Kendall, Sharon L; Withers, Mike; Soffair, Catherine N; Moreland, Nicole J; Gurcha, Sudagar; Sidders, Ben; Frita, Rosangela; Ten Bokum, Annemieke; Besra, Gurdyal S; Lott, J Shaun; +1 more... Stoker, Neil G; (2007) A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in Mycobacterium smegmatis and Mycobacterium tuberculosis. Molecular microbiology, 65 (3). pp. 684-699. ISSN 0950-382X DOI: https://doi.org/10.1111/j.1365-2958.2007.05827.x

Downloaded from: http://researchonline.lshtm.ac.uk/id/eprint/9132/

DOI: https://doi.org/10.1111/j.1365-2958.2007.05827.x

Usage Guidelines:

Please refer to usage guidelines at https://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by/2.5/
A highly conserved translational repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*

Sharon L. Kendall,1 Mike Withers,1 Catherine N. Soffair,1 Nicole J. Moreland,2 Sudagar Gurcha,3 Ben Sidders,1 Rosangela Frita,1 Annemieke ten Bokum,4 Gurdyal S. Besra3 J. Shaun Lott,2 and Neil G. Stoker1*

1Department of Pathology and Infectious Diseases, The Royal Veterinary College, Royal College Street, London NW1 0TU, UK.
2Laboratory of Structural Biology and Maurice Wilkins Centre for Molecular Biodiscovery, School of Biological Sciences, University of Auckland, Auckland, New Zealand.
3School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, and
4Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK.

Summary

The *Mycobacterium tuberculosis* TetR-type regulator Rv3574 has been implicated in pathogenesis as it is induced in vivo, and genome-wide essentiality studies show it is required for infection. As the gene is highly conserved in the mycobacteria, we deleted the Rv3574 orthologue in *Mycobacterium smegmatis* (MSMEG_6042) and used real-time quantitative polymerase chain reaction and microarray analyses to show that it represses the transcription both of itself and of a large number of genes involved in lipid metabolism. We identified a conserved motif within its own promoter (TnnAACnnGTTnnA) and showed that it binds as a dimer to 29 bp probes containing the motif. We found 16 and 31 other instances of the motif in intergenic regions of *M. tuberculosis* and *M. smegmatis* respectively. Combining the results of the microarray studies with the motif analyses, we predict that Rv3574 directly controls the expression of 83 genes in *M. smegmatis*, and 74 in *M. tuberculosis*. Many of these genes are known to be induced by growth on cholesterol in rhodococci, and palmitate in *M. tuberculosis*. We conclude that this regulator, designated elsewhere as kstR, controls the expression of genes used for utilizing diverse lipids as energy sources, possibly imported through the mce4 system.

Introduction

The success of *Mycobacterium tuberculosis* as a pathogen (Corbett et al., 2003) lies partly in its ability to adapt to varying conditions within the host. This adaptation depends on the co-ordination of gene expression via the regulation of transcription; in *M. tuberculosis* this is achieved by the collective action of the 190 transcriptional regulators that the genome encodes (Cole et al., 1998; Camus et al., 2002). The importance of these genes in pathogenesis is illustrated by the observations that in many cases, inactivation of genes encoding sigma factors (Chen et al., 2000; Ando et al., 2003; Sun et al., 2004; Calamita et al., 2005) or two-component regulatory systems (Perez et al., 2001; Zahrt and Deretic, 2001; Parish et al., 2003; Malhotra et al., 2004; Rickman et al., 2004; Martin et al., 2006; Walters et al., 2006) causes severe attenuation in vivo. However, the identities of the genes controlled by the majority of the transcription factors, and the functional roles of these genes in vivo, remain largely unknown.

The application of microarray technology to the study of bacterial gene expression during infection has allowed genome-wide analyses of genes important in pathogenesis. We previously reported a meta-analysis (Kendall et al., 2004) of data from studies in *M. tuberculosis*, and showed that there was (surprisingly) very little correlation between the lists of genes that were induced during infection (Schnappinger et al., 2003; Talaat et al., 2004) and those that were essential for infection (Sassetti and Rubin, 2003; Rengarajan et al., 2005). Indeed, only one gene was reported to be upregulated during macrophage infection, upregulated at the onset of acquired immunity in mice, and essential for infection in mice: *Rv3574*.

Rv3574 is a member of the TetR family of transcriptional regulators. These proteins are often repressors and are...
widely distributed among bacteria, regulating a number of diverse processes (Ramos et al., 2005). The prototype for this group is TetR from the Tn10 transposon of *Escherichia coli*, which regulates the expression of a tetracycline efflux pump in Gram-negative bacteria (Orth et al., 2000). Other members of the TetR family include *Staphylococcus aureus* QacR, which regulates the expression of a multidrug transporter (Schumacher et al., 2001), and *M. tuberculosis* EthR, which regulates the expression of ethA, a monoxygenase that catalyses the activation of ethionamide, an antibiotic used in tuberculosis treatment (Baulard et al., 2000; Dover et al., 2004).

In this work we have examined the function of *Rv3574* in order to clarify the importance implied by our metanalysis (Kendall et al., 2004). Our bioinformatic analyses indicate that *Rv3574* is highly conserved within the mycobacteria, and accordingly we have studied the function of orthologues in both *M. tuberculosis* and the fast-growing non-pathogen *M. smegmatis*. We inactivated the *Rv3574* orthologue in *M. smegmatis*, and used microarrays to identify a large number of genes that are de-repressed in the mutant. We identified a conserved regulatory motif present in the upstream regions of the genes in the regulon and also describe the same motif in *M. tuberculosis*. We show that recombinant *M. tuberculosis* Rv3574 binds as a dimer to short synthetic pieces of DNA containing this motif, and describe the likely regulons for *Rv3574* both in *M. tuberculosis* and in *M. smegmatis*. The functional relevance of the regulon in pathogenesis is discussed.

**Results**

*Rv3574* is a member of the TetR family of transcriptional regulators and is highly conserved in the mycobacteria

Orthologues of *Rv3574* were identified through a combination of sequence similarity and synteny (the conservation of adjacent genes). In all cases, *Rv3574* and its orthologues are transcribed divergently from orthologues of the *M. tuberculosis* fadE34, encoding an acyl-CoA dehydrogenase (Fig. 1). The *Rv3574* region is highly conserved within the mycobacteria and is also conserved in the closely related species *Nocardia farcinica* (all > 70% amino acid identity over the whole length of the protein, and > 90% amino acid identity over the DNA binding domain). No convincing orthologue was found in the corynebacteria, while in *Streptomyces coelicolor*, a possible orthologue was found (SCO2319, 32% amino acid identity over the whole length of the protein) but with no conservation of synteny. In *M. leprae*, *Rv3574* is present as a pseudogene.

While we were writing this manuscript, a paper was published in which the *Rhodococcus* sp. strain RHA1

**Fig. 1.** Conservation of the *Rv3574* region in the mycobacteria. *Rv3574* and its orthologues in *M. smegmatis* are shown in white, and other genes are shown in black. In all sequenced mycobacterial genomes and in *Nocardia farcinica*, a *fadE* gene encoding an acyl-CoA dehydrogenase was found adjacent to, but divergently transcribed from, *Rv3574* and its orthologues. The numbering for the *M. smegmatis* genes refers to the gene names (e.g. 6042 refers to MSMEG_6042).

*Rv3574* orthologue is referred to as *kstR* (Van der Geize et al., 2007). In order to aid clarity when we discuss orthologues from different species, we will use this name hereafter, and discuss the relevance of their work later.

**Deletion of *kstR*<sub>Msm</sub> (MSMEG_6042) causes a defect in growth in vitro**

A 646 bp pair deletion, removing the entire N-terminal DNA binding domain, was made in *kstR*<sub>Msm</sub> producing strain ΔkstR1. Axenic growth of ΔkstR1 was compared with the wild-type strain and showed that, although the mutant grew at a similar rate to the wild-type, a slight increase in the lag phase was repeatedly observed (data not shown). In order to confirm that the phenotype was not caused by a second-site mutation, the experiment was repeated with an independently derived mutant, with similar results.

**Deletion of *kstR*<sub>Msm</sub> leads to upregulation of adjacent genes**

To examine whether *kstR*<sub>Msm</sub> controls the expression of adjacent genes, the expression levels of the *fadE34* orthologue (MSMEG_6041) and *otsB* (Fig. 1) were measured in both wild-type and ΔkstR1 strains using real-time quantitative polymerase chain reaction (RTq-PCR). There is a 3 bp gap between the end of *kstR*<sub>Msm</sub> and *otsB*, so these genes are likely to form an operon. The results (Fig. 2A) show that both MSMEG_6041 and *otsB* are upregulated in the mutant strain (36-fold and 10-fold respectively). The experiment was repeated with the independently generated mutant, and confirmed the upregulation of MSMEG_6041 and *otsB* in the mutant (data not shown). These observations suggest that *kstR*<sub>Msm</sub> acts as a repressor of transcription of both MSMEG_6041 and an operon consisting of itself and *otsB*.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd, *Molecular Microbiology*, 65, 684–699
KstR<sub>Mtb</sub> binds to a conserved motif within its own promoter region

TetR-like proteins normally bind to short palindromic DNA sequences (Grkovic <em>et al</em>, 1998; Orth <em>et al</em>, 2000; Ramos <em>et al</em>, 2005). Because protein binding constrains the evolution of these nucleotides, regulatory motifs may be identifiable through their conservation relative to neighbouring DNA sequences. We therefore aligned the intergenic region from <em>kstR</em><sub>Mtb</sub>/<em>fadE34</em> (Fig. 1) from <em>M. tuberculosis</em> with the orthologous regions from other species, and found that there is an 18 bp region that is very highly conserved (Fig. 3A). Examination of the sequence showed that it contains a 14 bp palindrome [TAGAAC(N2)GTTCTA]. The other conserved nucleotides match known mycobacterial -10 and -35 regions (Gomez and Smith, 2000). The binding motif is upstream of, but partially overlapping, the -10 region, and this would efficiently block binding of the RNA polymerase.

In order to determine whether KstR<sub>Mtb</sub> binds directly to the motif we had identified, the protein was expressed as a His<sub>6</sub>-tagged form and used in electrophoretic mobility shift assays (EMSAs). His<sub>6</sub>-KstR<sub>Mtb</sub> was purified by Ni<sup>2+</sup>-affinity chromatography, followed by size exclusion chromatography (SEC) to > 95% purity as judged by SDS-PAGE. The purified protein showed clear binding to the entire <em>kstR</em><sub>Mtb</sub>/<em>fadE34</em> intergenic region (318 bp), but not to a random piece of DNA of the same size (data not shown). Additionally, the purified protein showed binding to a 29 bp DNA probe (Table 2: Rv3573c/Rv3574 pair) containing the highly conserved palindromic region identified above. Figure 4A shows a clear retardation of the labelled 29 bp probe in the presence of increasing amounts of protein. This binding was lost with a 100-fold excess of unlabelled probe as a specific competitor, but a non-specific competitor did not abolish binding (Fig. 4B). These observations show that His<sub>6</sub>-KstR<sub>Mtb</sub> binds directly and specifically within its own promoter region to a short region containing a highly conserved palindrome.

KstR<sub>Mtb</sub> binds to the motif as a dimer

In order to study the binding stoichiometry of His<sub>6</sub>-KstR<sub>Mtb</sub> to the motif, the elution of the protein alone and in the presence of the 29 bp fragment was analysed by SEC and compared with a standard curve of <em>v</em> vs. log Mr (Fig. 4C). The molecular mass of His<sub>6</sub>-KstR<sub>Mtb</sub> was determined to be 60.2 kDa, which is consistent with the protein forming a dimer in solution (the predicted monomeric molecular mass is 27.7 kDa). The apparent molecular mass of the 29 bp fragment alone was determined to be 58.9 kDa; note that this substantially exceeds its actual mass of 18.0 kDa due to the inflexible rod structure of DNA in comparison with the globular shape of standard proteins (Reuter <em>et al</em>, 1998). Only one species of KstR<sub>Mtb</sub>–DNA complex with an apparent molecular mass of 118.7 kDa was detected at

---

**Fig. 2.** Changes in the expression levels of selected genes in the ΔkstR1 mutant compared with wild-type mc2155. The expression levels were measured in mid-log phase aerated cultures using RTq-PCR as described in Experimental procedures. The results are expressed relative to sigA, which was not significantly different in the mutant compared with the wild-type. Error bars represent ± 1 standard deviation. Filled bars, mc<sup>155</sup> wild-type; empty bars, ΔkstR1 mutant.

A. Expression levels of genes adjacent to <em>kstR</em>. Both of the <em>fadE34</em> orthologue (MSMEG_6041) and <em>otsB</em> (MSMEG_6043) are significantly upregulated (de-repressed) in the mutant compared with the wild-type (unpaired Student’s <em>t</em>-test; <i>P</i> ≤ 0.05).

B. Expression levels of genes flanking 11 of the predicted KstR motifs in <em>M. smegmatis</em>. All genes tested were significantly de-repressed in the mutant compared with the wild-type strain, with levels of de-repression ranging from 6-fold (MSMEG_5932) to 155-fold (MSMEG_6038) (unpaired Student’s <em>t</em>-test; <i>P</i> ≤ 0.05).
protein : DNA ratios of 1:4, 1:1 and 4:1. This is consistent with a complex of dimeric His\textsubscript{6}-KstRMtb bound to one 29 bp fragment of DNA. Dimeric binding to palindromic DNA is characteristic of the TetR family of transcriptional regulators (Huffman and Brennan, 2002). Although we cannot exclude the possibility that an alternative DNA conformation is assumed in the protein–DNA complex, altering its apparent mass, structural analyses with TetR show that the DNA is generally straight (Huffman and Brennan, 2002), and we conclude that a dimeric state is the most likely.

The motif is present in the upstream regions of other genes in both M. tuberculosis and M. smegmatis

The experiments described above show that His\textsubscript{6}-KstRMtb binds as a dimer to a 29 bp sequence within its own promoter that contains a highly conserved palindromic sequence overlapping a putative –10 region. This is consistent with it acting as a direct repressor of transcription, and indicates that the de-repression seen in the M. smegmatis ΔkstR1 strain is due to the loss of binding of KstRMsm to its own promoter. In order to identify whether

### Table 1. Bacterial strains and plasmids used in this study.

| Strain/plasmid | Genotype/description | Source |
|----------------|----------------------|--------|
| Strain         |                      |        |
| E. coli        | supE44 ΔlacU169 (elacZΔM15) hsdR17 recA1 | Invitrogen |
| BL21(DE3)      | OmpT hsdS\textsubscript{B} (rB – mB –) gal dcm (DE3) | Novagen |
| M. smegmatis   | High-frequency transformation mutant ATCC 607 | Snapper et al. (1990) |
| mc\textsuperscript{c}155 | ΔkstR\textsubscript{Msm} | This study |
| ΔkstR1         |                      |        |
| Plasmid        |                      |        |
| p2NIL          | Gene manipulation vector, Kan | Parish and Stoker (2000) |
| pGOAL19        | P\textsuperscript{ac}acl cassette vector, \textit{hyg} P\textsuperscript{ac}lacZ P\textsuperscript{ac}lacZ-sacB, E. coli expression vector, Kan | Parish and Stoker (2000) |
| pET30a         |                      | Novagen |
| pCS1           | 3.5 kb fragment containing kstR\textsubscript{Msm} in p2NIL | This study |
| pCS2           | 646 bp deletion of pCS1 | This study |
| pCS3           | pCS2 with the pGOAL19 P\textsuperscript{ac}acl cassette inserted | This study |
| pSK35          | kstR\textsubscript{Msm} in pET30a expression vector | This study |
The motif predicted are also regulated by KstR

Two approaches were used in order to obtain experimental evidence for the motif predictions in *M. smegmatis* and *M. tuberculosis*. First, RTq-PCR was used to measure the levels of expression of the flanking genes in the ΔkstR1 mutant and compare them with those in wild-type *M. smegmatis*. Second, EMSAs were used to demonstrate the binding of His_{6}-KstR_{Mtb} to the predicted *M. tuberculosis* motifs.

The levels of expression from 11 of the predicted motifs in *M. smegmatis* were measured. If the predicted motif is biologically relevant, then the flanking genes should be de-repressed in the ΔkstR1 mutant. RTq-PCR analysis showed that all genes tested were significantly de-repressed in the mutant compared with the wild-type strain, with levels of de-repression ranging from 6-fold (*MSMEG_5932*) to 155-fold (*MSMEG_6038*) (Fig. 2B). EMSAs were carried out to look for binding of His_{6}-KstR_{Mtb} to 13 of the predicted *M. tuberculosis* motifs, and binding was observed in 12 of these (Table 3).

Microarray analysis indicates that a large number of genes are de-repressed in the ΔkstR1 mutant

In order to obtain a genome-wide picture of genes controlled by kstR, we carried out competitive hybridizations between cDNA from wild-type *M. smegmatis* and the mutant strain ΔkstR1 using *M. smegmatis* microarrays. The full results of the microarray analysis are given in Table S1. Using a *P*-value of 0.05 corrected for multiple testing, a total of 132 genes were significantly upregulated (6- to 1771-fold), and 27 were downregulated (6- to 18-fold).

The microarray analysis showed de-repression of genes flanking 26 of the 31 motifs that we had identified in *M. smegmatis* (Table 4). For the other five, although the computational analysis indicates the presence of a motif, a combination of low levels of de-repression, low levels of significance in terms of gene expression changes, and the absence of an orthologous gene with a motif in *M. tuberculosis* suggests that these instances of the motif may not be biologically relevant.
We identified four additional instances of the motif (MSMEG_0217, MSMEG_1410, MSMEG_3658 and MSMEG_5940) that had not been picked up in the original search (Table 4). Three of these were found to overlap with coding sequences of adjacent genes and one (MSMEG_1410) was within an operon, and therefore would have been excluded from our original search. We also searched M. tuberculosis in regions where motifs were present in M. smegmatis, and found possible matches for two of these (Rv3501c and Rv3536c) but with a relatively low probability as determined by MAST.

### Defining the kstR regulon in M. smegmatis

The genes with altered expression in the microarray analysis will be a combination of those where binding of KstR directly affects transcription (the kstR regulon), and those that are secondary effects. The genes in the kstR regulon were defined by using a combination of data from the motif search (Table 3), EMSA analyses (Table 3), RTq-PCR analyses (Fig. 2A and B) and genome-wide expression data from the microarray studies (looking both at fold-change and P-value; Table S1), comparative genomics with M. tuberculosis, and examination of operon organi-
Table 3. Intergenic sequences in M. smegmatis and M. tuberculosis with significant matches to the palindromic KstR motif.

| Motif sequence | P-value | Flanking genes | Experimental evidence |
|----------------|---------|----------------|-----------------------|
| **Instances in M. smegmatis** | | | RTq-PCR |
| ATGGAAACCTTTGATTC | 1.5e-09 | MSMEG_0305/MSMEG_0306 | ND |
| ACAGAAATGCCTAGCTTG | 2.06E-07 | MSMEG_0309 | ND |
| ACCTAGAAACTTTGAGAAA | 2.19E-09 | MSMEG_1098 | +ve |
| ACAGAAACCTTTGATTTG | 4.93E-08 | MSMEG_2645 | +ve |
| ACCTGGACGTCGCCGATAC | 5.40E-07 | MSMEG_2761/MSMEG_2763 | ND |
| ATTAGAACCTTTGAGCG | 1.65E-07 | MSMEG_2790 | ND |
| ATTAGAACCTTTGAGTTT | 9.65e-08 | MSMEG_3519 | ND |
| ACTGCAACCTTTGAGTTT | 1.32E-07 | MSMEG_3521 | ND |
| ATTAGAACCTTTGAGTTG | 1.24E-10 | MSMEG_3522 | ND |
| ACCTGGAAACCTTTGAGTTA | 4.14E-08 | MSMEG_3523 | ND |
| ACCTGGAAACCTTTGAGGA | 2.57E-07 | MSMEG_5228 | ND |
| ATTAGAACCTTTGAGCTCA | 1.04E-07 | MSMEG_5229 | ND |
| ACTGAAACGTTTACGTTCG | 6.96E-08 | MSMEG_5230 | ND |
| ATTAGAACCTTTGAGCG | 2.89E-07 | MSMEG_5231 | ND |
| ACTGCAACCTTTGAGTTT | 1.04E-07 | MSMEG_5232 | ND |
| ACCTGGAAACCTTTGAGTTG | 8.57E-09 | MSMEG_5233 | ND |
| ATTAGAACCTTTGAGTTT | 3.45E-08 | MSMEG_5234 | ND |
| ACTGCAACCTTTGAGTTG | 1.24E-09 | MSMEG_5235 | ND |

| **Instances in M. tuberculosis** | | | EMSA |
| ACAGAAACTCGTTTCA | 2.22E-07 | Rv0223c | -ve |
| ATTAGAACCTTTGAGCTTA | 7.54E-08 | Rv0687 | ND |
| ATTAGAACCTTTGAGCTTA | 7.20E-11 | Rv0940c | +ve |
| ATTAGAACCTTTGAGCTTA | 2.17E-08 | Rv0953c | ND |
| ACCTGGAAACTTTGAGTTT | 1.32E-07 | Rv1628c/Rv1629 | ND |
| ACCTGGAAACTTTGAGTTT | 6.83E-09 | Rv1894c/Rv1895 | +ve |
| ACCTGGAAACTTTGAGTTT | 8.57E-09 | Rv3050c/Rv3051 | +ve |
| ACCTGGAAACTTTGAGTTT | 3.45E-08 | Rv3515c/Rv3516 | +ve |
| ACCTGGAAACTTTGAGTTT | 3.40E-07 | Rv3520c/Rv3521 | +ve |
| ACCTGGAAACTTTGAGTTT | 9.47E-09 | Rv3525c/Rv3526 | ND |
| ATTAGAACCTTTGAGCTTA | 1.79E-08 | Rv3531c/Rv3532 | +ve |
| ATTAGAACCTTTGAGCTTA | 1.42E-07 | Rv3545c/Rv3546 | +ve |
| ATTAGAACCTTTGAGCTTA | 1.99E-08 | Rv3570c/Rv3571 | +ve |
| ATTAGAACCTTTGAGCTTA | 2.35E-11 | Rv3573c/Rv3574 | +ve |

a–s. Orthologous genes, e.g. MSMEG_0309, is an orthologue of Rv0223c.
ND means not determined whereas –ve means no binding was observed.

Transcriptional changes not associated with a KstR motif

We also analysed the genes where expression changes were not associated with a KstR motif, which are likely to be secondary effects. We examined fold-change, P-value and genomic organization, and took runs of modulated genes.
| Gene     | Fold-change | P-value | Gene     | Fold-change | P-value |
|----------|-------------|---------|----------|-------------|---------|
| MSMEG_0217 | 197.3       | 1.5E-03 | Rv0162c  | 8.5         | 2.8E-01 |
| MSMEG_0302 | 54.5        | 2.7E-03 | Rv1426c  | 61.7        | 1.6E-03 |
| MSMEG_0304 | 114.4       | 5.8E-03 | Rv1427c  | 3.9         | 1.4E-01 |
| MSMEG_0305 | 71.4        | 1.5E-03 | Rv1428c  | ND          | ND      |
| MSMEG_0309 | 82.8        | 1.3E-03 | Rv0223c  | ND          | ND      |
| MSMEG_1098 | 110.9       | 1.3E-03 | Rv0551c  | ND          | ND      |
| MSMEG_1410 | 223.7       | 1.3E-03 | Rv0687   | ND          | ND      |
| MSMEG_2644 | 2.0         | 2.9E-01 | Rv2800   | ND          | ND      |
| MSMEG_2645 | 24.6        | 4.8E-03 | Rv2799   | ND          | ND      |
| MSMEG_2789 | 3.0         | 1.4E-01 | Rv2699   | ND          | ND      |
| MSMEG_2790 | 7.8         | 4.0E-02 | Rv2668   | ND          | ND      |
| MSMEG_3515 | 62.0        | 1.6E-03 | ND       | ND          | ND      |
| MSMEG_3516 | 22.7        | 5.3E-03 | ND       | ND          | ND      |
| MSMEG_3519 | 168.0       | 1.3E-03 | Rv1894c  | ND          | ND      |
| MSMEG_3658 | 67.1        | 1.5E-03 | ND       | ND          | ND      |
| MSMEG_3843 | 100.8       | 1.3E-03 | Rv1628c  | ND          | ND      |
| MSMEG_3844 | 205.1       | 1.3E-03 | Rv1627c  | ND          | ND      |
| MSMEG_5202 | 36.4        | 3.3E-03 | Rv1132   | ND          | ND      |
| MSMEG_5228 | 100.8       | 1.3E-03 | Rv1106c  | ND          | ND      |
| MSMEG_5266 | 10.2        | 2.2E-02 | Rv1059   | ND          | ND      |
| MSMEG_5519 | 1.8         | 3.7E-01 | ND       | ND          | ND      |
| MSMEG_5520 | 4.4         | 8.2E-02 | ND       | ND          | ND      |
| MSMEG_5544 | 82.7        | 1.3E-01 | Rv0927c  | ND          | ND      |
| MSMEG_5584 | 250.7       | 1.3E-01 | Rv0926c  | ND          | ND      |
| MSMEG_5586 | 11.7        | 2.0E-02 | Rv3492c  | ND          | ND      |
| MSMEG_5589 | 11.7        | 2.4E-02 | Rv3493c  | ND          | ND      |
| MSMEG_5594 | 22.5        | 3.5E-03 | Rv3495c  | ND          | ND      |
| MSMEG_5595 | 12.1        | 2.5E-02 | Rv3496c  | ND          | ND      |
| MSMEG_5596 | 20.1        | 7.0E-03 | Rv3498c  | ND          | ND      |
| MSMEG_5599 | 27.1        | 4.7E-03 | Rv3500c  | ND          | ND      |
| MSMEG_5502 | 32.4        | 4.1E-03 | Rv3501c  | ND          | ND      |
| MSMEG_5503 | 74.7        | 1.5E-03 | Rv3502c  | ND          | ND      |
| MSMEG_5504 | 72.9        | 1.5E-03 | Rv3503c  | ND          | ND      |
| MSMEG_5506 | 113.4       | 1.3E-03 | Rv3504   | ND          | ND      |
| MSMEG_5507 | 84.9        | 1.3E-03 | Rv3505   | ND          | ND      |
| MSMEG_5508 | 59.2        | 1.7E-03 | Rv3506   | ND          | ND      |
| MSMEG_5509 | 22.4        | 5.8E-03 | Rv3511   | ND          | ND      |
| MSMEG_5913 | 138.6       | 1.3E-03 | Rv3515c  | ND          | ND      |
| MSMEG_5914 | 175.4       | 1.3E-03 | Rv3516   | ND          | ND      |
| MSMEG_5915 | 251.9       | 1.7E-03 | Rv3516   | ND          | ND      |
| MSMEG_5918 | 2.8         | 1.8E-01 | Rv3518c  | ND          | ND      |
| MSMEG_5919 | 19.4        | 1.6E-02 | Rv3519   | ND          | ND      |
| MSMEG_5920 | 38.5        | 4.2E-03 | Rv3520c  | ND          | ND      |
| MSMEG_5921 | 95.5        | 1.3E-03 | Rv3522   | ND          | ND      |
| MSMEG_5922 | 122.0       | 1.3E-03 | Rv3523   | ND          | ND      |
| MSMEG_5923 | 60.3        | 1.6E-03 | Rv3526   | ND          | ND      |
| MSMEG_5925 | 61.7        | 1.6E-03 | Rv3527   | ND          | ND      |
| MSMEG_5930 | 3.9         | 1.4E-01 | Rv3529c  | ND          | ND      |
| MSMEG_5931 | 8.5         | 2.8E-01 | Rv3530c  | ND          | ND      |
| MSMEG_5932 | 13.3        | 1.7E-01 | Rv3531c  | ND          | ND      |
Table 4. cont.

| M. smegmatis  | Motif  | Fold-change | P-value | M. tuberculosis | c | p | m | ø | e | Motif   | Gene Name | Function |
|--------------|--------|-------------|---------|-----------------|---|---|---|---|---|---------|-----------|----------|
| ↑ MSMEG_5937 | 47.1   | 2.3E-03     |         | Rv3534c         | hsaF  |   |   |   |   | 4-hydroxy-2-oxovalerate aldolase |
| ↑ MSMEG_5939 | 55.3   | 2.8E-03     |         | Rv3535c         | hsaG  |   |   |   |   | Acetaldehyde dehydrogenase |
| ↑ MSMEG_5940 | 125.9  | 2.7E-03     |         | Rv3536c         | hsaE  |   |   |   |   | 2-hydroxypentadecanoate |
| ↓ MSMEG_5941 | 85.5   | 1.3E-03     |         | Rv3537          | kstD  |   |   |   |   | 3-ketosteroid Δ1-dehydrogenase |
| ↓ MSMEG_5943 | 88.9   | 1.3E-03     |         | Rv3538          | hsd4  |   |   |   |   | 2-enoyl acyl-CoA hydratase |
| ↑ MSMEG_5990 | ND     | ND          |         | Rv3540c         | itp2  |   |   |   |   | Branched-chain 3-ketoacyl-CoA thiolase |
| ↑ MSMEG_5991 | 242.3  | 2.4E-03     |         | Rv3541c         | –    |   |   |   |   | CHP |
| ↑ MSMEG_5992 | 1771.8 | 1.2E-03     |         | Rv3542c         | fadE29 |   |   |   |   | Acyl-CoA dehydrogenase |
| ↑ MSMEG_5993 | 309.2  | 2.4E-03     |         | Rv3543c         | fadE28 |   |   |   |   | Short/branched chain acyl-CoA dehydrogenase |
| ↑ MSMEG_5994 | 138.3  | 2.4E-03     |         | Rv3544c         | –    |   |   |   |   | CHP |
| ↓ MSMEG_5995 | 170.2  | 2.4E-03     |         | Rv3545c         | cyp125 |   |   |   |   | Cytochrome P450 125 |
| ↓ MSMEG_5996 | 160.2  | 2.4E-03     |         | Rv3546          | fadA5  |   |   |   |   | Acetyl CoA acetyltransferase |
| ↓ MSMEG_5997 | 147.6  | 2.4E-03     |         | Rv3547          | –    |   |   |   |   | CysQ family |
| ↓ MSMEG_5998 | 115.7  | 2.4E-03     |         | Rv3547          | –    |   |   |   |   | CHP |
| ↑ MSMEG_6033 | 23.8   | 9.6E-03     |         | –              | –    |   |   |   |   | HP |
| ↑ MSMEG_6035 | 114.2  | 1.3E-03     |         | Rv3567c         | hsaB  |   |   |   |   | 3-HSA hydroxylase, reductase |
| ↑ MSMEG_6036 | 121.3  | 1.3E-03     |         | Rv3568c         | hsaC  |   |   |   |   | 3,4-DHSA dioxygenase |
| ↑ MSMEG_6037 | 315.4  | 1.7E-03     |         | Rv3569c         | hsaD  |   |   |   |   | 4,9-DHSA hydrolase |
| ↑ MSMEG_6038 | 136.5  | 1.7E-03     |         | Rv3570c         | hsaE  |   |   |   |   | 3-HSA hydroxylase, oxygenase |
| ↓ MSMEG_6039 | 66.7   | 1.6E-03     |         | Rv3571          | kshB  |   |   |   |   | Ketosteroid 9α-hydroxylase, reductase |
| ↓ MSMEG_6040 | 30.2   | 3.4E-03     |         | Rv3572          | –    |   |   |   |   | CHP |
| ↑ MSMEG_6041 | 93.2   | 1.3E-03     |         | Rv3573c         | fadE34 |   |   |   |   | Acyl-CoA dehydrogenase |
| ↓ MSMEG_6042 | 12.5   | 1.6E-02     |         | Rv3574          | kstR  |   |   |   |   | TetR regulator |
| ↓ MSMEG_6043 | 60.1   | 1.3E-03     |         | Rv3576          | –    |   |   |   |   | Trehalose phosphatase |
| ↑ MSMEG_6047 | 73.5   | 1.5E-03     |         | Rv0139          | –    |   |   |   |   | Oxidoreductase |
| ↑ MSMEG_6047 | 92.3   | 1.3E-03     |         | Rv0138          | –    |   |   |   |   | CHP |

In the shaded columns, dark shading indicates: c, induced by growth on cholesterol in Rhodococcus sp. strain RHA1 (Van der Geize et al., 2007); p, induced in palmitic acid at least 1.5 fold (Schnappinger et al., 2003); m, induced in macrophages (Schnappinger et al., 2003); ø, essential for survival in macrophages (Rengarajan et al., 2005; Rosas-Magallanes et al., 2007); e, essential for survival in mice (Chang et al., 2007; Sassetti and Rubin, 2003).

a. Arrows represent gene direction in relation to the rest of the genome. Sequential runs of M. smegmatis genes are grouped together by horizontal lines. In the most recent annotation of the M. smegmatis genome, the gene numbering was not sequential in a number of cases e.g. with MSMEG_0302/MSMEG_0303/MSMEG_0304 where the genes are adjacent to each other but are not numbered as such.

b. Bullets indicate the presence of a motif; two bullets indicate that there are two motifs. Where divergent genes share a common intergenic region, the motifs have been indicate for the first gene only, although they could function for either or both genes.

c. ND No data.

d. Motifs that were not originally predicted. The M. tuberculosis mce4 motif is detectable but with low significance.

e. MSMEG_5905 is an annotated ORF which would break up this operon, but is only 31 amino acids long, and its location conflicts with MSMEG_5906, so may be a mis-annotation.

f. Recently assigned to cholesterol degradation pathway and renamed (Van der Geize et al., 2007).

g. HP: hypothetical protein; CHP, conserved hypothetical protein.

into account and included 99 genes in this group. Most of these (87) were upregulated in the mutant (2- to 100-fold) and 11 were downregulated (5- to 16-fold) (Table S1); 74 of them (all but two induced in the mutant) lie in putative operons. The co-regulation of adjacent genes is particularly strong evidence that the effect is a genuine indirect effect of the kstR deletion, rather than being due to the noise inherent in microarray experiments.

Predicting the kstR regulon in M. tuberculosis

There are clear orthologues of most of the genes in the kstR<sub>main</sub> regulon in M. tuberculosis (Table 4), and these include all the genes in M. tuberculosis that were predicted to lie downstream of a motif (Table 3). The presence of the motif and the orthology with the de-repressed M. smegmatis gene is robust evidence for inclusion of these genes in the M. tuberculosis kstR regulon. The most striking observation from the data is that there is a large region in both mycobacterial genomes [Rv3492c to Rv3574 (kstR) and MSMEG_5893 to MSMEG_6043] that contains a number of operons that are de-repressed and associated with a motif.

Functional analysis of the kstR regulon

Analysis of the functions of the genes in the kstR regulon was carried out by a combination of BLAST analyses, as well as searching the Tuberculist database (http://genolist.pasteur.fr/Tuberculist/) and the literature. It was...
During infection, cell wall lipids play a variety of roles in M. tuberculosis in vivo. In addition to using fatty acids as a carbon source (Bishai, 2000), M. tuberculosis in vivo are thought to be particularly important for the survival of bacteria and other prokaryotes are able to use fatty acids as a sole carbon source via a pathway, mostly through bioinformatics, but also with experimental verification of some candidate genes. Although they named Rs574 kstR (for ketosteroid regulator), no experimental evidence was provided as to the role of this gene. Our results show that most of these genes are indeed controlled by kstR and have the kstR motif in their promoter regions. We have confirmed that the binding motif identified here is also present in appropriate sites in the Rhodococcus genome (data not shown).

Despite the proposed involvement of the rhodococcal kstR regulon in cholesterol degradation, we suggest that the situation is more complex in mycobacteria. First, we have demonstrated that the regulon is extremely large (83 genes in M. smegmatis), and it is unlikely that so many genes are required just for cholesterol utilization. Second, palmitic acid also induces 22 genes (including kstR itself) in the kstR regulon in M. tuberculosis (Table 4). We propose therefore that the kstR regulon is involved in the uptake and utilization of a variety of lipids, of which cholesterol is just one. It can be argued that it makes biological sense for the bacteria to have a mechanism that will enable degradation of a variety of lipids. We have elected to retain the name kstR because it is relevant to at least
part of the function, and in order to reduce confusion. In *Rhodococcus*, all 51 genes in the region orthologous to *Rv3492c–Rv3574* were upregulated in the presence of cholesterol, whereas we found that not all of these were induced in the *M. smegmatis kstR* mutant (Table 4). This suggests that part of the cholesterol response is under the control of regulators other than *kstR*. The induction ratios seen in the presence of cholesterol and palmitate are lower than we observed in this study, and this may reflect low intracellular concentrations of the molecules, or that they de-repress the regulon with different affinities.

It is noteworthy that 18 of the genes in the *kstR* regulon have been shown to be essential *in vivo* in mouse or macrophage models (Table 4). These include *kstR* itself, and some of the *mce4* operon genes. Many of these essentiality studies used the genome-wide TraSH screen, where methodological and statistical noise causes some error. However, the TraSH methodology has been shown to be reasonably robust through validation of individual genes. There is already good evidence that the *mce4* operon is required *in vivo* for survival in mice (Joshi et al., 2006) and macrophages (Rosas-Magallanes et al., 2007). In addition, deletion of the *Rv3540c–Rv3545c* operon causes attenuation of growth in macrophages and immunocompetent mice (Chang et al., 2007). Presumably the reason for the essentiality for *kstR* (where a mutant will express all the normally regulated genes constitutively) differs from the other genes (where a mutation results in loss of function). The induction levels we saw in the *kstR* regulon were extremely high, so the essentiality of *kstR* may merely reflect the energy cost of the elevated gene expression; alternatively, there may be times in the infection process where expression of a particular gene in the regulon is detrimental for another reason.

We identified 99 genes that were induced or repressed in the mutant but did not appear to be directly regulated by *kstR* (Table S4). Some of these are involved in lipid metabolism, suggesting the involvement of other regulatory systems, but many were ribosomal and chaperone genes. The induction of ribosomal and chaperone genes suggests that the transcription levels achieved by knocking out *kstR* place a strain on the translation apparatus of the cell. This may explain the slight growth defect seen in the mutant. It is possible that this situation does not occur in reality, and the transcription levels achieved by de-repression in the presence of an inducer will not be as high, so that less stress will be put on the translational apparatus.

The *mce4* operon appears to be a key part of the *kstR* regulon in *M. smegmatis*. Circumstantial data are accumulating that the *mce* operons (of which *M. tuberculosis* has four, and *M. smegmatis* at least five) function as lipid transport systems (Santangelo et al., 2002; Mitra et al., 2005; Uchida et al., 2007; Van der Geize et al., 2007). The results presented here show that the *mce4* operon is co-regulated with other genes involved in fatty acid metabolism, and support the hypothesis that the *mce* genes are involved in lipid uptake. Apart from its use as an energy source, cholesterol has been implicated in the uptake of *M. tuberculosis* by macrophages (Gatfield and Pieters, 2000), although the receptor for host cholesterol is unknown. It is tempting to suggest that the *mce4* system might play a role in the bacterial–host interaction, if it is also involved in internalizing cholesterol (Arruda et al., 1993; Casali et al., 2002; Mitra et al., 2005).

*KstR* is a TetR-type regulator; in this paradigm, repression is controlled by the binding of an inducer molecule. TetR itself binds tetracycline (Ramos et al., 2005), and ligands for other repressors are often hydrophobic molecules (Frenois et al., 2004). The induction of the *kstR* regulon by palmitate and cholesterol supports the hypothesis for a fatty acid ligand. Additionally, the induction of the regulon upon entry into the macrophage, and the essentiality of many of the genes in the regulon for *in vivo* survival (Table 4), suggests that the ligand(s) are present inside the host.

While it is likely that the *kstR* regulon has a major catabolic role, it is possible that some of the genes in the regulon are anabolic, although we did not see differences in quantity and abundance of the major cell wall lipids (data not shown). One gene that is present in the *kstR* regulon of *M. tuberculosis* but not that of *M. smegmatis* is the *nat* gene encoding arylamine N-acetyltransferase (Anderton et al., 2006). Mutants lacking *nat* are defective in mycolic acid synthesis (Bhakta et al., 2004), indicating a possible anabolic role for some genes in the *kstR* regulon. The Nat protein can bind to the antitubercular drug isoniazid, reducing its efficacy (Sandy et al., 2002). The induction of the *kstR* regulon in *M. tuberculosis in vivo* may therefore partially affect the antibiotic resistance of the bacteria.

In conclusion, we have described a large regulon within the mycobacteria. In *M. tuberculosis*, this makes up almost 2% of the genome. Although at least the core of this regulon is highly conserved in non-pathogens, many of the genes are critical in the pathogenesis of *M. tuberculosis*. Investigating both the regulation of *kstR* and the functions of the genes in the regulon is likely to provide important new information in our understanding of the adaptation of this major pathogen to its host.

**Experimental procedures**

**Bacterial strains and culture conditions**

Cultures of *M. smegmatis* mc²155 were grown at 37°C with shaking in Middlebrook 7H9 broth (Difco) containing 10% oleic acid-albumin-dextrose-catalase supplement (Becton Dickinson) and 0.05% Tween 80. Hygromycin (50 μg ml⁻¹).
kanamycin (20 µg µl⁻¹), 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (Xgal, 50 µg µl⁻¹) and sucrose (2% w/v) were used for selection as appropriate. *E. coli* DH5α was used as a host for cloning, and *E. coli* BL21(DE3) (Novagen) was used as a host for expression of recombinant KstR₄₄. Both *E. coli* strains were grown in Luria–Bertani, and kanamycin (50 µg µl⁻¹) was used for plasmid selection and maintenance. The strains and plasmids used in this study are described in Table 1.

**Deletion of kstR₄₄ by homologous recombination**

A 646 bp deletion in *MSMEG_6042* (kstR₄₄) was made in *M. smegmatis* mc²155 by homologous recombination (Parish and Stoker, 2000). Briefly, a 3.5 kb fragment containing the entire kstR₄₄ gene and flanking regions was PCR amplified from mc²155 genomic DNA using ΔkstR₄₄ forward and reverse primers (Table 2). The primers had BamHI–HindIII sites (shown in upper case in Table 2) introduced into them in order to enable cloning of the 3.5 kb fragment into p2NIL, resulting in plasmid pCS1. A deletion was made in pCS1 by inverse PCR using inv_kstR₄₄ forward and reverse primers (Table 2) and religation of the BgIII-digested PCR fragment. One of the BgIII sites was present in the genome, and the other was introduced in the inv_kstR₄₄ reverse primer. During the writing of this manuscript, the *M. smegmatis* genome was re-annotated and kstR₄₄ was designated as being 66 bp shorter than annotated previously. These primers were designed to remove 646 bp from the coding sequence. The deletion removes 39 bp upstream of the coding sequence according to the new annotation. Neither annotation has been confirmed experimentally. The deletion in the resulting plasmid pCS2 was confirmed by sequencing (sequencing reactions performed by MWG Biotech) across the junction (data not shown). Finally, the PacI cassette was inserted into pCS2, resulting in the suicide delivery vector pCS3.

pCS3 was electroporated into competent mc²155 (Parish and Stoker, 1998), and single cross-overs were selected for on medium containing hygromycin, kanamycin and Xgal. A single blue kanamycin and hygromycin-resistant colony was streaked onto fresh media without any selective markers, and incubated at 37°C for 3–5 days to allow the second cross-over to occur. Serial dilutions were plated onto media containing sucrose and Xgal to select for double cross-overs. Potential double cross-overs (white sucrose-resistant colonies) were screened for kanamycin sensitivity and confirmed by colony PCR. The resulting mutant was called ΔkstR₁. The intergenic region between fadE₃⁴ and kstR was sequenced in order to confirm that the promoter had not been affected by the mutagenesis.

**RNA extraction**

RNA for microarray analysis and RTq-PCR was extracted from both wild-type mc²155 and ΔkstR₁ strains by direct sampling into guanidinium thiocyanate (GTC). Briefly, 10 ml of aerated cultures in logarithmic phase (OD₆₀₀ 0.4–0.5) was added to 40 ml of 5 M GTC to prevent further transcription. The culture was pelleted by centrifugation (20 min, 4000 g, 4°C) and resuspended in 200 µl of water. The cultures were transferred to screwcap tubes containing 0.5 ml of 0.1 mm zirconia/silica beads (Biospec), and 700 µl of buffer RLT (Qiagen) was added. The bacteria were lysed using a MiniBeadBater (BioSpec), and cell lysates were recovered by centrifugation (5 min, 13 000 g, 4°C). RNA was purified from the lysate using an RNeasy kit (Qiagen) and treated with DNase (Qiagen) according to the manufacturer’s instructions. Finally, the samples were eluted in 30 µl of RNase-free water, and quantity was assessed using a NanoDrop (NanoDrop technologies).

**Reverse transcription reactions for RTq-PCR**

Real-time quantitative polymerase chain reaction was used for the analysis of the expression of single genes. Prior to reverse transcription, RNA was treated with DNase (Invitrogen) for 30 min at 37°C, followed by heat inactivation. Reverse transcription took place in a total volume of 20 µl containing 100 ng total RNA, 300 ng random primers (Invitrogen), 10 mM DTT, 0.5 mM each of dCTP, dATP, dGTP and dTTP, and 200 units of Superscript III (Invitrogen). For primer annealing, RNA and random primers were heated to 65°C for 10 min in a volume of 13 µl and then snap-cooled on ice prior to the addition of the remaining components. For reverse transcription, the reactions were incubated at 55°C for 50 min. A total of 1 µl (equivalent to 5 ng of RNA) of cDNA was used in the RTq-PCRs.

**Real-time quantitative polymerase chain reaction**

Real-time quantitative polymerase chain reactions were set up using the DyNAmo SYBR Green qPCR kit (MJ Research), and RTq-PCR was performed using the DNA Engine Opticon² System (GRI). 20 µl reactions were set up on ice containing 1× DNA Master SYBR Green I mix, 1 µl of cDNA product and 0.3 µM of each primer. Sequences of each primer are given in Table 2. Reactions were heated to 95°C for 10 min before cycling for 35 cycles of 95°C for 30 s, 62°C for 20 s, and 72°C for 20 s. Fluorescence was captured at the end of each cycle after heating to 80°C to ensure the denaturation of primer-dimers. At the end of the PCR, melting curve analysis was performed and PCR products were analysed on an agarose gel to ensure product specificity. The experiment was performed in triplicate and each gene was measured in duplicate, giving a total of six data points per gene.

**Expression and purification of recombinant KstR₄₄**

The kstR₄₄ gene was PCR amplified from *M. tuberculosis* H₃₇Rv genomic DNA using pET_kstR₄₄ forward and reverse primers (Table 2). These primers had Ncol–HindIII sites introduced into them to allow for cloning into the pET30a expression vector. The nucleotide sequences corresponding to the restriction sites are shown in upper case in Table 2, and the start site of kstR₄₄ is underlined (in accordance with the old annotation). The resulting plasmid, pSK35, was sequence verified and used for expression and purification of C-terminally His-tagged KstR₄₄. For expression, *E. coli* BL21(DE3) cultures containing plasmid pSK35 were grown at
37°C until mid-logarithmic phase. Cultures were induced with 1 mM IPTG (isopropyl-beta-D-thiogalactopyranoside) for 2 h at 37°C and harvested by centrifugation (10 min, 4000 g, 4°C). The cell pellet was resuspended in 5 ml of lysis buffer (20 mM HEPES pH 8.0, 150 mM NaCl, 1 mM β-mercaptoethanol, 10 mM imidazole) and lysed by passage through a cell disrupter (Constant Systems) set at 18 kpsi. The lysate was centrifuged (25 min, 16 000 g, 4°C) and His6-KstRm from the soluble fraction was purified by immobilized metal ion affinity chromatography using a HiTrap Ni-NTA column (GE Healthcare Biosciences), followed by SEC using a Superdex200 10/30 column (GE Healthcare Biosciences).

Electrophoretic mobility shift assays

Oligonucleotides (Table 2) were annealed by heating to 95°C for 10 min and allowed to cool slowly to room temperature. The resulting probes were end-labelled with DIG-11-ddUTP using the DIG gel shift kit, 2nd generation (Roche), according to the manufacturer's instructions. For the binding reaction, varying amounts of purified His6-KstRm were incubated with 0.66 pmol of labelled fragment in binding buffer [20 mM HEPES pH 8.0, 75 mM NaCl, 10 mM MgCl2, 0.1 μg of poly-L-lysine, 1 μg of poly(dl-dC)]. Specific and non-specific competitors were added for the control reactions. Specific competition reaction mixtures contained a 100-fold excess of unlabelled probe, and non-specific competition reaction mixtures contained a 150-fold excess of poly(dl-dC). Incubations were carried out for 30 min at room temperature, and reaction mixtures were loaded onto 8% polyacrylamide gels containing 0.5× TBE. Gels were run, with cooling at 80–100 V over 1.5–2 h. The DNA–protein complexes were contact blotted onto positively charged Hybond-N nylon membranes (Amer sham), and detected by anti-DIG-alkaline phosphatase and the chemiluminescent substrate CSPD as described by the manufacturer (Roche). Membranes were exposed to X-ray film at room temperature for 10–30 min.

Molecular weight determination of the protein–DNA complex by SEC

The molecular weight of His6-KstRm was determined by analytical SEC on a Superdex200 10/30 column. A standard curve of n, n, of the following standards: ovalbumin (43.0 kDa), ribonuclease A (13.7 kDa), albumin (67.0 kDa), chymotrypsinogen A (25.0 kDa) and catalase (232.0 kDa). The void volume (v0) of the column was determined with blue Dextran 2000. All SEC experiments were performed at a flow rate of 0.5 ml min⁻¹ in 20 mM HEPES pH 8.0, 75 mM NaCl, 10 mM MgCl2, and 1 mM β-mercaptoethanol. His6-KstRm was used at a concentration of 15 μM. Samples containing His6-KstRm and the 29 bp annealed probes (Table 2) were incubated on ice for 15 min prior to analysis. Collected fractions were analysed by SDS-PAGE and stained with Coomassie blue and ethidium bromide to confirm the presence of protein and DNA.

Microarray analysis of M. smegmatis ΔkstR1

Microarrays for genome-wide expression analysis of the mutant strain ΔkstR1 were obtained from the Pathogen Functional Genomics Resource Centre at TIGR (http://pfgrc.tigr.org/). The arrays consist of 6746 different 70-mer single-stranded oligonucleotides spotted onto glass slides. The oligonucleotides represent the entire M. smegmatis genome, and each oligonucleotide is spotted four times. Wild-type RNA was competitively hybridised against mutant RNA, and the design included a dye-swap. For the labelling reactions, 2–10 μg of RNA was labelled with either Cy3-dCTP or Cy5-dCTP (Amersham Pharmacia Biotech). In each case, 3 μg of random primers (Invitrogen™ Life Technologies) was annealed to the RNA by heating to 95°C for 5 min, followed by snap-cooling on ice. The labelling reaction contained 0.5 mM each of dATP, dGTP and dTTP, 10 mM DTT, 60 μmol of Cy3-dCTP (or Cy5-dCTP) and 500 units of Superscript II (Invitrogen™ Life Technologies) in a final volume of 25 μl. The samples were incubated in for 10 min at 25°C, followed by a 90 min incubation at 42°C in the dark.

The slides were prehybridised by incubating in prehybridisation buffer (3.5× SSC, 0.1% SDS, 10 mg ml⁻¹ BSA) at 65°C for 20 min. They were then washed in 400 ml of water, followed by 400 ml of isopropanol, for 1 min each. The slides were dried by centrifugation (1500 g, 5 min, room temperature) and stored in the dark until hybridization (< 1 h).

Microarray hybridisations

Labelled wild-type samples were combined with the corresponding labelled mutant samples, and were purified using a MinElute PCR Purification Kit from Qiagen. Samples were eluted in 25 μl of water and hybridised onto the array in hybridisation buffer (4× SSC, 40% formamide, 0.1% SDS). The samples were denatured by heating to 95°C for 2 min before being added to the array. Hybridization took place under a glass coverslips in a humidified slide chamber (Corning) submerged in a 65°C water bath for approximately 16 h. Coverslips were removed in wash buffer I (1× SSC, 0.05% SDS) prewarmed to 65°C, and slides were washed sequentially in buffer I at 65°C for 2 min, followed by two washes in buffer II (0.06× SSC) at room temperature for 2 min each. Slides were dried by centrifugation (1500 g, 5 min, room temperature), and were scanned using an Affymetrix 418 scanner. The image files were quantified using ImageGene 7.0 software (BioDiscovery). The whole experiment was performed in duplicate, and two arrays were used per experiment. As the oligonucleotides were spotted four times on the slides, this gave us a total of eight data points per open reading frame (ORF).

Microarray data analysis

Data analysis was performed using functions from the limma (linear models for microarray data analysis) (Smyth, 2005) (http://bioinf.wehi.edu.au/limma/) and yasma (Wernisch et al., 2003) software packages. Differentially expressed genes were identified by linear models using an experimental design for two-colour arrays which incorporated biological replicates with dye-swapped technical replicates. Data for control spots, and for spots with expression levels in the lower 10% quantile, were discarded. This was followed by background correction
Transcriptional repressor controlling a large lipid metabolism regulon in mycobacteria

Bailey, T.L., and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2: 28–36.

Bailey, T.L., and Gribskov, M. (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* 14: 48–54.

Baulard, A.R., Betts, J.C., Engoang-Ndong, J., Quan, S., McAdam, R.A., Brennan, P.J., et al. (2000) Activation of the pro-drug ethionamide is regulated in mycobacteria. *J Biol Chem* 275: 28326–28331.

Benjamini, Y., and Hochberg, Y. (1995) Controlling false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B* 75: 289–300.

Bhakta, S., Besra, G.S., Upton, A.M., Parish, T., Sholto-Douglas-Vernon, C., Gibson, K.J., et al. (2004) 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* 199: 1191–1199.

Bishai, W. (2000) Lipid lunch for persistent pathogen. *Nature* 406: 683–685.

Burguiere, A., Hitchen, P.G., Dover, L.G., Kremer, L., Ridell, M., Alexander, D.C., Liu, J., Morris, H.R., Minnikin, D.E., Dell, A., and Besra, G.S. (2005) LosA, a key glycosyltransferase involved in the biosynthesis of a novel family of glycosylated acetyltrehalose lipooligosaccharides from *Mycobacterium marinum*. *J Biol Chem* 280: 42124–42133.

Calamita, H., Ko, C., Tyagi, S., Yoshimatsu, T., Morrison, N.E., and Bishai, W.R. (2005) The *Mycobacterium tuberculosis* SigD sigma factor controls the expression of ribosome-associated gene products in stationary phase and is required for full virulence. *Cell Microbiol* 7: 233–244.

Camus, J.C., Pryor, M.J., Medigue, C., and Cole, S.T. (2002) Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* 148: 2967–2973.

Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., and Parkhill, J. (2005) ACT: the Artemis Comparison Tool. *Bioinformatics* 21: 3422–3423.

Casali, N., Konieczny, M., Schmidt, M.A., and Riley, L.W. (2002) Invasion activity of a *Mycobacterium tuberculosis* peptide presented by the *Escherichia coli* AIDA autotransporter. * Infect Immun* 70: 6846–6852.

Chang, J.C., Hanik, N.S., Liao, R.P., and Sherman, D.R. (2007) Identification of mycobacterial genes that alter growth and pathology in macrophages and in mice. *J Infect Dis* (in press).

Chen, P., Ruiz, R.E., Li, Q., Silver, R.F., and Bishai, W.R. (2000) Construction and characterization of a *Mycobacterium tuberculosis* mutant lacking the alternate sigma factor gene, sigF. * Infect Immun* 68: 5575–5580.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393: 537–544.

Corbett, E.L., Watt, C.J., Walker, N., Maher, D., Williams, B.G., Raviglione, M.C., and Dye, C. (2003) The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch Intern Med* 163: 1009–1021.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res* 14: 1186–1190.
Dobson, G., Minnikin, D.E., Minnikin, S.M., Parlett, J.H., Goodfellow, M., and Ridell, M.M.M. (1985) Systematic analysis of complex mycobacterial lipids. In Chemical Methods in Bacterial Systematics. Goodfellow, M. and Minnikin, D.E. (eds). London: Academic Press, pp. 237–265.

Dover, L.G., Corsino, P.E., Daniels, I.R., Cocklin, S.L., Tattituri, V., Besra, G.S., and Futterer, K. (2004) Crystal structure of the TetR/CamR family repressor Mycobacterium tuberculosis EthR implicated in ethionamide resistance. J Mol Biol 340: 1095–1105.

Frenois, F., Engohang-Ndong, J., Locht, C., Baulard, A.R., and Villeret, V. (2004) Structure of EthR in a ligand bound conformation reveals therapeutic perspectives against tuberculosis. Mol Cell 16: 301–307.

Gatfield, J., and Pieters, J. (2000) Essential role for cholesterol in entry of mycobacteria into macrophages. Science 288: 1647–1650.

Gomez, M., and Smith, I. (2000) Determinants of mycobacterial gene expression. In Molecular Genetics of Mycobacteria. Hatfull, G.F., and Jacobs, W.R. Jr (eds). Washington, DC: American Society for Microbiology Press, pp. 111–129.

Grkovic, S., Brown, M.H., Roberts, N.J., Paulsen, I.T., and Skurray, R.A. (1998) QacR is a repressor protein that regulates expression of the Staphylococcus aureus multidrug efflux pump QacA. J Biol Chem 273: 18665–18673.

Huffman, J.L., and Brennan, R.G. (2002) Prokaryotic transcription regulators: more than just the helix-turn-helix motif. Curr Opin Struct Biol 12: 98–106.

Joshi, S.M., Pandey, A.K., Capite, N., Fortune, S.M., Rubin, E.J., and Sassetti, C.M. (2006) Characterization of mycobacterial virulence genes through genetic interaction mapping. Proc Natl Acad Sci USA 103: 11760–11765.

Kendall, S.L., Rison, S.C., Movahedzadeh, F., Frita, R., and Stoker, N.G. (2004) What do microarrays really tell us about M. tuberculosis? Trends Microbiol 12: 537–544.

Malhotra, V., Sharma, D., Ramanathan, V.D., Shakila, H., Saini, D.K., Chakravorty, S., et al. (2004) Disruption of response regulator gene, devR, leads to attenuation in virulence of Mycobacterium tuberculosis. FEMS Microbiol Lett 231: 237–245.

Martin, C., Williams, A., Hernandez-Pando, R., Cardona, P.J., Gormley, E., Bordat, Y., et al. (2006) The live Mycobacterium tuberculosis phoP mutant strain is more attenuated than BCG and confers protective immunity against tuberculosis in mice and guinea pigs. Vaccine 24: 3408–3419.

McKinney, J.D., Honer zu Bentrup, K., Munoz-Elias, E.J., Miczak, A., Chen, B., Chan, W.T., et al. (2000) Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. Nature 406: 735–738.

Mitra, D., Saha, B., Das, D., Wiker, H.G., and Das, A.K. (2005) Correlating sequential homology of Mce1A, Mce2A, Mce3A and Mce4A with their possible functions in mammalian cell entry of Mycobacterium tuberculosis performing homology modeling. Tuberculosis (Edinb) 85: 337–345.

Orth, P., Schnappinger, D., Hillen, W., Saenger, W., and Hinrichs, W. (2000) Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. Nat Struct Biol 7: 215–219.

Parish, T., and Stoker, N.G. (1998) Electrophoresis of mycobacteria. Meth Mol Biol 101: 129–144.

Parish, T., Smith, D.A., Roberts, G., Betts, J., and Stoker, N.G. (2003) The senX3-regX3 two-component regulatory system of Mycobacterium tuberculosis is required for virulence. Microbiology 149: 1423–1435.

Parish, T., and Stoker, N.G. (2000) Use of a flexible cassette method to generate a double unmarked Mycobacterium tuberculosis tlyA plcABC mutant by gene replacement. Microbiology 146: 1969–1975.

Perez, E., Samper, S., Bordas, Y., Guilhout, C., Gicquel, B., and Martin, C. (2001) An essential role for phoP. Mycobacterium tuberculosis virulence. Mol Microbiol 41: 179–187.

Ramos, J.L., Martinez-Bueno, M., Molina-Henares, A.J., Teran, W., Watanabe, K., Zhang, X., et al. (2005) The TetR family of transcriptional repressors. Microbiol Mol Biol Rev 69: 326–356.

Rao, V., Gao, F., Chen, B., Jacobs, W.R. Jr and Glickman, M.S. (2006) Trans-cyclopropanation of mycolic acids on trehalose dimycolate suppresses Mycobacterium tuberculosis-induced inflammation and virulence. J Clin Invest 116: 1660–1667.

Rengarajan, J., Bloom, B.R., and Rubin, E.J. (2005) Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages. Proc Natl Acad Sci USA 102: 8327–8332.

Reuter, M., Kupper, D., Meisel, A., Schroeder, C., and Kruger, D.H. (1998) Cooperative binding properties of restriction endonuclease EcoRII with DNA recognition sites. J Biol Chem 273: 8294–8300.

Rickman, L., Saldanha, J.W., Hunt, D.M., Hoar, D.N., Colston, M.J., Millar, J.B., and Buxton, R.S. (2004) A two-component signal transduction system with a PAS domain-containing sensor is required for virulence of Mycobacterium tuberculosis in mice. Biochem Biophys Res Commun 314: 259–267.

Rosas-Magallanes, V., Stadthagen-Gomez, G., Rauzier, J., Barreiro, L.B., Tailleux, L., Boudou, F., et al. (2007) Signature-tagged transposon mutagenesis identifies novel Mycobacterium tuberculosis genes involved in the parasitism of human macrophages. Infect Immun 75: 504–507.

Russell, D.G., Mwandumba, H.C., and Rhoades, E.E. (2002) Mycobacterium and the coat of many lipids. J Cell Biol 158: 421–426.

Sandy, J., Mushraq, A., Kawamura, A., Sinclair, J., Sim, E., and Noble, M. (2002) The structure of arylamine N-acetyltransferase from Mycobacterium smegmatis – an enzyme which inactivates the anti-tubercular drug, isoniazid. J Mol Biol 318: 1071–1083.

Santangelo, M.P., Goldstein, J., Aito, A., Gioffre, A., Caimi, K., Zaball, O., et al. (2002) Negative transcriptional regulation of the mce3 operon in Mycobacterium tuberculosis. Microbiology 148: 2997–3006.

Sassetti, C.M., and Rubin, E.J. (2003) Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci USA 100: 12989–12994.

Schnappinger, D., Ehrt, S., Voskuil, M.I., Liu, Y., Mangan, J.A., Monahan, I.M., et al. (2003) Transcriptional adaptation of Mycobacterium tuberculosis within macrophages: insights into the phagosomal environment. J Exp Med 198: 693–704.
The following supplementary material is available for this article:

Table S1. All genes. Normalised microarray data for all M. smegmatis genes for which valid data were obtained (kstR mutant compared to the wild-type strain).

Table S2. Genes, P < 0.05. Normalised microarray data for all M. smegmatis genes found to be significantly altered in expression in the kstR mutant compared to the wild-type strain.

Table S3. kstR regulon. Normalised microarray data for M. smegmatis genes we predict to be directly controlled by kstR (kstR mutant compared to the wild-type strain).

Table S4. Secondary effects. Normalised microarray data for M. smegmatis genes were significantly altered in expression (kstR mutant compared to the wild-type strain), but which we believe not to be directly controlled by kstR.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2958.2007.05827.x

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.