SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
BIOCHEMISTRY
PATHOGENS
GLYCOBIOLOGY
ISOENZYMES

Received
3 October 2011

Accepted
15 November 2011

Published
30 November 2011

Correspondence and requests for materials should be addressed to N.E. (numenius@cnc.uc.pt)

Mycobacterium tuberculosis Rv2419c, the missing glucosyl-3-phosphoglycerate phosphatase for the second step in methylglucose lipopolysaccharide biosynthesis

Vítor Mendes[1], Ana Maranha[1], Susana Alarico[1], Milton S. da Costa[1,2] & Nuno Empadinhas[1,2]

[1]CNC-Center for Neuroscience and Cell Biology, University of Coimbra, 3004-517 Coimbra, Portugal, [2]Department of Life Sciences, University of Coimbra, 3001-401 Coimbra, Portugal.

Mycobacteria synthesize intracellular methylglucose lipopolysaccharides (MGLP) proposed to regulate fatty acid synthesis. Although their structures have been elucidated, the identity of most biosynthetic genes remains unknown. The first step in MGLP biosynthesis is catalyzed by a glucosyl-3-phosphoglycerate synthase (GpgS, *Rv1208* in *Mycobacterium tuberculosis* H37Rv). However, a typical glucosyl-3-phosphoglycerate phosphatase (GpgP, EC3.1.3.70) for dephosphorylation of glucosyl-3-phosphoglycerate to glucosylglycerate, was absent from mycobacterial genomes. We purified the native GpgP from *Mycobacterium vanbaalenii* and identified the corresponding gene deduced from amino acid sequences by mass spectrometry. The *M. tuberculosis* ortholog (*Rv2419c*), annotated as a putative phosphoglycerate mutase (PGM, EC5.4.2.1), was expressed and functionally characterized as a new GpgP. Regardless of the high specificity for glucosyl-3-phosphoglycerate, the mycobacterial GpgP is not a sequence homolog of known isofunctional GpgPs. The assignment of a new function in *M. tuberculosis* genome expands our understanding of this organism's genetic repertoire and of the early events in MGLP biosynthesis.

Introduction

Tuberculosis remains a leading cause of death by infectious disease worldwide[1]. The growing incidence of *Mycobacterium tuberculosis* drug-resistant strains urges for a deeper understanding of the pathogen's biology and discovery of novel essential pathways that can grant new drug targets. Although the pathogen's genome has been decoded more than a decade ago, a function remains to be assigned to many of the genes[2].

Mycobacteria synthesize unusual polysaccharides containing α–(1,4)-linked methylated hexoses that are slightly hydrophobic and with an helical conformation typical of amylose[3]. One of these polysaccharides is the methylglucose lipopolysaccharide (MGLP) and the other is a methylmannose polysaccharide (MMP)[4,5]. MGLP contains 10–20 hexose units (two branching glucose units), some of which are acylated with acetate, propionate, isobutyrate, succinate or octanoate[4]; the mycobacterial MMP is linear, also has variable chain length and methylation, but does not seem to contain acyl groups[5]. These polysaccharides form stable 1 : 1 complexes with fatty acids and modulate the activity of fatty acid synthase I *in vitro*[6]. While MGLP has been found in both slow- and rapid-growing mycobacteria and also in *Nocardia otitidiscaviarum*, MMP has been detected in rapid-growing mycobacteria and in *Streptomyces griseus*[4,7–10].

MGLP contains glyceric acid linked to the first glucose forming glucosylglycerate (GG)[4,11]. The GG moiety is linked to a second glucose to form glucosyl–(1,6)–glucosylglycerate (diglucosylglycerate, DGG), which is attached to the glucose that initiates the main α–(1,4)–linked MGLP chain[7]. While the genes for the first step (*Rv1208*) and for the elongation (*Rv3032*), and possibly the gene for 6-*O*-methylation (*Rv3030*) of MGLP have been identified[12,13], those for the additional steps remain to be identified (Fig. 1). These genes were considered essential for *M. tuberculosis* growth rendering this biosynthetic pathway a promising target for drug development[14]. Free GG and DGG were detected in *M. smegmatis* and proposed to be the precursors for MGLP synthesis[15]. More recently, GG has also been identified in different organisms ranging from methanogenic

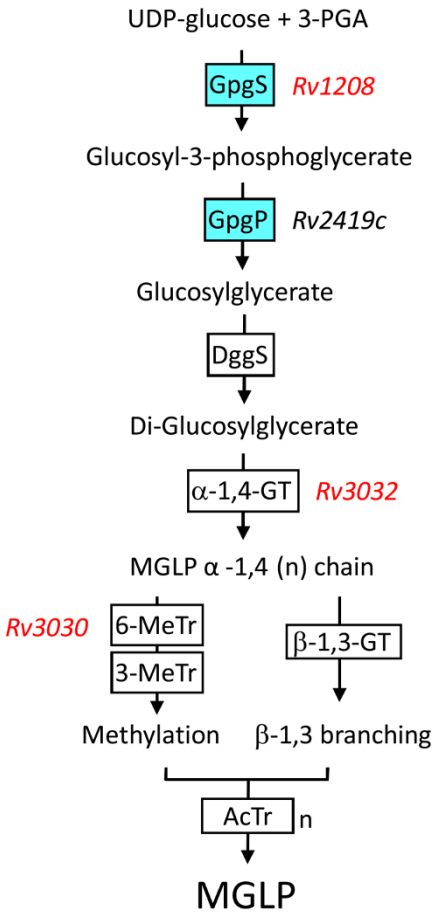

Figure 1 | Proposed pathway for the synthesis of the MGLP in *M. tuberculosis*. Confirmed activities are shaded in blue. White boxes indicate putative/deduced enzyme activities. Genes linked to the MGLP pathway by mutagenesis studies are indicated in red. GpgS, glucosyl-3-phosphoglycerate synthase; GpgP, glucosyl-3-phosphoglycerate phosphatase; DggS, di-glucosylglycerate synthase; GT, glucosyltransferases; MeTr, methyltransferases; AcTr, acyltransferases.

archaea and thermophilic bacteria to cyanobacteria and proteobacteria, where it serves as compatible solute during salt stress[16–18].

Glucosylglycerate can be synthesized by condensation of NDP-glucose and D-glycerate[19]. Alternatively, a phosphorylated intermediate (glucosyl-3-phosphoglycerate, GPG) is formed by a glucosyl-3-phosphoglycerate synthase (GpgS) and dephosphorylated by a glucosyl-3-phosphoglycerate phosphatase (GpgP)[20]. Mycobacterial GpgSs (Rv1208 in *M. tuberculosis* H37Rv) have very low sequence identity with GpgSs from the above mentioned organisms[12]. Since the *M. tuberculosis* gene had been considered essential for growth[14], we have determined the protein's three-dimensional structure, laying the molecular basis for structure-based drug design[21]. Given that GpgS synthesizes GPG, a phosphatase was deemed necessary to yield GG, the putative primer for MGLP synthesis. However, a gene for an archetypal GpgP was absent from mycobacterial genomes[20].

Herein, we report the purification of the native GpgP from *M. vanbaalenii* cell extracts and the identification of the corresponding gene. The *M. tuberculosis* GpgP gene (*Rv2419c*) was expressed in *E. coli* and the recombinant enzyme was purified and characterized. The assignment of a new function to a mycobacterial gene represents an important contribution into the knowledge of MGLP biosynthesis.

Results

Identification and sequence analyses of the mycobacterial GpgP (mGpgP). BLAST analyses with glucosyl-3-phosphoglycerate phosphatase (GpgP) or related mannosyl-3-phosphoglycerate phosphatase (MpgP) sequences showed no homologues in mycobacterial genomes[20,22]. To detect glucosyl-3-phosphoglycerate-dephosphorylating activity we tested cell-free extracts from two species of *Mycobacterium*. Unlike the *Mycobacterium smegmatis* extract that only partially dephosphorylated glucosyl-3-phosphoglycerate (GPG), that from *Mycobacterium vanbaalenii* completely dephosphorylated GPG to GG (results not shown) leading us to select this species for purification of the native GpgP. We performed chromatography to isolate fractions with GpgP activity and one of the purest fractions contained 8 protein bands that were analyzed by mass spectrometry (peptide mass fingerprinting) (Fig. 2A). One of the proteins below the 25 kDa standard was identified as a putative phosphoglycerate mutase (PGM, Mvan_3924) (Fig. 2A). Since PGMs are part of the large histidine phosphatase superfamily, which includes several phosphatases with different specificities[23], this protein was considered a likely GpgP candidate. The homolog from *M. tuberculosis* (*Rv2419c*) was expressed and functionally confirmed to be a glucosyl-3-phosphoglycerate phosphatase (GpgP, EC 3.1.3.70) since GPG was, by far, the preferred substrate (Table 1).

The *gpgP* gene from *M. vanbaalenii* contained 678 bp coding for a polypeptide with 225 amino acids with a calculated molecular mass of 24.2 kDa and a isoelectric point of 5.6, while the *M. tuberculosis gpgP* gene contained 672 bp encoding a protein with 223 amino acids with a calculated molecular mass of 24.2 kDa and a isoelectric point of 6.1. Gel filtration indicated that the recombinant His-tagged mGpgP behaved as a dimeric protein in solution, with a molecular mass of about 46.0 ± 2.8 kDa (results not shown).

mGpgP had close homologues in the available mycobacterial genomes and in other actinobacteria. BLAST analyses with the *M. tuberculosis* GpgP sequence revealed homologues in *M. bovis* (100% amino acid identity), *M. leprae* (86%), *M. marinum* (84%), *M. intracellulare* (84%), *M. ulcerans* (83%), *M. avium* (83%), *M. parascrofulaceum* (82%), *M. kansasii* (80%), *M. smegmatis* (77%), *M. vanbaalenii* (77%), *M. abscessus* (75%) and *M. gilvum* (73%) and also in *Rhodococcus equi* (62%), *Nocardia farcinica* (61%), *Corynebacterium diphtheriae* (49%) and *Streptomyces griseus* (42%). The amino acid identity of mGpgP (Rv2419c) with known GpgPs and MpgPs (EC 3.1.3.70), typically associated to the two-step pathways for synthesis of GG or MG[20,22], was negligible. Although the flanking genes have not been assigned to known functions, the genetic environment in the region containing the *gpgP* gene was highly conserved in mycobacterial genomes and in those of members of the genera *Rhodococcus*, *Nocardia* and *Corynebacterium* (Fig. 3).

Five paralogs with relevant homology to mGpgP (Rv2419c) were identified (Table 2); the C-terminal domain of Rv2228c (43% amino acid identity), Rv3214 (33%), Rv3837c (32%), Rv0525 (32%) and Rv0489 (30%), all contained the distinctive "RHG" motif of the histidine phosphatase superfamily[24–26] and some have assigned putative PGM functions. However, only the Rv0489 was structurally characterized as a PGM[26]. A sixth hypothetical PGM was annotated in the *M. tuberculosis* genome (Rv2135c) but no relevant homology with mGpgP was detected with the BLAST tool.

A phylogenetic tree constructed from the alignment between mGpgP and other known GpgPs/MpgPs shows that the mycobacterial protein groups within the histidine phosphatase superfamily cluster (Fig. 4A). The additional *M. tuberculosis* enzymes annotated as putative PGMs also cluster within this group while other known GpgPs cluster with the Haloacid dehalogenase-like hydrolases (HAD-like hydrolase) (Fig. 4A). However, mycobacterial GpgPs exhibit the "RHG" motif of the histidine phosphatase superfamily (Fig. 4B).

Properties of the recombinant mGpgP from *M. tuberculosis*. Expression of the *M. tuberculosis gpgP* gene in *E. coli* resulted in high level production of recombinant His-tagged protein, which

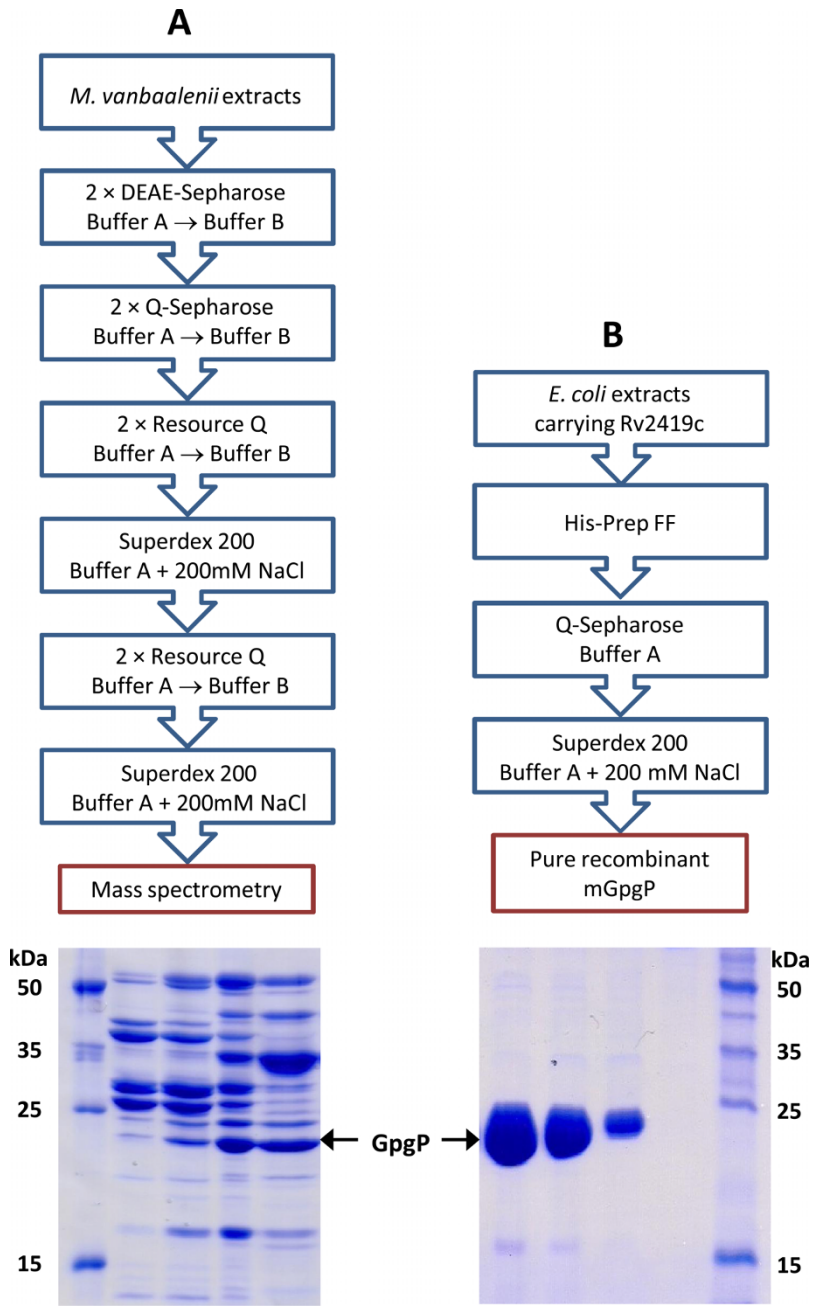

Figure 2 | Flow diagrams of the purification processes for the native and recombinant GpgPs. (A) Purification of the native GpgP from *M. vanbaalenii* and SDS-PAGE gel with the purest fractions obtained. (B) Purification from *E. coli* extracts of the *M. tuberculosis* recombinant GpgP and SDS-PAGE gel with the pure recombinant enzyme (three consecutive fractions eluting from the final step).

was purified to homogeneity (Fig. 2B). Among the compounds tested as possible substrates for the recombinant mGpgP, GPG was by far the preferred substrate. Lower activity could also be measured with MPG (<10%) and MGPG (<5%), while the activity with *p*-nitrophenyl phosphate was extremely low (Table 1, Fig. 5A). *Mycobacterium tuberculosis* mGpgP exhibited Michaelis-Menten kinetics at 37°C with a K_m for GPG of 0.35±0.03 mM and V_{max} of 67.2±1.4 μmol/min.mg protein (Table 3). A comparison between mGpgP and other isofunctional GpgPs (HAD-like) is indicated in Table 3[20,27,28]. mGpgP showed only trace PGM activity (0.96±0.08 $\mu mol/min^{-1}mg^{-1}$ protein) in the 3-PGA to 2-PGA direction representing <2% of the activity measured with GPG (Table 1, Fig. 5A), while the human PGM converted 3-PGA to 2-PGA with a specific activity of 23.0±1.5 μmol/min.mg protein. No PGM activity was detected in the 2-PGA to 3-PGA direction (Table 1, Fig. 5A).

The recombinant mGpgP was active between 20 and 50°C, with maximal activity at about 45°C (Fig. 5B). At 37°C, the enzyme was active between pH 4.0 and 10.0, optimally at pH 7.0 (Fig. 5C). mGpgP activity was not dependent on divalent cations , although the addition of 5 mM EDTA reduced the activity to about 70% of maximal rate. There were no relevant differences in mGpgP activity with Mg^{2+} (1–10 mM), Mn^{2+} (1–10 mM), Co^{2+} (1–5 mM) or water. However, higher Co^{2+} concentrations (10–50 mM) progressively inhibited enzyme activity, and Cu^{2+} was inhibitory at any concentration tested (1–50 mM). The enzyme retained full activity after 1 week storage on ice and at −20°C in 50 mM BTP with 50 mM NaCl (pH 7.0).

Discussion

The structural complexity and different linkages between hexoses in the mycobacterial methylglucose lipopolysaccharide (MGLP) require

Table 1 | Substrate specificity of the recombinant GpgP (Rv2419c) from *M. tuberculosis*

Substrate	Specific activity (μmol/min.mg protein)
Dephosphorylating activity	
Glucosyl-3-phosphoglycerate (GPG)	55.8±3.7
Mannosyl-3-phosphoglycerate (MPG)	5.2±0.6
Mannosylglucosyl-3-phosphoglycerate (MGPG)	2.8±0.1
3-phosphoglycerate (3-PGA)	0
2-phosphoglycerate (2-PGA)	0
p-nitrophenyl phosphate (pNPP)	0.02±0.01
Glucose-6-phosphate (G6P)	0
Trehalose-6-phosphate (T6P)	0
PGM activity	
3-PGA to 2-PGA	0.96±0.08
2-PGA to 3-PGA	0
PGM (from human muscle)	
3-PGA to 2-PGA	23.0±1.5

different enzymes for its biosynthesis[7,8,15,29]. Although most of the genes involved in this pathway remain unknown, the *M. tuberculosis* H37Rv gene (*Rv3032*) for the α-(1,4)-glycosyltransferase required for the elongation of MGLP, and the methyltransferase gene (*Rv3030*) for 6-*O*-methylations have been identified (Fig. 1)[13]. It was also suggested that other genes in the *Rv3030-Rv3037c* cluster including putative glycoside hydrolase and acyltransferase genes might catalyze additional steps in the pathway[13,30]. We have recently identified the gene (*Rv1208*) for the first step in MGLP biogenesis and characterized the corresponding glucosyl-3-phosphoglycerate synthase (GpgS)[12,21]. Biochemical evidence was supported by genetic confirmation of the involvement of *Rv1208* in the MGLP pathway[31]. This gene was located within a cluster (*Rv1208-Rv1213*) containing a α-(1,4)-glycosyltransferase gene (*Rv1212c*) likely involved in the elongation of MGLP, and a NDP-glucose pyrophosphorylase gene (*Rv1213*) that may supply NDP-glucose for polymerization[13,32].

Since GpgS synthesizes glucosyl-3-phosphoglycerate (GPG), the existence of a phosphatase to convert this phosphorylated precursor into GG was anticipated[12,30]. In most organisms that synthesize GPG, the dephosphorylation is carried out by a specific glucosyl-3-phosphoglycerate phosphatase (GpgP, EC 3.1.3.70) encoded by a gene normally contiguous or located at close distance from the GpgS gene[20,27]. Despite the fact that no typical GpgP could be detected in mycobacterial genomes, *M. vanbaalenii* extracts completely dephosphorylated GPG allowing the purification of a highly specific GpgP, which was designated mGpgP (mycobacterial GpgP). The corresponding gene (*Rv2419c* in *M. tuberculosis*) was annotated as a putative phosphoglycerate mutase (PGM) for the interconversion of 3-phosphoglycerate and 2-phosphoglycerate, but our results indicate that the enzyme retained only residual PGM activity. Phosphoglycerate mutases are part of a large group of functionally diverse enzymes included in the histidine phosphatase superfamily, as they share a conserved catalytic motif "RHG" where the histidine plays a crucial role in catalysis, being phosphorylated during the reaction[23]. Although PGMs are the best studied members of this superfamily, phosphatases largely dominate the group, which now also includes mGpgP with a confirmed function. Interestingly, all GpgPs characterized to date share conserved motifs of the haloacid dehalogenase-like hydrolase superfamily (HAD-like hydrolase)[20,27,33] and lack the "RHG" motif found in mGpgP[23]. HAD-like phosphatases, which also include mannosyl-3-phosphoglycerate phosphatases (MpgP, EC 3.1.3.70) that are GpgP homologues, are also sometimes referred to as "DDDD" phosphohydrolases due to the presence of four invariant aspartate residues[33,34]. Structural studies have attributed PGM function to Rv0489[26], which shares ~30% amino acid identity with Rv2419c (mGpgP). Four other proteins in *M. tuberculosis* have relevant amino acid identity to Rv0489 and to Rv2419c and all share the "RHG" motif. One of these paralogs, Rv3214, was annotated as a putative PGM but has recently been shown to be an acid phosphatase without PGM activity, although the specific substrate has not been identified[25]. While function based on sequence similarity alone is not a reliable approach to the unambiguous identification of enzyme activity, histidine phosphatases present additional challenges to automated genome annotation as they are remarkably diverse[23,35].

The mycobacterial GpgP had homologues in other actinobacterial genomes, namely in *Rubrobacter xylanophilus*, so far the only actinobacterium known to accumulate the GG-related solute mannosylglycerate (MG)[36]. In this organism, an actinobacterial-type MpgS/GpgS with dual specificity could synthesize the phosphorylated precursors for both MG and GG *in vitro* even though neither GG nor MGLP have, so far, been detected *in vivo*[37]. A classical MpgP (HAD-like) involved in MG synthesis in hyper/thermophiles, was also absent from this organism's genome[22,28]. The present study allowed the detection of two mGpgP homologues in this organism, which may dephosphorylate mannosyl-3-phosphoglycerate (MPG) to MG.

Several of the above MpgPs were functionally interchangeable as they efficiently dephosphorylated MPG and GPG alike[20], while mGpgP preferentially hydrolyzed GPG. Moreover, the mycobacterial GpgS was also specific for glucose donors[12]. Since both enzymes participate in the same pathway, it was not surprising that the temperature and pH profiles for the *M. tuberculosis* GpgS and GpgP activities *in vitro* were comparable. However, GpgS activity was absolutely dependent on magnesium ions, while mGpgP activity was independent of cations, unlike all known HAD-like GpgPs/MpgPs that require divalent cations for activity[12,20,22,27,28]. The GpgS for the first step in the MGLP pathway was predicted to be essential for *M.*

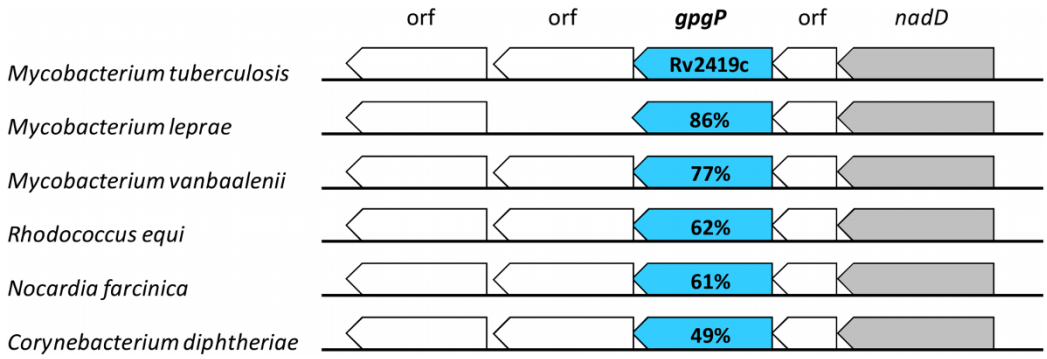

Figure 3 | Genetic context of the *gpgP* genes in mycobacteria and closely related actinobacteria. *gpgP* – glucosyl-3-phosphoglycerate phosphatase gene. *nadD* – probable nicotinate-nucleotide adenylyltransferase gene. orf, open reading frame of unknown function. Amino acid identity between GpgP homologues is indicated.

Table 2 | GpgP (*Rv2419c*) paralogs in *Mycobacterium tuberculosis* H37Rv genome

Gene	Protein Length (aa)	Identity (%)	Function	Reference
Rv2419c	223	100	GpgP	This study
Rv2228c	364	43 (in 119 aa)	Putative Rnase HI/α-ribazole phosphatase	24
Rv3214	203	33 (in 184 aa)	Acid phosphatase (pNPP)	25
Rv0489	249	30 (in 194 aa)	1,3BPG + 3PGA = 2,3BPG + 3PGA	26
Rv3837c	232	32 (in 172 aa)	Possible 3PGA = 2PGA	-
Rv0525	202	32 (in 123 aa)	Unknown	-

tuberculosis growth, while the mGpgP was not[14]. This suggests that one of the above Rv2419c paralogs, or a phosphatase with broader specificity, could replace mGpgP activity. However, we did not detect an alternative dephosphorylating activity in *M. vanbaalenii* extracts during purification.

The Rv2419c sequence motifs place this novel mGpgP in the histidine phosphatase superfamily. The high specificity for GPG points toward convergent evolution with the non-homologous isofunctional GpgPs of the HAD-like hydrolase superfamily, indicating that these types of enzymes have emerged in evolution more than once[38].

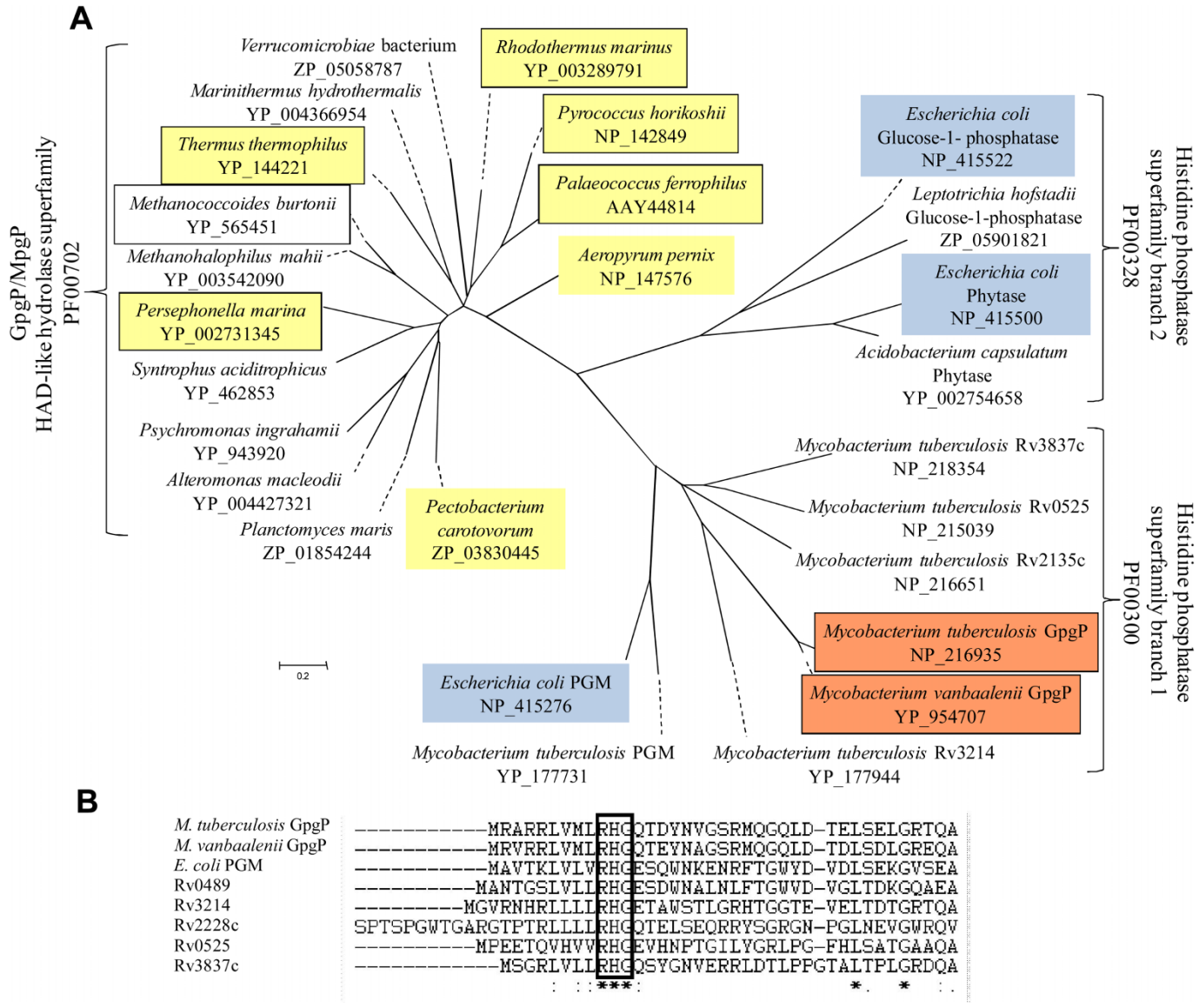

Figure 4 | (A) Unrooted phylogenetic tree based on the amino acid sequences of identified and putative GpgPs/MpgPs (EC 3.1.3.70), PGMs (EC 5.4.2.1) and of other enzymes of the histidine phosphatase superfamily. Organisms where GG (or MG) has been detected are shaded in yellow. Organisms with enzymes shown to dephosphorylate GPG (or MPG) are boxed. *E. coli* enzymes with confirmed function are highlighted in blue. The mycobacterial GpgPs studied in this work are highlighted in red. Peptide accession numbers (NCBI) are indicated. Scale bar, 0.2 changes per site. (B) Alignment of the N-terminal amino acid sequences of the PGM from *E. coli*, the GpgPs from *M. vanbaalenii* (Mvan_3924) and from *M. tuberculosis* (Rv2419c) and its paralogs. The typical "RHG" motif of the histidine phosphatase superfamily is boxed.

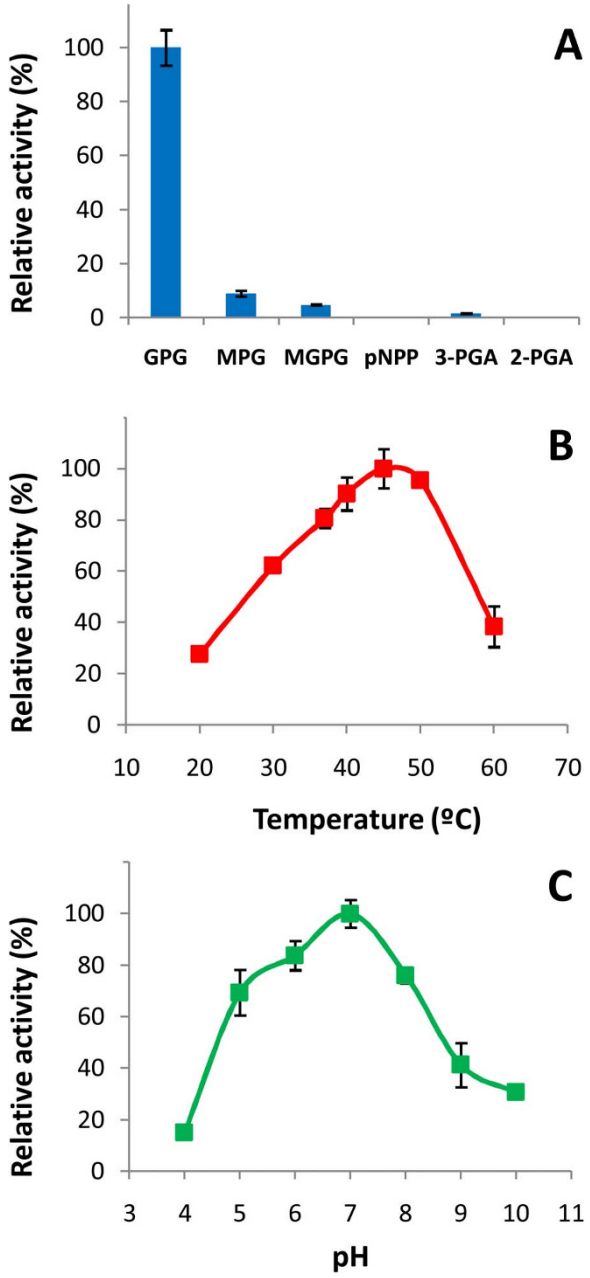

Figure 5 | (A) Substrate preference of the recombinant GpgP from *M. tuberculosis*, (B) Temperature profile and (C) pH dependence. GPG, MPG, MGPG and pNPP were tested as substrates for dephosphorylation; 3-PGA and 2-PGA were tested as substrates for PGM activity.

Ongoing structural studies will add further insights on the catalytic mechanism and evolution of both types of GpgPs.

The apparent ubiquity of MGLP in pathogenic mycobacteria and the predicted essentiality of some of the putative biosynthetic genes[14] suggest a fundamental role for this structure in mycobacterial physiology and that its absence or deficient assemblage could lead to cell death. Here, we provide biochemical evidence for the second step in MGLP biosynthesis[12,30]. The identification of this function in the genome of *M. tuberculosis* with the functional assignment of Rv2419c to GpgP, not only expands the already broad activity range of the histidine phosphatase superfamily, but also furthers our understanding of the MGLP biosynthetic pathway, which might help us devise new strategies to fight tuberculosis.

Methods

Bacterial strains and growth conditions. *Mycobacterium smegmatis* mc^{2}155 (ATCC 700084) was obtained from LGC Standards S.L.U. (Barcelona, Spain). *Mycobacterium vanbaalenii* (DSM 7251) was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (Germany). *M. smegmatis* and *M. vanbaalenii* biomass for enzymatic assays was obtained from 1-L cultures grown with a glycerol-based medium in an orbital shaker, at 30°C for 24 h. The medium contained per litre: 20 g glycerol (Merck), 5 g casaminoacids (Difco), 1 g fumaric acid (Sigma), 1 g K_2HPO_4, 0.3 g $MgSO_4$, 0.02 g $FeSO_4$ and 2 g Tween 80 (Sigma), at pH 7.0[39]. *Mycobacterium vanbaalenii* biomass for purification of the native glucosyl-3-phosphoglycerate phosphatase (GpgP) was obtained from an 18-L culture under the conditions described above. *Escherichia coli* BL-21 (Novagen) was grown at 37°C in LB medium at pH 7.0 in a 5-L fermentor with continuous aeration and stirred at 180 rpm. Kanamycin was added at a final concentration of 30 μg/mL.

Preparation of cell-free extracts. Cells were harvested by centrifugation (8000 × *g*, 10 min, 4°C) during the late exponential phase of growth and the pellet suspended in 20 mM Bis-tris propane buffer (BTP) (Molekula) at pH 7.5, containing 5 mM $MgCl_2$, 10 μg/ml DNAse I and protease inhibitors (Roche). Cells were disrupted twice in a French-press followed by centrifugation to remove debris (15000 × *g*, 30 min, at 4°C). For the analysis of GpgP activity, cell extracts were dialysed overnight against 50 mM BTP at pH 7.5.

GPG synthesis. Glucosyl-3-phosphoglycerate (GPG) was synthesized after overnight reaction of the pure recombinant *M. smegmatis* GpgS with a mixture containing 20 mM (each) of UDP-glucose, 3-PGA and $MgCl_2$ in 50 mM BTP, pH 7.5 at 37°C[12]. The concentration of GPG was determined as previously described[20]. The synthesis of GPG was monitored by thin-layer chromatography (TLC) with the solvent system ethyl acetate/acetic acid/water/ammonia 25% (6 : 6 : 2 : 1, v/v). Glucosylglycerate (GG), UDP–glucose and D-glucose standards were used for comparative purposes.

Enzyme assays. The activities of the GpgPs from *M. smegmatis* and from *M. vanbaalenii* in crude extracts and during purification, and the activity of the recombinant GpgP from *M. tuberculosis*, were detected with reaction mixtures (25 μl) containing 2.0 mM of GPG and 2 mM $MgCl_2$ in 20 mM BTP, pH 7.0, after 10 min incubations at 30°C and 37°C, respectively. The hydrolysis of GPG was monitored by TLC with a solvent system composed by chloroform/methanol/acetic acid/water (30 : 50 : 8 : 4, v/v). Standard GG and GPG were used for comparative purposes. Protein concentration was determined by the Bradford method[40].

Purification of the native GpgP from *Mycobacterium vanbaalenii*. *Mycobacterium vanbaalenii* cell extracts were loaded onto a DEAE-Sepharose column equilibrated with 20 mM Bis-Tris propane buffer (BTP) at pH 7.5 (Buffer A) (Fig. 2A). Elution was carried out with a linear gradient of 0 to 1 M NaCl and GpgP activity was determined by TLC as described above. The active fractions were pooled, concentrated by centrifugation in 30 kDa cutoff centricons (Amicon), loaded onto a DEAE-Sepharose equilibrated with 20 mM BTP, pH 6.5 (Buffer B) and eluted by a linear gradient of 0 to

Table 3 | Properties of known GpgPs and one MpgP (with GpgP activity). The *M. tuberculosis* GpgP belongs to the histidine phosphatase superfamily (His-Phos) while the other proteins belong to the HAD-like hydrolase superfamily

GpgP activity	Optimum pH	[a]Optimum temperature	K_m GPG (mM)[b]	V_{max} (μmol/min.mg)[b]	Cation dependence	Reference
Mycobacterium tuberculosis (His-Phos)	7.0	45°C	0.35	67.2	Independent	This study
Methanococcoides burtonii (HAD-like)	6.0	50°C	0.08	0.10	Co^{2+}>Mg^{2+}>Ni^{2+}	20
Persephonella marina (HAD-like)	7.0	85°C	0.36	97.0	Mg^{2+}>Co^{2+}>Mn^{2+}	27
Thermus thermophilus (HAD-like)	6.0	95°C	0.83	46.7	Co^{2+}>Mg^{2+}>Mn^{2+}	20, 28

[a]Temperatures at which GpgPs show maximal activity.
[b]K_m and V_{max} values were determined at each organism's optimum growth temperature (*M. tuberculosis*, 37°C; *M. burtonii*, 30°C; *P. marina*, 70°C; *T. thermophilus*, 70°C).

0.7 M NaCl. Active fractions were pooled and diluted 10-fold with buffer A, loaded onto a Q-Sepharose column equilibrated with the same buffer followed by a linear gradient of 0 to 0.7 M NaCl and by a second Q-Sepharose column equilibrated with buffer B and a linear gradient of 0 to 0.5 M NaCl. Desalted samples were loaded onto two sequential Resource-Q columns, equilibrated with buffer A and buffer B, respectively, and eluted with linear gradients of 0 to 0.4 M NaCl. Active fractions were equilibrated with buffer A containing 200 mM NaCl and loaded onto a Superdex 200 column equilibrated with the same buffer. Two additional Resource-Q columns and a final Superdex 200 column were used as final purification steps as described above (Fig. 2A). GpgP activity was assessed by TLC and fraction purity was determined by SDS-PAGE (Fig. 2A). Each fraction (50 µl) was frozen and stored at −20°C for protein identification by mass spectrometry as described below.

Identification and sequence analyses of the mycobacterial *gpgP* gene. Each of the *M. vanbaalenii* purest fractions (50 µl) obtained after the last purification step were loaded onto a large SDS-PAGE gel (20 cm×20 cm) and separated under constant voltage (100 V). A total of 8 bands were excised for peptide mass fingerprinting analyses (Proteomics Unit, Biocant, Cantanhede). Each of the peptide sequences obtained from the 8 bands was analyzed with the BLAST tool at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The ClustalX2 program (http://www.clustal.org) was used for sequence alignments and the MEGA4 program (www.megasoftware.net) was used to generate the phylogenetic trees.

Cloning and functional overexpression of *gpgP* from *M. tuberculosis* in *E. coli*. The *M. tuberculosis Rv2419c* (*gpgP*) gene sequence was optimized for expression in *E. coli* and synthesized (GenScript Corp.). The artificial gene was cloned between the NdeI and HindIII restriction sites of pET30a (Novagen) and transformed into *E. coli* BL21 using standard molecular biology procedures. Recombinant *E. coli* were grown to mid-exponential phase of growth (OD_{610}=0.8); IPTG (Zymo Research) was added at a final concentration of 0.5 mM to induce gene expression and the temperature was lowered to 20°C. Cells were harvested 18 h later by centrifugation (8000×*g*, 10min, 4°C).

Purification of the recombinant mGpgP from *E. coli* extracts. The His-tagged recombinant GpgP from *M. tuberculosis* was purified with a pre-packed Ni-Sepharose high-performance column (His-Prep FF 16/10) equilibrated with 20 mM sodium phosphate, pH 7.4, 0.5 M NaCl and 20 mM imidazole (Fig. 2B). Elution was carried out with 500 mM imidazole and the purity of the fractions was determined by SDS-PAGE. The purest active fractions were pooled, concentrated, equilibrated with Buffer A and loaded onto a Q-Sepharose column equilibrated with the same buffer and eluted by a linear gradient of NaCl (0 to 1 M). Purity was determined by SDS-PAGE and the purest active fractions were pooled, concentrated, equilibrated with Buffer A with 200 mM NaCl, and loaded onto a Superdex 200 column equilibrated with the same buffer (Fig. 2B). After SDS-PAGE analysis (Fig. 2B), the pure active fractions were concentrated and equilibrated with 50 mM BTP, pH 7.0 with 50 mM NaCl. Protein content of the samples was determined by the Bradford assay[40].

Characterization of the recombinant mGpgP from *M. tuberculosis*. The substrate specificity of the recombinant GpgP was determined using 2-phosphoglycerate, 3-phosphoglycerate, 2,3-bisphosphoglycerate, 3-glycerol phosphate, phosphoenolpyruvate, α-glucose-1-phosphate, β-glucose-1-phosphate, glucose-6-phosphate, glucose-1,6-diphosphate, mannose-1-phosphate, mannose-6-phosphate, fructose-6-phosphate, trehalose-6-phosphate, ATP, UTP and ADP (all from Sigma-Aldrich), glucosyl-3-phosphoglycerate (GPG), mannosyl-3-phosphoglycerate (MPG) and mannosyl-(1,2)-glucosyl-3-phosphoglycerate (MGPG). GPG, MPG and MGPG were synthesized as previously described[41]. The general phosphatase substrate *p*-nitrophenyl phosphate (Sigma-Aldrich) was also tested as a possible substrate for GpgP.

Reaction mixtures (25 µl) containing pure recombinant GpgP (0.025 µg), 3 mM of substrate and 2.5 mM $MgCl_2$ in 20 mM BTP at pH 7.0, were incubated at 50°C for 10 min. The reaction products were visualized by TLC as described above; the dephosphorylation of 2-phosphoglycerate (2-PGA), 3-phosphoglycerate (3-PGA), 2,3-bisphosphoglycerate, 3-glycerol-phosphate and phosphoenolpyruvate was quantified by the release of free phosphate measured by the Ames method[42]. The *p*-nitrophenol released was determined at 405 nm as previously described and alkaline phosphatase was used as positive control[43].

To assess whether the recombinant GpgP retained PGM activity, both in the forward (3-PGA to 2-PGA) or reverse (2-PGA to 3-PGA) directions, reactions were carried out as previously described[44]. A recombinant PGM from human muscle (Affymetrix) was used as positive control. Phosphoglycerate mutase activities were monitored using a standard method coupled to NADH oxidation[45].

Temperature, pH and cation dependence profiles were determined by the addition of known amounts of pure GpgP to 25 µl reaction mixtures containing the appropriate buffer and 2 mM GPG, and stopped at different times by cooling on ethanol-ice. Free phosphate was quantified using the Ames reaction[42]. The effect of cations was examined by incubation of the reaction mixtures with the chloride salts of Mg^{2+}, Mn^{2+}, Co^{2+} (1 to 10 mM), with 5 mM EDTA or with no additions, at 37°C. The temperature profile was determined between 20 and 60°C in 20 mM BTP, pH 7.0. The effect of pH was determined at 37°C in 20 mM acetate buffer (pH 4.0 and 5.0), 20 mM BTP (pH 6.0 to 9.0) and 20 mM CAPS buffer (pH 10.0).

The K_m value for GPG was determined at 37°C as follows: the reactions were initiated by the addition of the *M. tuberculosis* GpgP to reaction mixtures with 50 mM BTP at pH 7.5, containing GPG (0.25–20.0 mM) and stopped at different times. The phosphate released was measured as described above and the kinetic parameters K_m and V_{max} were calculated from Lineweaver–Burk plots. All experiments were performed in triplicate.

The molecular mass of the recombinant GpgP was estimated by gel filtration on a Superdex 200 column and molecular mass standards used were albumin (67 kDa), aldolase (158 kDa), catalase (232 kDa), ferritin (440 kDa). Blue dextran 2000 was used to determine the void volume (GE Healthcare).

1. Dye, C. Doomsday postponed? Preventing and reversing epidemics of drug-resistant tuberculosis. *Nat Rev Microbiol* **7**, 81–87 (2009).
2. Cole, S. T., *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 (1998).
3. Yabusaki, K. K. & Ballou, C. E. Interaction of mycobacterial polymethylpolysaccharides with paranaric acid and palmitoyl-coenzyme A: structural specificity and monomeric dissociation constants. *Proc Natl Acad Sci U S A* **75**, 691–695 (1978).
4. Smith, W. L. & Ballou, C. E. The 6-*O*-methylglucose-containing lipopolysaccharides of *Mycobacterium phlei*. Locations of the neutral and acidic acyl groups. *J Biol Chem* **248**, 7118–7125 (1973).
5. Maitra, S. K. & Ballou, C. E. Heterogeneity and refined structures of 3-*O*-methyl-D-mannose polysaccharides from *Mycobacterium smegmatis*. *J Biol Chem* **252**, 2459–2469 (1977).
6. Ilton, M. *et al.* Fatty acid synthetase activity in *Mycobacterium phlei*: regulation by polysaccharides. *Proc Natl Acad Sci U S A* **68**, 87–91 (1971).
7. Tuffal, G., Albigot, R., Riviere, M. & Puzo, G. Newly found 2-N-acetyl-2,6-dideoxy-β-glucopyranose containing methyl glucose polysaccharides in *M. bovis* BCG: revised structure of the mycobacterial methyl glucose lipopolysaccharides. *Glycobiology* **8**, 675–684 (1998).
8. Tuffal, G., Ponthus, C., Picard, C., Riviere, M. & Puzo, G. Structural elucidation of novel methylglucose-containing polysaccharides from *Mycobacterium xenopi*. *Eur J Biochem* **233**, 377–383 (1995).
9. Pommier, M. T. & Michel, G. Isolation and characterization of an *O*-methylglucose-containing lipopolysaccharide produced by *Nocardia otitidis-caviarum*. *J Gen Microbiol* **132**, 2433–2441 (1986).
10. Harris, L. S. & Gray, G. R. Acetylated methylmannose polysaccharide of *Streptomyces*. *J Biol Chem* **252**, 2470–2477 (1977).
11. Saier, M. H. Jr. & Ballou, C. E. The 6-*O*-methylglucose-containing lipopolysaccharide of *Mycobacterium phlei*. Identification of D-glyceric acid and 3-*O*-methyl-D-glucose in the polysaccharide. *J Biol Chem* **243**, 992–1005 (1968).
12. Empadinhas, N., Albuquerque, L., Mendes, V., Macedo-Ribeiro, S. & da Costa, M. S. Identification of the mycobacterial glucosyl-3-phosphoglycerate synthase. *FEMS Microbiol Lett* **280**, 195–202 (2008).
13. Stadthagen, G. *et al.* Genetic basis for the biosynthesis of methylglucose lipopolysaccharides in *Mycobacterium tuberculosis*. *J Biol Chem* **282**, 27270–27276 (2007).
14. Sassetti, C. M., Boyd, D. H. & Rubin, E. J. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**, 77–84 (2003).
15. Kamisango, K., Dell, A. & Ballou, C. E. Biosynthesis of the mycobacterial *O*-methylglucose lipopolysaccharide. Characterization of putative intermediates in the initiation, elongation, and termination reactions. *J Biol Chem* **262**, 4580–4586 (1987).
16. Klahn, S., Steglich, C., Hess, W. R. & Hagemann, M. Glucosylglycerate: a secondary compatible solute common to marine cyanobacteria from nitrogen-poor environments. *Environ Microbiol* **12**, 83–94 (2010).
17. Goude, R., Renaud, S., Bonnassie, S., Bernard, T. & Blanco, C. Glutamine, glutamate, and α-glucosylglycerate are the major osmotic solutes accumulated by *Erwinia chrysanthemi* strain 3937. *Appl Environ Microbiol* **70**, 6535–6541 (2004).
18. Robertson, D. E., Lai, M. C., Gunsalus, R. P. & Roberts, M. F. Composition, Variation, and Dynamics of Major Osmotic Solutes in *Methanohalophilus* Strain FDF1. *Appl Environ Microbiol* **58**, 2438–2443 (1992).
19. Fernandes, C., Empadinhas, N. & da Costa, M. S. Single-step pathway for synthesis of glucosylglycerate in *Persephonella marina*. *J Bacteriol* **189**, 4014–4019 (2007).
20. Costa, J. *et al.* Characterization of the biosynthetic pathway of glucosylglycerate in the archaeon *Methanococcoides burtonii*. *J Bacteriol* **188**, 1022–1030 (2006).
21. Pereira, P. J. *et al. Mycobacterium tuberculosis* glucosyl-3-phosphoglycerate synthase: structure of a key enzyme in methylglucose lipopolysaccharide biosynthesis. *PLoS ONE* **3**, e3748 (2008).
22. Empadinhas, N., Marugg, J. D., Borges, N., Santos, H. & da Costa, M. S. Pathway for the synthesis of mannosylglycerate in the hyperthermophilic archaeon *Pyrococcus horikoshii*. Biochemical and genetic characterization of key enzymes. *J Biol Chem* **276**, 43580–43588 (2001).
23. Rigden, D. J. The histidine phosphatase superfamily: structure and function. *Biochem J* **409**, 333–348 (2008).
24. Watkins, H. A. & Baker, E. N. Structural and functional characterization of an RNase HI domain from the bifunctional protein Rv2228c from *Mycobacterium tuberculosis*. *J Bacteriol* **192**, 2878–86 (2010).
25. Watkins, H. A. & Baker, E. N. Structural and functional analysis of Rv3214 from *Mycobacterium tuberculosis*, a protein with conflicting functional annotations, leads to its characterization as a phosphatase. *J Bacteriol* **188**, 3589–3599 (2006).
26. Muller, P. *et al.* The 1.70 angstroms X-ray crystal structure of *Mycobacterium tuberculosis* phosphoglycerate mutase. *Acta Crystallogr D Biol Crystallogr* **61**, 309–315 (2005).

27. Costa, J., Empadinhas, N. & da Costa, M. S. Glucosylglycerate biosynthesis in the deepest lineage of the Bacteria: characterization of the thermophilic proteins GpgS and GpgP from *Persephonella marina*. *J Bacteriol* **189**, 1648–1654 (2007).
28. Empadinhas, N., Albuquerque, L., Henne, A., Santos, H. & da Costa, M. S. The bacterium *Thermus thermophilus*, like hyperthermophilic archaea, uses a two-step pathway for the synthesis of mannosylglycerate. *Appl Environ Microbiol* **69**, 3272–3279 (2003).
29. Saier, M. H., Jr. & Ballou, C. E. The 6-*O*-methylglucose-containig lipopolysaccharide of *Mycobacterium phlei*. Complete structure of the polysaccharide. *J Biol Chem* **243**, 4332–4341 (1968).
30. Jackson, M. & Brennan, P. J. Polymethylated polysaccharides from *Mycobacterium* species revisited. *J Biol Chem* **284**, 1949–1953 (2009).
31. Kaur, D. *et al*. Initiation of methylglucose lipopolysaccharide biosynthesis in mycobacteria. *PLoS ONE* **4**, e5447 (2009).
32. Sambou, T. *et al*. Capsular glucan and intracellular glycogen of *Mycobacterium tuberculosis*: biosynthesis and impact on the persistence in mice. *Mol Microbiol* **70**, 762–774 (2008).
33. Koonin, E. V. & Tatusov, R. L. Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity. Application of an iterative approach to database search. *J Mol Biol* **244**, 125–132 (1994).
34. Thaller, M. C., Schippa, S. & Rossolini, G. M. Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase superfamily. *Protein Sci* **7**, 1647–1652 (1998).
35. Gerlt, J. A. & Babbitt, P. C. Can sequence determine function? *Genome Biol* **1**, REVIEWS0005 (2000).
36. Empadinhas, N. *et al*. Organic solutes in *Rubrobacter xylanophilus*: the first example of di-*myo*-inositol-phosphate in a thermophile. *Extremophiles* **11**, 667–673 (2007).
37. Empadinhas, N. *et al*. Functional and structural characterization of a novel mannosyl-3-phosphoglycerate synthase from *Rubrobacter xylanophilus* reveals its dual substrate specificity. *Mol Microbiol* **79**, 76–93 (2011).
38. Omelchenko, M. V., Galperin, M. Y., Wolf, Y. I. & Omelchenko, M. V., Galperin, M. Y., Wolf, Y. I. & Koonin, E. V. Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution. *Biol Direct* **5**, 31 (2010).
39. Brennan, P. & Ballou, C. E. Biosynthesis of mannophosphoinositides by *Mycobacterium phlei*. The family of dimannophosphoinositides. *J Biol Chem* **242**, 3046–3056 (1967).
40. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248–254 (1976).
41. Fernandes, C. *et al*. Two alternative pathways for the synthesis of the rare compatible solute mannosylglucosylglycerate in *Petrotoga mobilis*. *J Bacteriol* **192**, 1624–1633 (2010).
42. Ames, B. N. Assay of inorganic phosphate, total phosphate and phosphatases. *Methods Enzymol* **8**, 115–118 (1966).
43. Takai, A. & Mieskes, G. Inhibitory effect of okadaic acid on the *p*-nitrophenyl phosphate phosphatase activity of protein phosphatases. *Biochem J* **275**, 233–239 (1991).
44. Fraser, H. I., Kvaratskhelia, M., & White, M. F. The two analogous phosphoglycerate mutases of *Escherichia coli*. *FEBS Lett* **455**, 344–348 (1999).
45. Hill, B. & Attwood, M. M. Purification and characterization of phosphoglycerate mutase from methanol-grown *Hyphomicrobium* X and *Pseudomonas* AM1. *J Gen Microbiol* **96**, 185–193 (1976).

Acknowledgements

This work was supported by Fundação para a Ciência e a Tecnologia (FCT), Portugal and FEDER, projects PTDC/BIA-PRO/110523/2009 and PTDC/BIA-BCM/112459/2009. V. Mendes, A. Maranha and S. Alarico acknowledge scholarships from FCT (SFRH/BD/36373/2007, SFRH/BD/74845/2010 and SFRH/BPD/43321/2008). We acknowledge Helena Santos (ITQB, Oeiras, Portugal) for supplying pure GG. We thank Bruno Manadas (Proteomics Unit, Center for Neuroscience and Cell Biology, University of Coimbra, Portugal) for the mass spectrometry analyses.

Author contributions

VM, AM, and SA performed the experiments. VM and NE conceived and designed the experiments and analyzed the data. VM, MdC and NE wrote the paper. NE and MdC contributed with reagents and materials.

Additional information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Mendes, V., Maranha, A., Alarico, S., da Costa, M.S. & Empadinhas, N. *Mycobacterium tuberculosis* Rv2419c, the missing glucosyl-3-phosphoglycerate phosphatase for the second step in methylglucose lipopolysaccharide biosynthesis. *Sci. Rep.* **1**, 177; DOI:10.1038/srep00177 (2011).