Mce3R, a TetR-type transcriptional repressor, controls the expression of a regulon involved in lipid metabolism in *Mycobacterium tuberculosis*

María de la Paz Santangelo,1 Laura Klepp,1 Javier Nuñez-García,2 Federico C. Blanco,1 Marcelo Soria,3 María del Carmen García-Pelayo,2 María Verónica Bianco,1 Angel A. Cataldi,1 Paul Golby,2 Mary Jackson,4 Stephen V. Gordon2‡ and Fabiana Bigi1

1Instituto de Biotecnología, CICVyA-INTA, N. Repetto and De los Reseros, 1686 Hurlingham, Argentina
2Veterinary Laboratories Agency (Weybridge), New Haw, Addlestone, Surrey KT15 3NB, UK
3Microbiología Agrícola Facultad de Agronomía, Universidad de Buenos Aires, Av. San Martin 4453, 1417 Buenos Aires, Argentina
4Mycobacteria Research Laboratories, Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO 80523-1682, USA

The *mce* operons constitute four homologous regions in the *Mycobacterium tuberculosis* genome, each of which has 8–13 ORFs. Although the function of the Mce protein family has not been clearly established, its members are believed to be membrane lipid transporters. Based on functional experiments, we found that the regulator of the *mce3* locus, Mce3R, negatively regulates the expression of the *Rv1933c–Rv1935c* and *Rv1936–Rv1941* transcriptional units. These operons are adjacent to one another and divergently transcribed. The predicted functions of most of these genes are related to either lipid metabolism or redox reactions. Bioinformatic analysis of the 5′ UTR sequences of the differentially expressed genes allowed us to define a putative Mce3R motif. Importantly, the Mce3R motif was present six and three times in the *mce3R–yrbE3A* and *Rv1935c–Rv1936* intergenic regions, respectively. Two occurrences of this motif mapped within the two regions of the *mce3* operon that were protected by Mce3R in a footprinting analysis, thus indicating that this motif is likely to serve as an operator site for the Mce3R regulator in the promoter. In addition, alterations in the lipid content of *M. tuberculosis* were detected in the absence of Mce3R. Taken together, these results suggest that Mce3R controls the expression of both the putative transport system encoded in the *mce3* operon and the enzymes implicated in the modification of the Mce3-transported substrates.

INTRODUCTION

Tuberculosis (TB) remains one of the leading causes of mortality throughout the world. The HIV/AIDS pandemic, deterioration in public health systems in developing countries, and the emergence of multiresistant and extremely drug resistant forms of TB have further contributed to that spread. *Mycobacterium tuberculosis*, the agent of human TB, has a remarkable ability to survive in the diverse conditions encountered during infection. However, our understanding of how *M. tuberculosis* regulates its gene expression to effect survival in the host remains poor.
The genome of *M. tuberculosis* contains four *mce* operons (*mce1–4*). Each *mce* operon contains 8–13 genes, with a similar arrangement within each operon (Cole *et al.*, 1998). The MceA protein has been suggested to be involved in the entry of the pathogen into non-phagocytic cells (Arruda *et al.*, 1993), and the role of the Mce proteins during the replication of *M. tuberculosis* in mice has been reported by several groups, including ours (Sassetti & Rubin, 2003; Gioffré *et al.*, 2005; Shimono *et al.*, 2003).

The availability of *Mycobacterium avium* and *Mycobacterium smegmatis* genome sequences (http://cmr.jcvi.org/) revealed that orthologues of the *mce* genes are present in these opportunistic species. Moreover, *mce* operons with an identical structure have been identified in all *Mycobacterium* species examined, as well as in five other species of the *Actinomycetales* (Casali & Riley, 2007).

There is evidence for the differential expression of *mce* operons under different environmental and experimental conditions. It has been found that there are growth-phase- and tissue-specific differences in the expression of the *mce* operons in *M. tuberculosis* (Kumar *et al.*, 2003), which is in agreement with the presence of regulatory mechanisms controlling *mce* transcription. It is tempting to speculate that such differential regulation, together with gene duplication, ensures the production of Mce in different host environments.

Although the function of the members of the Mce protein family has not been clearly established, it has been hypothesized that they are membrane lipid transporters. Based on the identification of conserved domains in many Mce proteins, previous studies strongly suggest that the *mce* and *yrbE* genes encode components of ABC transport systems (Kumar *et al.*, 2005; Casali & Riley, 2007). Moreover, it has recently been demonstrated that *mce4* encodes a cholesterol import system that enables *M. tuberculosis* to derive both carbon and energy from host cholesterol (Pandey & Sassetti, 2008). Interestingly, Kendall *et al.* (2007) have shown that the *mce4* operon is regulated by KstR, a TetR-type regulator, and co-regulated with other genes involved in fatty acid metabolism.

We have previously reported that Mce3R, a TetR family transcriptional regulator, downregulates the *mce3* operon during *in vitro* growth of *M. tuberculosis* (Santangelo *et al.*, 2002). We have found that this regulation is specific for the *mce3* operon (among all the *mce* genes) and that the Mce3R repressor regulates its own expression (Santangelo *et al.*, 2008). In this study, we investigated additional genes that are regulated by Mce3R in order to define the Mce3R regulon. We reasoned that the identification of genes with a putative function that are regulated by Mce3R could give clues as to the functional role of the *mce3*-encoded proteins.

In this work, we used microarray analysis to compare the transcriptional profile of an *M. tuberculosis* Δ*mce3R* mutant strain with that of wild-type *M. tuberculosis* H37Rv. We found that Mce3R controls the expression of a number of genes involved in lipid metabolism and β-oxidation in *M. tuberculosis*, thus supporting the hypothesis that the *mce* operons encode lipid transporters. We validated the microarray data for the genes whose expression varied greatly between strains by qRT-PCR. As a complementary approach we compared the protein expression patterns of the wild-type strain and the Δ*mce3R* mutant by 1D and 2D PAGE. In addition, we identified a region within the *mce3* promoter, protected in a footprinting assay, that is conserved in most of the promoter regions of the Mce3R-regulated genes. Finally we analysed the lipid profile of both strains in order to confirm the association of the *mce3* regulon with lipid transport and/or metabolism.

**METHODS**

**Bacterial strains and culture media.** The construction of the Δ*mce3R* mutant is described by Santangelo *et al.* (2008). All cloning steps were performed in *Escherichia coli* DH5α. *E. coli* BL21(DE3) was used for recombinant protein production. *E. coli* was grown either in Luria–Bertani (LB) broth or on LB agar. *M. tuberculosis* strains were grown in Middlebrook 7H9 medium supplemented with 0.05% Tween 80, or Middlebrook 7H11 supplemented with 0.5% albumin, 0.4% dextrose (glucose) and 0.5% glycerol (M7H9-AD-G). When necessary, 50 μg hygromycin ml⁻¹ was added to the media.

**General DNA methodology.** PCR amplifications from genomic DNA templates were performed as previously described (Santangelo *et al.*, 2002). Each primer contained base mismatches that introduced a restriction site suitable for directional cloning (Table 1). Chromosomal DNA samples were obtained as described by van Soolingen *et al.* (1991). Purification of plasmids and DNA fragments was performed using the GFX Micro Plasmid Prep kit (GE Healthcare) and the DNA and Gel Band Purification kit (GE Healthcare), respectively, according to the manufacturer’s instructions.

**RNA preparation.** DNA-free RNA was extracted from 50 ml mid-exponential-phase cultures of *M. tuberculosis* strains as described by Santangelo *et al.* (2002).

**cDNA labelling.** Fluorescently labelled cDNA was produced using either total RNA or genomic DNA as a template as described previously (Golby *et al.*, 2007). Briefly, total *M. tuberculosis* RNA (8 μg) was reverse-transcribed with Superscript II (Invitrogen) in the presence of Cy5-dCTP (Amersham Bioscience) using random hexamer oligonucleotides (Invitrogen) to prime cDNA synthesis.

Whole genomic DNA of *M. tuberculosis* H37Rv (3 μg) was used as a template for a randomly primed polymerization reaction with the Klenow fragment of DNA polymerase (New England Biolabs) in the presence of Cy3-dCTP (Amersham Bioscience) and random hexamer oligonucleotides.

**DNA microarray hybridizations and scanning.** A DNA microarray containing non-redundant CD5 from the two sequenced *M. tuberculosis* strains, CDC1551 and H37Rv, and from *M. bovis* AF2122/97, was used in these experiments. The microarrays were developed by the Bacterial Microarray Group (BuG®S, St Georges, University of London) in collaboration with the Veterinary Laboratories Agency (Weybridge). The array design is available in BuG®Sbase, accession no. A-BUGS-31 (http://bugs.sgal.ac.uk/A-
**Table 1.** Primers used in this study

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer (5’–3’)*</th>
<th>Reverse primer (5’–3’)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>echA (Rv1935c)</td>
<td>gacctcctgggaggtgcg</td>
<td>gatgtcgacgccccaaac</td>
</tr>
<tr>
<td>fadE18 (Rv1933c)</td>
<td>aacgggaaactgcggcag</td>
<td>gggttagaactttcagg</td>
</tr>
<tr>
<td>ephB (Rv1938)</td>
<td>ccaagactatcgccgaacta</td>
<td>gctagggactgcagcaaaac</td>
</tr>
<tr>
<td>ribA (Rv1940)</td>
<td>cggccggtctgttaatgcct</td>
<td>agtncagcgtcgcggatag</td>
</tr>
<tr>
<td>Rv1942</td>
<td>gttggtgatagctggatgat</td>
<td>accgtgtgcaagaaaccct</td>
</tr>
<tr>
<td>mce3R (Rv1963)</td>
<td>gggccatgctggcctgcg</td>
<td>gatctctatgccgcaagggttaacct</td>
</tr>
</tbody>
</table>

*Restriction enzyme site added at the end of each primer is underlined.

BUGS-31 and also ArrayExpress (http://www.ebi.ac.uk/microarray-as/aef), accession no. A-BUGS-31.

Prehybridization, hybridization and washing were performed as described previously (Golby et al., 2007). Microarrays were hybridized with a combination of Cy3-cDNA generated from genomic DNA of *M. tuberculosis* H37Rv and Cy5-cDNA obtained from total RNA of either *M. tuberculosis* H37Rv or the Δmce3R mutant.

Microarray slides were incubated in prehybridization solution (3.5 × SSC, 0.1 % SDS and 10 mg ml⁻¹ BSA) at 65 °C for 20 min. Slides were rinsed in water for 1 min and in propan-2-ol for 1 min before drying by centrifugation at 400 g for 5 min. The purified Cy3/Cy5-labelled DNA was adjusted to 60 μl in 4 × SSC and 0.3 % SDS. The hybridization mixture was denatured at 95 °C for 2 min, cooled to room temperature, applied to the array and covered with a 2.2 × 5 cm coverslip. The slide was placed in a waterproof hybridization chamber and incubated at 65 °C in the dark for 16–20 h. After hybridization, slides were washed twice at 65 °C for 2 min in 1 × SSC buffer with 0.05 % SDS, followed by two washes of 2 min at room temperature in 0.06 × SSC, and then dried by centrifugation.

Eight sets of microarray data, consisting of four biological replicates (cells from independent cultures) in duplicate (technical replicates), were produced for each *M. tuberculosis* strain.

The microarrays were scanned with an Affymetrix 428 scanner. Fluorescent spot intensities were quantified using BlueFuse for Microarrays v3.2 (BlueGnome, www.cambridgebluegnome.com). For each spot, background fluorescence was subtracted from the average spot fluorescence to produce a channel-specific ratio.

**Data processing and statistical analysis.** For subsequent calculations, log2 cy5:cy3 (test:control) ratios were used. Within each microarray, block median normalization, excluding control and empty spots, was carried out using the BlueFuse software. Median absolute deviation using Mathematica 5.2 (Wolfram Research) was applied to bring the histograms of all microarrays into the same scale. Details of the calculation of the median absolute deviation are given in the supplementary data available with the online version of this paper. Technical replicates were averaged. Genes differentially expressed between the strains were detected by applying t-tests with a Benjamini and Hochberg adjusted P-value.

**RT-qPCR.** DNA-free RNA (1 μg) was mixed with 50 ng of random primers (Invitrogen) in a final volume of 20 μl and reverse-transcribed to total cDNA with SuperScript II reverse transcriptase (Invitrogen), following the manufacturer’s instructions. Identiﬁcal reactions lacking reverse transcriptase were also performed to confirm the absence of genomic DNA in all samples.

RT-qPCR was performed in the Applied Biosystems 7000 DNA sequence detection system (Perkin-Elmer), with Master Mix QuantiTect SYBR Green (Qiagen), 1 μl of template cDNA and the pairs of primers listed in Table 1. Each reaction was performed in duplicate. Results are presented as ratios calculated with the Relative Expression Software Tool (REST®) application described by Pfaffl et al., (2002), based on four biological replicates for in vitro studies. Relative quantification of each target gene was performed by using sigA as a reference gene, with real-time PCR efficiencies of target and reference genes considered as 2.

A subsequent test for signiﬁcance of the results was performed by using the pair-wise fixed reallocation randomization test (www.rest.gene-quantification.info).

**Expression and purification of recombinant Mce3R.** Full-length Rv1963 (mce3R) was PCR-ampliﬁed from *M. tuberculosis* H37Rv with the primers shown in Table 1 and cloned as a His-fusion protein into pRSET-A (Invitrogen). The resulting plasmid, pRSETmce3R, was introduced into *E. coli* BL21(DE3). Recombinant *E. coli* were grown in 500 ml LB medium containing 125 μg ampicillin ml⁻¹ at 28 °C. When the OD₆₀₀ reached 0.3–0.5, expression of the gene encoding the recombinant protein was induced with 0.1 mM IPTG overnight. Cells were then harvested by centrifugation and resuspended in 100 mM Tris/HCl, pH 7.5, 1 M NaCl, 20 % glycerol, 1 % NP-40 and 0.5 mM PMSF. Soluble cell extracts from the cultures were prepared as described elsewhere (Santangelo et al., 2008) and His-Mce3R recombinant protein was puriﬁed from the supernatants by using Ni⁺⁺ resin (Ni-NTA agarose, Qiagen), following the manufacturer’s recommendation (Invitrogen).

**DNase I footprinting.** A DNA region encompassing P₄₀₀ was PCR-ampliﬁed by using speciﬁc primers in which the antisense primer was end-labelled. rMce3R was incubated with the ³²P-labelled fragment. DNA binding was performed in a 25 μl reaction volume containing binding buffer [4 % (v/v) glycerol, 1 mM MgCl₂, 0.5 mM EDTA, 0.5 mM DTT, 50 mM NaCl, 10 mM Tris/HCl (pH 7.5), 0.05 mg salmon sperm DNA ml⁻¹ and 2.5 μg poly(dI-dC) ml⁻¹ (Amersham Bioscience)], 25 fmol labelled DNA and 200–750 pmol rMce3R. After incubation at room temperature for 40 min, DNase digestion was performed as described in the Promega Technical Bulletin (www.promega.com/tbs/tb137/tb137.pdf). Sequencing was performed using the fmol DNA Cycle Sequencing System kit (Promega). Samples were run in a urea-denaturing 6 % polyacrylamide gel. After drying, the gels were exposed to Storage Phosphor screen (Molecular Dynamics) for 1 h and scanned with a Typhoon Trio scanner (GE-Bioscience).

**Search for regulatory motifs.** The computer programs MEME (Bailey & Elkan, 1994) and MAST (Bailey & Gribskov, 1998) running on a web server (http://meme.sdsc.edu/meme4_1/intro.html) were used for motif discovery and alignment analysis of the putative promoter-containing intergenic regions of genes or operons that were upregulated in the microarray experiments. The intergenic sequences of *M. tuberculosis* H37Rv were obtained from the MibRegList R. 1.1
database (http://www.usherbrooke.ca/vers/MtbRegList; Jacques et al., 2005). MEME was used with a set of training sequences for motif discovery. The program was set to identify any number of occurrences of motifs because a preliminary analysis showed that certain motifs were repeated over some intergenic regions. The minimum and maximum width for motifs was set at 9 and 20 nucleotides, respectively. MAST was used to align significant motifs over the training sequences and to search for new occurrences in the entire genome.

**Macrophage infections.** The murine macrophage-like cell line J774 was infected as described previously (Santangelo et al. 2002) with some modifications. Briefly, 5 x 10⁶ cells per well in 24-well plates were infected at an m.o.i. of 10 in RPMI medium. The *M. tuberculosis* strains were grown in Middlebrook 7H9 liquid medium, and cultures were passed 10 times through a syringe needle (25 gauge) so as to obtain bacterial suspensions free of clumps. The OD₆₀₀ was taken and cell viability was estimated with the Live/Dead BacLight Bacterial Viability kit (Invitrogen), according to the manufacturer’s instructions. Infected cells were incubated for 3 h, and then washed four times with fresh RPMI medium. At 0, 4, 24 and 72 h post-infection, the J774 cells were scraped and lysed with 100 μl of 1% Triton X-100 for 15 min. Serial dilutions in PBS of the recovered intracellular bacteria were plated onto Middlebrook 7H10 agar to count c.f.u. Two independent experiments were carried out, each in duplicate. Means of c.f.u. counts were tested for differences with Student’s t-test. Values were determined to be statistically significant at P<0.05.

**Determination of the lipid composition in *M. tuberculosis* strains.** [1-14C]Acetate (50 mCi; specific activity 56.7 Ci mol⁻¹, 2.1 TBq mol⁻¹) (American Radiolabeled Chemicals) was added to 30 ml of 10–12-day-old cultures (OD₆₀₀ 1.4–1.6) of *M. tuberculosis* strains, and incubation was continued for another 24 h at 37°C in roller bottles. Cells were collected, suspended in PBS and disrupted using a Fastprep FP120 bead-beater (40 s at 6 m s⁻¹) following the protocol of Bligh & Dyer (1959) and subjected to 1D TLC on silica gel 60F254 (precoated, 20 μm), to resolve lipids of various polarities. Plates were then exposed to Storage Phosphor screen (Molecular Dynamics) and the signal was detected with a Typhoon Trio scanner (GE).

**Analysis of the proteomic profiles of *M. tuberculosis* strains.** Pellets from 1 litre of 15-day-old cultures of each strain were washed and resuspended in cold 1× PBS pH 7.4, 1 mM EDTA, 100× protease inhibitor mix (GE Healthcare) and disrupted using a Fastprep FP120 bead-beater (40 s at 6 m s⁻¹, with Lysing Matrix B). The cell extract was clarified centrifugation at 3000 g for 20 min and cell wall proteins were obtained by further centrifugation at 23000 g for 30 min. The supernatant obtained was ultracentrifuged for 2 h at 50000 g to separate proteins from the cytoplasm and membrane.

Proteins from the cell wall pellet were obtained by the addition of 2 ml buffer (1× PBS pH 7.4, 1 mM EDTA, 2× SDS%), incubation for 2 h at 60°C and centrifugation at 15 min at 18000 g. Cell wall proteins present in the supernatant were then concentrated and desalted twice by precipitation with 10% TCA. Cytoplasmic proteins were also cleaned up by precipitation with TCA. Membrane proteins were obtained as described previously (Gu et al., 2003). The protein pellets were resuspended with rehydration buffer (8 M urea, 2% CHAPS, 0.5% IPG buffer pH 4–7 and 20 mM DTT), and 800 μg protein in a final volume of 450 μl was used to rehydrate a 24 cm immobilized pH 3–7 linear gradient strip (Immobiline DryStrips, GE Healthcare) for 16 h at room temperature, following the manufacturer’s instructions. Isoelectric focusing was carried out with an IETT IPI Ghor 3 system (GE Healthcare), using the following programme: 1, 0.5 kV h at 500 V; 2, 5.2 kV h (gradient) at 1000 V; 3, 13.5 kV h (gradient) at 8000 V; 4, 30 kV h at 8000 V.

After focusing, the strips were equilibrated for 20 min in equilibration buffer (2% SDS, 50 mM Tris/HCl pH 8.8, 6 M urea, 30% glycerol, 0.002% bromphenol blue and 0.5% DTT). The strips were then overlaid onto 12% SDS-polyacrylamide gels, and after electrophoresis, proteins were stained with colloidal Coomassie (2 g Coomassie brilliant blue G250 1 l⁻¹; 0.5 g Coomassie brilliant blue R250 1 l⁻¹; 5% methanol, 42.5% ethanol, 10% acetic acid).

Proteins from total cellular extracts and culture supernatants from *M. tuberculosis* strains were prepared as described previously (Bigi et al., 1997) and resolved on 12% SDS-polyacrylamide gels, which were then silver-stained (Blum et al., 1987). The differential bands or spots identified were cut from the gels, digested with modified porcine trypsin (Promega) and extracted as previously described (Bienvenut et al., 1999). MS analysis was carried out on a Bruker Ultraflex II mass spectrometer, with the matrix x-cyano-4-hydroxycinnamic acid. Data interpretation and protein identification were performed with the MS/MS spectra datasets using the Mascot search algorithm (Version 1.6b9, Matrix Science, available at http://www.matrixscience.com).

**RESULTS**

**Replication of the Δmce3R mutant in vitro and inside macrophages**

We first assessed the replication of mutant Δmce3R and the wild-type strain *in vitro* and intracellularly. The growth profiles of both strains under standard *in vitro* culture conditions and inside the murine macrophage cell line J774 showed similar doubling times (see Supplementary Fig. S1, available with the online version of this paper). Thus, the mutation of the *mce3R* gene does not appear to compromise the *in vitro* growth or the intracellular replication of *M. tuberculosis*.

**Microarray analysis**

To test the role of Mce3R in the regulation of genes other than those in the *mce3* operon, we performed whole-genome expression profiling on the Δmce3R knockout mutant of *M. tuberculosis* and the parental H37Rv strain. We found that 22 genes were significantly overexpressed in Δmce3R (*P<0.01*, fold changes >1). The difference in expression between the mutant and the wild-type strain for the 22 differentially expressed genes was low (fold changes from 1 to 3) for six of the genes, with the remaining 16 displaying a major alteration in expression (fold changes >3). As expected, all of the genes whose expression changed dramatically between strains were upregulated in the Δmce3R mutant, which is consistent with the absence of a transcriptional repressor (Table 2). Among the group of genes whose expression was significantly upregulated in the Δmce3R mutant were several *mce3* genes, as well as genes predicted to be involved in lipid metabolism.

We reasoned that only those genes whose expression was highly increased in the mutant strain would be regulated by Mce3R. Therefore only this latter group of genes was considered for subsequent analysis. A total of 16 genes...
showed at least a threefold greater expression in the mutant strain compared to their expression in the wild-type strain (Table 2). Interestingly, all of these differentially expressed genes were localized in the same genomic region of M. tuberculosis. With the exception of Rv1942c, the differentially expressed genes are clustered in three different loci, one of which is the mce3 operon. The other two sets of genes mapped to two genomic regions encompassing Rv1936–Rv1941 and Rv1933c–Rv1935c, which, on the basis of genome annotation, were expected to be transcribed as part of operon structures. The Rv1936–Rv1941 and Rv1933c–Rv1935c operons were adjacent to one another and divergently transcribed (Fig. 1).

### Validation of microarray analysis by RT-qPCR

The microarray data for the genes whose expression was dramatically upregulated in the Δmce3R mutant strain were validated by RT-qPCR. A subset of four genes, belonging to the two transcriptional units (other than the mce3 operon), as well as Rv1942c, were tested, and the RT-qPCR results corroborated the microarray data for the five genes (Fig. 2, Table 2). The concordance between the microarray results and the RT-qPCR results was acceptable, thus supporting the statistical approach used in this study.

### Identification of the Mce3R binding motifs

To define the binding sites of Mce3R, the area of DNA bound by recombinant Mce3R in the mce3 promoter region was determined by DNase I footprint mapping. Mce3R was overexpressed as a His-fusion in E. coli (rMce3R) and purified under native conditions. We then used this purified protein for a DNase I footprinting experiment. The DNA targets consisted of 400 bp DNA fragments corresponding to the mce3 operon promoter (Pmce3). As shown in Fig. 3, rMce3R protects two regions separated by 39 bp in the mce3 promoter, extending from position –214 to –182 and from –142 to –111 relative to the transcription start site of yrbE3A (the first ORF of the mce3 operon). These protected regions of 33 and 32 bp were called RII and RI, respectively.

Because protein binding to promoter elements constrains the evolution of the target nucleotides, regulatory motifs may be identified through their conservation among promoters. Based on this premise, we searched for conserved motifs in the upstream regions of the differentially expressed genes. A first cycle of motif discovery was performed using MEME with all those intergenic regions that contained the putative promoter regions of CDS that

### Table 2. Genes differentially transcribed between M. tuberculosis Δmce3R and H37Rv as determined by microarray assays

<table>
<thead>
<tr>
<th>Rv no.</th>
<th>Gene symbol</th>
<th>Description*</th>
<th>Fold change†</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv1930c</td>
<td>None</td>
<td>CHP</td>
<td>2.49</td>
<td>0.0058</td>
</tr>
<tr>
<td>Rv1931c</td>
<td>None</td>
<td>Transcriptional regulator</td>
<td>2.58</td>
<td>0.0095</td>
</tr>
<tr>
<td>Rv1933c</td>
<td>fade18</td>
<td>Probable acyl-CoA dehydrogenase</td>
<td>7.22</td>
<td>0.0006</td>
</tr>
<tr>
<td>Rv1934c</td>
<td>fade17</td>
<td>Probable acyl-CoA dehydrogenase</td>
<td>1.83</td>
<td>0.0004</td>
</tr>
<tr>
<td>Rv1935c</td>
<td>echa13</td>
<td>Possible enoyl-CoA hydratase Echa13 (enoyl hydrase) (unsaturated acyl-CoA hydratase) (Crotonase)</td>
<td>9.10</td>
<td>0.0001</td>
</tr>
<tr>
<td>Rv1936</td>
<td>None</td>
<td>Possible monoxygenase</td>
<td>1.04</td>
<td>0.0005</td>
</tr>
<tr>
<td>Rv1937</td>
<td>None</td>
<td>Possible electron transfer oxygenase (including ring-hydroxylating dioxygenase); 2Fe–2S ferredoxin-type iron–sulfur binding domain profile</td>
<td>5.94</td>
<td>0.0005</td>
</tr>
<tr>
<td>Rv1938</td>
<td>ephB</td>
<td>Probable epoxide hydrolase (epoxide hydratase).</td>
<td>6.09</td>
<td>0.0001</td>
</tr>
<tr>
<td>Rv1939</td>
<td>None</td>
<td>Probable FMN oxidoreductase</td>
<td>2.96</td>
<td>0.0010</td>
</tr>
<tr>
<td>Rv1940</td>
<td>ribA1</td>
<td>Probable bifunctional enzyme: GTP cyclohydrolase II and 3,4-dihydroxy-2-butanone-4-phosphate synthase</td>
<td>3.48</td>
<td>0.0001</td>
</tr>
<tr>
<td>Rv1941</td>
<td>None</td>
<td>Probable short-chain type dehydrogenase/reductase; similar to 5-cyclohexadiene-1,4-diol dehydrogenase.</td>
<td>4.40</td>
<td>0.0023</td>
</tr>
<tr>
<td>Rv1942c</td>
<td>None</td>
<td>CHP, possible toxin (TA system)</td>
<td>9.85</td>
<td>0.0014</td>
</tr>
<tr>
<td>Rv1943c</td>
<td>None</td>
<td>CHP, possible antitoxin (TA system)</td>
<td>4.18</td>
<td>0.0026</td>
</tr>
<tr>
<td>Rv1964</td>
<td>yrbE3A</td>
<td>Mce-family protein YrbE3A</td>
<td>4.87</td>
<td>0.0035</td>
</tr>
<tr>
<td>Rv1965</td>
<td>yrbE3B</td>
<td>Mce-family protein YrbE3B</td>
<td>5.12</td>
<td>0.0079</td>
</tr>
<tr>
<td>Rv1966</td>
<td>mce3A</td>
<td>Mce-family protein Mce3A</td>
<td>1.04</td>
<td>0.0021</td>
</tr>
<tr>
<td>Rv1967</td>
<td>mce3B</td>
<td>Mce-family protein Mce3B</td>
<td>4.49</td>
<td>0.0068</td>
</tr>
<tr>
<td>Rv1970</td>
<td>lprM</td>
<td>Mce-family lipoprotein LprM (Mce3E)</td>
<td>6.60</td>
<td>0.0059</td>
</tr>
<tr>
<td>Rv1971</td>
<td>mce3F</td>
<td>Mce-family protein Mce3F</td>
<td>6.40</td>
<td>0.0055</td>
</tr>
<tr>
<td>Rv1972</td>
<td>None</td>
<td>Mce-associated membrane protein</td>
<td>5.06</td>
<td>0.0018</td>
</tr>
<tr>
<td>Rv1973</td>
<td>None</td>
<td>Mce-associated membrane protein</td>
<td>5.36</td>
<td>0.0020</td>
</tr>
<tr>
<td>Rv1975</td>
<td>None</td>
<td>Protein expressed in hypoxia</td>
<td>4.35</td>
<td>0.0044</td>
</tr>
</tbody>
</table>

*According to TubercuList (http://genolist.pasteur.fr/TubercuList/). CHP, conserved hypothetical protein.
†M. tuberculosis Δmce3R vs M. tuberculosis H37Rv.
were upregulated in the microarray experiments with a fold change of $\geq 3$ and that had $P$-values in the modified $t$-test $\leq 0.05$. This initial step did not provide conclusive or well-conserved motifs.

A second training set was prepared with the intergenic regions of genes that had a fold change of $\geq 5$ and a $P$-value $\leq 0.01$. Several of the CDS that satisfied this condition belonged to the same operon; in these cases the intergenic region upstream of the first gene was used. The resulting training set was composed of the intergenic regions $mce3R$–$yrbE3A$, $Rv1935c$–$Rv1936$ and $Rv1944c$–$Rv1945$. Among the several motifs found by MEME, there was one, named motif 1, that had six repeats clustered in two groups in the intergenic region $mce3R$–$yrbE3A$ (Fig. 4a) and two more repeats in $Rv1935c$–$Rv1936$ (Fig. 4b).

The occurrences of motif 1 proximal to $mce3R$ showed $P$-values ranging between $1.3 \times 10^{-7}$ and $8.9 \times 10^{-7}$, while those proximal to $yrbE3A$ were more similar to each other and to the putative consensus, with $P$-values $2.1 \times 10^{-9}$, $3.7 \times 10^{-8}$ and $4.2 \times 10^{-7}$. The occurrences of motif 1 proximal to $Rv1935c$–$Rv1936$ showed $P$-values of $1.7 \times 10^{-8}$ and $1.4 \times 10^{-7}$, respectively. Interestingly, the distance that separates the two occurrences of motif 1 on the positive strand proximal to both $mce3R$ and $yrbE3A$ and in the intergenic region $Rv1935$–$Rv1936$ is 60 nucleotides, suggesting that this length enables the binding of a unit of Mce3R to each motif.

While there is only a weak consensus shown by the sequence logo (Fig. 4c), two occurrences of motif 1 close to $yrbE3A$ overlap with the regions protected in the DNase I footprinting experiment (Fig. 4a).

These findings suggest that, in addition to the $mce3$ operon and $mce3R$, which are already known to be under Mce3R regulation (Santangelo et al., 2002, 2008), the expression of the $Rv1936$–$Rv1941$ and $Rv1933c$–$Rv1935c$ loci is directly regulated by the binding of Mce3R to the operator motifs identified.

### Proteomic analysis

In a complementary approach to identify the genes regulated by Mce3R, we compared the protein expression patterns of the wild-type strain and the $\Delta mce3R$ mutant. Evaluation of protein fractions by 1D-PAGE revealed that EphB and Rv1936 were more abundant in the culture supernatant of $\Delta mce3R$ mutant than in the corresponding fractions of the wild-type strain (Fig. 5a, lanes 3 and 4).
also detected a higher level of EphB in cell extracts of the Δmce3R mutant than in the wild-type (Fig. 5a, lanes 1 and 2).

To improve the resolution of the Mce3R-dependent bands, membrane, cell wall and cytoplasmic fractions were separated by 2D-PAGE. Analysis of two sets of 2D-PAGE gels revealed that Rv1936, Rv1498c, Rv2315c, Rv0234c, Rv1323 (FadA4), Rv0046c and Rv1416 were overexpressed in the mutant strain. All of these proteins were localized in the cytoplasmic fraction, except Rv1936, which was identified in the cell wall fraction (Fig. 5b, Table 3). No differential spots were found in the cell membrane fraction, probably due to the low resolution of the protein spots in this cell fraction. In agreement with the microarray data, Rv1936 and EphB showed higher expression in the mutant strain (Table 2).

No occurrences of motif 1 were found in the regions upstream of genes encoding the proteins identified as overexpressed in the mutant strain.

Characterization of the genes encoded in the Mce3R regulon

Based on the expression analysis of the Δmce3R mutant and the presence of a regulatory motif in the promoter sequences, we defined the mce3, Rv1936–Rv1941 and Rv1933c–Rv1935c operons as members of the Mce3R regulon. Analysis of the functions of the proteins encoded in the Mce3R regulon was carried out by searching the TubercuList database (http://genolist.pasteur.fr/TubercuList/) and the literature. Their predicted functions are shown in Table 2. Most of these genes are related either to lipid metabolism or to redox reactions.

Fig. 3. DNase I footprints of the mce3 promoter. A 32P-labelled DNA fragment (25 fmol) carrying P400-mce3 was incubated in the absence or presence (200–750 pmol) of rMce3R and treated with 0.3 units DNase. The protected regions RI and RII are indicated by brackets. A better resolution of protected region RII is depicted on the right.

Fig. 4. (a, b) Occurrence of motif 1 in the intergenic region mce3R–yrbE3A (a) and Rv1935c–Rv1936 (b). The left- and right-pointing arrows indicate occurrences of motif 1 on the negative and positive strands respectively. The sequences of the positive strands are shown. RI and RII protected regions in the DNase I footprinting experiment are shadowed in grey in (a). (c) Sequence logo of motif 1 built from the six occurrences in the intergenic region mce3R–yrbE3A and from two in Rv1935c–Rv1936.
**Fig. 5.** Differential proteins identified by 1D (a) and 2D (b) gel electrophoresis. (a) Equal amounts of proteins from cell extracts (lanes 1 and 2) and culture supernatants (lanes 3 and 4) of *M. tuberculosis* H37Rv (lanes 1 and 3) and the Δmce3R mutant strain (lanes 2 and 4) were resolved on 12% polyacrylamide gels, which were silver-stained. Proteins more abundant in the Δmce3R mutant strain than in the wild-type strain are indicated. (b) The first-dimension isoelectric focusing used a pH range of 3–7. The second-dimension SDS-PAGE was on 12% polyacrylamide gels, which were stained with colloidal Coomassie blue. Proteins present in the cell wall (panel A) and cytoplasm (panels B and C) fractions of the Δmce3R strain and absent in the wild-type strain are indicated. One spot consisted of a mixture of two proteins (Rv2315c and Rv0234c).

**Table 3.** Differentially expressed proteins between *M. tuberculosis* Δmce3R and *M. tuberculosis* H37Rv

<table>
<thead>
<tr>
<th>Rv no.</th>
<th>Gene symbol</th>
<th>Description*</th>
<th>Cellular localization</th>
<th>T pI</th>
<th>E pI</th>
<th>T MM (kDa)</th>
<th>E MM (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv1938</td>
<td>ephB</td>
<td>Probable epoxide hydrolase (epoxide hydratase)</td>
<td>Cellular extract</td>
<td>4.8</td>
<td>–</td>
<td>39.2</td>
<td>42.0</td>
</tr>
<tr>
<td>Rv1936</td>
<td>None</td>
<td>Possible monoxygenase</td>
<td>Culture supernatant</td>
<td>4.8</td>
<td>–</td>
<td>39.2</td>
<td>40.0</td>
</tr>
<tr>
<td>Rv1498c</td>
<td>None</td>
<td>Possible methyltransferase</td>
<td>Culture supernatant</td>
<td>5.4</td>
<td>–</td>
<td>41.8</td>
<td>41.0</td>
</tr>
<tr>
<td>Rv2315c</td>
<td>None</td>
<td>Conserved hypothetical protein</td>
<td>Cell wall</td>
<td>5.4</td>
<td>5.0</td>
<td>41.8</td>
<td>40.0</td>
</tr>
<tr>
<td>Rv0234c</td>
<td>gabD1</td>
<td>Probable succinate-semialdehyde dehydrogenase, NADP+ dependent</td>
<td>Cytoplasm</td>
<td>6.5</td>
<td>4.5</td>
<td>23.1</td>
<td>55.0</td>
</tr>
<tr>
<td>Rv1323</td>
<td>fabA4</td>
<td>Probable acetyl-CoA acetyltransferase</td>
<td>Cytoplasm</td>
<td>4.7</td>
<td>4.8</td>
<td>40.0</td>
<td>45.0</td>
</tr>
<tr>
<td>Rv0046c</td>
<td>ino1</td>
<td>myo-Inositol-1-phosphate synthase involved in phosphatidylinositol biosynthetic pathway</td>
<td>Cytoplasm</td>
<td>4.8</td>
<td>5.0</td>
<td>40.0</td>
<td>45.0</td>
</tr>
<tr>
<td>Rv1416</td>
<td>ribH</td>
<td>Riboflavin synthase</td>
<td>Cytoplasm</td>
<td>5.1</td>
<td>5.2</td>
<td>16.3</td>
<td>20.0</td>
</tr>
</tbody>
</table>

*According to TubercuList (http://genolist.pasteur.fr/TubercuList/).
had suffered a deletion of the genes Rv1928c–Rv1936
Wheeler et al. (2008) were able to show a role for this operon in propanoate metabolism. Hence, the
role of this locus in lipid metabolism is further strengthened by the recent description of a function for the Rv1933c–Rv1935c transcriptional unit (Wheeler et al., 2008). El-Etr et al. (2004) found that several of the orthologous clusters that display the highest similarity to CDS in the Rv1936–Rv1941 of M. marinum are eukaryotic clusters (http://www.ncbi.nlm.nih.gov/COG/), suggesting that this locus is involved in metabolism of complex fatty acids obtained from eukaryotic cells. However, our study shows that the overexpression of Mce3R-dependent proteins did not improve the replication of M. tuberculosis in an epidemic clone of M. bovis in Great Britain that had suffered a deletion of the genes Rv1928c–Rv1936, Wheeler et al. (2008) were able to show a role for this deleted locus in propanoate metabolism. Hence, the available evidence supports the role of the Mce3R-controlled genes in lipid metabolism.

Lipid analysis
The DNA microarray and proteomics results revealed that many genes that are negatively regulated by Mce3R could be involved in lipid metabolism (Tables 2 and 3). We therefore analysed the lipid content of the wild-type and mutant strains biochemically. The lipid profiles of the cell extracts obtained from the wild-type and the Δmce3R mutant in the exponential phase of growth were compared. Analysis of the radiolabelled lipids fractioned by TLC showed that the lipid profile of the mutant Δmce3R was altered as compared with that of the wild-type strain, with several lipid bands absent in the mutant (Fig. 6).

DISCUSSION
Increasing evidence suggests that the mce operons encode transport systems in M. tuberculosis. Using in silico analysis, Kumar et al. (2005) identified the presence of domains for toluene tolerance in mce genes. In addition, using a bioinformatic approach, Casali & Riley (2007) suggested that mce genes encode ABC uptake systems with a role in the remodelling of the cell envelope. Joshi et al. (2006) identified a number of genes that seem to function in concert with the mce genes. One of these genes, mceG, encodes an ATPase, and many of these genes are predicted to be involved in lipid metabolism (Joshi et al., 2006).

An interesting alternative hypothesis has been proposed by Pandey & Sassetti (2008), who have recently demonstrated the involvement of Mce4 proteins in cholesterol uptake. They have shown that mce4 encodes an import system that enables M. tuberculosis to derive both carbon and energy from host cell membrane cholesterol.

In this study we found that the expression of the mce3 operon of M. tuberculosis is directly regulated by Mce3R together with two transcriptional units, Rv1936–Rv1941 and Rv1933c–Rv1935c, thus suggesting a functional relationship between the products encoded in the three operons.

The analysis of the putative function of the proteins encoded in the Rv1936–Rv1941 operon leads us to postulate that a cyclic compound with a lateral chain and an epoxy group may be a substrate of this group of enzymes. The localization of Rv1936 and EphB either in the cell wall or in the culture supernatant of M. tuberculosis suggests that these enzymes probably modify the Mce3 substrates before they are transported inside the bacteria. In this hypothetical model, Rv1936 oxidizes the substrate for degradation and EphB opens the epoxy group in a ring molecule. The ability of EphB of M. tuberculosis to hydrolyse small aromatic epoxide-containing substrates has recently been demonstrated (Biswal et al., 2006).

In addition, the orthologous Rv1936–Rv1941 operon in Mycobacterium marinum has been shown to be important for the ability of the bacteria to infect host cells (El-Etr et al., 2004). El-Etr et al. (2004) found that several of the orthologous clusters that display the highest similarity to CDS in the Rv1936–Rv1941 of M. marinum are eukaryotic clusters (http://www.ncbi.nlm.nih.gov/COG/), suggesting that this locus is involved in metabolism of complex fatty acids obtained from eukaryotic cells.
inside macrophages, thus indicating that, in the conditions assayed, the uptake of nutrients required for intracellular growth is independent of the Mce3R system.

The other transcriptional unit, \textit{Rv1933c–Rv1935c}, encodes two FadE proteins (FadE17 and FadE18) and EchA13. Although the function of these proteins has not been completely resolved, Wheeler \textit{et al.} (2008) have recently demonstrated that the \textit{Rv1933c–Rv1935c} operon is involved in propanoate metabolism. In addition, FadE proteins are acyl-CoA dehydrogenases that catalyse the first step of \(\beta\)-oxidation (oxidative degradation of fatty acids). EchA is an enoyl-CoA hydratase also involved in \(\beta\)-oxidation. The \(\beta\)-oxidation cycle is the dominant route in bacteria for oxidative degradation of fatty acids to acetyl-CoA and propionyl-CoA, which can be further oxidized via the citric acid cycle (Muñoz-Elias & McKinney, 2006). In agreement with our findings, the co-expression of the \textit{mce} operons and genes involved in lipid metabolism has been recently reported. KstR, a TetR-type regulator, has been previously shown to control the expression of the \textit{mce4} operon and that of other genes involved in fatty acid metabolism (Kendall \textit{et al.}, 2007). Moreover, the first gene in the \textit{mce1} operon is \textit{fadE5} (Casali \textit{et al.}, 2006). Therefore, all of these findings are clearly consistent with the proposed role of Mce proteins as components of ABC-like transport systems whose substrates may be lipids. This idea is also supported by the fact that the lipid profile of \textit{M. tuberculosis} was modified in the absence of Mce3R, as demonstrated in this study, indicating an alteration in lipid metabolism.

We also found that the transcription of \textit{Rv1942c} and the presence of the proteins \textit{Rv1948c}, \textit{Rv2315c}, \textit{Rv0234c} (\textit{GabD1}), \textit{Rv1323} (\textit{FadA4}), \textit{Rv0046c} (\textit{Ino1}) and \textit{Rv1416} (\textit{RibH}) in the bacterial cytoplasm seem to be dependent on the absence of Mce3R. However, the analysis of the upstream regions of their coding sequences did not reveal the presence of the Mce3R regulatory motif identified in this study, suggesting that Mce3R exerts an indirect effect on the expression of these proteins. Interestingly, FadA4 and Ino1 are thought to be involved in lipid degradation and the phosphatidylinositol biosynthetic pathway, respectively. \textit{Rv1942c}, which appears to be transcribed together with \textit{Rv1943c}, encodes part of a toxin–antitoxin (TA) system homologous to that of the MazEF family of \textit{E. coli} (Pandey & Gerdes, 2005).

\textit{M. tuberculosis} is a prototrophic and metabolically flexible organism, capable of oxidizing a variety of carbon substrates, including sugars, tricarboxylic acids and fatty acids (Wheeler & Ratledge, 1994). Emerging evidence suggests that fatty acids, rather than carbohydrates, are the dominant carbon substrates used by \textit{M. tuberculosis} during infection. Based on these observations, as well as on the findings of this study and the fact that Mce3 proteins are relevant for the replication of \textit{M. tuberculosis} in mice, we hypothesize that the Mce3R regulon encodes a lipid transport system and the enzymes involved in the modification/degradation of its transported substrate, which provides the pathogen with a carbon/energy source during its replication inside host cells.

In this study, we defined a motif that is present three times in the promoter regions of both \textit{mce3R} and the \textit{mce3} operon, and twice in the intergenic region \textit{Rv1935c–Rv1936}. Importantly, this motif is included in both DNase-protected regions identified in the \textit{mce3} promoter. This finding, together with the fact that only these promoters showed a major dysregulation in the absence of Mce3R, led us to define this motif as the site of recognition by Mce3R. Work is in progress to define the affinity of Mce3R to this motif.

In conclusion, the present investigation provides new evidence that the Mce3R regulon plays a role in the adaptation \textit{M. tuberculosis} for survival in the host. However, further research is necessary to decipher the mechanism through which the Mce3R regulon contributes to \textit{M. tuberculosis} virulence.

\textbf{REFERENCES}


Edited by: M. Daffe