



HAL
open science

MmpS4 promotes glycopeptidolipids biosynthesis and export in *Mycobacterium smegmatis*

Caroline Deshayes, Horacio Bach, Daniel Euphrasie, Rodgoun Attarian, Mathieu Coureuil, Wladimir P Sougakoff, Françoise Laval, Yossef Av-Gay, Mamadou Daffe, Gilles Etienne, et al.

► **To cite this version:**

Caroline Deshayes, Horacio Bach, Daniel Euphrasie, Rodgoun Attarian, Mathieu Coureuil, et al.. MmpS4 promotes glycopeptidolipids biosynthesis and export in *Mycobacterium smegmatis*. *Molecular Microbiology*, 2010, 78 (4), pp.989. 10.1111/j.1365-2958.2010.07385.x . hal-00580703

HAL Id: hal-00580703

<https://hal.science/hal-00580703>

Submitted on 29 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MmpS4 promotes glycopeptidolipids biosynthesis and export in *Mycobacterium smegmatis*

Journal:	<i>Molecular Microbiology</i>
Manuscript ID:	MMI-2010-09792.R2
Manuscript Type:	Research Article
Date Submitted by the Author:	30-Aug-2010
Complete List of Authors:	<p>Deshayes, Caroline; University Paris-Descartes, School of Medecine, Inserm U570</p> <p>Bach, Horacio; University of British Columbia, Medicine</p> <p>Euphrasie, Daniel; INSERM, U570</p> <p>Attarian, Rodgoun; University of British Columbia, Medicine</p> <p>Coureuil, Mathieu; University Paris-Descartes, School of Medecine, Inserm U570</p> <p>Sougakoff, Wladimir; University Pierre et Marie Curie, Inserm UMRS-872</p> <p>Laval, Françoise; CNRS, IPBS</p> <p>Av-Gay, Yossef; University of British Columbia, Medicine</p> <p>Daffe, Mamadou; CNRS-University Paul Sabatier Mixed Research Lab (UMR 5089), Department of Molecular Mechanisms of Mycobacterial Infections</p> <p>Etienne, Gilles; CNRS-University Paul Sabatier Mixed Research Lab (UMR 5089), Department of Molecular Mechanisms of Mycobacterial Infections</p> <p>Reyrat, Jean-Marc; University Paris-Descartes, School of Medecine, Inserm U570</p>
Key Words:	mycobacteria, glycopeptidolipids, cell wall, membrane protein

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

MmpS4 promotes glycopeptidolipids biosynthesis and export in *Mycobacterium smegmatis*

Caroline Deshayes^{1, 2, 3,*}, Horacio Bach⁴, Daniel Euphrasie², Rodgoun Attarian⁴, Mathieu Coureuil^{1, 2}, Wladimir Sougakoff⁵, Françoise Laval^{6, 7}, Yossef Av-Gay⁴, Mamadou Daffé^{6, 7}, Gilles Etienne^{6, 7} and Jean-Marc Reyrat^{1, 2, §}

¹ INSERM-UMR 570, Unité de Pathogénie des Infections Systémiques, Paris Cedex 15, F-75730, France

² Université Paris Descartes, Faculté de Médecine, site Necker, Paris Cedex 15, F-75730, France

³ Groupe d'Etude des Interactions Hôte-Pathogène, Université d'Angers, 4 rue Larrey, Angers, France

⁴ Department of Medicine, Division of Infectious Diseases, University of British Columbia, 2733 Heather St., Vancouver, V5Z 3J5 British Columbia, Canada

⁵ INSERM UMRS-872, équipe 12, Laboratoire de Recherche Moléculaire sur les Antibiotiques (LRMA), faculté de Médecine Pierre et Marie Curie, site Pitié-Salpêtrière, 91 bd de l'Hôpital, F-75634 Paris cedex 13, France

⁶ CNRS, IPBS (Institut de Pharmacologie et de Biologie Structurale), Département Mécanismes Moléculaires des Infections Mycobactériennes, 205 route de Narbonne, F-31077 Toulouse, France

⁷ Université de Toulouse, Université Paul Sabatier (Toulouse III), IPBS, F-31077 Toulouse, France

***: corresponding author.** Dr Deshayes, C. Groupe d'Etude des Interactions Hôte-Pathogène, Université d'Angers, 4 rue Larrey, Angers, France. deshayes@necker.fr

§: this paper is dedicated to Dr Jean-Marc Reyrat, who deceased before the submission of the manuscript, for his excellent contribution to the mycobacterial genetics.

Key words: mycobacteria, glycopeptidolipids, cell wall, membrane protein

1 **Summary.** The MmpS family (mycobacterial membrane protein small) includes over one hundred
2 small membrane proteins specific to the genus *Mycobacterium* that have not yet been studied
3 experimentally. The genes encoding MmpS proteins are often associated with *mmpL* genes, which are
4 homologous to the RND (resistance nodulation cell division) genes of Gram-negative bacteria that
5 encode proteins functioning as multidrug efflux system. We showed by molecular genetics and
6 biochemical analysis that MmpS4 in *M. smegmatis* is required for the production and export of large
7 amounts of cell surface glycolipids, but is dispensable for biosynthesis *per se*. A new specific and
8 sensitive method utilizing single chain antibodies against the surface-exposed glycolipids was
9 developed to confirm that MmpS4 was dispensable for transport to the surface. Orthologous
10 complementation demonstrated that the MmpS4 proteins are exchangeable, thus not specific to a
11 defined lipid species. MmpS4 function requires the formation of a protein complex at the pole of the
12 bacillus, which requires the extracytosolic C-terminal domain of MmpS4. We suggest that MmpS
13 proteins facilitate lipid biosynthesis by acting as a scaffold for coupled biosynthesis and transport
14 machinery.

1 Introduction

2 The number of complete genome sequences of high GC % Gram-positive bacteria available has
3 increased considerably in the recent years, leading to the accumulation of unprecedented amounts of
4 information. For example, 16 mycobacterial genomes have been sequenced to date and the species
5 sequenced include both rapid and slow growers (Bernal *et al.*, 2001). *Mycobacterium* is a large genus
6 containing over 150 species, including *M. tuberculosis*, the etiologic agent of tuberculosis (WHO), *M.*
7 *ulcerans*, which causes a devastating necrotic disease of the skin, the so-called Buruli ulcer (WHO),
8 and many opportunistic pathogens (De Groote & Huitt, 2006). Recent sequencing efforts have,
9 through comparative and functional genomics, led to advancements in the characterization of the
10 mycobacterial core genome and of genes involved in virulence. For example, genes encoding a new
11 type of secretion system were characterized, the type VII (Abdallah *et al.*, 2007), in addition to genes
12 required for the biosynthesis of arabinogalactan, a specific attribute restricted to this genus (Crick &
13 Brennan, 2008) and related genera. Sequencing of the mycobacterial genome has also led to the
14 identification of the MmpSL family of membrane proteins, the function of which remains unclear. The
15 name of this protein family is short for mycobacterial membrane protein large (MmpL) and small
16 (MmpS).

17 The genes encoding MmpS are often found closely associated with *mmpL* genes and nothing is
18 currently known about the function of these proteins. MmpL proteins consist of approximately 1000
19 amino acids organized into twelve predicted transmembrane domains and two large periplasmic
20 domains. MmpL are not restricted to mycobacteria, and are also found in related streptomycetes and
21 *Rhodococcus*. This family of proteins is distantly related to the resistance nodulation cell division
22 (RND) superfamily (Tekaia *et al.*, 1999), members of which are present in all domains of life and may
23 serve as efflux pumps for toxic environmental molecules (Tseng *et al.*, 1999). Two of the fourteen
24 MmpL members in *M. tuberculosis* have been shown to be involved in the biosynthesis and/or
25 transport of mycobacterial lipid metabolites. MmpL7 is required together with an ABC transporter and
26 a lipoprotein, for the specific transport of phthiocerol dimycoserate (PDIM) to the cell surface
27 (Camacho *et al.*, 2001, Cox *et al.*, 1999). MmpL8 plays a role in the synthesis of sulfolipid-1 (SL-1),
28 another cell surface polyketide, possibly in the transport of a SL-1 precursor from the cytoplasm to the

1 periplasm (Converse *et al.*, 2003, Domenech *et al.*, 2004). In *M. smegmatis* and *M. abscessus*, the
2 MmpL4a and MmpL4b proteins (previously known as TmtpB and TmtpC, respectively) are
3 implicated in the biosynthesis of cell surface polyketides, the glycopeptidolipids (GPLs) (Recht *et al.*,
4 2000, Sonden *et al.*, 2005, Medjahed & Reyrat, 2009).

5 As previously mentioned, some *mmpL* genes are associated with *mmpS* genes, which encode
6 proteins predicted to have only one N-terminal transmembrane domain with an extracytoplasmic C-
7 terminus (Domenech *et al.*, 2005). Two (*M. leprae*) to twenty-seven (*M. abscessus*) such genes may be
8 present in the genome, depending on the species. Most *mmpL* and *mmpS* genes are located close to
9 genes involved in the synthesis or modification of polyketides. This close proximity suggests a
10 possible role in the transport and/or biosynthesis of molecules synthesized by the neighbouring
11 synthase gene products (Tekaiia *et al.*, 1999). Little is known about MmpS function, but transcriptional
12 regulation of these genes has been observed. For example, the *mmpS4* and *mmpS5* genes of *M.*
13 *tuberculosis* are down regulated during nutrient starvation (Betts *et al.*, 2002). More recently,
14 resistance to azole in a *M. tuberculosis* mutant strain was correlated with increased transcription of
15 *mmpS5-mmpL5* genes (Milano *et al.*, 2009).

16 In *M. smegmatis*, *mmpS4* is organized into a putative operon with the *mmpL4a* and *mmpL4b*
17 genes, which have been shown to be involved in GPL biosynthesis (Recht *et al.*, 2000, Sonden *et al.*,
18 2005). GPLs are the predominant glycolipids found at the surface of many non-tuberculous
19 mycobacteria, including *M. avium* subsp. *avium* (Brennan *et al.*, 1981) and *M. abscessus* (Lopez-
20 Marin *et al.*, 1994), an emerging pathogen predominantly infecting young cystic fibrosis patients.
21 These surface polyketides are required for sliding motility and biofilm formation, and in some cases
22 their level of production is correlated with strain virulence (Byrd & Lyons, 1999, Howard *et al.*,
23 2006). Upon engineering an in-frame unmarked deletion of the *mmpS4* gene in *M. smegmatis*, reduced
24 levels of GPLs were observed. However, we found that MmpS4 is not required for GPLs localization.
25 This finding was confirmed with a new immunofluorescent method for specific detection of GPLs at
26 the cell surface. Our results suggest that MmpS are not specific for structurally defined lipid
27 molecules. By studying fusions of red fluorescent protein (Rfp) with MmpS4 or two GPL biosynthesis
28 enzymes, we show that MmpS4 is required for the formation of a protein complex at the pole of the

1 bacillus, presumably *via* the C-terminal domain. Thus, we suggest that MmpS proteins promote
2 efficient biosynthesis and secretion by acting as a scaffold for the enzymatic and export machinery.
3 This report constitutes the first experimental investigation of a member of the MmpSL family of
4 proteins.

For Peer Review

1 Results

2 **The *mmpS* family in mycobacteria.** *mmpS* genes are specific to mycobacteria and are most often
3 associated with *mmpL* genes. Indeed, our homology searches in genomes of closely related members
4 of the *Actinomycetales*, such as *Streptomyces* and *Rhodococcus*, show that these bacteria contain
5 *mmpL* homologues, but no *mmpS* genes (Table 1). The MmpS family is currently composed of more
6 than one hundred proteins and new members are added with the release of each new mycobacterial
7 genome sequence. The number of *mmpS* paralogs per genome varies from two to twenty seven and is
8 correlated with the number of *mmpL* genes, rather than genome size (Table 1). These members can be
9 organized into five classes, based on the clusters obtained by phylogenetic analysis, corresponding to
10 the *M. tuberculosis mmpS* paralog-numbering (Fig. 1A).

11 MmpS proteins have an overall identity of 22 to 81 % within the same class and 8 to 37 % between
12 members of different classes. Each *mmpS* gene encodes a predicted protein of approximately 150
13 amino acids in length, anchored in the membrane by a 15-amino-acid-transmembrane domain at its N-
14 terminus. It has been suggested that the C-terminus of the protein is extracytosolic (Tekaiia et al., 1999,
15 Domenech et al., 2005) and this hypothesis has been supported by recent computer-assisted predicting
16 of membrane protein topology (Fig. 1B). An alkaline phosphatase gene fusion *mmpS4-phoA* was used
17 to confirm this predicted extracytosolic localization. The activity of bacterial alkaline phosphatase
18 (PhoA) is indeed dependent on it being exported across the plasma membrane. *M. smegmatis*
19 expressing the *mmpS4-phoA* fusion showed blue phenotype on plates containing X-P, the chromogenic
20 substrate of PhoA, reflecting an activity of the alkaline phosphatase (Fig. 1C). In contrast, *M.*
21 *smegmatis* carrying the empty pJEM11 plasmid was white, clearly confirming the bioinformatical
22 prediction and indicating that the C-terminal domain of MmpS4 is extracytoplasmic.

23 The *mmpS3* genes differ from other *mmpS* genes in encoding predicted proteins of approximately 300
24 amino acids in size. These proteins are thus twice the size of other MmpSs proteins and likely arose
25 from an ancient complete *mmpS3* copy through an event of duplication and fusion. *mmpS3* is present
26 in both rapid and slow growing mycobacterial species, suggesting that the duplication event occurred
27 in the common ancestor of these two phylogenetic clades. *M. smegmatis* has ten *mmpS* genes (Table
28 1), and is specifically enriched in paralogs of the *mmpS5* families (Table S1). By contrast, the *mmpS4*

1 gene is present as only a single copy in all the sequenced mycobacterial genomes, with the exception
2 of *M. abscessus* genome, which includes five paralogs.

3

4 **MmpS4 is required for sliding motility and biofilm formation.** We investigated the function of
5 MmpS4 by generating a mutant strain for this gene in *M. smegmatis*. The *mmpS4*, *mmpL4a* and
6 *mmpL4b* genes are probably transcribed as an operon, as these genes either overlap or are separated by
7 only few nucleotides. We prevented possible polar effects on downstream genes by constructing an in
8 frame unmarked deletion of the *mmpS4* gene by homologous recombination, using a two steps strategy
9 and a counter-selectable marker (Pelicic *et al.*, 1996, Reyrat *et al.*, 1998). The mutant strain was
10 verified by PCR, sequencing and Southern-blotting, all of which confirmed its identity and the
11 occurrence of the anticipated event, a deletion of about 90% of the coding sequence of the *mmpS4*
12 gene (Fig. S1). The resulting $\Delta mmpS4$ mutant had rough colony morphology readily visible by the
13 naked eye. This mutant was unable to slide (Fig. 2A) or form biofilms (data not shown).
14 Complementation of this mutant strain with the wild-type *mmpS4* allele from *M. smegmatis* restores
15 the wild-type smooth phenotype and the ability to slide (Fig. 2A) and form biofilms. This excluded the
16 possibility that the observed phenotype was due to secondary mutations or a polar effect on
17 downstream genes.

18 In *M. smegmatis*, changes in sliding motility and biofilm formation are often associated with
19 quantitative or qualitative changes in GPL biosynthesis, due to the lack of biochemical modifications
20 by acetylation, methylation or glycosylation of the GPL backbone for instance (Deshayes *et al.*, 2005,
21 Recht & Kolter, 2001, Sonden *et al.*, 2005). We therefore analyzed GPL production in the mutant
22 strain. As shown in Figure 2B, the $\Delta mmpS4$ strain produced small amounts of GPL-like compounds.
23 Only the major species remained detectable by TLC, whereas the complemented strain produced GPL
24 compounds in amounts similar of those of the wild-type. For the precise quantification of GPL
25 production in the $\Delta mmpS4$ strain, we grew large amounts of cells and purified GPLs, making use of
26 their alkali-stable character (Belisle *et al.*, 1993). In these conditions the *mmpS4* mutant produced
27 about 20% as much alkali-stable lipids as the wild-type and complemented strains did, accounting for

1 the rough morphotype of the mutant strain (Fig. 2C). Complementation of the mutant strain with the
2 *mmpS4* wild-type allele of *M. smegmatis* restored the wild-type level of GPL production (Fig. 2B and
3 2C). The levels of other lipids (triacyl glycerols, glycerol monomycolate, trehalose dimycolate,
4 trehalose monomycolate, phosphatidylinositol mannosides) were similar (Fig. S2). Thus MmpS4 plays
5 a specific role in controlling the level of GPLs biosynthesis in *M. smegmatis*.

6

7 **MmpS4 is required for high level of GPL production.** We checked for changes in GPL structure, by
8 analyzing lipid extracts of the $\Delta mmpS4$ strain using MALDI-TOF mass spectrometry, a highly
9 sensitive and accurate method of mycobacterial complex glycolipids detection and characterization
10 (Villeneuve *et al.*, 2003). Total lipid extract prepared from the wild-type strain displayed a major peak
11 at 1257 *m/z*, corresponding to the main diglycosylated (tri-*O*-Methyl Rhamnose and di-*O*-Acetylated
12 6-deoxyTalose) GPL of *M. smegmatis* mc²155 (Etienne *et al.*, 2005). A similar pattern of GPLs was
13 also detected in the mutant strain, but in relatively smaller amounts and with the peaks 14 Da higher
14 than those of the wild-type (Fig. S3). The major peak was observed at 1271 *m/z*. This 14 atomic mass-
15 units increase is likely due to an additive *O*-methylation on the fatty acid moiety, as the only spot
16 detectable by TLC (Fig. 2B) was the most apolar fully *O*-methylated GPL I (Patterson *et al.*, 2000).
17 Complementation of the $\Delta mmpS4$ strain fully restored the wild-type profile with a major peak at 1257
18 *m/z*, demonstrating that this 14 atomic mass-units increase was dependent on the absence of the
19 *mmpS4* gene.

20

21 **MmpS4 is not involved in GPL subcellular localization.** We previously showed that sliding motility
22 requires not only the normal level of GPL production but also a localization of these molecules on the
23 cell surface (Sonden *et al.*, 2005). As the $\Delta mmpS4$ strain is not motile, we hypothesized that MmpS4
24 might be involved in GPL export to the cell surface. We tested this hypothesis, by preparing surface
25 and cell-associated (after removal of the surface lipids with glass beads) fractions, extracting lipids
26 from each fraction and analyzing them by TLC (Fig. 3A). As the $\Delta mmpS4$ strain produced less than
27 20% the amount of GPL produced by the wild-type strain, we used larger amounts of cells and

1 concentrated lipid metabolites through solvent evaporation to visualize them by TLC. A Δgap mutant
2 strain was used as a control because this mutant produces wild-type levels of GPLs but does not export
3 them to the surface (Sonden *et al.*, 2005). Consistent with quantification results, the $\Delta mmpS4$ strain
4 produced much less GPL than did the wild-type and Δgap strains, but small amounts of GPL were still
5 present in the surface fraction of the mutant strain. Despite the low level of production of GPL by the
6 mutant, the glycolipid could be visualized in the surface compartment, although barely, for the
7 $\Delta mmpS4$ mutant but not for Δgap strain, demonstrating the presence of GPL at the surface of the
8 $\Delta mmpS4$ mutant (Fig. 3A). An overloaded TLC showed selective disappearance in the cellular extract
9 of the lower spots, corresponding to polar GPLs in the $\Delta mmpS4$ mutant. Mass spectrometry analyses
10 unambiguously confirmed the presence of GPLs in the cell surface of the $\Delta mmpS4$ strain while GPLs
11 were not detectable on the cell surface of the Δgap mutant (Fig. S3).

12
13 **Development of a new sensitive immunofluorescent method for lipid localization.** To further
14 investigate the *in vivo* localization of GPLs, we developed a new and specific method. GPL
15 localization in the $\Delta mmpS4$ mutant was characterized by an immunofluorescent method that
16 selectively detected surface-exposed GPLs. Briefly, purified GPLs coupled to ovalbumin were used as
17 antigens to screen a naïve human-derived antibody library using phage display technology. After four
18 biopanning cycles and screening against ovalbumin, four clones with a high affinity for GPLs were
19 selected. The gene encoding the antibody with the highest affinity was fused to the gene encoding the
20 DSred fluorescent protein. The construct was expressed in *E. coli*, and the fused protein fully retained
21 the fluorescent properties of the DSred as observed by epifluorescence microscopy (Figure 3B). The
22 detection of GPLs was performed in a single step as secondary antibodies were not needed. Bacteria
23 were first non-specifically labelled with FITC. The resultant green staining made it possible to
24 visualize the bacterial bodies and served as a control for evaluating of the number of bacteria present
25 on the slide. Two control strains were used in this experiment: a *pks* mutant strain producing no GPLs
26 (*pks*⁻) (Sonden *et al.*, 2005) used to evaluate labelling specificity, and the Δgap mutant strain with its
27 complemented companion strain (Sonden *et al.*, 2005). The DSred-antibody labelling perfectly co-

1 localized with the FITC-labelled bacteria in the case of the wild-type and the complemented Δgap
2 strain, resulting in a yellow staining after image overlapping (Fig. 3B). The strain producing no GPL
3 and the Δgap strain were not immunolabelled with the anti-GPL antibody. This antibody was therefore
4 highly specific for GPL and able to bind surface-exposed GPLs only, as the Δgap strain, which
5 produces wild-type level but does not export them on the surface, was not labelled by the antibody.
6 Interestingly, a weak anti-GPL fluorescence signal was detected in the $\Delta mmpS4$ strain, corresponding
7 to approximately 60% of the relative fluorescence compared to the wild-type strain. A clear signal was
8 observed for the $\Delta mmpS4$ strain but not for the GPL- and Δgap strains. These results confirm that the
9 $\Delta mmpS4$ strain produces smaller amounts of GPL than the wild-type strain but also demonstrate that
10 GPLs are present on the bacterial surface. The $\Delta mmpS4$ complemented strain gave a stronger signal,
11 although not as strong as that of the wild-type strain, in approximately 95% of the bacteria. These
12 experiments demonstrate that GPLs reach the cell surface in the $\Delta mmpS4$ strain but also show that that
13 reintroduction of the *mmpS4* gene increases both the total and surface levels of GPLs.

14
15 **MmpS4 proteins assemble into a complex at the pole of the bacillus.** We investigated the function
16 and location of MmpS4 by constructing gene fusions based on the red fluorescent protein (RFP). The
17 *rfp* gene was fused to the 5' or the 3' end of the *mmpS4* gene, yielding RFP-MmpS4 and MmpS4-RFP
18 proteins, respectively. The fusion of RFP to the N-terminal domain did not affect the function of
19 MmpS4. Indeed, the construct encoding RFP-MmpS4 fully complemented the $\square mmpS4$ mutant strain,
20 resulting in the production of large amounts of GPLs (data not shown). In contrast, the fusion of RFP
21 to the C-terminal domain of MmpS4 abolished its function, as this construct did not complement the
22 $\square mmpS4$ mutant strain. This result suggests that the extracytosolic C-terminal domain of MmpS4 is
23 crucial for its function and therefore does not tolerate fusion with RFP. Nevertheless, it is still possible
24 that the formation of a fused protein may lead to misfolding of MmpS4 and this in turn affects its
25 topology and membrane localization. To dissect the role of the MmpS4 extracytosolic C-terminal
26 domain, truncated proteins were constructed. None of them were able to complement the $\Delta mmpS4$
27 strain (Fig. S4), strengthening our hypothesis that the C-terminal domain of MmpS4 is necessary for

1 its function. Furthermore, this domain needs to be localized in the extracytosolic compartment. Indeed
2 the intracellular expression of this domain could not restore the *mmpS4* function, since the MmpS4
3 protein deleted of its transmembrane region was not able to complement the mutant strain.

4 The strains expressing fusion proteins were observed by fluorescence microscopy and a
5 control strain expressing cytosolic GFP was used (Barker *et al.*, 1998). In strains expressing MmpS4-
6 RFP or the cytosolic GFP, the fluorescence was diffused and spread throughout the bacillus. In
7 contrast, in the strain carrying the RFP-MmpS4 construct, the fluorescence was concentrated into
8 small dot-like area (spot). This punctuate fluorescence was absent from strains expressing the MmpS4-
9 RFP fusion and the cytosolic GFP, but present in approximately 25% of the bacilli expressing the
10 RFP-MmpS4 fusion (Fig. 4A). There was never more than one spot per bacillus and the spots were
11 located either at the site of division or at the pole of the cell (Fig. 4B). The presence of this complex at
12 the division site may enable each of the two daughter cells to inherit half the complex, therefore
13 becoming located at the pole after division, or can be correlated to the presence of the division
14 machinery. Although we can not completely exclude that this apparent MmpS4 localization was an
15 artefact of the RFP fusion, this seems unlikely since a polar or septum localization is rather expected
16 for a complex involved in the synthesis of an envelope constituent. In conclusion, these findings
17 strongly suggest that MmpS4 proteins are not randomly distributed in the membrane but instead
18 assemble into a complex located at the pole of the bacterium. A free C-terminal extracytosolic domain
19 is required for both complex formation and biological function.

20

21 **MmpS4 proteins are necessary for the formation of a complex of GPL biosynthesis enzymes at**
22 **the pole of the bacillus.** Based on the data presented herein and previous studies reporting interaction
23 between MmpL and enzymes required for the biosynthesis of small metabolites (Cox *et al.*, 1999,
24 Straight *et al.*, 2007), we suggest that MmpS4 acts as a scaffold for the assembly of the GPL
25 biosynthesis megacomplex (Fig. 5A). Proteins fusions were constructed, fusing Rfp in the C-terminal
26 domain of two GPL synthesis proteins (MbtH and FadD23) predicted to be cytosolic. Their
27 localization was analyzed in the wild-type and $\Delta mmpS4$ mutant strains. Punctuate fluorescence was
28 found at the poles only in the wild-type strain in roughly 25% and 35% of the bacilli for the MbtH-Rfp

1 and FadD23-RFP constructs, respectively (Fig. 5B and C). These percentages were similar to the
2 percentage obtained for the strain expressing the *rfp-mmpS4* fusion. In contrast, the fluorescence was
3 mainly diffuse in the $\Delta mmpS4$ mutant strain. These results show that the absence of MmpS4 proteins
4 affects the localization of other proteins involved in GPL synthesis and reinforce our megacomplex
5 model.

6
7 **MmpS4s are functionally exchangeable.** MmpS4 orthologs are 50 to 60% identical and many of the
8 conserved amino acids are located in the predicted extracytosolic region (Fig. S5). We assessed the
9 activity of these MmpS4 orthologs, by transforming the $\Delta mmpS4$ strain with a plasmid carrying an
10 ortholog of the *mmpS4* gene originating from other species (genes from several different species were
11 used). For *M. abscessus*, which carries five paralogs of *mmpS4*, we used the copy (MAB4117c) most
12 similar to *mmpS4* of *M. smegmatis*. This gene mapped to the GPL locus of *M. abscessus* (Ripoll *et al.*,
13 2007). We used orthologs from GPL-producing species such as *M. abscessus* and *M. avium* subsp.
14 *avium* (MAV3247), but also from non-GPL-producing organisms such as *M. tuberculosis* (Rv0451c).
15 In all cases (MAB4117c, MAV3247, and Rv0451c) complementation with the *mmpS4* gene restored a
16 smooth wild-type phenotype. The mutant strain complemented with the *M. tuberculosis* ortholog
17 exhibited wild-type sliding motility, consistent with functional complementation (Fig. 6A). As a
18 control experiment we carried out complementation experiments with the *mmpS1* gene of *M.*
19 *tuberculosis* (*mmpS1Tb*). Expression of this *mmpS1Tb* gene did neither restore the wild-type smooth
20 phenotype, nor the sliding motility, consistent with the notion of orthology for the MmpS4 family
21 (Fig. 6A). To avoid any possibility of non-expression of the *mmpS1* gene, its expression in the
22 $\Delta mmpS4/mmpS1Tb$ strain was confirmed by reverse transcription (Fig. S6). Biochemical analyses of
23 the $\Delta mmpS4$ mutant strain expressing the *mmpS4* gene of *M. tuberculosis* (*mmpS4Tb*) showed fully
24 functional complementation. The *mmpS4* gene of *M. tuberculosis* therefore conferred to the mutant
25 strain, upon complementation, the ability to produce more GPL, both in the cellular compartment and
26 at the surface (Fig. 6B). Quantitative analyses showed that the *mmpS4Tb*-complemented strain
27 produced significantly larger amounts of GPLs (2.09 ± 0.72 mg of alkali-stable lipids for 100 mg of

1 cells, dry weight) than did the $\Delta mmpS4$ mutant strain (0.54 ± 0.33 mg). Furthermore,
2 immunofluorescence experiments with the anti-GPL DSred antibody showed much more GPLs on the
3 surface of the *mmpS4*Tb complemented strain (Fig. 6C) than on the $\Delta mmpS4$ mutant strain. Indeed,
4 the *mmpS4*Tb complemented strain produced a stronger signal in approximately 95% of the bacilli.
5 The signal obtained was similar to that for the *mmpS4*Ms complemented strain (Fig. 3B). As stated
6 above, this complemented strain displayed greater sliding motility compared to the mutant, indicating
7 a gain of function (Fig. 6A). In conclusion, these findings indicate that the *mmpS4* gene of a species
8 that does not produce GPL can complement the *M. smegmatis* $\Delta mmpS4$ mutant strain, both in terms of
9 production and localization of GPLs. Thus, MmpS4 are functionally exchangeable and constitute a
10 true protein family. Our findings also suggest that MmpS proteins are not specific to the lipid
11 compound itself.

1 Discussion

2 The synthesis and export of complex polyketides present in the outermost structure of the cell
3 envelope such as GPLs require the expression of large biosynthetic pathways, involving proteins
4 encoded by grouped genes or genes dispersed throughout the genome, and the assembly of dedicated
5 machineries (Gokhale *et al.*, 2007). The *mmpS* family was first identified a decade ago with the
6 sequencing of the first mycobacterial genome (Cole *et al.*, 1998). These genes encode proteins
7 involved in various specific pathways and are often associated with *mmpL* genes. *M. tuberculosis* has
8 five *mmpS* genes and fourteen predicted *mmpL* genes. Consequently, not all biosynthetic pathways
9 include proteins encoded by *mmpS* gene. This seems to be the case for PDIM and SL-1 loci for
10 example (Camacho *et al.*, 2001, Converse *et al.*, 2003). The location of the *mmpS4* gene within the
11 GPL biosynthesis locus suggests a role in GPL biosynthesis and/or transport to the surface, but our
12 data do not support this hypothesis. Indeed, the $\Delta mmpS4$ mutant strain produced smaller amounts of
13 GPLs than did the wild-type, but these compounds were still targeted to the bacterial surface as shown
14 by biochemical fractionation and the new immunofluorescence sensitive technique for the specific
15 tracking of GPLs at the bacterial surface. The low level of production of GPL in the $\Delta mmpS4$ strain
16 accounts for its phenotype, lack of sliding motility and inability to form biofilms, as all these features
17 are dependant on GPL production (Recht *et al.*, 2000, Kocincova *et al.*, 2008). However, although
18 produced in smaller amounts, the GPLs present in the mutant were chemically identical to those in the
19 wild-type, excluding a possible role of MmpS4 in the synthesis, *per se*. The MmpS4 protein of *M.*
20 *smegmatis* therefore seems to determine the amount of GPL produced. An alternative explanation
21 would be that there was another protein, maybe another MmpS protein that provides a similar function
22 in *M. smegmatis*, albeit not as efficient as MmpS4. The fact that functional complementation was
23 observed when an orthologous gene (*mmpS4Tb*) from the same family was introduced into the
24 $\Delta mmpS4$ strain, but not with a paralogous gene (*mmpS1Tb*), makes this last hypothesis unlikely, since
25 only one *mmpS4* gene is present in the *M. smegmatis* genome. The functional complementation with
26 an ortholog from *M. tuberculosis* was particularly surprising, as this species does not produce GPL. In
27 *M. tuberculosis*, *mmpS4* is organized into an apparent operon with a single *mmpL4* gene and this

1 operon is located close to genes encoding a probable cyclopropane fatty acyl phospholipid synthase
2 and a putative enoyl-CoA hydratase. The lipid metabolite specified by this pathway has not yet been
3 characterized, but a *M. tuberculosis* *mmpL4* mutant has been shown to be strongly attenuated *in vivo*
4 (Domenech et al., 2005). One possible explanation for this orthologous complementation is that
5 MmpS proteins are not specific for their lipid substrate but are, instead, specific for their cognate
6 MmpL proteins, via molecular interactions involving amino acids conserved throughout the family
7 (Fig. S5).

8 The MmpL members are distantly related to the RND superfamily, but differ from members of
9 this superfamily in several ways. For example, RND proteins have broad substrate specificity, whereas
10 MmpL proteins seem to be much more specific, interacting with only one lipid metabolite. MmpL7
11 and MmpL8 affect specifically PDIM and SL-1 biosynthesis in *M. tuberculosis*, respectively (Cox *et al.*,
12 1999, Converse *et al.*, 2003), and MmpL4 affects GPL biosynthesis in *M. smegmatis* (Recht *et al.*,
13 2000). Furthermore, whereas RND family members from Gram-negative bacteria act purely as efflux
14 pumps, the RND protein ActII-orf3 from the Gram-positive *Streptomyces coelicolor*, is involved in the
15 biosynthesis of a blue pigment, γ -actinorhodin, a heterocyclic antibiotic synthesized by a polyketide
16 synthase (Bystrykh *et al.*, 1996). In *M. tuberculosis*, MmpL8 plays a role in the synthesis of sulfolipid-
17 1 (SL-1), possibly by transporting a precursor of this molecule to the surface (Converse *et al.*, 2003,
18 Domenech *et al.*, 2004). In the same species, MmpL7 has been shown to be required for the transport
19 of PDIM to the cell surface (Cox *et al.*, 1999, Camacho *et al.*, 2001). This transport process requires
20 another protein, DrrC, a component of the putative ABC transporter, DrrABC (Camacho *et al.*, 2001),
21 and secretion into the medium requires LppX, a lipoprotein (Sulzenbacher *et al.*, 2006). Studies based
22 on the two hybrid approach have shown that MmpL7 interacts with PpsE (Jain & Cox, 2005), which
23 interacts with TesA (Rao & Ranganathan, 2004), both of them being enzymes required for PDIM
24 synthesis. It has thus been suggested that MmpL7 interacts with the PDIM synthetic machinery to
25 form a complex that coordinately synthesizes and transports PDIM across the cell membrane (Jain &
26 Cox, 2005). In *Bacillus subtilis*, a polyketide synthase synthesizing an antibiotic, bacillaene, has been
27 shown to assemble into a single megacomplex at the membrane (Straight *et al.*, 2007). The authors
28 suggested that this organization is established *via* interaction with an unknown membrane protein.

1 Based on such models, we suggest that MmpS4 proteins act as a scaffold with MmpL4
2 proteins for the assembly of the GPL biosynthesis megacomplex (Fig. 5A). Several enzymes needed
3 for GPL synthesis/export are predicted to be membrane-bound or -inserted: Atf, a sugar
4 acetyltransferase, Gap, needed for GPL surface localization, MmpL4a and MmpL4b, whose precise
5 role is currently unknown. It is likely that these membrane proteins associate into a complex to
6 maximize GPL biosynthesis and export. In this model, MmpS4 interacts with MmpL4a, MmpL4b and
7 GPL biosynthetic enzymes such as the polyketide synthase (Pks, involved in the lipid moiety
8 synthesis), non-ribosomal protein synthetases (Nrp, involved in the tetrapeptide moiety synthesis),
9 acetyltransferase (Atf), methyltransferases (Mtf) or glycosyltransferases (Gtf), thereby increasing the
10 synthesis and export of GPLs. Our data are consistent with the C-terminal extracytosolic domain of
11 MmpS4 being essential for the function of this protein. The extracytosolic domain of MmpS4 could
12 specifically interact with the extracytosolic domains of MmpL4 proteins, explaining the lack of
13 specificity of MmpS4 towards the lipid moiety. MmpS4 is not randomly distributed throughout the
14 membrane. Instead, it is found principally at the pole of the bacterium. This suggests that MmpS4
15 assembles into a complex that interacts with the biosynthetic machinery. In the absence of MmpS, the
16 biosynthetic complex does not adopt the correct configuration and the enzymes are distributed
17 throughout the cytoplasm (Fig. 5A), resulting in lower levels of GPL production than the wild-type
18 strain. The absence of the scaffold would lead to inappropriate contacts with enzymes of the pathway,
19 leading to the artefactual detection of GPL₁₂₇₁ as the main product. To strengthen our megacomplex
20 model, we constructed fluorescent fusion proteins with two GPL biosynthesis enzymes, MbtH and
21 FadD23. The role of the *mbtH* genes, whose homologs are found associated with *nrp* genes, is unclear
22 to date but it has been shown that the *mbtH* gene is required for the chlorobiocin biosynthesis in
23 *Streptomyces coelicolor* (Wolpert *et al.*, 2007). The *FadD23* gene is a close homolog to the *M.*
24 *tuberculosis fadD28* gene involved in acyl transfer of mycocerosic acid (Cox *et al.*, 1999) and is
25 probably involved in the lipid attachment to the GPL tetrapeptide moiety. Our results showed that
26 MmpS4 is required for them to localize as a punctuate spot at the poles of the bacilli. Colocalization
27 studies should be carried out to prove interactions between GPL biosynthesis/export enzymes. The
28 megacomplex model is also supported by the finding that some enzymes predicted to be cytosolic and

1 involved in PDIM (PapA5, Mas, FadD28) or mycolactone (MlsA1, MlsB) biosynthesis are found in
2 the membrane fraction of *M. ulcerans* (Tafelmeyer *et al.*, 2008). It can be speculated that the
3 localization of these proteins is due to their anchoring to a biosynthetic membrane megacomplex.

4 The presence of several paralogs, as in *M. abscessus*, is consistent with the diversification of
5 lipid metabolism. The minimalist species *M. leprae* has only two *mmpS* members (*mmpS3* and
6 *mmpS4*), suggesting that other pathways are not required for a strict intracellular lifestyle. The new
7 sensitive immunofluorescent method developed for this study is clearly transposable to other lipid
8 molecules. The method is therefore of considerable interest, because many lipid molecules are poorly
9 immunogenic, making antibody-based detection difficult. This new method allows the sensitive and
10 specific tracking of a single lipid only and is clearly of potential use for biochemical purposes and for
11 investigating lipid trafficking during host cell infection. It is also potentially useful for diagnostic
12 purposes, as some lipids are species- or even strain-specific.

13 In conclusion, according to our current working model of MmpS function, these small
14 transmembrane proteins are not absolutely required for metabolite synthesis and export, but may
15 promote interactions between the various partners involved in the pathway, possibly *via* their
16 extracytosolic C-terminal domains, therefore stabilizing the complex and optimizing synthesis and
17 export. MmpS proteins display no similarity to membrane fusion proteins of Gram-negative bacteria.
18 However, the close association of *mmpS* and *mmpL* genes suggests that MmpS proteins may be the
19 functional homologs of MFPs. Further biophysical studies are required to analyze the assembly and
20 interactions required to stabilize this complex machinery consisting of a dozen enzymes.

1 **Experimental procedures**

2

3 **Bioinformatical analysis.** The NCBI website (<http://www.ncbi.nlm.nih.gov/BLAST/>) was searched
4 with the BLASTP program, using *M. tuberculosis* MmpS proteins as queries. The sequences of MmpS
5 proteins from the *M. tuberculosis* H37Rv, *M. bovis* AF2122/97, *M. ulcerans* Agy99 and *M. leprae* TN
6 genomes were inspected at the Pasteur GenoList website (<http://genolist.pasteur.fr/>). The sequences of
7 the MmpS proteins of *M. abscessus*, *M. avium subsp. avium* 104, *M. avium subsp. paratuberculosis* K-
8 10, *M. gilvum* PYR-GCK, *M. smegmatis* mc²155, *M. sp.* KMS and *M. vanbaalenii* PYR-1 were
9 downloaded from the NCBI website. Protein sequences were aligned using the MultAlin program
10 (Corpet, 1988), and a phylogenetic tree was inferred. Secondary structure prediction to identify
11 membrane domains was carried out with Sosui software (Hirokawa *et al.*, 1998) and HMMTOP
12 website (<http://www.enzim.hu/hmmtop/>).

13

14 **Bacterial strains, medium and growth.** *E. coli* DH5 α was used for plasmids propagation and was
15 grown in LB medium like *M. smegmatis* strain mc²155 (Snapper *et al.*, 1990). When required,
16 antibiotics were added to the medium at the following concentrations: hygromycin 50 μ g/ml (200
17 μ g/ml for *E. coli*) and kanamycin 25 μ g/ml. For the alkaline phosphatase assay, LB plates
18 supplemented with 25 μ g/ml kanamycin and 50 μ g/ml 5-bromo-4-chloro-3-indolyl phosphate (X-P), a
19 chromogenic alkaline phosphatase substrate, were used (Lim *et al.*, 1995).

20

21 **Disruption of the *mmpS4* gene of *M. smegmatis*.** We used an in-frame deletion to disrupt the *mmpS4*
22 gene by homologous recombination with the pPR27 vector to prevent polar effects (Pelicic *et al.*,
23 1996). The primers used to amplify the upstream and downstream regions for each gene are listed in
24 Table S2. The chromosomal DNA of clones obtained by these procedures was compared with the
25 wild-type DNA by PCR and Southern blotting. The mutant was named Δ *mmpS4*.

26

27 **Construction of the *mmpS4* expression plasmids.** The wild-type *mmpS4* gene coding sequence
28 (accession number AY439015) was amplified by PCR with *Pfu* Turbo DNA polymerase (Stratagene)

1 using *M. smegmatis* mc²155 genomic DNA as a template and the primers *mmpS4*_trans.5 and
2 *mmpS4*_trans.3 (Table S2). PCR products were purified with a Qiagen PCR purification kit, digested
3 with *Xba*I and inserted into the single *Xba*I site of the dephosphorylated integrative expression vector
4 pNIP40b to generate pNIP*mmpS4* (de Mendonca-Lima *et al.*, 2001). Enzymatic digestions were used
5 to select clones in which the *mmpS4* gene was inserted in the opposite orientation of the hygromycin
6 resistant gene. One clone was selected, sequenced and named Δ *mmpS4/mmps4Ms*. A similar strategy
7 was used to insert the *mmpS4* genes of *M. avium* (MAV3247), *M. abscessus* (MAB4117c), *M.*
8 *tuberculosis* (Rv0451c) and the *mmpS1* gene (Rv0403c) of *M. tuberculosis* in pNIP40b yielding
9 pNIP*mmpS4*Av, pNIP*mmpS4*Ab, pNIP*mmpS4*Tb, and pNIP*mmpS1*Tb respectively. These plasmids
10 were used to electroporate *M. smegmatis* Δ *mmpS4* and transformants were selected on hygromycin
11 and named Δ *mmpS4/mmpS4*Av, Δ *mmpS4/mmpS4*Ab, Δ *mmpS4/mmpS4*Tb and Δ *mmpS4/mmpS1*Tb
12 respectively. For the construction of plasmid expressing the *mmpS4-phoA* fusion gene, the pJEM11, a
13 *phoA* reporter shuttle plasmid that replicates in *E. coli* and in *M. smegmatis* (Lim *et al.*, 1995), was
14 generously provided by B. Gicquel. The *M. smegmatis* *mmpS4* gene was PCR amplified (primers
15 *mmpS4phoA*Cter.5 and *mmpS4phoA*Cter.3, Table S2) and inserted into the *Bam*HI site of pJEM11
16 plasmid to give pJEM11*mmpS4*. The pJEM11 and pJEM11*mmpS4* plasmids were then used to
17 electoporate *M. smegmatis* strain mc²155.

18

19 **Sliding motility assay.** A 3 μ l aliquot of liquid culture was dispensed onto plates containing 7H9
20 medium plus 0.3% agar with no added carbon source. These plates were then incubated at 37°C for
21 one week.

22

23 **Extraction and purification of mycobacterial lipids.** Lipids were extracted from cell pellets with a
24 mixture of chloroform and methanol as previously described (Villeneuve *et al.*, 2003). The extracts
25 were dried under vacuum and partitioned between water and chloroform (1:1, vol/vol). The organic
26 phases was extensively washed with distilled water and evaporated to dryness. The lipid extracts were
27 dissolved in chloroform and analyzed by thin-layer chromatography (TLC) on plates coated with

1 Durasil 25- silica gel (0.25 mm thickness; Macherey-Nagel). The GPLs (250 µg each deposit) were
2 resolved by TLC run in chloroform-methanol (9:1, vol/vol) and visualised by spraying the plates with
3 0.2% anthrone in concentrated sulfuric acid, and heating at 110°C. Identification of GPLs was
4 achieved by matrix-assisted laser-desorption/ionization-time-of-flight (MALDI-TOF) mass
5 spectrometry analysis (Villeneuve et al., 2003).

6
7 **Quantification of GPL production.** Three 200 mL independent cultures of each strain were grown in
8 LB medium, centrifuged and the pellets were weighted. Lipids were extracted and GPLs were purified
9 as described above. GPL purification was completed by deacylating the lipids with 0.1 M KOH as
10 described by Brennan and Goren (Brennan & Goren, 1979). We expressed the amount of GPLs
11 produced as a function of bacterial pellet weight.

12
13 **Surface-exposed material preparation.** The surface-exposed material was recovered from
14 mycobacteria cells treated with 10 g of glass beads as previously described (Sonden *et al.*, 2005). The
15 surface-exposed and the cell-associated lipids were extracted with chloroform and methanol from the
16 surface-exposed material and from the residual cells, respectively, and their GPLs components
17 identified by TLC as described above. The amount of lipid spotted on the TLC plates corresponded to
18 the amount of lipid produced by 5 mg of cells, dry weight.

19
20 **Immobilization of GPL and selection of single-chain antibodies.** We coupled 1 mg of purified GPL
21 to 6 mg of ovalbumin using 1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride (Pierce)
22 according to the manufacturer's instructions. A Tomlinson I single-chain Fv antibodies library was
23 kindly supplied by Geneservice, Cambridge, UK. Fv antibodies were screened according to the
24 instructions supplied with the library. GPL-coupled ovalbumin was used as the antigen for screening.
25 After screening, the gene encoding the antibody identified (sc051) was inserted into the pMal-2X
26 vector (NEB) (Bach *et al.*, 2001) containing the *ds-red2* gene (Clontech) downstream from the
27 multiple cloning site. This construct was used to produce a fusion protein consisting of the antibody
28 fused to the fluorescent DS-Red2 protein.

1
2 **Immunofluorescence microscopy.** Bacteria were labelled with 10 µg/ml FITC (Sigma) by incubating
3 with gentle shaking for 1h at 37°C. The bacteria were washed three times with PBS and three times
4 with double distilled water and immobilized on cover slip by flaming. The cover slip was covered with
5 100 µg of sc051 in PBS and incubated for 30 min. Unbound antibody was then washed away, and the
6 cover slip was washed three times, for three minutes each with double distilled water, and mounted on
7 a glass slide containing FluorSave (Calbiochem). Samples were analyzed by epifluorescence
8 microscopy, as previously described (Sendide *et al.*, 2004). The averages of intensities were measured
9 for each panel and the $\Delta mmpS4$ strain relative fluorescence was calculated compared to the wild-type
10 strain fluorescence.

11
12 **Construction of the red fluorescent fusion expression plasmids.** For construction of the chimeric
13 *rfp-mmpS4* gene, the wild-type *mmpS4* gene coding sequence was amplified by PCR with *Pfu* Turbo
14 DNA polymerase (Stratagene) using *M. smegmatis* mc²155 genomic DNA as the template and the
15 primers *mmpS4rfpNter.5* and *mmpS4_trans.3* (Table S2). The *rfp* gene was amplified using the
16 plasmid pMal-2X as a template and the primers *rfpNter.5* and *rfpNter.3*. The PCR products were
17 purified with a Qiagen PCR purification kit, digested with *NdeI* and *XbaI* and inserted into the
18 dephosphorylated integrative expression vector pNip40b at the single *XbaI* site to generate pNip*Rfp-*
19 *mmpS4*. Enzymatic digestion was used to select clones on which the *rfp-mmpS4* fusion gene was
20 inserted in the opposite orientation to the hygromycin resistant gene. One clone was selected and
21 sequenced. A similar strategy was used to clone the *mmpS4-rfp* (*mmpS4_trans.5* and *mmpS4rfpCter.3*
22 primers), *mbtH-rfp* (*mbtH_trans.5* and *mbtH_rfpCter.3*) and *fadD23-rfp* (*fadD23_trans.5* and
23 *fadD23_rfpCter.3*) chimeric genes. In this case, *rfpCter.5* and *rfpCter.3* primers were used for the
24 amplification of the *rfp* gene. The resulting plasmids were named pNip*mmpS4-rfp*, pNip*mbtH-rfp* and
25 pNip*fadD23rfp*. Plasmids were used to electroporate the wild-type and mutant *M. smegmatis* $\Delta mmpS4$
26 strains and transformants were selected on hygromycin.

27

1 **Construction of the truncated *mmpS4* expression plasmids.** For construction of the truncated
2 *mmpS4* genes, the *M. smegmatis* mc²155 *mmpS4* gene coding sequence was amplified by PCR as
3 described above using the primers *mmpS4*_trans.5 and *mmpS4*-48.3 or *mmpS4*-82.3 or *mmpS4*-116.3
4 (Table S2) to generate respectively the plasmids pNip*mmpS4*-48, pNip*mmpS4*-82 or pNip*mmpS4*-116.
5 For the amplification of the extracytosolic domain of *mmpS4* gene the primers *mmpS4*cyto.5 and
6 *mmpS4*cyto.3 were used.

7
8 **RNA isolation and RT-PCR assay.** Total RNA was extracted from 10 ml of log phase cultures of *M.*
9 *smegmatis* strains grown in LB medium as previously described (Deshayes et al., 2005).
10 Contaminating DNA was removed by digestion with DNase I according to the manufacturer's
11 instructions (Promega). The DNase I enzyme was removed with two phenol-chloroform-
12 isoamylalcohol extractions, followed by ethanol precipitation. Reverse transcription was carried out
13 with both primers *mmpS4*-R and *mmpS1*-R. A PCR run on cDNA was then performed by using the
14 primers *mmpS4*-F/*mmpS4*-R and *mmpS1*-F/*mmpS1*-R. The PCR products were resolved by horizontal
15 electrophoresis on a 1.5% agarose gel.

16

1 **Acknowledgements.**

2 We thank C. Jeanneau and M. Bertili for bacterial medium preparation. CD is funded by a doctoral
3 grant from “FRM” and a post-doctoral grant form Conseil Général Maine-et-Loire and Angers Loire
4 Métropole. We gratefully acknowledge “CEFIPRA” (grant 3403-B) and INSERM for funding this
5 project. We thank V. Pelicic for critical reading of the manuscript. JMR was Directeur de Recherches
6 at INSERM. We also thank Jeffrey Helm for technical assistance.

7

For Peer Review

1 **Legends to Figures.**

2

3 **Table 1.** Number of *mmpS* and *mmpL* orthologs per bacterial species considered. The asterisk
4 indicates the presence of a gene containing frameshift (the *mmpL13a* and *mmpL13b* genes in *M.*
5 *tuberculosis*, *mmpL1a*, *mmpL1b*, *mmpL9a* and *mmpL9b* in *M. bovis*, *mmpL1_1*, *mmpL13_1*, *mmpL2*,
6 *mmpL4*, *mmpL4_5*, *mmpL4_6*, *mmpL4_7*, *mmpL4_8* and *mmpL4_9* in *M. ulcerans*).

7

8 **Figure 1. A-** Phylogenetic tree of the *mmpS* paralogs of *M. smegmatis* (MSMEG), *M. avium* subsp.
9 *avium* (MAV), *M. bovis* (MB), *M. ulcerans* (MUL), *M. leprae* (ML) and *M. tuberculosis* H37Rv (Rv).

10 **B-** Topology of MmpS4 from *M. smegmatis* generated though the prediction obtained by the
11 **HMMTOP** software. The hydrophobic residues are indicated in black, the polar residues are indicated
12 in light blue, the positively charged residues are indicated in dark blue and the negative residues are
13 indicated in red. **C-** Phenotype on X-P containing LB plates of wild-type *M. smegmatis* strains
14 carrying pJEM11 or pJEM11*mmpS4* reflecting the alkaline phosphatase activity.

15

16 **Figure 2. A-** Sliding motility of the wild-type and Δ *mmpS4* *M. smegmatis* strains on low agar
17 containing plate. **B-** TLC analyses of GPLs (250 μ g each deposit) produced by the wild-type, Δ *mmpS4*
18 and Δ *mmpS4/mmpS4Ms* strains. **C-** Quantification of alkali-stable lipids in wild-type, mutant and
19 complemented strains. The percentages of alkali-stable lipids differ significantly between the mutant
20 and the wild-type and complemented strains ($p < 0.01$).

21

22 **Figure 3: A-** TLC analyses of the lipid extracts of the wild-type, Δ *mmpS4*, and Δ *mmpS4/mmpS4Ms*
23 strains **C**, cell-associated (non-surface-exposed) GPLs; **S**, surface-exposed GPLs. Each deposit
24 corresponds to the estimated amount of lipid produced by 5 mg of cells, dry weight. Apolar and polar
25 GPLs are indicated with open and closed boxes, respectively. **B-** Immunofluorescence microscopy
26 analyses. Bacteria were labelled with FITC (green) and incubated with purified anti-GPL antibodies
27 coupled to DSred. The merge image is indicated on the panel on the right.

1

2 **Figure 4. A-** Percentage of punctuate fluorescence in *M. smegmatis* producing cytosolic GFP (pG13),
3 the MmpS4-RFP or the RFP-MmpS4 fusion protein. **B-** Immunofluorescence of *M. smegmatis*
4 producing the MmpS4-RFP and RFP-MmpS4 fusion proteins. The bar indicates 2.5 μ m. The merge
5 image is indicated on the panel on the right.

6

7 **Figure 5. A-** Hypothetical model of the GPL biosynthesis enzymatic megacomplex stabilized and
8 anchored to the bacterial membrane by MmpS and MmpL proteins. **B-** Immunofluorescence of *M.*
9 *smegmatis* producing the MbtH-Rfp and FadD23-Rfp fusion proteins. The merge image is indicated
10 on the right panel. The bar indicates 2.5 μ m. **C-** Percentage of punctuate fluorescence in the wild-type
11 or $\Delta mmpS4$ strains producing the MbtH-RFP or FadD23-Rfp fusion proteins.

12

13 **Figure 6.** Orthologous complementation. **A-** Sliding motility of $\Delta mmpS4/mmpS4Tb$ and
14 $\Delta mmpS4/mmpS4Tb$ strains on low agar containing plate. **B-** TLC analysis of the lipid extracts from the
15 $\Delta mmpS4/mmpS4Tb$ strain. C, cell-associated (non-surface-exposed) GPLs; S, surface-exposed GPLs.
16 Apolar and polar GPLs are indicated with open and closed boxes, respectively. **C-** Immunofluorescent
17 microscopy analyses of the $\Delta mmpS4/mmpS4Tb$ strain. Bacteria were labelled with FITC (green) and
18 incubated with purified anti-GPL antibodies coupled to DSred. The merge image is indicated on the
19 panel on the right.

20

21

1 **Table 1.** Number of *mmpS* and *mmpL* orthologs per considered bacterial species. The asterisk
 2 indicates the presence of gene containing frameshift (The *mmpL13a* and *mmpL13b* in *M. tuberculosis*,
 3 *mmpL1a*, *mmpL1b*, *mmpL9a* and *mmpL9b* in *M. bovis*, *mmpL1_1*, *mmpL13_1*, *mmpL2*, *mmpL4*,
 4 *mmpL4_5*, *mmpL4_6*, *mmpL4_7*, *mmpL4_8* and *mmpL4_9* in *M. ulcerans*).

5
6
7

Bacterial strains	<i>mmpS</i>	<i>mmpL</i>	Genome size (Mb)
<i>M. smegmatis</i> mc ² 155	10	17	6.9
<i>M. abscessus</i> ATCC 19977	27	31	5.1
<i>M. gilvum</i> PYR-GCK	6	16	5.6
<i>M. sp.</i> KMS	6	15	6.3
<i>M. vanbaalenii</i> PYR-1	5	14	6.4
<i>M. avium subsp. avium</i> ATCC25291	10	15	5.5
<i>M. avium subsp. paratuberculosis</i> K-10	12	18	4.8
<i>M. bovis</i> AF2122/97	6	16 *	4.3
<i>M. leprae</i> TN	2	7	3.3
<i>M. tuberculosis</i> H37Rv	5	14 *	4.4
<i>M. ulcerans</i> Agy99	6	16 *	5.8
<i>Rhodococcus jostii</i> RHA1	0	9	9.7
<i>Streptomyces coelicolor</i> A3	0	13	8.6

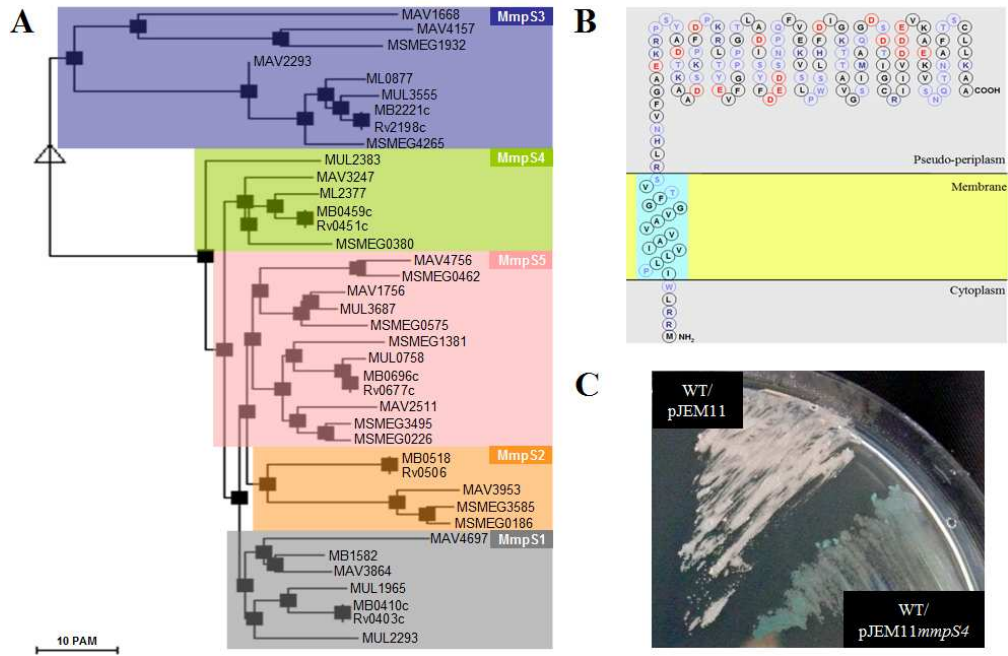
8
9
10

1 **References**

- 2
- 3 Abdallah, A. M., N. C. Gey van Pittius, P. A. Champion, J. Cox, J. Luirink, C. M. Vandenbroucke-Grauls, B. J.
4 Appelmek & W. Bitter, (2007) Type VII secretion--mycobacteria show the way. *Nat Rev Microbiol* **5**:
5 883-891.
- 6 Bach, H., Y. Mazor, S. Shaky, A. Shoham-Lev, Y. Berdichevsky, D. L. Gutnick & I. Benhar, (2001) Escherichia
7 coli maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-
8 chain antibodies. *J Mol Biol* **312**: 79-93.
- 9 Barker, L. P., D. M. Brooks & P. L. Small, (1998) The identification of Mycobacterium marinum genes
10 differentially expressed in macrophage phagosomes using promoter fusions to green fluorescent
11 protein. *Mol Microbiol* **29**: 1167-1177.
- 12 Belisle, J. T., M. R. McNeil, D. Chatterjee, J. M. Inamine & P. J. Brennan, (1993) Expression of the core
13 lipopeptide of the glycopeptidolipid surface antigens in rough mutants of Mycobacterium avium. *J Biol*
14 *Chem* **268**: 10510-10516.
- 15 Bernal, A., U. Ear & N. Kyrpides, (2001) Genomes OnLine Database (GOLD): a monitor of genome projects
16 world-wide. *Nucleic Acids Res* **29**: 126-127.
- 17 Betts, J. C., P. T. Lukey, L. C. Robb, R. A. McAdam & K. Duncan, (2002) Evaluation of a nutrient starvation
18 model of Mycobacterium tuberculosis persistence by gene and protein expression profiling. *Mol*
19 *Microbiol* **43**: 717-731.
- 20 Brennan, P. J., G. O. Aspinall & J. E. Shin, (1981) Structure of the specific oligosaccharides from the
21 glycopeptidolipid antigens of serovars in the Mycobacterium avium-Mycobacterium intracellulare-
22 Mycobacterium scrofulaceum complex. *J Biol Chem* **256**: 6817-6822.
- 23 Brennan, P. J. & M. B. Goren, (1979) Structural studies on the type-specific antigens and lipids of the
24 *Mycobacterium avium-Mycobacterium intracellulare-Mycobacterium scrofulaceum* serocomplex.
25 *Mycobacterium intracellulare* serotype 9. *J Biol Chem* **254**: 4205-4211.
- 26 Byrd, T. F. & C. R. Lyons, (1999) Preliminary characterization of a *Mycobacterium abscessus* mutant in human
27 and murine models of infection. *Infect Immun* **67**: 4700-4707.
- 28 Bystrykh, L. V., M. A. Fernandez-Moreno, J. K. Herrema, F. Malpartida, D. A. Hopwood & L. Dijkhuizen,
29 (1996) Production of actinorhodin-related "blue pigments" by *Streptomyces coelicolor* A3(2). *J*
30 *Bacteriol* **178**: 2238-2244.
- 31 Camacho, L. R., P. Constant, C. Raynaud, M. A. Laneelle, J. A. Triccas, B. Gicquel, M. Daffe & C. Guilhot,
32 (2001) Analysis of the phthiocerol dimycocerosate locus of *Mycobacterium tuberculosis*. Evidence that
33 this lipid is involved in the cell wall permeability barrier. *J Biol Chem* **276**: 19845-19854.
- 34 Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E.
35 Barry, 3rd, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K.
36 Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels & B. G. Barrell, (1998)
37 Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*
38 **393**: 537-544.
- 39 Converse, S. E., J. D. Mougous, M. D. Leavell, J. A. Leary, C. R. Bertozzi & J. S. Cox, (2003) MmpL8 is
40 required for sulfolipid-1 biosynthesis and *Mycobacterium tuberculosis* virulence. *Proc Natl Acad Sci U*
41 *S A* **100**: 6121-6126.
- 42 Corpet, F., (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**: 10881-10890.
- 43 Cox, J. S., B. Chen, M. McNeil & W. R. Jacobs, Jr., (1999) Complex lipid determines tissue-specific replication
44 of *Mycobacterium tuberculosis* in mice. *Nature* **402**: 79-83.
- 45 Crick, D. C. & P. J. Brennan, (2008) Biosynthesis of the arabinogalactan peptidoglycan complex of
46 *Mycobacterium tuberculosis*.; Daffe M, Reyrat JM, editors: ASM press. 395 p.
- 47 De Groote, M. A. & G. Huitt, (2006) Infections due to rapidly growing mycobacteria. *Clin Infect Dis* **42**: 1756-
48 1763.
- 49 de Mendonca-Lima, L., M. Picardeau, C. Raynaud, J. Rauzier, Y. O. de la Salmoniere, L. Barker, F. Bigi, A.
50 Cataldi, B. Gicquel & J. M. Reyrat, (2001) Erp, an extracellular protein family specific to mycobacteria.
51 *Microbiology* **147**: 2315-2320.
- 52 Deshayes, C., F. Laval, H. Montrozier, M. Daffe, G. Etienne & J. M. Reyrat, (2005) A glycosyltransferase
53 involved in biosynthesis of triglycosylated glycopeptidolipids in *Mycobacterium smegmatis*: impact on
54 surface properties. *J Bacteriol* **187**: 7283-7291.
- 55 Domenech, P., M. B. Reed & C. E. Barry, 3rd, (2005) Contribution of the *Mycobacterium tuberculosis* MmpL
56 protein family to virulence and drug resistance. *Infect Immun* **73**: 3492-3501.
- 57 Domenech, P., M. B. Reed, C. S. Dowd, C. Manca, G. Kaplan & C. E. Barry, 3rd, (2004) The role of MmpL8 in
58 sulfatide biogenesis and virulence of *Mycobacterium tuberculosis*. *J Biol Chem* **279**: 21257-21265.

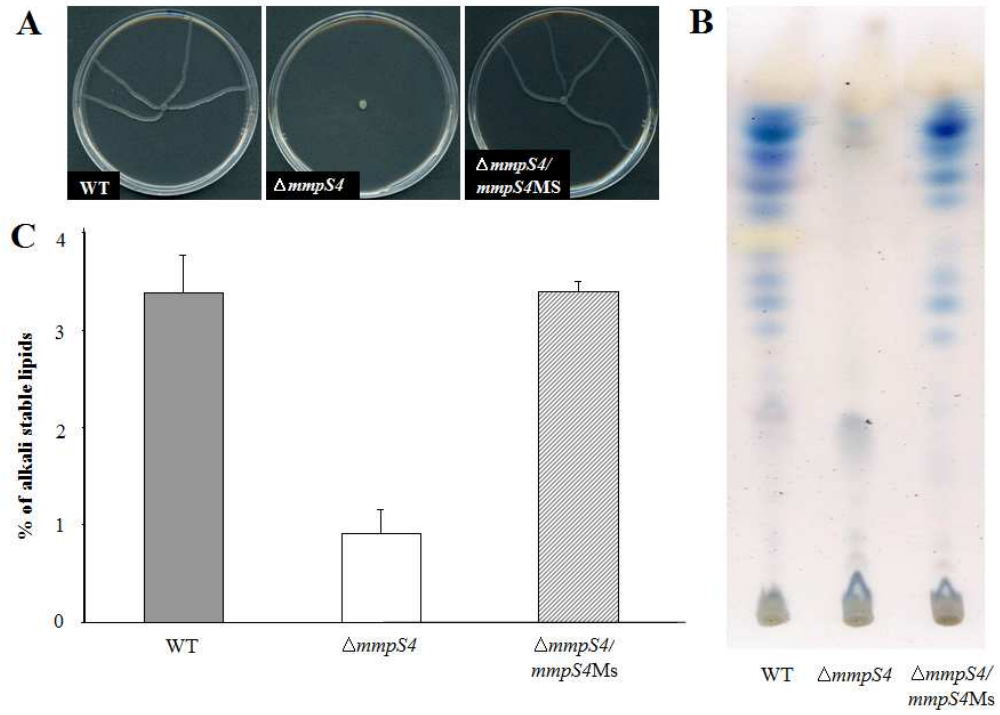
- 1 Etienne, G., F. Laval, C. Villeneuve, P. Dinadayala, A. Abouwarda, D. Zerbib, A. Galamba & M. Daffe, (2005)
2 The cell envelope structure and properties of *Mycobacterium smegmatis* mc(2)155: is there a clue for
3 the unique transformability of the strain? *Microbiology* **151**: 2075-2086.
- 4 Gokhale, R. S., P. Saxena, T. Chopra & D. Mohanty, (2007) Versatile polyketide enzymatic machinery for the
5 biosynthesis of complex mycobacterial lipids. *Nat Prod Rep* **24**: 267-277.
- 6 Hirokawa, T., S. Boon-Chieng & S. Mitaku, (1998) SOSUI: classification and secondary structure prediction
7 system for membrane proteins. *Bioinformatics* **14**: 378-379.
- 8 Howard, S. T., E. Rhoades, J. Recht, X. Pang, A. Alsup, R. Kolter, C. R. Lyons & T. F. Byrd, (2006)
9 Spontaneous reversion of *Mycobacterium abscessus* from a smooth to a rough morphotype is associated
10 with reduced expression of glycopeptidolipid and reacquisition of an invasive phenotype. *Microbiology*
11 **152**: 1581-1590.
- 12 Jain, M. & J. S. Cox, (2005) Interaction between Polyketide Synthase and Transporter Suggests Coupled
13 Synthesis and Export of Virulence Lipid in *M. tuberculosis*. *PLoS Pathog* **1**: e2.
- 14 Kocincova, D., A. K. Singh, J. L. Beretti, H. Ren, D. Euphrasie, J. Liu, M. Daffe, G. Etienne & J. M. Reyrat,
15 (2008) Spontaneous transposition of IS1096 or ISMsm3 leads to glycopeptidolipid overproduction and
16 affects surface properties in *Mycobacterium smegmatis*. *Tuberculosis (Edinb)* **88**: 390-398.
- 17 Lim, E. M., J. Rauzier, J. Timm, G. Torrea, A. Murray, B. Gicquel & D. Portnoi, (1995) Identification of
18 mycobacterium tuberculosis DNA sequences encoding exported proteins by using phoA gene fusions. *J*
19 *Bacteriol* **177**: 59-65.
- 20 Lopez-Marin, L. M., N. Gautier, M. A. Laneelle, G. Silve & M. Daffe, (1994) Structures of the
21 glycopeptidolipid antigens of *Mycobacterium abscessus* and *Mycobacterium chelonae* and possible
22 chemical basis of the serological cross-reactions in the *Mycobacterium fortuitum* complex.
23 *Microbiology* **140 (Pt 5)**: 1109-1118.
- 24 Medjahed, H. & J. M. Reyrat, (2009) Construction of *Mycobacterium abscessus* defined glycopeptidolipid
25 mutants: comparison of genetic tools. *Appl Environ Microbiol* **75**: 1331-1338.
- 26 Milano, A., M. R. Pasca, R. Provvedi, A. P. Lucarelli, G. Manina, A. L. Ribeiro, R. Manganelli & G. Riccardi,
27 (2009) Azole resistance in *Mycobacterium tuberculosis* is mediated by the MmpS5-MmpL5 efflux
28 system. *Tuberculosis (Edinb)* **89**: 84-90.
- 29 Patterson, J. H., M. J. McConville, R. E. Haites, R. L. Coppel & H. Billman-Jacobe, (2000) Identification of a
30 methyltransferase from *Mycobacterium smegmatis* involved in glycopeptidolipid synthesis. *J Biol Chem*
31 **275**: 24900-24906.
- 32 Pelicic, V., J. M. Reyrat & B. Gicquel, (1996) Generation of unmarked directed mutations in mycobacteria,
33 using sucrose counter-selectable suicide vectors. *Mol Microbiol* **20**: 919-925.
- 34 Rao, A. & A. Ranganathan, (2004) Interaction studies on proteins encoded by the phthiocerol dimycocerosate
35 locus of *Mycobacterium tuberculosis*. *Mol Genet Genomics* **272**: 571-579.
- 36 Recht, J. & R. Kolter, (2001) Glycopeptidolipid acetylation affects sliding motility and biofilm formation in
37 *Mycobacterium smegmatis*. *J Bacteriol* **183**: 5718-5724.
- 38 Recht, J., A. Martinez, S. Torello & R. Kolter, (2000) Genetic analysis of sliding motility in *Mycobacterium*
39 *smegmatis*. *J Bacteriol* **182**: 4348-4351.
- 40 Reyrat, J. M., V. Pelicic, B. Gicquel & R. Rappuoli, (1998) Counterselectable markers: untapped tools for
41 bacterial genetics and pathogenesis. *Infect Immun* **66**: 4011-4017.
- 42 Ripoll, F., C. Deshayes, S. Pasek, F. Laval, J. L. Beretti, F. Biet, J. L. Risler, M. Daffe, G. Etienne, J. L. Gaillard
43 & J. M. Reyrat, (2007) Genomics of glycopeptidolipid biosynthesis in *Mycobacterium abscessus* and
44 *M. chelonae*. *BMC Genomics* **8**: 114.
- 45 Sendide, K., A. E. Deghmane, J. M. Reyrat, A. Talal & Z. Hmama, (2004) *Mycobacterium bovis* BCG urease
46 attenuates major histocompatibility complex class II trafficking to the macrophage cell surface. *Infect*
47 *Immun* **72**: 4200-4209.
- 48 Snapper, S. B., R. E. Melton, S. Mustafa, T. Kieser & W. R. Jacobs, Jr., (1990) Isolation and characterization of
49 efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* **4**: 1911-1919.
- 50 Sonden, B., D. Kocincova, C. Deshayes, D. Euphrasie, L. Rhayat, F. Laval, C. Frehel, M. Daffe, G. Etienne & J.
51 M. Reyrat, (2005) Gap, a mycobacterial specific integral membrane protein, is required for glycolipid
52 transport to the cell surface. *Mol Microbiol* **58**: 426-440.
- 53 Straight, P. D., M. A. Fischbach, C. T. Walsh, D. Z. Rudner & R. Kolter, (2007) A singular enzymatic
54 megacomplex from *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **104**: 305-310.
- 55 Sulzenbacher, G., S. Canaan, Y. Bordat, O. Neyrolles, G. Stadthagen, V. Roig-Zamboni, J. Rauzier, D. Maurin,
56 F. Laval, M. Daffe, C. Cambillau, B. Gicquel, Y. Bourne & M. Jackson, (2006) LppX is a lipoprotein
57 required for the translocation of phthiocerol dimycocerosates to the surface of *Mycobacterium*
58 *tuberculosis*. *Embo J* **25**: 1436-1444.

- 1 Tafelmeyer, P., C. Laurent, P. Lenormand, J. C. Rousselle, L. Marsollier, G. Reysset, R. Zhang, A. Sickmann, T.
2 P. Stinear, A. Namane & S. T. Cole, (2008) Comprehensive proteome analysis of *Mycobacterium*
3 *ulcerans* and quantitative comparison of mycolactone biosynthesis. *Proteomics* **8**: 3124-3138.
- 4 Tekaia, F., S. V. Gordon, T. Garnier, R. Brosch, B. G. Barrell & S. T. Cole, (1999) Analysis of the proteome of
5 *Mycobacterium tuberculosis* in silico. *Tuber Lung Dis* **79**: 329-342.
- 6 Tseng, T. T., K. S. Gratwick, J. Kollman, D. Park, D. H. Nies, A. Goffeau & M. H. Saier, Jr., (1999) The RND
7 permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and
8 development proteins. *J Mol Microbiol Biotechnol* **1**: 107-125.
- 9 Villeneuve, C., G. Etienne, V. Abadie, H. Montrozier, C. Bordier, F. Laval, M. Daffe, I. Maridonneau-Parini &
10 C. Astarie-Dequeker, (2003) Surface-exposed glycopeptidolipids of *Mycobacterium smegmatis*
11 specifically inhibit the phagocytosis of mycobacteria by human macrophages. Identification of a novel
12 family of glycopeptidolipids. *J Biol Chem* **278**: 51291-51300.
- 13 WHO, Buruli Ulcer Fact Sheets. [<http://www.who.int/mediacentre/factsheets/fs199/en/>].
14 WHO, Tuberculosis Fact Sheets. [<http://www.who.int/mediacentre/factsheets/fs104/en/>].
- 15 Wolpert, M., B. Gust, B. Kammerer & L. Heide, (2007) Effects of deletions of mbtH-like genes on clorobiocin
16 biosynthesis in *Streptomyces coelicolor*. *Microbiology* **153**: 1413-1423.
17
18



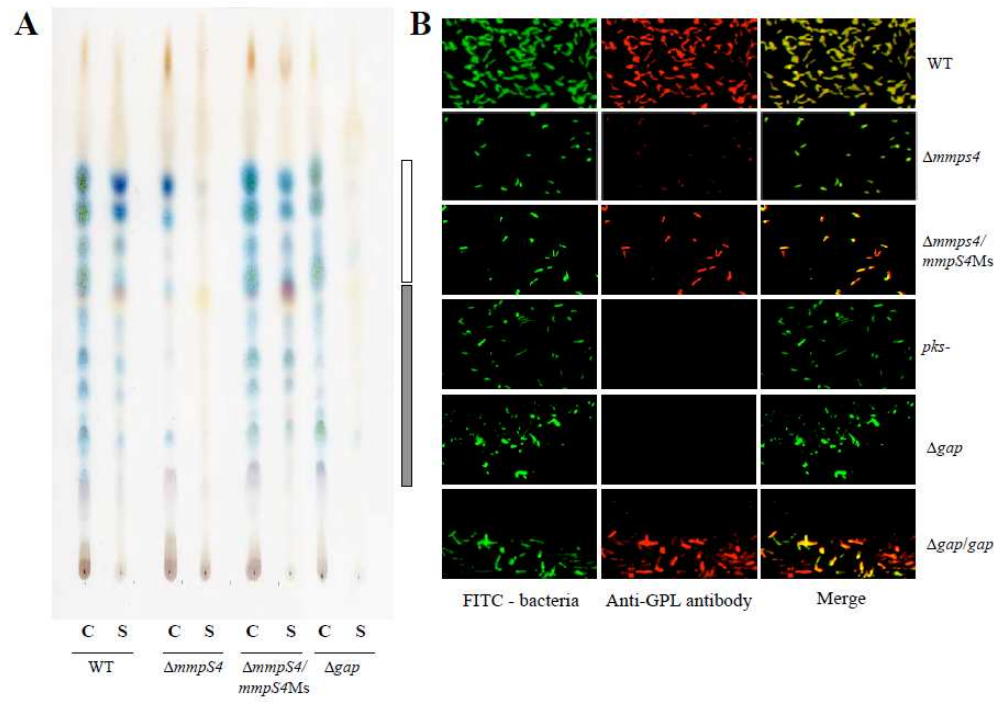
348x230mm (72 x 72 DPI)

Review



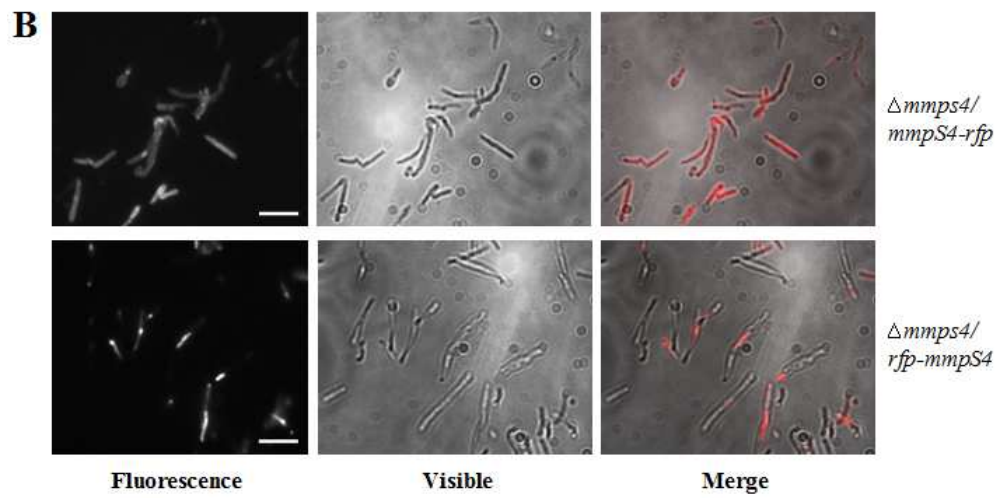
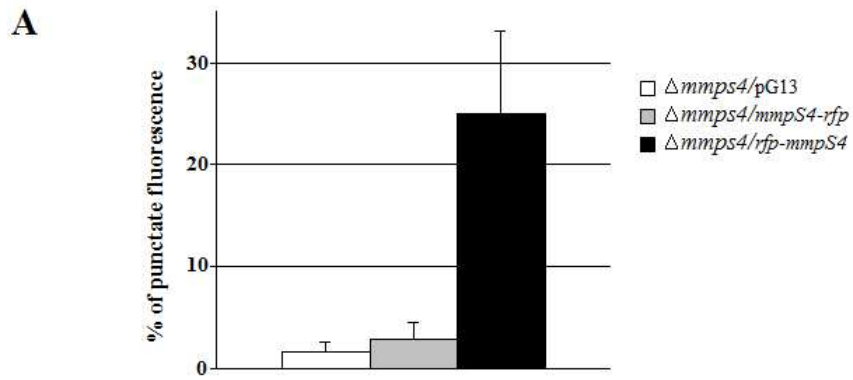
309x222mm (72 x 72 DPI)

review



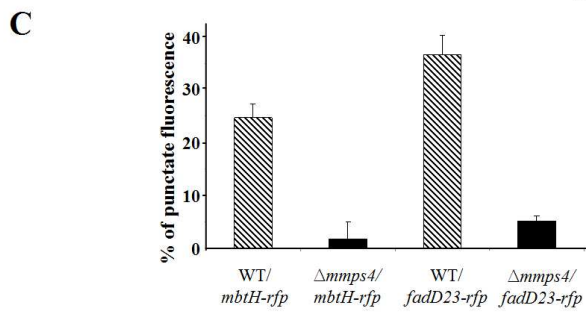
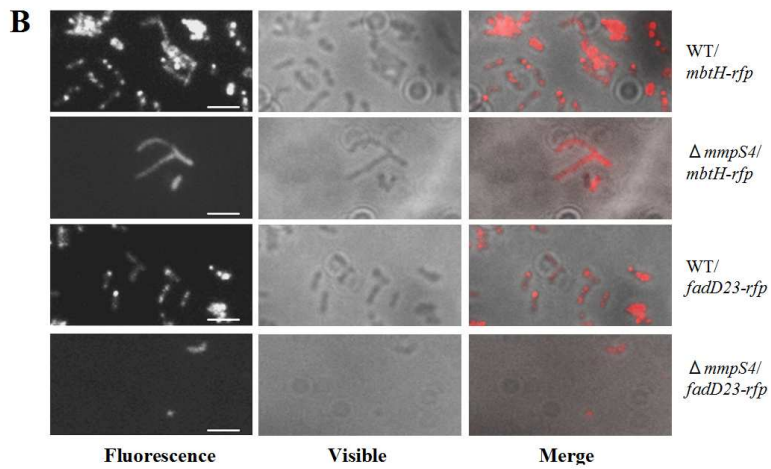
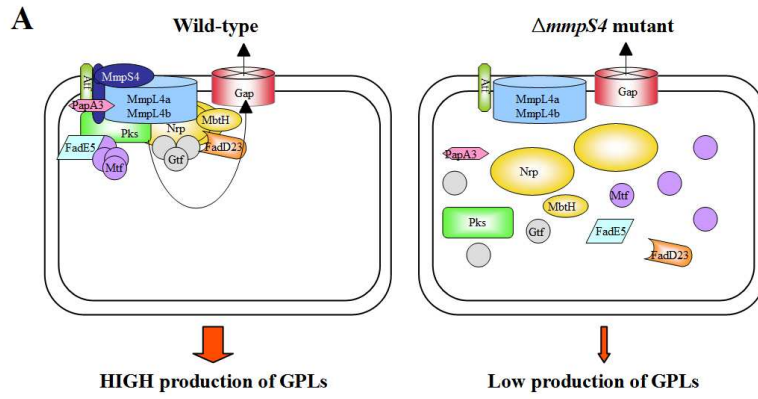
321x228mm (72 x 72 DPI)

review

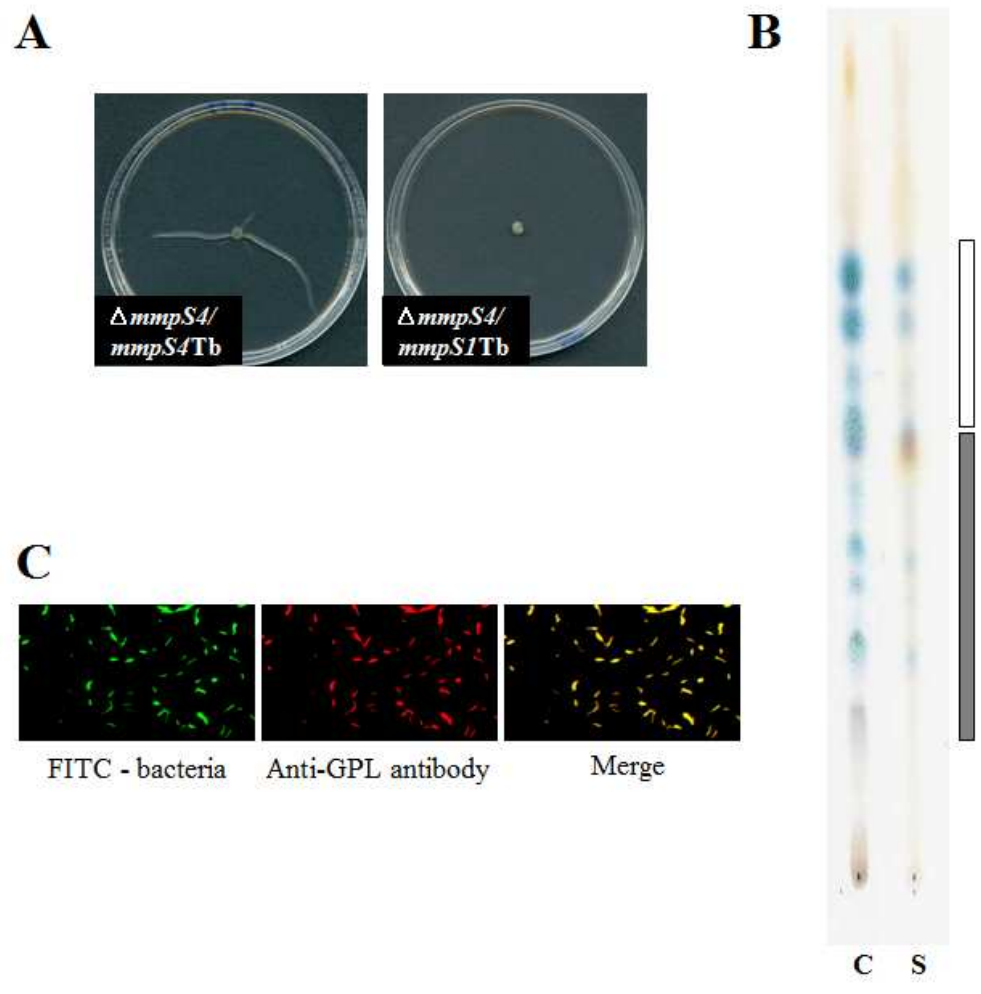


251x231mm (72 x 72 DPI)





376x567mm (72 x 72 DPI)



219x221mm (72 x 72 DPI)