

DR. APOORVA BHATT (Orcid ID : 0000-0002-6655-1636)

Article type : Research Article

**The mycolic acid reductase Rv2509 has distinct structural motifs and is essential for growth in slow growing mycobacteria**

Asma Javid<sup>1†</sup>, Charlotte Cooper<sup>1†</sup>, Albel Singh<sup>1</sup>, Steffen Schindler<sup>2</sup>, Milena Hänisch<sup>2</sup>, Robert Marshall<sup>1</sup>, Rainer Kalscheuer<sup>2</sup>, Vassiliy N. Bavro<sup>3\*</sup> and Apoorva Bhatt<sup>1\*</sup>

<sup>1</sup>School of Biosciences and Institute of Microbiology and Infection, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom

<sup>2</sup>Institute of Pharmaceutical Biology and Biotechnology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

<sup>3</sup>School of Life Sciences, University of Essex, Colchester, CO43SQ, United Kingdom.

† Joint first authors

\*Joint corresponding authors

a.bhatt@bham.ac.uk

v.bavro@essex.ac.uk

**Running Title:** Mycolic acid reduction in mycobacteria

**Key words:** *Mycobacterium*, mycolic acid, reductase, dehydrogenase, tuberculosis

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/MMI.14437](https://doi.org/10.1111/MMI.14437)

This article is protected by copyright. All rights reserved

## Summary

The final step in mycolic acid biosynthesis in *Mycobacterium tuberculosis* is catalysed by a mycolyl reductase encoded by the *Rv2509* gene. Sequence analysis and homology modelling indicates that *Rv2509* belongs to the short-chain fatty acid dehydrogenase/reductase (SDR) family, but with some distinct features that warrant its classification as belonging to a novel family of short-chain dehydrogenases. In particular, the predicted structure revealed a unique  $\alpha$ -helical C-terminal region which we demonstrated to be essential for *Rv2509* function, though this region did not seem to play any role in protein stabilisation or oligomerisation. We also show that unlike the *M. smegmatis* homologue which was not essential for growth, *Rv2509* was an essential gene in slow growing mycobacteria. A knockdown strain of the *BCG2529*, the *Rv2509* homologue in *Mycobacterium bovis* BCG was unable to survive following conditional depletion of *BCG2529*. This conditional depletion also led to a reduction of mature mycolic acid production and accumulation of intermediates derived from 3-oxo-mycolate precursors. Our studies demonstrate novel features of the mycolyl reductase *Rv2509* and outline its role in mycobacterial growth, highlighting its potential as a new target for therapies.

**Key words:** *Mycobacterium*, mycolic acid, reductase, dehydrogenase, tuberculosis

## Introduction

The cell walls of mycobacteria including the tuberculosis-causing *Mycobacterium tuberculosis* contain distinct long chain fatty acids termed mycolic acids. These  $\alpha$ -alkyl,  $\beta$ -hydroxy fatty acids form an integral part of the cell wall of mycobacteria, and of other related bacteria of the suborder Corynebacterianae, including the genera *Nocardia*, *Corynebacterium* and *Rhodococcus* (Marrakchi *et al.*, 2014). Mycolic acids can be found covalently-linked to cell wall arabinogalactan, and as part of the glycolipids trehalose monomycolate (TMM), trehalose dimycolate (TDM) and glucose monomycolate. They are essential for mycobacterial viability (Marrakchi *et*

*al.*, 2014, Nataraj *et al.*, 2015) and virulence and are synthesised by a complex array of enzymes that includes a mammalian-like Fatty Acid Synthase I (FAS-I), and a multienzyme complex of Fatty Acid Synthase-II (FAS-II) (Marrakchi *et al.*, 2014). One of the late stages of mycolic acid biosynthesis involves the Claisen condensation of a FAS-II-derived long meromycolate chain (C<sub>42</sub>-C<sub>62</sub>) with a short FAS-I-derived fatty acid (C<sub>24</sub>-C<sub>26</sub>) to yield an  $\alpha$ -alkyl,  $\beta$ -keto fatty acid intermediate, catalysed by the polyketide synthase Pks13 (Gande *et al.*, 2004, Portevin *et al.*, 2004). The final step in the formation of a mature mycolic acid involves the enzymatic reduction of the  $\beta$ -keto group of the product of Pks13 to produce a  $\alpha$ -alkyl,  $\beta$ -hydroxy fatty acid (mature mycolic acid). In *M. tuberculosis*, the reductase required for this final step is encoded by *Rv2509* (Bhatt *et al.*, 2008, Lea-Smith *et al.*, 2007). Initial studies in *Corynebacterium glutamicum*, a species that can survive loss of mycolic acids, indicated that deletion of *cmrA* (*NCgl2385*), the *Rv2509* orthologue, resulted in a strain that accumulated the  $\alpha$ -alkyl,  $\beta$ -keto fatty acid precursor instead of mature mycolic acids (Lea-Smith *et al.*, 2007). Mycolic acid biosynthesis genes are essential in mycobacteria; however, surprisingly, we were able to generate a viable mutant of *MSMEG4722*, the homologue of *Rv2509* in the fast growing *Mycobacterium smegmatis* (Bhatt *et al.*, 2008). Similar to the *C. glutamicum* *NCgl2385* mutant, the *M. smegmatis* *MSMEG4722* mutant produced  $\alpha$ -alkyl,  $\beta$ -keto fatty acyl precursors of mycolic acids (3-oxo-mycolic acid precursors) which were transported and esterified to the arabinogalactan in the cell wall (Bhatt *et al.*, 2008). While the precise knowledge of the substrate for *Rv2509* is lacking, functional analysis of the preceding enzyme in the biosynthetic pathway, Pks13, indicated that while this polyketide synthase primarily catalyses the formation of the  $\alpha$ -alkyl,  $\beta$ -keto fatty acyl precursor, it also contains enzymatic motifs that facilitate the release of the nascent fatty acyl chain and its subsequent transfer to a trehalose residue to produce a trehalose residue esterified with the  $\alpha$ -alkyl,  $\beta$ -keto fatty acyl precursor of mycolic acid (Gavalda *et al.*, 2014). These findings suggest that the trehalose-bound mono 3-oxo-mycolic acid precursor may likely be the substrate for *Rv2509*, which catalyses its conversion to trehalose monomycolate (Fig.1). In this study we set out to do a detailed study of the amino acid sequence of *Rv2509* with the aim of deducing its structure and functional connections, which led to the identification of series of unique sequences, and by proxy, novel structural features. We then probed the role of one distinct structural feature of the mycobacterial mycolyl reductase *in vitro* by deletion analysis. We also probed the

essentiality of *Rv2509* for growth and viability of the slow growing *M. tuberculosis* complex using *Mycobacterium bovis* BCG.

## Results

### Homologues of *M. tuberculosis Rv2509* are found across mycolic acid-producing species.

While mycolic acids are produced by a number of species belonging to the suborder Corynebacterineae, the biosynthesis of mycolates has been studied in detail exclusively in the genera *Mycobacterium* and *Corynebacterium*. While the two genera share a common set of enzymes required for mycolic acid biosynthesis, select differences also exist. For example, while the meromycolate chains are synthesised by a Type-II fatty acid synthase complex in species of *Mycobacterium*, the same function is carried out by a mammalian-like Type-I fatty acid synthase in *Corynebacterium* species (Radmacher *et al.*, 2005). Similarly, differences exist in the transport of mycolic acids between the two genera (Varela *et al.*, 2012, Yang *et al.*, 2014). To investigate whether the reduction of premycolates by an exclusive mycolyl reductase, was an enzymatic process conserved across other mycolic acid producing species, we used the amino acid sequence of *Rv2509* as a source template for a BLASTp search of genome sequences of other known mycolic acid-producing bacteria. Homologues of *Rv2509* were found across the mycolate-producing genera (Fig.2). A phylogenetic analysis of *Rv2509*, its homologues in other mycobacteria and in select mycolic acid producing species revealed that mycolyl reductases from corynebacterial species branched early (Supporting Information Fig.1). Corynebacteria produce relatively short chain mycolic acids (Radmacher *et al.*, 2005). To test whether mycolyl reductases across the mycolata had evolved specificities to accommodate differing chain lengths, we transformed the *M. smegmatis*  $\Delta$ MSMEG4722 strain with a plasmid containing *C. glutamicum* mycolyl reductase gene *NCgl2385* to generate the strain  $\Delta$ MSMEG4722-CNCgl2385. Mature mycolic acid production was restored in the transformed strain indicating that *NCgl2385* could

reduce *M. smegmatis* 3-oxo-mycolic acid precursors and consequently fully complement the mutant *M. smegmatis* strain. This outcome suggested that mycolyl reductases across the mycolata likely did not evolve specificities for differing chain lengths (Fig.3).

### **Predicted structure of *M. tuberculosis* Rv2509**

We attempted to deduce possible functional properties of Rv2509 using a comparative sequence and structural analysis with proteins of known function. Sequence analysis of Rv2509 using the NCBI conserved domain database (Marchler-Bauer *et al.*, 2015) categorises the protein as a part of the DltE family of short-chain dehydrogenases (cluster COG0300), which are in turn members of the cl27753 short-chain dehydrogenase/reductase (SDR) superfamily of NAD(P)-dependent oxidoreductases. This is a diverse group, containing over 80,000 unique sequences (Fujisawa *et al.*, 2003). However, despite this diversity, all SDR-superfamilies display the classical Rossmann fold typical of NAD<sup>+</sup>/NADP<sup>+</sup>-binding proteins and hence can be classified as belonging to the cl21454 family of topologically and structurally related proteins.

Next, we screened the peptide sequence of *Rv2509* against the PDB database to find structural templates for homology modelling. The serine dehydrogenase YdfG (Uniprot ID P39831; PDB ID 3ASU/3ASV) (Yamazawa *et al.*, 2011), had the highest identity score, however structural analysis revealed several gaps in the pairwise alignment, and the C-terminal region was shorter than that observed in *Rv2509*, which necessitated additional template input. Therefore, the structures of the four highest-scoring experimental SDRs (3ASU/3ASV, 1XG5, 4X54 and 4BMV) were used in parallel as a basis for comparative homology modelling of *Rv2509* (Fig.4) using the I-TASSER package (Yang *et al.*, 2015) with specific template selection enabled. Notably, while, three of the templates yielded models with roughly equivalent (confidence) C-scores, the one based on the SDR from *Sphingobium yanoikuyae* (4BMV.pdb; GenBank ID ACB78183.1) stood out with the highest C-score of 0.6, indicating high quality of prediction.

As visualized by the structural alignment (Fig.2) which shows the predicted secondary structure elements of Rv2509 aligned to its corynebacterial homologues, the core structure of the Rv2509 model presents a typical NADP-binding protein Rossmann fold (residues 1-235) (Persson *et al.*, 2003) with 7 parallel  $\beta$ -strands forming a central  $\beta$ -sheet, which is sandwiched between two layers of  $\alpha$ -helices – three on each side (Fig. 4). Although an earlier study has correctly classified Rv2509 within the short-chain reductase (SDR) superfamily and identified crucial sequence signatures such as the potential NAD(H)/NADP(H) binding motif and active site in residues 157-

161 (Lea-Smith *et al.*, 2007), it did not include further structural prediction and analysis, and was limited in terms of sequences used. We therefore sought to expand of this initial analysis by performing a systematic analysis of the Rv2509 orthologues from a wide range of *Corynebacterineae*, so as to represent all known mycolate-producing genera.

We initiated these studies by analysing the co-factor binding site. SDRs can be divided into two large families, ‘Classical’ SDRs which are around 250 amino acid residues long and ‘Extended’ SDRs which contain around 350 residues (Jornvall *et al.*, 1995), exhibiting differences in the arrangement of glycine residues within the co-enzyme binding sites. Coenzyme preference-predictions based on the strong conservation of the signature residues (Persson *et al.*, 2003) classified the Rv2509 into the group of the NADP(H)-binding SDRs, consistent with its mycolyl reductase function. However, it was difficult to categorically allocate Rv2509 into a specific group of SDRs as the predicted NADP-binding site of Rv2509 (G<sub>16</sub>XXQ<sub>19</sub>N<sub>20</sub>I<sub>21</sub>G<sub>22</sub>) displays a unique and novel variation of the (T)-G-X-X-X-G-I-G motif described by Brakoulis and Jackson (Brakoulis & Jackson, 2004) (Fig.2). The variation seems to be restricted to the long-chain mycolate producing species of *Mycobacteriaceae*, as short chain mycolate producing *Corynebacteria* retain the more established G-X-X-X-G-I-G motif. (Fig.2). Key residue positions in the NADP(H) site suggest Rv2509 is a cP2 member of the “Classic” SDR family, according to the classification system of Persson *et al.*, (Brakoulis & Jackson, 2004), which is further reinforced by the presence of two specific discriminator residues, R-R, found in positions 41-42 in both Rv2509 and YdfG from *E. coli* (3ASV.pdb), as these residues are associated with coordination of the phosphate group of the NADP(H) cofactor. Further analysis of Rv2509 suggests that has a Class IV NADP(H) binding site (with consensus [AVIC]-[LIV]-[VIL]-T-G-[ASGC]-X<sub>2</sub>-[GR]-[ILF]-G-X<sub>6</sub>-[LFY] (Hua *et al.*, 2014). Thus, our analysis firmly places the Rv2509 family within the NADP(H) binding sub-family of SDRs.

The active site in SDRs is formed by a tight cleft, lined by  $\alpha$ -helix 4,  $\beta$ -strand 4 and  $\alpha$ -helix 5 (Filling *et al.*, 2002, Oppermann *et al.*, 1997). Rv2509 exhibits the “classical” SDR motif Yx[AS][ST]K in  $\alpha$ -helix 5 (Y157 and K161 being the central catalytic residues), while the GxgxxgSS/T motif in  $\beta$ -strand 4 (G137, S144) and the central residue of  $\alpha$ -helix 4, N116 (Fig.2) are also instantly recognisable. In addition, it is possible that some contact with the substrate may be provided by the  $\beta$ -strand 4.

Rather intriguingly, the consensus residues ‘NNAG’ involved in the stabilization of the central  $\beta$ -sheet of the Rossmann fold in the “Classical” SDRs and located at the core of the

protein, are not fully conserved in *Rv2509*, but take the form of a novel ‘ANAG’ sequence (see Fig.2). Furthermore, this novel sequence is paired with a preceding Cys residue, which is also correlated with the presence of a C-terminal Thr (in position 91 in *M. tuberculosis*). This unique ‘CANAGT’ signature makes *Rv2509* orthologues in *Mycobacteriaceae* instantly recognisable (Fig.2).

In SDR proteins, the substrate binding loop links the sixth  $\beta$ -sheet with the seventh  $\alpha$ -helix, and is one of the larger loops. In the predicted structure of *Rv2509* this loop spans roughly from P<sub>187</sub> to S<sub>214</sub> and appears largely conserved within the mycolic acid producing species (Fig.2). While the loop’s length and position corresponds closely to the length of the substrate binding loop in both 4BMV (P183-E206); and 3ASV (P178-T208), there was little conservation of the loop’s sequence relative to *Rv2509* in these and other divergent SDRs (see ConSurf analysis below (Ashkenazy *et al.*, 2016); Supporting information Fig.2), suggesting, as expected, major differences in the structure of the substrate for *Rv2509*.

The most intriguing finding upon the analysis of the *Rv2509* orthologues was that all of them seem to present a C-terminal domain which seems to be highly specific to the *Corynebacterineae*. It proved also the most difficult to model reliably, as in the majority of the available SDR structural templates, the C-terminal region, appears to be either truncated or disorganized. ConSurf mapping using the multiple sequence alignment shown in Fig. 2 and multiple sequence alignment based on 150 non-redundant sequences of remote SDR homologues revealed that C-terminus of *Rv2509* is uniquely a feature of the subfamily of *Corynebacteriales* (Supporting information Fig.3). Both pairwise and multiple sequence analyses indicated that the C-terminal region of the 4BMV.pdb is the closest related to *Rv2509* and thus can be tentatively used as a template for that region. This was further corroborated that by comparative modelling (Song *et al.*, 2013) and pure *ab-initio* modelling (Raman *et al.*, 2009), as implemented in the NewRobetta server (<http://new.robetta.org>). To validate this hypothesis, we conducted additional *ab initio* secondary structure predictions of *Rv2509* (236-268) using JPred4 (Drozdetskiy *et al.*, 2015) and PSIPRED (Jones, 1999; Buchan and Jones, 2019) Both of these bioinformatics tools supported a high alpha-helical propensity of the C-terminal region of *Rv2509*, suggesting that 4BMV as a reliable template for the region (Supporting information Fig.3). Indeed, the C-terminus of the 4BMV.pdb structure, similarly to *Rv2509*, appears to be extended and is well-structured, presenting an  $\alpha$ -helical organisation.

### **The extended C-terminal of *Rv2509* is essential for function**

Mature mycolic acid production could be restored in the *M. smegmatis*  $\Delta$ *MSMEG4722* mutant on complementation with the *M. tuberculosis* orthologue *Rv2509* (Bhatt *et al.*, 2008). The mutant strain thus provided us with a means of identifying residues and domains required for *Rv2509* function, by assessing the ability of mutated constructs of plasmid-borne *Rv2509* to complement the  $\Delta$ *MSMEG4722* strain by restoring mature mycolic acid biosynthesis. To test if the extended C-terminal of *Rv2509* was required for function, we generated truncated versions of *Rv2509* cloned in the replicative shuttle plasmid pMV261 (Stover *et al.*, 1991). Individual plasmids were then introduced by electroporation into the *M. smegmatis*  $\Delta$ *MSMEG4722* mutant and the transformed strain analysed for restoration of mature mycolic acid biosynthesis. Deletion of 14 amino acid residues (Ala<sub>255</sub>-Ser<sub>268</sub>) from the C-terminal of *Rv2509* resulted in a complete loss of function as the strains transformed with the truncated construct failed to restore mature mycolic acid biosynthesis in the  $\Delta$ *MSMEG4722* strain (Fig.5;  $\Delta$ *MSMEG4722*-CRvD1). A construct with a shorter truncation (Tyr<sub>262</sub>-Ser<sub>268</sub> deleted) also failed to complement the  $\Delta$ *MSMEG4722* strain (Fig.5;  $\Delta$ *MSMEG4722*-CRvD2). Expression of *Rv2509* and truncated constructs could be detected by RT-PCR confirming that the loss of *Rv2509* function was linked to deletion of the C-terminal residues (Supporting Information Fig.4). These results demonstrated that the C-terminus of *Rv2509* spanning terminal residues Ala<sub>255</sub>-Ser<sub>268</sub> was critical for function. The findings suggested an essential role for the unique, C-terminal extension in the activity of *Rv2509*.

To further query whether this C-terminal extension played a role in either protein stability, protecting *Rv2509* from degradation, or in oligomerisation, we expressed N-terminal His-tagged versions of *Rv2509* and *Rv2509*-D1 (*Rv2509* with Ala<sub>255</sub>-Ser<sub>268</sub> deleted) in *Escherichia coli*. Western blot analysis of native polyacrylamide gels showed there were no patterns of degradation detected in *Rv2509*-D1 indicating that the loss of function due to the deletion of this region was not related to protein stability (Supporting Information Fig.5). Furthermore, we did not observe additional slower migrating bands indicating dimerization (or oligomerisation) of full length or truncated *Rv2509*. Furthermore, crosslinking *Rv2509* with glutaraldehyde did not reveal additional slower migrating bands suggesting that *Rv2509* did not oligomerise in solution (Supporting Information Fig.5).

### ***Rv2509* is an essential mycobacterial gene**



Given the non-essentiality of the mycolyl reductase gene in *M. smegmatis*, and prediction of transposon-site hybridisation screens of a slow growing *M. tuberculosis* Rv2509 mutant, we aimed to generate a null mutant of Rv2509 in *M. tuberculosis* using Specialized Transduction (Bardarov *et al.*, 2002). Such a mutant would also allow us to test the role of the mycolyl reductase in virulence using macrophage and mouse models of infection. However, repeated attempts failed to generate transductants, suggesting that Rv2509 was essential for growth in *M. tuberculosis*. Indeed, subsequent transposon mutagenesis screens utilising deep sequencing predicted Rv2509 to be essential for *in vitro* growth of *M. tuberculosis* (Griffin *et al.*, 2011). We then utilised a promoter replacement strategy to demonstrate the essentiality of Rv2509 in slow growing mycobacterial vaccine strain *Mycobacterium bovis* BCG. The *P<sub>myc1</sub>* promoter from *M. smegmatis* engineered to contain four *tetO* operator sites (Korte *et al.*, 2016) was inserted immediately upstream of the start codon of BCG2529, the Rv2509 homologue in *M. bovis* BCG-Pasteur, using Specialised Transduction generating the strain BCG::P<sub>Tet</sub>-Mb2537. Controlled gene expression of the BCG2529 gene was achieved using a plasmid borne synthetic gene (*rev-tetR*) derived from Tn10 *tetR* encoding a mutated TetR protein with reversed binding affinity to *tetO* sites upon binding of tetracycline. Addition of anhydrotetracycline (ATc) results in loss of expression.

Addition of ATc resulted in loss of growth of BCG::P<sub>Tet</sub>-BCG2529 in broth indicating that expression of the gene encoding the mycolyl reductase was essential for *in vitro* growth of slow growing mycobacteria (Fig.6A). Furthermore, while concentrated cultures of BCG::P<sub>Tet</sub>-BCG2529 showed confluent growth on 7H10 agar plates, no confluent growth was observed on plates containing ATc (Fig.6B). We did observe a few scattered colonies on the ATc-containing plates and these are likely to be suppressors allowing leaky expression of P<sub>Tet</sub> driven BCG2529. These results demonstrated that the mycolyl reductase was essential for growth and viability of slow growing mycobacteria.

### **Conditional depletion of the mycolyl reductase results in loss of mature mycolic acids**

The *M. smegmatis* mycolyl reductase mutant was unable to synthesise fully mature mycolic acids, and instead produced 3-oxo-mycolate precursors that were found esterified to the cell wall arabinogalactan and to trehalose (Bhatt *et al.*, 2008). These labile precursors gave rise to palmitone derivatives following alkali treatment, which were visualised on Thin Layer Chromatography (TLC) plates migrating close to the solvent front in a solvent system used for separating methyl esters of mycolic acids and fatty acids (FAMES and MAMES) (Bhatt *et al.*,

2008). To test the effects of conditional depletion of the mycolyl reductase we labelled cultures of BCG::P<sub>Tet</sub>-BCG2529 with <sup>14</sup>[C]-acetic acid following grown in the presence and absence of ATc, and extracted MAMES from the cultures. TLC analysis of the extracted MAMES showed a decrease in the levels of  $\alpha$  and keto MAMES, and an accumulation of the degradation products derived from 3-oxo-mycolate precursors close to the solvent front (Fig.7). These results showed that, as in *M. smegmatis*, loss of the mycolyl reductase function led to the loss of mature mycolic acid production and accumulation of the premature 3-oxo-mycolate precursors in slow grow mycobacteria.

## Discussion

As summarised above, the predicted structure of the Rv2509 displays a Rossmann fold with 7 parallel  $\beta$ -strands forming a central beta-sheet, which is sandwiched between two layers of alpha-helices – three on each side (Fig. 4). While largely consistent with the “classical” family of SDRs (Persson *et al.*, 2003), in *Mycobacteriaceae*, the NADP(H) binding motif (G-x-x-Q-N-I-G), alongside the core stabilisation consensus (C-A-N-A-G-T) are significantly divergent from the consensus, to instantly identify this group of SDRs, and in our view is distinct enough to be separated into a distinct family.

While we currently lack knowledge about the exact mode of co-factor and ligand binding within the Rv2509 family, some information could be inferred from the structurally related FabG family of  $\beta$ -ketoacyl reductases which are classified as ‘Complex family SDRs’ (Persson *et al.*, 2003). Given that the structural overlay of Rv2509 models with existing ligand-bound FabG members yields an RMSD of around 1 Å over the C-alpha backbone atoms, and that FabG share common active site residues with *Rv2509* it is tempting to suggest that these structures share common ligand-binding and NADP(H) orientations towards the active site residues with that observed in the aceto-acyl-CoAs. Furthermore, the predicted 3D structure of Rv2509 also revealed a unique  $\alpha$ -helical C-terminal extension, which we subsequently showed to be essential for function - constructs with deletions in the C-terminal failed to complement (restore mycolate biosynthesis) in the *M. smegmatis* mutant.

The role of this C-terminal domain is difficult to assign with certainty, however in related SDRs similar C-terminal extensions play a number of functions. In the NADP(H)-dependent tetrameric serine-dehydrogenase YdfG from *E.coli* (3ASV.pdb) (Yamazawa *et al.*, 2011), the C-terminal region protrudes from the main core of the protein and is required for tetramer formation

also plays a role in the substrate binding via stabilisation of the substrate loop. It is also notable that in FabG family of related SDRs (see below) also organise as tetrameric assemblies with the participation of their C-terminal domains (Javidpour *et al.*, 2014). Furthermore, in the mycobacterial FabG4 dimer, Arg146 and Arg445 of one protomer interact with the C-terminus of the second protomer and play an essential role in substrate association and catalysis (Dutta *et al.*, 2013). However, our experiments with purified full length and truncated Rv2509 do not seem to suggest that the mycolyl reductase forms oligomers.

While it isn't in itself surprising that Rv2509 presents a conserved SDR fold, our analysis has identified it for the first time decisively as a member of the NADPH-binding family and indicated that it possesses a number of unique features. Taken together the unique signatures of the NADPH-binding site, core stabilisation domains, as well as C-terminal domain extension possibly merit its inclusion into a distinct family.

Our previous studies with *M. smegmatis* suggested that the mycolyl reductase would likely not be essential for growth in *M. tuberculosis* and related slow-growing mycobacteria. However, this work demonstrated that gene encoding the reductase, *Rv2509*, was an essential gene required for growth of slow growing mycobacteria in laboratory media. This, along with the identification of unique domains in the predicted structure of Rv2509 highlights its potential as a novel anti-TB drug target.

## **Experimental Procedures**

### **Growth conditions, strains, phages and plasmids**

*M. smegmatis* mc<sup>2</sup>155 and derived strains was cultured in Tryptic Soy Broth (TSB) or on TSB-agar, *M. bovis* BCG strains were grown in 7H9 broth or on 7H10 agar plates. *Escherichia coli* strains were grown in LB broth or LB agar. Hygromycin was used at a concentration of 100 mg ml<sup>-1</sup> for *E. coli* and for mycobacterial strains. Kanamycin was used at a concentration of 25 mg ml<sup>-1</sup> and 50 mg ml<sup>-1</sup> for mycobacteria and *E. coli* respectively. Recombinant mycobacterial strains, phages and plasmid constructs used in this study listed and described in Table.1.

### **Sequence analysis and homology modelling**

The multiple sequence alignments (MSA) were prepared using MAFFT and NJ/UPGMA phylogeny algorithms as implemented in MAFFT v.7 server (<https://mafft.cbrc.jp/>) (Kato *et al.*,

2002)) and the resulting phylogenetic trees were visualized using Archaeopteryx (Han & Zmasek, 2009). Multiple sequence alignment visualisation and structural annotations were performed using Esript 3 (Gouet *et al.*, 2003). For the purposes of homology modelling, we employed I-TASSER (Yang *et al.*, 2015) with assignment of templates. Outputs were independently corroborated by *ab-initio* modelling using the NewRobetta server (<http://new.robetta.org>). Sequence conservation analysis was done using ConSurf (Ashkenazy *et al.*, 2016) and visualised using Pymol (The PyMOL Molecular Graphics System, Version 1.71 Schrödinger, LLC) with additional local manual refinement and structural superposition performed in Coot (Emsley *et al.*, 2010). Secondary structure predictions for the C-terminal domain performed with JPred4 (Cole *et al.*, 2008) and PSIPRED (Buchan *et al.*, 2013).

### **Expression, purification and analysis of Rv2509 and its C-terminal truncated derivative**

Rv2509 was cloned in the expression vector pET28a by PCR from *M. tuberculosis* H37Rv genomic DNA using the primer pair Pet28a\_Rv2509\_F2 (5'-TTTACATATGCCGATACCCGCGCCC-3') and Pet28a\_Rv2509\_R1 (5'-ATTAAAGCTTCTAGCTGCCCCCAA GCCTCTTG-3'). Similarly, the C-terminal truncated version of Rv2509 (missing the last 14 amino acids) was PCR amplified using the primer pair Pet28a\_Rv2509\_F2 (5'-TTTACATATGCCGATACCCGCGCCC-3') and Pet28a\_Rv2509\_delR1 (5'-ATTAAAGCTTCTACACGATGGCGCGCGGAGC-3'). The constructs were co-transformed into BL21 cells with *M. tuberculosis* chaperone GroES 60.2. Cultures were seeded 1:100 with an overnight starter culture and grown at 37°C in terrific broth supplemented with ampicillin (100µg/mL) and kanamycin (50µg/mL) to OD600 = 0.6 and induced with 0.5mM IPTG overnight at 37°C or 16°C respectively. Cells were harvested at 4,000rpm at 4°C, washed with PBS and frozen at -80°C until further use. Full length protein was purified on cobalt-IMAC in buffer 50mM NAH2PO4 pH 7.5, 500mM NaCl, 10% glycerol and truncated in the same buffer composition at pH 7.0. The eluted fraction of 200mM imidazole was taken for native-PAGE analysis.

A native polyacrylamide gel with the separated samples was transferred to a Hybond™ nitrocellulose membrane (GE Healthcare Life Sciences) which was then blocked in 5% milk 20mM Tris pH7.5, 150mM NaCl TBS 0.05% Tween for 1h at room temperature followed by incubation with 0.1% milk TBS-tween with penta-his antibody BSA free (Qiagen). Following washing with TBS-tween the membrane was incubated with 0.1% milk TBS-tween with anti-mouse IgG-Alkaline Phosphatase (Sigma Aldrich) for 30 min at room temperature, washed with

TBS-tween and a final wash with TBS. The membrane was stained for 5 min with BCIP/NBT (Sigmafast™) in water for visualization.

Cross linking of full length Rv2509 was carried out in 50mM NaH<sub>2</sub>PO<sub>4</sub> pH 7.5, 10% glycerol at final concentration of 0.1% glutaraldehyde and allowed to progress at room temperature for 2 minutes. The reaction was halted with 1M Tris pH 8 and subsequently run on SDS-PAGE for visualisation using SYPRO-ruby (Molecular Probes).

### **Conditional depletion of the Rv2509 homologue *BCG2529* in *M. bovis* BCG-Pasteur**

For establishing regulated expression of the *BCG2529* gene, a synthetic gene cassette (*hyg-Pmyc1-4XtetO*) comprising a hygromycin resistance gene and the *Pmyc1* promoter from *M. smegmatis* engineered to contain four *tetO* operator sites (Korte *et al.*, 2016) was inserted immediately upstream of the *BCG2529* start codon in *M. bovis* BCG-Pasteur. Targeted gene knock-in was achieved by specialized transduction employing temperature-sensitive mycobacteriophages essentially as described previously (Korte *et al.*, 2016). Briefly, for generation of an allelic exchange construct for site-specific insertion of the *hyg-Pmyc1-4XtetO* cassette in *M. bovis* BCG-Pasteur, upstream- and downstream DNA regions flanking the *BCG2529* start codon were amplified by PCR employing the oligonucleotide pairs *Mb2537\_LL* 5'-TTTTTCCATAAATTGGAACCGCTACCTGACATGAAACCC-3' and *Mb2537\_LR* 5'-TTTTTCCATTTCTTGGGCCGATGTTCTGCGAAGCCCCGG-3' as well as *Mb2537\_RL* 5'-TTTTTCCATAGATTGGATGCCGATACCCGCGCCCAGCCC-3' and *Mb2537\_RR* 5'-TTTTTCCATCTTTTGGCGGTGTGGGAGGAGATACTCAAG-3', respectively. The gene *BCG2528c* is localized in antilinear orientation upstream of *BCG2529* with only 84 bp between the start codons of both genes. Since it thus could not be excluded that the *Mb2536c* promoter region might overlap with the *BCG2529* coding region, the upstream flanking region contained a duplication of 66 bp of the 5'-end of *BCG2529* to conserve 150 bp in front of *BCG2529* that likely comprised its promoter. Subsequently, the upstream and downstream flanks were digested with *Van91I* (restriction sites underlined), and ligated with *Van91I*-digested pcRv1327c-4XtetO vector arms (Korte *et al.*, 2016). The resulting knock-in plasmid was then linearized with *PacI* and cloned and packaged into the temperature-sensitive phage phAE159 (Jain *et al.*, 2014), yielding a knock-in phage which was propagated in *M. smegmatis* at 30°C. Allelic exchange in *M. bovis* BCG-Pasteur using the knock-in phage at the non-permissive temperature of 37°C was achieved by specialized transduction using hygromycin (50 mg/l) for selection, resulting in site-specific

insertion of the *hyg-PmycI-4XtetO* cassette. The obtained *M. bovis* c-BCG\_2529-4XtetO knock-in mutant was verified by diagnostic PCR of genomic DNA using the oligonucleotide pair BCG\_2529\_L\_fw 5' GTCAGGTAGACGGAGAACAC-3' and BCG\_2529\_L\_rev 5' AGCTCACCGCGCAGAGATTC-3' binding outside the allelic exchange substrates used to generate this mutant and subsequent sequencing of PCR products (Supporting Information Fig.6).

For achieving controlled gene expression of the *Mb2537* gene, a synthetic gene (*rev-tetR*) derived from Tn10 *tetR* encoding a mutated TetR protein with reversed binding affinity to *tetO* sites upon binding of tetracycline (Klotzsche *et al.*, 2009) was heterologously expressed in the knock-in mutant by electroporation of the episomal *E. coli*-mycobacterium shuttle plasmid pMV261::*rev-tetR*-RBS-D providing constitutive *rev-tetR* gene expression from the HSP60 promoter in mycobacteria using solid medium containing 50 mg/l hygromycin and 20 mg/l kanamycin for selection (Famulla *et al.*, 2016). This yielded the conditional mutant BCG-Pasteur c- *BCG2529-4×tetO* pMV261::*rev-tetR*-RBS-D (referred to as BCG::P<sub>Tet</sub>- *BCG2529*) allowing silencing of the *BCG2529* gene in presence of anhydrotetracycline (ATc).

For silencing experiments, precultures of the BCG::P<sub>Tet</sub>- *BCG2529* strain were first grown in liquid medium containing 0.3 µg/ml ATc, before these precultures were used to inoculate test cultures containing 0 - 5 µg/ml ATc. Growth in 96-well microtiter plates was quantified after incubation at 37°C for five days by adding 10% of a 100 µg/mL resazurin solution. After further incubation at ambient temperature for 16 h, cells were fixed for 30 min by formalin addition (5%, v/v, final concentration), and fluorescence was quantified using a microplate reader (excitation 540 nm, emission 590 nm). For growth on 7H10 agar, non-permissive conditions were obtained by adding ATc to a final concentration of 5 µg/ml.

### Functional complementation studies

Deletion constructs for *Rv2509* were generated by PCR amplification of selected regions of *Rv2509* using pMV261-*Rv2509* as a template, incorporating a premature stop codon in the reverse primer. The plasmid constructs generated in this manner were called pMV261-*Rv2509*D1 and pMV261-*Rv2509*D2, encoding a truncated protein missing the C-terminal residues Ala<sub>255</sub>-Ser<sub>268</sub> and Tyr<sub>262</sub>-Ser<sub>268</sub> respectively. The *C. glutamicum* *Rv2509* orthologue, *NCgl2385* was PCR amplified using *C. glutamicum* genomic DNA as a template and the PCR product was cloned into pMV261 using primer incorporated BamHI and EcoRI sites generating the plasmid pMV261-*NCgl2385*. Plasmid clones were electroporated into the *M. smegmatis*  $\Delta$ MSMEG4722 mutant

strain using previously described protocols (Snapper *et al.*, 1990). Kanamycin resistant transformants were cultured in TSB and FAMES and MAMES were extracted from the cell pellets and separated by TLC as previously described (Vilcheze & Jacobs, 2007).

### **RT-PCR of *Rv2509* transcripts**

*M. smegmatis* strains were grown to mid-log phase, and RNA was extracted by acid phenol and chloroform extraction. The purified RNA was treated with DNase and converted to cDNA, which was used as a template for quantitative PCR. PCR products were run on a 1% agarose gel to detect the expression of *Rv2509*. The gene *sigA* was monitored as a constitutively expressed gene. The primers used for PCR amplification were Rv2509RT-f (5'-ACAAGTACCGCGTCACGGTC -3') and Rv2509RT-r (5'-AAGTCCGGCACC AGCTTCTC-3') for amplification of non-deleted sections of *Rv2509*.

### **Acknowledgements**

CC is funded by a BBSRC PhD studentship provided by the BBSRC Midlands Integrative Biosciences Training Partnership (MIBTP). AB acknowledges previous grant support from the Medical Research Council (UK). RK acknowledges financial support from the Jürgen Manchot Foundation.

### **Author Contributions**

AB, RK and VB designed the study. AJ, CC, AS, SS, MH, RM conducted experiments and acquired data. AJ, CC, AS, SS, MH, RM, AB, RK and VB interpreted the data. AB, RK and VB wrote the manuscript.

### **Abbreviated Summary**

The *Mycobacterium tuberculosis* gene *Rv2509* encodes a reductase enzyme involved in the last step of synthesis of mycolic acids, a key cell wall lipid. By generating a predicted model for the protein encoded by *Rv2509*, we show that the enzyme has distinct structural motifs including a distinct C-terminal region, which we show by complementation studies to be vital for function.

The gene is also essential for growth and viability underlining its potential as a new drug target.

## Figure Legends

**Fig.1** Proposed reaction catalysed by Rv2509. The putative substrate is trehalose monomycolate and the product is trehalose monomycolate. The examples shown here is that of an  $\alpha$ -mycolic acid from *M. tuberculosis*.

**Fig.2** A structural alignment of the predicted Rv2509 structure from *M. tuberculosis*, with annotated secondary structure elements (top) with the sequences of the Mycobacterial and Corynebacterial homologues highlighting the unique sequence motifs of the Rv2509 subfamily of SDRs. The GxxQNIG motif along with RR pair identifies a novel NADP(H) binding consensus, limited to *Mycobacteriaceae* (highlighted in yellow). GxxSGIG is the wider consensus seen in the more remotely related Corynebacteria. Similarly, the unique Rossmann core-stabilisation signature CANAGT reported here is also restricted to the *Mycobacteriaceae* (highlighted in yellow), and reverts to the more common (C)NNAG(I/F) in the more remote homologues. Active site residues N-G-S-Y-K in alpha-helix 4 and alpha-helix 5 are preserved throughout the family, however, a unique additional tyrosine residue can be seen to be present in *Mycobacteriaceae* (highlighted in yellow). The C-terminal extension is also uniquely conserved across the wider Corynebacterial species.

**Fig.3** Complementation of the *M. smegmatis*  $\Delta$ MSMEG4722 mutant with the corynebacterial mycolyl reductase. TLC analysis of  $^{14}$ C-labelled mycolic acid methyl esters (MAMEs) isolated from different strains.  $\alpha$ ;  $\alpha$  MAMEs,  $\alpha'$ ;  $\alpha'$  MAMEs, e; epoxy MAMEs, O; origin, D, degradation



products derived from 3-oxo-mycolate precursors. Solvent system; Petroleum ether:acetone (95:5, v:v).

**Fig.4** Visualisation of the predicted structure of *M. tuberculosis* Rv2509 monomer. Two different orientations are given; side-view and top-down view towards the co-enzyme binding site. The protein is coloured as a rainbow from blue (N-terminus) to red (C-terminus) and main secondary structure elements are labelled, as well as the predicted C-terminal extension domain is circled in a red oval. The location of key sequence and structural motifs discussed in the text is also provided.

**Fig.5** TLC analysis of <sup>14</sup>C-labelled MAMEs isolated from strains of *M. smegmatis* ΔMSMEG4722 mutant transformed with the different deletion constructs of *Rv2509*. α; α MAMEs, α'; α' MAMEs, e; epoxy MAMEs, O; origin, D, degradation products derived from 3-oxo-mycolate precursors. Solvent system; Petroleum ether:acetone (95:5, v:v).

**Fig.6** Conditional depletion of the mycolyl reductase in BCG::P<sub>Tet</sub>- *BCG2529* leads to loss of growth. A; growth in different concentrations of ATc in 7H9 broth, B; growth on 7H10 agar containing 5 μg/ml ATc. Control strain; BCG transformed with pMV261::rev-tetR-RBS-D.

**Fig.7** TLC analysis of <sup>14</sup>C-labelled MAMEs isolated from BCG::P<sub>Tet</sub>- *BCG2529* strain. α; α MAMEs, keto; keto MAMEs, O; origin, D, degradation products derived from 3-oxo-mycolate precursors. Solvent system; Petroleum ether:acetone (95:5, v:v).

#### **Supporting information Fig.1**

Phylogenetic analysis based on the sequences of *Rv2509* homologues and showing the positions of some of the structural templates used for their modelling. Length of branches corresponds to evolutionary distance. Note the tight clustering of *Mycobacteriaceae* *Rv2509* homologues.

#### **Supporting information Fig. 2**

Consurf mapping of sequence conservation onto the predicted *Rv2509* structure.

A. Mapping of conservation within the Corynebacterial sequences presented in the MSA shown in Figure 2, highlights the high level of conservation of both substrate loop and C-terminal extension indicative of preservation of functional residues associated with these structural elements. B.

Mapping of conservation across wider SDR family, covering 150 unique sequences from UNIREF database, shows that both C-terminal extension and the substrate binding loop are highly variable outside of the immediate corynebacterial grouping. Note that the active site residues and coenzyme binding residues – (deep purple in the core) are instantly recognisable and at conserved positions across SDR family.

### **Supporting information Fig. 3**

Secondary structure prediction for the C-terminal domain of Rv2509 suggests strongly alpha-helical propensity (A) Output from PSIPRED and (B) JPRED4 output incorporating JNET.

### **Supporting information Fig.4**

Detection of *Rv2509* transcripts in *M. smegmatis* strains transformed with C-terminal deletion constructs by RT-PCR. (A) shows products for *Rv2509* while (B) shows products obtained for control (*sigA*).

### **Supporting information Fig.5**

Analysis of full length and C-terminal truncated Rv2509 by Native PAGE (A) and glutaraldehyde crosslinking (B).

### **Supporting information Fig.6**

Generation of a conditional *M. bovis* BCG-Pasteur *BCG2529* mutant.

(A) Organization of the *BCG2529* locus in *M. bovis* BCG-Pasteur wild-type as well as in a *BCG2529* conditional mutant. The sizes of relevant fragments as well as the location of the oligonucleotides used for diagnostic PCR are indicated. WT, wild-type; fw, oligonucleotide *BCG\_2529\_L\_fw*; rev, oligonucleotide *BCG\_2529\_L\_rev*; hyg, hygromycin resistance gene. (B) Diagnostic PCR of genomic DNA using oligonucleotides binding to the position indicated in A, showing the insertion of a synthetic gene cassette (*hyg-Pmyc1-4XtetO*) that contained a duplication of 66 bp of the 5'-end of *BCG2529*.

## References

- Ashkenazy, H., S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko & N. Ben-Tal, (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**: W344-350.
- Bardarov, S., S. Bardarov, Jr., M.S. Pavelka, Jr., V. Sambandamurthy, M. Larsen, J. Tufariello, J. Chan, G. Hatfull & W.R. Jacobs, Jr., (2002) Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in *Mycobacterium tuberculosis*, *M. bovis* BCG and *M. smegmatis*. *Microbiology* **148**: 3007-3017.
- Bhatt, A., A.K. Brown, A. Singh, D.E. Minnikin & G.S. Besra, (2008) Loss of a mycobacterial gene encoding a reductase leads to an altered cell wall containing beta-oxo-mycolic acid analogs and accumulation of ketones. *Chem Biol* **15**: 930-939.
- Brakoulias, A. & R.M. Jackson, (2004) Towards a structural classification of phosphate binding sites in protein-nucleotide complexes: an automated all-against-all structural comparison using geometric matching. *Proteins* **56**: 250-260.
- Buchan, D.W.A. & Jones, D.T. (2019) The PSIPRED Protein Analysis Workbench: 20 years on *Nucleic Acids Res* **47**:402-407.
- Drozdetskiy, A., Cole, C., Procter, J., Barton, G.J. (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* **43**:389-394.

- Dutta, D., S. Bhattacharyya, A. Roychowdhury, R. Biswas & A.K. Das, (2013) Crystal structure of hexanoyl-CoA bound to beta-ketoacyl reductase FabG4 of *Mycobacterium tuberculosis*. *Biochem J* **450**: 127-139.
- Emsley, P., B. Lohkamp, W.G. Scott & K. Cowtan, (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**: 486-501.
- Famulla, K., P. Sass, I. Malik, T. Akopian, O. Kandror, M. Alber, B. Hinzen, H. Ruebsamen-Schaeff, R. Kalscheuer, A.L. Goldberg & H. Brotz-Oesterhelt, (2016) Acyldepsipeptide antibiotics kill mycobacteria by preventing the physiological functions of the ClpP1P2 protease. *Mol Microbiol* **101**: 194-209.
- Filling, C., K.D. Berndt, J. Benach, S. Knapp, T. Prozorovski, E. Nordling, R. Ladenstein, H. Jornvall & U. Oppermann, (2002) Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *J Biol Chem* **277**: 25677-25684.
- Fujisawa, H., S. Nagata & H. Misono, (2003) Characterization of short-chain dehydrogenase/reductase homologues of *Escherichia coli* (YdfG) and *Saccharomyces cerevisiae* (YMR226C). *Biochim Biophys Acta* **1645**: 89-94.
- Gande, R., K.J. Gibson, A.K. Brown, K. Krumbach, L.G. Dover, H. Sahm, S. Shioyama, T. Oikawa, G.S. Besra & L. Eggeling, (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.
- Gavalda, S., F. Bardou, F. Laval, C. Bon, W. Malaga, C. Chalut, C. Guilhot, L. Mourey, M. Daffe & A. Quemard, (2014) The polyketide synthase Pks13 catalyzes a novel mechanism of lipid transfer in mycobacteria. *Chem Biol* **21**: 1660-1669.
- Gouet, P., X. Robert & E. Courcelle, (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* **31**: 3320-3323.
- Griffin, J.E., J.D. Gawronski, M.A. Dejesus, T.R. Ioerger, B.J. Akerley & C.M. Sassetti, (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog* **7**: e1002251.
- Han, M.V. & C.M. Zmasek, (2009) phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* **10**: 356.

- Hua, Y.H., C.Y. Wu, K. Sargsyan & C. Lim, (2014) Sequence-motif detection of NAD(P)-binding proteins: discovery of a unique antibacterial drug target. *Sci Rep* **4**: 6471.
- Jain, P., T. Hsu, M. Arai, K. Biermann, D.S. Thaler, A. Nguyen, P.A. Gonzalez, J.M. Tufariello, J. Kriakov, B. Chen, M.H. Larsen & W.R. Jacobs, Jr., (2014) Specialized transduction designed for precise high-throughput unmarked deletions in *Mycobacterium tuberculosis*. *MBio* **5**: e01245-01214.
- Javidpour, P., J.H. Pereira, E.B. Goh, R.P. McAndrew, S.M. Ma, G.D. Friedland, J.D. Keasling, S.R. Chhabra, P.D. Adams & H.R. Beller, (2014) Biochemical and structural studies of NADH-dependent FabG used to increase the bacterial production of fatty acids under anaerobic conditions. *Appl Environ Microbiol* **80**: 497-505.
- Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Bio.* **292**:195-202.
- Jornvall, H., B. Persson, M. Krook, S. Atrian, R. Gonzalez-Duarte, J. Jeffery & D. Ghosh, (1995) Short-chain dehydrogenases/reductases (SDR). *Biochemistry* **34**: 6003-6013.
- Katoh, K., K. Misawa, K. Kuma & T. Miyata, (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059-3066.
- Klotzsche, M., S. Ehrt & D. Schnappinger, (2009) Improved tetracycline repressors for gene silencing in mycobacteria. *Nucleic Acids Res* **37**: 1778-1788.
- Korte, J., M. Alber, C.M. Trujillo, K. Syson, H. Koliwer-Brandl, R. Deenen, K. Kohrer, M.A. DeJesus, T. Hartman, W.R. Jacobs, Jr., S. Bornemann, T.R. Ioerger, S. Ehrt & R. Kalscheuer, (2016) Trehalose-6-Phosphate-Mediated Toxicity Determines Essentiality of OtsB2 in *Mycobacterium tuberculosis* In Vitro and in Mice. *PLoS Pathog* **12**: e1006043.
- Lea-Smith, D.J., J.S. Pyke, D. Tull, M.J. McConville, R.L. Coppel & P.K. Crellin, (2007) The reductase that catalyzes mycolic motif synthesis is required for efficient attachment of mycolic acids to arabinogalactan. *J Biol Chem* **282**: 11000-11008.
- Marchler-Bauer, A., M.K. Derbyshire, N.R. Gonzales, S. Lu, F. Chitsaz, L.Y. Geer, R.C. Geer, J. He, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng & S.H. Bryant, (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res* **43**: D222-226.
- Marrakchi, H., M.A. Laneelle & M. Daffe, (2014) Mycolic acids: structures, biosynthesis, and beyond. *Chem Biol* **21**: 67-85.

- Nataraj, V., C. Varela, A. Javid, A. Singh, G.S. Besra & A. Bhatt, (2015) Mycolic acids: deciphering and targeting the Achilles' heel of the tubercle bacillus. *Mol Microbiol* **98**: 7-16.
- Oppermann, U.C., C. Filling, K.D. Berndt, B. Persson, J. Benach, R. Ladenstein & H. Jornvall, (1997) Active site directed mutagenesis of 3 beta/17 beta-hydroxysteroid dehydrogenase establishes differential effects on short-chain dehydrogenase/reductase reactions. *Biochemistry* **36**: 34-40.
- Persson, B., Y. Kallberg, U. Oppermann & H. Jornvall, (2003) Coenzyme-based functional assignments of short-chain dehydrogenases/reductases (SDRs). *Chem Biol Interact* **143-144**: 271-278.
- Portevin, D., C. De Sousa-D'Auria, C. Houssin, C. Grimaldi, M. Chami, M. Daffe & C. Guilhot, (2004) A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms. *Proc Natl Acad Sci U S A* **101**: 314-319.
- Radmacher, E., L.J. Alderwick, G.S. Besra, A.K. Brown, K.J. Gibson, H. Sahm & L. Eggeling, (2005) Two functional FAS-I type fatty acid synthases in *Corynebacterium glutamicum*. *Microbiology* **151**: 2421-2427.
- Raman, S., Vernon, R., Thompson, J., Tyka, M., Sadreyev, R., Pei, J., Kim, D., Kellogg, E., DiMaio, F., Lange, O., Kinch, L., Sheffler, W., Kim, B.H., Das, R., Grishin, N.V., Baker, D. (2009) Structure prediction for CASP8 with all-atom refinement using Rosetta *Proteins* **9**:89-99.
- Snapper, S.B., R.E. Melton, S. Mustafa, T. Kieser & W.R. Jacobs, Jr., (1990) Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* **4**: 1911-1919.
- Song, Y., DiMaio, F., Wang, R.Y., Kim, D., Miles, C., Brunette, T., Thompson, J., Baker, D. (2013) High-resolution comparative modeling with RosettaCM *Structure* **21**:1735-1742.
- Stover, C.K., V.F. de la Cruz, T.R. Fuerst, J.E. Burlein, L.A. Benson, L.T. Bennett, G.P. Bansal, J.F. Young, M.H. Lee, G.F. Hatfull & et al., (1991) New use of BCG for recombinant vaccines. *Nature* **351**: 456-460.
- Varela, C., D. Rittmann, A. Singh, K. Krumbach, K. Bhatt, L. Eggeling, G.S. Besra & A. Bhatt, (2012) MmpL genes are associated with mycolic acid metabolism in mycobacteria and corynebacteria. *Chem Biol* **19**: 498-506.

Vilcheze, C. & W.R. Jacobs, (2007) Isolation and analysis of Mycobacterium tuberculosis mycolic acids. *Curr Protoc Microbiol* **Chapter 10**: Unit 10A 13.

Yamazawa, R., Y. Nakajima, K. Mushiake, T. Yoshimoto & K. Ito, (2011) Crystal structure of serine dehydrogenase from Escherichia coli: important role of the C-terminal region for closed-complex formation. *J Biochem* **149**: 701-712.

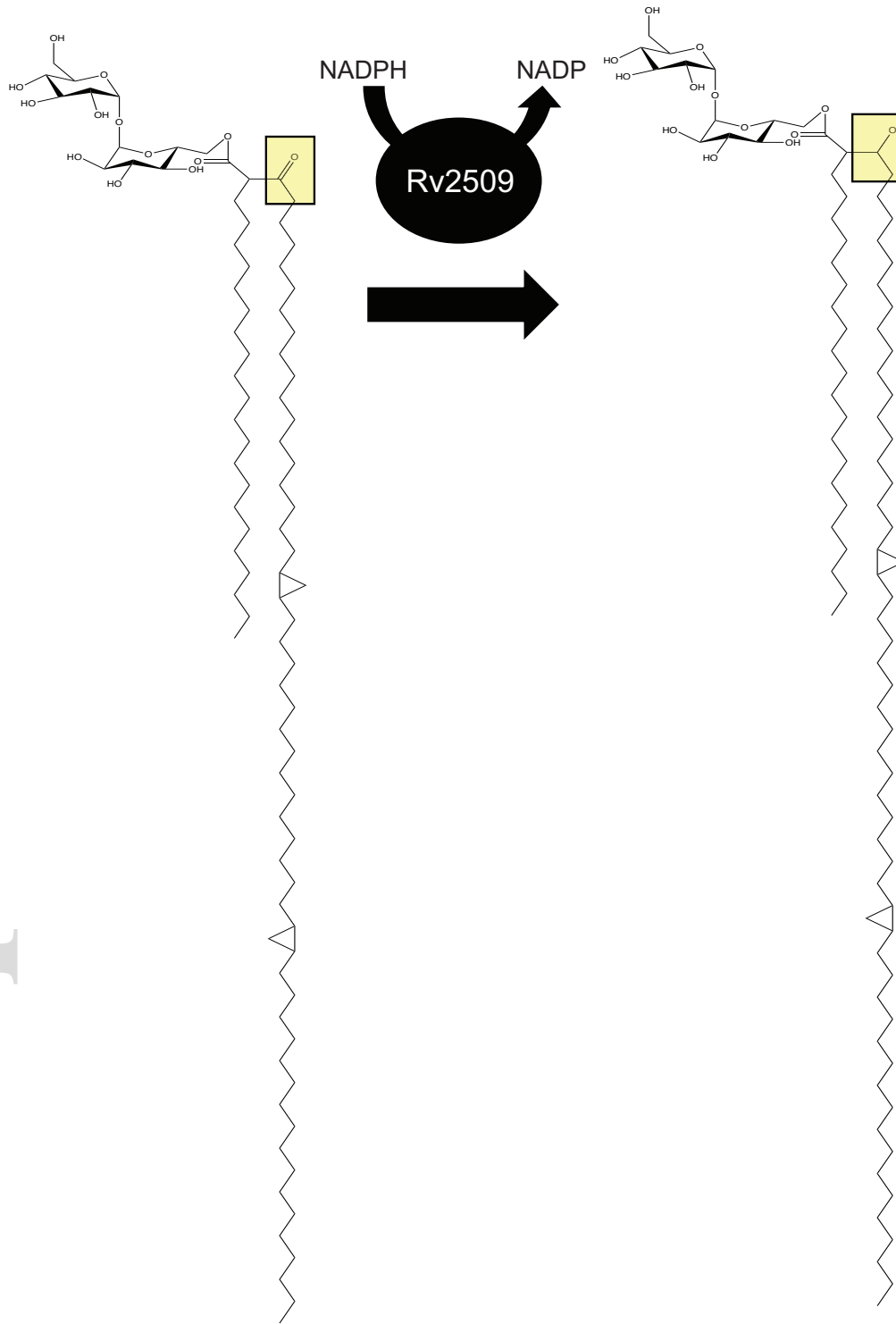
Yang, J., R. Yan, A. Roy, D. Xu, J. Poisson & Y. Zhang, (2015) The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**: 7-8.

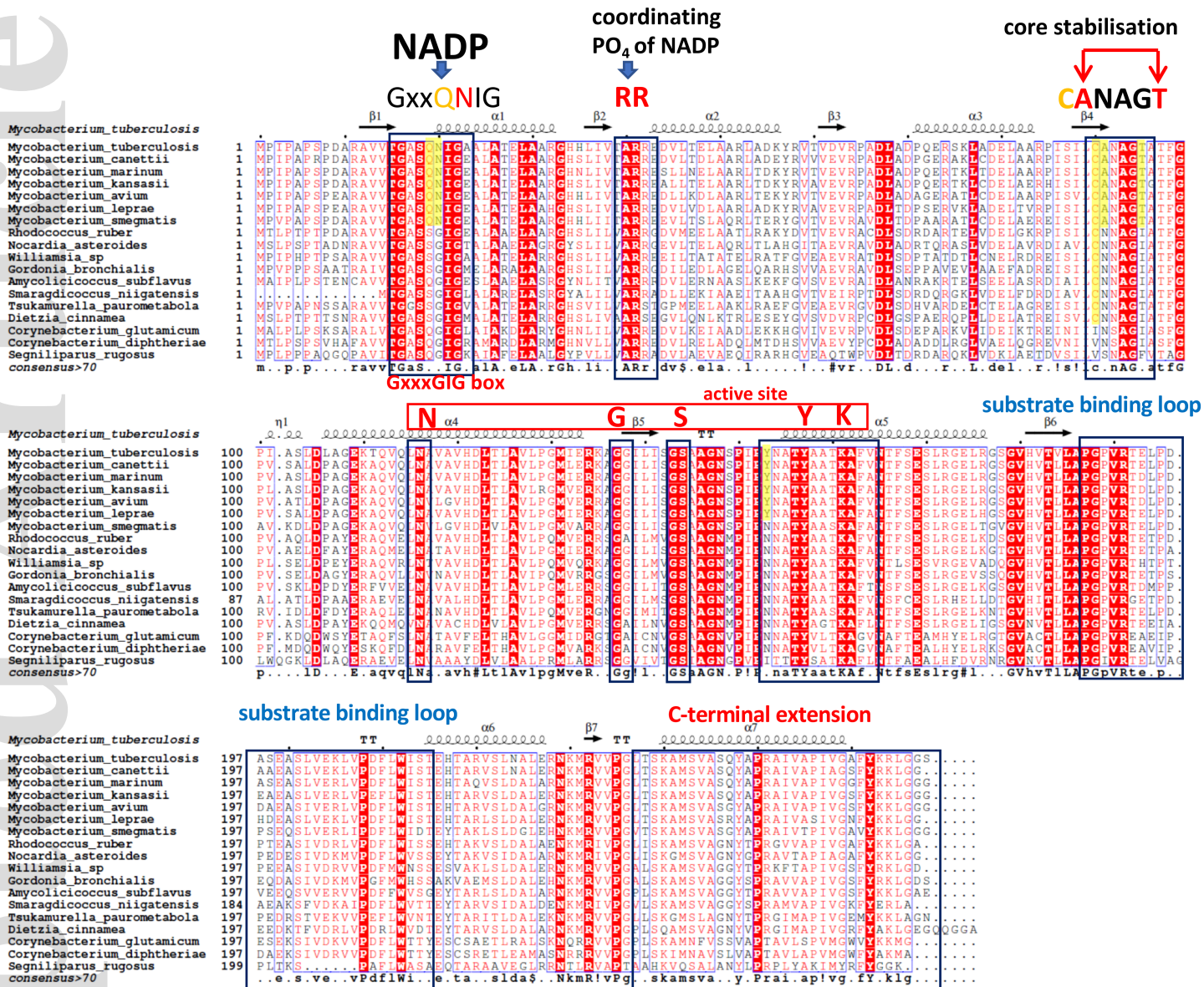
Yang, L., S. Lu, J. Belardinelli, E. Huc-Claustre, V. Jones, M. Jackson & H.I. Zgurskaya, (2014) RND transporters protect Corynebacterium glutamicum from antibiotics by assembling the outer membrane. *Microbiologyopen* **3**: 484-496.

**Table 1. Plasmids and bacterial strains used in this study.**

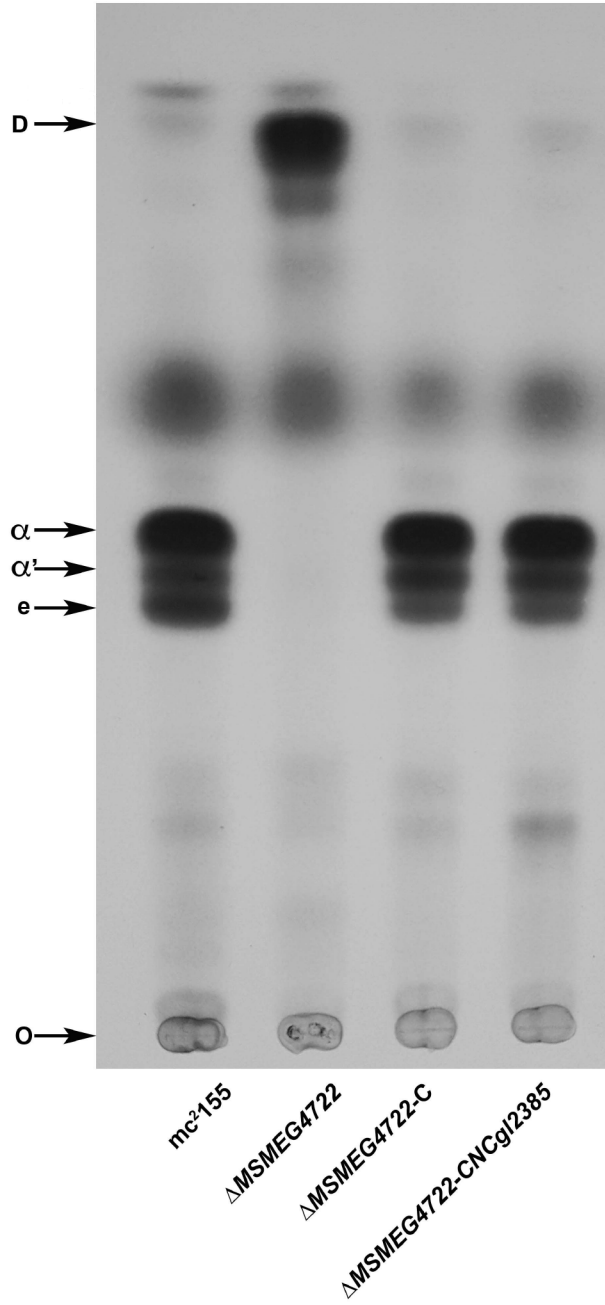
	Description	Reference/Source
<b>Plasmids</b>		
pMV261	<i>E. coli</i> - <i>Mycobacterium</i> shuttle plasmid vector with <i>hsp60</i> promoter and Kan <sup>R</sup> cassette ( <i>aph</i> )	(Stover <i>et al.</i> , 1991)
pMV261- <i>Rv2509</i>	<i>Rv2509</i> cloned in pMV261	(Bhatt <i>et al.</i> , 2008)
pMV261- <i>NCgl2385</i>	<i>NCgl2385</i> cloned in pMV261	This work
pMV261- <i>Rv2509D1</i>	Deletion construct of <i>Rv2509</i> cloned in pMV261-encodes a truncated protein missing the C-terminal residues Ala <sup>255</sup> -Ser <sup>268</sup>	This work
pMV261- <i>Rv2509D2</i>	Deletion construct of <i>Rv2509</i> cloned in pMV261-encodes a truncated protein missing the C-terminal residues Tyr <sup>262</sup> -Ser <sup>268</sup>	This work
pMV261:: <i>rev-tetR</i> -RBS-D	Episomal <i>E. coli</i> -mycobacterium shuttle plasmid pMV261:: <i>rev-tetR</i> -RBS-D providing constitutive <i>rev-tetR</i> gene expression from the HSP60 promoter in mycobacteria	(Korte <i>et al.</i> , 2016)
pET28a- <i>Rv2509</i>	Plasmid for expressing <i>Rv2509</i> in <i>E. coli</i>	This work
pET28a- <i>Rv2509D1</i>	Plasmid for expressing C-terminal truncated <i>Rv2509</i> in <i>E. coli</i>	This work
<b>Bacterial strains</b>		
mc <sup>2</sup> 155	Electroporation-proficient <i>ept</i> mutant of <i>M. smegmatis</i> strain mc <sup>2</sup> 6	(Snapper <i>et al.</i> , 1990)
ΔMSMEG4722	Deletion mutant of mc <sup>2</sup> 155 in which MSMEG4722 is replaced by <i>hyg</i>	(Bhatt <i>et al.</i> , 2008)
ΔMSMEG4722-CRv	ΔMSMEG4722 containing pMV261- <i>Rv2509</i>	(Bhatt <i>et al.</i> , 2008)
ΔMSMEG4722- <i>CNCgl2385</i>	ΔMSMEG4722 containing pMV261- <i>NCgl2385</i>	This work
ΔMSMEG4722-CRvD1	ΔMSMEG4722 containing pMV261- <i>Rv2509D1</i>	This work
ΔMSMEG4722-CRvD2	ΔMSMEG4722 containing pMV261- <i>Rv2509D2</i>	This work
BCG:: <i>P</i> <sub>Tet</sub> -BCG2529	BCG Pasteur strain containing the <i>Pmyc1</i> promoter from <i>M. smegmatis</i> engineered to contain four <i>tetO</i> operator sites, inserted immediately upstream of the start codon of BCG2529, the BCG homologue of <i>Rv2509</i> .	This work



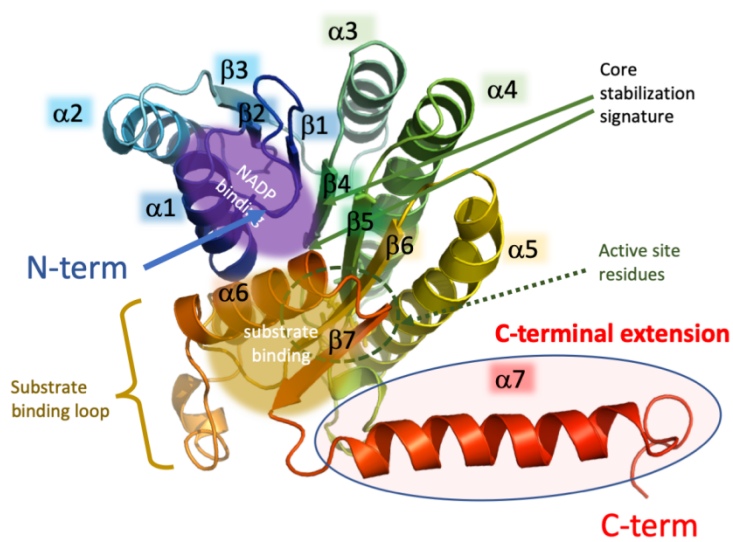
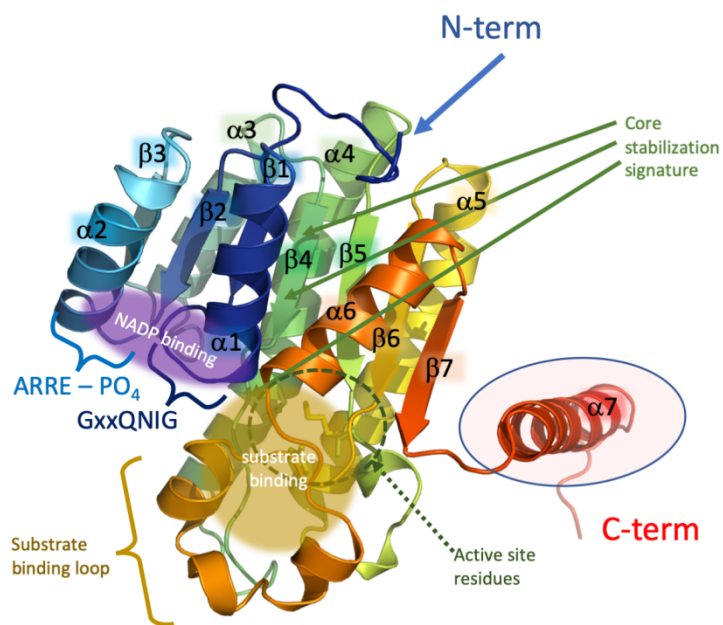




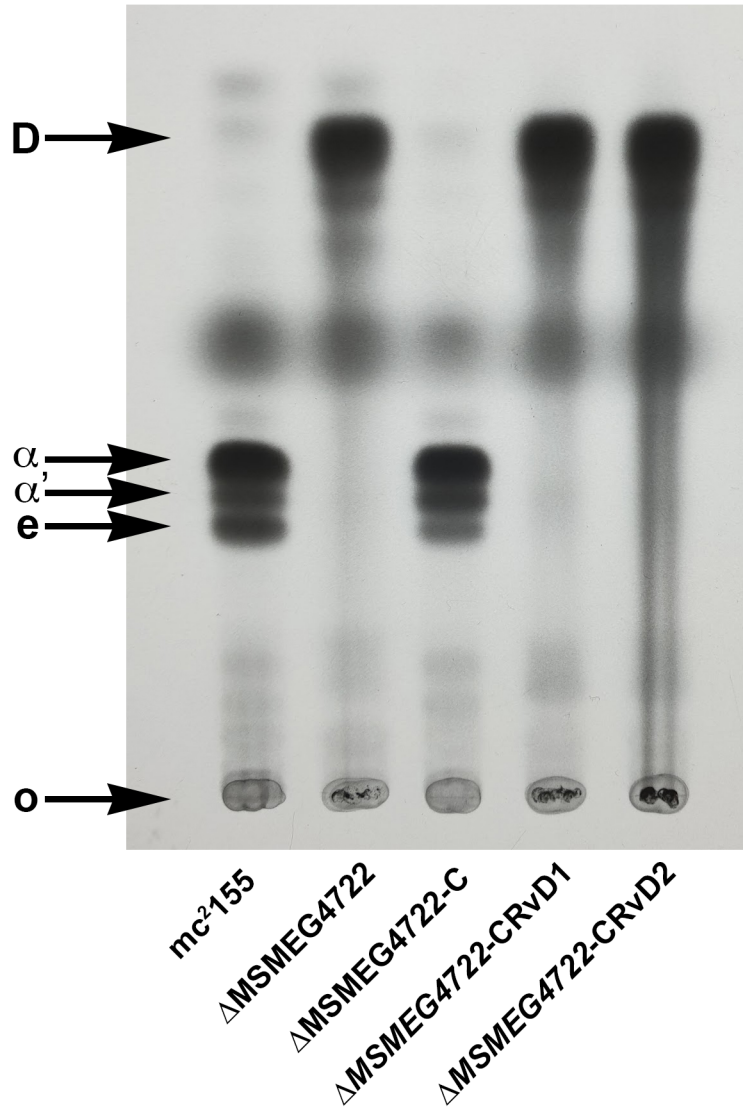
mmi\_14437\_f2.tif



mmi\_14437\_f3.tif

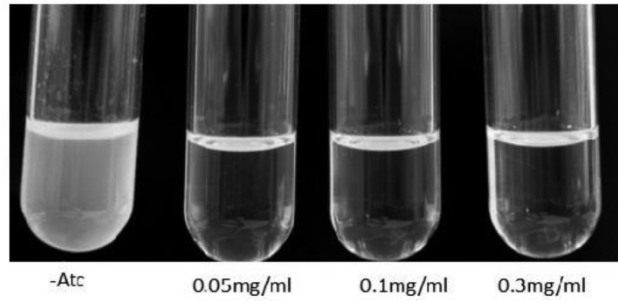


mimi\_14437\_f4.tiff

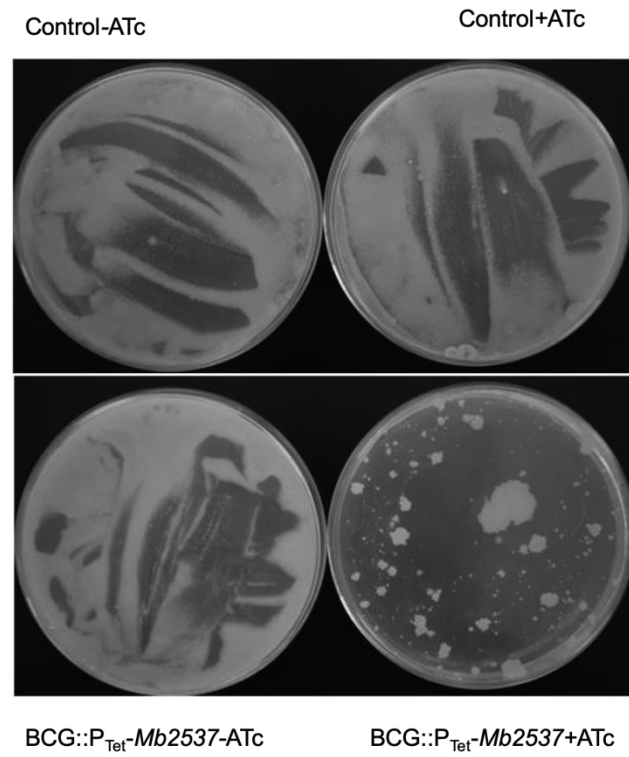


mmi\_14437\_f5.tif

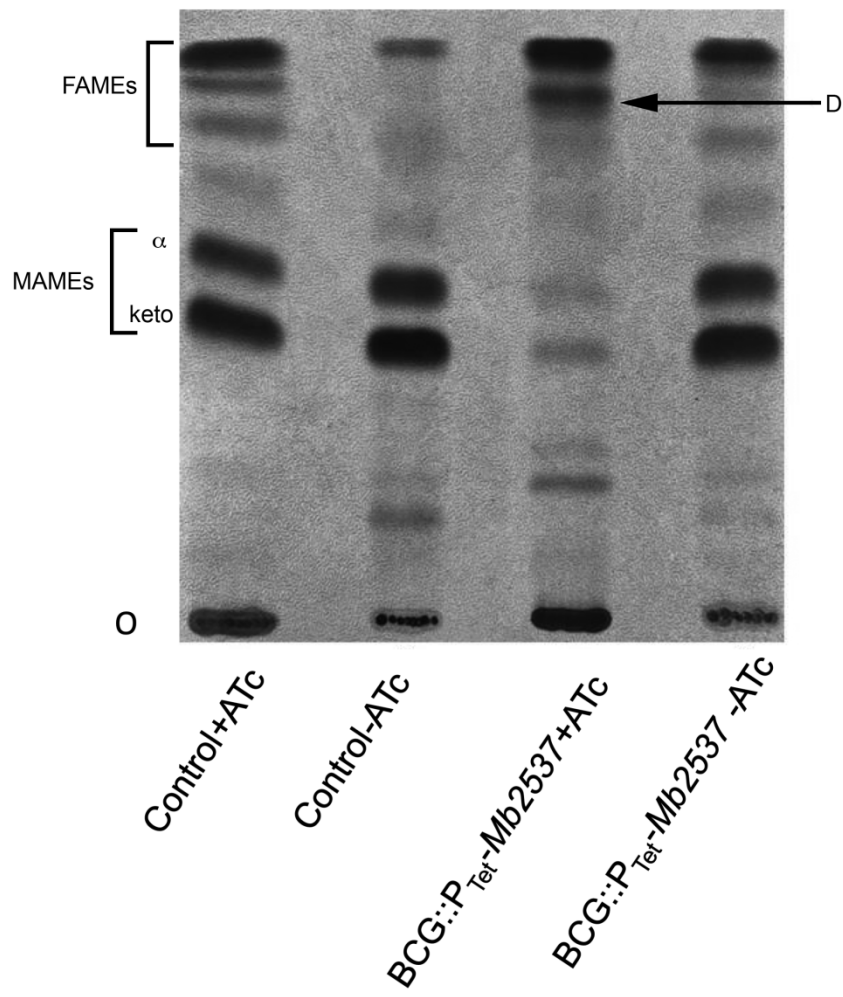
A



B



mmi\_14437\_f6.tiff



mmi\_14437\_f7.tiff