Molecular and functional analysis of the *mce4* operon in *Mycobacterium smegmatis*

Julia García-Fernández[1], Kadamba Papavinasasundaram[2], Beatriz Galán[1], Christopher M. Sassetti[2] and José L. García[1]*

[1]Centro de Investigaciones Biológicas, (CIB-CSIC), Madrid, Spain.

[2]University of Massachusetts Medical School, Worcester, MA, USA.

* Corresponding author: Environmental Biotechnology, Centro de Investigaciones Biológicas, Ramiro de Maeztu 9, 28040 Madrid, Spain. Tel: 34-918373112; e-mail: jlgarcia@cib.csic.es

Abstract

***Mycobacterium smegmatis* contains 6 homologous *mce* (mammalian cell entry) operons which have been proposed to encode ABC-like import systems. The *mce* operons encode up to 10 different proteins of unknown function that are not present in conventional ABC transporters. We have analyzed the consequences of individually deleting each of the genes of the *mce4* operon of *M. smegmatis,* which mediates the transport of cholesterol. None of the *mce4* mutants were able to grow in cholesterol suggesting that all these genes are required for its uptake and that none of them can be replaced by the homologous genes of the other *mce* operons. This result suggests that different *mce* operons do not provide redundant capabilities and that *M. smegmatis,* in contrast with *Mycobacterium tuberculosis*, is not able to use alternative import systems to import cholesterol in the analyzed culture conditions. Either deletion of the entire *mce4* operon or single point mutations that eliminate the transport function cause a phenotype similar to the one observed in a mutant lacking all 6 *mce* operons suggesting a pleiotropic role for this system.**

Introduction

Cholesterol is a polycyclic sterol highly abundant in biosphere with a great relevance in biology, medicine and chemistry since it plays essential roles as structural component of animal cell membranes and as precursor of vitamins, steroid hormones and bile acids (Slaytor and Bloch, 1965). Its environmental ubiquity is due in part to its high resistance to microbial degradation. Cholesterol is used as a carbon and energy source by a few genera of microorganisms, some of them very important pathogens such as *Mycobacterium tuberculosis*. Cholesterol is recalcitrant to biodegradation not only because of its high hydrophobicity (*i.e*., solubility in water of 3×10^{-8} M), but also because it does not freely diffuse to the cytoplasm and needs to be transported by a specific uptake system to become metabolized. This transport appears to be critical mainly when the cell envelope structure is as complex as that of mycobacteria (Minnikin, 1982). The cell wall of mycobacteria is composed of a peculiar asymmetric external lipid bilayer of exceptional thickness (≥10 nm). The inner leaflet of this layer contains mycolic acids (up to 90 carbons in length) that are covalently linked to arabinogalactan, and the outer leaflet contains other extractable lipids (Liu *et al*., 1996; Niederweis *et al*., 2010). Thus, the unique complexity and hydrophobicity of the mycobacterial cell envelope presents a challenge in understanding how nutrients such as cholesterol are transported across it.

Genome sequence analyses have revealed the presence of *mce* operons in mycobacteria and other Actinobacteria, which are peculiar transport systems homologous to ABC importers (Casali and Riley, 2007). ABC importers have only been identified in prokaryotes and are recognized by the presence of substrate binding proteins (SBP), nucleotide binding proteins that bind and hydrolyze ATP, and transmembrane permeases that span the inner membrane forming a pore for translocation. The *mce* operons contain 10–13 genes that encode two transmembrane proteins homologous to the permease subunits of ABC transporters, along with several putative secreted or cell-surface proteins whose function is still unknown (Casali and Riley, 2007). In this sense, it has been proposed that the Mce systems require many more proteins than do classical ABC transporters probably because these proteins might form a large complex structure necessary for the movement of hydrophobic compounds across membranes and cell walls (Song *et al*., 2008). Although there is increasing evidence showing that these operons are important for the virulence of *M. tuberculosis* (Kumar *et al*., 2003), the *mce* operons are also present in

saprophytic species as *M. smegmatis* (with 6 *mce* operons) and *R. jostii* RHA1 (with 2 *mce* operons), suggesting additional roles for these transporters. The reason for the high number of different Mce systems in the same cell is a question that remains unanswered.

It has been demonstrated that the *mce4* operon is involved in the cholesterol uptake in *M. tuberculosis* (Pandey and Sassetti, 2008), *R. jostii* RHA1 (Mohn *et al*., 2008), *R. equi* (van der Geize *et al*., 2008) and *M. smegmatis* (Klepp *et al*., 2012). Remarkably, in *M. tuberculosis* the uptake of this steroid has been linked to its ability to maintain chronic infection in the host (Pandey and Sassetti, 2008). The *mce4* locus of *R. jostii* RHA1 is up-regulated by cholesterol, and mutants with deletions in this operon lose their ability to grow in this compound (Mohn *et al*., 2008). On the contrary, mutants of *M. tuberculosis* with deletions within the *mce4* operon suffer a restriction for cholesterol dependent-growth relative to glycerol but are still able to grow at a reduced rate. This result supported the role of Mce4 in cholesterol transport but at the same time suggested that *M. tuberculosis* might have an alternative transport system for cholesterol (Pandey and Sassetti, 2008). Klepp *et al*. (2012) have deleted the six *mce* operons of *M. smegmatis* demonstrating that *mce4* is involved in cholesterol uptake; however, whether this is the only cholesterol uptake system still remains to be clarified. In addition, the study of Santangelo *et al*. (2016) concludes that the lack of Mce4 proteins abrogates the recycling of certain cell wall undefined lipids that trigger a redox stress response, suggesting an additional role for this complex. In spite of these studies, a systematic genetic analysis of the *mce* loci has yet to be performed and no precise function or substrate for most of these systems has been proposed up to date.

In this work, we analyse the essentiality of important domains as well as each one of the *mce4* genes in the metabolism of cholesterol in *M. smegmatis* in order to identify the putative presence of alternative cholesterol transport systems in mycobacteria. Our results indicate that all of the genes belonging to the Mce4 system are needed for the cholesterol uptake and that *M. smegmatis* has only one active cholesterol transport system in the conditions tested. Moreover, we suggest that this active transport system is most probably involved in other physiological functions not initially related to the uptake of sterols. Finally, we show that the homologous Mce systems present in the same cell do not substitute for Mce4, suggesting that each one of the complex systems has been maintained to perform a specific role.

Results

In sílico analysis of the mce4 genes

In sílico analyses of the *mce4* locus in *M. smegmatis* showed that it has the conserved configuration of two *yrbE*, six *mce* and two additional *mas* genes found in the other *mce4* loci described to date (*MSMEG_5902-MSMEG_5893*). In addition, the four genes upstream of *yrbE4A* are conserved in all of the organisms studied and are putatively related to steroid metabolism as they encode a 3-ketoacyl-(acyl-carrier-protein) reductase, a ferredoxin and two acyl-CoA dehydrogenases (Fig. 1A). These observations, suggested that the Mce4 system in *M. smegmatis*, like in *M. tuberculosis* and *R. jostti*, constitutes the Mce system involved in sterol uptake.

To investigate possible redundancy between the Mce systems, we performed an *in sílico* comparison of the *mce* operons present in *M. smegmatis.* Of the six *mce* operons, three of them (*mce3*, *mce4* and *mce7* operon) possess the same genetic arrangement described above. On the contrary, *mce1* differs from these in having two additional *mas* genes, like the *mce1* operon found in *M. tuberculosis*. The *mce5* and *mce5b* operons are peculiar since they have insertions between the *mce* genes (Fig. S1). The presence of additional genes in these last *mce* operons compared to the other ones could suggest a further adaptation to provide more specialised transport systems.

Sequence comparisons showed that the *yrbEAB* genes are the most conserved within the *mce* operons, showing over 54% sequence identity (Table S1), on the contrary, the *mce* and *mas* genes are less conserved (Table S2A). Remarkably, the lowest identity was found when comparing the percentage of identity of the *mce4* genes with the homologous genes in the same operon, showing around 25% sequence identity (Table S2B). The sequence divergence of paralogous genes of different operon suggests that distinct Mce systems have evolved to play different functions. Whether the similarities observed within the encoded proteins provide redundancy and robustness to the Mce systems has not been investigated.

Expression analysis of the mce4 genes

The *in sílico* analysis showed putative transcriptional promoters upstream of the *hsd* and *yrbE4A* genes but none of them were found within the 10 genes of *mce4*, suggesting co-transcription of the genes

as an operon (Fig. 1A). In addition, the automatic *in silico* analyses performed by Kendall *et al*. (2007) detected the presence of a KstR operator region located 232 nt upstream the putative translational initiation site suggesting a cholesterol-dependent expression of the whole Mce4 system in *M. smegmatis,* as observed in *R. jostii* (Mohn *et al*., 2008). Although this KstR operator is distal to the proposed promoter we have found by manual inspection a putative KstR operator overlapping the -10 and -35 promoter boxes. Remarkably, the presence of both KstR operator regions was found also in *R. jostii* (Fig. 1A). However, in *M. tuberculosis* we have found only one putative KstR operator sequence that differs significantly from the consensus operator sequence (Fig. 1B).

In order to determine the induction level of the *mce4* operon in *M. smegmatis*, we analysed by qRT-PCR the transcription levels of the *mce4* operon (*MSMEG_5902-MSMEG_5893*) using the *MSMEG_5899* gene as a probe. The expression of this gene showed an induction of 2.2-fold in the presence of cholesterol. Interestingly, its expression was higher in the exponential phase of growth compared to the stationary phase, contrasting with the observation in *M. tuberculosis* (Kumar *et al*., 2003) (Fig. 2). In addition, the specificity of this slight cholesterol induction of *mce4* operon was also confirmed by measuring the expression of the *MSMEG_0134* gene belonging to the *mce1* operon. The expression of this gene was not induced by cholesterol. Finally, the expression of *MSMEG_0134* in stationary phase was higher in glycerol than in cholesterol, suggesting the need of this Mce system in different carbon sources or metabolic conditions (Fig. 2).

Construction and analysis of the mce4 genes deletion mutants

The work reported by Mohn *et al.* (2008) in *R. jostii* RHA1 showed by deleting the functional blocks of the *mce4* operon that the *yrbE4AB*, *mce4ABCDEF* and *mas4AB* blocks were required to grow in cholesterol, but the need of all the individual components of Mce4 transport system still remained to be characterized. In order to confirm that the Mce4 transport system is the only cholesterol uptake system in *M. smegmati*s and to investigate whether all its proteins are indispensable to grow in cholesterol, we generated ten individual gene deletion mutants. The resulting strains named ΔyrbE4AB (Δ*MSMEG_5902*Δ*MSMEG_5901*), ΔyrbE4B (Δ*MSMEG_5901*), Δmce4A (Δ*MSMEG_5900*), Δmce4B (Δ*MSMEG_5899*), Δmce4C (Δ*MSMEG_5898*), Δmce4D (Δ*MSMEG_5897*), Δmce4E (Δ*MSMEG_5896*),

Δmce4F (Δ*MSMEG_5895*), Δmas4A (Δ*MSMEG_5894*) and Δmas4B (Δ*MSMEG_5893*), were unable to grow in cholesterol as the sole carbon and energy source even after 120 h of culture. This inability to utilize cholesterol was proved to be due to the specific deletions, because the complementation *in trans* of each mutant with a wild type copy of the gene expressed from an episomal vector restored its ability to grow in the sterol (Fig. S2). Thus, all of the Mce4 proteins are indispensable for cholesterol growth, reinforcing the theory of the requirement of a large and specific protein complex. But even more interesting, these results suggest that there is not a trans-complementation between the Mce4 proteins and the homologous proteins from the other Mce systems of *M. smegmatis*.

Construction and analysis of the complete mce4 operon deleted mutant

The presence of different Mce systems in the same bacteria and the expression pattern found suggested possible redundant functions of these transporters that might be needed in different growth or metabolic conditions. To confirm the absolute requirement of the Mce4 system of *M. smegmatis* to grow in cholesterol we have generated a complete Δmce4op operon mutant. As expected, this mutant showed no growth in cholesterol containing media (Fig. 3A).

To restore the cholesterol uptake in the Δmce4op mutant we took advantage of the pmce4 plasmid which contains the homologous *mce4* operon from *M. tuberculosis* (GenBank accession number DQ823233) and that has been previously tested in the pathogen as a functional plasmid (Pandey and Sassetti, 2008). Unexpectedly, the pmce4-complemented Δmce4op-C strain started to grow but only after 150 h of culture. The expression of the Mce4 system by the pmce4 plasmid appears to be somehow toxic to the cells since the WT strain transformed with the pmce4 plasmid also showed a growth lag in cholesterol (Fig. 3A). In this sense, additional experiments showed that the Δmce4op-C strain grows better in cholesterol at 30 ºC than at 37 ºC suggesting that a slower growth rate favours the synthesis of the Mce4 system or decreases its toxicity (Fig. 3B). Finally, the addition of a small amount of glycerol (0.5 mM) facilitates the catabolism of cholesterol by the Δmce4op-C complemented strain (Fig. 3C and 3D), suggesting that the toxic effect of pmce4 plasmid might be caused by a depletion of the minimal energy levels required to start the catabolism of cholesterol.

Analysis of the cholesterol uptake

To further confirm that the *mce4* operon is responsible of an active cholesterol transport we determined the ability to actively import ^{14}C-labeled cholesterol by the *M. smegmatis* wild type and ΔMce4A mutant strains. Although due to the high hydrophobicity of cholesterol it was impossible to completely eliminate the background of cholesterol adsorbed to the cell envelope, we observed that the wild type cells imported approximately twice the amount of cholesterol per minute compared to the mutant strain (Fig. 4). To determine the background levels of cholesterol bound to the cell envelope we used DCCD to inhibit the active transport of cholesterol (Mohn *et al.*, 2008) and we observed that the ΔMce4A strain was unable to import cholesterol over the background level observed in wild type DCCD-treated cells. Taking into account that all the *mce4* mutants are unable to grow in cholesterol, we can conclude that the large amount of cholesterol retained non-specifically in the cell envelop of the *mce4* mutants cannot be internalized into the cytoplasm by passive diffusion in these conditions. These results reinforced the hypothesis that the Mce4 is the only active transport system of cholesterol in *M. smegmatis*.

Morphological analysis of the *mce4* operon mutants

All the individual Mce4 mutants of *M. smegmatis* cultured in rich media showed changes in the morphology of the cells similar to those observed in the mutant lacking the six Mce systems (Klepp *et al*., 2012). The mutant cultures grew in clumps in media with glycerol that was devoid of detergent, and displayed a characteristic morphology when plated on 7H9 with Congo Red (Fig. S3A and S3B). The mutant phenotypes were reverted by plasmid complementation of the specific mutations (Fig. 5A and 5B). This phenotype indicates that the Mce4 system is involved in other processes in addition to cholesterol transport.

Similar phenotypes have been attributed to changes in different components of the cell envelope. For example, the *mce1*, *mce2* and *mce3* operons in *M. tuberculosis* have been proposed to be responsible of the lipid content present in the cell wall (Dunphy *et al*., 2010; Marjanovic *et al*., 2011; Santangelo *et al*., 2009). Nevertheless, our analyses of the total content of mycolic acids (MA) in several conditions showed no difference between the wild type and the Mce4 mutants of *M. smegmatis* (data not shown).

Analysis of the EExDA motif found in the YrbE4 permeases

Apart from the high number of proteins present in the *mce* operons, additional characteristics make these transporters unconventional ABC uptake systems. Dassa *et al.* (1985) described a highly-conserved sequence named the EAA motif that is found in the final cytoplasmic loop of some SBP-dependent ABC permeases and is proposed to interact with the cognate ATPase (Mourez *et al*., 1997). This sequence is not found in the Mce permeases, however, examination of the multiple alignments of YrbE proteins revealed a conserved sequence motif located in the penultimate cytoplasmic loop termed the EExDA motif that has been proposed to serve as the site of interaction with the putative cognate MkI ATPase of the Mce systems, in a manner analogous to the EAA loop (Casali and Riley, 2007).

In order to investigate the function of this motif and further test the model that Mce systems represent non-canonical ABC-like transporters, we generated three different mutants in the glutamic (E) residues of the EExDA motif that were substituted for alanine (A) in one or both permeases (Fig. 6A). These mutated permeases were introduced in a replicating vector to test the complementing ability of the constructions generated. Upon transformation of the ΔyrbE4AB strain, we found that the permease point mutants were unable to grow in cholesterol as the only carbon and energy source (Fig. 6B). Interestingly, these mutants showed a phenotype similar to that observed for the Mce4 deletion mutants (Fig. 6C). The importance of this putative ATPase interaction site further supports the functional similarity between Mce and ABC transporters. Moreover, taking into account that these mutants are functional and not structural like those caused by the Mce4 deletions, the observation of this cholesterol-independent phenotype reinforces the idea of the existence of a functional relationship of the Mce4 system with the active transport in the cell envelope of other compounds different from cholesterol.

Discussion

The *M. smegmatis mce4* operon has been associated with cholesterol catabolism because it is slightly induced during growth on this sterol (van der Geize *et al*., 2007; Kendall *et al*., 2007; Uhía *et al*., 2012). The constitutive expression and the modest 2-fold up-regulation of *mce4* observed in this work are consistent with the need of a basal expression of the transporter. This basal expression may be important to detect the presence of cholesterol in their environment which activates the cytosolic KstR regulatory

system. Moreover, this constitutive expression of *mce4* appears to be required to fulfil other roles in addition to cholesterol uptake. Considering the results presented in this work and those of other authors, the transcriptional regulation mechanism of the *mce4* operon of *R. jostii* RHA1 seems to be similar to that observed in *M. smegmatis* (van der Geize *et al*., 2007). However, the presence of an additional regulator binding site in *M. smegmatis* and *R. jostii* and its absence in *M. tuberculosis*, suggests differences in the regulation of this uptake system in saprophytic and pathogenic species.

Living organisms evolved to be functional in variable and unpredictable environments. One common source of this phenotypic robustness is genetic redundancy, in which the failure of a specific component through mutation or environmental challenge can be compensated by duplicate components capable of performing the same function (Kitano, 2004). It has been proposed that the Mce systems likely arose by duplications of a single ancestral locus (Cole *et al*., 1998) and the fact that the highly reduced genome of *Mycobacterium leprae* still contains one *mce* operon, suggests an important role for these systems. Although *M. smegmatis* possesses 6 Mce systems, the results presented in this work confirmed that the *MSMEG_5902-MSMEG_5893* operon constitutes the only active cholesterol uptake system in *M. smegmatis*. This result also confirms that cholesterol cannot be transported by a passive diffusion mechanism in these conditions, in spite of the fact that the cell envelope of *M. smegmatis* is able to capture a high amount of cholesterol by hydrophobic interactions. However, this finding contrast with the behaviour of the Mce4 mutant of *M. tuberculosis* that is still able to grow at a reduced rate in cholesterol (Pandey and Sassetti, 2008). These results suggest that the pathogen could have developed other alternative mechanisms to uptake this compound, due to its importance for growth and survival during infection.

The results obtained in *M. smegmatis* confirm that all 10 proteins found in the Mce4 uptake system are essential for its function, including two permease subunits (YrbE4AB) and eight additional proteins of unknown function (Mce4ABCDEFMas4AB). The absolute requirement of all the genes for cholesterol uptake and their identical growth phenotypes indicate that a partially active transport system does not exist in the absence of any of the proteins. The need for all of the *mce4* genes to grow in cholesterol also indicates that the high number of proteins present in this operon is not due to a functional redundancy or to the need to respond to different uptake specificities. In this sense, these results reinforce

the hypothesis that the Mce4ABCDEFMas4AB proteins form a single complex that fulfils the role of substrate-binding proteins (Casali and Riley, 2007). These studies also demonstrate that none of the deleted genes can be replaced by the homologous genes present in other *mce* operons of *M. smegmatis*. In spite of the high similarity between these genes, ranging from 37% to 72% identity, the specificity of the Mce4 system for cholesterol appears to be also very high suggesting that the components of each Mce import system are devoted to specific and non-redundant tasks.

The inactivation of specific genes of the *mce4* operon caused a cumpling phenotype when grown in glycerol similar to that observed in a Δ6mce mutant which lacks all the *mce* operons of *M. smegmatis* (Klepp *et al*., 2012). Thus, the Mce4 transport system has an additional, cholesterol-independent role in the physiology of *M. smegmatis* to maintain cell wall homeostasis that is observed in normal growth media. This is consistent with previous reports which indicate a role for Mce systems in recycling of lipids (Dunphy *et al*., 2010; Forrellad *et al*., 2014), and alterations of the metabolic profile due to Mce4 deletion (Santangelo *et al*., 2016

The Mce systems possess an EExDA conserved motif that differentiates them from the conventional ABC importers. The mutants obtained in this work allowed us to conclude that the integrity of this sequence in both permeases is essential for the cholesterol uptake function, reinforcing the hypothesis of its possible role in the binding of the ATPase (Casali and Riley, 2007). This result not only proves the essentiality of this motif for the cholesterol uptake process but also suggests that the ATPase subunits associated with the two permease proteins likely act symmetrically, in contrast to the asymmetric orientation described for other ABC transporters (Mourez *et al*., 1997). In addition, the similar phenotype of these point mutants and Mce4 deletion mutant indicates that this phenotype is due to a lack of transporter activity and not a secondary effect on the structure of the cell envelope caused for the lack of Mce4 proteins.

Understanding the biological function of the Mce4 system provides new insight to the apparent role of Mce4 proteins in the pathogenesis of *M. tuberculosis*, suggesting other possible mechanisms to the already described ones. Moreover, the knowledge of the minimal system required to uptake cholesterol can give valuable information in order to transfer this function to heterologous systems for a biotechnological use in the steroid industry.

Experimental Procedures

Bacterial strains and culture conditions

The strains as well as the plasmids used in this work are listed in Table S3. *M. smegmatis* mc^2155 strain was grown at 37°C in an orbital shaker at 250 rpm. Middlebrook 7H9 broth medium (Difco) without albumin-dextrose supplement and without glycerol was used as minimal medium. As carbon source 1.8 mM cholesterol (Sigma) dissolved in 3.6% Tyloxapol (Sigma), was added. Due to the low solubility of steroids, stock solutions were warmed at 80ºC in agitation and then autoclaved. As rich medium, 7H9 broth containing 10% albumin-dextrose supplement and 0.05% Tween 80 was used. 7H10 agar (Difco) plates supplemented with 10% albumin-dextrose were used. Gentamycin (5 μg ml^{-1}), kanamycin (25 μg ml^{-1}) and hygromycin (50 μg ml^{-1}) were used for selection of mycobacteria. Kanamycin (50 μg ml^{-1}) and ampicillin (100 μg ml^{-1}) were used for plasmid selection and maintenance in *Escherichia coli* DH10B strain.

Gene deletions and complementation

The knock-out strains, ΔyrbE4AB (Δ*MSMEG_5902*Δ*MSMEG_5901*), ΔyrbE4B (Δ*MSMEG_5901*) and Δmce4A (Δ*MSMEG_5900*) were constructed by homologous recombination using the pJQ200x plasmid (Jackson *et al*., 2001). The primers used are listed in Table S4. The knock-out strains, Δmce4B (Δ*MSMEG_5899*), Δmce4C (Δ*MSMEG_5898*), Δmce4D (Δ*MSMEG_5897*), Δmce4E (Δ*MSMEG_5896*), Δmce4F (Δ*MSMEG_5895*), Δmas4A (Δ*MSMEG_5894*), Δmas4B (Δ*MSMEG_5893*) and Δmce4op were generated by allelic exchange using the recombineering system (van Kessel and Hatfull, 2007). Briefly, a PCR fragment was constructed that contained the hygromycin resistance cassette flanked by two *lox*P sites cloned in between a sequence upstream and downstream of each gene (-120 bp). This PCR fragment was introduced by electroporation in competent *M. smegmatis*/pJV53 cells that have been induced for the expression of the RecET proteins encoded on the pJV53 plasmid (Murphy *et al*., 2015). Transformants were first selected on 7H10 plates containing hygromycin B (50 μg ml^{-1}). Individual antibiotic-resistant colonies were subcultured and single colonies were analyzed for loss of Km resistance in order to select colonies which lost pJV53. The correct insertion of the hygromycin cassette was verified

by PCR analysis. The hygromycin resistance cassette was cured from the mutant strains by transforming an episomal plasmid (pCre-SacB-Km) that expresses the Cre recombinase. Finally the mutant was cured of the pCre-SacB-Km plasmid by plating on 7H10 medium supplemented with 12.5% sucrose and screening a few colonies on plates containing and lacking kanamycin. Loss of the hygromycin resistance cassette and generation of the deletions were confirmed by both PCR and plating in the presence or absence of hygromycin.

The mutant strains were complemented resulting in the strains ΔyrbE4AB-C, ΔyrbE4B-C, Δmce4A-C, Δmce4B-C, Δmce4C-C, Δmce4D-C, Δmce4E-C, Δmce4F-C, Δmas4A-C and Δmas4B-C respectively. The different genes were amplified using primers listed in Table S4 and the amplicons were digested with the appropriate restriction enzymes and cloned into pMV261. The resulting constructions were introduced in their corresponding mutant strains by electroporation. The complemented Δmce4op-C strain was generated by transforming the mutant strain with the pJS5 plasmid (GenBank accession number DQ823233).

Permease point mutations

The Site Directed Mutagenesis (SDM) approach (Agilent technologies) was used to generate the permease point mutations. First, a PCR was performed using the *yrbE4AB* sequence cloned into pUC18 as template and specific primers designed for site directed mutagenesis (SDM) (Table S4). At the end of PCR, 1 μl of *Dpn*I was added for 1 h to destroy the circular WT template, and the PCR product was transformed into DH5α competent cells. Individual antibiotic-resistant colonies were picked up and sequenced to verify the desired mutation. Once the mutation was confirmed, the *yrbE4AB* mutated sequence was subcloned into pMV261 expression vector and transformed into the ΔyrbE4AB strain.

RNA extraction and real-time PCR (RT-qPCR)

RNA for RT-qPCR was extracted from the *M. smegmatis* mc^2155 strain in logarithmic phase (OD_{600} 0.8-0.9 in medium containing 18 mM glycerol or 0.5–0.6 in medium containing 1.8 mM cholesterol) or in stationary phase (OD_{600} 3-3.2 in medium containing 18 mM glycerol or 1.6–1.8 in medium containing

1.8 mM cholesterol). The culture was centrifuged and the pellet resuspended in 1 ml of Kirby solution (1% SDS, 0.2 M EDTA) with lysozyme (15 mg ml^{-1}) (Sigma). The mixes were incubated at room temperature for 5 min and then transferred to Falcon tubes were solid glass beads (150-212 μm diameter, Sigma) had previously being added. After the incubation, 200 μl of PCIA (phenol-chloroform-isoamyl alcohol acid) and 800 μl of buffer RLT (Qiagen) were added to each sample and several steps of 5 min agitation were performed. The cell lysates were recovered by centrifugation and the supernatant was added to new Falcon tubes with 700 μl of absolute ethanol (Merck). RNA was purified from the lysate using the Rneasy kit (Qiagen) according to the manufacturer's instructions. Quantity was measured using a NanoDrop (NanoDrop Technologies). RT-qPCR was performed as described previously (García-Fernández *et al*., 2017) using primers listed in Table S4.

Macroscopic studies

Bacterial clumping was analyzed by culturing the different strains in minimal medium with 18 mM glycerol devoid of detergents. Colony morphology was analyzed by plating approximately 100 colony forming units (CFU) of the parental and the mutant strains on 7H10 solid agar medium. Congo Red, an azo-dye with a well-known affinity for lipids and lipoproteins, was added to improve visualization of colony morphology alterations as described by Cangelosi *et al*. (1999). Plates were inspected after two days at 37 ºC either by naked eye or under a binocular scope at low magnification.

Cholesterol uptake assay

Cholesterol uptake was directly measured in resting cell suspensions. Cells were grown to mid-log phase on 18 mM glycerol, washed, and resuspended at a final cell density of A_{600} = 8 in 7H9 (minimal medium) buffer. Aliquots of 0.5 ml of the cells were placed in 1.5 ml vials, and 0.01 μCi ml^{-1} of [4-^{14}C] cholesterol (53 mCi/mmol, PerkinElmer Life Sciences) and 0.03 mM of non-labelled cholesterol were added. The vials were incubated for different times at 37 ºC. Cholesterol uptake was stopped by adding 5 ml of cold PBS buffer and collecting the cells immediately after on a 0.45-μm Millipore nitrocellulose filter (Fisher Scientific). Filtered cells were washed twice with 10 ml of PBS buffer with 5% Tween 20 (Fisher

Scientific) and 10 ml of 50% ethanol. The filters were dried and placed in Beckman Ready-Safe scintillation mixture (Beckman Coulter) and counted in a Beckman LS-600IC scintillation counter to determine the amount of cholesterol taken up by the cells. Where indicated, 2.0 mM dicyclohexylcarbodiimide (DCCD) was added to cell suspensions 10 min prior to the addition of the labeled cholesterol.

Bioinformatic analysis

The local sequence alignment was performed using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Homology search was carried out using pBLAST in NCBI database.

Acknowledgements

The technical work of A. Valencia is greatly appreciated. This work was supported by grants from the Ministry of Science and Innovation (BFU2009-11545-C03-03; BIO2012-39695-C02-01). JGF was supported by an FPI fellowship from the Spanish Ministry of Science and Innovation.

Conflict of interest

None declared

References

Cangelosi G.A., Palermo C.O., Laurent J.P., Hamlin A.M. and Brabant W.H. (1999) Colony morphotypes on Congo red agar segregate along species and drug susceptibility lines in the *Mycobacterium avium*-intracellular complex. Microbiology 145: 1317-1324.

Casali N. and Riley L. (2007). A phylogenomic analysis of the *Actinomycetales mce* operons. BMC Genomics 8: 60.

Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., *et al.* (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393: 537-544.

Dassa E. and Hofnung M. (1985) Sequence of gene *malG* in *E. coli* K12: Homologies between integral membrane components from binding protein-dependent transport systems. EMBO J 9: 2287–2293.

Dunphy K.Y., Senaratne R.H., Masuzawa M., Kendall L.V. and Riley L.W. (2010) Attenuation of *Mycobacterium tuberculosis* functionally disrupted in a fatty acyl-coenzyme A synthetase gene *fadD5*. J Infect Dis 201: 1232-1239.

Forrellad M.A., McNeil M., Santangelo M., Blanco F.C., García E., Klepp L.I., *et al*. (2014) Role of the Mce1 transporter in the lipid homeostasis of *Mycobacterium tuberculosis*. Tuberculosis 94: 170-177.

García-Fernández J., Papavinasasundaram K., Galán B., Sassetti C.M. and García J.L. (2017) Unravelling the pleiotropic role of the MceG ATPase in *Mycobacterium smegmatis*. Environ microbial doi: 10.1111/1462-2920.13771.

Jackson M., Reinaldo Camacho L., Gicquel B. and Guilhot C. (2001) Gene replacement and transposon delivery using the negative selection marker *sacB*. *Mycobacterium tuberculosis* Protocols. T. Parish and N. Stoker, Humana Press 54: 59-75.

Joshi S.M., Pandey A.K., Capite N., Fortune S.M., Rubin E.J. and Sassetti C.M. (2006) Characterization of mycobacterial virulence genes through genetic interaction mapping. Proc Natl Acad Sci USA 103: 11760-11765.

Kendall S.L., Withers M., Soffair C.N., Moreland N.J., Gurcha S., Sidders B., *et al*. (2007) A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. Mol Microbiol 65: 684-699.

Kitano H. (2004) Biological robustness. Nat Rev Genet 5: 826–837.

Klepp L.I., Forrellad M.A., Osella A.V., Blanco F.C., Stella E.J., Bianco M.V., *et al*. (2012) Impact of the deletion of the six *mce* operons in *Mycobacterium smegmatis*. Microbes Infect 14: 590-599.

Kumar A., Bose M. and Brahmachari V. (2003) Analysis of expression profile of mammalian cell entry (*mce*) operons of *Mycobacterium tuberculosis*. Infect Immun 71: 6083-6087.

Liu J., Barry III C.E., Besra G.S. and Nikaido H. (1996). Mycolic acid structure determines the fluidity of the mycobacterial cell wall. J Biol Chem 271: 29545-29551.

Marjanovic O., Iavarone A.T. and Riley L.W. (2011) Sulfolipid accumulation in *Mycobacterium tuberculosis* disrupted in the *mce2* operon. J Microbiol 49: 441-447.

Minnikin D.E. (1982) Lipids: Complex lipids, their chemistry, biosynthesis, and roles. The Biology of the Mycobacteria, eds. Ratledge C., Stanford J. (Academic, New York), pp 95–184.

Mohn W.W., van der Geize R., Stewart G. R., Okamoto S., Liu J., Dijkhuizen L. *et al.* (2008). The actinobacterial *mce4* locus encodes a steroid transporter. J Biol Chem 283: 35368-35374.

Mourez M., Hofnung M. and Dassa E. (1997) Subunit interactions in ABC transporters: A conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. EMBO J 16: 3066–3077.

Murphy K.C., Papavinasasundaram K. and Sassetti C.M. (2015) Mycobacterial recombineering. Methods Mol Biol 1285:177-199.

Niederweis M., Danilchanka O., Huff J., Hoffmann C. and Engelhardt H. (2010) Mycobacterial outer membranes: in search of proteins. Trends Microbiol 18: 109-116.

Pandey A. K. and Sassetti C. M. (2008) Mycobacterial persistence requires the utilization of host cholesterol. Proc Natl Acad Sci USA 105: 4376-4380.

Santangelo P., Klepp L., Nun J., Blanco F. C., Soria M., García-Pelayo M., *et al*. (2009) Mce3R, a TetR-type transcriptional repressor, controls the expression of a regulon involved in lipid metabolism in *Mycobacterium tuberculosis*. Microbiology 155: 2245–2255.

Santangelo M.P., Heuberger A., Blanco F., Forrellad M., Taibo C., Klepp L., *et al*. (2016) Metabolic profile of *Mycobacterium smegmatis* reveals Mce4 proteins are relevant for cell wall lipid homeostasis. Metabolomics 12: 97.

Slaytor M. and Bloch K. (1965) Metabolic transformation of cholestenediols. J Biol Chem 240: 4598-602.

Song H., Sandie R., Wang Y., Andrade-Navarro M.A. and Niederweis M. (2008) Identification of outer membrane proteins of *Mycobacterium tuberculosis*. Tuberculosis 88: 526–544.

Snapper S.B., Melton R.E., Mustafa S., Kieser T. and Jacobs W.R. Jr. (1990) Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. Mol Microbiol 4: 1911-1919.

Stover C.K., de la Cruz V.F., Fuerst T.R., Burlein J.E., Benson L.A., Bennett L.T., *et al*. (1991) New use of BCG for recombinant vaccines. Nature 351: 456-60.

Uhía I., Galán B., Kendall S.L., Stoker N.G. and García J.L. (2012). Cholesterol metabolism in *Mycobacterium smegmatis*. Environ Microbiol Reports 4: 168-182.

van der Geize R., Yam K., Heuser T., Wilbrink M. H., Hara H., Anderton M. C. *et al.* (2007) A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. Proc Natl Acad Sci USA 104: 1947-1952.

van der Geize R., de Jong W., Hessels G.I., Grommen A.W.F., Jacobs A.A.C. and Dijkhuizen L. (2008) A novel method to generate unmarked gene deletions in the intracellular pathogen *Rhodococcus equi* using 5-fluorocytosine conditional lethality. Nucleic Acids Res 36: e151.

van Kessel J.C. and Hatfull G.F. (2007) Recombineering in *Mycobacterium tuberculosis*. Nat Methods 4: 147-152.

Legends to the figures

Figure 1. (A) Schematic representation of the genomic region corresponding to the *mce4* operon in *M. smegmatis* mc²155, *M. tuberculosis* H37Rv and *R. jostii* RHAI. The percentage of identity in amino acids of each gene with respect to the ones found in *M. smegmatis* mc²155 is shown. The triangles located in the *Pyrb* area represent operator sequences recognized by the KstR repressor and the squares represent the putative -10 and -35 boxes. The new putative KstR operator sequences found in this work are highlighted in red **(B) Operator sequences recognized by KstR located in the P_{yrb} area of *M. smegmatis* mc²155, *M. tuberculosis* H37Rv and *R. jostii* RHAI.** The new operator sequences identified in this work are indicated with an asterisk. The bases identical to the consensus sequences are in bold while the ones different are underlined.

Figure 2. Differential expression of *mce4B and mce1A* genes. Expression of the *MSMEG_5899* (*mce4B*) and *MSMEG_0134* (*mce1A*) genes in exponential (lined bars) or in early stationary (dotted bars) growth phase. Transcription levels were measured using RT-qPCR as described in Experimental procedures. The values indicate the ratios of mRNA levels observed for the wild-type strain growing on cholesterol relative to glycerol, and normalized to *sigA* levels. Data represent averages of three

independent experiments, error bars indicate ± standard deviation and asterisks indicate that results are statistically significant (one-way ANOVA with a Student-Newman-Keuls test; $^{**}P < 0.01$; $^{***}P < 0.001$).

Figure 3. Growth analysis of mycobacterial strains. Growth curves of wild-type (pMV261) (black circles), Δmce4op (black triangles), wild-type (pmce4op) (grey circles) and Δmce4op (pmce4op) (grey triangles) strains on 1.8 mM cholesterol as the sole carbon and energy source and 37ºC (A) or 30ºC (B) or on 1.8 mM cholesterol plus 0.5 mM glycerol and 37ºC (C) or 30ºC (D). Growth was monitored by measuring the OD at 600 nm. Data represent averages of three independent experiments.

Figure 4. Radioactive cholesterol uptake by wild-type and Δmce4A strains. DCCD refers to measurements obtained in the presence of the ATPase inhibitor. Data represent averages of two independent experiments.

Figure 5. Phenotypic studies of the wild-type, Δmce4op and Δmce4op-C strains. (A) Aggregation assay in 18 mM glycerol medium devoid of detergents. (B) Colony morphology studies made in 7H10 plates containing 1 μg ml^{-1} Congo Red dye.

Figure 6. Analysis of the EExDA domains present in the YrbE4 permeases. (A) Schematics of the point mutations generated in the EExDA loop. (B) Growth of the ΔyrbE4AB mutant expressing the point mutated versions of the permeases in a replicating vector on 1.8 mM of cholesterol as the sole carbon source. Data represent averages of three independent experiments. (C) Aggregation studies of wild-type, ΔyrbE4AB, Δ01Δ02 (ΔyrbE4AB (pMVΔ01Δ02)), Δ01WT02 (ΔyrbE4AB (pMVΔ01WT02)), WT01Δ02 (ΔyrbE4AB (pMVWT01Δ02)) strains cultured on 18 mM glycerol containing medium devoid of any detergents.

Supplementary material

Table S1. Sequence comparisons of the YrbE4AB proteins. The percentage of identity in amino acids of the YrbE4A and YrbE4B proteins with respect to their homologues inter-operon in *M. smegmatis* mc^2155 is shown. The sequence coverage for each comparison is shown in brackets.

Table S2. Sequence comparisons of the Mce4 proteins. The percentage of identity in amino acids of each of the Mce4 proteins with respect to their homologues inter-operon (A) or intra-operon (B) in *M. smegmatis* mc^2155 is shown. The sequence coverage for each comparison is shown in brackets.

Table S3. Bacterial strains and plasmids used in this study.

Table S4. Primers used in this study.

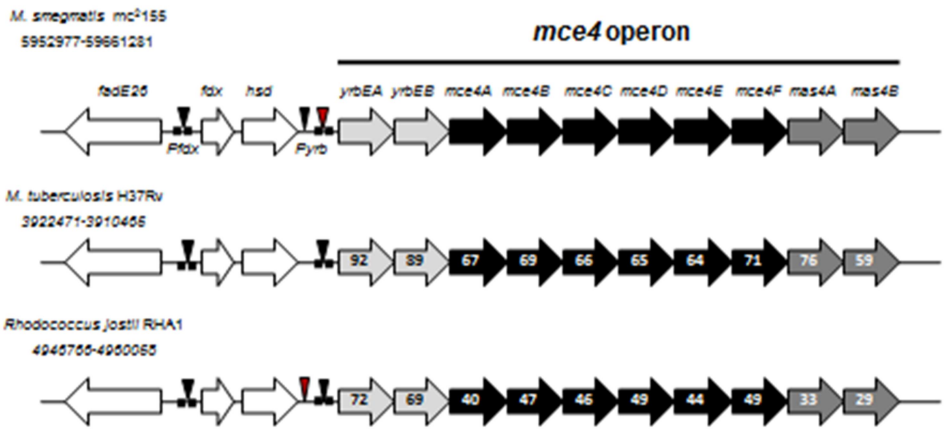
Figure S1. Genetic organization of the six *mce* operons found in *M. smegmatis* mc^2155

Figure S2. **Growth analysis of different mycobacterial strains.** Growth curves of the different mutant and complemented strains on 1.8 mM cholesterol as the sole carbon and energy source. Growth was monitored by measuring the OD at 600 nm. Data represent averages of three independent experiments.

Figure S3. Phenotypic studies of different mycobacterial strains. (A) Aggregation assay in 18 mM glycerol medium devoid of detergents. (B) Colony morphology studies made in 7H10 plates containing 1 μg ml^{-1} Congo Red dye.

Fig. 1

A.



1

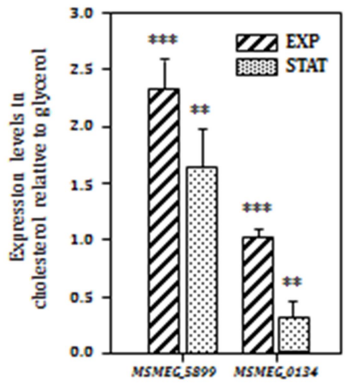
B.

Consensus sequence	tnnaacnngttnna
M. smegmatis mc²155	tagaacacgtttca
M. smegmatis mc²155 (-10)*	gcgaacttgttcta
R. jostii RHA1*	tcgaacg_gttcga
R. jostii RHA1 (-10)	gtgaacacgttcta
M. tuberculosis H37Rv (-10)	gcaaactcgttctg

2

3

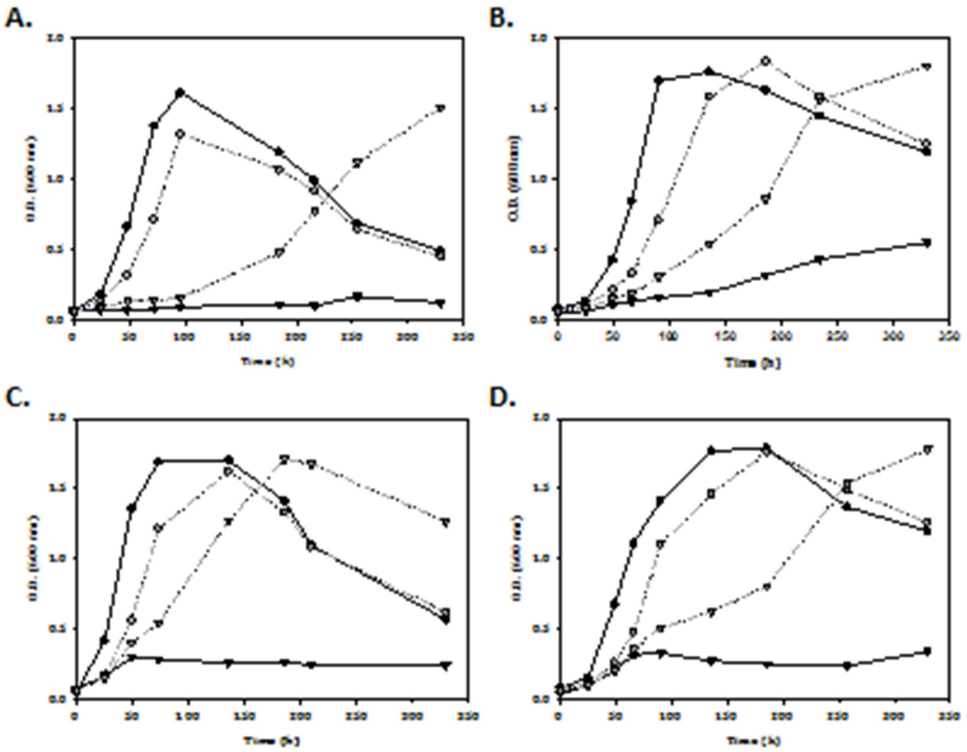
Fig. 2


Expression levels in cholesterol relative to glycerol
3.0
2.5
2.0
1.5
1.0
0.5
0.0
***
**
***
**
EXP
STAT
MSMEG_5899
MSMEG_0134

1

2

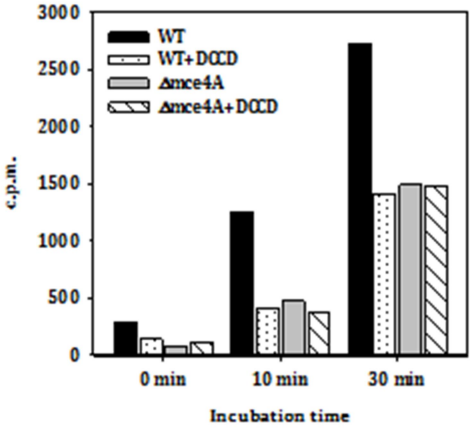
Fig. 3



1

2

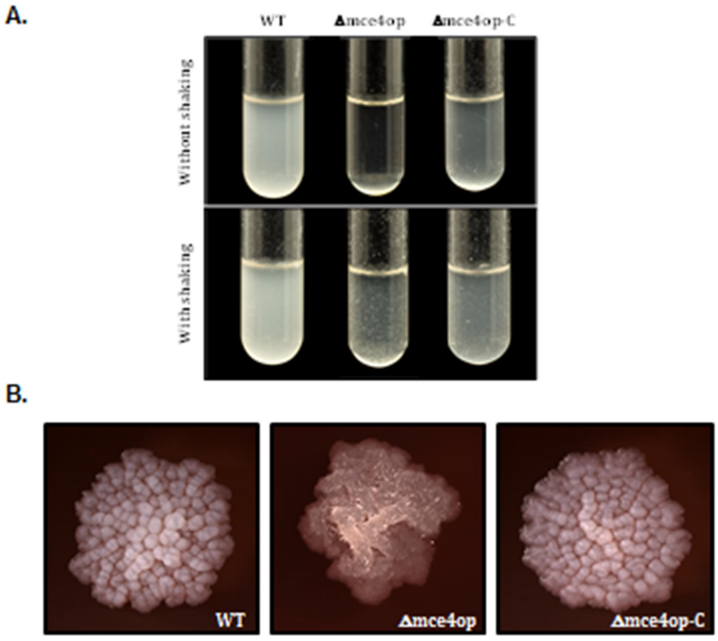
Fig. 4


3000
2500
2000
1500
1000
500
0
c.p.m.
WT
WT+DCCD
Δmce4A
Δmce4A+DCCD
0 min
10 min
30 min
Incubation time

1

2

Fig. 5


A.
WT
Δmce4op
Δmce4op-C
Without shaking
With shaking
B.
WT
Δmce4op
Δmce4op-C

1

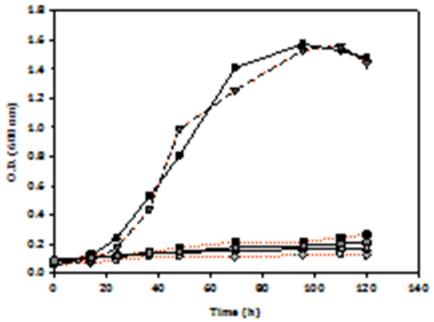
2

Fig. 6

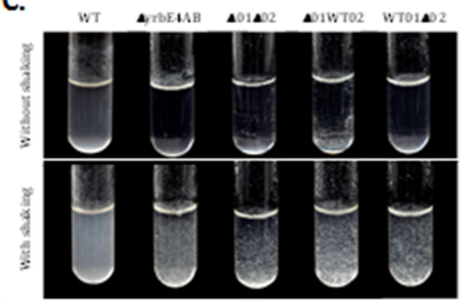
A.

M. smegmatis mc²155 strain name	YrbE4A permease sequence	YrbE4B permease sequence	Graph legend
Wild type	EExDA	EExDA	
ΔyrbE4AB	-	-	
ΔyrbE4AB (pMVyrbE4AB)	EExDA	EExDA	
ΔyrbE4AB (pMVΔ01Δ02)	AAxDA	AAxDA	
ΔyrbE4AB (pMVΔ01WT02)	AAxDA	EExDA	
ΔyrbE4AB (pMVWT01Δ02)	EExDA	AAxDA	

B.



C.



1

2

1