JBC Papers in Press. Published on December 27, 2019 as Manuscript RA119.011373
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.RA119.011373

A universal stress protein in *Mycobacterium smegmatis* sequesters the cAMP-regulated lysine acyltransferase and is essential for biofilm formation

Sintu Samanta[1,2], Priyanka Biswas[1], Arka Banerjee, Avipsa Bose, Nida Siddiqui, Subhalaxmi Nambi, Deepak Kumar Saini and Sandhya S. Visweswariah*

Department of Molecular Reproduction, Development and Genetics, Indian Institute of Science, Bengaluru, India

[1] Contributed equally to this study
[2] Current address Department of Applied Sciences, IIIT Allahabad, India
* To whom correspondence should be addressed
E mail: sandhya@iisc.ac.in; Tel: +91 80 22932542

Running Title: Universal stress protein and lysine acylation

Keywords: universal stress protein; *Mycobacterium smegmatis*; biofilm; protein lysine acylation; cyclic AMP; KATms; MSMEG_4207; tricarboxylate-regulated operon

Abstract

Universal stress proteins (USPs) are present in many bacteria and their expression is enhanced under various environmental stresses. We have previously identified a USP in *Mycobacterium smegmatis* that is a product of the *msmeg_4207* gene and is a substrate for a cAMP-regulated protein lysine acyltransferase, KATms (MSMEG_5458). Here, we explored the role of this USP (USP^{4207}) in *M. smegmatis*, finding that its gene is present in an operon that also contains genes predicted to encode a putative tripartite tricarboxylate transporter (TTT). Transcription of the *TTT-usp^{4207}* operon was induced in the presence of citrate and tartrate, perhaps by the activity of a divergent histidine kinase-response regulator gene pair. A *usp^{4207}*-deleted strain had rough colony morphology and reduced biofilm formation compared with the WT strain; however, both normal colony morphology and biofilm formation were restored in a Δ*usp^{4207}*Δ*katms* strain. We identified several proteins whose acetylation was lost in the Δ*katms* strain, and whose transcript levels increased in *M. smegmatis* biofilms along with that of USP^{4207}, suggesting that USP^{4207} insulates KATms from its other substrates in the cell. We propose that USP^{4207} sequesters KATms from diverse substrates whose activities are down-regulated by acylation but which are required for biofilm formation, thus providing a defined role for this USP in mycobacterial physiology and stress responses.

Universal stress proteins (USPs) are found in bacteria, archaea, fungi and plants (1). Their name derives from the fact that many of these proteins are induced under various stresses such as oxidative stress, salt stress, nutrient starvation, heat shock and insult by DNA damaging agents (2,3). USPs are classified into two groups based on their ability to bind ATP (UspFG-like), while UspA-like proteins do not bind ATP (4). While a number of these USP proteins have been characterized biochemically and structurally (4,5), the roles of the majority of these USPs remain unknown, with many

usp genes being non-essential for regular bacterial growth (6).

Mycobacteria harbor genes for a number of USP-like proteins and many of them are conserved across both fast, as well as slow-growing pathogenic mycobacteria, such as *M. tuberculosis (5-7)*. In a manner similar to that reported in other bacteria, USP-like genes are induced under different stresses, and Rv2623 has been shown to regulate growth of *M. tuberculosis* and establish latent infection (7). We have shown that Rv1636 and its ortholog MSMEG_3811 bind cAMP and ATP and could serve as a sink for cAMP in these organisms (5).

Protein acylation on lysine residues is an important post-translational modification that occurs in bacteria (8-10), and a number of proteins have been shown to be modified in proteome-wide analyses in many bacteria (11,12). The proteins that are acylated include those involved in central carbon metabolism, transcription, fatty acid metabolism and cell wall functions (10,13-17). A number of bacteria harbor multiple enzymes of the GNAT family that serve to acylate proteins (15), suggesting an inbuilt redundancy. However, individual GNATs can also acylate only specific proteins (15). Bacteria also express promiscuous sirtuin-like deacetylases (9,18-20) that allow dynamic changes in protein acylation (21), depending on the physiological requirements of the cell.

Protein lysine acetylation in mycobacteria was first identified by the biochemical characterization of a cAMP-dependent protein lysine acyltransferase (KAT; to distinguish it from N-terminal acetylation) (8). The fusion of a cAMP-binding domain to a GNAT-like acyltransferase domain in a single protein is found only in mycobacterial species (22) and could have evolved due to the abundance of genes involved in cAMP generation (adenylyl cyclases) in both pathogenic and non-pathogenic mycobacteria (23,24). We identified the MSMEG_4207 USP (USP^{4207}) as a substrate for the KAT found in *M. smegmatis* (KATms) (8), thus providing a novel link between protein acetylation and USPs. A number of additional protein substrates were subsequently identified for KATms and KATmt, and include those involved in metabolic pathways, fatty acid degradation or transcription (13,14,17,19).

The identification of USP^{4207} as a substrate for KATms utilized pull-down approaches, using GST-KATms as bait and crude cytosol from *M. smegmatis (8)*. The presence of significant amounts of USP^{4207} interacting with GST-KATms indicates its relatively high expression in *M. smegmatis*. Nevertheless, the tight association of USP^{4207} with KATms is surprising, since enzyme-substrate complexes are rarely identified by pull-down approaches. We argued at that time that tight binding of USP^{4207} may preclude the binding of other substrates to KATms, and thus USP^{4207} may act as a regulator of protein acylation brought about by KATms (8).

In the current study, we explore the role of USP^{4207} as a modulator of KATms-mediated acylation. We show that usp^{4207} is part of a tricarboxylic acid-regulated operon, and a strain deleted for usp^{4207} is defective in biofilm formation. By identifying substrates for KATms, and observing that deletion of *katms* in the Δusp^{4207} strain restores biofilm formation, we propose that USP^{4207} is the first proteinaceous inhibitor identified for a GNAT-like enzyme, and has evolved

specifically to regulate KATms activity in the cell.

Results

usp^{4207} and its gene neighborhood. The *usp^{4207}* gene is predicted to be in an operon (Fig. 1A) based on overlapping stop and start codons in the genome sequence. Using primers across the junctions of genes in the predicted operon, we obtained products across all junctions, suggesting that *msmeg_4210* to *msmeg_4207* could be made as a single RNA (Fig. 1B). Analysis of the sequences of these genes in NCBI predicted their protein products to be members of a tripartite tricarboxylate transporter (TTT) family (25). MSMEG_4210 may be a secreted protein that could bind the tricarboxylic acid in the periplasmic space of the cell. Indeed, bioinformatic analysis of its amino acid sequence (http://www.cbs.dtu.dk/services/SignalP/) predicted a signal peptide cleavage site after residue 24. Two integral membrane proteins are predicted to be the products of the *msmeg_4209* and *msmeg_4208* genes. MSMEG_4208 is predicted to have 12 transmembrane spanning domains, and along with MSMEG_4209 (with 4 predicted transmembrane domains) could serve as a channel to transport tricarboxylic acids into the cell. TTT operons have been characterized from other bacteria, where a similar three gene arrangement exists (26-28). Interestingly, adjacent to this operon and divergently orientated, is another operon comprising of a putative histidine kinase (*msmeg_4211*) and a response regulator (*msmeg_4212*) (Fig. 1A) that may function as a two-component system that could regulate the *msmeg_4207* operon, with the histidine kinase serving as a sensor for citrate (29,30). Occurrence of such a two-component system regulating the expression of TTT operons is also common (25,31).

Orthologs of USP are found in a number of fast-growing mycobacteria, and the lysine residue that is acetylated in USP^{4207} (8) is conserved (Fig. 1C).

The USP^{4207} operon is induced by tricarboxylic acids. None of the genes in the USP^{4207} operon or the two-component regulatory system have been characterized earlier in any mycobacterial species, and therefore we asked if tricarboxylic acids, such as citrate, could regulate the expression of the USP operon. Citrate was indeed able to increase by 10-fold, the expression of the *msmeg_4207* operon (Fig. 2A). Moreover, USP^{4207} protein levels were also increased in the presence of citrate (Fig. 2B). Tartrate could also increase levels of USP^{4207} in cells, while another dicarboxylic acid, succinate failed to do so (Fig. 2B). Therefore, it appears that C4 dicarboxylates with a hydroxyl group at the C2 position serve as activators of the operon, and the presence of a third carboxyl group is not essential for operon induction. Thus, proteins encoded in this operon may serve to transport both tri-and specific dicarboxylic acids into the cell. A similar operon in *Corynebacterium* is induced in the presence of citrate and its protein products serve in transporting citrate into the cell (31). However, the ability to transport tartrate was not tested.

Messenger RNA levels of *katms* (*msmeg_5458*) or two sirtuin-like genes (*msmeg_4620* and *msmeg_5175*) were unaltered in the presence of citrate (Fig. 2A), indicating that the availability of large amounts of USP^{4207} could result in an increase in absolute levels of *acetylated*

USP in the cell, when intracellular levels of citrate are high.

We cloned the intergenic region between *msmeg_4210* and *msmeg_4211* (~ 103 bp) upstream of luciferase in two orientations and monitored luciferase activity in the presence and absence of citrate. As shown in Fig. 2C, luciferase activity was increased in the presence of citrate, in an orientation that would lie 5' to the *usp*^{4207} operon. In a direction that would lie 5' to the two-component system, luciferase activity remained constant in the presence and absence of citrate (Fig. 2C), suggesting constitutive expression of *msmeg_4211* and *msmeg_4212*. This was in agreement with data shown in Fig. 2A, where RNA levels of these two genes were unchanged in the presence of citrate. Therefore, this intergenic region serves as a promoter for both the TTT-*usp*^{4207} operon and the two-component system.

Based on these findings, and as observed in *C. glutamicum*, (26), it is possible that citrate may bind to the periplasmic domain of the adjacent histidine kinase, MSMEG_4211. This would activate the kinase and allow phosphorylation of the response regulator MSMEG_4212, since most cognate histidine kinase-response regulator pairs are found in operons (Fig. 1B)(32). MSMEG_4212 may therefore bind to a region upstream of *msmeg_4210*. In order to test this possibility, we cloned, expressed and purified the putative response regulator (Fig. 3A) and performed EMSAs with a radiolabelled fragment representing the promoter. As shown in Fig. 3B, increasing concentrations of the purified MSMEG_4212 showed a series of shifts in the EMSA, perhaps indicating different oligomers of the protein binding to the intergenic region, or the presence of multiple DNA-binding sites.

It is important to note that we used MSMEG_4212 directly as purified from *E. coli*. This may either indicate that the non-phosphorylated form of the response regulator can bind to the region upstream of *msmeg_4210*, or that the response regulator is phosphorylated as purified from *E. coli* (33). In order to test this, we aligned the sequence of MSMEG_4212 with OmpR from *E. coli*, where the aspartate residue that is phosphorylated has been identified (34). We mutated residue D54 to an alanine and purified the mutant protein. Interestingly, even low concentrations of the D54A mutant protein bound to the probe and showed only a single shift (Fig. 3B). This indicated that the non-phosphorylated form of MSMEG_4212 bound to the intergenic sequence with a higher affinity than the wild type protein and formed a complex comprising a single oligomeric species.

To confirm that the shift seen in the EMSA was specific, and also to identify a smaller region that could serve as the binding site for MSMEG_4212, we amplified two fragments, shown as dotted and filled lines in Fig. 3C. We used these purified PCR products, as well as the full length intergenic sequence, to determine if they could compete for binding of the radiolabelled intergenic region to MSMEG_4212 and the D54A mutant protein. As shown in Fig. 3D, the entire intergenic region could inhibit binding of the radiolabelled probe to both complexes seen with the wild type protein, and also the single complex seen with the mutant protein. The PCR product closer to *msmeg_4211* could completely inhibit binding of MSMEG_4212 and the D54A mutant to the intergenic fragment. The

product closer to *msmeg_4201* showed a slight degree of competition, as seen by the enhanced intensity of the faster moving complex with the wild type protein. In the case of the D54A mutant protein, there was a marginal decrease in intensity of shifted band. We therefore suggest that MSMEG_4212 could act as a regulator of the *msmeg_4210-msmeg_4207* operon by binding to residues in the intergenic region proximal to *msmeg_4211*. It is unclear at present whether there are two binding sites in the intergenic region, since the PCR products contained ~ 20 bp overlapping region. However, affinity binding of mutant MSMEG_4212 suggests that in the absence of citrate, unphosphorylated MSMEG_4212 could act as a repressor of the *usp*^{4207} operon.

Deletion of msmeg_4207 alters colony morphology. We have earlier shown that USP^{4207} is acetylated on a single lysine residue by KATms in *M. smegmatis* (8). What is the role of USP^{4207} and its acetylation in the cell? In order to study this, we deleted the *usp*^{4207} gene by homologous recombination (35) and confirmed deletions using PCR (Fig. 4A). Therefore, in order to test the role of acetylation of USP^{4207}, we also generated a *usp*^{4207} deletion strain in the background of a Δ*katms* strain. We complemented the deletion strains with either wild type *usp*^{4207} or with a mutant *usp*^{4207}, where K104 (the site for acetylation) was mutated to R. Wild type and mutant strains grew similarly in shaking cultures (data not shown). However, as shown in Figure 4B, colonies formed by the *Δusp*^{4207} strain were smaller and more compact, with a rougher surface than those formed by wild type cells. This altered morphology was less evident when Tween 80 was included in the agar plates, suggesting that alterations in the cell wall glycopeptidolipids (GPLs) and cell-to-cell interactions may determine the morphology of the colony (36). It is possible that this was a consequence of aberrant acylation and inhibition of proteins/enzymes which could include those involved in GPL synthesis (37,38). Complementation with wild type USP as well as the K104R mutant protein restored normal colony morphology, indicating that the *presence* of USP protein and not the absence of *acetylated* USP, was responsible for normal colony morphology (Fig. 4B). While the Δ*katms* strain showed no change in colony morphology, and < 10% increase in USP levels (Fig. 4C) when normalized to CRP across duplicate experiments, deletion of *katms* in the *Δusp*^{4207} strain restored normal colony morphology and smooth appearance (Fig. 4B). This indicated that in the absence of USP^{4207}, the presence of KATms, and presumably its catalytic activity, altered colony appearance.

USP^{4207} *forms a stable complex with KATms and is the preferred substrate*. USP^{4207}is an abundantly expressed protein in *M. smegmatis*, as evidenced in our earlier experiments using GST-KATms as bait (8). We directly tested the stability of complex formation between KATms and USP^{4207} in pull-down experiments. Indeed, both wild type and USP^{4207}K104R could interact with GST-KATms bound to glutathione beads, indicating that the mutation in USP^{4207} did not alter binding of USP^{4207}K104R to KATms (Fig. 5A). Therefore, in the cell, we hypothesized that USP^{4207}K104R could serve to sequester KATms and regulate the extent of acetylation of other substrates, thereby providing a mechanistic

explanation for the rescue by USP^{4207}K104R of the small-colony phenotype that was seen on deletion of *usp^{4207}* (Fig. 4B).

To confirm that such substrate competition could occur, we monitored the acetylation of Acs, a known substrate of KATms (20), in the presence and absence of varying concentrations of USP, *in vitro*. As shown in Fig. 5B, acetylation of acetyl CoA synthase (Acs), was steadily reduced in the presence of increasing concentrations of USP^{4207}. Given the high expression levels of USP^{4207} in the cell, we propose that *M. smegmatis* can modulate acylation levels, and thereby the activity, of different substrates of KATms, depending on the levels of USP in the cell.

Deletion of usp^{4207} inhibits biofilm formation. We then asked what the potential substrates for KATms could be that may remain poorly acylated in the presence of USP^{4207}. We prepared lysates from planktonic cultures of isogenic mutant strains of *M. smegmatis* and subjected them to Western blot analysis with antibodies to acetyl-lysine. As seen in Fig. 6A, a band of Mr ~ 80 kDa was absent in the *Δkatms* strain. Acetylation of this protein (or proteins) was similar in strains that harbored a wild type copy of *katms*, and in the Δ*usp^{4207}*strain, equivalent to that seen in lysates prepared from wild type *M. smegmatis*. Interestingly, acetylation of the ~ 80 kDa band was markedly reduced (~ 3-fold) in the Δ*usp^{4207}* strain complemented with the *usp^{4207}K104R* mutant. When we checked expression levels of USP protein in different strains, it was evident that levels of USP^{4207}K104R were much higher (~ 5-fold) in the complemented strain (Fig. 6A), perhaps due to enhanced stability of the non-acetylated form of USP. Therefore, the reduced acetylation seen in the protein band at ~ 80 kDa supports the hypothesis that USP^{4207} could indeed modulate the extent of acetylation of substrates of KATms in the cell, while its own acetylation state is dispensable. Enhanced acetylation in the Δ*usp^{4207}* strain was not observed, perhaps due to the limiting amounts of KATms substrates in the cell, which remain fully acetylated even in wild type cells.

The prominent acetylated bands we observed here, which were absent in a *Δkatms* strain, were identified in an earlier report utilizing similar biochemical approaches (19), and included propionyl CoA synthetase (PrpE) and acetyl CoA synthetase (Acs). PrpE and Acs were also shown to be acetylated in a global acylome analysis of *M. smegmatis* (39).

We cut out a 0.5 cm piece of the gel corresponding to the region that contained proteins that were acetylated in planktonic cultures and subjected the gel piece to tryptic digestion and mass spectrometry. We specifically looked for additional peptides whose acetylation status changed in the *Δkatms* strain. We were able to identify the same peptides in Acs and PrPE as reported earlier, but also identified a peptide from MSMEG_0401 (product of *mps*), that is part of the cluster of genes involved in mycobacterial GPL synthesis (40-42). Deletion of genes in this cluster results in a rough colony phenotype (40,41,43), similar to that seen in the *Δusp^{4207}* strain.

In a proteomic analysis of the biofilm pellicle in *M. smegmatis*, propionyl CoA synthetase levels were reported to be increased (44). Moreover, alteration in GPL biosynthesis has been shown to affect biofilm formation (38,43). It is therefore possible that synthesis of GPLs are

modulated by acylation during biofilm formation, and consequently are dependent on the presence of USP and KATms in the biofilm. Indeed, biofilm formation was dramatically lost in the *Δusp*^{4207} strain and cells became flocculent and disperse (Fig. 6B). Biofilm formation was restored and marginally enhanced in the *usp*^{4207} complemented strains. Interestingly, deletion of *katms* in the background of *Δusp*^{4207} restored biofilm formation, while deletion of *katms* alone had no effect in biofilm formation. Moreover, complementation with the K104R mutant of *usp4*^{207} was able to restore the biofilm formation, emphasizing again that association of USP with KATms, and not acylation of USP, is sufficient for preventing acylation by KATms of its other substrates.

To confirm that deletion of *usp4*^{207} altered the levels of GPLs in the biofilm, we prepared lipids from strains that formed a biofilm and from flocculent cells seen in the *Δusp*^{4207}strain (Fig. 6C). GPLs were dramatically reduced in the *usp*^{4207} strain, but restored on deletion of *katms*, concomitant with biofilm formation. Therefore, it appears that GPLs play a critical role in biofilm formation and their production is regulated by the presence of KATms and levels of USP.

We monitored transcript levels of *prpE, katms, mps,* the sirtuin genes and *usp*^{4207} in the pellicle and planktonic cultures. Levels of transcripts of *usp*^{4207}*, prpE, mps* and *msmeg_4620* were found to be higher (~ 3 to 50-fold) in the pellicle than in planktonic cells. A modest increase in *katms* or sirtuin gene expression (~ 1.5-fold) was observed (Fig. 7A). Comparing the qPCR efficiency of all primer pairs for individual genes, and generating a standard curve using known amounts of genomic DNA (data not shown), we could estimate the amount of gene-specific cDNA in samples taken for qPCR (Fig. 7B). Based on this analysis, mRNA levels of *usp*^{4207} were higher than other transcripts, which could reflect a higher concentration of protein, assuming half-lives of protein and mRNA are similar across the genes tested. We could estimate levels of USP^{4207} by performing quantitative Western blot analysis using anti MSMEG_4207 antibodies and defined concentrations of USP. Indeed, USP^{4207} protein levels in the pellicle reached ~ 10 µg/ mg protein, while levels were ~ 1 µg/ mg protein in planktonic cultures (Fig. 7C).

Our results therefore strongly suggest that proteins/enzymes that may be required for biofilm formation (such as PrpE and MSMEG_0401) should remain non-acylated in order to remain active. The high concentrations of USP^{4207} could serve to sequester KATms from these and other substrates. In the absence of USP in a KATms-expressing strain (*Δusp*^{4207}), increased acylation of these proteins required for biofilm formation would occur, thus inhibiting their activity and affecting biofilm formation.

To provide support for this, we performed Western blot analysis of lysates prepared from the pellicle of wild type, *Δusp*^{4207}, *Δkatms* and complemented strains. As shown in Fig. 7D, the major acetylated band at ~ 80 kDa band was absent in pellicle fractions of all strains that formed the biofilm. However, fractions prepared from flocculent cells formed by the *Δusp*^{4207} strain showed a prominent band at ~ 80 kDa, corresponding to the band seen in planktonic cultures (Fig. 6A). Therefore, reduced acetylation of proteins

such as PrpE and Acs is critical for biofilm formation. Interestingly, transcript levels of the sirtuin genes increased in the biofilm (Fig. 7A), suggesting that acylation of proteins in general may be detrimental to biofilm formation.

The blot also revealed changes in KATms-*independent* protein acetylation in different strains. A band of 140 kDa is absent in the *Δusp*^{4207} strain but present in the *Δkatms* strain. 55 kDa, 65 kDa and 110 kDa proteins remain acetylated. Therefore, protein acetylation seems to be exquisitely regulated during biofilm formation

Attempts to normalize protein loading using CRP antiserum revealed that intriguingly, CRP levels were markedly reduced in the *Δusp*^{4207} strain (Fig. 7D). Nevertheless, the increased intensity of the band at ~ 80 kDa and equivalent intensity of other bands allowed us to conclude that KATms-mediated acetylation is indeed regulated by high levels of USP^{4207} during biofilm formation.

Cyclic AMP levels increase during biofilm formation. Our hypothesis related to the role of USP^{4207} in regulating KATms activity arises from genetic analysis which showed that the presence of KATms in a *Δusp*^{4207} strain was inhibitory to biofilm formation. Since KATms activity can be enhanced in the presence of cAMP, we asked if levels of cAMP were altered in the pellicle. We found that in fact, cAMP levels were increased during biofilm formation (Fig. 7E). RNA levels of *usp*^{4207} were increased in the pellicle (Fig. 7A), suggesting transcriptional upregulation of the operon, perhaps by elevation of citrate and other tricarboxylates during biofilm formation.

In summary, cells have devised a means to regulate KATms activity in the presence of cAMP, by increasing USP^{4207} levels, rather than by altering KATms levels. Cyclic AMP may be required during biofilm formation to modulate the activities of other cNMP-binding proteins, including CRP. The presence of USP^{4207}, serving as a substrate and 'inhibitor' of KATms, could bring in a high degree of specificity in terms of regulating the activity of a specific cNMP-binding protein (i.e. KATms) during biofilm formation, following the elevation of cAMP.

Discussion

In this study, we propose a novel mechanism of regulation of protein acylation, that is, by cells providing a proteinaceous competitive inhibitor of an acyltransferase. The inhibitor in this case is a universal stress protein, which also serves as a substrate for a cAMP-regulated acyltransferase. As shown in Fig. 8A, *usp*^{4207} is part of an operon containing genes that probably encode a TTT, since expression of genes in that operon are upregulated in the presence of citrate. We also propose that the divergent two-component, histidine kinase-response regulator gene pair encode a signaling system to transcriptionally activate the *TTT-usp*^{4207} operon in the presence of tricarboxylic acids. The putative response regulator can bind in a non-phosphorylated state, to a region present between the *TTT-usp*^{4207} operon and the genes of the two-component signaling system (Fig. 3B), suggesting that it could serve as a repressor, or allow only weak transcription, under conditions when citrate levels are low. The multiple concentration-dependent shifts seen with the wild type protein (Fig. 3B) could suggest that phosphorylation of the wild type protein may stabilize oligomers that bind to the DNA, without altering DNA

binding affinity or specificity (45) and/or represent binding of the wild type protein to two sites in the intergenic region. The single shift seen with the D54A mutant protein suggests that this protein forms a stable oligomer following mutation, or adopts a DNA-binding conformation more readily (46). In a manner similar to that proposed for CpxRA from *Legionella pneumophila* (47) MSEG_4212 may adopt different oligomeric states which bind to different sequences contained in the intergenic region to either activate or repress transcription of the *usp*^{4207}operon. Mutational analysis of this intergenic region could determine the recognition motif for this newly identified response regulator.

Proteins containing a cAMP-binding domain fused to a GNAT-like acyltransferase domain are found only in mycobacteria (8). We explored the distribution of USP^{4207}-like proteins and the TTT operon in various genera of Actinobacteria, and found that in mycobacteria, the presence of the TTT operon in fast-growing species is always seen along with genes encoding USP^{4207}-like and KATms-like proteins (Fig. 8B). Therefore, we predict that in these mycobacterial species, the level of acylation of substrates of KATms-like proteins is modulated by USP^{4207} orthologs, as we report here in *M. smegmatis*. Slow-growing mycobacteria do not contain either a USP^{4207}-like protein, or a TTT operon. The acyltransferase activity of KATms is substantial even in the absence of cAMP (8). However, the KAT protein from *M. tuberculosis* (KATmt) obligatorily requires cAMP for its activation and the structural basis for these differences has been elucidated (8,13,22,48,49). We therefore speculate that in slow-growing mycobacteria, where the KAT protein has been exquisitely tuned to respond to levels of cAMP, there is no requirement for additional levels of regulation, such as the presence of proteinaceous competitive inhibitors. Indeed, in preliminary pull-down experiments using GST-KATmt as a bait and lysates prepared from *M. tuberculosis*, no major protein could be identified that interacts with KATmt (data not shown), in contrast to our results with KATms (8). The levels of KAT, cAMP and protein acylation during biofilm formation in slow-growing mycobacteria, such as in *M. tuberculosis* which lack the USP^{4207} ortholog, remain to be explored.

Is there additional information on possible links between protein acylation, colony morphology and biofilm formation, that would validate our hypothesis on the interplay of USP^{4207} and KATms? In a study performed to characterize the acetylome of *M. tuberculosis,* a sirtuin-like gene (*rv1151c*) was deleted which should result in hyperacetylation of proteins (11). It was observed that the Δ*rv1151c* strain showed a rough colony morphology and did not form biofilm. This data agrees with our findings in *M. smegmatis* where the deletion of *usp*^{4207}, which in our model would result in high activity of KATms and therefore increased acylation of its substrates (as should be observed in a sirtuin-deletion strain), also altered colony morphology and biofilm formation (Figs. 4B and 6B).

The promiscuous nature of sirtuins poses a difficulty in deacetylating a *specific* subset of acetylated proteins (13,18,50,51). In *M. smegmatis*, there exists 25 putative acyltransferases and 3 lysine deacetylases (15,19). In *M. smegmatis*, we show that a specific USP^{4207} could regulate acylation of substrates by directly interacting with its

cognate acyltransferase, KATms, thereby overcoming the lack of specificity of sirtuins to regulate the KATms-specific acylome.

The *M. smegmatis* gene cluster *msmeg_0400-msmeg_0402* encodes a non-ribosomal peptide synthase which is involved in synthesizing the tetrapeptide in GPLs that are found in non-tuberculosis causing mycobacteria as well as certain pathogenic mycobacteria (37,38,40,43,52). MSMEG_0401 is a member of the FadD enzyme family, some of which we have shown earlier are acetylated by KATmt (13). MSMEG_0401 harbors a lysine residue (residue 432) in the first condensation domain (https://pfam.xfam.org/) that catalyzes the formation of a peptide bond in non-ribosomal peptide synthesis, that could be the site for acetylation by KAT.

In *M. abscessus*, rough, cording and non-biofilm phenotypes are associated with low GPL levels (53). Smooth, biofilm-competent strains show increased GPL biosynthesis and higher expression of *mps1*. A KAT-like enzyme is present in this species (Fig. 8B) but resembles the KATmt protein which does not contain the PP-loop that contributes to the reduced requirement for cAMP binding to activate the acyltransferase domain of KATms (48). This may account for the absence of the USP^{4207}-like protein encoding gene in *M. abscessus* (Fig. 8B). In summary we propose that in non-tuberculosis causing mycobacteria, GPL biosynthesis is regulated by protein lysine acylation, and in turn, through the interplay between KATms, USP^{4207} and intracellular cAMP levels.

No TTT-like transporter has been characterized in mycobacteria, and citrate is usually added to media along with iron in culturing mycobacteria where it aids in the uptake of iron by the bacteria (54,55) . Our findings here have identified a putative transporter, based on the fact that genes in the operon are up-regulated by citrate, as is seen in other organisms where such TTT transporters have been characterized (25). However, in other bacteria, such as *Salmonella* Typhimurium, the tricarboxylate-binding protein can interact with a variety of citrate analogs and organic acids (56). Therefore, in *M. smegmatis*, it is possible that the TTT operon is regulated by a number of substances that may be found in environments where *M. smegmatis* thrives. The presence of *usp^{4207}* in the TTT operon, and therefore, linking protein acylation to cellular metabolism, is worth investigating in future.

USPs are found in many bacteria and some have been shown to be important in regulating biofilm formation. For example, deletion of the *uspA* gene in *Porphyromonas gingivitis* resulted in poor biofilm formation, and levels of UspA were increased during biofilm formation (57). A *uspA* gene in *Staphylococcus epidermidis* is upregulated in planktonic aggregates that show biofilm-like behaviour (58). Finally, deletion of a *usp* gene in *Micrococcus luteus* regulated the expression of a number of genes including those involved in central carbon metabolism (59). Therefore, since the expression of various USPs is increased in a number of stress conditions, they may play a direct role in regulating genes involved in overcoming stress such as during infection (2) , perhaps independently of a role in regulating protein acylation.

In conclusion, we have provided new information on the role of protein acylation in *M. smegmatis*, a link to colony morphology and biofilm formation and the involvement of a universal stress protein in

acting as a rheostat to tune levels of protein acylation in the cell. This complex interplay of small molecules like cAMP and citrate, unique enzymes such as KATms that can regulate the activities of enzymes as diverse as those involved in non-ribosomal peptide synthesis and metabolism, and a widely expressed small protein like USP^{4207} is depicted in Fig. 9. What remains unknown is how cAMP regulates biofilm formation and if citrate levels are increased during biofilm formation. Importantly, it appears that *usp*^{4207}-like genes are found exclusively in an operon with TTT-transporter-like genes in Actinobacteria, even in those genera who do not have a KAT-like enzyme (Fig. 8B). This raises the interesting possibility that USP^{4207} could regulate the activity of other GNAT-like enzymes present in bacteria that do not contain KAT-like enzymes, because of structural similarity amongst acyltransferase enzymes (60).

The action of USP^{4207} as a competitive inhibitor of additional protein substrates of GNAT-like enzymes is distinct from the inhibitory actions of small motifs that inhibit acetyl CoA binding to histone acetyltransferases (61), for example. Indeed, our findings have more general implications when considering the activity of an enzyme against multiple substrates. The relative concentrations of different protein substrates and their varying affinities for the enzyme will determine which protein becomes modified by enzymatic activity in the cell. Therefore, promiscuity in enzyme-substrate interactions may actually provide an additional level of regulation within the cell, as we have shown here.

Experimental Procedures

Mycobacterial strains, culture conditions and preparation of cell lysates

Mycobacterium smegmatis mc^2 155 (ATCC 700084) was grown at 37 °C in Middlebrook 7H9 broth (BD Biosciences) supplemented with 0.2% glycerol and 0.05% Tween 80 with shaking at 200 rpm or in Middlebrook 7H10 agar (BD Biosciences) supplemented with 0.5% glycerol. Wherever necessary, cells were grown in Sauton's medium containing 0.2% glycerol and 0.05% Tween 20, further supplemented with 2 g/L of citrate, succinate or tartrate as required.

For the preparation of cell lysates, stationary phase cells were harvested and washed with TBST containing 10 mM Tris-Cl (pH 8.2), 0.9% NaCl and 0.1% Tween 80. Cells were then lysed by sonication in buffer containing 50 mM Tris-Cl (pH 8.2), 100 mM NaCl, 10% glycerol, 5 mM 2-mercaptoethanol (2-ME), 2 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM benzamidine hydrochloride, 3 mM nicotinamide and 1 µM trichostatin A, followed by centrifugation at 30,000 x *g* for 30 min at 4 °C. The supernatant was collected and protein estimation was performed by Bradford method.

RNA isolation and reverse transcription

Cells were harvested, resuspended in Tri reagent (Sigma-Aldrich) and lysed by bead beating (BioSpec Products). The lysate was then heated at 65 °C for 5 min and centrifuged at 16,000 x *g* for 10 min at 4 °C. The supernatant was collected and mixed with chloroform followed by centrifugation at 16,000 x *g* for 15 min at 4 °C. The upper aqueous phase was collected and RNA was precipitated with isopropyl alcohol. The RNA pellet was washed with 75% ethanol, dissolved in RNase-free Milli-Q water and treated with RNase-free DNase (Thermo Scientific). 2 µg of RNA was used for

reverse transcription using 200 units of reverse transcriptase (Thermo Scientific). Sequences of primers used to study the transcript level of different genes are shown in Table 1. Real-time PCR was performed using SYBR Premix *Ex Taq* (Tli RNase H Plus) on a CFX96 Touch real-time PCR detection system (Bio-Rad, USA). The housekeeping gene 16s was used as internal normalization control.

Western blotting

Protein samples were electrophoresed on an SDS-polyacrylamide gel prior to transfer to a polyvinylidene difluoride membrane (Immobilon-P, Millipore). Both MSMEG_4207 and CRP (Rv3676) polyclonal antibodies were generated in the laboratory and were used at a dilution of 1:5000. Anti-acetyl lysine antibody (Cell Signalling Technology, Inc.) was used at a dilution of 1:1000 for whole cell lysate and 1:2500 for recombinant protein. Horseradish peroxidase-conjugated secondary antibody (GE Healthcare) was used at a dilution of 1:50,000 and detected by enhanced chemiluminescence (Luminata Crescendo, Millipore) following the manufacturer's protocol. A standard curve was generated by measuring the intensity of bands obtained on Western blot analysis using known concentrations of USP, from which levels of USP could be measured in planktonic cultures and the pellicle.

Luciferase reporter assay

The intergenic region (103 bp) present upstream of *msmeg_4210/4211* was amplified by PCR using primers MSMEG_4210-4211 INT FWD *SpeI* and MSMEG_4210-4211 INT REV *BamHI*. The PCR amplicon was cloned into pBluescript II KS vector and the sequence of the insert was verified (Macrogen, S. Korea). The *Bam*HI-*Spe*I digested insert from the above pBluescript II KS construct was ligated separately with *Xba*I-*Bam*HI digested pMV-no prom-LuxAB and *Bam*HI-*Spe*I digested pMV-no prom-LuxAB, respectively. Wild type *M. smegmatis* Δ*katms* strains was electroporated with the plasmids. Cultures from individual single colonies were grown till OD_{600} of ~1.0 and diluted into fresh medium to an OD_{600} of 0.02. Aliquots of cultures at indicated time points were taken, and luciferase counts were measured as previously described (62).

To generate shorter fragments for gel shift assays, PCR was performed on plasmid harbouring the entire intergenic region using primers USP-prom_short_*SpeI* fwd and MSMEG_4210-4211INT REV BamH1 (to generate a PCR product containing 56 bp of the intergenic region proximal to *msmeg_4210*) and primers MSMEG_4210-4211 INT REV *BamHI* and MSMEG_4210-4211 INT FWD *SpeI* (to generate a PCR product containing 60 bp of the intergenic region proximal to *msmeg_4211*). PCR products were purified and used in EMSA.

Cloning, expression and purification of proteins

PCR was carried out on genomic DNA of *M. smegmatis* mc^2 155 using MSMEG_4212 FWD and MSMEG_4212 REV primers (Table 1). The amplified product was digested with *Eco*RI-*Not*I and cloned into similarly digested pPROEx-HTa to generate plasmid pPRO-MS4212. The clone was confirmed by sequencing (Macrogen, S. Korea).

The mutant MSMEG_4212 (D54A) was generated by site directed mutagenesis (63) of the pPRO-MS4212 plasmid. The mutation was confirmed by sequencing. Hexahistidine-tagged wild type and mutant MSMEG_4212 were purified by Ni-NTA

affinity chromatography as described earlier (8).

Cloning, expression and purification strategies for KATms, USP^{4207}, GST-KATms, Rv3667 (Acs) and generation of USP^{4207}K104R mutant have been described earlier (8,13).

Electrophoretic mobility shift assay (EMSA)

A double-stranded oligonucleotide harbouring the intergenic region (104 bp) present upstream of *msmeg_4210/4211* was end labeled using [γ- ^{32}P]-ATP by T4 polynucleotide kinase (PNK). The labelled oligonucleotide was purified and ~ 50 fmol (~ 10,000 cpm) was incubated with varying concentration of either wild type or D54A mutant MSMEG_4212 in buffer containing 25 mM Tris-Cl (pH 7.8), 5 mM $MgCl_2$, 50 mM KCl, 10 µM EDTA, 10% glycerol, 1 µg of poly-dI/dC and 50 µg/mL BSA in a 20 µL reaction volume (62) at 37° C for 20 min. Gels were electrophoresed at 50v for 1h at 4°C prior to loading the samples on 6% polyacrylamide gels in Tris-borate/ EDTA (45 mM Tris-borate and 1 mM EDTA) buffer at 100 V for 1.5 h at 4 °C for the separation of bound complex from the free probe. The gels were dried and scanned using a phosphoimager (Azure Biosystems).

For competition with unlabelled DNA fragments, wild type MSMEG_4212 (300 ng; 12.6 pmol) or the D54A mutant protein (30 ng; 1. 26 pmol) were incubated at 37 °C for 20 min in the absence or presence of the unlabelled double-stranded oligonucleotide (104 bp intergenic region or two halves as indicated in Figure 3C; 20 pmol each), along with the radiolabelled probe.

Generation of Δusp^{4207} strain

A strain with deletion of *usp^{4207}* was generated using the homologous recombination-based suicide vector approach (35). The region 5' to *usp^{4207}*consisting of ~988 bp upstream was amplified by PCR using primers up4207 FWD and up4207REV, and a fragment of ~1035 bp downstream of *usp^{4207}*was amplified using down4207FWD and down4207 REV primers (Table 1). The 5'amplicon was cloned into pBluescript II KS vector using *Pst*I-*Hind*III sites, and the 3'amplicon was cloned into pBluescript II KS vector using *Hind*III-*Not*I sites to generate plasmids pBKS-MS4207–5'KO and pBKS-MS4207–3'KO, respectively. The clones were confirmed by sequencing (Macrogen, S. Korea). The *Pst*I-*Hind*III digested pBKS-MS4207–5'KO and *Hind*III-*Not*I digested pBKS-MS4207–3'KO inserts were ligated into *Pst*I-*Not*I digested p2NIL plasmid to generate p2NILMS4207–5'3'KO plasmid. The *Pac*I fragment from pGOAL19 containing the marker gene cassette (*Ag85p-lacZ*, *hyg^r*, and *hsp60p-sacB*) was cloned into p2NIL-MS4207-5'3'KO to generate plasmid p2NIL-MS4207–5'3'KO-*Pac*I. 5 µg of the plasmid p2NIL-MS4207–5'3'KO-*Pac*I was electroporated into *M. smegmatis* mc^2 155, and single crossovers and double crossovers were obtained essentially as described in the original method (35). Double crossovers were further tested by genomic PCR and Western blotting to obtain the *Δusp^{4207}* strain. The strategy for generation of Δ*katms* has been described earlier (8). Similar strategy was taken to generate *Δusp^{4207} Δkatms* strain.

To construct the *Δusp^{4207}* strain complemented with wild type *usp^{4207}*driven by *sigA* promoter, the *Kpn*I-*Bam*HI fragment from pBKS-sigA and *Bam*HI-*Hind*III fragment from pPRO-MS4207 were cloned into *Kpn*I-*Hind*III digested pMV 10-25 plasmid to generate pMV-

MS4207Comp. Identical cloning strategy was taken to generate pMV-MS4207$_{K104R}$Comp. The pMV-MS4207Comp or pMV-MS4207$_{K104R}$Comp plasmids were electroporated into the *Δusp*^{4207}strain. The positive transformants carrying the required plasmid were screened by colony PCR and further validated by RT-PCR and Western blot.

Study of colony morphology

Mid-log phase cells grown in Middlebrook 7H9 medium supplemented with 0.2% glycerol and 0.05% Tween 80 were harvested and resuspended in fresh 7H9 medium to an OD_{600} of 0.5. 5 µL of neat and serially diluted (10^{-1} and 10^{-2} dilution) cultures were spotted on the 7H10 agar plates, with or without 0.05% Tween 80 and incubated at 37 °C for 48-72 h until visible growth was observed.

GST pulldown assays

GST-KATms was expressed in the *E. coli* SP850 strain upon induction by 0.5 mM isopropyl 1-thio-β-D-galactopyranoside. Following purification, GST or GST fusion protein bound to glutathione beads were interacted with 0.5 mg of cytosolic proteins from *Δusp*^{4207} strain complemented either with wild type or mutant (*usp*^{4207}*K104R*) form of *Δusp*^{4207}at 4 °C for 1 h. The beads were then washed five times with 1 mL buffer containing 100 mM Tris-Cl (pH 8.2), 150 mM NaCl, 10% glycerol, 5 mM 2-ME, 2 mM PMSF and 1 mM benzamidine hydrochloride. The bound proteins were analyzed on a 13.5% SDS-polyacrylamide gel.

Mass spectrometric analysis of acetylated proteins

Lysates were subjected to SDS gel electrophoresis analysed by Western blotting using anti-acetyl lysine antibodies. The blot was aligned with the gel and a region containing acetylated proteins from the wild type strain, and the corresponding region from the Δ*katms* strain were excised. Pieces were directly shipped to the Taplin Mass Spectrometric Facility in Harvard University, Boston, USA, where they were reduced with 10 mM DTT, followed by alkylation with iodoacetamide. Samples were then washed, dried, and subjected to tryptic digestion at 37ºC for 16 h. Peptides were collected and sequenced using an Orbitrap Mass Spectrometer from Thermo Scientific. Peptides, and the proteins they were present in, were identified based on the predicted proteome of *M. smegmatis*. Further analysis was performed to detect acetylated peptides differing by a mass of 42 Da, representing the addition of an acetyl group to a lysine residue in the sequence. It was ensured that the same peptides were identified in a non-acetylated form, in the Δ*katms* strain. Here, we report only peptides whose acetylation was completely lost in the Δ*katms* strain.

Pellicle formation assay

Pellicle formation was monitored by growing static cultures of mycobacteria with an initial OD_{600} of 0.1 in Middlebrook 7H9 medium (without Tween 80) at 37 °C for 72-96 h in either polystyrene 3.5 cm dishes or T-25 tissue culture flasks. The formation of pellicle was identified visually. To study the expression profile of different genes during the biofilm formation, the pellicle formed at the air-liquid interface was collected, followed by RNA and/or protein isolation as required.

To prepare whole cell lysates for anti-acetyl lysine Western blot, cells, planktonic cultures at OD ~ 2 or those harvested from strains after pellicle formation (flocculent cells seen in the Δ*usp* strain) were washed with TBS containing

10 mM Tris-Cl (pH 8.2), and 0.9% NaCl. Cells were lysed by bead beating using 0.5 mm diameter glass beads (BioSpec Products) in buffer containing 50 mM Tris-Cl (pH 8.2), 100 mM NaCl, 10% glycerol, 5 mM 2-mercaptoethanol (2-ME), 2 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM benzamidine hydrochloride and 3 mM nicotinamide. Samples were centrifuged at 1000 x g for 5 min at 4 °C to separate the glass beads. The supernatant was centrifuged again at 3000 x g for 5 min at 4 °C to separate the unlysed cells and the supernatant was collected and protein estimation performed by Bradford method. Protein (50 µg) was loaded for Western blot analysis.

Analysis of GPLs in pellicle

Cells from biofilm or flocculent cells (in the case of Δ*usp*^{4207}) were harvested in phosphate buffered saline and the suspension sonicated at 25 °C for 60 min to disrupt clumps. A fraction of the cell suspension was used to prepare lysates for protein estimation. The remainder was centrifuged at 3000 x g for 10 min, supernatant was discarded and the biomass dried at 65 °C in a glass tube. To the dried biomass, 5 ml of chloroform/ methanol (2:1, v/v) was added, vortexed and sonicated at 25 °C for 45 min. The tubes were then incubated overnight in an end-over rocker. The suspension was centrifuged at 4000 rpm for 20 min and the supernatant was collected. 0.2 volume of 0.9% NaCl was added to the supernatant, the tube vortexed and mixed on a rotator for 20 min. The mixture was centrifuged for 5 min at 4000 rpm allowing two phases to form. The lower organic phase containing glycopeptidolipids (GPLs) was collected, transferred to a separate tube and evaporated to dryness at 65 °C. Dried GPLs were resuspended in different volumes of solvent chloroform/ methanol (2:1, v/v), according to the protein content in the sample (final concentration of 64 mg/ ml protein). 5 µl from each sample was spotted on a silica gel 60 F_{254} plate (Merck) and resolved using chloroform/ methanol (100: 7, v/v) as the solvent system. The plate was dried for 3 h followed by spraying with 0.1% orcinol in 40% sulfuric acid. The plate was incubated at 95 °C for 45 min until bands were revealed.

In vitro acetylation assay

Assays were carried out in a 20 µL total reaction volume containing 25 mM Tris-Cl (pH 8.2), 100 mM NaCl, 5 mM EDTA, 30 µM acetyl-CoA, 10 µM of cAMP, KATms (15 pmol), Acs (Rv3667; 75 pmol) and varying concentrations of USP^{4207}. Reactions were incubated at 25 °C for 20 min, stopped by heating in SDS sample buffer, and analyzed by Western blotting using anti-acetyl lysine antibody, quantitated by densitometric analysis of the bands and normalized for protein content in a lane by blotting with anti-his antibody. To determine the extent of acetylation of Acs, the intensity of the band in each lane of the blot probed with acetyl lysine antibodies was estimated and normalized to the amount of Acs in that lane based on the intensity of bands seen in the blot with anti-hexahistidine antibodies. The value obtained was then expressed as a fraction of the extent of acetylation seen in the absence of USP^{4207}.

Measurement of cAMP

Both the planktonic and pellicle cells were harvested by centrifugation; cell pellets were resuspended in 0.1 N HCl and heated for 10 min at 95 °C. Samples were stored at -70 °C until further use. Aliquots were taken for estimation of cAMP by radioimmunoassay (64).

All data shown was analysed using GraphPad Prism 8 using the two-tailed t test.

Acknowledgements: Support from the Department of Biotechnology, Government of India is acknowledged (BT/PR15216/COE/34/02/2017) and funding from DBT-IISc Partnership Program Phase-II BT/PR27952/INF/22/212/2018/21.01.2019. SSV is a JC Bose National Fellow (SB/S2/JCB-18/2013) and a Margdarshi Fellow supported by the Wellcome Trust India Alliance. SS is a DS Kothari Post-doctoral fellow, and AB is a recipient of a Senior Research Fellowship from the Council of Scientific and Industrial Research, Government of India.

Authors declare no conflict of interest.

References

1. Vollmer, A. C., and Bark, S. J. (2018) Twenty-Five Years of Investigating the Universal Stress Protein: Function, Structure, and Applications. *Adv Appl Microbiol* **102**, 1-36
2. O'Connor, A., and McClean, S. (2017) The Role of Universal Stress Proteins in Bacterial Infections. *Curr Med Chem* **24**, 3970-3979
3. O'Toole, R., Smeulders, M. J., Blokpoel, M. C., Kay, E. J., Lougheed, K., and Williams, H. D. (2003) A two-component regulator of universal stress protein expression and adaptation to oxygen starvation in Mycobacterium smegmatis. *J Bacteriol* **185**, 1543-1554
4. Tkaczuk, K. L., I, A. S., Chruszcz, M., Evdokimova, E., Savchenko, A., and Minor, W. (2013) Structural and functional insight into the universal stress protein family. *Evol Appl* **6**, 434-449
5. Banerjee, A., Adolph, R. S., Gopalakrishnapai, J., Kleinboelting, S., Emmerich, C., Steegborn, C., and Visweswariah, S. S. (2015) A universal stress protein (USP) in mycobacteria binds cAMP. *J Biol Chem* **290**, 12731-12743
6. Hingley-Wilson, S. M., Lougheed, K. E., Ferguson, K., Leiva, S., and Williams, H. D. (2010) Individual Mycobacterium tuberculosis universal stress protein homologues are dispensable in vitro. *Tuberculosis (Edinb)* **90**, 236-244
7. Glass, L. N., Swapna, G., Chavadi, S. S., Tufariello, J. M., Mi, K., Drumm, J. E., Lam, T. T., Zhu, G., Zhan, C., Vilcheze, C., Arcos, J., Chen, Y., Bi, L., Mehta, S., Porcelli, S. A., Almo, S. C., Yeh, S. R., Jacobs, W. R., Jr., Torrelles, J. B., and Chan, J. (2017) Mycobacterium tuberculosis universal stress protein Rv2623 interacts with the putative ATP binding cassette (ABC) transporter Rv1747 to regulate mycobacterial growth. *PLoS Pathog* **13**, e1006515
8. Nambi, S., Basu, N., and Visweswariah, S. S. (2010) cAMP-regulated protein lysine acetylases in mycobacteria. *J Biol Chem* **285**, 24313-24323
9. Carabetta, V. J., and Cristea, I. M. (2017) Regulation, Function, and Detection of Protein Acetylation in Bacteria. *J Bacteriol* **199**
10. Hentchel, K. L., and Escalante-Semerena, J. C. (2015) Acylation of Biomolecules in Prokaryotes: a Widespread Strategy for the Control of Biological Function and Metabolic Stress. *Microbiol Mol Biol Rev* **79**, 321-346
11. Liu, F., Yang, M., Wang, X., Yang, S., Gu, J., Zhou, J., Zhang, X. E., Deng, J., and Ge, F. (2014) Acetylome analysis reveals diverse functions of lysine acetylation in Mycobacterium tuberculosis. *Mol Cell Proteomics* **13**, 3352-3366

12. Soufi, B., Soares, N. C., Ravikumar, V., and Macek, B. (2012) Proteomics reveals evidence of cross-talk between protein modifications in bacteria: focus on acetylation and phosphorylation. *Curr Opin Microbiol* **15**, 357-363
13. Nambi, S., Gupta, K., Bhattacharyya, M., Ramakrishnan, P., Ravikumar, V., Siddiqui, N., Thomas, A. T., and Visweswariah, S. S. (2013) Cyclic AMP-dependent protein lysine acylation in mycobacteria regulates fatty acid and propionate metabolism. *J Biol Chem* **288**, 14114-14124
14. Yang, H., Sha, W., Liu, Z., Tang, T., Liu, H., Qin, L., Cui, Z., Chen, J., Liu, F., Zheng, R., Huang, X., Wang, J., Feng, Y., and Ge, B. (2018) Lysine acetylation of DosR regulates the hypoxia response of Mycobacterium tuberculosis. *Emerg Microbes Infect* **7**, 34
15. Favrot, L., Blanchard, J. S., and Vergnolle, O. (2016) Bacterial GCN5-Related N-Acetyltransferases: From Resistance to Regulation. *Biochemistry* **55**, 989-1002
16. Vergnolle, O., Xu, H., Tufariello, J. M., Favrot, L., Malek, A. A., Jacobs, W. R., Jr., and Blanchard, J. S. (2016) Post-translational Acetylation of MbtA Modulates Mycobacterial Siderophore Biosynthesis. *J Biol Chem* **291**, 22315-22326
17. Noy, T., Xu, H., and Blanchard, J. S. (2014) Acetylation of acetyl-CoA synthetase from Mycobacterium tuberculosis leads to specific inactivation of the adenylation reaction. *Arch Biochem Biophys* **550-551**, 42-49
18. Anand, C., Garg, R., Ghosh, S., and Nagaraja, V. (2017) A Sir2 family protein Rv1151c deacetylates HU to alter its DNA binding mode in Mycobacterium tuberculosis. *Biochem Biophys Res Commun* **493**, 1204-1209
19. Hayden, J. D., Brown, L. R., Gunawardena, H. P., Perkowski, E. F., Chen, X., and Braunstein, M. (2013) Reversible acetylation regulates acetate and propionate metabolism in Mycobacterium smegmatis. *Microbiology* **159**, 1986-1999
20. Xu, H., Hegde, S. S., and Blanchard, J. S. (2011) Reversible acetylation and inactivation of Mycobacterium tuberculosis acetyl-CoA synthetase is dependent on cAMP. *Biochemistry* **50**, 5883-5892
21. Weinert, B. T., Satpathy, S., Hansen, B. K., Lyon, D., Jensen, L. J., and Choudhary, C. (2017) Accurate Quantification of Site-specific Acetylation Stoichiometry Reveals the Impact of Sirtuin Deacetylase CobB on the E. coli Acetylome. *Mol Cell Proteomics* **16**, 759-769
22. Nambi, S., Badireddy, S., Visweswariah, S. S., and Anand, G. S. (2012) Cyclic AMP-induced conformational changes in mycobacterial protein acetyltransferases. *J Biol Chem* **287**, 18115-18129
23. Shenoy, A. R., Srinivasan, N., and Visweswariah, S. S. (2002) The ascent of nucleotide cyclases: conservation and evolution of a theme. *J Biosci* **27**, 85-91
24. Shenoy, A. R., and Visweswariah, S. S. (2006) Mycobacterial adenylyl cyclases: biochemical diversity and structural plasticity. *FEBS Lett* **580**, 3344-3352
25. Rosa, L. T., Bianconi, M. E., Thomas, G. H., and Kelly, D. J. (2018) Tripartite ATP-Independent Periplasmic (TRAP) Transporters and Tripartite Tricarboxylate Transporters (TTT): From Uptake to Pathogenicity. *Front Cell Infect Microbiol* **8**, 33
26. Brocker, M., Schaffer, S., Mack, C., and Bott, M. (2009) Citrate utilization by Corynebacterium glutamicum is controlled by the CitAB two-component system through positive regulation of the citrate transport genes citH and tctCBA. *J Bacteriol* **191**, 3869-3880
27. Rosa, L. T., Springthorpe, V., Bianconi, M. E., Thomas, G. H., and Kelly, D. J. (2018) Massive over-representation of solute-binding proteins (SBPs) from the tripartite tricarboxylate transporter (TTT) family in the genome of the alpha-proteobacterium Rhodoplanes sp. Z2-YC6860. *Microb Genom*

28. Antoine, R., Huvent, I., Chemlal, K., Deray, I., Raze, D., Locht, C., and Jacob-Dubuisson, F. (2005) The periplasmic binding protein of a tripartite tricarboxylate transporter is involved in signal transduction. *J Mol Biol* **351**, 799-809
29. Kaspar, S., Perozzo, R., Reinelt, S., Meyer, M., Pfister, K., Scapozza, L., and Bott, M. (1999) The periplasmic domain of the histidine autokinase CitA functions as a highly specific citrate receptor. *Mol Microbiol* **33**, 858-872
30. Kaspar, S., and Bott, M. (2002) The sensor kinase CitA (DpiB) of Escherichia coli functions as a high-affinity citrate receptor. *Arch Microbiol* **177**, 313-321
31. Bott, M., and Brocker, M. (2012) Two-component signal transduction in Corynebacterium glutamicum and other corynebacteria: on the way towards stimuli and targets. *Appl Microbiol Biotechnol* **94**, 1131-1150
32. Salazar, M. E., and Laub, M. T. (2015) Temporal and evolutionary dynamics of two-component signaling pathways. *Curr Opin Microbiol* **24**, 7-14
33. Willett, J. W., and Crosson, S. (2017) Atypical modes of bacterial histidine kinase signaling. *Mol Microbiol* **103**, 197-202
34. Delgado, J., Forst, S., Harlocker, S., and Inouye, M. (1993) Identification of a phosphorylation site and functional analysis of conserved aspartic acid residues of OmpR, a transcriptional activator for ompF and ompC in Escherichia coli. *Mol Microbiol* **10**, 1037-1047
35. Parish, T., and Stoker, N. G. (2000) Use of a flexible cassette method to generate a double unmarked Mycobacterium tuberculosis tlyA plcABC mutant by gene replacement. *Microbiology* **146 (Pt 8)**, 1969-1975
36. Etienne, G., Villeneuve, C., Billman-Jacobe, H., Astarie-Dequeker, C., Dupont, M. A., and Daffe, M. (2002) The impact of the absence of glycopeptidolipids on the ultrastructure, cell surface and cell wall properties, and phagocytosis of Mycobacterium smegmatis. *Microbiology* **148**, 3089-3100
37. Kocincova, D., Singh, A. K., Beretti, J. L., Ren, H., Euphrasie, D., Liu, J., Daffe, M., Etienne, G., and Reyrat, J. M. (2008) Spontaneous transposition of IS1096 or ISMsm3 leads to glycopeptidolipid overproduction and affects surface properties in Mycobacterium smegmatis. *Tuberculosis (Edinb)* **88**, 390-398
38. Recht, J., Martinez, A., Torello, S., and Kolter, R. (2000) Genetic analysis of sliding motility in Mycobacterium smegmatis. *J Bacteriol* **182**, 4348-4351
39. Xu, J. Y., Zhao, L., Liu, X., Hu, H., Liu, P., Tan, M., and Ye, B. C. (2018) Characterization of the Lysine Acylomes and the Substrates Regulated by Protein Acyltransferase in Mycobacterium smegmatis. *ACS Chem Biol* **13**, 1588-1597
40. Fujiwara, N., Ohara, N., Ogawa, M., Maeda, S., Naka, T., Taniguchi, H., Yamamoto, S., and Ayata, M. (2015) Glycopeptidolipid of Mycobacterium smegmatis J15cs Affects Morphology and Survival in Host Cells. *PLoS One* **10**, e0126813
41. Billman-Jacobe, H., McConville, M. J., Haites, R. E., Kovacevic, S., and Coppel, R. L. (1999) Identification of a peptide synthetase involved in the biosynthesis of glycopeptidolipids of Mycobacterium smegmatis. *Mol Microbiol* **33**, 1244-1253
42. Chen, J., Kriakov, J., Singh, A., Jacobs, W. R., Jr., Besra, G. S., and Bhatt, A. (2009) Defects in glycopeptidolipid biosynthesis confer phage I3 resistance in Mycobacterium smegmatis. *Microbiology* **155**, 4050-4057
43. Recht, J., and Kolter, R. (2001) Glycopeptidolipid acetylation affects sliding motility and biofilm formation in Mycobacterium smegmatis. *J Bacteriol* **183**, 5718-5724
44. Sochorova, Z., Petrackova, D., Sitarova, B., Buriankova, K., Bezouskova, S., Benada, O., Kofronova, O., Janecek, J., Halada, P., and Weiser, J. (2014) Morphological and proteomic analysis of early stage air-liquid interface biofilm formation in Mycobacterium smegmatis. *Microbiology* **160**, 1346-1356

45. Hung, D. C., Downey, J. S., Kreth, J., Qi, F., Shi, W., Cvitkovitch, D. G., and Goodman, S. D. (2012) Oligomerization of the response regulator ComE from Streptococcus mutans is affected by phosphorylation. *J Bacteriol* **194**, 1127-1135
46. Gao, R., Bouillet, S., and Stock, A. M. (2019) Structural Basis of Response Regulator Function. *Annu Rev Microbiol* **73**, 175-197
47. Feldheim, Y. S., Zusman, T., Speiser, Y., and Segal, G. (2016) The Legionella pneumophila CpxRA two-component regulatory system: new insights into CpxR's function as a dual regulator and its connection to the effectors regulatory network. *Mol Microbiol* **99**, 1059-1079
48. Podobnik, M., Siddiqui, N., Rebolj, K., Nambi, S., Merzel, F., and Visweswariah, S. S. (2014) Allostery and conformational dynamics in cAMP-binding acyltransferases. *J Biol Chem* **289**, 16588-16600
49. Lee, H. J., Lang, P. T., Fortune, S. M., Sassetti, C. M., and Alber, T. (2012) Cyclic AMP regulation of protein lysine acetylation in Mycobacterium tuberculosis. *Nat Struct Mol Biol* **19**, 811-818
50. Bi, J., Wang, Y., Yu, H., Qian, X., Wang, H., Liu, J., and Zhang, X. (2017) Modulation of Central Carbon Metabolism by Acetylation of Isocitrate Lyase in Mycobacterium tuberculosis. *Sci Rep* **7**, 44826
51. Gu, L., Chen, Y., Wang, Q., Li, X., Mi, K., and Deng, H. (2015) Functional Characterization of Sirtuin-like Protein in Mycobacterium smegmatis. *J Proteome Res* **14**, 4441-4449
52. Richards, J. P., and Ojha, A. K. (2014) Mycobacterial Biofilms. *Microbiol Spectr* **2**
53. Pawlik, A., Garnier, G., Orgeur, M., Tong, P., Lohan, A., Le Chevalier, F., Sapriel, G., Roux, A. L., Conlon, K., Honore, N., Dillies, M. A., Ma, L., Bouchier, C., Coppee, J. Y., Gaillard, J. L., Gordon, S. V., Loftus, B., Brosch, R., and Herrmann, J. L. (2013) Identification and characterization of the genetic changes responsible for the characteristic smooth-to-rough morphotype alterations of clinically persistent Mycobacterium abscessus. *Mol Microbiol* **90**, 612-629
54. Harshey, R. M., and Ramakrishnan, T. (1977) Rate of ribonucleic acid chain growth in Mycobacterium tuberculosis H37Rv. *J Bacteriol* **129**, 616-622
55. Messenger, A. J., and Ratledge, C. (1982) Iron transport in Mycobacterium smegmatis: Uptake of iron from ferric citrate. *J Bacteriol* **149**, 131-135
56. Shimamoto, T., Negishi, K., Tsuda, M., and Tsuchiya, T. (1996) Mutational analysis of the CitA citrate transporter from Salmonella typhimurium: altered substrate specificity. *Biochem Biophys Res Commun* **226**, 481-487
57. Chen, W., Honma, K., Sharma, A., and Kuramitsu, H. K. (2006) A universal stress protein of Porphyromonas gingivalis is involved in stress responses and biofilm formation. *FEMS Microbiol Lett* **264**, 15-21
58. Bottagisio, M., Soggiu, A., Piras, C., Bidossi, A., Greco, V., Pieroni, L., Bonizzi, L., Roncada, P., and Lovati, A. B. (2019) Proteomic Analysis Reveals a Biofilm-Like Behavior of Planktonic Aggregates of Staphylococcus epidermidis Grown Under Environmental Pressure/Stress. *Front Microbiol* **10**, 1909
59. Havis, S., Bodunrin, A., Rangel, J., Zimmerer, R., Murphy, J., Storey, J. D., Duong, T. D., Mistretta, B., Gunaratne, P., Widger, W. R., and Bark, S. J. (2019) A Universal Stress Protein That Controls Bacterial Stress Survival in Micrococcus luteus. *J Bacteriol* **201**
60. Salah Ud-Din, A. I., Tikhomirova, A., and Roujeinikova, A. (2016) Structure and Functional Diversity of GCN5-Related N-Acetyltransferases (GNAT). *Int J Mol Sci* **17**

61. Shi, S., Liu, K., Chen, Y., Zhang, S., Lin, J., Gong, C., Jin, Q., Yang, X. J., Chen, R., Ji, Z., and Han, A. (2016) Competitive Inhibition of Lysine Acetyltransferase 2B by a Small Motif of the Adenoviral Oncoprotein E1A. *J Biol Chem* **291**, 14363-14372
62. Sharma, R., Zaveri, A., Gopalakrishnapai, J., Srinath, T., Varshney, U., and Visweswariah, S. S. (2014) Paralogous cAMP receptor proteins in Mycobacterium smegmatis show biochemical and functional divergence. *Biochemistry* **53**, 7765-7776
63. Shenoy, A. R., and Visweswariah, S. S. (2003) Site-directed mutagenesis using a single mutagenic oligonucleotide and DpnI digestion of template DNA. *Anal Biochem* **319**, 335-336
64. Brooker, G., Harper, J. F., Terasaki, W. L., and Moylan, R. D. (1979) Radioimmunoassay of cyclic AMP and cyclic GMP. *Adv Cyclic Nucleotide Res* **10**, 1-33

Table 1 Primers used in this study

Primer sequences are shown 5' to 3' and a brief description of their use is provided.

SL No.	Primer name	Primer sequence (5'-3')	Purpose
1	MSMEG_4207-4208 RT FWD	CGGCCCAGTCGTGATCGTGG	*msmeg_4207-4208* junction
2	MSMEG_4208-4209 RT FWD	GCTGAGCTTCTACGCGTTCTAC	*msmeg_4208-4209* junction
3	MSMEG_4208-4209 RT REV	GCGTACAGCAGGTTCATCG	*msmeg_4208-4209* junction
4	MSMEG_4209-4210 RT FWD	CAGTTCCTCAAGGACCAGGACG	*msmeg_4209-4210* junction
5	MSMEG_4209-4210 RT REV	GCGCACACGAGGTACTGC	*msmeg_4209-4210* junction
6	MSMEG_4211-4212 RT FWD	AGGTGACCGTTCATGCGTGACGTGC	*msmeg_4211-4212* junction
7	MSMEG_4211-4212 RT REV	GCAGATACACGTCGAGCAGGATCAGG	*msmeg_4211-4212* junction
8	16S rRNA RT FWD	CGAGCGTTGTCCGGAATTA	*16s* gene for qRTPCR
9	16S rRNA RT REV	TCCACCGCTACACCAGGAAT	
10	MSMEG_4207 RT FWD	GAGCACGGGATCGAAGAGGC	For qRTPCR
11	MSMEG_4207 RT REV	CTCGACGTCATGGACCTCGC	For qRTPCR. Also as a junction primer used in conjunction with primer 1.
12	MSMEG_4210 RT FWD	TTGCGCCGCCTGGTATTTCG	For qRTPCR
13	MSMEG_4210 RT REV	CGTGGTGAAGGCGTCGCTCC	
14	MSMEG_5458 RT FWD	GGGCATCGGCAGCTTCTTGATGGG	
15	MSMEG_5458 RT REV	CGCACCCACACCGCCCCCAG	
16	MSMEG_4620 RT FWD	GCGGTGATGCTGGAGGAGGCC	
17	MSMEG_4620 RT REV	CGAATACGCCTGCTCCACACG	
18	MSMEG_5175 RT FWD	GCGCGGGCAGCACCAATGTG	
19	MSMEG_5175 RT REV	CCACGTCGGCGCTGCTCACC	
20	MSMEG_4211 RT FWD	CGGATTCGGGTCCCGGAGTGCCCG	
21	MSMEG_4211 RT REV	CCGGGAGGCGGACGACGAACGACG	
22	MSMEG_4212 RT FWD	GGAGGTACTCAAACGACTTCGGGCG	
23	MSMEG_4212 RT REV	ACGAGCTTGGCTCGCAACTGGTCG	
24	MSMEG_5458 FWD	ACGTCGGATCCTGGGAATGTGGCGGAAC	To amplify *msmeg_5458* from genomic DNA
25	MSMEG_5458 REV	CCTGAATTCCGCCCCACTCACTGGCTCAC	
26	MSMEG_4207 FWD	GGCGGCGGATCCGTGATCGTGGTCGGTTAC	To amplify *msmeg_4207* from genomic DNA
27	MSMEG_4207 REV	GTCACTCGAGATCAGAACCCATGTGGCTTG	
28	MSMEG_4210-4211 INT FWD *SpeI*	GGACTAGTAGACCAAAACTTCCTTTGCG	To generate a PCR product containing 60 bp of the intergenic region proximal to *msmeg_4211*
29	MSMEG_4210-4211 INT REV *BamHI*	GCGGATCCTGCAGTCTCCGATCACTGC	To generate a PCR product either

			containing 56 bp of the intergenic region proximal to *msmeg_4210*
30	MSMEG_4212 FWD	GTGAGAATTCATGCGTGACGTGCTGG	To amplify *msmeg_4212* from genomic DNA
31	MSMEG_4212 REV	CACGAGGCGGCCGTAACCCCATTCGACTGG	To amplify *msmeg_4212* from genomic DNA
32	up4207 FWD	TGCCCCTGCAGAGCGACGGAATCG	To amplify a fragment of ~988 bp upstream of *msmeg_4207*
33	uP4207 REV	CCGGCCAAGCTTGTCCGCGCTGTAACCGACC	To amplify a fragment of ~988 bp upstream of *msmeg_4207*
34	down4207 FWD	TGTCCAAAGCTTGTGCTCGCGGTCAAGC	To amplify a fragment of ~1035 bp downstream of *msmeg_4207*
35	down4207 REV	CGATCATGGCGGCCGCGGTGCGGATGTCC	To amplify a fragment of ~1035 bp downstream of *msmeg_4207*
36	MSMEG_5404 RT FWD	GAAGGCGCGATCTGCATCAA	For qRTPCR
37	MSMEG_5404 RT REV	GAAGGCGCGATCTGCATCAA	
38	MSMEG_0401 RT FWD	GTACCGGCGATTCGTCTCGT	
39	MSMEG_0401 RT REV	CGCTGACACCTGGAACGACT	
40	USP-PROM_short_*SpeI* FWD	GGACTAGTTCACGTGCGGGTCCAGTAT	To generate a PCR product containing 56 bp of the intergenic region proximal to *msmeg_4210*
41	MSMEG_5404_*EcoRI* FWD	GGACTAGTTCACGTGTCGGGTCCAGTAT	Cloning of msmeg_5404 for recombinant protein expression
42	MSMEG_5404_*HindIII* REV	CCCAAGCTTTCAGCCCCGGCGCAACACCGACTTG	
43	USP_PROM_short_*BamHI*_REV	GCGGATCCATACTGGACCCGACACGTGA	
44	MSMEG_4208 RT FWD	TGGTCTACGTCCTGATGGCG	For qRTPCR
45	MSMEG_4208 RT REV	CGCCTTTTCCTTCTGATCCGC	
46	MSMEG_4209 RT FWD	GGCAACCACCACTACGTCCG	
47	MSMEG_4209 RT REV	ACAGAATCCCGTCCAGAATGCC	
48	MSMEG_4212_D54A_*NheI* FWD	CGACCTGATCCTGCTAGCCGTGTATCTGCCCGAC	For site-directed mutagenesis

Legends to Figures

Figure 1 USP^{4207} is in an operon with a putative tripartite tricarboxylic acid transporter (TTT) and adjacent to a two-component signaling system (A) Schematic of the arrangement of the *TTT-usp^{4207}* operon and the two-component system. Arrows indicate primers used for amplification across the junctions of two genes as annotated in the genome. Indicated above the black arrows are suggested functions of the genes encoded in the operon, as well as the two-component system that is divergently transcribed. Indicated by an open arrow is the intergenic region that was considered as a putative promoter (~100 bp) for the TTT-*usp^{4207}* operon, and in a reverse orientation, the promoter for the two-component system. (B) Reverse transcription (RT) and PCR of RNA prepared from logarithmic growing cultures of *M. smegmatis*, using primers across the junctions of the putative *TTT-usp^{4207}* and two-component system operons. Data shown is representative of RNA prepared from three independently grown cultures. Numbers above lanes containing a product indicate the predicted sizes of amplicons. RNA was either subjected to RT (+) or not (-) prior to PCR. (C) Alignment of the USP^{4207} sequence across Actinobacteria. Highlighted in red is the lysine residue that is acylated by KATms.

Figure 2. The *TTT-USP* operon is induced by tricarboxylic acids (A) Reverse transcription and PCR of RNA prepared from cultures grown in the presence of 0.1 g/L sodium citrate (low) or 2 g/L sodium citrate (high). Primers were designed to amplify the genes indicated, and primers to the *16S* gene were used for normalization in real time PCR reactions. *usp^{4207}* and other genes in the operon (*msmeg_4210-msmeg_4208*) are induced by high levels of citrate. Experiments are representative of RNA prepared from three independent experiments. Data was analysed by the t test. *** $p < 0.001$. (B) Western blot analysis using USP^{4207} -specific antisera (8) and lysates prepared from *M. smegmatis* cells grown in the presence of high or low amounts of citrate and indicated concentrations of di- or tricarboxylic acids. Normalization of protein loading in the Western blot was indicated by equivalent levels of CRP in lysates, using a CRP-specific antiserum (62). Data shown is representative of three independently grown cultures. (C) The ~ 150 bp intergenic region between *msmeg_4210* and *msmeg_4211* genes was cloned in two orientations upstream of the *luxAB* gene, and plasmids electroporated into *M. smegmatis*. Luciferase activity was measured at the indicated times in cultures grown in low (white bars) or high levels (black bars) of sodium citrate. Citrate induction was observed only when the promoter was cloned in an orientation that would drive transcription of the *TTT-usp^{4207}* operon (left panel) and not of the two-component system (right panel). Values shown are from two independent experiments, with each sample taken in duplication, and are the mean ± SD.

Figure 3. MSMEG_4212 binds to the intergenic region between *msmeg_4210* and *msmeg_4211* (A) The putative response regulator (RR) encoded by the *msmeg_4212* gene was expressed in *E. coli* and purified (left panel). A mutant of the RR where an aspartate residue was converted to an alanine was also expressed and purified. (B) Varying concentrations of purified proteins were used to monitor binding to the radiolabelled ~ 100 bp promoter fragment by EMSA (lane 2). Increasing amounts of the wild type or D54A mutant protein (as indicated

above the gels) used were 7.5 ng, 15 ng, 30 ng, 50 ng, 100 ng, 500 ng, 1 µg, 2µg and 4 µg. The filled arrow shows the shifted radiolabelled band representing the protein-bound fraction. The open arrow marks the mobility of the free probe. (C) The sequence between the start of *msmeg_4210* and *msmeg_4211* is shown. The sequences representing the two PCR products generated to identify the binding region of MSMEG_4212 are shown as dotted and filled lines under the sequence. Arrows represent the positioning of adjacent genes to the promoter region. Note that the two PCR products contain a 20 bp overlapping region. (D) Wild type (300 ng) or mutant MSMEG_4212 (50 ng) were incubated with the radiolabelled intergenic region in the absence or presence of unlabelled entire intergenic region, or PCR products representing two halves of the intergenic region. The gel shown is representative of experiments performed twice with two different batches of purified protein.

Figure 4 A *Δusp*^{4207} strain shows altered colony morphology. (A) PCR analysis of genomic DNA prepared from wild type *M. smegmatis* and strains deleted for *usp*^{4207} in the wild type strain (*Δusp*^{4207}), or in a strain where *katms* has been deleted (*Δusp*^{4207}*Δkatms*). Primer pairs used for the genes are indicated and confirm the specific gene deletions. (B) Cultures of the indicated strains were grown and resuspended to an A_{600} of 0.5 and 5 µL of each culture spotted on 7H10-agar plates in the absence or presence of 0.05% Tween 80. *Δusp*^{4207} strains were transformed with a plasmid expressing either wild type USP^{4207} or the USP^{4207}K104R mutant protein driven by the *sigA* promoter to generate complemented strains (*Δusp*^{4207}*/usp*^{4207} or *Δusp/usp*^{4207}*K104R*, respectively). Data shown is representative of at least three independently grown cultures. (C) Western blot analysis to determine expression levels of USP^{4207} in the wild type and *Δkatms* strains using USP^{4207}-specific antibody. Strains were grown independently at least three times, and the blot shown is representative of lysates prepared from paired cultures.

Figure 5 USP^{4207} binds to KATms and competitively inhibits acetylation of Acs (A) GST-KATms pull-down of USP^{4207}. Lysates prepared from *E. coli* strains expressing either wild type USP^{4207} or the USP^{4207}K104R mutant were interacted with GST-KATms or GST bound to glutathione beads. Protein bound to beads was analysed either by Western blot using the USP^{4207} -specific antibody or staining the membrane with Commassie to normalize for amounts of GST-KATms and GST used for interaction. The experiment was performed at least thrice with independent batches of purified proteins. Data shown is from one representative experiment. (B) USP^{4207} acts a competitive inhibitor of KATms. Acetylation assays were performed using purified KATms (15 pmol) and Acs (75 pmol) in the absence or presence of indicated concentrations of USP^{4207}. Samples were subjected to Western blot analysis using an acetyl lysine-specific antibody, or an antibody to hexahistidine (anti-his) found at the N-terminus of the purified proteins to normalize for Acs amounts taken for assay. The graph shows a quantitation of acetylated Acs as a fraction of that seen in the absence of USP. The x-axis is the mole ratio of USP^{4207} to Acs in the reaction tube. The experiment shown is representative of two assays performed with different batches of purified protein.

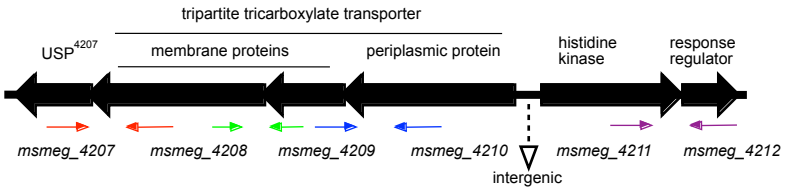
Figure 6 Identification of substrates of KATms and biofilm formation in the *Δusp*^{4207} strain. (A) Samples prepared from planktonic cultures of the indicated strains were subjected to Western blot analysis using either an antibody to acetyl lysine, USP^{4207} or CRP (to normalize for protein loading). Blots were repeated thrice with independently grown cultures. USP^{4207} expressed in the complemented strains contained a hexahistidine N-terminal tag, so the proteins migrate at a higher molecular weight than USP^{4207} in the wild type strain. The lower molecular weight band in the complemented strains could represent protein in which the hexahistidine tag had been cleaved within the bacteria. Parallel gels were run and 0.5 cm pieces corresponding to a region of ~ 80 kDa was cut from the lanes containing lysates of wild type or *Δkatms* strains. The gel piece was analysed by mass spectrometry to identify acetylated peptides found in the wild type strain, but absent in the *Δkatms* strain. Peptides thus identified are shown in the table below with the acetylated lysine residue (K) in bold and italicized. The proteins in which these peptides are present are indicated. (B) Strains as indicated were grown in 7H9 media containing 0.2% glycerol, in the absence of Tween 80 and biofilm was allowed to develop over 3 days. Pictures are shown from the top (left hand side) and from the side of the individual dishes (right hand side). Biofilm was formed in all but the *Δusp*^{4207} strains. The experiment was performed four times with independently grown cultures. (C) Lipids were extracted from cultures after formation of the biofilm, or from flocculent cells from the *Δusp*^{4207} strain. Lipids were spotted after normalizing for protein content present in cells taken for lipid extraction. GPLs were resolved by performing one dimensional thin layer chromatography using chloroform/ methanol (100: 7; vol:vol) as the solvent system, and detected by spraying the plate with 0.1% orcinol in 40% sulfuric acid.

Figure 7 USP^{4207} is upregulated in the biofilm and regulates protein acetylation in the pellicle. (A) RNA was prepared from planktonic cultures of *M. smegmatis* and from the biofilm pellicle. Levels of mRNA transcripts were monitored following reverse transcription and real time PCR, with *16S* rRNA levels used for normalization. Cultures were grown independently Data shown is from three independently grown cultures. . *, $p< 0.05$; **, $p< 0.01$; ***, $p< 0.001$. (B) Gene-specific primer efficiency was determined using varying amounts of DNA and used to estimate the amount of specific cDNA in planktonic and pellicle fractions. Data shown in mean ± SD of three independent experiments. *, $p< 0.05$; **, $p< 0.01$; ***, $p< 0.001$. (C) Western blot analysis was performed on lysates prepared from planktonic cultures and pellicles of biofilm formed by wild type *M. smegmatis*. CRP was used to check if its expression changed in the pellicle or planktonic cultures. The blot shown is representative of two independent experiments. (D) Samples were prepared from the pellicle (or flocculent cells in the case of the *Δusp*^{4207}). Samples were subjected to Western blot analysis with antibodies against acetyl lysine or CRP. Note the absence of a band of ~ 80 kDa from pellicles of wild type and *Δkatms* strains. (E) *M. smegmatis* cells were harvested from planktonic cultures and at the times indicated following pellicle formation. Cyclic AMP was measured by radioimmunoassay. Cultures were grown twice and extracts prepared for independently for measurement. (** $p < 0.01$; Student's t test).

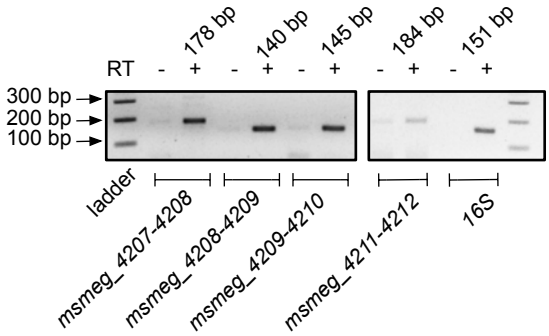
Figure 8. Proposed role of USP^{4207} in regulating KATms-mediate acylation. (A) Predicted regulation of the *TTT-usp^{4207}* operon by the divergent histidine kinase-response regulator gene pair. The *TTT-usp^{4207}* operon is regulated by tricarboxylic acids such as citrate. Citrate can bind to the secreted protein (encoded by the *msmeg_4210* gene represented by a brown arrow) or directly to the histidine kinase (MSMEG_4211), resulting in its phosphorylation. Phosphotransfer to the response regulator may prevent binding of the response regulator to the promoter of the *TTT-usp^{4207}* operon, thereby increasing transcription of the 3 genes that encode the tripartite transporter, (two membrane proteins encoded by genes represented in purple (*msmeg_4208*) and red (*msmeg_4209*) and the secreted protein). *usp^{4207}* (green) is the last gene in the operon. (B) Distribution of the *usp^{4207}* gene, *kat* orthologs and presence of the *TTT-usp^{4207}* operon across actinobacterial members including slow and fast-growing mycobacteria. Green boxes represent the presence of a particular gene(s), and purple boxes the absence of the gene(s). To the right of the boxes is a graphical representation of the organization of the *TTT-usp^{4207}* operon in organisms where it is present. The two-component gene pair is shown in brown arrows, and the *usp* gene in a red arrow.

Figure 9. The complex interplay between protein acylation, cAMP and citrate in *M. smegmatis* biofilm formation. Cyclic AMP binds to MSMEG_3811, which is a ATP-cAMP-binding USP (5) and also to KATms. Cyclic AMP can regulate transcription via CRP (62). Given the abundance of putative cAMP-binding proteins in mycobacteria, cAMP is likely to have a number of additional roles, which are not known at present (indicated by a question mark). KATms can acylate USP^{4207}, and also PrpE, Acs and MSMEG_0401. Transcript levels of *prpE* and *msmeg_0401* increase in biofilm pellicles, suggesting their requirement in biofilm formation. Acylated PrpE (19) and MSMEG_0401 are inactive. During biofilm formation, both cAMP and USP^{4207} levels are increased (Fig. 7). To prevent the unwanted acylation of PrpE, MSMEG_0401 (and perhaps other substrates of KATms), USP^{4207} sequesters KATms from its substrates. The thickness of the blue arrows represent the extent of enzymatic activity of the indicated enzymes, KATms and sirtuins, towards their substrates.

A



B



C

```
gi|500048922 M. smeg HDVEAHLQDSGVPFEIRQPV-GVDATEELLTAMDSPDAELLVIGIRHRNPVGKLLLGSVA
gi|738500577 M. vul. HDVEERLRESGVPFELRQPV-GVDAATELLDAMDAADAELLVIGIRHRSPVGKLLLGSVS
gi|746273817 M. set  HDVEERLRESGVPFELRQPV-GVDAATELLDAMDAPEAELLVIGIRHRSPVGKLLLGSVS
gi|764953514 M. sep. HDVEERLRESGVPFELRQPV-GVDAATELLDAMDAADAELLVIGIRHRSPVGKLLLGSVS
gi|500223792 M. gil. RDVEARLAECEVPYELVQPV-GVDPADELLAAMDRADAEMLVVGIRHRNPVGKLLLGSVA
gi|653362753 Rhodoc. DQLQSTLESSGVPFTVEQDT-TTDPVEALLTAMEKPDADVLVIGIRHRTPVGKLLMGSTA
gi|497366015 Gordon. VEIENRLSESGVKFRITQPI-GATPAEELLNAMDDPEAELLIIGMRRRTQVGKLLLGSTS
gi|515765718 Arthro. QSIEERLADQGIQHIIKQPVRGHDAAAEVLDAADENDADLIVIGLRRRSPVGKLIMGSVS
gi|655023680 Nocard. ATLEEELTASGLEFVVRRPI-AADVAEAVVEAADEESADVIVIGLRRRSPVGKLVMGSVA
gi|490099597 Strept. THVREDLADMGIEFDIRQVLGARDAAEEIIDLAERMDASLVVIGLRRRSAVGKLIMGSAA
gi|754178748 Modest. ERLEQRLGESGVPFEVRQPVRGRDVAEELAAAADDTSAELVVIGLRRRSAVGKLLMGSAA
                          *                                 *     * * *  ****  **
```

Figure 1

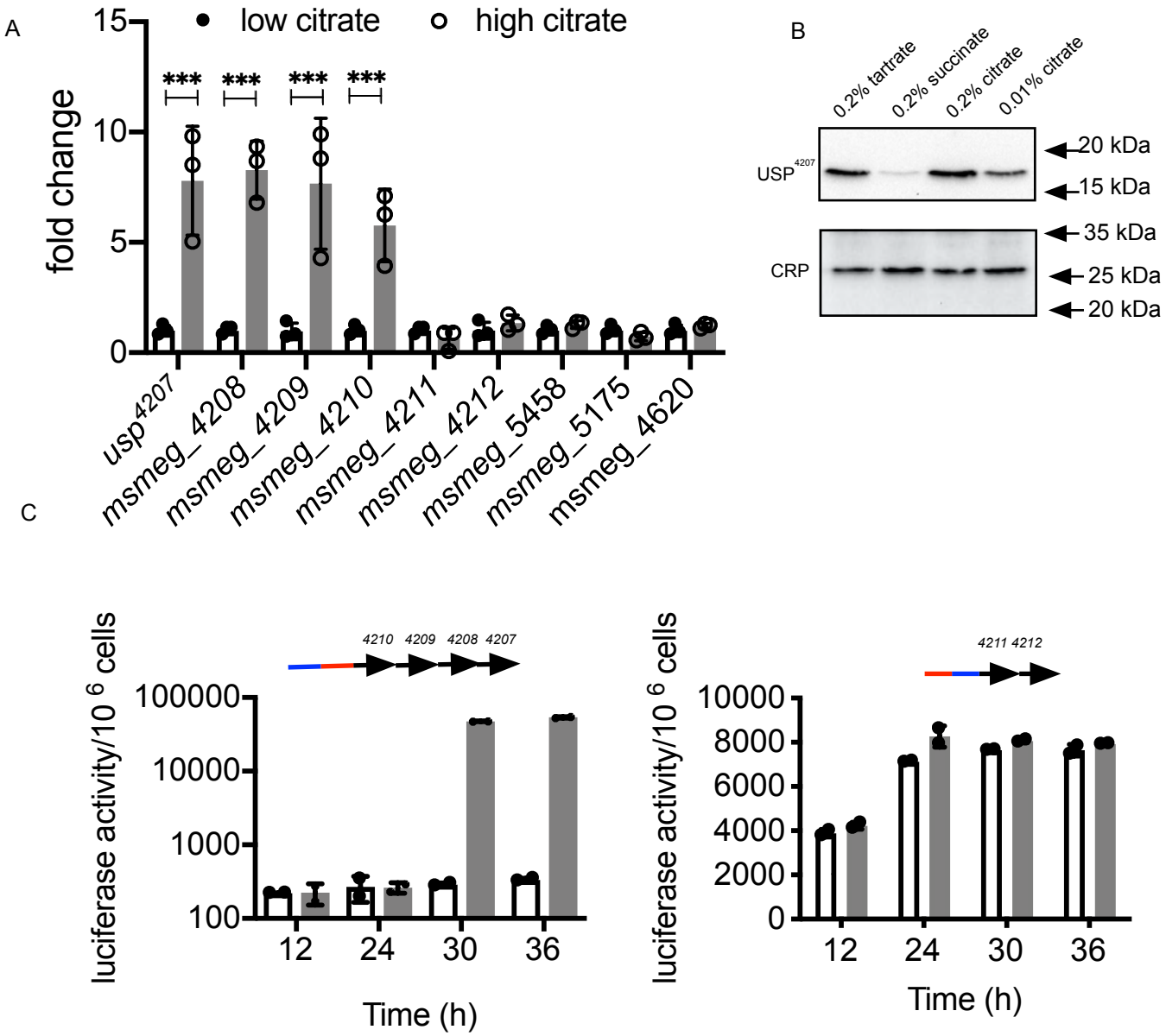


Figure 2

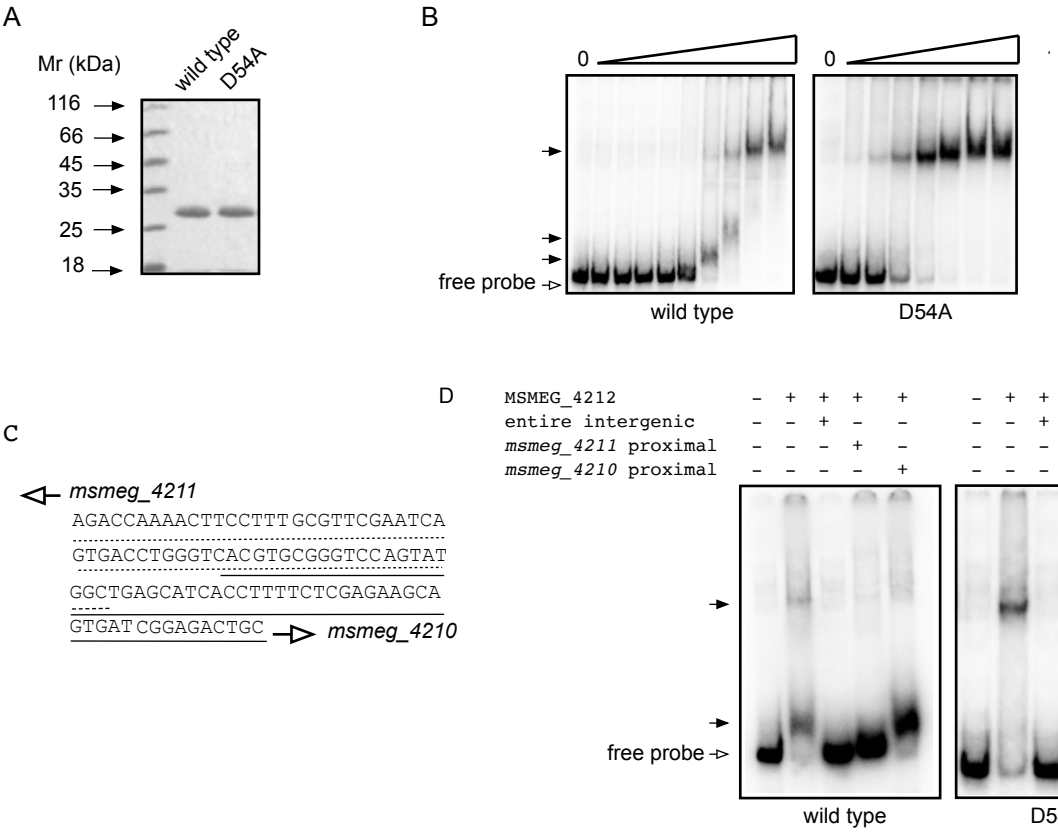


Figure 3

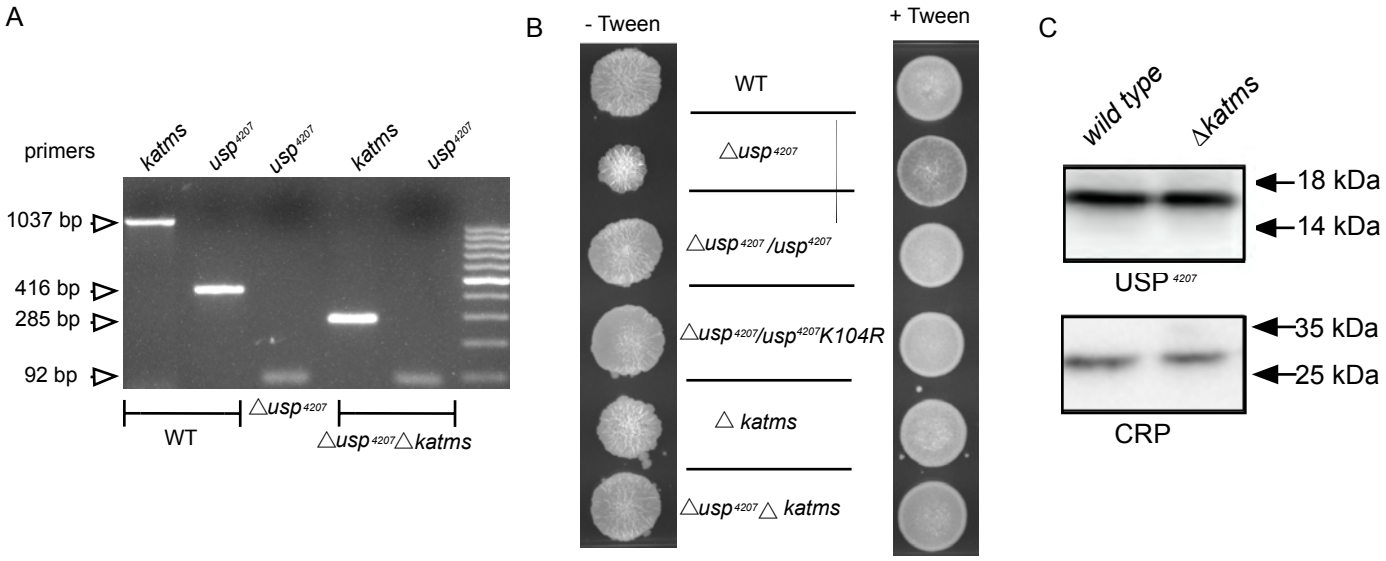
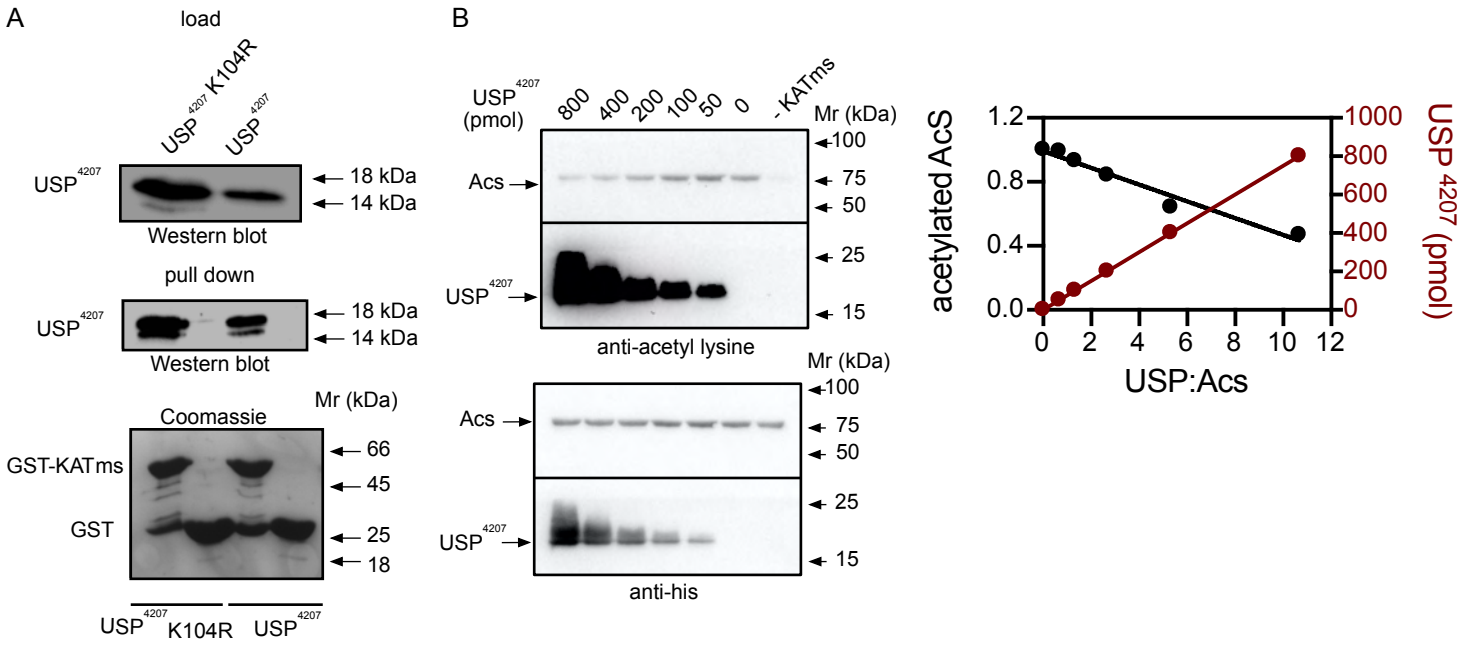


Figure 4



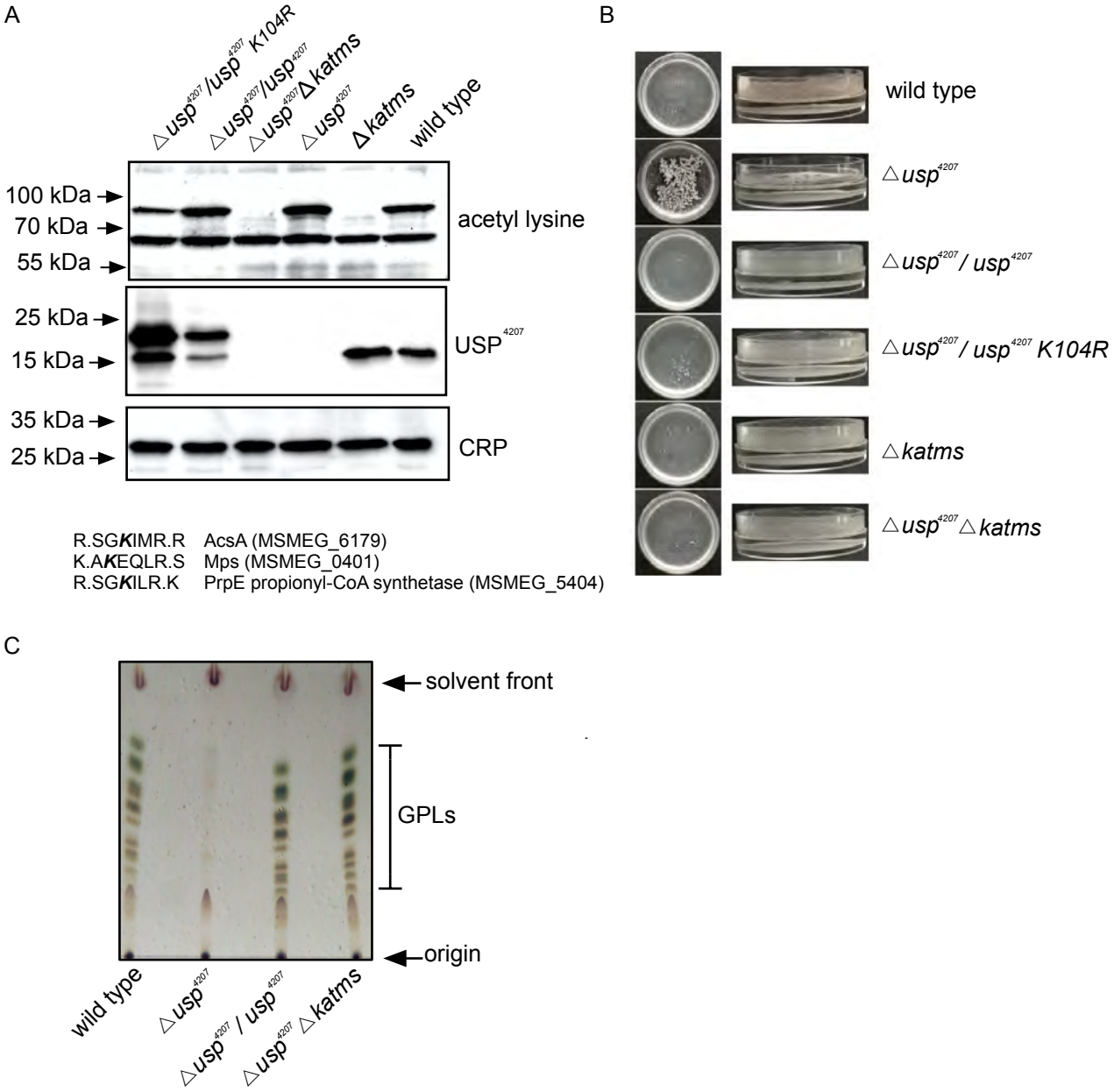


Figure 6

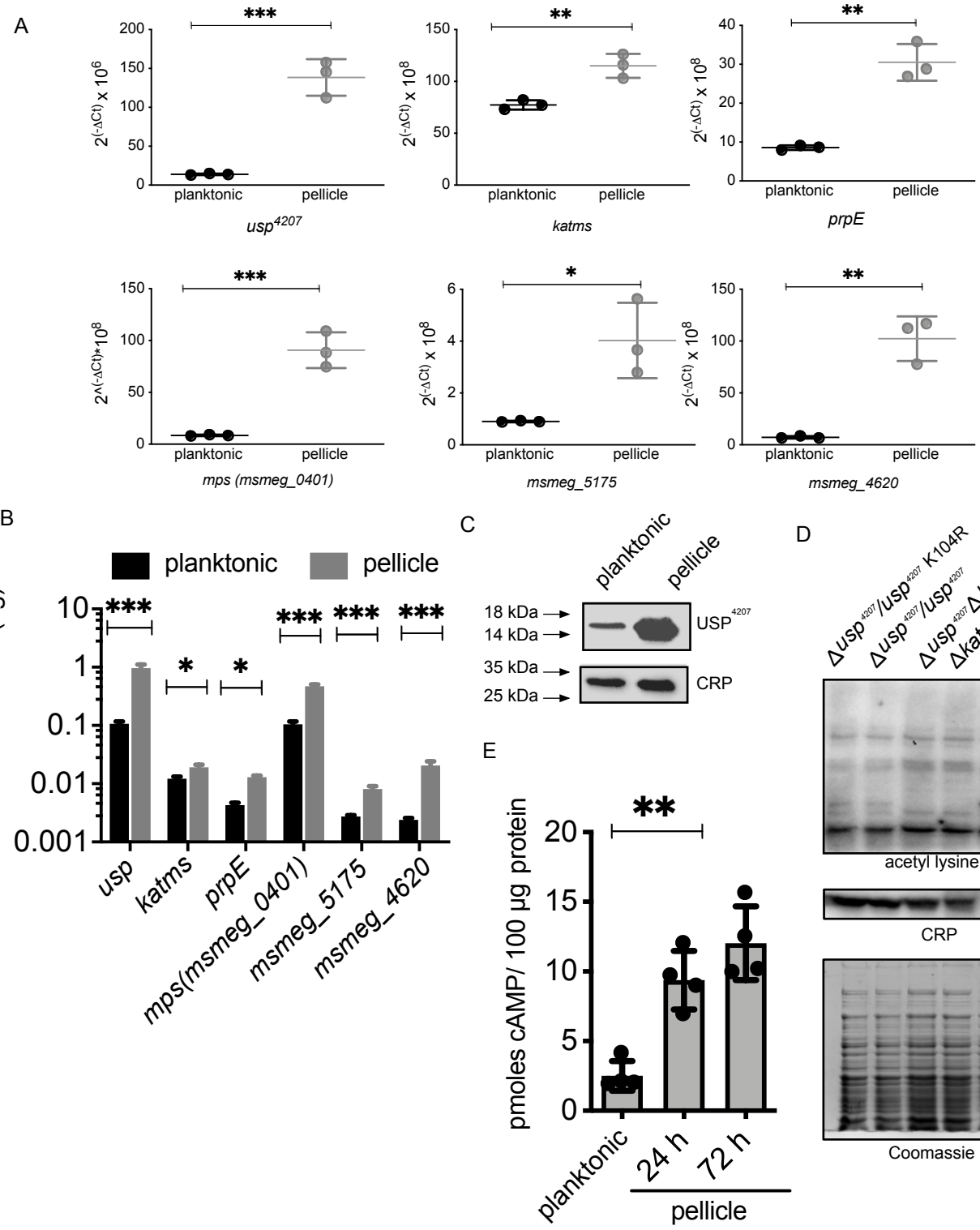


Figure 7

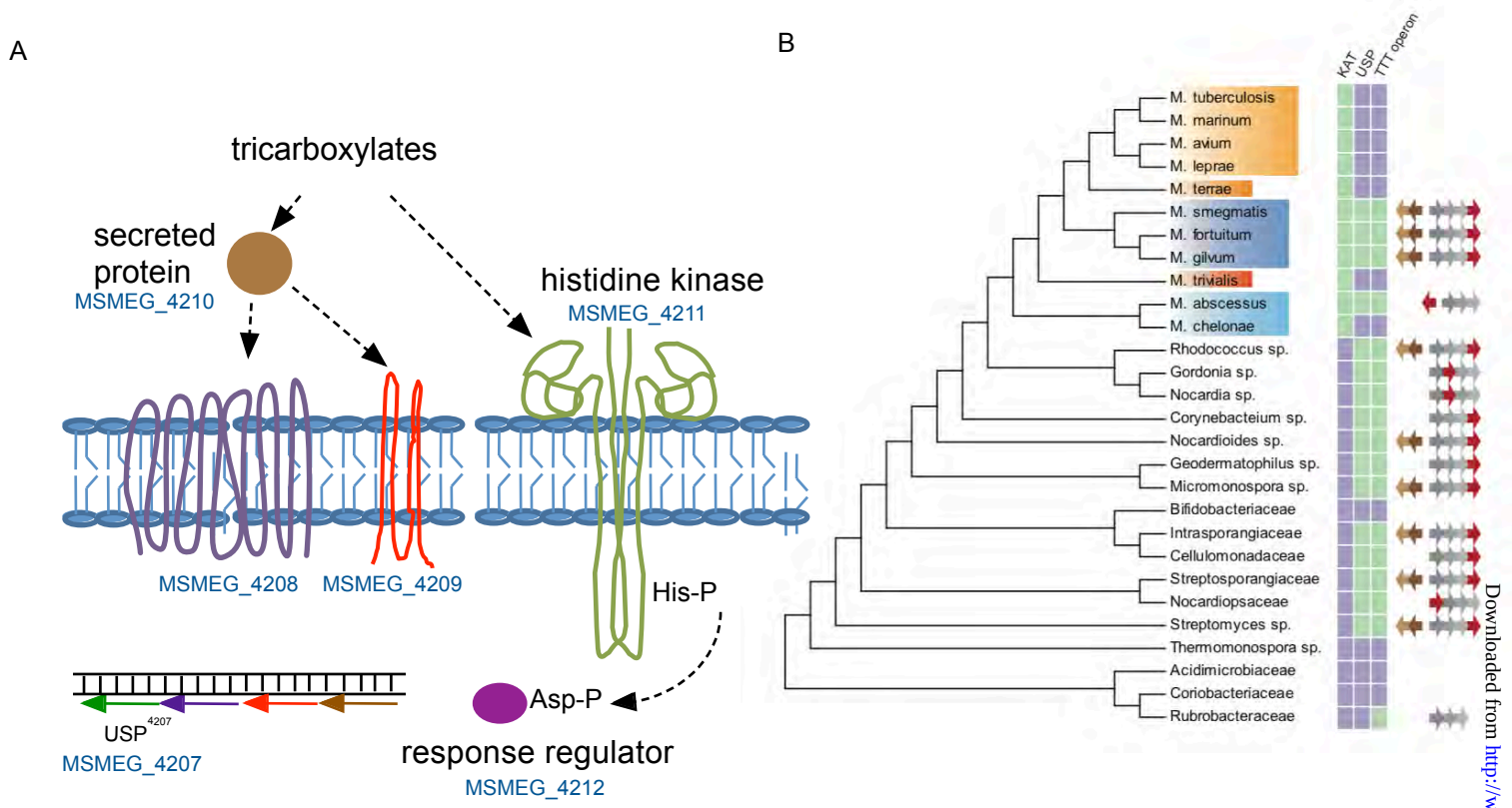


Figure 8

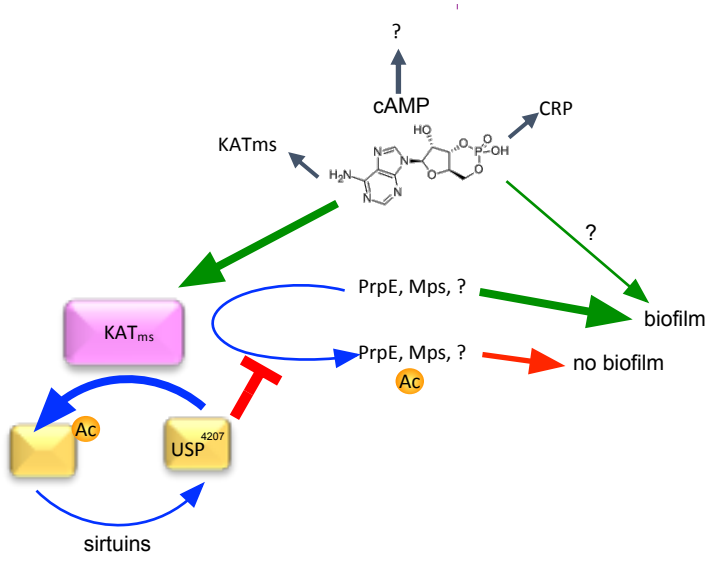


Figure 9

A universal stress protein in *Mycobacterium smegmatis* sequesters the cAMP-regulated lysine acyltransferase and is essential for biofilm formation

Sintu Samanta, Priyanka Biswas, Arka Banerjee, Avipsa Bose, Nida Siddiqui, Subhalaxmi Nambi, Deepak Kumar Saini and Sandhya S Visweswariah

J. Biol. Chem. published online December 27, 2019
