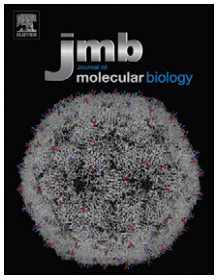

Structure and Function of AmtR in *Mycobacterium smegmatis*: Implications for Post-Transcriptional Regulation of Urea Metabolism through a Small Antisense RNA

Michael Petridis[1,†], Chelsea Vickers[2,†], Jennifer Robson[1], Joanna L. McKenzie[2], Magdalena Bereza[2], Abigail Sharrock[2], Htin Lin Aung[1], Vickery L. Arcus[2] and Gregory M. Cook[1,3]

*1 - **Department of Microbiology and Immunology,** Otago School of Medical Sciences, University of Otago, Dunedin 9054, New Zealand*
*2 - **Department of Biological Sciences,** University of Waikato, Hamilton 3240, New Zealand*
*3 - **Maurice Wilkins Centre for Molecular Biodiscovery,** The University of Auckland, Private Bag 92019, Auckland 1042, New Zealand*

Correspondence to Vickery L. Arcus and Gregory M. Cook: *varcus@waikato.ac.nz; gregory.cook@otago.ac.nz*
http://dx.doi.org/10.1016/j.jmb.2016.09.009
Edited by M Gottesman

Abstract

Soil-dwelling bacteria of the phylum actinomycetes generally harbor either GlnR or AmtR as a global regulator of nitrogen metabolism. *Mycobacterium smegmatis* harbors both of these canonical regulators; GlnR regulates the expression of key genes involved in nitrogen metabolism, while the function and signal transduction pathway of AmtR in *M. smegmatis* remains largely unknown. Here, we report the structure and function of the *M. smegmatis* AmtR and describe the role of AmtR in the regulation of nitrogen metabolism in response to nitrogen availability. To determine the function of AmtR in *M. smegmatis*, we performed genome-wide expression profiling comparing the wild-type *versus* an *ΔamtR* mutant and identified significant changes in the expression of 11 genes, including an operon involved in urea degradation. An AmtR consensus-binding motif (CTGTC-N_4-GACAG) was identified in the promoter region of this operon, and ligand-independent, high-affinity AmtR binding was validated by both electrophoretic mobility shift assays and surface plasmon resonance measurements. We confirmed the transcription of a *cis*-encoded small RNA complementary to the gene encoding AmtR under nitrogen excess, and we propose a post-transcriptional regulatory mechanism for AmtR. The three-dimensional X-ray structure of AmtR at 2.0 Å revealed an overall TetR-like dimeric structure, and the alignment of the *M. smegmatis* AmtR and *Corynebacterium glutamicum* AmtR regulatory domains showed poor structural conservation, providing a potential explanation for the lack of *M. smegmatis* AmtR interaction with the adenylylated P_{II} protein. Taken together, our data suggest an AmtR (repressor)/GlnR (activator) competitive binding mechanism for transcriptional regulation of urea metabolism that is controlled by a *cis*-encoded small antisense RNA.

Introduction

Nitrogen is an essential component of macromolecules (e.g., nucleic acids, proteins, and cell wall components), and bacteria have developed elaborate mechanisms for the uptake, assimilation, and regulation of nitrogen source utilization. Actinomycetes inhabit diverse environments, and some members of the soil-dwelling actinomycetes, for example, *Streptomyces coelicolor, Corynebacterium glutamicum*, and *Mycobacterium smegmatis* harbor two distinct global nitrogen regulatory proteins, GlnR and AmtR [1–3]. GlnR is an OmpR-like transcriptional regulator, which has been characterized in *S. coelicolor* [1,4]. The global regulator of nitrogen metabolism in *C. glutamicum* is AmtR, which mediates the expression of at least 35 genes in response to changing nitrogen levels, including genes implicated in ammonium uptake (*amtA, amtB*), ammonium assimilation (*glnA, gltBD*), urea metabolism (*ureABCEFGD*), and regulatory proteins (*glnK, glnD*) [2,5,6].

M. smegmatis has homologs for both GlnR and AmtR, whereas members of the *Mycobacterium tuberculosis* complex encode only for GlnR [7,8]. In *M. smegmatis*, GlnR was reported to regulate the expression of over 100 genes in response to changing

nitrogen levels, including multiple nitrogen uptake systems, and ammonium assimilation mechanisms [3]. Transcription of the urease-encoding operon is not under the regulatory control of GlnR, whereas an operon encoding proteins involved in a putative, alternative ATP-dependent urea-degrading metabolic pathway (*msmeg_2184-msmeg_2189)* was shown to be regulated by GlnR, and the expression of this operon was upregulated in response to nitrogen limitation [3]. While the genome-wide AmtR regulon of *M. smegmatis* is unknown, this operon has also been shown to be upregulated in a *M. smegmatis* Δ*amtR* mutant in response to nitrogen limitation [9].

AmtR belongs to the TetR-like protein family, which is abundant in bacteria that are exposed to environmental changes, such as those found in soil. The members of TetR family are DNA-binding proteins and function generally as transcriptional repressors. This family of proteins are involved in regulating cellular processes such as antibiotic resistance, efflux pumps, osmotic stress, and pathogenicity [10]. Members of the TetR family usually bind small effector molecules; however, in *C. glutamicum*, AmtR interacts with the P_{II} protein GlnK to regulate the expression of the genes within its regulon in response to changing nitrogen levels [2,5,6]. In this model, an AmtR:GlnK interaction depends on the nitrogen supply of the cell, which has an immediate effect on the state of GlnK (unmodified form or adenylylated form). GlnK is adenylylated (AMP) in response to nitrogen limitation, and in this form, AmtR interacts with GlnK-AMP, which results in abolishment of the repression of the genes in the AmtR regulon [2]. In regard to the mechanism of AmtR regulation in *M. smegmatis*, it is still unknown whether the signaling pathway involves an effector molecule, another protein, or a novel mode of regulation. Here, we report the structure and function of AmtR in *M. smegmatis* and propose a model of post-transcriptional regulation of urea metabolism mediated by a *cis*-encoded small antisense RNA (asRNA) that is complementary to *msmeg_4300 (amtR)*.

Results and Discussion

Molecular analysis of mycobacterial AmtR and phenotypic characterization of a *M. smegmatis* Δ*amtR* mutant

The AmtR protein encoded by *msmeg_4300* is 225 aa in length and belongs to the TetR/AcrR family of transcriptional regulators (Fig. S1). Bioinformatic analysis demonstrated that AmtR shares the highest amino acid sequence identity with AmtR homologs from other *Mycobacterium* species such as *Mycobacterium mageritense* (81%; Uniprot X5L5Y0) and *Mycobacterium vulneris* (79%; Uniprot X5LSM6), but it shares only 41% sequence identity with the well-studied AmtR from *C. glutamicum* (Uniprot Q79VH8; Fig. S1). The region of highest sequence similarity (~65%) is observed in the N-terminal DNA-binding domain, while the sequence similarity in the C-terminal ligand-binding domain is only ~20% to AmtR from *C. glutamicum*. A BLAST search using *C. glutamicum* AmtR and *M. smegmatis* AmtR against the *M. tuberculosis* complex shows no homologs with an amino acid identity over 33%, suggesting the lack of AmtR-like homologs in bacteria that belong to the *M. tuberculosis* complex [8].

In *M. smegmatis*, the *amtR* gene is flanked by genes that encode for an acyl-CoA synthase (*msmeg_4301*) and an enoyl-CoA hydratase/isomerase (*msmeg_4299*) with proposed roles in fatty acid and phospholipid metabolism, respectively (Fig. S2). To investigate the role of AmtR in *M. smegmatis*, we first confirmed whether the *amtR* gene was polycistronic with the flanking genes. Reverse transcriptase PCR and DNA sequencing of amplicons with primers that bind within the *amtR* gene alone (Fig. S2, PCR product II) and on either side of the gene of interest (Fig. S2, PCR products I and III) confirmed that *amtR* was a single, independent transcript under the conditions tested. We then created a markerless *amtR* deletion mutant (*M. smegmatis* JR258 Δ*amtR*; Table S1) and confirmed the deletion of 65% of the internal region of *amtR* with Southern hybridization and PCR (Fig. S3). To identify a phenotype of the JR258 Δ*amtR* mutant, we compared the growth between the wild-type strain and the JR258 Δ*amtR* mutant in Hartmans-de Bont (HdB) minimal medium with different nitrogen sources. Growth of the wild-type and JR258 Δ*amtR* mutant on the Phenotypic Micro-Array plate PM3 from Biolog Omnilog Systems™ (i.e., 95 different sole nitrogen sources tested) was comparable for the majority of the nitrogen sources tested [e.g., $(NH_4)_2SO_4$ shown in Fig. 1, and for other nitrogen sources tested, see Fig. S4]. We did observe a growth phenotype between the wild-type and JR258 Δ*amtR* mutant when lysine was added as the sole nitrogen source (Fig. 1b). Under these growth conditions, the JR258 Δ*amtR* mutant revealed a slower doubling time compared to the wild-type strain (wild-type 18.6 ±1.4 h *versus* JR258 Δ*amtR* 28.4 ± 2.7 h), and we confirmed this slow growth rate phenotype with colony-forming unit (CFU ml^{-1}) measurements (Fig. 1c). To validate the growth rate phenotype of the JR258 Δ*amtR* mutant, we successfully performed a complementation of JR258 Δ*amtR* with a tetracycline-inducible vector carrying the *amtR* gene (pMind-*amtR*) of *M. smegmatis* (Fig. 1d). To determine the metabolic basis for the lower growth rate of the JR258 Δ*amtR* mutant, we measured the cellular dry weight and molar growth yield ($Y_{glycerol}$) when cells were grown under these conditions (Fig. S5a and b). This analysis revealed that the cell dry weights for the wild-type (0.93 ± 0.02 g) and the

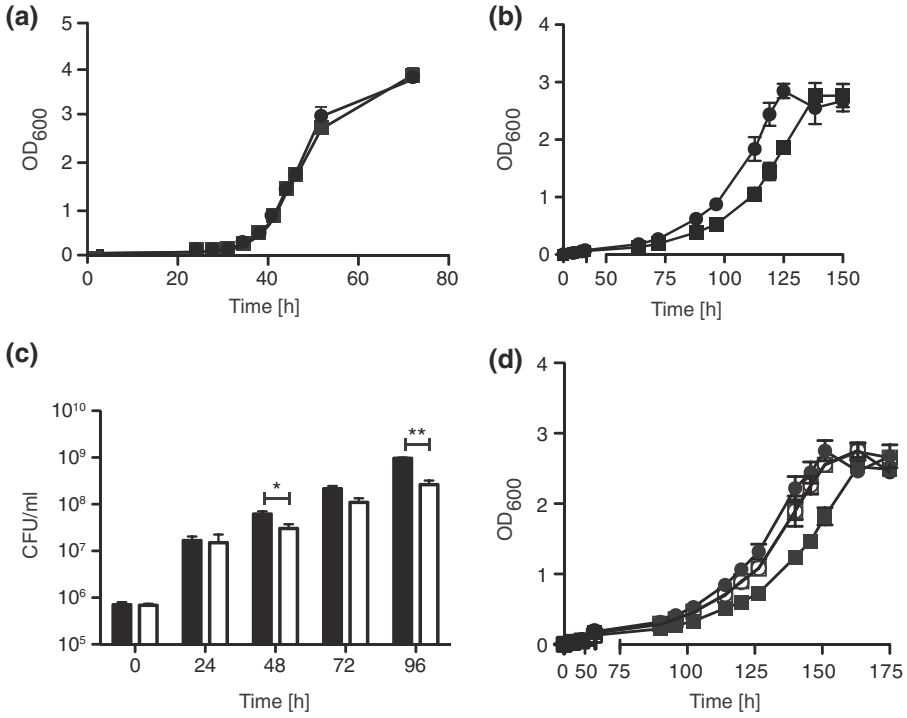

Fig. 1. Growth comparison of strains wild-type mc^{2}155 and JR258 Δ*amtR*. Comparative growth analyses by OD_{600} measurements were performed by growing the strains in HdB minimal medium containing 0.2% (wt/vol) glycerol as carbon source and (a) 1.5 mM $(NH_4)_2SO_4$ or (b) 1 mM L-lysine as sole nitrogen source (wild-type, closed circles; JR258, closed squares). (c) Colony-forming unit (CFU) counts of panel (b) comparing CFU of wild-type strain mc^{2}155 (black bars) and mutant JR258 Δ*amtR* (white bars). (d) Complementation of JR258 Δ*amtR* (open squares) with a tetracycline-inducible pMind-*amtR* construct. Wild-type control with pMind-*amtR* (open circles) and empty vector controls (pMind) were used as negative controls in the background of both the wild-type (closed circles) and JR258 (closed squares). Error bars represent the standard error of the mean of three independent biological replicates. *$P < 0.05$; **$P < 0.01$.

JR258 Δ*amtR* mutant (0.90 ± 0.02 g) were comparable (Fig. S5a), consistent with the OD_{600} measurements (Fig. 1b). However, the molar growth yield on glycerol ($Y_{glycerol}$) was significantly lower for the JR258 Δ*amtR* mutant (35.4 ± 0.5 g cells mol^{-1} glycerol utilized) compared to the wild-type strain (38.2 ± 0.8 g cells mol^{-1} glycerol utilized; Fig. S5b), suggesting a greater energetic demand (ATP consumption) on the JR258 Δ*amtR* mutant during growth with lysine as the sole nitrogen source.

Genome-wide regulation by AmtR: repression of a pathway for urea degradation and biotin synthesis

To gain further insight into the role of AmtR and the molecular basis for the lower $Y_{glycerol}$ in the Δ*amtR* mutant, we compared the transcriptional response of the JR258 Δ*amtR* mutant with the wild-type using microarray analysis. Cells of both strains were grown in HdB minimal medium containing 0.2% (wt/vol) glycerol as carbon source and 1 mM L-lysine as the sole nitrogen source and were harvested at an OD_{600} between 0.15 and 0.18 to avoid a general stress response linked to the slower growth of the JR258 Δ*amtR* mutant. Comparison of the transcriptional response of the JR258 Δ*amtR* mutant with the wild-type revealed that a total of 11 genes were differentially expressed (>2.0× change, $P < 0.05$, and four independent biological replicates, including two dye swaps). Nine genes were upregulated and two genes were downregulated in the JR258 Δ*amtR* mutant (Table 1). The microarray data were validated by quantitative real-time PCR (qPCR) of seven selected genes (Fig. S6).

Five of the most highly upregulated genes, *msmeg_2184*, *msmeg_2185*, *msmeg_2186*, *msmeg_2187*, and *msmeg_2189*, were co-transcribed within an operon (Table 1). The genes *msmeg_2187* and *msmeg_2189* (*atzF*) were upregulated 30-fold and 56-fold, respectively, and are annotated as urea amidolyase (UA) and allophanate hydrolase (AH), while *msmeg_2185* and *msmeg_2186* encode for hypothetical proteins that were upregulated 26-fold and 42-fold, respectively, in the JR258 Δ*amtR* mutant. A BLAST search against the Uniprot database revealed an identity between 70 and 80% of these two putative proteins to a urea carboxylase (UC)-associated

Table 1. List of genes that are differentially expressed in strain JR258 Δ*amtR* compared to strain wild-type mc^{2}155

mc^{2}155 locus[a]	Expression ratio[b]	*p* value	Description
msmeg_2189	56.04	7.50E-02	allophanate hydrolase
msmeg_2186	42.72	3.28E-04	conserved hypothetical protein
msmeg_2184	39.79	1.43E-02	amino acid permease
msmeg_2187	29.86	1.67E-02	urea amidolyase
msmeg_2185	26.15	4.13E-03	conserved hypothetical protein
msmeg_4588	2.90	4.63E-02	ABC nitrate/sulfonate/bicarbonate family protein transporter, inner membrane subunit
msmeg_5594	2.27	3.50E-02	ferredoxin-dependent glutamate synthase
msmeg_4925	2.20	4.77E-02	transcriptional regulator
msmeg_6770	2.07	7.69E-02	conserved hypothetical protein
msmeg_6241	0.50	3.76E-03	ATPase associated with various cellular activities
msmeg_6240	0.46	1.18E-02	conserved hypothetical protein

[a] Locus number of gene in *M. smegmatis* mc^{2}155
[b] Mean of the gene expression ratio of four biological replicates

protein 1 (*msmeg_2186*) and UC-associated protein 2 (*msmeg_2185*) of *Gordonia bronchialis* and some other *Gordonia* species. Genes encoding for a UA and an AH are, like AmtR, only present in some soil-dwelling actinobacteria, for example, *M. smegmatis*, *Mycobacterium vanbaalenii*, *Mycobacterium abscessus*, and *Streptomyces avermitilis*.

UA consists of two distinct enzyme activities: UC and AH. This UC/AH enzyme complex provides a biotin- and ATP-dependent pathway for urea degradation, catalyzing a two-step reaction to recover ammonium via the intermediate allophanate. Along with the aforementioned genes, we also observed a 1.7-fold upregulation of genes encoding for a biotin synthase and a biotin ligase in the JR258 Δ*amtR* mutant (Fig. S6). The UC harbors several domains, a biotin carboxylase, a carboxyltransferase, and a biotin carboxyl carrier protein. In an ATP-dependent step, the biotin carboxylase domain carboxylates biotin that is covalently linked to the biotin carboxyl carrier protein domain and the carboxyltransferase domain subsequently transfers this carboxyl group from biotin to urea, producing allophanate. However, *M. smegmatis* encodes for a Ni^{2+}-dependent urease enzyme that catalyzes a very similar but energy-independent reaction, converting urea into ammonium [7]. Furthermore, the gene *msmeg_2184* is annotated as an amino acid permease and was upregulated 40-fold in the JR258 Δ*amtR* mutant. However, its specific function is unknown, while the genes *msmeg_6240* (conserved hypothetical protein) and *msmeg_6241* (AAA-ATPase) are downregulated 2-fold in the JR258 Δ*amtR* mutant. Expression of these two genes has been previously shown to be affected by growth rate and hypoxia, both conditions that alter the energetic level of bacteria similar to the situation in the JR258 Δ*amtR* mutant [11].

These data suggest that AmtR represses genes for urea degradation, and the ATP-consuming reaction of this pathway may indeed explain the lower $Y_{glycerol}$ observed in the JR258 Δ*amtR* mutant. We also compared the growth of wild-type to the JR258 Δ*amtR* mutant in a carbon-limited, continuous culture with a doubling time (t_d) of 24 h to simulate severe energy (carbon)-limiting conditions [11]; however, we did not observe significant differences in either the final OD_{600} or the $Y_{glycerol}$ measurements during steady-state growth, suggesting that AmtR was not controlling carbon utilization under these conditions (data not shown).

M. smegmatis AmtR binds to a palindromic sequence in the promoter region of the *msmeg_2184–msmeg_2189* operon that encodes for the ATP-dependent urea-degrading pathway

Previous work has described both a GlnR- and an AmtR-binding site in the promoter region of *msmeg_2184*, the first gene in the operon that encodes for the ATP-dependent urea-degrading pathway [9]. To measure AmtR binding, a 186-bp digoxigenin (DIG)-labeled PCR product including the promoter region of *msmeg_2184* (P_{2184}) was amplified and purified; the AmtR protein from *M. smegmatis* was used for electrophoretic mobility shift assays (EMSA; Fig. 2a). The AmtR protein caused a mobility shift of P_{2184} at a molar DNA:protein ratio of 1:25, with a complete shift occurring at a ratio of 1:50 (Fig. 2a, lane 4). To determine the specificity of binding to P_{2184}, we added unlabeled P_{2184} DNA in 120-fold excess (Fig. 2a, lane 5), resulting in the complete abolition of the shift. Unlabeled, non-specific DNA (promoter region of the *M. smegmatis phoH2* gene) was added in 103-fold excess and did not affect the mobility shift of P_{2184} (Fig. 2a, lane 6). This clearly demonstrates that DNA retardation is specific for the promoter region of *msmeg_2184*. We identified a palindromic sequence (CTGTC-N_4-GACAG, −55 to −42; Fig. 2b and c) that is located in close proximity to the GlnR-binding site (−107 to −91) [3] in the promoter region of *msmeg_2184* (Fig. 2c). To verify AmtR binding at the palindromic sequence, we mutated this putative

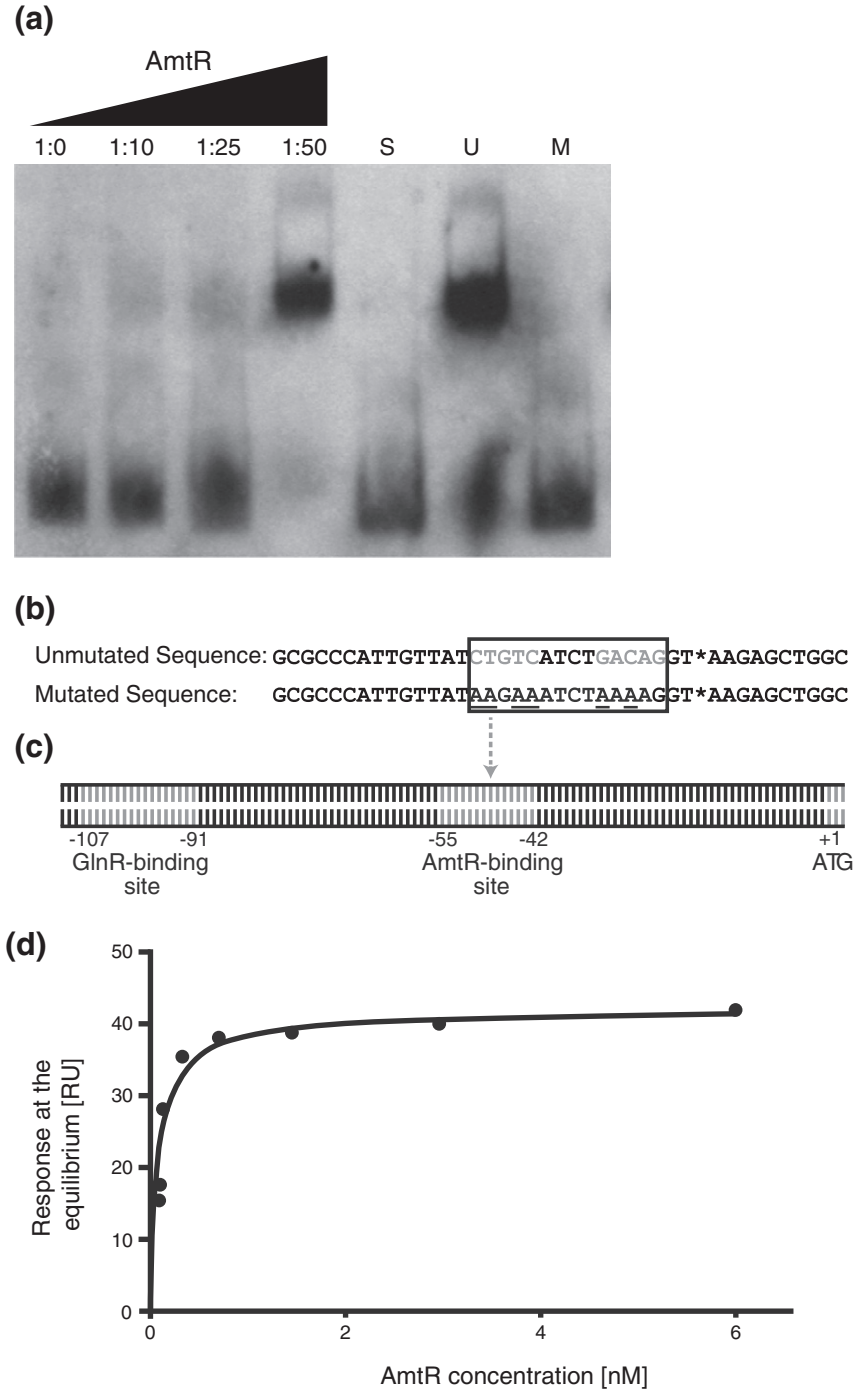

Fig. 2. *In vitro* DNA binding of purified AmtR protein. (a) Electrophoretic mobility shift assay (EMSA) with AmtR protein and a DIG-labeled 186-bp DNA fragment including the putative AmtR-binding site in P_{msmeg_2184}. The molar ratios of protein to DNA are indicated. Controls for binding specificity include: lane S, 103-fold excess of unlabeled P_{msmeg_2184}; lane U, 103-fold excess of unlabeled promoter region of the *phoH2* gene; lane M, EMSA using a mutated (see panel b) promoter region of *msmeg_2184*. (b) Unmutated AmtR-binding region with an inverted repeat (box) in P_{msmeg_2184}. Nucleotides in the inverted repeat were mutated (underlined) using site-directed mutagenesis. (c) The promoter region of *msmeg_2184* including the confirmed GlnR- and AmtR-binding sites. (d) AmtR–DNA interaction as determined by SPR. The AmtR responses were double-referenced by subtracting the data from the reference flow cell and buffer responses. The equilibrium dissociation binding constant (K_D) of 111 ± 16 pM was calculated using the steady-state/equilibrium binding model including the standard error.

AmtR-binding motif by replacing nucleotides in the inverted repeats using site-directed mutagenesis (Fig. 2b). DNA retardation was abolished with the mutated promoter region of *msmeg_2184*, validating the binding of AmtR to this sequence-specific site (Fig. 2a, lane 7).

EMSA demonstrates that AmtR binds to the palindromic sequence in the promoter region of the UC/amidase hydrolase operon. To quantify the affinity of the AmtR interaction with this DNA fragment, we performed surface plasmon resonance (SPR) measurements. Due to a very weak dissociation of AmtR from the DNA (Fig. S7), we used the steady-state binding model at the equilibrium to calculate the dissociation constant (K_D). For the AmtR–DNA interaction, the K_D calculated from binding at equilibrium was 111 ± 16 pM (Fig. 2d). Furthermore, this high-affinity interaction is acutely specific, as an alternative DNA sequence was immobilized in the control flow cell. This very tight and specific binding of AmtR to its target DNA suggests that even very low levels of AmtR protein in the cell will repress the urea degradation operon that lies downstream of this recognition sequence.

Using the Motif Alignment & Search Tool of the MEME software, we analyzed the conservation of an AmtR consensus-binding motif in the promoter region of genes in the genome of *M. smegmatis*. We used the confirmed motifs of *C. glutamicum* (ATCTATAG-N_4-ATAGN) [5] and *M. smegmatis* (TATCTGTC-N_4-GACAG in the promoter region of *msmeg_2184*) as input to screen the upstream regions (250 bp upstream of +1) of all annotated *M. smegmatis* genes. An *M. smegmatis* AmtR-binding motif was only present upstream of *msmeg_2184, msmeg_3966*, and *msmeg_0426*; however, we observed only the changes in expression for *msmeg_2184* and *msmeg_3966* in our microarray analysis, suggesting that the regulation of genes differentially expressed in the JR258 Δ*amtR* mutant without an AmtR-binding motif may be due to indirect effects involving other regulators. No *C. glutamicum* AmtR-binding motif was detected in the genome of *M. smegmatis*.

A *trans*-acting small asRNA is upregulated under nitrogen excess and post-transcriptionally regulates the expression of *msmeg_4300*

Recent work revealed an extensive number of non-coding RNAs including small RNAs (sRNAs) in mycobacteria that are between 50 and 500 nt in length [12,13]. More than 40 sRNAs with a regulatory function have been experimentally confirmed in *M. smegmatis,* while the presence of more than 100 sRNAs is computationally predicted [14]. Previous work on defining the nitrogen-regulated transcriptome of *M. smegmatis* using nitrogen-limited chemostat culture demonstrated the presence of a novel *cis*-encoded small asRNA that is complementary to *msmeg_4300* that was induced 30-fold under nitrogen excess [15]. To verify this prediction and localize the asRNA, we used sequence-specific primers for cDNA synthesis, followed by qPCR with three primer pairs spanning 150-bp regions within the *msmeg_4300* asRNA (Fig. 3a). Primer pair 1 spans the region nc*msmeg_4300*c-1 that is located in the 5′-end of the antisense strand and that was expressed under nitrogen excess, while primer pairs 2 and 3 span the regions nc*msmeg_4300*c-2 and nc*msmeg_4300*c-3 that are located in the middle and at the 3′-end of the antisense strand and that did not show expression in RNA sequencing, respectively [15]. We confirmed an 11-fold upregulation of nc*msmeg_4300*c-1 under nitrogen excess, while we could not detect significant changes in expression for nc*msmeg_4300*c-2 and nc*msmeg_4300*c-3, respectively (Fig. 3b). The presence of a *cis*-encoded asRNA nc*msmeg_4300*c will impact upon the transcription of the mRNA encoded opposite and might result in the blocking of translation and in directing mRNA processing, suggesting a post-transcriptional regulatory mechanism for AmtR, where the *cis*-encoded asRNA

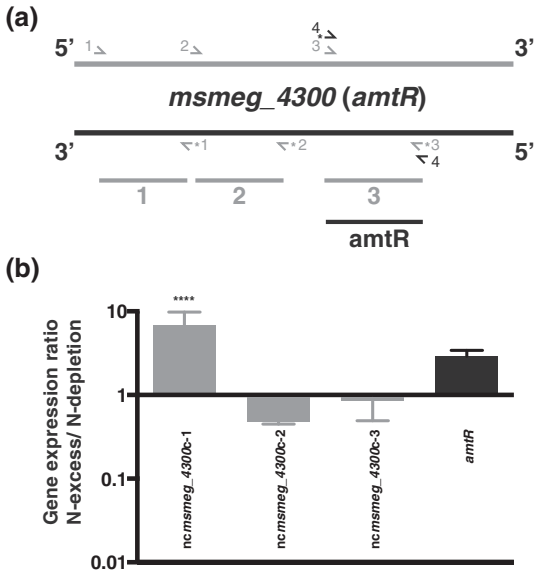

Fig. 3. qPCR validation of the *cis*-encoded small asRNA nc*msmeg_4300*c. (a) The *amtR* gene with the sense strand (black) and antisense strand (gray). The primer pairs (Table S1) 1–3 (gray arrows) span 150-bp regions (1: nc*msmeg_4300*c-1; 2: nc*msmeg_4300*c-2; 3: nc*msmeg_4300*c-3; gray bars) within the *msmeg_4300* asRNA, while primer pair 4 (black arrows) spans a 157-bp region (*amtR*; black bar) within the *msmeg_4300* mRNA. Strand-specific primers for cDNA synthesis are indicated (*). (b) Validation of the expression of nc*msmeg_4300*c by qPCR. Results shown represent the expression ratio in response to nitrogen excess as mean ± SD of three independent biological replicates. ****$P < 0.0001$.

nc*msmeg_4300*c base pairs with the complementary target RNA of *msmeg_4300 (amtR)*.

Structure of *M. smegmatis* homodimeric AmtR is consistent with a TetR-like regulator and lack of interaction with GlnK

The 3D X-ray structure of AmtR was solved at 2.0-Å resolution in two steps. Firstly, the putative ligand-binding domain was solved using single anomalous diffraction (SAD) of a mercury derivative. The structure of the ligand-binding domain was then used as a molecular replacement model to solve the full-length AmtR structure (Fig. 4a). A critical step in obtaining high-quality crystals of the full-length AmtR protein was a DNase step to remove all the DNA that co-purified with the recombinant protein (see Materials and Methods). The structure of AmtR is a homodimer with each monomer composed of 12 α-helices (with connecting turns and loops) constituting two domains: a DNA-binding domain and a regulatory domain responsible for dimerization (Fig. 4a). The structure conforms to the archetypal TetR-like fold [16].

The N-terminal domain contains a helix-turn-helix (HTH) DNA-binding motif and includes helices α1, α2, and α3. This domain contains highly conserved residues involved in DNA binding (residues 27–48) that are typical for TetR proteins [10]. A fourth helix, α4, is responsible for connecting the HTH motif to the regulatory domain. The C-terminal regulatory domain is composed of helices α5–α10 and is less conserved, consistent with the diversity of signaling molecules recognized by different TetR-like proteins involved in an array of different regulatory roles [10]. Helix α8 and helix α9 constitute the dimerization interface forming an orthogonally packed four-helix bundle with the same helices (helix α8′ and helix α9′) from the other monomer (Fig. 4a). Each monomer contains a central cavity that is approximately 126 $Å^3$ and is defined by nine residues, of which three are aromatic (Asp_{105}, Leu_{109}, Tyr_{120}, Phe_{133}, Leu_{140}, Arg_{141}, Tyr_{144} and Glu_{170}, and Iso_{173}), which are accessed by a long entrance tunnel (Fig. 4b). Glu_{170} on helix α9 shows electron density for alternative confirmation and appears to form a gate to the pocket that extends into a well-defined entrance site that is formed by helices α6–α9 (Fig. 4c). The position of Glu_{170} may be dictated by the conformation of AmtR (i.e., unbound or bound to DNA) and may in turn play a stabilization role in ligand binding.

The AmtR structure was compared with similar structures in the PDB database. The closest alignments were the AmtR structure from *C. glutamicum* (PDB code 5DY1; RMSD = 1.40 Å, sequence identity = 41%), a putative TetR-like protein from *Rhodococcus jostii* RHA1 (PDB code 3HIM; RMSD = 2.4 Å; sequence identity = 23%), and the phenylacetic acid and fatty acid metabolism regulator (PfmR) from the thermophilic bacterium, *Thermus thermophilus* HB8 (PDB code 3VPR; RMSD = 2.3 Å, sequence identity = 18%). Each of the above structures is in the apo form (not in complex with DNA), and structural conservation for all alignments was observed primarily in the DNA-binding domain, particularly helix α3 of the HTH motif. For example, the alignment of the apo-*M. smegmatis* AmtR and *C. glutamicum* AmtR structures is given in Fig. 4d. Of the 11 DNA-binding residues (Arg_9, Arg_{10}, Tyr_{42}, Arg_{52}, Gln_{53}, Ala_{54}, Ser_{55}, Tyr_{57}, Tyr_{58}, His_{59}, and Lys_{63}) present in the DNA-binding domain of the *C. glutamicum* AmtR [17], 9 are conserved in the *M. smegmatis* AmtR DNA-binding domain (Arg_{11}, Tyr_{44}, Arg_{54} Gln_{55}, Ala_{56}, Ser_{57}, Tyr_{59}, His_{61}, and Lys_{65}, respectively). The *M. smegmatis* AmtR has an additional two residues in the N-terminal region at position 8.

The *M. smegmatis* apo-AmtR structure was then aligned against the *C. glutamicum* AmtR structure in complex with DNA (PDB code 5DY0; RMSD = 1.94 Å, sequence identity = 37%). The most notable region of poor alignment between these structures in the DNA-binding domain occurs in helix α3 of the HTH motif, which sits into the major groove of the DNA and contains most of the identified DNA-binding residues. The most apparent residue displacement is at the arginine residue (Arg_{52} and Arg_{54} in the *C. glutamicum* and *M. smegmatis* structures, respectively) located in the N terminus of helix α3. In the alignment between the apo-AmtR structures, this arginine residue is similarly orientated and differs by a distance of 1.8 Å at Cβ (Fig. 4e). In the alignment of the *M. smegmatis* apo-AmtR structure with the *C. glutamicum* AmtR structure in complex with DNA, the orientation of this arginine differs significantly with a distance of 4.2 Å at Cβ (Fig. 4f). In the *C. glutamicum* AmtR structure in complex with DNA, the change in orientation of Arg_{52} results in the residue projecting directly into the major groove of the DNA to interact with both the DNA phosphate backbone and the adjacent nucleotides (Fig. 4f). This observation implicates Arg_{54} in the *M. smegmatis* AmtR as an important residue for DNA binding. Interestingly, Palanca and Rubio described a 19-bp, arginine-rich N-terminal extension of the DNA-binding domain as an essential anchor for DNA binding [17]. They confirmed the anchoring role by truncating the N terminus of AmtR to exclude the anchor sequence, which resulted in abolished DNA binding. They suggest that the role of the anchor with its inherent flexibility and apparent binding to AT pairs is to "search" for these regions in high GC genomes and to enhance DNA binding. The extension was exclusively observed in the structure in complex with DNA, suggesting the disorder of this region until it can bind into the minor groove of DNA and anchor the rest of the HTH motif to the major groove of DNA. A similar arginine-rich N-terminal extension is also observed in

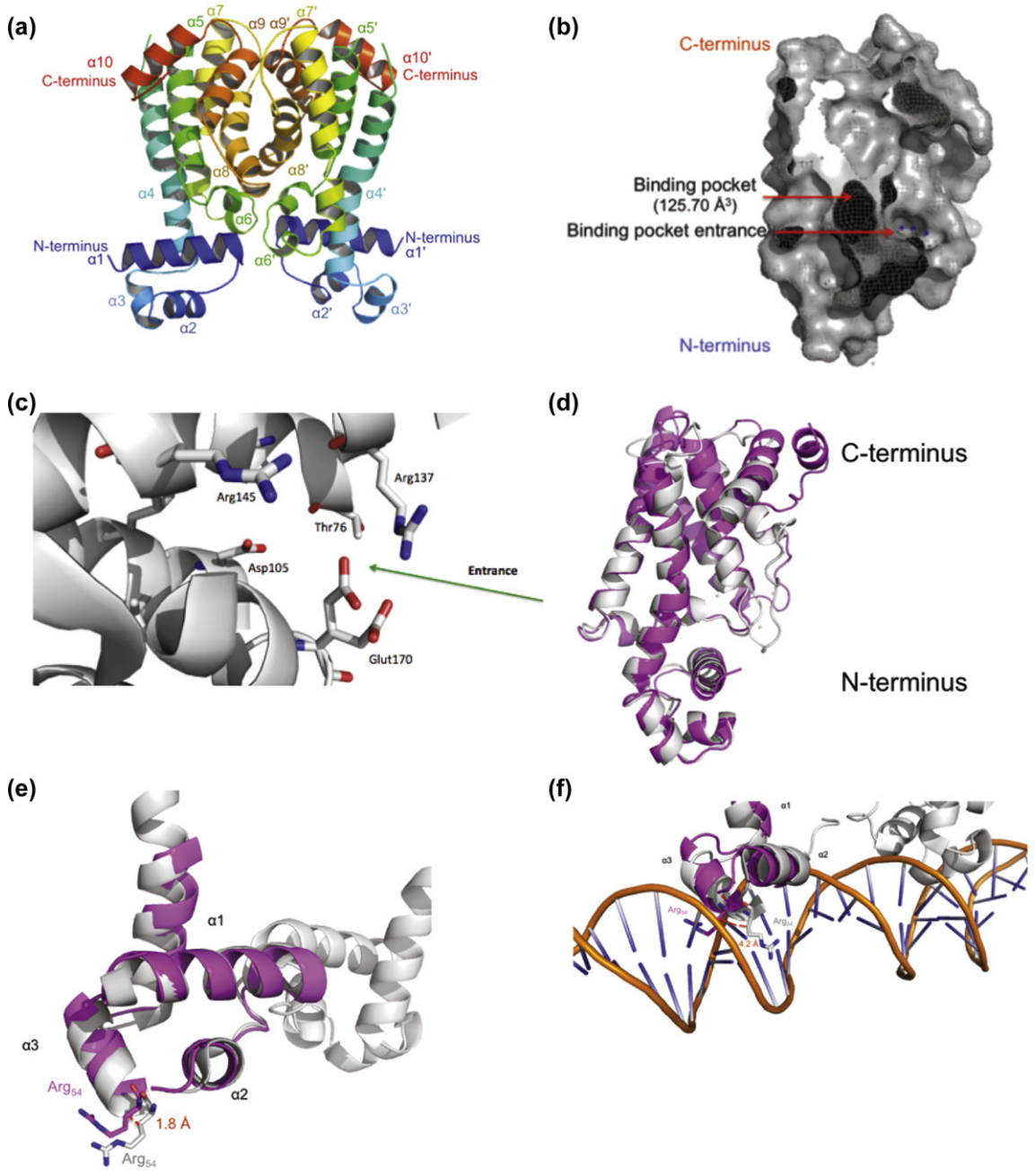

Fig. 4. Overall structure of the AmtR protein from *M. smegmatis*. (a) The full-length X-ray crystallographic structure of *M. smegmatis* AmtR. The structure was solved by molecular replacement to 1.98 Å with C 2 symmetry. The *M. smegmatis* AmtR structure is a homodimer with each monomer composed of 10 α-helices constituting an N-terminal HTH DNA-binding domain and a C-terminal ligand-binding domain. (b) A mesh-lined space-filling model of the *M. smegmatis* AmtR LBD monomer structure showing the long entrance tunnel along the gray surface with three water molecules (in blue). The ligand-binding site is shown in black with a mesh surface. (c) The ligand-binding pocket of *M. smegmatis* AmtR showing the putative interacting residues in binding pocket as sticks from helices α5, α6, α7, and α8. Glu_{170} is shown with its alternate confirmations illustrating its potential role as a gate to the binding pocket. (d) Alignment of the *M. smegmatis* AmtR monomer (magenta) to the *C. glutamicum* AmtR monomer (gray). Structural conservation is observed primarily in N-terminal DNA-binding domain and not in C-terminal regulatory domain. (e) Alignment of the *M. smegmatis* apo-AmtR DNA-binding domain (magenta) to the *C. glutamicum* apo-AmtR DNA-binding domain (gray). In the apo form, the orientation of the conserved N-terminal arginine (Arg_{54} and Arg_{52} in *M. smegmatis* and *C. glutamicum*, respectively) on helix α3 is similarly orientated and differs by 1.8 Å at the cβ of the residue. (f) Alignment of the *M. smegmatis* apo-AmtR DNA-binding domain (magenta) to the complexed *C. glutamicum* AmtR DNA-binding domain (gray). This alignment shows that the conserved N-terminal arginine (Arg_{54} and Arg_{52} in *M. smegmatis* and *C. glutamicum*, respectively) on helix α3 is no longer similarly orientated and differs by 4.2 Å at the cβ of the residue.

the *M. smegmatis* AmtR, which shares 10 of the 19 residues with the anchor region of *C. glutamicum* AmtR. This anchor region is not observed in the structure of *M. smegmatis* apo-AmtR, suggesting that this region is also disordered in the absence of DNA and may be playing a similarly essential role in DNA binding as the *C. glutamicum* AmtR. The alignment of the *M. smegmatis* AmtR and *C. glutamicum* AmtR regulatory domains shows poor structural conservation (Fig. 4d). This may be expected given the inherent differences of each regulatory region, that is, the *C. glutamicum* AmtR regulatory domain functions to interact with the adenylylated P_{II} signaling protein GlnK, which in turn dictates the DNA-binding state of AmtR. Further investigation is required to identify a potential signaling mechanism that has not yet been elucidated in *M. smegmatis* AmtR regulation.

Summary

We propose a regulatory model in which AmtR regulates the transcription of the *msmeg_2184-msmeg_2189* operon encoding for urea degradation through a very tight binding of AmtR at its promoter region (Fig. 5). The likely sensor in this regulation is a *cis*-encoded (*trans*-acting) sRNA nc*msmeg_4300*c that responds to nitrogen availability. Under conditions of nitrogen excess, strongly upregulated nc*msmeg_4300*c base pairs with the complementary target mRNA *msmeg_4300*, blocking the translation and mRNA processing. As cells become depleted for nitrogen, the transcript level of nc*msmeg_4300*c decreases, resulting in the effective translation of the *amtR* mRNA into functional protein and in the repression of urea metabolism through an AmtR/GlnR competitive binding mechanism (Fig. 5) [3]. The dual regulation of the putative ATP-dependent urea-degrading metabolic pathway (*msmeg_2184-msmeg_2189)* by GlnR (activation) and AmtR (repression) in response to nitrogen suggests that urea degradation is an important source of nitrogen for the soil saprophyte *M. smegmatis* (Fig. 5). Further work is needed to measure the relative affinity of GlnR and AmtR binding at this operon to determine the mechanism of dual regulatory control in response to nitrogen supply.

Materials and Methods

Bacterial strains, media, and growth conditions

Escherichia coli DH10B was grown in Luria-Bertani (LB) medium at 37 °C with agitation (200 rpm) or on LB agar plates. *M. smegmatis* mc^2155 and derived strains were grown in LB supplemented with 0.05% (wt/vol) Tween-80 (LBT) or HdB minimal medium [18] supplemented with 0.2% (wt/vol) glycerol (unless otherwise stated) and 0.05% (wt/vol) Tyloxapol at 37 °C with agitation (200 rpm). Growth curves were performed in triplicate. Bacterial cell viability was monitored by cell counts based on CFUs per ml, where serial dilutions of bacterial cell culture in phosphate-buffered saline with 0.05% Tween-80 (PBST) were plated on LBT agar plates. Chemostat growth conditions under nitrogen excess and nitrogen depletion were as previously described [15]. For *in vivo* nitrogen-limiting-induced adenylylation of GlnK, the *glnK gene* (msmeg_2426) was inserted into the *M. smegmatis* expression vector pYUB28b (pYUB28b + glnK). Primers for the amplification of the *glnK* gene from *M. smegmatis* mc^2155 genomic DNA were designed to include the 5′ restriction site NdeI and the 3′ restriction site HindIII necessary for ligation into pYUB28b with an N-terminal 6xHis tag. Extra 3–6 bases were included to allow for efficient cleavage, depending on the restriction enzyme used. The gene was digested with NdeI and HindIII and inserted into pYUB28b. The vector was transformed into *E. coli* TOP10 cells prior to transformation and protein expression in the *M. smegmatis* 4517 expression strain. The transformation was streaked out on a low-salt LBT agar plate. Selective media contained ampicillin (100 μg ml^{-1} for *E. coli*), kanamycin (50 μg ml^{-1} for *E. coli*, 20 μg ml^{-1} for *M. smegmatis*), hygromycin B (200 μg ml^{-1} for *E. coli*, 50 μg ml^{-1} for *M. smegmatis*), and gentamycin (25 μg ml^{-1} for *E. coli*, 5 μg ml^{-1} for *M. smegmatis*). All bacterial strains, plasmids, and primers used in this study are listed in Table S1.

Dry weight measurements and glycerol concentration determinations

Cultures were grown in HdB with 0.2% (wt/vol) glycerol as carbon source and 1 mM L-lysine as the sole nitrogen source until stationary phase, and 20 ml of each culture was harvested by rapid filtration through a nitrocellulose filter (pore size 0.22 μm; Millipore) with an applied vacuum of approximately 7.5 psi. Filters were washed twice with 20 ml PBST and dried at 60 °C until no change in weight could be determined, and dry weights were measured. For glycerol concentration determinations, samples were taken after culture inoculation to determine the starting glycerol concentration for each biological replicate and after entering stationary growth phase. Culture samples were centrifuged (5 min, 14,400*g*) to obtain cell-free supernatants. Glycerol concentrations were measured by detecting NADH oxidation (absorbance = 340 nm) as previously described [19].

Alignment and bioinformatics of AmtR proteins

Amino acid sequences of AmtR proteins from *C. glutamicum* (CAF19569.1), *S. avermitilis* (BAC74412.1), *M. smegmatis* (WP_011729715.1), *M. mageritense* (WP_036438965.1 and WP_019346862.1), *M. vulneris* (WP_036449407.1), *M. farcinogenes* (WP_036393868.1), *M. septicum* (WP_044523525.1), and *M. abscessus* (WP_005071231.1) were obtained from the National Center for Biotechnology Information. Sequences were aligned using Clustal Omega [20].

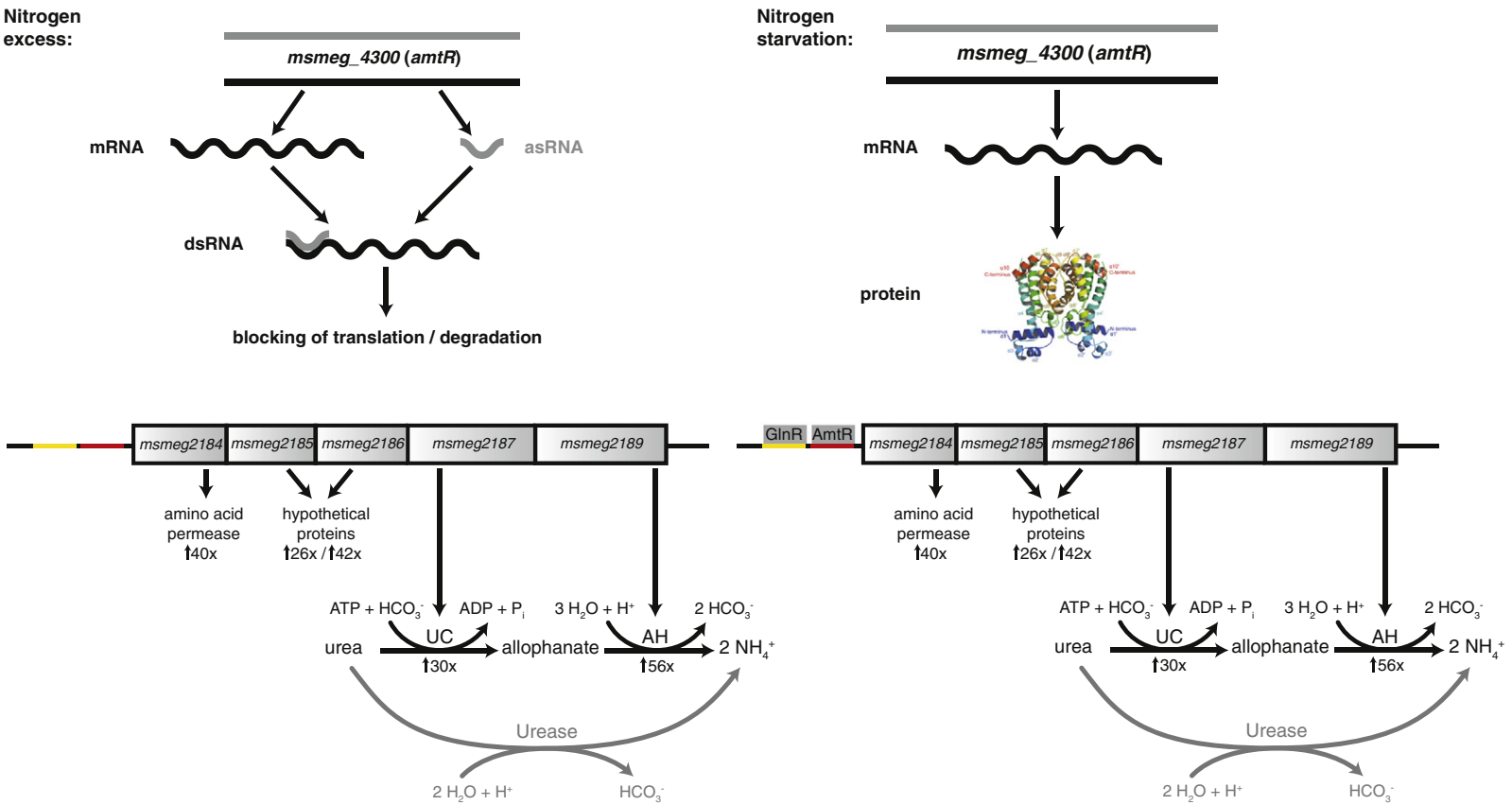

Fig. 5. Model for an AmtR/GlnR competitive binding mechanism for transcriptional regulation of urea metabolism. (a) During nitrogen excess conditions, transcription of a *cis*-encoded sRNA nc*msmeg_4300*c is enhanced, suggesting the formation of a double-stranded RNA molecule and subsequent mRNA degradation. (b) As cells become depleted for nitrogen, the transcription of nc*msmeg_4300*c is abolished, therefore resulting in translation of *amtR* mRNA into AmtR protein. The AmtR protein functions as transcriptional repressor of the UC/AH encoding operon through tight binding at an AmtR-binding site (red box) in the promoter region of *msmeg_2184*, while GlnR functions as a transcriptional activator of the same operon through binding at a GlnR-binding site under nitrogen limitation (yellow box). This dual regulatory control would allow for fine-tuning the expression of urea metabolism under changing environmental conditions.

Construction of *amtR* deletion mutant and complementation

The creation of an *amtR* markerless deletion was performed using the two-step method for integration and excision as previously described [21]. In brief, the regions flanking either side of the *amtR* gene were amplified by PCR using primer pairs AmtR LF F and AmtR LF o/lap R to obtain the left flank (LF; 920 bp) and AmtR RF o/lap F and AmtR RF R to obtain the right flank (RF; 883 bp) regions (Table S1). To create the markerless deletion insert, we used both left and right flanks as the template for an overlap PCR reaction. The assembled insert containing the left and right flank DNA was then cloned into the SpeI site of pPR23-derived vector pX33 generating pJR500, which was electroporated into *M. smegmatis* mc^2155 background [22,23]. For the selection of single recombination events, a culture of *M. smegmatis* mc^2155 harboring pJR500 was grown at 28 °C with agitation (200 rpm) to an OD_{600} of approximately 0.6–0.8. Aliquots of culture were plated onto LBT gentamicin and grown at 42 °C. Single colonies were selected and grown in LBT at 37 °C to an OD_{600} of 0.6. Aliquots of culture were plated onto low-salt (2 g NaCl l^{-1}) LBT plates containing 10% sucrose and were incubated at 42 °C, selecting for double-recombinant events. To confirm that the correct double recombination event had occurred, we performed Southern blot hybridization on PvuII-digested genomic DNA using the Amersham Gene Images AlkPhos Direct Labeling and Detection System with CDP-Star detection reagent (GE Healthcare) according to manufacturer's instructions, generating *M. smegmatis* JR258 Δ*amtR*.

To create an AmtR complementation construct, we amplified a PCR product from *M. smegmatis* containing the *amtR* gene using primers AmtR pMind RBS R and AmtR pMind F (Table S1). The reverse primer introduced a synthetic ribosome-binding sequence (GGAGG) upstream of the *amtR* gene, which was obtained from a nonpolar kanamycin resistance cassette (*aphA*-3) to ensure translation. The PCR product containing the *amtR* gene was then cloned into the tetracycline-inducible vector pMind, resulting in pMind-*amtR* [24].

Biolog phenotypic microarray analysis

Phenotype microarray (PM) was performed using Biolog Omnilog Systems™ (Biolog Inc., Hayward, CA). Strains were grown to an OD_{600} of 0.4–0.45 in 7H9 media containing 0.1% glycerol and 0.05% tyloxapol. Cells were washed twice in PBST, then resuspended in PBST and transferred to a sterile flask and incubated for 22 h at 37 °C and at 200 rpm. The cells were then harvested by centrifugation and were resuspended in PM IF-0a media (Biolog). The volume was adjusted to result in the same optical density for each culture (OD_{600} = 0.5). PM additive solution was made as per protocol for PM plate PM3. The protocol was modified to add cells to a final OD_{600} of 0.005 instead of 85% turbidity. Growth was determined by monitoring the purple color development due to tetrazolium dye as an indicator. The plates were incubated in a shaking incubator at 37 °C for 48 h, and metabolic activity was monitored at OD_{590} in a plate reader (Infinite® 200/ Tecan).

Microarray analysis and qPCR

To identify the AmtR regulon in *M. smegmatis*, we performed microarray analysis comparing strains JR258 Δ*amtR* and wild-type mc^2155 grown in HdB minimal medium with 0.2% (wt/vol) glycerol as the sole carbon source and 0.05% (wt/vol) Tyloxapol. We replaced $(NH_4)_2SO_4$ as the sole nitrogen and sulfur source with 11.4 mM K_2SO_4 as a sulfur source and 1 mM L-lysine as the sole nitrogen source. Single colonies were used as an inoculum into HdB for initial starter cultures and were grown overnight at 37 °C to an OD_{600} of 0.2–0.3. This was used as an inoculum for HdB to an OD_{600} of 0.001. Cells were harvested at an OD_{600} between 0.17 and 0.2. Total RNA was extracted using TRIzol reagent (Ambion) according to the manufacturer's instructions. The cells were lysed by four cycles in a mini-Beadbeater (Biospec Products) at 4800 rpm for 30 s and on ice between each of the cycles for 30 s. DNA was removed by treatment with 3 U RNase-free DNase using the TURBO DNA-free kit (Ambion) according to the manufacturer's instructions. The quality of the RNA was checked with the Bioanalyzer, and the concentration was determined using a NanoDrop ND-100 spectrophotometer. Reverse transcriptase PCR was performed using SuperScript III (Invitrogen), according to the manufacturer's instructions for cDNA synthesis, and Phusion High-Fidelity PCR Kit (New England Biolabs) for PCR. Microarray analysis was performed using arrays provided by the Pathogen Functional Genomics Research Center funded by the National Institute of Allergy and Infectious Diseases using protocols SOP# M007 and M008 from The Institute of Genomic Research [25]. All data have been deposited in ArrayExpress with the accession number E-MTAB-4857. RT-qPCR was performed as described previously [11]. SigA was used as an internal standard, and ddCt method was used for the calculation of gene expression ratios. Error bars represent standard deviations from three biological replicates. For qPCR, cDNA was synthesized using 800 ng of RNA for each sample using the SuperScript III reverse transcriptase kit (Invitrogen). After cDNA synthesis, qPCR was carried out using Platinum SYBR Green qPCR SuperMix-UDG with ROX (Invitrogen). For the verification of the *cis*-encoded sRNA, chemostat-grown cells (nitrogen excess and nitrogen depletion) were used for qPCR [15]. For asRNA verification, cDNA synthesis was performed using strand-specific primers. Primers were designed with the Primer 3 software, and optimization was performed (for primers, see Table S1). The qPCR reactions were carried out on a 7500 Fast Real-Time PCR System (Applied Biosystems). As an internal control and for the normalization of results, we used the gene *sigA*.

Expression and purification of AmtR

The *amtR* gene (*msmeg_4300*) was synthesized (GeneArt) with the stop codon removed to allow for a C-terminal 6xHis tag. The gene was then inserted into the pYUB28b expression plasmid and cloned into the *M. smegmatis* 4517 expression strain [26]. Cells were harvested and centrifuged at 3900*g* for 20 min. Cell pellets were resuspended in lysis buffer [20 mM phosphate, 200 mM NaCl, and 20 mM imidazole (pH 7.4)], and cells

were sonicated on ice using a large probe for 30-s bursts at setting 7–8.5 with 1 min cooling in between using a Missonix Sonicator (USA) until cells were lysed (3–8 min). The cell suspension was centrifuged at 25,860*g* for 20 min at 4 °C. The supernatant was filtered through 1.2- and 0.45-µm filters and loaded onto a pre-equilibrated 5-ml nickel affinity column (His-trap HP™) for initial purification [27]. The eluted protein was combined 1:1 (volume) with high-salt phosphate-buffered saline (PBS) buffer [20 mM phosphate and 4 M NaCl (pH 7.4)] to facilitate DNA removal. The solution was concentrated to 700 µl and filtered through a small 0.2-µm filter. The protein was then further purified by S200 10/300 size-exclusion chromatography in buffer [20 mM phosphate and 2 M NaCl (pH 7.4)]. To reduce salt concentration, we dialyzed the protein in 1 l of buffer [20 mM phosphate and 200 mM NaCl (pH 7.4)] for a minimum of 12 h with regular buffer changes. SDS-PAGE gels were cast in a Hoefer mini gel casting system. SDS-PAGE gels consisted of a 5% acrylamide stacking gel overlaid on a 16.5% acrylamide resolving gel. All SDS-PAGE gels were made up with 30% acrylamide with an acrylamide:bisacrylamide ratio of 37.5:1 (BioRad Laboratories), including 0.1% filtered SDS, and were polymerized by the addition of 0.05% ammonium persulphate and tetramethylethylenediamine.

Protein samples were mixed in a 3:1 ratio with 4 x SDS loading buffer and were denatured for 5 min at 95 °C prior to loading onto the gel. Gels were run with 1 x SDS-PAGE running buffer at 70 V until the samples entered the resolving gel, then at 150 V until the dye front reached the end of the gel. For protein molecular weight estimation, 10 µl of Precision Plus Protein™ Unstained Standard (BioRad Laboratories) was run alongside each gel.

Crystallization of AmtR and collecting X-ray diffraction data

Initial protein crystallization trials employed sitting-drop vapor diffusion in 96-well plates and were performed at 18 °C. AmtR protein solution [100 nl, 10.5 mg ml^{-1}, 20 mM phosphate, and 200 mM NaCl (pH 7.4)] with 100 nl of precipitant solution (Hampton screens HR2–086, HR2–130, HR2–136, and HR2–144) was mixed and equilibrated beside a 100-µl reservoir of precipitant solution. Further crystallization screens employed hanging-drop vapor diffusion in 24-well plates at 18 °C. Crystals of AmtR grew after 4 days by mixing 1 µl of protein solution [10 mg ml^{-1}, 20 mM phosphate, and 200 mM NaCl (pH 7.4)] with 1 µL of precipitant solution [100 mM sodium-acetate (pH 5.0), and 1.6 M sodium-formate) and by equilibrating over a 500-µl reservoir of precipitant solution. The AmtR C-terminal domain was crystallized using in-drop limited trypsin digestion at a ratio of 1:100 with AmtR and hanging-drop vapor diffusion in 24-well plates at 18 °C. Crystals of AmtR C-terminal domain grew after 14 days by mixing 1 µl of protein solution [10 mg ml^{-1}, 20 mM phosphate, and 200 mM NaCl (pH 7.4)] with 1 µL of precipitant solution [100 mM sodium-acetate (pH 5.0), and 1.6 M sodium-formate] and by equilibrating over a 500-µl reservoir of precipitant solution. These crystals were subsequently soaked in mercury chloride (final concentration of 5 mM) in order to solve the structure using SAD.

AmtR protein crystals were flash-frozen in liquid nitrogen with up to 30% glycerol prior to data collection. For data collection, crystals were mounted in a stream of cold N_2 gas at 100 K. C-terminal (Hg soaked) and full-length AmtR data were collected at the macromolecular crystallography beamline (MX1) at the Australian Synchrotron, Melbourne, Australia using an ADSC Quantum 210r detector. For the C-terminal domain, data were collected using radiation at a wavelength of 12,290.0 eV, (6 eV above the L_{III} edge of Hg) to a resolution of 1.84 Å. For full-length AmtR, data were collected at a wavelength of 12,669.5 eV to a resolution of 1.98 Å. All data sets were indexed and integrated using iMosflm 7.0.9 [28], then scaled and truncated using Aimless [29] in the CCP4 program suite 6.4.0. (Collaborative Computational Project, Number 4, 1994). The resolution limit of 1.84 Å for the C-terminal domain data set gives a very high R_{merge} value of 0.201 (1.662). This is possibly the result of the very high redundancy. This resolution limit of 1.84 Å can be justified by the CC½ of 0.859 in the outer shell and the mean I/σ(I) 2.9 in outer shell.

AmtR structure solution and refinement

The C-terminal domain structure was solved using SAD. Initial Hg sites were identified and refined using SHELX [30], and the structure was solved using Auto-rickshaw [31]. The full-length AmtR structure was solved by molecular replacement using the C-terminal domain structure as a model. Molecular replacement was carried out using Phaser [32] in the CCP4 suite 6.4.0 (Collaborative Computational Project, Number 4, 1994). The N-terminal domain was then built using Buccaneer [33,34], followed by an initial restrained refinement in refmac5 [35,36] in the CCP4 suite 6.4.0. Subsequent cycles of manual model building using COOT 0.7 [37], followed by refinement in Refmac5 on the CCP4 suite 6.4.0 were carried out. Full data collection and refinement statistics are provided in

Table 2. AmtR data collection and refinement statistics

	LBD structure	Full Structure
Data collection		
Wavelength	1.0072 Å	0.95370 Å
Space group	I 2 3	C 2
Unit cell dimensions (Å)	111.1, 111.1, 111.1	181.8, 105.8, 39.57
Unit cell dimensions (°)	90, 90, 90	90, 94.5, 90
Resolution (Å)	32.07–1.84, (1.88–1.84)	37.25–1.98 (2.03–1.98)
R_{merge}	0.201 (1.662)	0.100 (0.694)
CC (1/2)	0.999 (0.859)	0.997 (0.769)
Completeness (%)	100 (100)	97.1 (90.5)
Redundancy	42.7 (39.7)	6.7 (5.7)
No. of observations	856,794 (48,790)	336,413
No. of unique reflections	20,068 (1230)	50,329
Mean I/σ(I)	69.1 (2.9)	2.1
Refinement		
R_{factor}	0.183	0.212
R_{free}	0.210	0.252
Protein atoms	1097	4549
Metal atoms	9 Hg	
Solvent atoms	147	126
Average *B* value ($Å^2$)	25.7	28.8
RMSD		
Bond angles (°)	2.670	1.851
Bond lengths (Å)	0.028	0.018

Table 2. Coordinate files and structure factors for the full structure have been deposited in the Protein Data Bank with PDB code (5E57).

EMSA and SPR

Purified AmtR that had been treated for co-purified DNA (see Materials and Methods) was used in EMSA with DIG-labeled DNA using the DIG Gel Shift kit, 2nd Generation (Roche, Germany). Assay protein:DNA ratios began at equal molar and subsequent higher ratios of protein to DNA. Assays were conducted as per manufacturer's instructions (Roche, Germany; for primers, see Table S1). To identify the specificity of AmtR for target DNA rather than DIG label, binding assays were set up with unlabeled target DNA added in 120× excess at the highest labeled DNA: protein ratio. This assay was labeled the specific control. To determine the specificity of the protein for target DNA, binding assays were set up with unlabeled non-specific DNA (the promoter region of the *M. smegmatis phoH2* gene that contains no putative AmtR recognition sequence) added in 103× excess (to account for size difference of the DNA fragments). This assay was labeled the non-specific control. Assays were prepared on ice and incubated at room temperature for 20 min and then run on 6% native acrylamide Tris-borate-EDTA gel. DNA was transferred from the gel to a nylon transfer membrane by Southern blotting and was then fixed to the membrane by UV crosslinking. The membrane was washed, blocked, and equilibrated in detection buffer [0.1 M Tris and 0.1 M NaCl (pH 9.5)] as per manufacturer's instructions (Roche, Germany) before doing the chemiluminescent immunoassay. A chemiluminesence reaction using the alkaline phosphatase-conjugated antibody bound to the DIG-labeled DNA and disodium 3-(4-methoxyspiro {1,2-dioxetane-3,2′-(5′-chloro)tricyclo [3.3.1.1^{3,7}]decan}-4-yl)phenyl phosphate was done as per the manufacturer's instructions and was incubated at room temperature for 5 min, followed by a 10-min incubation at 37 °C to enhance the reaction. The light emission was captured by exposing the membrane to X-ray film in a light-protective cassette for a minimum of 1 h. The film was then developed and fixed for visualization and was captured using white light.

The SPR experiments were performed using a BIACORE 3000 (GE Healtcare) equipped with a research-grade SA sensor chip. The ligand-specific, biotinylated double-stranded DNA was immobilized on the FC2 surface of the SA sensor chip. Non-specific, biotinylated double-stranded DNA was used as a reference and was immobilized on the FC1 surface of the SA chip. To collect the kinetic binding data, we injected the analyte (AmtR) in 50 mM Na_2PO_4, 200 mM NaCl, and 0.05% Tween (pH 7.4) over the two flow cells at concentrations of 6, 3, 1.5, 0.75, 0.375, 0.187, and 0.09 nM at a flow rate of 50 μl/min and at a temperature of 25 °C. Duplicate injections (in random order) of each sample and a buffer blank were flowed over the two surfaces. The AmtR responses were double-referenced by subtracting the data from the reference flow cell and buffer responses. Data were analyzed using GraphPad PRISM using a non-linear regression association and dissociation analysis; constraints included the nM concentration of AmtR and the start of the dissociation phase. The equilibrium dissociation binding constant (K_D) was calculated using the steady-state/equilibrium binding model including the standard error.

Identification of AmtR-binding motif and accession numbers

The MEME/ Motif Alignment & Search Tool‡ was used to detect a conserved AmtR consensus-binding motif in the upstream regions of genes that showed differential gene expression in *M. smegmatis* JR258 Δ*amtR*. This consensus-binding motif was then used to search for further putative AmtR-binding regions in the genome of *M. smegmatis* mc^{2}155 [38].

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 5E57. Microarray data have been deposited in ArrayExpress with the Accession Number E-MTAB-4857 [39].

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2016.09.009.

Acknowledgments

This work was supported by the Maurice Wilkins Centre for Molecular Biodiscovery and a Marsden Grant from the Royal Society of New Zealand. G.M.C. was supported by a James Cook Fellowship from the Royal Society of New Zealand. M.P. was supported by a University of Otago Doctoral Scholarship, the Webster Center for Infectious Diseases and the Otago School of Medical Sciences. C.V. was supported by a University of Waikato Doctoral Scholarship.

Author contributions: M.P., C.V., J.R., J.L.M., M.B., and A.S. designed and performed the experiments and analyzed and interpreted the data. H.L.A. helped with the data analysis. M.P., C.V., and J.R. wrote the manuscript. V.L.A. and G.M.C. supervised the project, performed data interpretation, and wrote the manuscript. All authors have approved the final manuscript for publication.

Competing interests: The authors declare that they have no competing interests.

Received 16 July 2016;
Received in revised form 2 September 2016;
Accepted 7 September 2016
Available online 15 September 2016

Keywords:
AmtR;
mycobacteria;
urea degradation;
crystal structure;
regulatory small antisense RNA

†M.P. and C.V. contributed equally to this work.
‡http://meme-suite.org/

Abbreviations used:
asRNA, antisense RNA; CFU, colony-forming unit; qPCR,

quantitative real-time PCR; UA, urea amidolyase; UC, urea carboxylase; SPR, surface plasmon resonance; AH, allophanate hydrolase; EMSA, electrophoretic mobility shift assay; sRNA, small RNA; SAD, single anomalous diffraction; HTH, helix-turn-helix; PBST, phosphate-buffered saline with 0.05% Tween-80; HdB, Hartmans-de Bont; DIG, Digoxigenin; LB, Luria-Bertani; PM, Phenotype Microarray; PBS, phosphate-buffered saline.

References

[1] D. Fink, N. Weissschuh, J. Reuther, W. Wohlleben, A. Engels, Two transcriptional regulators GlnR and GlnRII are involved in regulation of nitrogen metabolism in *Streptomyces coelicolor* A3(2), Mol. Microbiol. 46 (2002) 331–347.

[2] G. Beckers, J. Strösser, U. Hildebrandt, J. Kalinowski, M. Farwick, R. Krämer, et al., Regulation of AmtR-controlled gene expression in *Corynebacterium glutamicum*: mechanism and characterization of the AmtR regulon, Mol. Microbiol. 58 (2005) 580–595.

[3] V.A. Jenkins, G.R. Barton, B.D. Robertson, K.J. Williams, Genome wide analysis of the complete GlnR nitrogen-response regulon in *Mycobacterium smegmatis*, BMC Genomics 14 (2013) 301.

[4] Y. Tiffert, P. Supra, R. Wurm, W. Wohlleben, R. Wagner, J. Reuther, The *Streptomyces coelicolor* GlnR regulon: identification of new GlnR targets and evidence for a central role of GlnR in nitrogen metabolism in actinomycetes, Mol. Microbiol. 67 (2008) 861–880.

[5] M. Jakoby, L. Nolden, J. Meier-Wagner, R. Krämer, A. Burkovski, AmtR, a global repressor in the nitrogen regulation system of *Corynebacterium glutamicum*, Mol. Microbiol. 37 (2000) 964–977.

[6] S. Buchinger, J. Strösser, N. Rehm, E. Hänssler, S. Hans, B. Bathe, et al., A combination of metabolome and transcriptome analyses reveals new targets of the *Corynebacterium glutamicum* nitrogen regulator AmtR, J. Biotechnol. 140 (2009) 68–74.

[7] J. Amon, F. Titgemeyer, A. Burkovski, A genomic view on nitrogen metabolism and nitrogen control in mycobacteria, J. Mol. Microbiol. Biotechnol. 17 (2009) 20–29.

[8] C. Harper, D. Hayward, I. Wiid, P. van Helden, Regulation of nitrogen metabolism in *Mycobacterium tuberculosis*: a comparison with mechanisms in *Corynebacterium glutamicum* and *Streptomyces coelicolor*, IUBMB Life 60 (2008) 643–650.

[9] N. Jessberger, Y. Lu, J. Amon, F. Titgemeyer, S. Sonnewald, S. Reid, et al., Nitrogen starvation-induced transcriptome alterations and influence of transcription regulator mutants in *Mycobacterium smegmatis*, BMC Res. Notes 6 (2012) 482.

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

[11] M. Berney, G.M. Cook, Unique flexibility in energy metabolism allows mycobacteria to combat starvation and hypoxia, PLoS One 5 (2010), e8614.

[12] K.B. Arnvig, D.B. Young, Identification of small RNAs in *Mycobacterium tuberculosis*, Mol. Microbiol. 73 (2009) 397–408.

[13] K.B. Arnvig, I. Comas, N.R. Thomson, J. Houghton, H.I. Boshoff, N.J. Croucher, et al., Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of *Mycobacterium tuberculosis*, PLoS Pathog. 7 (2011), e1002342.

[14] K. Haning, S.H. Cho, L.M. Contreras, Small RNAs in mycobacteria: an unfolding story, Front. Cell Infect. Microbiol. 4 (2014) 96.

[15] M. Petridis, A. Benjak, G.M. Cook, Defining the nitrogen regulated transcriptome of *Mycobacterium smegmatis* using continuous culture, BMC Genomics 16 (2015) 821.

[16] Z. Yu, S.E. Reichheld, A. Savchenko, J. Parkinson, A.R. Davidson, A comprehensive analysis of structural and sequence conservation in the TetR family transcriptional regulators, J. Mol. Biol. 400 (2009) 847–864.

[17] C. Palanca, V. Rubio, Structure of AmtR, the global nitrogen regulator of *Corynebacterium glutamicum*, in free and DNA-bound forms, FEBS J. 283 (2016) 1039–1059.

[18] S. Hartmans, J.A.M. Bont, E. Stackebrandt, The Genus Mycobacterium - Nonmedical, in: M. Dworkin, S. Falkow, E. Rosenberg, K.H. Schleifer, E. Stackebrandt (Eds.), The Prokaryotes Third Edition: Archaea. Bacteria: Firmicutes, Actinomycetes, Springer, New York 2006, pp. 889–918.

[19] P.B. Garland, P.J. Randle, A rapid enzymatic assay for glycerol, Nature 196 (1962) 987–988.

[20] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, et al., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol. 7 (2011) 539.

[21] M. Berney, M.R. Weimar, A. Heikal, G.M. Cook, Regulation of proline metabolism in mycobacteria and its role in carbon metabolism under hypoxia, Mol. Microbiol. 84 (2012) 664–681.

[22] S. Gebhard, S.L. Tran, G.M. Cook, The Phn system of *Mycobacterium smegmatis*: a second high-affinity ABC-transporter for phosphate, Microbiology 152 (2006) 3453–3465.

[23] V. Pelicic, M. Jackson, J.M. Reyrat, W.R. Jacobs, B. Gicquel, C. Guilhot, Efficient allelic exchange and transposon mutagenesis in *Mycobacterium tuberculosis*, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 10,955–10,960.

[24] M.C.J. Blokpoel, H.N. Murphy, R. O'Toole, S. Wiles, E.S.C. Runn, G.R. Stewart, et al., Tetracycline-inducible gene regulation in mycobacteria, Nucleic Acids Res. 33 (2005), e22.

[25] P. Hegde, R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, et al., A concise guide to cDNA microarray analysis, BioTechniques 29 (2000) 548–562.

[26] G. Bashiri, A.M. Rehan, D.R. Greenwood, J.M.J. Dickson, E.N. Baker, Metabolic engineering of cofactor F420 production in *Mycobacterium smegmatis*, PLoS One 5 (2010), e15803.

[27] J. Robson, J.L. McKenzie, R. Cursons, G.M. Cook, V.L. Arcus, The *vapBC* operon from *Mycobacterium smegmatis* is an autoregulated toxin-antitoxin module that controls growth via inhibition of translation, J. Mol. Biol. 390 (2009) 353–367.

[28] A.G.W. Leslie, Powell, Processing diffraction data with mosflm, in: R.J. Read, J.L. Sussman (Eds.), Evolving Methods for Macromolecular Crystallography, 1st ed., 245, Springer, Netherlands 2007, pp. 41–51 chapter 4.

[29] P.R. Evans, G.N. Murshudov, How good are my data and what is the resolution? Acta Crystallogr. D 69 (2013) 1204–1214.

[30] G.M. Sheldrick, A short history of SHELX, Acta Crystallogr. A 64 (2007) 112–122.

[31] S. Panjikar, V. Parthasarathy, V.S. Lamzin, M.S. Weiss, P.A. Tucker, Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation

of an X-ray diffraction experiment, Acta Crystallogr. D 61 (2005) 449–457.
[32] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, J. Appl. Crystallogr. 40 (2007) 658–674.
[33] K. Cowtan, The Buccaneer software for automated model building. 1. Tracing protein chains, Acta Crystallogr. D 62 (2006) 1002–1011.
[34] K. Cowtan, Fitting molecular fragments into electron density, Acta Crystallogr. D 64 (2008) 83–89.
[35] G.N. Murshudov, A.A. Vagin, E.J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method, Acta Crystallogr. D 53 (1997) 240–255.
[36] G.N. Murshudov, P. Skubák, A.A. Lebedev, N.S. Pannu, R.A. Steiner, R.A. Nicholls, et al., REFMAC5 for the refinement of macromolecular crystal structures, Acta Crystallogr. D 67 (2011) 355–367.
[37] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr. D 66 (2010) 486–501.
[38] T.L. Bailey, M. Gribskov, Combining evidence using *p*-values: application to sequence homology searches, Bioinformatics 14 (1998) 48–54.
[39] M. Petridis, G.M. Cook, The AmtR regulon on *Mycobacterium smegmatis*—genome-wide expression profile of an *M. smegmatis amtR* mutant (JR258) compared to *M. smegmatis* mc^2155, Data in Brief (2016) (submitted).