21(DE3) pLysS were grown in Luria–Bertani (LB) broth with rotation at 200 rpm or in LB agar at 37 °C, unless specified otherwise. *M. smegmatis* mc<sup>2</sup>155 [19] and the modified strains were grown in Middlebrook 7H9 broth supplemented with 0.2% (v/v) glycerol, 0.05% (w/v) Tween-80 and 10% ADC (Albumin-Dextrose-Catalase, BD) with rotation at 180 rpm or in Middlebrook 7H10 agar supplemented with 0.5% (v/v) of glycerol and 10% OADC (Oleic acid-Albumin-Dextrose-Catalase, BD) at 37 °C, unless specified otherwise. Comparative growth curves of *M. smegmatis* mc<sup>2</sup>155 and the  $\Delta$ nat,  $\Delta$ MSMEG\_0307 and  $\Delta$ MSMEG\_0308 strains were performed using mycobacterial minimal medium (0.5 g l<sup>-1</sup> of L-asparagine, 1 g l<sup>-1</sup> of KH<sub>2</sub>PO<sub>4</sub>, 2.5 g l<sup>-1</sup> of Na<sub>2</sub>HPO<sub>4</sub>, 50 mg l<sup>-1</sup> of ferric ammonium citrate, 0.5 g l<sup>-1</sup> of MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.5 mg l<sup>-1</sup> of CaCl<sub>2</sub> and 0.1 mg l<sup>-1</sup> of ZnSO<sub>4</sub>) supplemented with a carbon source (glycerol, glucose, melibiose or cholesterol) at 0.2% (v/v). Antibiotics were used at the following concentrations (µg ml<sup>-1</sup>): kanamycin (30), chloramphenicol (34), hygromycin B (100 for *E. coli* and 50 for *M. smegmatis*) and trimethoprim (12). All chemicals were purchased from Sigma–Aldrich (Poole, UK) unless otherwise stated. High-fidelity Phusion DNA polymerase (NEB) was employed in all cloning related PCR reactions whereas Taq DNA polymerase (NEB) was used in all other PCR reactions in this study. Restriction digestion enzymes were purchased from New England Biolabs (Hitchin, UK). All constructs were confirmed by sequencing (Gene Service, UCL).

### 2.2. Identification promoter sites and DNA motifs

All mycobacterial genome sequences that were used in this study were obtained from the NCBI (<http://www.ncbi.nlm.nih.gov>). BLAST analysis was performed using the NCBI BLAST algorithm [20]. Multiple sequence alignments were done using ClustalW algorithm at the EBI server [21]. For viewing, annotating and comparing mycobacterial genomes, the java-based software packages Artemis and ACT, from the Sanger Institute, were used [22,23]. The sequence viewer, PromView (<http://www.comlab.ox.ac.uk/~activities/compbio/bioinformatics-/software/index-.htm#PromView>) was used for the search of consensus promoter sequences obtained from previously published mycobacterial promoters [24]. The MEME algorithm [25] was used to discover conserved palindromic DNA motifs among mycobacterial species and MAST [26] was used to identify the presence of these motifs in the genome of *M. smegmatis* mc<sup>2</sup>155.

### 2.3. Cloning, overexpression and purification of MS0307

The gene encoding the AraC protein (MSMEG\_0307) from *M. smegmatis* mc<sup>2</sup>155 was PCR amplified from genomic DNA and cloned into pET28b (+) vector (Novagen). The N-terminus of the AraC-family transcriptional regulator MSMEG\_0307 was co-transcribed with a thrombin cleavage site followed by a hexahistidine tag. The recombinant AraC-family transcriptional regulator MSMEG\_0307 was produced in BL21(DE3)pLysS cells at 18 °C following induction with 0.5 mM IPTG overnight. Cells were lysed by sonication on wet ice (5 cycles of 45sec on, 45sec off) and the AraC-family transcriptional regulator MSMEG\_0307 was purified using nickel affinity chromatography (Invitrogen). The His-tagged MSMEG\_0307 protein was then further purified on a HiLoad

16/60 Superdex™ 75 pg (Pharmacia) preparative gel filtration column, equilibrated with 20 mM Tris–HCl pH 8, 100 mM NaCl. Fractions contained pure (>99%) His-tagged MSMEG\_0307 protein was pooled and concentrated using an Amicon Ultra concentrator (Millipore) at 5 mg ml<sup>-1</sup> and stored in 50% (v/v) glycerol in –80 °C for further use.

### 2.4. Electrophoretic mobility shift assays (EMSAs)

The DNA fragments (~300bp) containing the binding sequence motifs were amplified using PCR and primers (Table S1) and further purified. The reactions had a final volume of 10 µl and contained 100 ng of DNA, 1× EMSA buffer (20 mM Tris.HCl pH8, 75 mM NaCl, 10 mM MgCl<sub>2</sub>) and increasing concentrations of recombinant His-tagged MSMEG\_0307 protein (0.01 µg to 1 µg). The reactions were incubated at room temperature for 30 min and then were loaded onto a 5% (v/v) native polyacrylamide gel. Following electrophoresis the gels were stained with ethidium bromide and the bands were visualised using BioDoc-It™ imaging system (UVP, Cambridge, UK).

### 2.5. Generation of $\Delta$ araC in *M. smegmatis* and complementation studies

The deletion of the MSMEG\_0307 gene (MSMEG\_0307) from *M. smegmatis* mc<sup>2</sup>155 was performed using the method of specialised transduction as described previously [27]. Briefly, the left (898 bp) and right (782 bp) arms of the MSMEG\_0307 gene were PCR amplified using the primers given in table S1 and cloned into the suicide delivery vector p0004S to create the allelic-exchange plasmid p0004S-MSMEG\_0307. The p0004S-MSMEG\_0307 was then *PacI* digested and packed into the temperature sensitive mycobacteriophage phAE159 to generate the allelic-exchange phage ph $\Delta$ araC. Wild-type *M. smegmatis* mc<sup>2</sup>155 was transduced using high-titre ph $\Delta$ MSMEG\_0307 phages as described [27]. Following the specialised transduction, hygromycin resistant colonies were screened by PCR using a gene internal and external primer set (Table S1) and  $\Delta$ araC mutants were further confirmed by DNA sequencing. Complementation of the  $\Delta$ MSMEG\_0307 as well as overexpression of the WT *M. smegmatis* mc<sup>2</sup>155 were performed using the pMV261 plasmid [28] with the native MSMEG\_0307 expressed under hsp60 mycobacterial promoter located in the plasmid.

### 2.6. Drug susceptibility assay

The susceptibility of mycobacterial strains against various antibiotics was determined using the resazurin redox indicator assay as described previously [29]. Briefly, wild-type *M. smegmatis* mc<sup>2</sup>155 and the  $\Delta$ MSMEG\_0307 mutant were grown until mid-exponential phase (1 OD<sub>600</sub>) and then 100 µl of diluted cells (10<sup>4</sup> CFUs) were added into a 96 well plate that contained 100 µl of two-fold dilutions of antibiotics at various concentrations in µg/mL [isoniazid (INH 50 to 0.09), pyrazinamide (PZA 50 to 0.09), rifampicin (RMP 50 to 0.09), ethambutol (EMB 150 to 0.29), streptomycin (STM 50 to 0.09), kanamycin (KAN 50 to 0.09), ampicillin (AMP 150 to 0.29) and chloramphenicol (CLP 50 to 0.09)]. The plates were incubated at 37 °C for 2 days. Following 24 h of incubation, 50 µl of 0.01% (w/v) sterile resazurin solution in presence of 1% (v/v) Tween-80 was added to all wells of the plate and left overnight at 37 °C. The minimum inhibitory concentrations (MICs) were defined as the lowest antibiotic concentration of the well where bacterial cells were not able to grow and thus did not reduce the resazurin dye.

### 2.7. RNA extraction and cDNA synthesis

Wild-type *M. smegmatis* mc<sup>2</sup>155,  $\Delta$ MSMEG\_0307, wild-type *M. smegmatis* mc<sup>2</sup>155 with either the empty vector (pMV261) or overexpressing MSMEG\_0307 (pMVaraC) and the  $\Delta$ MSMEG\_0307 complemented mutant cells were grown in Middlebrook 7H9 broth until OD<sub>600</sub> was 0.8 and total RNA was extracted from using the GTC method as previously described [17] and cDNA was synthesised using SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. A control was set up to assess genomic DNA contamination by replacing the Superscript III Reverse Transcriptase with water.

### 2.8. Operon analysis

The boundaries of the *nat* operon were identified by PCR amplification of the intergenic regions of the operon and adjacent genes using *M. smegmatis* mc<sup>2</sup>155 cDNA as a template and specific primers (Table S1) for each region. Positive (gDNA from *M. smegmatis* mc<sup>2</sup>155) and negative controls (cDNA made without Reverse Transcriptase) were also used. The amplicons were then analysed using agarose gel electrophoresis.

### 2.9. RT-qPCR

Real-time quantitative polymerase chain reactions (RT-qPCR) were performed using the DyNAmo SYBR Green qPCR kit (NEB) on the MJ Research Bio-Rad Real Time PCR Opticon Engine 2 System (GRI). *M. smegmatis* mc<sup>2</sup>155 gDNA was used for the generation of a standard curve and *sigA* gene (MSMEG\_2758; a mycobacterial sigma factor) was used as a reference gene for the relative quantification method. Briefly, a 20  $\mu$ l reaction was set up on ice containing 1  $\times$  DNA Master SYBR Green I mix, 1  $\mu$ l of cDNA and 0.3  $\mu$ M of each primer (Table S1). The PCR reactions were initially heated to 95 °C for 10 min before 35 cycles of 95 °C for 30 s, 62 °C for 20 s, and 72 °C for 20 s were performed. Fluorescence was measured at the end of each cycle following a heating step to 80 °C to ensure the denaturation of any primer-dimers. At the end of the PCR, melting curve analysis was performed to verify the product specificity. The experiment was performed in duplicate and each gene was measured in triplicates (three biological replicates, two experimental replicates) giving a total of six data points per gene. Fold changes were calculated using the 2<sup>- $\Delta\Delta$ Ct</sup> statistical method [30].

## 3. Results

### 3.1. Defining the *nat* operon in *M. smegmatis*

In order to define the presence and extent of a *nat* operon in *M. smegmatis*, RT-PCR analysis was performed on the basis that any amplicons obtained using intergenic primers would indicate that the two genes are co-transcribed together and thus belong to the same operon (Figure 1). Using this rationale, we obtained amplicons for the intergenic regions between *nat*-MSMEG\_0307 (MSMEG\_0306-MSMEG\_0307) and MSMEG\_0307-MSMEG\_0308 (MSMEG\_0307-MSMEG\_0308) but not between the MSMEG\_0305-*nat* and MSMEG\_0308-MSMEG\_0309, indicating that in *M. smegmatis* the *nat* operon consists of three genes (Figure 1). Using bioinformatic analyses, we search for a putative promoter sequence using as input known mycobacterial promoters; however we were not able to identify a conserved promoter sequence upstream of *nat* gene at the start of the operon. A comparative analysis on different mycobacterial genomes indicated that a similar gene organisation was seen in the *nat* operon among fast-growing

mycobacteria as opposed to the *nat* and *hsaA-D* clusters found in slow-growing mycobacteria (Figure S1).

Thorough sequence analysis using the MEME algorithm [25] was carried out on the *M. smegmatis* *nat* operon as well as a 1.5 kb DNA fragment upstream of the *nat* gene in order to identify the presence of regulatory DNA sequences that could be affected by the MSMEG\_0307 gene product. Comparative genome analyses were performed on the closely related fast-growing mycobacterial species that have similar *nat* gene clusters and an MSMEG\_0307 orthologue (*M. smegmatis*, *Mycobacterium gilvum*, *Mycobacterium vanbaalenii* and *Mycobacterium* sp. MCS/KMS/JLS). This approach identified two different regulatory motifs. The motif designated from now on as Motif 1 is located between -25 and -5 upstream of the *nat* gene start codon (Figure 2A) and another motif designated from now on as Motif 2 is located between -57 and -34 upstream of the MSMEG\_0307 gene (Figure 2A). Motif 1 consists of a 20 bp palindromic DNA sequence (ACCTCGACAGCAGTTCAGGT) (Figure 2A) and the Motif 2 consists of a 23 bp DNA sequence (GTCAGGACATGACTTTTCTTGCT) (Figure 2A). In order to identify the presence of these two motifs elsewhere in the genome of *M. smegmatis* as well as possible sites of action of the MSMEG\_0307 protein, the MAST algorithm was employed to search a database of intergenic regions of the *M. smegmatis* genome [26]. Motif 1 was found to be present in twenty additional instances in the *M. smegmatis* genome (Table S2) and the Motif 2 in five additional instances (Table S3). This indicates the existence of a regulon controlled by the AraC transcription regulator MSMEG\_0307.

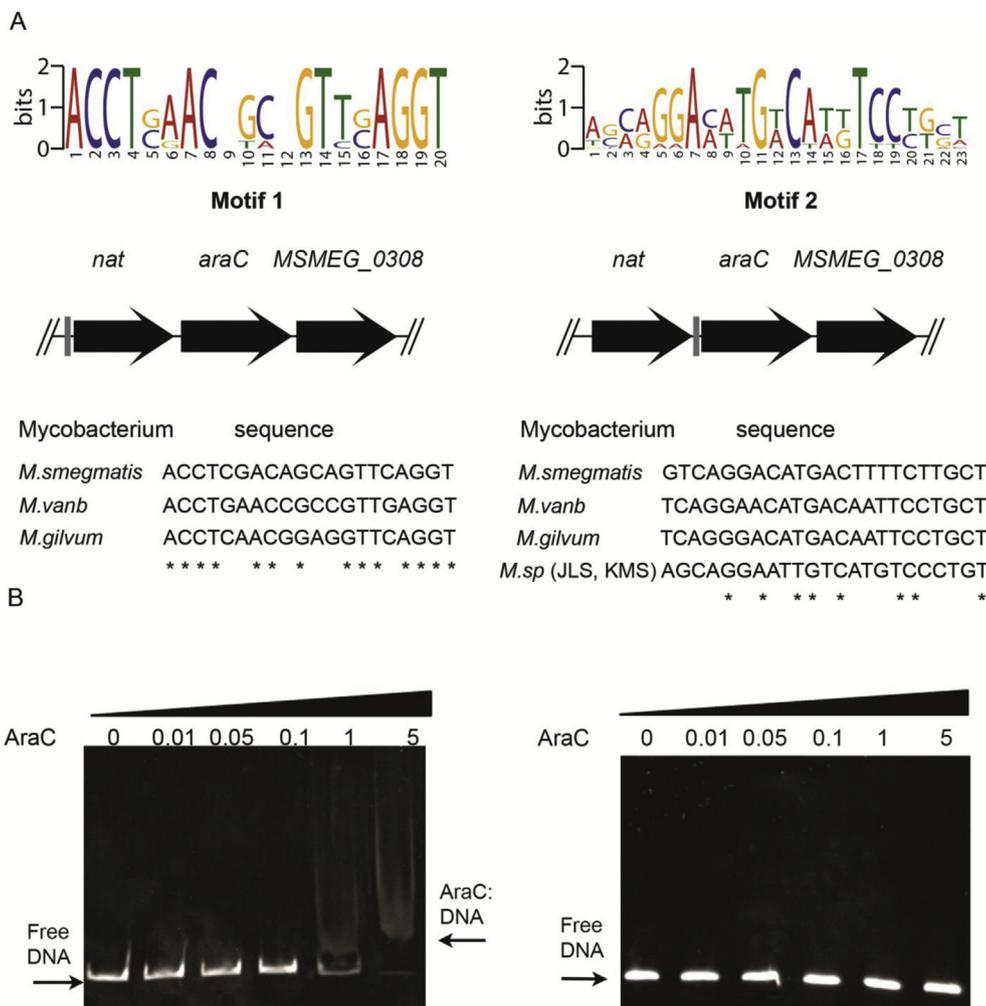
### 3.2. Binding of AraC<sub>Msmeg</sub> to the identified DNA motifs

In order to determine whether the MSMEG\_0307 protein binds to any of the two identified motifs, the MSMEG\_0307 protein from *M. smegmatis* mc<sup>2</sup>155 was cloned, over-expressed and purified as a recombinant protein (Figure S2). DNA fragments containing Motif 1 or 2 were amplified using PCR and used for Electrophoretic Mobility Shift Assays (EMSA). As shown in Figure 2B the presence of recombinant MSMEG\_0307 was able to produce a band shift indicating the binding of the protein to Motif 1. Interestingly, in a parallel experiment no DNA band shift was observed when Motif 2 was used as a substrate for the MSMEG\_0307 (Figure 2B) indicating that the MSMEG\_0307 protein binds specifically to the Motif 1.

### 3.3. Characterisation of a $\Delta$ *araC* strain

In order to investigate the role of the MSMEG\_0307 gene product of the *nat* operon in *M. smegmatis* the gene (MSMEG\_0307) was deleted from the genome using specialised transduction [27]. Resistant colonies were screened for the presence of the deletion by PCR using one set of primers which amplified either an internal region of the gene or a region flanking the deleted MSMEG\_0307 gene. As expected, the  $\Delta$ MSMEG\_0307 strain did not show any PCR amplification using the internal set and had an amplicon with a size difference using the external primer set indicating the presence of the hygromycin resistance cassette in the genome in the place of the MSMEG\_0307 gene (Figure S3). A similar methodology was applied for the generation of the MSMEG\_0308 gene deletion mutant ( $\Delta$ MSMEG\_0308) as well.

The ability to generate a  $\Delta$ MSMEG\_0307 strain indicates that the MSMEG\_0307 gene is not essential for *in vitro* growth of *M. smegmatis*. Comparative growth curves in enriched media between wild type and the  $\Delta$ MSMEG\_0307 strain indicated that the loss of the MSMEG\_0307 gene did not significantly affect the growth of the mutant strain (Figure 3A). Subsequently the ability of the  $\Delta$ MSMEG\_0307 strain to grow in minimal media with glycerol,



**Figure 2. Investigation of DNA regulatory motifs found in the *nat* operon in *M. smegmatis*.** (A) The palindromic DNA sequences are shown on the graphs. The big letters illustrate the consensus nucleotides in the motifs. The grey boxes in the operon genes below the motifs show the positions of the motifs in the relationship to the operon. The nucleotide sequences of each of the motifs in different mycobacterial genomes are shown with the consensus nucleotides highlighted with an asterisk (\*). (B) EMSA assays with MSMEG\_0307 recombinant protein and the identified motifs. DNA containing the Motif 1 or the Motif 2 were incubated with increasing concentration of MSMEG\_0307 protein [0, 0.01, 0.05, 0.1, 1 and 5  $\mu\text{g}/\text{mL}$ ] and then were loaded onto an 8% polyacrylamide gel.

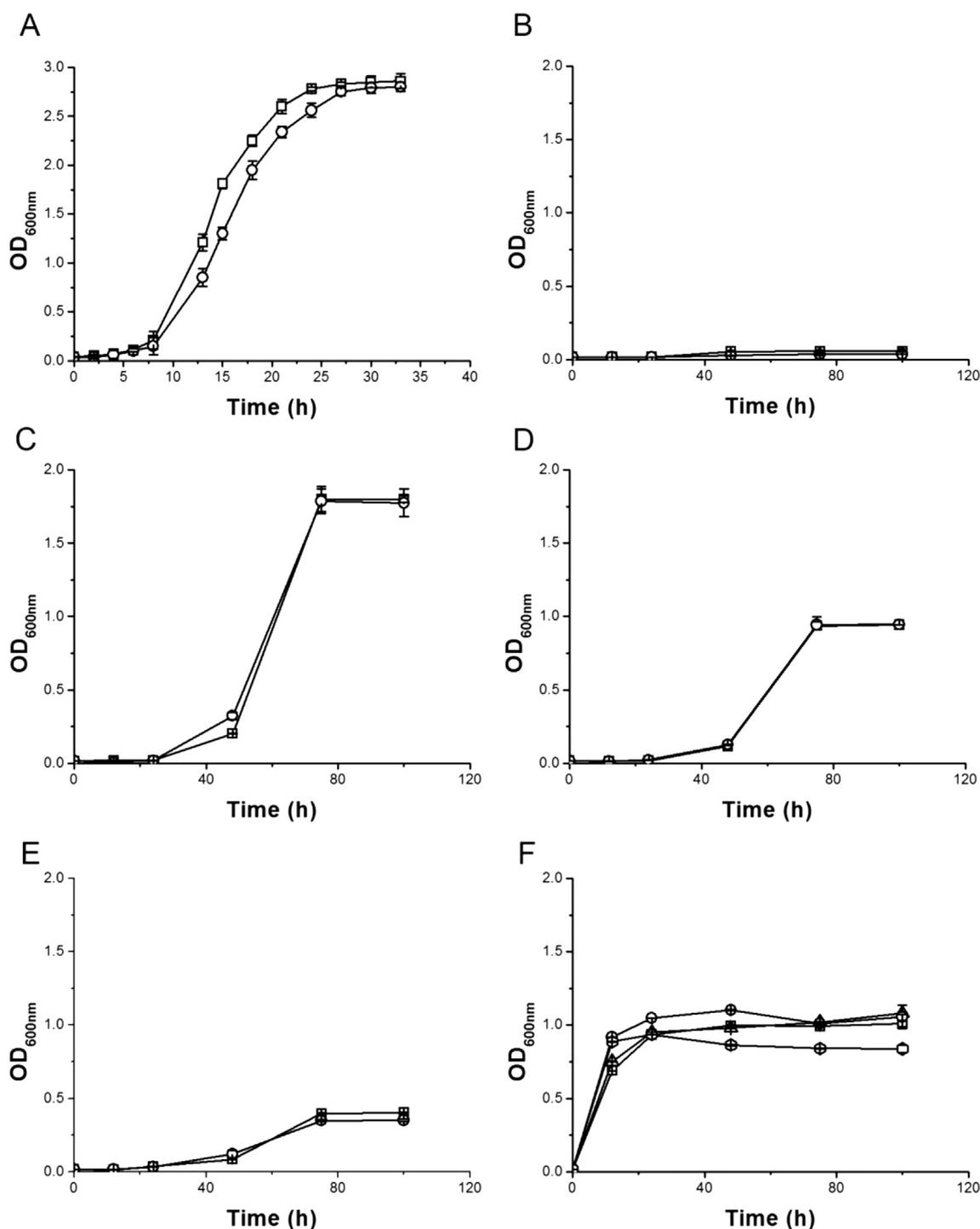
glucose, melibiose [31] was tested as several bacterial AraC transcription factors are known to control genes that are responsible for the degradation of complex carbon sources, such as arabinose metabolism in *E. coli* [32]. It was found that the deletion of the MSMEG\_0307 transcription regulator had little effect on the growth on these carbon sources (Figure 3B–E).

As the genes of the *nat* operon in *M. tuberculosis* are involved in cholesterol metabolism [13,33], the ability of the  $\Delta\text{MSMEG}_0307$  as well as  $\Delta\text{nat}$  [34] and  $\Delta\text{MSMEG}_0308$  strains to grow in presence of cholesterol as the sole carbon source was also examined. The results showed that the proteins encoded by *nat* operon in *M. smegmatis* are not essential for degradation of cholesterol (Figure 3F) in contrast to those present in *M. tuberculosis*.

Furthermore, in order to investigate the possibility that the MSMEG\_0307 regulator is involved in the development of multi-drug resistance as other studied AraC proteins are known to be [35], we tested the susceptibility of the  $\Delta\text{MSMEG}_0307$  strain to a variety of antibiotics including all the first line anti-tuberculosis drugs (isoniazid, ethambutol, rifampicin and pyrazinamide). The susceptibility pattern of the two strains, WT and  $\Delta\text{MSMEG}_0307$ , was similar apart from negligible differences observed in the MIC values on rifampicin, kanamycin and chloramphenicol (Figure S4).

### 3.4. MSMEG\_0307 protein regulates the expression of the *nat* operon in *M. smegmatis*

As the recombinant MSMEG\_0307, a putative transcription regulator, was able to bind specifically to the Motif 1 upstream of the *nat* operon in *M. smegmatis*, the effect of the MSMEG\_0307 gene deletion on the expression of the other genes of the operon was studied using RT-qPCR. The comparison of the relative expression levels of *nat* (MSMEG\_0306) and MSMEG\_0308 genes between the WT and the  $\Delta\text{MSMEG}_0307$  mutant revealed that the deletion of the MSMEG\_0307 gene has a significant effect on the expression of the other operon genes (Figure 4) suggesting that the AraC protein controls the gene expression of the operon. Both *nat* and MSMEG\_0308 genes were down-regulated in the  $\Delta\text{MSMEG}_0307$  mutant by 18 and 13 fold respectively indicating that the AraC transcription factor acts as an activator for the expression of *nat* and MSMEG\_0308 genes. On the contrary, overexpression of the MSMEG\_0307 gene in the wild-type *M. smegmatis* from the pMVaraC plasmid increased the expression of both *nat* and MSMEG\_0308 by 23 and 16 fold respectively (Figure 4) confirming the activator effect of the MSMEG\_0307 protein on the operon. The overexpression of the MSMEG\_0307 gene using the pMVaraC



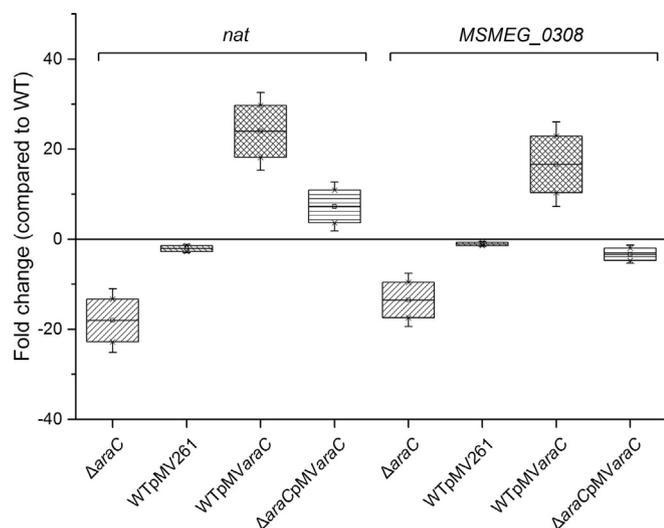
**Figure 3.** Growth curves of WT and  $\Delta$ MSMEG\_0307 *M. smegmatis* in enriched and minimal media. The cultures were inoculated with the same amount of bacteria ( $\sim 10^7$ ) and the growth was detected by measuring the OD<sub>600nm</sub> over time. The volume of each culture was 100 mL. Growth curves of WT (□),  $\Delta$ nat (Δ),  $\Delta$ MSMEG\_0307 (○) and  $\Delta$ MSMEG\_0308 (◇) *M. smegmatis* in (A) Middlebrook 7H9 medium, (B) minimal medium with no added carbon source (C) minimal medium with glycerol (D) minimal medium with glucose (E) minimal medium with melibiose and (F) minimal medium supplemented with cholesterol.

construct in *M. smegmatis* was also confirmed by qPCR (data not shown). Complementation of the  $\Delta$ MSMEG\_0307 strain with the same plasmid (pMVaraC) was able to compensate for the down regulation of the *nat* gene resulting in a 7 fold up regulation; however, this was not the case for the *MSMEG\_0308* gene that remained down regulated 3 fold. The presence of the internal cassette in place of *MSMEG\_0307* gene might partially explain the

reason why *MSMEG\_0308* gene expression remained down regulated in the complemented  $\Delta$ MSMEG\_0307 strain.

#### 4. Discussion

The regulation of gene expression is essential for all living organisms to adapt to various environmental and physiological



**Figure 4. Differential expression of *nat* and *MSMEG\_0308* genes in *M. smegmatis* strains.** The fold changes of the  $\Delta$ *MSMEG\_0307*, WT containing the empty pMV261 vector (WTpMV261), WT overexpressing *MSMEG\_0307* (WTpMVaraC) and the complementation strain  $\Delta$ *MSMEG\_0307* containing the pMaraC vector (WTpMVaraC) compared to WT. Relative quantification of the gene expression levels were applied using the  $\Delta\Delta$ Ct statistical method and the expression of *sigA* gene from *M. smegmatis* was used as reference gene. The range of the box plots represent the SE with the mean outlined as the small box inside each box plot. Confidence intervals at 1% and 99% are mentioned by the x symbols at each box.

conditions and stresses. In this study, we describe the role of a transcriptional regulator from the AraC/XylS family in *M. smegmatis*, a fast growing, environmental species of *Mycobacterium*. We have identified that the AraC-family transcriptional regulator MSMEG\_0307 protein recognises and binds to a palindromic motif upstream of the *nat* operon in *M. smegmatis* and that it is also involved in the regulation of the expression of the *nat* and the *MSMEG\_0308* genes of the same operon. It is also very likely that it auto-regulates its own expression as other AraC proteins are known to do [36,37] and thus controls the gene expression of the *nat* operon in *M. smegmatis*. In addition, the palindromic DNA sequence that the MSMEG\_0307 binds to, has been found in 20 loci across the genome and this might represent an early view of the regulon that is under the control of the MSMEG\_0307 protein.

The exact roles of these 20 genes that might be under the control of the MSMEG\_0307 regulator have not been fully assigned; however, they can be divided into five main categories based on their function excluding the 6 hypothetical proteins (Table S2). There are four genes encoding for ATP-binding cassette (ABC) transporters and two transmembrane proteins with unknown functions. ABC transporters that might be under the control of AraC constitute universal transporter systems that are responsible for the transfer of a variety of substrates across the cell membrane [38]. In addition, one of the mechanisms of antibiotic resistance in bacteria is due to specific ABC transporters. The fluoroquinolone efflux pump encoded by the Rv2686c-Rv2688c operon in *M. tuberculosis* is an example [39]. Four of the genes that might be controlled by the MSMEG\_0307 protein, encode proteins that are involved in metabolic pathways, including the NAT protein that has been shown to be involved in lipid biosynthesis [10] and a cholesterol degradation pathway [14]. In addition, there are three genes involved in oxidation/reduction reactions and three more that are involved in the regulation of gene expression. One of these 20 genes encodes an alternative sigma factor (*MSMEG\_0574*) *rpoE1*, a putative extracytoplasmic (ECF) function alternative sigma factor. *M. tuberculosis* contains 10 copies of ECF alternative sigma factors

and they are believed to act in a similar manner to the two component system enhancing the adaptation of bacteria under different physiological stages and pathogenesis [40]. It is clear that the binding of the MSMEG\_0307 protein to the conserved Motif 1 located upstream of the *nat* gene has a direct effect on the expression of *nat* and *MSMEG\_0308* transcripts.

Our preliminary studies using protein-fragment complementation [41] revealed the presence of a small protein complex made from the gene products of the *nat* operon (Figure S5). We also hypothesise that the MSMEG\_0307, transcription factor, might also interact with sigma factors as parts of the RNA polymerase in order to support the initiation of the transcription on this genomic area.

Although there is no clear evidence of an orthologue of the MSMEG\_0307 AraC protein in *M. tuberculosis* it is clear from comparative genomic analyses that all fast-growing environmental mycobacteria sequenced to date possess a similar gene architecture in their *nat* gene clusters and that the AraC-family transcriptional regulator MSMEG\_0307 protein and its preferred DNA binding motif are highly conserved, suggesting that this genomic area plays an important role in the adaptation of these mycobacterial species to their specific environment. It will be interesting to identify the external stimuli as well as the ligands that bind to the MSMEG\_0307 protein. Furthermore, ascertaining the biological significance of the MSMEG\_0307 protein being situated next to NAT in environmental mycobacteria and providing an explanation of the different evolutionary pathways adopted by fast-growing environmental and slow-growing pathogenic mycobacteria, will give us an insight into the unique characteristics of the adaptation of this genus to multiple environments. This is the first report of the role of the AraC-family transcriptional regulator MSMEG\_0307 protein from the *M. smegmatis* *nat* operon.

#### Acknowledgements

DE acknowledges Ricardo Balhana, The Royal Veterinary College, for helpful discussions on the MEME/MAST software. We are grateful to Professor William R. Jacobs, Jr., Albert Einstein College of Medicine, for his generous gift of the plasmid and phasmid used in this study and support during the generation of the gene deletion mutants. DE acknowledges a Sybil Evans research studentship supports from Birkbeck, University of London and the Department of Pharmacology at the University of Oxford.

**Funding:** This work was supported by Birkbeck Faculty of Science Research Grant, University of London Central Research Grant and a UK Medical Research Council New Investigators Research Grant (G0801956) to SB.

**Competing interests:** None declared.

**Ethical approval:** Not required.

#### Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.tube.2014.08.007>.

#### References

- [1] Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos JL. AraC/XylS family of transcriptional regulators. *Microbiol Mol Biol Rev*: MMBR 1997;61:393–410.
- [2] Munson GP, Scott JR. Binding site recognition by Rns, a virulence regulator in the AraC family. *J Bacteriol* 1999;181:2110–7.
- [3] Camus JC, Pryor MJ, Medigue C, Cole ST. Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* 2002;148:2967–73.

- [4] Camacho LR, Ensergueix D, Perez E, Gicquel B, Guilhot C. Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. *Mol Microbiol* 1999;34:257–67.
- [5] Frota CC, Papavinasasundaram KG, Davis EO, Colston MJ. The AraC family transcriptional regulator Rv1931c plays a role in the virulence of *Mycobacterium tuberculosis*. *Infect Immun* 2004;72:5483–6.
- [6] Singh A, Jain S, Gupta S, Das T, Tyagi AK. *mymA* operon of *Mycobacterium tuberculosis*: its regulation and importance in the cell envelope. *FEMS Microbiol Lett* 2003;227:53–63.
- [7] Singh A, Gupta R, Vishwakarma RA, Narayanan PR, Paramasivan CN, Ramanathan VD, et al. Requirement of the *mymA* operon for appropriate cell wall ultrastructure and persistence of *Mycobacterium tuberculosis* in the spleens of guinea pigs. *J Bacteriol* 2005;187:4173–86.
- [8] Singh R, Singh A, Tyagi AK. Deciphering the genes involved in pathogenesis of *Mycobacterium tuberculosis*. *Tuberculosis* 2005;85:325–35.
- [9] Evangelopoulos D, Cronin N, Daviter T, Sim E, Keep NH, Bhakta S. Characterization of an oxidoreductase from the arylamine N-acetyltransferase operon in *Mycobacterium smegmatis*. *FEBS J* 2011;278:4824–32.
- [10] Bhakta S, Besra GS, Upton AM, Parish T, Sholto-Douglas-Vernon C, Gibson KJ, et al. Arylamine N-acetyltransferase is required for synthesis of mycolic acids and complex lipids in *Mycobacterium bovis* BCG and represents a novel drug target. *J Exp Med* 2004;199:1191–9.
- [11] Anderton MC, Bhakta S, Besra GS, Jeavons P, Eltis LD, Sim E. Characterization of the putative operon containing arylamine N-acetyltransferase (*nat*) in *Mycobacterium bovis* BCG. *Mol Microbiol* 2006;59:181–92.
- [12] Yam KC, D'Angelo I, Kalscheuer R, Zhu H, Wang J-X, Snieckus V, et al. Studies of a ring-cleaving dioxygenase illuminate the role of cholesterol metabolism in the pathogenesis of *Mycobacterium tuberculosis*. *PLoS Pathog* 2009;5:e1000344.
- [13] Van der Geize R, Yam K, Heuser T, Wilbrink MH, Hara H, Anderton MC, et al. A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. *Proc Natl Acad Sci U S A* 2007;104:1947–52.
- [14] Lack NA, Kawamura A, Fullam E, Laurieri N, Beard S, Russell AJ, et al. Temperature stability of proteins essential for the intracellular survival of *Mycobacterium tuberculosis*. *Biochem J* 2009;418:369–78.
- [15] Abuhammad A, Fullam E, Lowe ED, Staunton D, Kawamura A, Westwood IM, et al. Piperidinols that show anti-tubercular activity as inhibitors of arylamine N-acetyltransferase: an essential enzyme for mycobacterial survival inside macrophages. *PLoS ONE* 2012;7:e52790.
- [16] Griffin JE, Pandey AK, Gilmore SA, Mizrahi V, McKinney JD, Bertozzi CR, et al. Cholesterol catabolism by *Mycobacterium tuberculosis* requires transcriptional and metabolic adaptations. *Chem Biol* 2012;19:218–27.
- [17] Kendall SL, Withers M, Soffair CN, Moreland NJ, Gurcha S, Sidders B, et al. A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *Mol Microbiol* 2007;65:684–99.
- [18] Kendall SL, Burgess P, Balhana R, Withers M, Ten Bokum A, Lott JS, et al. Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: *kstR* and *kstR2*. *Microbiology* 2010;156:1362–71.
- [19] Snapper SB, Melton RE, Mustafa S, Kieser T, Jacobs Jr WR. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* 1990;4:1911–9.
- [20] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–10.
- [21] McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res* 2013;41:W597–600.
- [22] Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. ACT: the artemis comparison tool. *Bioinformatics* 2005;21:3422–3.
- [23] Carver T, Berriman M, Tivey A, Patel C, Bohme U, Barrell BG, et al. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 2008;24:2672–6.
- [24] Mulder MA, Zappe H, Steyn LM. Mycobacterial promoters. *Tuber Lung Dis* 1997;78:211–23.
- [25] Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 1994;2:28–36.
- [26] Bailey TL, Gribskov M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* 1998;14:48–54.
- [27] Bardarov S, Bardarov Jr S, Pavelka Jr MS, Sambandamurthy V, Larsen M, Tufariello J, et al. Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in *Mycobacterium tuberculosis*, *M. bovis* BCG and *M. smegmatis*. *Microbiology* 2002;148:3007–17.
- [28] Stover CK, de la Cruz VF, Fuerst TR, Burlein JE, Benson LA, Bennett LT, et al. New use of BCG for recombinant vaccines. *Nature* 1991;351:456–60.
- [29] Taneja NK, Tyagi JS. Resazurin reduction assays for screening of anti-tubercular compounds against dormant and actively growing *Mycobacterium tuberculosis*, *Mycobacterium bovis* BCG and *Mycobacterium smegmatis*. *J Antimicrob Chemother* 2007;60:288–93.
- [30] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>(-Delta Delta C(T))</sup> Method. *Methods* 2001;25:402–8.
- [31] Wade JT, Belyaeva TA, Hyde EI, Busby SJ. Repression of the *Escherichia coli* *meIR* promoter by *MeIR*: evidence that efficient repression requires the formation of a repression loop. *Mol Microbiol* 2000;36:223–9.
- [32] Schleif R. AraC protein: a love-hate relationship. *Bioessays* 2003;25:274–82.
- [33] Lack NA, Yam KC, Lowe ED, Horsman GP, Owen RL, Sim E, et al. Characterization of a carbon-carbon hydrolase from *Mycobacterium tuberculosis* involved in cholesterol metabolism. *J Biol Chem* 2010;285:434–43.
- [34] Payton M, Gifford C, Schartau P, Hagemeyer C, Mushtaq A, Lucas S, et al. Evidence towards the role of arylamine N-acetyltransferase in *Mycobacterium smegmatis* and development of a specific antiserum against the homologous enzyme of *Mycobacterium tuberculosis*. *Microbiology* 2001;147:3295–302.
- [35] Veleba M, Higgins PG, Gonzalez G, Seifert H, Schneiders T. Characterisation of *RarA*, a novel AraC-family multidrug resistance regulator in *Klebsiella pneumoniae*. *Antimicrob Agents Chemother* 2012;56:4450–8.
- [36] Rosenfeld N, Elowitz MB, Alon U. Negative autoregulation speeds the response times of transcription networks. *J Mol Biol* 2002;323:785–93.
- [37] Madar D, Dekel E, Bren A, Alon U. Negative auto-regulation increases the input dynamic-range of the arabinose system of *Escherichia coli*. *BMC Syst Biol* 2011;5:111.
- [38] Braibant M, Gilot P, Content J. The ATP binding cassette (ABC) transport systems of *Mycobacterium tuberculosis*. *FEMS Microbiol Rev* 2000;24:449–67.
- [39] Pasca MR, Gugliera P, Arcesi F, Bellinzoni M, De Rossi E, Riccardi G. Rv2686c-Rv2687c-Rv2688c, an ABC fluoroquinolone efflux pump in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* 2004;48:3175–8.
- [40] Staron A, Sofia HJ, Dietrich S, Ulrich LE, Liesegang H, Mascher T. The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) sigma factor protein family. *Mol Microbiol* 2009;74:557–81.
- [41] Singh A, Mai D, Kumar A, Steyn AJ. Dissecting virulence pathways of *Mycobacterium tuberculosis* through protein-protein association. *Proc Natl Acad Sci U S A* 2006;103:11346–51.