## Title

Crystal structure of the toxin Msmeg_6760, the structural homolog of Mycobacterium tuberculosis Rv2035, a novel type II toxin involved in the hypoxic response



## Journal

Acta Crystallographica Section F: Structural Biology Communications, 72(12)

## ISSN

2053-230X

## Authors

Bajaj, R Alexandra
Arbing, Mark A
Shin, Annie
et al.

## Publication Date

2016-12-01

## DOI

10.1107/s2053230x16017957

Peer reviewed

STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

# Crystal structure of the toxin Msmeg_6760, the structural homolog of *Mycobacterium tuberculosis* Rv2035, a novel type II toxin involved in the hypoxic response

**R. Alexandra Bajaj, Mark A. Arbing, Annie Shin, Duilio Cascio and Linda Miallau*‡**

UCLA–DOE Institute and Departments of Biological Chemistry and Chemistry and Biochemistry, University of California, Los Angeles, CA 90095-1570, USA. *Correspondence e-mail: linda.miallau@gmail.com

Received 25 September 2016
Accepted 8 November 2016

Edited by W. N. Hunter, University of Dundee, Scotland

‡ Present address: Department of Biochemistry and Molecular Biology, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Drive, Bethesda, MD 20814, USA

**Keywords:** *Mycobacterium smegmatis*; *Mycobacterium tuberculosis*; X-ray crystallography; protein structure; toxin–antitoxin complexes; bacterial latency; macrophage infection; novel folds.

**PDB reference**: Msmeg_6760, 3uid

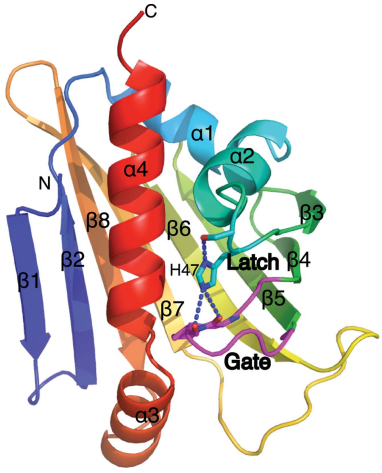
**Supporting information**: this article has supporting information at journals.iucr.org/f

The structure of Msmeg_6760, a protein of unknown function, has been determined. Biochemical and bioinformatics analyses determined that Msmeg_6760 interacts with a protein encoded in the same operon, Msmeg_6762, and predicted that the operon is a toxin–antitoxin (TA) system. Structural comparison of Msmeg_6760 with proteins of known function suggests that Msmeg_6760 binds a hydrophobic ligand in a buried cavity lined by large hydrophobic residues. Access to this cavity could be controlled by a gate–latch mechanism. The function of the Msmeg_6760 toxin is unknown, but structure-based predictions revealed that Msmeg_6760 and Msmeg_6762 are homologous to Rv2034 and Rv2035, a predicted novel TA system involved in *Mycobacterium tuberculosis* latency during macrophage infection. The Msmeg_6760 toxin fold has not been previously described for bacterial toxins and its unique structural features suggest that toxin activation is likely to be mediated by a novel mechanism.

## 1. Introduction

Toxin–antitoxin (TA) systems were originally discovered in the 1980s as mediators of plasmid stability by acting as 'addiction' molecules. Since then, a vast number of TA systems have been discovered on the chromosomes of almost all bacteria, with some bacteria having only a few while others have multiple dozens (Yamaguchi *et al.*, 2011). For instance, only four TA systems are encoded in the chromosome of *Mycobacterium smegmatis*, while *M. tuberculosis* has 88 reported chromosomal TA systems (Makarova *et al.*, 2009; Ramage *et al.*, 2009).

Expression of the toxin and antitoxin, which are typically encoded in a bicistronic operon, is upregulated in response to environmental or nutritional stressors (Gerdes *et al.*, 2005). In the abundant type II TA systems, both the toxin and the antitoxin are proteins that assemble into a stable complex and autoregulate the expression of their own operon (Gerdes *et al.*, 2005). Under conditions of stress, such as starvation or hypoxia, intracellular proteases such as the Clp family of proteases are upregulated and target the labile antitoxin for degradation (Hansen *et al.*, 2012; Gerdes *et al.*, 2005). Toxin activation and accumulation in the cell results in variable toxic effects ranging from decreased cellular growth rates to cell death. Toxins cause cell-growth arrest by targeting various essential cell processes such as DNA replication (*e.g.* the PemK toxin; Ruiz-Echevarría *et al.*, 1995), cell division (*e.g.* the YeeV toxin; Tan *et al.*, 2011), mRNA degradation (*e.g.* the VapC toxin; Miallau *et al.*, 2009; Min *et al.*, 2012) or inhibition



of protein synthesis (*e.g.* the RelE toxin; Korch *et al.*, 2009; Miallau *et al.*, 2013) or cell-wall synthesis (*e.g.* the PezT toxin; Mutschler *et al.*, 2011). It has been shown that toxins mediate programmed cell death or, more commonly, the induction of reversible bacteriostasis, which may become irreversible if toxin exposure is prolonged (Syed & Lévesque, 2012). Bacteriostasis is especially important to the survival and pathogenicity of dormant phenotype pathogens such as *M. tuberculosis*, which display persistence and multidrug resistance (Pandey & Gerdes, 2005).

Here, we present the crystal structure of Msmeg_6760, a homolog of *M. tuberculosis* Rv2035 (Makarova *et al.*, 2009) from the nonpathogenic organism *M. smegmatis*. A prior study of Rv2035 revealed that it contains a polyketide cyclase 2 domain, as found in proteins involved in polyketide cyclase/dehydrase and lipid transport, implying a potential role for Msmeg_6760/Rv2035 proteins in the regulation of lipid metabolism-related genes (Gao *et al.*, 2012). The Msmeg_6760 and Rv2035 genes are located in operons that contain two additional components: an ArsR-type transcriptional regulator (Msmeg_6762/Rv2034) positioned upstream of Msmeg_6760/Rv2035 and a protein of unknown function (Msmeg_6763/Rv2036).

Msmeg_6762 (115 amino acids; pI 9.2) is predicted to contain a helix–turn–helix DNA-binding domain. This information, taken in the context of its genetic location upstream of *msmeg_6760*, opens up the possibility that together Msmeg_6760 (161 amino acids; pI 4.3) and Msmeg_6762 are the components of a toxin–antitoxin system. Indeed, our biochemical data reveal that Msmeg_6760 and Msmeg_6762 interact and form a complex. We propose that Msmeg_6760 and Msmeg_6762 comprise a novel toxin–antitoxin system, with the DNA-binding domain contained in Msmeg_6762 serving as an antitoxin which inhibits the Msmeg_6760 toxin. Structural and sequence analyses suggest that the toxin binds an unknown hydrophobic ligand, which may be a key regulator of adaptation to hypoxic conditions.

**Table 1**
Data-collection, processing and refinement statistics.

Values in parentheses are for the highest resolution shell.

| | |
|---|---|
| Data collection and processing | |
| Diffraction source | ID-24-C, Advanced Photon Source |
| Wavelength (Å) | 0.9791 |
| Temperature (K) | 100 |
| Detector | ADSC Quantum 315 |
| Total rotation range per image (°) | 1 |
| Total rotation range (°) | 180 |
| Exposure time per image (s) | 1 |
| Space group | $P2_12_12_1$ |
| Unit-cell parameters (Å, °) | $a$ = 48.32, $b$ = 60.21, $c$ = 100.98, $\alpha = \beta = \gamma$ = 90 |
| Resolution range (Å) | 19.63–1.57 (1.61–1.57) |
| $R_{merge}$ (%) | 4.8 (48.9) |
| Total No. of reflections | 130502 |
| No. of unique reflections | 39602 |
| Completeness (%) | 94.9 (68.4) |
| Multiplicity | 3.32 |
| $\langle I/\sigma(I)\rangle$ | 16.79 (2.11) |
| Overall $B$ factor from Wilson plot (Å$^2$) | 13.93 |
| Structure refinement | |
| Resolution range (Å) | 19.63–1.57 (1.61–1.57) |
| Completeness (%) | 94.9 |
| No. of reflections, working set | 35637 |
| No. of reflections, test set | 3965 |
| Final $R_{work}$ | 0.192 |
| Final $R_{free}$ | 0.213 |
| Molecules in asymmetric unit | 2 |
| No. of non-H atoms | |
| Protein | 2536 |
| Ion | 0 |
| Ligand | 0 |
| Water | 197 |
| Total | 2733 |
| R.m.s. deviations | |
| Bonds (Å) | 0.006 |
| Angles (°) | 1.043 |
| Average $B$ factors (Å$^2$) | |
| Protein | 17.77 |
| Solvent | 20.86 |
| Solvent content (%) | 37.8 |
| Ramachandran plot | |
| Favored regions (%) | 91.5 |
| Additionally allowed (%) | 8.5 |
| Outliers (%) | 0.0 |

## 2. Materials and methods

### 2.1. Construct design and cloning

The bicistronic operon encoding Msmeg_6762 and Msmeg_6760 was cloned into pMAPLe3, a modified pET vector harboring kanamycin resistance. The resulting vector allows the expression of the first gene in the operon as a maltose-binding protein fusion that is subsequently cleaved *in vivo* by constitutively co-expressed TEV protease, leading to an unmodified protein (Arbing *et al.*, 2013). The gene product in the 3′ position is expressed with a C-terminal tag with the sequence THHHHHH for affinity purification. Using the SLIC cloning method (Li & Elledge, 2007), the operon was inserted into pMAPLe3 using the gene-specific primers Msmeg_6762-6760.pM3.For (5′-AACCTGTATTTCCAGAGTATGTACGTTGTGTGGGTGACC) and Msmeg_6762-6760.pM3.Rev (5′-GTGATGGTGATGGTGATGAGTTGCAGTGAGCAGCGCGTC).

### 2.2. Protein expression and purification

The pMAPLe3 plasmid containing the Msmeg_6762–Msmeg_6760 operon was transformed into *Escherichia coli* Rosetta (DE3) cells. The cells were grown at 315 K to an $OD_{600}$ of 1.0 in Luria–Bertani broth with 35 µg ml$^{-1}$ kanamycin. Protein expression was induced with 0.4 m*M* IPTG and the temperature was lowered to 293 K for 15.5 h. The cells were harvested at an $OD_{600}$ of 3.0 and immediately cooled to 253 K. Cell pellets were resuspended in 50 m*M* Tris–HCl pH 6.0, 500 m*M* NaCl, 20 m*M* imidazole pH 7.0, 10%(*v*/*v*) glycerol (buffer *A*) containing protease-inhibitor cocktail (Sigma) and phenylmethylsulfonyl fluoride. Lysis was performed by sonication and the lysate was then centrifuged at 40 000*g* for 1 h. The supernatant was filtered through 0.45 and 0.2 µm polyethersulfone (PES) filters and loaded onto a 5 ml Ni–NTA HisTrap chelating column (GE Healthcare) previously equilibrated with buffer *A*. The TA complex was coeluted from the column using a buffer consisting of 50 m*M* Tris–HCl pH 6.0,

500 m*M* NaCl, 500 m*M* imidazole pH 7.0, 10%(*v*/*v*) glycerol. The relevant eluates were pooled and dialyzed against 50 m*M* Tris–HCl pH 6.0, 100 m*M* NaCl, 10%(*v*/*v*) glycerol (buffer *B*). The eluates were concentrated and injected onto a Superdex 75 size-exclusion column (GE Healthcare) equilibrated with buffer *B*. The two subunits of the complex copurify in the first purification step, but partial complex dissociation occurs during or before the size-exclusion chromatography (SEC) step, with two peaks being obtained from the SEC run. SDS–PAGE analysis reveals that the first peak contains both Msmeg_6760 and Msmeg_6762, while the second peak contains Msmeg_6760 alone. Fractions corresponding to first and second elution peaks were pooled separately and were concentrated to 2.6 and 15 mg ml$^{-1}$, respectively.

### 2.3. Crystallization and structure determination

Both the Msmeg_6762–Msmeg_6760 complex and Msmeg_6760 alone were subjected to crystallization screening at concentrations of 2.6 and 15.0 mg ml$^{-1}$, respectively. However, only Msmeg_6760 alone yielded crystals of a size suitable for diffraction analysis. Msmeg_6760 crystals formed at 291 K within 14 days in 0.1 *M* sodium acetate pH 4.5, 20%(*w*/*v*) PEG 3000 using a 1:4 ratio of protein:reservoir solution. All crystals were cryoprotected using 20%(*v*/*v*) glycerol and then flash-cooled in liquid nitrogen. A data set at a resolution of 1.57 Å was collected on beamline 24-ID-C at the Advanced Photon Source. All data were processed using *XDS* (Kabsch, 2010). Phases for the native crystal form of Msmeg_6760 were obtained by molecular replacement (MR) using the hypothetical protein MM0500 from *Methanosarcina mazei* (PDB entry 1xuv; 22% sequence identity to Msmeg_6760; Northeast Structural Genomics Consortium, unpublished work) as a model. The model, which was identified as the best homology model by *MrBUMP* (*MOLREP*; Vagin & Teplyakov, 2010), was truncated using *CHAINSAW* (Stein, 2008) and manually with *Coot* (Emsley *et al.*, 2010) to remove nonconserved loop regions. *Phaser* (McCoy *et al.*, 2007) was used to perform the molecular-replacement solution, with an *R* factor of 53.84%. Additional residues were built manually in iterative cycles using *Coot* interspersed with refinement in *phenix.refine* (Adams *et al.*, 2010) using TLS groups (one per chain) and individual anisotropic atomic displacement parameters. The geometry of the structure was checked using the *Structure Analysis and VErification Server* (*SAVES*; http://nihserver.mbi.ucla.edu/SAVES), which integrates the programs *PROCHECK*, *WHAT CHECK*, *ERRAT*, *VERIFY3D* and *PROVE*. Graphical representations of the structures were prepared with *PyMOL* (v.1.2r3pre; Schrödinger). Data-collection and refinement statistics are summarized in Table 1.

## 3. Results and discussion

### 3.1. Overall structure of Msmeg_6760

The Msmeg_6760 structure has excellent geometry and was refined to $R_{work}$ and $R_{free}$ values of 19.17 and 21.30%,

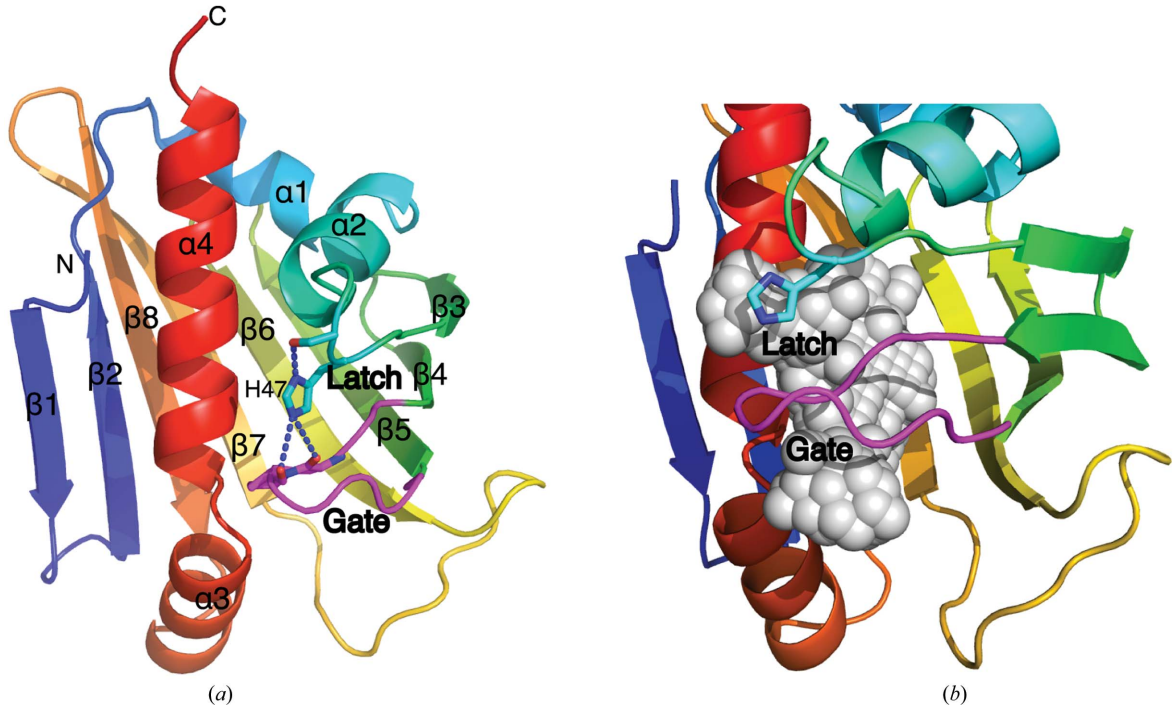



**Figure 1**
Cartoon representation of the structure of Msmeg_6760. In both figures, secondary structures are colored using a rainbow gradient from blue for the N-terminus (labelled N) to red for the C-terminus (labelled C). Residues involved in the latch and gate are shown in cyan and magenta, respectively. (*a*) Hydrogen bonds between the latch residue, His47, and main-chain atoms of residues that belong to the gate loop are shown as blue dotted lines. (*b*) The cavity, estimated using *POCASA* (Yu *et al.*, 2010), is represented as gray spheres.

**Table 2**
Details of selected structures homologous to Msmeg_6760 (PDB entry 3uid) and superimposed in Fig. 2(*a*).

| Protein | Ligand | R.m.s.d. with 3uid (Å) | No. of $C^{\alpha}$ pairs | Sequence identity (%) | Sequence similarity (%) | PDB code | Reference |
|---|---|---|---|---|---|---|---|
| Birch pollen allergen Bet v 1l | (3α,5β,12α)-3,12-Dihydroxycholan-24-oic acid | 3.4 | 147 | 11.8 | 24.2 | 1fm4 | Marković-Housley *et al.* (2003) |
| Plant pathogenesis-related 10 | 8-Anilino-1-naphthalene sulfonate | 3.0 | 143 | 6.4 | 22.0 | 1txc | F. Wu, Z. Wei, Z. Zhou & W. Gong (unpublished work) |
| Cytokinin-specific binding protein from mung bean | (2*E*)-2-Methyl-4-(9*H*-purin-6-ylamino)-but-2-en-1-ol | 3.2 | 137 | 11.8 | 24.2 | 2flh | Pasternak *et al.* (2006) |
| Norcoclaurine synthase | *p*-Hydroxybenzaldehyde | 3.2 | 139 | 10.9 | 21.8 | 2vq5 | Ilari *et al.* (2009) |
| Abscisic acid receptor | (2*Z*,4*E*)-5-[(1*S*)-1-Hydroxy-2,6,6-trimethyl-4-oxocyclohex-2-en-1-yl]-3-methylpenta-2,4-dienoic acid | 3.4 | 126 | 10.4 | 17.0 | 3kb0 | Melcher *et al.* (2009) |

respectively (Table 1). The asymmetric unit contains two Msmeg_6760 molecules with a root-mean-square deviation of 0.82 Å: chain *A* contains residues Pro2–His165 with six alternative conformations, chain *B* contains residues Pro2–Pro102 and Val108–His163 with only three alternative conformations and there are a total of 197 water molecules. Although the asymmetric unit contains two molecules, Msmeg_6760 appears to be a monomer, as suggested by the 160° rotation between chain *A* and chain *B*. Calculation of the buried surface area using *PDBePISA* (Krissinel & Henrick, 2007) indicates that only 3.6% of the total exposed surface is involved in making contacts between the non-crystallographic monomers, further suggesting that Msmeg_6760 is a monomer in solution.

The Msmeg_6760 structure belongs to the Bet v 1 fold family (Radauer *et al.*, 2008) and is comprised of a curved β-sheet consisting of eight antiparallel β-strands with two long kinked α-helices packed against the concave face of the sheet; one helix is located between β-strands 2 and 3, while the other is at the C-terminus of the structure (Fig. 1). The position of the helices relative to the curved sheet forms a large hydrophobic cavity lined with multiple hydrophobic residues that may be involved in binding hydrophobic ligands (Figs. 2*b* and 2*c*).

### 3.2. Search for function

**3.2.1. Msmeg_6762–Msmeg_6760 is a novel TA system.** Msmeg_6760 and Msmeg_6762 are predicted to belong to a large TA-system family of unclassified function, which is supported by the presence of genes orthologous to those for Msmeg_6760 and Msmeg_6762 in many prokaryotic genomes (Makarova *et al.*, 2009; Keren *et al.*, 2011). The upstream Msmeg_6762 gene encodes an ArsR-type repressor; ArsR is a well characterized regulator of the cellular response to stress induced by heavy metals. ArsR also binds its own promoter and represses its own expression, like most antitoxins (Hayes & Kędzierska, 2014). The orthologous *M. tuberculosis* operon, Rv2034–Rv2035, is one of the most upregulated systems in *M. tuberculosis* drug-resistant persisters (Keren *et al.*, 2011). Rv2034 autoregulates its own expression and was also found to positively regulate the dormancy regulon, *dosR*, responsible for hypoxic adaptation in latent *M. tuberculosis* during macrophage infection (Shiloh *et al.*, 2008; Keren *et al.*, 2011; Gao *et al.*, 2012).

Msmeg_6760 and Msmeg_6762 form a stable complex on copurifying untagged Msmeg_6762 antitoxin *via* its association with $His_6$-tagged Msmeg_6760 toxin using affinity chromatography. Moreover, the presence of a stable complex was further confirmed by the coelution of both subunits of the complex from a size-exclusion column with an elution volume of 43.7 ml, while Msmeg_6760 alone eluted at 69.4 ml. We are confident that Msmeg_6760 and Msmeg_6762 encode a TA system in which the function of the ArsR-like antitoxin can be inferred from existing biochemical data. However, the function of the Msmeg_6760 toxin, and its homologs, has not been previously characterized.

**3.2.2. Msmeg_6760 is likely to bind an unidentified hydrophobic ligand.** Type II toxins represent a group of structurally diverse proteins that regulate microbial growth through the inhibition of essential cellular processes, often by acting as gyrase poisons or mRNA interferases/ribonucleases. The folds of type II toxins vary considerably and include all-α-helical folds (Fic/Doc toxins; Arbing *et al.*, 2010), microbial RNase-like folds consisting of a central β-sheet flanked by α-helices (RelE/YoeB toxins; Miallau *et al.*, 2013) and the PIN domain-like α/β/α sandwich topology of VapB toxins (Miallau *et al.*, 2009). However, the fold of Msmeg_6760 and analysis of its electrostatic surface potential suggest that Msmeg_6760 has a function that dramatically differs from other type II toxins. The external surface of Msmeg_6760 shows the presence of multiple patches of negative surface potential and only one rather limited patch of positive surface potential that is likely to be insufficient for interaction with a negatively charged nucleic acid. A *DALI* search for nearest structural neighbors revealed similarity to over 100 proteins of unknown function for which the *Z*-score is superior to 12 (Holm & Sander, 1995). The fold of Msmeg_6760 belongs to the common START/RHO_alpha_C/PITP/Bet_v1/CoxG/CalC (SRPBCC) ligand-binding domain superfamily, which comprises a large number of members known to bind a diverse array of hydrophobic ligands, including lipids, hormones, polycyclic aromatic hydrocarbons and RNA (Radauer *et al.*, 2008; Table 2). This superfamily includes the large family of polyketide cyclases, the steroidogenic acute regulatory protein (StAR)-related lipid transfer (START) domains and the activator of Hsp90 ATPase homolog 1-like protein (AHSA1) family. The large number of bulky hydrophobic residues lining the interior of the Msmeg_6760 cavity (Figs. 2*b* and 2*c*) suggests that it too

may bind a hydrophobic ligand. The volume of the cavity is estimated to be around 390 Å$^3$ (Yu *et al.*, 2010; Fig. 1*b*). Although we did not observe any extra density indicating the presence of a ligand in the Msmeg_6760 cavity, superposition of homologous ligand-bound structures deposited in the Protein Data Bank (PDB) indicate which part of the cavity is

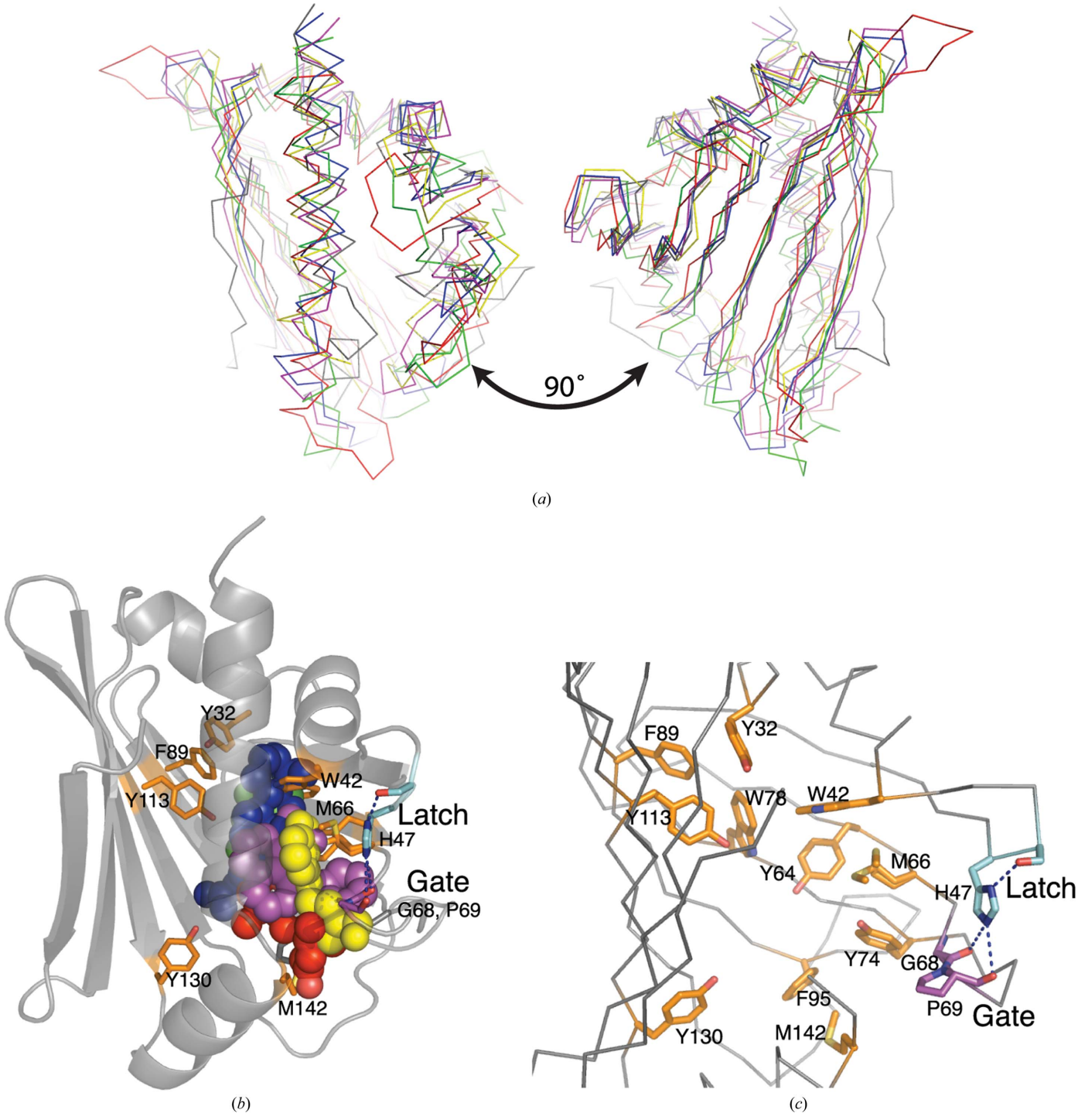



**Figure 2**
Structural homologs and putative binding-site residues. (*a*) Two views of the superposition of Msmeg_6760 structural homologs in ligand-bound conformation (ligands are omitted for clarity). Msmeg_6760 is shown in dark gray and superimposed with PDB entries 1fm4 (blue), 1txc (magenta), 2flh (yellow), 2vq5 (green) and 3kb0 (red). Details of the superimposed structures are given in Table 2. (*b*) The structure of Msmeg_6760 is shown as a gray cartoon with ligands of homologous proteins represented as spheres: (3$\alpha$,5$\beta$,12$\alpha$)-3,12-dihydroxycholan-24-oic acid (PDB entry 1fm4), blue; 8-anilino-1-naphthalene sulfonate (PDB entry 1txc), magenta; (2*E*)-2-methyl-4-(9*H*-purin-6-ylamino)but-2-en-1-ol (PDB entry 2flh), yellow; *p*-hydroxybenzaldehyde (PDB entry 2vq5), green; (2*Z*,4*E*)-5-[(1*S*)-1-hydroxy-2,6,6-trimethyl-4-oxocyclohex-2-en-1-yl]-3-methylpenta-2,4-dienoic acid (PDB entry 3kb0), red. (*c*) Close-up view of the putative binding-site residues of Msmeg_6760 shown as orange sticks with helix $\alpha$4 removed for clarity. The latch and gate loops are shown in magenta and red, respectively. Hydrogen bonds are shown as blue dotted lines.

most likely to be involved in ligand binding (Figs. 2*a*, 2*b* and 2*c*). Structural superposition of the structures listed in Table 1 and Fig. 2(*a*) with Msmeg_6760 suggests that five tyrosine residues (Tyr32, Tyr64, Tyr74, Tyr113 and Tyr130), two tryptophan residues (Trp42 and Trp78), two phenylalanine residues (Phe89 and Phe95) and two methionine residues (Met66 and Met142) could be involved in ligand binding. The substrate could be a lipid, an antibiotic or RNA, as suggested by the classes of compounds bound by the members of the SRPBCC family. It is noteworthy that the ligand-binding cavity of Msmeg_6760 is not solvent-accessible in our structure and thus conformational changes are likely to be involved in substrate binding.

**3.2.3. The hydrophobic ligand-binding cavity could be accessed by a gate–latch mechanism**. A gate–latch lock mechanism has been described in the structure of the abscisic acid receptor, a vital plant START protein receptor that mediates response to environmental stresses such as drought, cold or salinity (Melcher *et al.*, 2009). Entry to the ligand-binding pocket is obstructed by a gate-like loop on one side and by a latch-like region on the other. Analysis of the structure of Msmeg_6760 suggests that a similar mechanism may regulate Msmeg_6760 activity. A hydrogen bond between the latch residue His47 (situated in the loop between helix $\alpha2$ and strand $\beta3$) and the main-chain atoms of the gate-loop residues Gly68 and Pro69 would lock the cavity (Fig. 3*a*). The latching residue, His47, is also held in position by accepting a hydrogen bond from the OG atom of Ser46. In our apo structure the gate loop is locked in the closed state through these intramolecular interactions with the latch and by intermolecular interactions with residues of the C-terminus of a symmetry-related molecule. Thus, in the presence of ligand the side chain of His47 of the apo protein would flip to expose NE2 and allow ligand binding; His47 ND1 would then be in an unfavorable deprotonated state that is unable to interact with the Thr67 and Gly68 carboxyl groups. The interaction between the latch and the gate loop would be destabilized, which would result in release of the latch and opening of the gate (Fig. 3*b*).

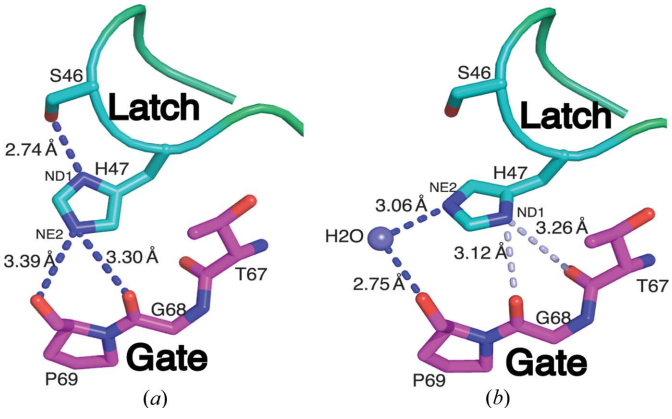



**Figure 3**
Representation of the hydrogen-bond network between the latch and gate loops. (*a*) His47 of the latch loop maintains the gate loop in a closed conformation as seen in the crystal structure. (*b*) Modelled structure with the His47 side chain flipped so that NE2 donates a hydrogen bond to a water molecule while the protonated His47 ND1 interacts with the Gly68 and Thr67 main-chain carboxyl atoms in our structure. At physiological pH it is likely that ND1 is deprotonated so the hydrogen bonds between His47 and Gly68 and Thr67 would be disrupted (pale blue dotted lines), opening the gate.

**3.2.4. Homologs of Msmeg_6760**. To identify a potential function for Msmeg_6760, we performed a *PSI-BLAST* search using its primary amino-acid sequence. This identified a eukaryotic protein: activator of Hsp90 ATPase homolog 1-like protein (AHSA1). The AHSA1 protein family is comprised of 4271 sequences and its members are named in accordance with their sequence similarity to the C-terminal region of AHSA1. The function of the prokaryotic members of this family is ambiguous and they are classified as general stress proteins. The eukaryotic homologs stimulate Hsp90 function (Lotz *et al.*, 2003) and are likely to increase the efficiency of protein folding under conditions of increased stress (Panaretou *et al.*, 2002). Although no ortholog of Hsp90 could be detected in *M. smegmatis* mc$^2$155, it was recently discovered that DnaK/Hsp70 is the major chaperone mediating mycobacterial native protein folding and that it also limits protein aggregation under stress. DnaK is essential and its depletion results in the formation of large protein aggregates that mediate growth arrest and result in bacterial filamentation (Fay & Glickman, 2014). It is thus possible that Msmeg_6760 modulates DnaK activity in response to stress.

Analysis of the Msmeg_6760 amino-acid sequence also identified a polyketide_cyc2 domain, which is found in polyketide cyclases/dehydrases and proteins involved in lipid transport, which corroborates with the large hydrophobic ligand-binding cavity described earlier. This class of proteins participates in building biologically active natural products, produced by both bacterial and fungal species, to impair the growth of other microbes and gain a competitive advantage (Ames *et al.*, 2008). Given the abundance of polyketide-type antibiotics, it is possible that a polyketide-type antibiotic or other secondary metabolite may be the ligand that triggers Msmeg_6760 activity.

However, the most relevant result was the identification of a plant protein family of the Bet v 1-like superfamily called the PR-10 (pathogenesis related-10) family that has over 100 members. Proteins of this family have been experimentally determined to bind and cleave RNA and induce apoptotic processes in pathogen-infected cells in order to limit pathogen invasion (Zubini *et al.*, 2009). Moreover, the concentration of PR-10 is highly increased by pathogen invasion and results in cell apoptosis. The functional and structural similarities between plant PR-10 proteins and bacterial toxins are striking and it is very likely that PR-10 proteins and toxins such as Msmeg_6760 and Rv2035 are related.

## 4. Conclusions

Toxins from bacterial toxin–antitoxin systems have transient effects on metabolic processes to alter bacterial physiology in order to withstand environmental stress and ensure survival. The Msmeg_6762–Msmeg_6760 protein complex fits the criteria for bacterial TA systems; however, the fold of the

Msmeg_6760 toxin, which belongs to the SRPBCC superfamily, deviates significantly from the folds of those toxins which have been structurally characterized. The structural similarity of the Msmeg_6760 toxin extends to subfamilies of the SRPBCC superfamily, which have functions that could be associated with the bacterial response to environmental stress, including polyketide binding, RNase activity and stimulation of chaperone activity. Given the importance of the homologous system, Rv2034–Rv2035, in *M. tuberculosis* latency (Gao *et al.*, 2012; Keren *et al.*, 2011), further study of the molecular details of the Msmeg_6762–Msmeg_6760 TA system to identify the physiological stress, the ligand, the sequence of complex dissociation and potential protein interaction partners may be valuable for the design of antimycobacterial therapeutics.

## Acknowledgements

We thank the staff of the UCLA–DOE X-ray Crystallography Core Facility (supported by Department of Energy Grant DE-FC02-02ER63421) for assistance with crystallization screening. The assistance of the staff of NE-CAT beamline 24-ID-C at the Advanced Photon Source is greatly appreciated. We also thank David Eisenberg for his constant support. This work was supported by National Institutes of Health Grant TBSGC R01 Grant No. 1P01AI095208-01A1 and TBSGC R01 subaward No. 99-S150606-1 (to LM).

## References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Ames, B. D., Korman, T. P., Zhang, W., Smith, P., Vu, T., Tang, Y. & Tsai, S.-C. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 5349–5354.
Arbing, M. A., Chan, S., Harris, L., Kuo, E., Zhou, T. T., Ahn, C. J., Nguyen, L., He, Q., Lu, J., Menchavez, P. T., Shin, A., Holton, T., Sawaya, M. R., Cascio, D. & Eisenberg, D. (2013). *PLoS One*, **8**, e81753.
Arbing, M. A. *et al.* (2010). *Structure*, **18**, 996–1010.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fay, A. & Glickman, M. S. (2014). *PLoS Genet.* **10**, e1004516.
Gao, C., Yang, M. & He, Z.-G. (2012). *PLoS One*, **7**, e36255.
Gerdes, K., Christensen, S. K. & Løbner-Olesen, A. (2005). *Nature Rev. Microbiol.* **3**, 371–382.
Hansen, S., Vulić, M., Min, J., Yen, T.-J., Schumacher, M. A., Brennan, R. G. & Lewis, K. (2012). *PLoS One*, **7**, e39185.
Hayes, F. & Kędzierska, B. (2014). *Toxins*, **6**, 337–358.
Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.
Ilari, A., Franceschini, S., Bonamore, A., Arenghi, F., Botta, B., Macone, A., Pasquo, A., Bellucci, L. & Boffi, A. (2009). *J. Biol. Chem.* **284**, 897–904.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Keren, I., Minami, S., Rubin, E. & Lewis, K. (2011). *mBio*, **2**, e00100–11.
Korch, S. B., Contreras, H. & Clark-Curtiss, J. E. (2009). *J. Bacteriol.* **191**, 1618–1630.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Li, M. Z. & Elledge, S. J. (2007). *Nature Methods*, **4**, 251–256.
Lotz, G. P., Lin, H., Harst, A. & Obermann, W. M. J. (2003). *J. Biol. Chem.* **278**, 17228–17235.
Makarova, K. S., Wolf, Y. I. & Koonin, E. V. (2009). *Biol. Direct*, **4**, 19.
Marković-Housley, Z., Degano, M., Lamba, D., von Roepenack-Lahaye, E., Clemens, S., Susani, M., Ferreira, F., Scheiner, O. & Breiteneder, H. (2003). *J. Mol. Biol.* **325**, 123–133.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Melcher, K. *et al.* (2009). *Nature (London)*, **462**, 602–608.
Miallau, L., Faller, M., Chiang, J., Arbing, M., Guo, F., Cascio, D. & Eisenberg, D. (2009). *J. Biol. Chem.* **284**, 276–283.
Miallau, L., Jain, P., Arbing, M. A., Cascio, D., Phan, T., Ahn, C. J., Chan, S., Chernishof, I., Maxson, M., Chiang, J., Jacobs, W. Jr & Eisenberg, D. S. (2013). *Structure*, **21**, 627–637.
Min, A. B., Miallau, L., Sawaya, M. R., Habel, J., Cascio, D. & Eisenberg, D. (2012). *Protein Sci.* **21**, 1754–1767.
Mutschler, H., Gebhardt, M., Shoeman, R. L. & Meinhart, A. (2011). *PLoS Biol.* **9**, e1001033.
Panaretou, B. *et al.* (2002). *Mol. Cell*, **10**, 1307–1318.
Pandey, D. P. & Gerdes, K. (2005). *Nucleic Acids Res.* **33**, 966–976.
Pasternak, O., Bujacz, G. D., Fujimoto, Y., Hashimoto, Y., Jelen, F., Otlewski, J., Sikorski, M. M. & Jaskolski, M. (2006). *Plant Cell*, **18**, 2622–2634.
Radauer, C., Lackner, P. & Breiteneder, H. (2008). *BMC Evol. Biol.* **8**, 286.
Ramage, H. R., Connolly, L. E. & Cox, J. S. (2009). *PLoS Genet.* **5**, e1000767.
Ruiz-Echevarría, M. J., de la Torre, M. A. & Díaz-Orejas, R. (1995). *FEMS Microbiol. Lett.* **130**, 129–135.
Shiloh, M. U., Manzanillo, P. & Cox, J. S. (2008). *Cell Host Microbe*, **3**, 323–330.
Stein, N. (2008). *J. Appl. Cryst.* **41**, 641–643.
Syed, M. A. & Lévesque, C. M. (2012). *Can. J. Microbiol.* **58**, 553–562.
Tan, Q., Awano, N. & Inouye, M. (2011). *Mol. Microbiol.* **79**, 109–118.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Yamaguchi, Y., Park, J.-H. & Inouye, M. (2011). *Annu. Rev. Genet.* **45**, 61–79.
Yu, J., Zhou, Y., Tanaka, I. & Yao, M. (2010). *Bioinformatics*, **26**, 46–52.
Zubini, P., Zambelli, B., Musiani, F., Ciurli, S., Bertolini, P. & Baraldi, E. (2009). *Plant Physiol.* **150**, 1235–1247.