

bkaR* is a TetR-type repressor that controls an operon associated with branched-chain keto-acid metabolism in *Mycobacteria

Ricardo J.C. Balhana¹, Sade N. Swanston¹, Stephen Coade², Mike Withers¹, Mahmudul Hasan Sikder¹, Neil G. Stoker¹ & Sharon L. Kendall¹

¹Department of Pathology and Pathogen Biology, The Royal Veterinary College, Camden, London, UK; and ²Division of Mycobacterial Research, MRC National Institute for Medical Research, London, UK

Correspondence: Sharon L. Kendall, Department of Pathology and Pathogen Biology, The Royal Veterinary College, Royal College Street, Camden, London NW1 0TU, UK. Tel.: +44 020 7468 5272; fax: +44 020 7468 5306; e-mail: skendall@rvc.ac.uk

Received 23 May 2013; revised 6 June 2013; accepted 6 June 2013. Final version published online 8 July 2013.

DOI: 10.1111/1574-6968.12196

Editor: Roger Buxton

Keywords

Rv2506; MSMEG_4718; accA1; accD1; transcriptional regulation; microarray.

Abstract

This study describes how *bkaR*, a highly conserved mycobacterial TetR-like transcriptional repressor, regulates a number of nearby genes that have associations with branched-chain keto-acid metabolism. *bkaR* (*MSMEG_4718*) was deleted from the nonpathogenic species *Mycobacterium smegmatis*, and changes in global gene expression were assessed using microarray analysis and reporter gene studies. *bkaR* was found to directly control the expression of 10 genes in *M. smegmatis*, and its ortholog in *Mycobacterium tuberculosis* (*Rv2506*) is predicted to control at least 12 genes. A conserved operator motif was identified, and binding of purified recombinant *M. tuberculosis* BkaR to the motif was demonstrated. Analysis of the stoichiometry of binding showed that BkaR binds to the motif as a dimer.

Introduction

TetR regulators are abundant in the *Mycobacterium* genus, which includes several pathogenic species, and comprise over 30% of the total DNA-binding regulators in both *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. In *M. tuberculosis*, the few whose regulons have been described include Mce3R, a repressor that controls an operon involved in lipid metabolism (de la Paz Santangelo *et al.*, 2009), EthR, involved in the control of the activation of the pro-drug ethionamide (Baulard *et al.*, 2000) and KstR and KstR2, involved in the control of cholesterol catabolism (Kendall *et al.*, 2007, 2010). Although many of the regulons defined to date are conserved in the environmental species *M. smegmatis*, their importance in the life style of the pathogenic species *M. tuberculosis* is clear. The Mce3R regulon has been implicated in the survival of *M. tuberculosis* in mice (Senaratne *et al.*, 2008) as have many of the genes in the

KstR and KstR2 regulons (Hu *et al.*, 2009; Nesbitt *et al.*, 2009; Griffin *et al.*, 2011).

TetR regulators often bind to palindromic motifs in operators using the N-terminal end of the protein to repress transcription. In the presence of a ligand that binds to the C-terminus, the regulator is removed from the operator allowing access to RNA polymerase and transcription. Here, we focus on a mycobacterial TetR regulator, which we name *branched-chain keto-acid regulator* (BkaR). This is highly conserved in both pathogenic and nonpathogenic *Mycobacteria* and is encoded by the gene *Rv2506* in *M. tuberculosis* and *MSMEG_4718* in *M. smegmatis*. BkaR was previously found to have potential associations with pathogenesis in a whole-genome transposon screen (Stewart *et al.*, 2005). We show that BkaR_{Msm} controls the expression of 10 genes likely to be involved in branched-chain keto-acid metabolism. We show relevance to *M. tuberculosis* through protein-binding experiments, regulatory motif analysis and reporter

gene assays. The potential role of the BkaR regulon is discussed.

Materials and methods

Bacterial strains and culture conditions

The strains and plasmids used in this study are described in Supporting Information, Table S1 and were grown as described previously (Kendall *et al.*, 2010).

Microarray analysis of *M. smegmatis* $\Delta bkaR_{Msm}$

$\Delta bkaR_{Msm}$ was available from a previous study (Balhana *et al.*, 2010). RNA from both wild-type and $\Delta bkaR_{Msm}$ strains was prepared by direct sampling into guanidine thiocyanate (GTC) followed by the use of the RNAeasy kit (Qiagen) as previously described (Kendall *et al.*, 2010). Microarrays were provided by the Pathogen Functional Genomics Resource Centre at TIGR (<http://pfgrc.jcvi.org>). cDNA from wild-type and mutant strains were labelled and competitively hybridized onto the arrays. All methods including scanning, data analysis and significance criteria were as previously described (Kendall *et al.*, 2010). Fully annotated microarray data have been deposited in BμG@Sbase (accession number E-BUGS-116; <http://bugs.sgul.ac.uk/E-BUGS-116>) and also ArrayExpress (accession number E-BUGS-116).

RT-PCR for analysis of operon structure

RNA isolated from the $\Delta bkaR_{Msm}$ mutant was treated twice with DNase I (Invitrogen) (30 min at 37 °C). The reaction was inactivated at 65 °C for 10 min in the presence of 1.25 mM EDTA. DNA-free RNA (150 ng) was mixed with 300 ng of random primers, 10 mM DTT, 0.5 mM dNTPs and reverse-transcribed with 200 units of Superscript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions. Control reactions were performed without reverse transcriptase. PCR was carried out using Phusion[®] High-Fidelity PCR Master Mix with GC buffer (New England Biolabs).

Reporter analysis

This was performed using the *lacZ* integrative reporter construct pEJ414 (Papavinasundaram *et al.*, 2001). Upstream intergenic regions of genes of interest were either PCR-amplified or synthesized as oligonucleotides (Table S2), cloned using NotI and XbaI and electroporated into wild-type and $\Delta bkaR_{Msm}$ strains. Reporter assays were carried out as described previously (Papavinasundaram *et al.*, 2001).

Expression and purification of recombinant BkaR_{Mtb}

bkaR_{Mtb} was PCR-amplified from *M. tuberculosis* H37Rv genomic DNA (Table S2) and inserted into pNIC28-Bsa4 (GenBank accession no. EF198106) using ligation-independent cloning with T4 DNA polymerase. The final construct, pNbkaR-MTB was transformed into *Escherichia coli* BL21(DE3). Expression and induction were achieved in 400 mL of autoinducing medium (Studier, 2005) at 37 °C overnight. Cultures were harvested by centrifugation (30 min, 3070 g, 4 °C), and the pellet was resuspended in 5 mL of lysis buffer (70 mM HEPES, 20 mM imidazole, 650 mM NaCl, 0.5 mM β-mercaptoethanol, 10% glycerol, pH 8). Cells were lysed on ice using sonication (Soniprep 150) at 20 μm for 5 min with 30-s rest period every 1 min. Soluble fractions were isolated through centrifugation (11 337 g, 20 min, 4 °C), and His₆-BkaR_{Mtb} was purified using immobilized metal ion affinity chromatography on HisTrap FF Ni-Sepharose columns (GE Healthcare Life Sciences). The protein was eluted with histidine elution buffer (250 mM histidine, 60 mM HEPES, 150 mM NaCl, 3% glycerol, pH 8).

Electrophoretic mobility shift assays

DNA oligonucleotides or PCR amplicons were used to assay the binding of His₆-BkaR_{Mtb}. Probes used for competition assays were end-labelled with DIG-11-ddUTP using the DIG gel shift kit, 2nd generation (Roche). Binding reactions were performed by incubating varying concentrations of protein with 0.03 pmol of labelled probe in 1X binding buffer (20 mM HEPES, 1 mM EDTA, 10 mM (NH₄)₂SO₄, 1 mM DTT, 0.2% Tween-20, 30 mM KCl, pH 7.6) together with 0.1 μg of poly-L-lysine and 1 μg of poly[d(I-C)] in a total volume of 20 μL. Specific competitor (nonlabelled probe) was added in 150-fold excess. Mixtures also contained an excess 125-fold poly [d(I-C)] (nonspecific competitor). Reactions were incubated, separated by polyacrylamide gel electrophoresis and blotted onto Hybond-N nylon membranes as previously described (Kendall *et al.*, 2010). The membrane was washed according to the manufacturer's instructions and detection followed the chemiluminescent method using anti-DIG-alkaline phosphatase and the substrate CSPD. The luminescent membranes were exposed to X-ray film for varying time points between 5 min and 1 h.

Determination of the binding stoichiometry of the His₆-BkaR_{Mtb}-DNA complex

A Ferguson plot assay was used to determine the molecular weight of the protein-DNA complex (Orchard & May,

1993). Electrophoretic mobility shift assays (EMSA) reactions were run alongside protein standards through a series of gels differing in acrylamide concentration (range of 6–12%). Gels were run at 10 V cm^{-1} in 0.4X TBE buffer, and once finished, the distance migrated by the bromophenol blue was measured for each sample. Each gel was cut in half, isolating the tracks containing the protein–DNA complex from the tracks with protein standards. The gel fragments containing DNA were stained for 30 min at room temperature in 0.4X TBE with $0.68 \mu\text{g mL}^{-1}$ ethidium bromide, whereas the gel fragments containing the protein standards were Coomassie-stained (30% methanol, 8% acetic acid, 0.25% w/v Coomassie blue R). The distances migrated by the protein–DNA complexes and by each standard were measured and divided by the distance migrated by the bromophenol blue, giving the relative mobility (R_f) for each species. The logarithms of the relative mobilities were plotted against gel concentration, and the retardation coefficients (K_r , slopes of the trend curves) were calculated and plotted again vs. the molecular weight of each standard. The Ferguson plot obtained allowed determination of the molecular weight of the complex.

Bioinformatic analyses

Genome sequences were compared using Artemis Comparison Tool (ACT) (Carver *et al.*, 2005), and using CLUSTALW (Thompson *et al.*, 1994), operator motif discovery and analysis were carried out using MEME and MAST (Bailey & Elkan, 1994; Bailey & Gribskov, 1998).

Results

Genomic analyses of the *bkaR* region show conservation in *Mycobacteria* and closely related species

Using genomic alignment, *bkaR* was found to be clearly conserved in *M. tuberculosis*, *M. bovis* BCG, *Mycobacterium marinum*, *Mycobacterium ulcerans*, *Mycobacterium avium*, *M. smegmatis*, *Rhodococcus jostii* and *Nocardia farcinica* (Fig. 1). The percentage amino acid sequence identities between BkaR in *M. tuberculosis* and its orthologs were > 50% in all cases. Most neighbouring genes were also conserved, although *fadD35*, *scoA* and *scoB* were only present in species more closely related to *M. tuberculosis*.

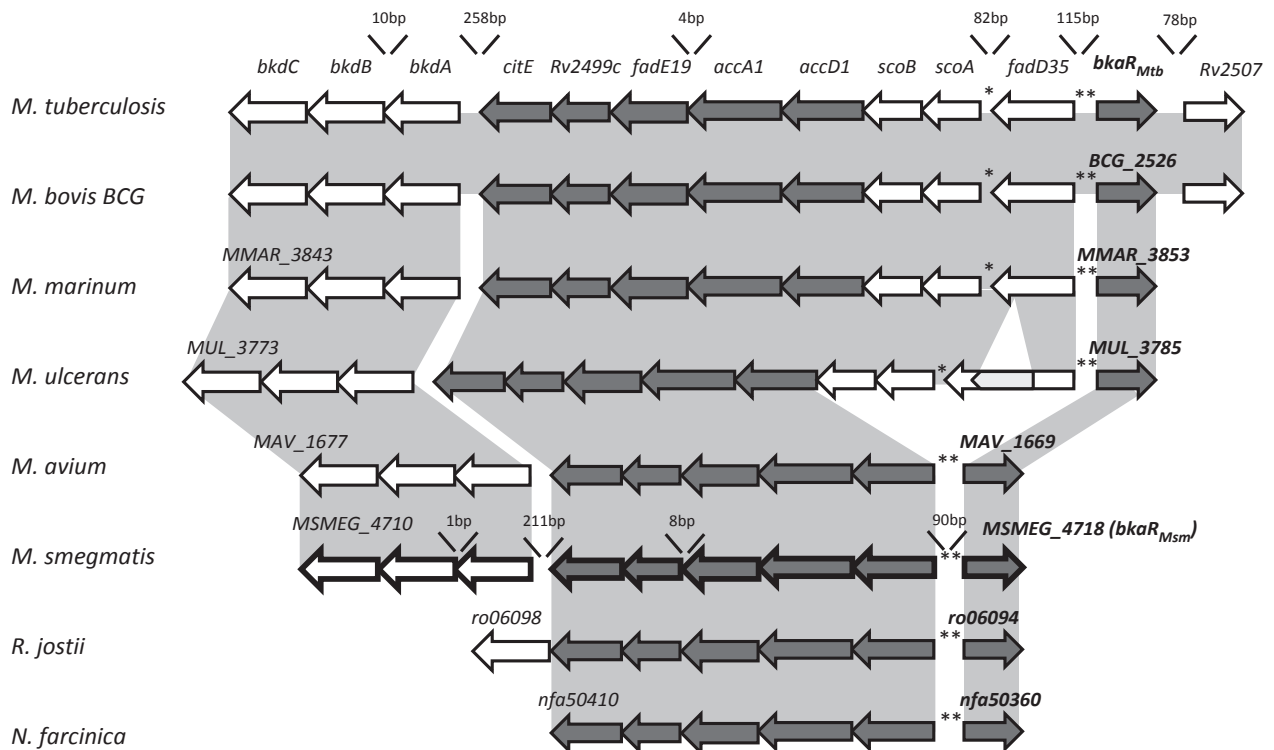


Fig. 1. Genomic context of *bkaR* in *Mycobacteria* and close relatives. The grey shading represents regions of conservation between genomes. Arrows represented with a dark grey background correspond to genes that are conserved amongst all species, whereas genes that find no ortholog in at least one species are shown with white arrows. The light grey arrow inserted in *MUL_3785* represents a transposon insertion. Asterisks represent binding motifs identified by the programs MEME and MAST, and the sizes of the intergenic gaps in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis* are indicated. Genes that were derepressed in the $\Delta bkaR_{Msm}$ mutant are highlighted by thickened lines.

BkaR is autoregulatory and binds to a conserved palindromic motif within its own promoter

TetR regulators are generally autoregulatory. Measurement of expression from the *bkaR*_{Msm} and *bkaR*_{Mtb} promoters in *lacZ* reporter constructs in wild-type *M. smegmatis* and $\Delta bkaR$ _{Msm} showed that transcription was approximately fourfold higher in the strain lacking the *bkaR* regulator, indicating that both *bkaR*_{Mtb} and *bkaR*_{Msm} are autoregulatory and repress their own expression (Fig. 2).

Transcriptional repressors of the TetR family tend to bind to palindromic DNA motifs (Yu *et al.*, 2010). Computational analysis of the *bkaR* promoter using MEME identified a 24-bp palindromic motif with a highly conserved 16-bp core GTTA(N)T(N4)A(N)TAAC that was present twice in the promoter regions of all the species tested (Fig. 3a). The promoter regions of some of the *Mycobacteria* and *Nocardia* species were aligned to identify the position of the motif (Fig. 3b). Both copies of the motif were clearly visible with the 16-bp core being more conserved than the rest of the intergenic region. Additionally, there was another region that showed more conservation. This could possibly be the -35 site as it is a short conserved GC-rich region as seen in other mycobacterial promoters (Gomez & Smith, 2000).

To test if BkaR binds to the two motifs, 30-bp double-stranded probes containing the motifs were used in EMSAs. The probes were DIG-labelled, and the binding

assays were carried out in the presence of 125-fold excess of nonspecific competitor poly[d(I-C)] and with specific competition with unlabelled probe (Fig. 3c). Clear shifts were seen in the presence of the nonspecific competitor at all molar ratios. At a 1 : 10 molar ratio of DNA/protein, the presence of excess unlabelled probe successfully competed with the labelled probe (lanes marked with an asterisk). BkaR appears to show lower affinity for probe 2 (motif near *fadD35* in *M. tuberculosis* and *accD1* in *M. smegmatis*) as a complete shift was not achieved for any of the DNA/protein ratios, whereas with probe 1, a complete shift was observed at a 1 : 2 molar ratio.

BkaR binds to the motif as a dimer

The stoichiometry of binding of BkaR_{Mtb} with the motif DNA was examined using Ferguson plot analysis (Ferguson, 1964; Orchard & May, 1993). The logarithms of the relative mobility (R_f) of protein standards and BkaR_{Mtb}-DNA complex were plotted against the percentage gel concentration (Data S1), and the slopes (retardation coefficient) were calculated. The retardation coefficients were subsequently plotted as a function of molecular weight. Using the equation of the adjusted curve obtained, we estimated the molecular weight of the complex to be 76.9 kDa. Subtracting the molecular weight of the 30-bp DNA oligonucleotide used (18.4 kDa), the mass accounted for the protein alone is 58.5 kDa, which approximately corresponds to the mass of an His₆-BkaR_{Mtb} dimer (monomer is 25.8 kDa).

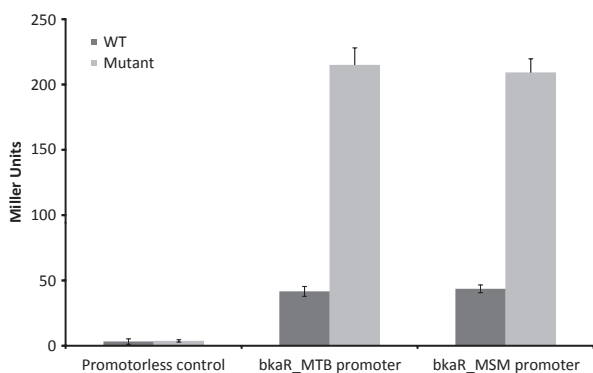


Fig. 2. Expression of the *bkaR*_{Mtb} and *bkaR*_{Msm} promoter in wild-type mc^2 155 and $\Delta bkaR$ _{Msm} mutant background. Beta-galactosidase activity was calculated in Miller Units (measured as the amount of O-nitrophenol produced (nmol) over time (min) per mg of protein). Deletion of *bkaR*_{Msm} causes a statistically significant increase in the expression of *lacZ* from the *bkaR*_{Mtb} and the *bkaR*_{Msm} promoters (unpaired Student's *t*-test; $P < 0.05$, *). The values result from means of biological triplicates and technical duplicates (this experiment was repeated independently with similar results). Error bars represent standard deviations.

Genomewide analysis of the presence of the motif in *M. smegmatis* and *M. tuberculosis* and global gene expression analyses

Computational analyses for further instances of the motif using MAST showed that there were four more instances of the motif in the *M. smegmatis* genome and three more instances in *M. tuberculosis* (Table 1). In *M. smegmatis*, these were found upstream of *MSMEG_4920*, *MSMEG_4524* and between the divergently oriented *MSMEG_3414/MSMEG_3415* (two copies). However, the significances of these motifs were much lower than the motif originally found within the *bkaR* promoter. DNA-binding studies provided experimental corroboration for the motif upstream of *MSMEG_4920* only (Data S3).

In *M. tuberculosis*, the additional motifs were found upstream of *Rv2503c* (*scoA*) and between the divergently oriented *Rv0575c/Rv0576* (two copies). The putative motif upstream of *scoA* (which lies close to *bkaR* in the *M. tuberculosis* genome, but is not present in *M. smegmatis*; see Fig. 1) had comparable significance to the original motifs identified by MEME, and we were able to observe

Table 1. Occurrences of the motif in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*

Motif sequence	E-value	Flanking genes	EMSA	Reporter assay	Microarray analysis
<i>M. smegmatis</i>					
CTCG GTTAAT CGTG ACTAAC GTAC	1.4e-05	bkaR_{Msm} / <i>MSMEG_4717</i>	+	+	+
CTGAG GTTAAT GACG ACTAAC TCGT	3.9E-05	MSMEG_4717/bkaR_{Msm}	+	ND	+
GCCG GTTAAT CATCGT TAAC ACAG	0.6	<i>MSMEG_4920</i>	+	ND	+
CCAG GTTAGC CAAA CTA TTTCAG	2	MSMEG_3414/MSMEG_3415	–	–	–
CTGAAA TAGT TTTGG CTAAC CTGG	3.4	<i>MSMEG_3414/MSMEG_3415</i>	–	–	–
CTAG GTTA GGCTAC ATTAC TTGC	8.2	<i>MSMEG_4524</i>	–	–	–
<i>M. tuberculosis</i>					
CTAC GTTAGT GACG ATTAA CCGAA	6.1e-06	bkaR_{Mtb} / <i>fadD35</i>	+	+	ND
TTCG GTTA ACATCG AATAG CTTAG	1.1e-03	<i>bkaR_{Mtb}</i> / fadD35	+	ND	ND
CTCAG GTTAAT GATA ATTAAC TGAA	9.8E-06	<i>scoA</i>	+	ND	ND
ATAC GTTA GCGCTC ACTAAC GTAT	2.7	Rv0575c/Rv0576	–	–	ND
ATAC GTTAGT GAGCG CTAAC GTAT	3.1	<i>Rv0575c/Rv0576</i>	–	–	ND

ND, not done; +, complete shift or derepression in the case of reporter and microarray analysis; –, no shift observed or no derepression of gene expression; underlined nucleotides represent conserved base pairs of the motif; gene numbers in bold represent the ones that are closer to the motif for divergent arrangements.

Table 2. The regulon of *bkaR_{Msm}*

<i>Mycobacterium smegmatis</i>	Fold change	P-value	Gene name	Ortholog in <i>Mtb</i>	Annotated function
<i>MSMEG_5576</i>	3.4	2.2E-02	–	–	D-mannonate oxidoreductase
MSMEG_4920	4.9*	7.9E-04	fadA4	<i>Rv1323</i>	Acetyl-CoA acetyltransferase
MSMEG_4718	–1.8*	4.8E-01	bkaR	<i>Rv2506</i>	TetR transcriptional regulator (bkaR)*
MSMEG_4717	9.0*	3.7E-05	accD1	<i>Rv2502c</i>	Acetyl/propionyl-coenzyme A carboxylase (β subunit)
MSMEG_4716	15.0	1.5E-05	accA1	<i>Rv2501c</i>	Acetyl/propionyl-coenzyme A carboxylase (α subunit)
MSMEG_4715	13.7	5.8E-06	fadE19	<i>Rv2500c</i>	Acyl-CoA dehydrogenase
MSMEG_4714	9.5	1.1E-04	Rv2499c	<i>Rv2499c</i>	Hydratase
MSMEG_4713	16.5	4.6E-06	citE	<i>Rv2498c</i>	HpcH/Hpal aldolase/citrate lyase family protein
MSMEG_4712	8.4	1.1E-04	bkdA	<i>Rv2497c</i>	Part of branched-chain keto-acid dehydrogenase complex†
MSMEG_4711	3.3	3.3E-02	bkdB	<i>Rv2496c</i>	Part of branched-chain keto-acid dehydrogenase complex†
MSMEG_4710	1.8	4.9E-01	bkdC	<i>Rv2495c</i>	Part of branched-chain keto-acid dehydrogenase complex†
<i>MSMEG_4005</i>	–3.2	2.5E-02	–	–	Calcium-binding protein
<i>MSMEG_2080</i>	3.1	3.3E-02	<i>fadE23</i>	<i>Rv3140</i>	Putative acyl-CoA dehydrogenase
<i>MSMEG_1885</i>	3.3	2.2E-02	<i>Rv3230c</i>	<i>Rv3230c</i>	Iron-sulphur cluster binding domain protein
<i>MSMEG_1548</i>	3.0	4.3E-02	–	–	Dehydratase (propanediol utilization)
<i>MSMEG_1543</i>	3.2	2.3E-02	<i>Rv0458</i>	<i>Rv0458</i>	EPTC-Inducible aldehyde dehydrogenase
<i>MSMEG_0881</i>	–3.4	2.2E-02	–	–	Hypothetical protein
<i>MSMEG_0066</i>	3.3	2.2E-02	<i>esxA</i>	<i>Rv3875</i>	Early secretory antigenic target

The genes in bold are those directly regulated by *bkaR*.

*Occurrence of the motif.

†Venugopal *et al.* (2011).

binding of purified BkaR_{Mtb} to this motif using EMSA (Data S3). However, the occurrences of the motifs in the intergenic region between *Rv0575c* and *Rv0576* were less significant, and no DNA binding was observed (data not shown). These results indicate that the motifs upstream of *MSMEG_4920* and *scoA* are bound by BkaR but those identified upstream of the other genes are either not functional binding sites or motifs for an unidentified transcriptional regulator.

Global gene expression changes as a result of deleting *bkaR* were examined by microarray analysis of wild-type and Δ *bkaR_{Msm}* strains. A total of 14 genes were found to

be significantly derepressed for a *P*-value cut-off of 0.05 (Table 2). Additionally, *MSMEG_4710* (*bkdC*) was also derepressed but did not meet the significance criteria. Although this gene did not meet the significance cut-off, it was 1.8-fold derepressed, and other evidence suggests that it is part of the regulon; therefore, it was included in the table (see below).

The highest fold change (> 8) was observed in genes that are part of the operon divergently oriented to *bkaR_{Msm}* (*accD1-citE*) (Fig. 1). With the exception of an 8-bp gap between *fadE19* and *MSMEG_4714*, the genes *accD1-citE* are contiguous and are likely to be expressed

as a single transcript. Therefore, the microarray data suggest that, as predicted by the presence of the motif, the regulator controls the expression of the divergently transcribed genes between *accD1-citE* in *M. smegmatis*. This correlates with the observation that TetR regulators often control adjacent genes (Ahn *et al.*, 2012).

The genes between *bkdA* and *bkdC* were also derepressed in the mutant (although the last gene *bkdC* was not significant at the chosen *P*-value cut-off). This was unexpected because of the presence of a 211-bp gap upstream of *bkdA* with no associated regulatory motif. Therefore, RT-PCR was used to assess whether *bkdABC* were cotranscribed with the upstream genes. Reactions were performed with primers annealing across runs of genes between *fadE19* and *bkdA* incorporating the 211-bp gap (Data S2). The presence of a band of the expected size in the samples that were reverse-transcribed samples only (+) supports the microarray data and indicates that these genes are indeed cotranscribed. Large gaps (> 200 bp) have been reported to separate genes that form an operon in other bacterial species (Krause *et al.*, 1992). *MSMEG_4920* was also significantly up-regulated with over fourfold change and has a proximal motif to which the regulator bound (Data S3). These combined evidence strongly suggest that *MSMEG_4920* is also directly under the control of the regulator.

The remaining genes that are either significantly derepressed (*MSMEG_5576*, *MSMEG_2080*, *MSMEG_1885*, *MSMEG_1548*, *MSMEG_1543*, *MSMEG_0066*) or significantly repressed (*MSMEG_4005*, *MSMEG_0881*) are not associated with a motif. We conclude that these are not likely to be under the direct control of BkaR. The changes in expression in these genes may be an indirect effect of knocking out the regulator.

Discussion

The *bkaR* regulon in *M. smegmatis* and prediction of the regulon in *M. tuberculosis*

This study shows that the highly conserved transcriptional regulator *bkaR* binds to a 16-bp palindromic motif and to act as a repressor to directly control expression of itself, and of the divergently oriented operon (Fig. 1). In *M. smegmatis*, this consists of the genes from *accD1-bkdC*. Additionally, *bkaR_{Msm}* controls the expression of *MSMEG_4920*, which is de-repressed in the microarray analysis and has an associated upstream motif.

While writing this manuscript, a paper was published describing a single motif in the intergenic region of *bkaR_fadD35* in *M. tuberculosis*. Binding was demonstrated to this region together with affinity measurements, and it was shown that both tetracycline and palmitoyl-coA

could interfere with binding (Anand *et al.*, 2012). These authors suggested that *bkaR* (which they call *fad35R* and consider to be a homologue of *E. coli* FadR) controls the expression of *fadD35* in *M. tuberculosis*. However, the palindrome described by Anand *et al.* was only partially identified (the authors describe a diffuse and poorly conserved palindrome), only a single motif in the *bkaR_fadD35* region was described, the motif upstream of *scoA* was not identified, and no gene expression studies were carried out to support the work. In contrast, our data clearly show the presence of two palindromic motifs in the *bkaR-fadD35* intergenic region, a motif upstream of *scoA*, and we provide gene expression data to support the identification of the regulon in *M. smegmatis*.

Gene expression studies in combination with motif analysis have previously allowed us to use data from *M. smegmatis* to predict the regulon in *M. tuberculosis*, and these predictions have been subsequently experimentally verified (Kendall *et al.*, 2007; Nesbitt *et al.*, 2009). Similarly, in this study, we predict that in addition to the orthologs shown in Table 2, *fadD35* and the cotranscribed genes *scoA-scoB* will be controlled by BkaR in *M. tuberculosis*, and we show that BkaR_{Mtb} binds to the motifs upstream of *fadD35* and *scoA*. The presence of the two motifs in the *bkaR_fadD35* intergenic region strongly suggests that *bkaR* acts like the paradigm TetR where the repressor binds as a dimer to each motif to repress expression in both directions (Hillen & Berens, 1994).

Possible functions of the genes in the *bkaR* regulon

Many of the genes in the regulon have annotated functions that could be involved in β -oxidation. However, much of the *M. tuberculosis* genome is dedicated to fatty acid β -oxidation, and it is often difficult to work out the precise substrates and exactly what role each of the enzymes play in the pathogens lifestyle. This is compounded by re-annotation of some of the genes in the regulon. The *bkdABC* genes were originally annotated as encoding for the pyruvate dehydrogenase complex (*pdhABC*), but have been found not to possess such activity (Tian *et al.*, 2005). More recent work suggests that they form part of a complex that has branched-chain keto-acid dehydrogenase (BCKADH) activity, which is the second stage in the catabolism of the branched-chain amino acids leucine, valine and isoleucine (Venugopal *et al.*, 2011).

This latter observation has led us to speculate that the other genes in the regulon also catalyse reactions in this pathway (Data S4). The branched-chain keto-acid derivatives are activated by the addition of coenzyme A (the product of BCKADH activity) and are then degraded in a

series of reactions that involve dehydrogenase (*fadE19*), hydratase (*Rv2499c*), and, specifically in the case of leucine, carboxylase (*accA1accD1*) activity. The ultimate endpoint of the catabolism of branched-chain amino acids is acetyl-coA and propionyl-coA. Acetyl-coA can enter the glyoxylate cycle where it is used for energy generation, while propionyl-coA enters the methyl citrate cycle and the methyl malonyl pathway. Intermediates from these cycles can enter the TCA cycle; in this way, energy can be derived from the breakdown of branched-chain amino acids. Interestingly, other recent studies have also found evidence for the role of *accA1accD1* in branched-chain amino acid catabolism in *Mycobacteria* (Ehebaur *et al.*, 2013).

The branched-chain keto-acid derivatives (namely isobutyryl-CoA and 2-methyl-butryryl-CoA) also act as precursors for branched-chain fatty acid synthesis, and so, it is possible that the other genes in the *bkaR* regulon are involved in a synthetic pathway. In this alternative scenario, the *acc* genes act as other biotin-dependent carboxylases used for the synthesis of multi-methyl-branched fatty acids that can be units for mycolic acids, as in the case of *accD4* and *accD5* in *M. tuberculosis* (Gande *et al.*, 2004, 2007; Daniel *et al.*, 2007).

In conclusion, we have described the regulation of a set of genes likely to be involved in branched-chain keto-acid metabolism genes in *Mycobacteria*. The genes are repressed by an adjacent autoregulatory TetR regulator, *bkaR*, which binds to a conserved palindromic motif in its own promoter region and the promoter regions of the genes it controls.

Acknowledgements

This project was funded by the Wellcome trust (Grant 073237) and the MRC U117581288 (SC). The microarrays were performed using the Bacterial Microarray facility at St. Georges Hospital Medical School (BμG@S). We would also like to thank Dr Elaine Davis for the plasmid pEJ414. RVC manuscript number P/PID/000132/XDH26.

References

- Ahn SK, Cuthbertson L & Nodwell JR (2012) Genome context as a predictive tool for identifying regulatory targets of the TetR family transcriptional regulators. *PLoS ONE* **7**: 1–13.
- Anand S, Singh V, Singh AK, Mittal M, Datt M, Subramani B & Kumaran S (2012) Equilibrium binding and kinetic characterization of putative tetracycline repressor family transcription regulator Fad35R from *Mycobacterium tuberculosis*. *FEBS J* **279**: 3214–3228.
- Bailey TL & 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 TL & Gribskov M (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**: 48–54.
- Balhana R, Stoker NG, Sikder MH, Chauviac FX & Kendall SL (2010) Rapid construction of mycobacterial mutagenesis vectors using ligation-independent cloning. *J Microbiol Methods* **83**: 34–41.
- Baulard AR, Betts JC, Engohang-Ndong J, Quan S, McAdam RA, Brennan PJ, Loch C & Besra GS (2000) Activation of the pro-drug ethionamide is regulated in mycobacteria. *J Biol Chem* **275**: 28326–28331.
- Carver TJ, Rutherford KM, Berriman M, Rajandream MA & Barrell J (2005) ACT: the Artemis Comparison Tool. *Bioinformatics* **21**: 3422–3423.
- de la Paz Santangelo M, Klepp L, Nunez-Garcia J *et al.* (2009) Mce3R, a TetR-type transcriptional repressor, controls the expression of a regulon involved in lipid metabolism in *Mycobacterium tuberculosis*. *Microbiology* **155**: 2245–2255.
- Daniel J, Oh TJ, Lee CM & Kolattukudy PE (2007) AccD6, a member of the Fas II locus, is a functional carboxyltransferase subunit of the acyl-coenzyme A carboxylase in *Mycobacterium tuberculosis*. *J Bacteriol* **189**: 911–917.
- Ehebaur M, Noens EE, Zimmerman M, Marrakchi H, Laneelle M-A, Anandhakrishnan M, Daffe M, Saur U & Wilmanns M (2013) *Catabolism of Branched Chain Amino Acids in Mycobacteria: Identification of Novel Carboxylase Activities in the Acyl-coA Carboxylase Family*. Paper presented at Tuberculosis: Understanding the Enemy, British Columbia, Canada.
- Ferguson KA (1964) Starch-gel electrophoresis—application to the classification of pituitary proteins and polypeptides. *Metabolism* **13**(suppl): 985–1002.
- Gande R, Gibson KJ, Brown AK, Krumbach K, Dover LG, Sahn H, Shioyama S, Oikawa T, Besra GS & Eggeling L (2004) Acyl-CoA carboxylases (*accD2* and *accD3*), together with a unique polyketide synthase (*Cg-pks*), are key to mycolic acid biosynthesis in Corynebacteriaceae such as *Corynebacterium glutamicum* and *Mycobacterium tuberculosis*. *J Biol Chem* **279**: 44847–44857.
- Gande R, Dover LG, Krumbach K, Besra GS, Sahn H, Oikawa T & Eggeling L (2007) The two carboxylases of *Corynebacterium glutamicum* essential for fatty acid and mycolic acid synthesis. *J Bacteriol* **189**: 5257–5264.
- Gomez M & Smith I (2000) Determinants of mycobacterial gene expression. *Molecular Genetics of Mycobacteria* (Hatfull GF & Jacobs WR Jr, eds.), pp. 111–129. ASM Press, Washington, DC.
- Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ & Sasseti CM (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog* **7**: e1002251.
- Hillen W & Berens C (1994) Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annu Rev Microbiol* **48**: 345–369.

- Hu Y, van der Geize R, Besra GS, Gurcha SS, Liu A, Rohde M, Singh M & Coates A (2009) 3-Ketosteroid 9 α -hydroxylase is an essential factor in the pathogenesis of *Mycobacterium tuberculosis*. *Mol Microbiol* **75**: 107–121.
- Kendall SL, Withers M, Soffair CN *et al.* (2007) A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *Mol Microbiol* **65**: 684–699.
- Kendall SL, Burgess P, Balhana R, Withers M, Ten Bokum A, Lott JS, Gao C, Uhiia-Castro I & Stoker NG (2010) Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: kstR and kstR2. *Microbiology* **156**: 1362–1371.
- Krause M, Fang FC & Guiney DG (1992) Regulation of plasmid virulence gene-expression in salmonella-dublin involves an unusual operon structure. *J Bacteriol* **174**: 4482–4489.
- Nesbitt NM, Yang X, Fontan P, Kolesnikova I, Smith I, Sampson NS & Dubnau E (2009) A thiolase of *Mycobacterium tuberculosis* is required for virulence and production of androstenedione and androstadienedione from cholesterol. *Infect Immun* **78**: 275–282.
- Orchard K & May GE (1993) An EMSA-based method for determining the molecular weight of a protein–DNA complex. *Nucleic Acids Res* **21**: 3335–3336.
- Papavinasundaram KG, Anderson C, Brooks PC, Thomas NA, Movahedzadeh F, Jenner PJ, Colston MJ & Davis EO (2001) Slow induction of RecA by DNA damage in *Mycobacterium tuberculosis*. *Microbiology* **147**: 3271–3279.
- Senaratne RH, Sidders B, Sequeira P *et al.* (2008) *Mycobacterium tuberculosis* strains disrupted in *mce3* and *mce4* operons are attenuated in mice. *J Med Microbiol* **57**: 164–170.
- Stewart GR, Patel J, Robertson BD, Rae A & Young DB (2005) Mycobacterial mutants with defective control of phagosomal acidification. *PLoS Pathog* **1**: 269–278.
- Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**: 207–234.
- Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.
- Tian J, Bryk R, Shi S, Erdjument-Bromage H, Tempst P & Nathan C (2005) *Mycobacterium tuberculosis* appears to lack alpha-ketoglutarate dehydrogenase and encodes pyruvate dehydrogenase in widely separated genes. *Mol Microbiol* **57**: 859–868.
- Venugopal A, Bryk R, Shi S, Rhee K, Rath P, Schnappinger D, Ehrt S & Nathan C (2011) Virulence of *Mycobacterium tuberculosis* depends on lipoamide dehydrogenase, a member of three multienzyme complexes. *Cell Host Microbe* **9**: 21–31.
- Yu Z, Reichheld SE, Savchenko A, Parkinson J & Davidson AR (2010) A comprehensive analysis of structural and sequence conservation in the TetR family transcriptional regulators. *J Mol Biol* **400**: 847–864.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Bacterial strains and plasmids used in this study.

Table S2. Primers and oligonucleotides used in this study. Sites used in cloning, or, in the case of EMSAs, the location of the motif, are underlined.

Data S1. Stoichiometry of the His₆-BkaR_{Mtb}-DNA complex.

Data S2. RT-PCR results for the run of genes *fadE19-bkdA* performed with cDNA derived from *AbkaR_{Msm}*.

Data S3. EMSA with purified His₆-BkaR and probes containing the motifs predicted by MAST in *M. smegmatis* and *M. tuberculosis*.

Data S4. Overview of branched chain amino acid metabolism and the possible involvement of the *bkaR* regulon.