Cg1246, a new player in mycolic acid biosynthesis in Corynebacterium glutamicum

Célia de Sousa-d'Auria, Florence Constantinesco, Nicolas Bayan, Patricia Constant, Maryelle Tropis, Mamadou Daffé, Marc Graille, Christine Houssin

RESEARCH ARTICLE
de Sousa-d'Auria *et al.*, *Microbiology* 2022;168:001171
DOI 10.1099/mic.0.001171

Cg1246, a new player in mycolic acid biosynthesis in *Corynebacterium glutamicum*

Célia de Sousa-d'Auria[1], Florence Constantinesco[1], Nicolas Bayan[1], Patricia Constant[2], Maryelle Tropis[2], Mamadou Daffé[2], Marc Graille[3] and Christine Houssin[1,*]

Abstract

Mycolic acids are key components of the complex cell envelope of *Corynebacteriales*. These fatty acids, conjugated to trehalose or to arabinogalactan form the backbone of the mycomembrane. While mycolic acids are essential to the survival of some species, such as *Mycobacterium tuberculosis*, their absence is not lethal for *Corynebacterium glutamicum,* which has been extensively used as a model to depict their biosynthesis. Mycolic acids are first synthesized on the cytoplasmic side of the inner membrane and transferred onto trehalose to give trehalose monomycolate (TMM). TMM is subsequently transported to the periplasm by dedicated transporters and used by mycoloyltransferase enzymes to synthesize all the other mycolate-containing compounds. Using a random transposition mutagenesis, we recently identified a new uncharacterized protein (Cg1246) involved in mycolic acid metabolism. Cg1246 belongs to the DUF402 protein family that contains some previously characterized nucleoside phosphatases. In this study, we performed a functional and structural characterization of Cg1246. We showed that absence of the protein led to a significant reduction in the pool of TMM in *C. glutamicum*, resulting in a decrease in all other mycolate-containing compounds. We found that, *in vitro*, Cg1246 has phosphatase activity on organic pyrophosphate substrates but is most likely not a nucleoside phosphatase. Using a computational approach, we identified important residues for phosphatase activity and constructed the corresponding variants in *C. glutamicum*. Surprisingly complementation with these non-functional proteins fully restored the defect in TMM of the Δ*cg1246* mutant strain, suggesting that *in vivo*, the phosphatase activity is not involved in mycolic acid biosynthesis.

INTRODUCTION

The *Corynebacteriales* order includes many important members among which, the causative agents of tuberculosis, leprosy and diphtheria. Their cell envelope exhibits an atypical diderm organization with a conventional cytoplasmic membrane surrounded by an unusual matrix made of peptidoglycan (PG) and arabinogalactan (AG) linked to a mycolic acid-containing outer membrane (hence named the mycomembrane) [1, 2]. Mycolic acids (MAs) play a central role in *Corynebacteriales* envelope, constituting the skeleton of the mycomembrane [3, 4]. They are also essential to the survival of mycobacteria, and therefore, are the subject of numerous studies [5]. MAs are α-branched and *β*-hydroxylated fatty acids of long, but various chain length (e.g. C_{60}-C_{90} in *Mycobacterium*, C_{22}-C_{36} in *Corynebacterium*), that are covalently linked to different envelope components [2, 6, 7]. They are found as esters of the pentaarabinosyl ends of the AG, forming the inner leaflet of the mycomembrane, or as free lipids esterified to trehalose [Trehalose MonoMycolate (TMM) and Trehalose DiMycolate (TDM)] constituting the framework of the mycomembrane outer leaflet. In addition to their key role in the building of the cell wall, recent discoveries that pore-forming small proteins were post-translationally modified by mycolates (O-acylation of a serine residue) in the *Corynebacterium* genus, suggest a larger role for these lipids in protein activity and trafficking [8–10].

Received 05 January 2022; Accepted 04 March 2022; Published 08 April 2022
Author affiliations: [1]Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette, France; [2]Institut de Pharmacologie et de Biologie Structurale, IPBS, Université de Toulouse, CNRS, UPS, Toulouse, France; [3]Laboratoire de Biologie Structurale de la Cellule (BIOC), CNRS, Ecole polytechnique, IP Paris, F-91128 Palaiseau Cedex, Paris, France.
***Correspondence:** Christine Houssin, christine.houssin@i2bc.paris-saclay.fr
Keywords: mycolic acid; *Corynebacterium glutamicum*; lipid biosynthesis; phosphatase; mycomembrane.
Abbreviations: AG, arabinogalactan; FPP, farnesyl-pyroPhosphate; GMPMC, guanosine-5'-[(α,*β*)-methyleno]diphosphate; GPP, geranyl-pyroPhosphate; MA, mycolic acid; Myt, mycoloyltransferase; PG, peptidoglycan; PP, pyrophosphate; TDCM, trehalose dicorynomycolate; TDM, trehalose dimycolate; TLC, thin-layer chromatography; TMCM, trehalose monocorynomycolate; TMM, trehalose monomycolate.
Four supplementary figures and one supplementary table are available with the online version of this article.

All *Corynebacteriales* synthetize their MA in the cytoplasm by a well-conserved pathway ([2, 6, 11], Fig. S1, available in the online version of this article). In corynebacteria, synthesis begins with the fatty acid synthases FAS-IA and B, that produce the two C_{16} to C_{18} fatty acid precursors of MA molecules [12]. These precursors must, at first, be activated, one by carboxylation and the other by acyl-adenylation, to serve as substrates for a polyketide synthase to condense them into an α-alkyl *β*-keto fatty acid [13, 14]. In *C. glutamicum* the enzymes involved in this reaction sequence are the AccD2/AccD3 acyl-CoA carboxylase, the AMP-ligase FadD2 and the polyketide synthase Cg-Pks, respectively [13–15]. Although there is no experimental evidence, it is very likely that Cg-Pks works like its mycobacterial orthologue Pks13, transferring the condensation product onto a trehalose, *via* its thio-esterase domain, forming a keto-Trehalose MonoCorynoMycolate (keto-TMCM) molecule [16]. Keto-TMCM is then reduced by a distinct enzyme, the CmrA reductase, giving the mature hydroxylated TMCM [17]. Transport of TMCM to the periplasm is then carried out by two redundant non-essential lipid transporters of the resistance-nodulation-cell division (RND) family, CmpL1 and CmpL4 [18]. Optimal transport of TMCM requires its transient acetylation, a modification that involves TmaT, a membranous acetyltransferase that acetylates the mycoloyl chain of TMCM [19] and two other proteins: MtrP, a cytoplasmic methyltransferase [20] and MmpA a membrane protein [21], whose functions have not been elucidated to date. Once translocated into the periplasm, TMCM is used as a mycolate donor by a family of cell-wall enzymes, the mycoloyltransferases (Myts [22]). In *C. glutamicum,* five partially redundant Myts (Myt A, B, D, E and F) catalyse the covalent attachment of mycolates onto TMCM, to form Trehalose DiCorynoMycolate (TDCM), or onto AG [23, 24], while MytC specifically O-mycolates proteins [9].

Several factors have been described to control MA biosynthesis, modulating the TMM amount and consequently all the other mycolate-containing compounds. In *C. glutamicum*, MA biosynthesis is regulated by transcriptional factors that either influence fatty acid synthesis (FasR [25, 26], WhcD [27]), or control the expression of genes involved in MA metabolism (WhiA [28], σ^D [29, 30]). In mycobacteria a large number of enzymes involved in MA biosynthesis and transport were shown to be regulated by phosphorylation/dephosphorylation [31, 32]. Although not demonstrated in *C. glutamicum*, it is very likely that this type of protein modification also contributes to fine-tuning some enzymatic and transport activities involved in corynomycolate metabolism in this bacterium. A stress-induced factor was also found to regulate the mycomembrane lipid composition by an unknown mechanism (Elrf, [33]).

Very recently, using a random transposition mutagenesis, combined with a screen that allowed the identification of mutants with an altered cell-wall permeability, we identified *cg1246*, the genetic deletion of which leads to a significant decrease in the TMCM pool, suggesting that Cg1246 is involved in MA metabolism [34]. Cg1246 is an uncharacterized protein, carrying a DUF402 domain (for domain of unknown function), that encompasses all its sequence. DUF402 domain-containing proteins, the representative of which is FomD, a protein found in fosfomycin-producing *Streptomyces* species [35], are mostly found in *Actinobacteria* and *Firmicutes*. Only three proteins of this family have been functionally characterized to date, the FomD proteins from *Streptomyces wedmorensis* (*Sw*FomD) and *Streptomyces fradiae* (*Sf*FomD) and the SA1684 protein from *Staphylococcus aureus* [36, 37]. All these proteins display a nucleoside phosphatase activity *in vitro*. FomD proteins were shown to catalyse the hydrolysis of cytidylyl (S)-2-hydroxypropylphosphonate to give (S)−2-hydroxypropylphosphonate, the direct precursor of the antibiotic fosfomycin [36], while SA1684 was shown to hydrolyse nucleoside diphosphates into nucleoside monophosphates, an activity required for bacterial virulence through an unknown mechanism [37]. Structural data are also available for *Sf*FomD [36] and *Sc*FomD (FomD from *Streptomyces coelicolor*).

In the present study, we explored the functional and structural characteristics of Cg1246. We studied the structure–function relationships of this protein using a computational approach and characterized its phosphatase activity and substrate preferences *in vitro*. We present evidence that the protein most likely acts upstream to the TMCM transport step, regulating the homeostasis of the periplasmic pool of TMCM for the synthesis of all mycolate-containing species in the cell envelope. Surprisingly the phosphatase activity of Cg1246 is entirely dispensable to its function in mycolate synthesis.

METHODS

Strains and media

The bacterial strains used in this study are shown in Table 1. *C. glutamicum* strains were grown in brain–heart infusion (BHI) liquid medium (DIFCO) with shaking (250 r.p.m.) or in BHI-agar at 30 °C. *Escherichia coli* strains were grown at 37 °C in Luria–Bertani (LB) medium. When necessary, antibiotics were added to final concentrations of 25 µg ml^{-1} kanamycin (Km), and 6 µg ml^{-1} or 30 µg ml^{-1} chloramphenicol (Cm) for *C. glutamicum* and *E. coli*, respectively. *E. coli* BL21(DE3) was used for protein expression, and *E. coli* DH5α was used for all cloning steps. Transformation of *C. glutamicum* was performed by electroporation as previously described in de Sousa-d'Auria *et al.* [34].

DNA manipulations

Recombinant DNA techniques were performed as described by Sambrook *et al.* [38]. Plasmid DNA was isolated from *C. glutamicum* or *E. coli* using the Macherey-Nagel Nucleospin Plasmid kit. *C. glutamicum* chromosomal DNA was extracted as described by Ausubel *et al.* [39]. PCR were performed with a SimpliAmp thermocycler (Applied Biosystem) using DreamTaq or Phusion™ High Fidelity

Table 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Relevant genotype and/or phenotype	Source or reference
E. coli strains		
DH5α	F- φ80*lacZ*ΔM15 Δ(*lacZYA-arg*F)U169 *recA1 endA1 hsd*R17(r_K^-,m_K^+) *pho*A *sup*E44 λ- *thi-1 gyr*A96 *rel*A1	Invitrogen
BL21(DE3) GOLD	F- *ompT hsd*S_B ($r_B^-m_B^-$) *gal dcm* (DE3)	ThermoFisher Scientific
C. glutamicum strains		
RES167	Restrictionless derivative of ATCC13032	[43]
Δ*cg1246*	Δ*cg1246* derivative of RES167	[34]
Δ*cg1247*	Δ*cg1247* derivative of RES167	This work
Δ*cg1247-46*	Δ*cg1247-46* derivative of RES167	This work
Plasmids		
pK18MobSac	Km^r *sacB* RP4 *oriT* ColE1 *ori*	[40]
pK18MobSac-Δ*cg1247*	pK18MobSac containing the upstream and downstream regions of the *cg1247* ORF (for *cg1247* deletion)	This work
pK18MobSac-Δ*cg1247-46*	pK18MobSac containing the upstream region of cg1247 and the downstream region of the *cg1246* ORF (for *cg1247-1246* deletion)	This work
pET9sn1	Commercial expression vector pET9 with some minor modifications, Km^R	Novagen, laboratory modified
pET9sn-*cg1246*his	pET9sn1 expressing Cg1246 with a N-ter His-tag, Km^R	This work
pET9sn-*cg1246*-D98A	pET9sn1 expressing Cg1246 with D98 replaced by A and with a N-ter His-tag, Km^R	This work
pET9sn-*cg1246*-D115A	pET9sn1 expressing Cg1246 with D115 replaced by A and with a N-ter His-tag, Km^R	This work
pET9sn-*cg1246*-D102A-D115A	pET9sn1 expressing Cg1246 with D115 replaced by A and D102 replaced by A and with a N-ter His-tag, Km^R	This work
pET9sn-*cg1246*-K23A	pET9sn1 expressing Cg1246 with K23 replaced by A and with a N-ter His-tag, Km^R	This work
pET9sn-*cg1246*-R43A	pET9sn1 expressing Cg1246 with R43 replaced by A and with a N-ter His-tag, Km^R	This work
pET9sn-*cg1246*-Y81A	pET9sn1 expressing Cg1246 with Y81 replaced by A and with a N-ter His-tag, Km^R	This work
pCGL482	Shuttle vector *E coli*/*C. glutamicum*, Cm^R	[41]
pCGL824	Derivative of pCGL482 containing the *cspB* gene of *C. glutamicum* ATCC 17965, Cm^R	[41]
pCGL2420	Derivative of pCGL482 containing the *cg1246* gene under its own promoter (304 bp upstream the ATG of *cg1247*), Cm^R	[34]
pCGL2402	Derivative of pCGL482 containing the *cg1247* gene under the p*cspB* promoter, Cm^R	This work
pCGL2420-D98A	Derivative of pCGL2420, expressing the variant Cg1246-D98A, Cm^R	This work
pCGL2420-D115A	Derivative of pCGL2420, expressing the variant Cg1246-D115A, Cm^R	This work
pCGL2420- D102A-D115A	Derivative of pCGL2420, expressing the variant Cg1246-D115A-D102A, Cm^R	This work
pCGL2420-R43A	Derivative of pCGL2420, expressing the variant Cg1246-R43A, Cm^R	This work

DNA polymerases (ThermoFischer Scientific). All DNA purifications were performed using the Macherey-Nagel Nucleospin gel and PCR clean-up kit. Standard procedures for DNA digestion and ligation were used in conditions recommended by the enzyme manufacturer (Fermentas). Oligonucleotide primers synthesis and DNA sequencing were carried out by Eurofins Genomics.

Construction of plasmids and bacterial strains

The different plasmids used in this study are described in Table 1. *C. glutamicum* mutant strains Δ*cg1247* and Δ*cg1247-46* strains were constructed using the strategy described by Schäfer *et al.* [40]. Two DNA fragments overlapping the *cg1247* gene or the *cg1247-cg1246* operon, at its 5′ and 3′ extremities, were amplified by PCR from *C. glutamicum* total DNA using appropriate primers (1247-del1/1247-del2 and 1247-del3/1247-del4 for *cg1247* deletion and 1246-del1/1246-del2 and 1247-del3/1247-del4 for *cg1247-46* deletion, see Table S1). Fragments were cloned in the non-replicative vector pK18MobSac to give plasmids pK18MobSac-Δ*cg1247* and pK18MobSac-Δ*cg1247-46*. Plasmids were sequenced and transferred into *C. glutamicum* RES167 by electroporation. Transformants, in which the construct was integrated into the chromosome by single crossing-over, were selected on BHI plates containing Km. The second crossing-over event was selected by plating Km-resistant clones on BHI plates containing 10% sucrose. Km-sensitive among sucrose-resistant colonies were selected and screened by PCR for DNA excision, using appropriate primers. After verification of PCR products by sequencing, one clone carrying the *cg1247* deletion (Δ*cg1247* strain) or *cg1247-46* deletion (Δcg*1247-46* strain), was selected for further studies.

The expression vector pET9sn-*cg1246*his was constructed as follows: PCR primers 1246-Nde and 1246-His were used to amplify the *cg1246* ORF, from chromosome of the RES167 strain. The reverse primer (1246-His) contained a sequence encoding an in-frame C-terminal hexa-histidine tag (His_6) and a stop codon. The resulting PCR product was digested with NdeI and NotI and ligated into the same restriction sites of the expression vector pET9sn1 (a pET9 derivative) to obtain plasmid pET9sn-*cg1246*his. Six plasmids were constructed for expression of variant proteins in *E. coli* (pET9sn-*cg1246*-D98A, pET9sn-*cg1246*-D115A, pET9sn-*cg1246*-D102A-D115A, pET9sn-*cg1246*-K23A, pET9sn-*cg1246*-R43A and pET9sn-*cg1246*-Y81A) and four were constructed for complementation experiments with Cg1246 variants in *C. glutamicum* Δ*cg1246* strain (pCGL2420-D98A, pCGL2420-D115A, pCGL2420-D102A-D115A and pCGL2420-R43A). These vectors were obtained by site-directed mutagenesis, using the quickChange method (Agilent technologies) and plasmids pET9sn-*cg1246*his or pCGL2420 [34] as templates. The mutagenic oligonucleotide primers are listed in Table S1. Sequences of all cloned inserts were systematically controlled by DNA sequencing.

Complementation vector encoding Cg1247 was constructed using pCGL482 as the cloning vector [41]. The coding sequence of *cg1247* from *C. glutamicum* RES167 chromosomal DNA was amplified by PCR, using the primer pair 1247-Rca/1247-Xho, and digested with RcaI and XhoI. A DNA fragment containing the *cspB* promoter was generated using the primer pair pcspB-Bam/pcspB-Nco and pCGL824 as template DNA [41] and treated with BamHI and NcoI. Both amplicons were ligated between the compatible sites BamHI and XhoI of pCGL482, generating the pCGL2402.

Production and purification of Cg1246-His and its variants

E. coli BL21(DE3) GOLD cells harbouring the pET9sn-*cg1246*his or the mutated derivatives were grown in LB/Km medium at 37 °C. Protein expression was induced with 1 mM Isopropyl *β*-D-thiogalactopyranoside when the OD_{600} of the culture reached 0.5. Cultivation was continued at 25 °C for 4 h (200 r.p.m.). Cells were then harvested by centrifugation and washed with 25 mM Tris-HCl buffer, pH 8. The pellet was suspended in 4.5 ml of the same buffer, and 0.5 ml of cell lytic solution (Sigma) was added, with lysozyme at 0.02 mg ml^{-1} (final concentration) and 0.1 mg ml^{-1} of protease inhibitor AEBSF (Sigma). After 1 h incubation at 4 °C, 0.4 mg ml^{-1} of DNase and 1 mM $MgCl_2$ were added, and the incubation was continued for an additional 15 min. The lysate was then clarified by centrifugation, at 5000 *g* for 20 min at 4 °C, and imidazole and NaCl were added to the supernatant to obtain final concentrations of 10 and 300 mM, respectively. The lysate was then mixed with 0.75 ml of Ni-NTA resin (Macherey-Nalgen), pre-equilibrated with 25 mM Tris-HCl buffer, pH 8, 300 mM NaCl (wash buffer) containing 10 mM imidazole. After 1 h incubation at 4 °C, the resin was washed with the same buffer, and then with 10 ml of wash buffer containing 35 mM imidazole. Proteins were eluted with wash buffer containing 100 and 250 mM imidazole. The purity and homogeneity of eluted protein fractions were analysed by SDS-PAGE. The recombinant purified His-tagged Cg1246 (Cg1246-His) was desalted and concentrated with an Amicon Ultra-15 centrifugal filter (10 kDa cutoff, Merk Millipore) at 4500 *g* and 4 °C 15 min, until imidazole was eliminated. The concentration of Cg1246-His was determined by ultraviolet (UV) absorption at 280 nm (ε_m : 29 280 M^{-1} cm^{-1}). Cg1246 variants were expressed and purified in the same manner as Cg1246-His.

Lipid analysis

All thin-layer chromatography (TLC) analysis were performed using silica gel-coated plates (G-60, Macherey-Nagel). Radiolabeled plates were systematically analysed by phosphorimaging (Variable Mode Imager Typhoon TRIO, Amersham Biosciences) and lipid bands quantified using the ImageQuant software.

Extractable lipids

Cells were grown to exponential phase (OD_{600} ~ 4) and lipids were extracted from wet pellets twice, with $CHCl_3/CH_3OH$ (1:2 v/v, then 2:1 v/v) for 24 h at room temperature. The organic phases were pooled and evaporated to dryness and lipids were solubilized in $CHCl_3$ (typically 100 µl for lipids extracted from 20 ml culture). Lipids were analysed by TLC, developed with $CHCl_3/CH_3OH/H_2O$ (65:25:4, v/v/v). Detection of lipids was performed by immersion of the TLC plates in 10% H_2SO_4 in ethanol, followed by

heating at 110 °C. Extractable lipids were also analysed by TLC after radiolabelling of the bacteria. Briefly, BHI medium was inoculated with bacteria at an OD_{600} = 0.2 and cultivated 3 h, before being divided in 3 ml independent cultures. Then 3 µl of [1-^{14}C] palmitic acid (1.11 10^4 bq, Perkin Elmer) were added to each culture and the growth is continued for 2 h. Radiolabelled whole bacteria were harvested by centrifugation, and lipids were extracted and analysed by TLC as described above after adjustment of the lipid extract concentration at 20 mg ml^{-1} in $CHCl_3$ and deposit of 10 µl, except for lipid detection, which was obtained by phosphorimaging.

Mycolic acid contents

Strains were inoculated at an OD_{600} = 0.2, in 5 ml BHI medium, and cultivated 6 h with 5 µl [1-^{14}C] acetic acid (1.85 10^5 bq, Amersham) to radiolabel fatty acids. Radiolabelled whole bacteria were harvested by centrifugation and saponified with 1 ml of KOH (40%)/2-methoxyethanol (1:7; v/v) 16 h at 80 °C. The suspension was then acidified by adding sulfuric acid (20%) to reach a pH 1~2. Mycolic acids and fatty acids were extracted three times with diethyl ether, and the organic phase was washed two times with H_2O. After drying, the resulting fatty acids were methylated using diazomethane and analysed by TLC run in CH_2Cl_2 as developing solvent. TLC bands of radiolabeled lipids (fatty acids and mycolic acids) were quantified after phosphorimaging.

To quantify mycolic acids covalently linked to arabinogalactan or to trehalose, radiolabeled bacteria were first extracted by $CHCl_3$/CH_3OH (1:2, then 2:1 v/v) for 24 h prior saponification of the resulting delipidated residues or total lipid extract, respectively.

Pulse-chase experiment

Strains were inoculated at an OD_{600} = 0.2, in 20 ml BHI medium and cultivated for 3 h, before addition of 20 µl of [1-^{14}C] palmitic acid (7.4 10^4 bq, Perkin Elmer). After 15 min incubation, radiolabeled bacteria were washed with cold BHI medium, suspended in 20 ml fresh medium, and divided in six culture aliquots (3 ml each). After 0, 15 min, 30 min, 1 h, 2 h or overnight incubation, cells were harvested by centrifugation and lipids were extracted and analysed as described above.

Porin extraction and mass spectrometry analysis

Bacterial cultures were harvested by centrifugation and pellets were extracted with organic solvents CH_3OH and $CHCl_3$ (2:1 v/v). Extracts containing PorA, PorH and protein X (ProtX) were mixed with an equal volume of sinapinic acid matrix (Fluka, 10 mg ml^{-1}, Acetonitrile 70%, Triluoroacetic acid 1%) and 1 µl were spotted onto a metal target and dried at air condition. Mass spectrometry spectra of each sample were acquired in positive reflectron mode with the MALDI-TOF/TOF 5800 mass spectrometer (Sciex) and 5000 mass spectra were averaged per spot.

Size exclusion chromatography

Size exclusion chromatography was performed using an AKTA FPLC system (GE Healthcare, Buckinghamshire, UK). A superdex 75 Increase 10/300 GL column (GE Healthcare) was equilibrated with 25 mM Tris-HCl pH 8, 100 mM NaCl. The column was calibrated using a gel filtration calibration kit (GE Healthcare). Each standard protein was dissolved in the equilibration buffer and chromatographed on the column separately. The protein samples were fractionated on the column (0.5 ml min^{-1} ; 1 ml/fraction).

Enzymatic assays

The phosphatase activity of Cg1246-His was first assessed using *p*-NitroPhenyl Phosphate as a substrate (10 to 50 mM of substrate). Different conditions were tested (pH, NaCl, cations) and allowed to define the optimal buffer conditions: Tris 100 mM, pH 9.5, NaCl 100 mM, $MgCl_2$ 10 mM (assay buffer). The activity of the protein was then measured for different substrates, using the colorimetric free phosphate detection assay Biomol Green (Enzo Life Sciences). Reactions were performed at 37 °C in 96-well microplate. Each well contained the substrate (up to 350 µM) in assay buffer (50 µl final volume) and reaction was initiated by addition of Cg1246-His (between 1 and 16 µg ml^{-1} of enzyme in the reaction mixture, except for the variant proteins of Cg1246, for which the enzyme concentration was increased to 32 µg ml^{-1}). As controls, the same measurements were done without adding any enzyme. Time points were obtained by addition of 0.02 M EDTA (final concentration, to stop the reaction) following by 100 µl of the Biomol Green reagent. After 30 min incubation at room temperature, absorbance of the mix in the wells was measured in a labsystems Multiskan EX microplate reader at 620 nm. Because geranyl-geranyl-pyrophosphate is poorly soluble in water, dodecylmaltoside (a non-ionic detergent) was added to the reaction mixture (0.6% w/v final concentration). As a control, activity measurements were done with farnesyl-pyrophosphate in the same condition and showed a slight decrease in the substrate hydrolysis, due to the presence of the detergent. To determine phosphate concentration in the reaction solutions, absorbance values were compared to absorbance of standard Pi solutions. Initial velocities were calculated from the change in absorbance as a function of time after subtracting values of the corresponding controls obtained in the absence of Cg1246, using the online curve-fitting tool MyCurveFit (https://mycurvefit.com/).

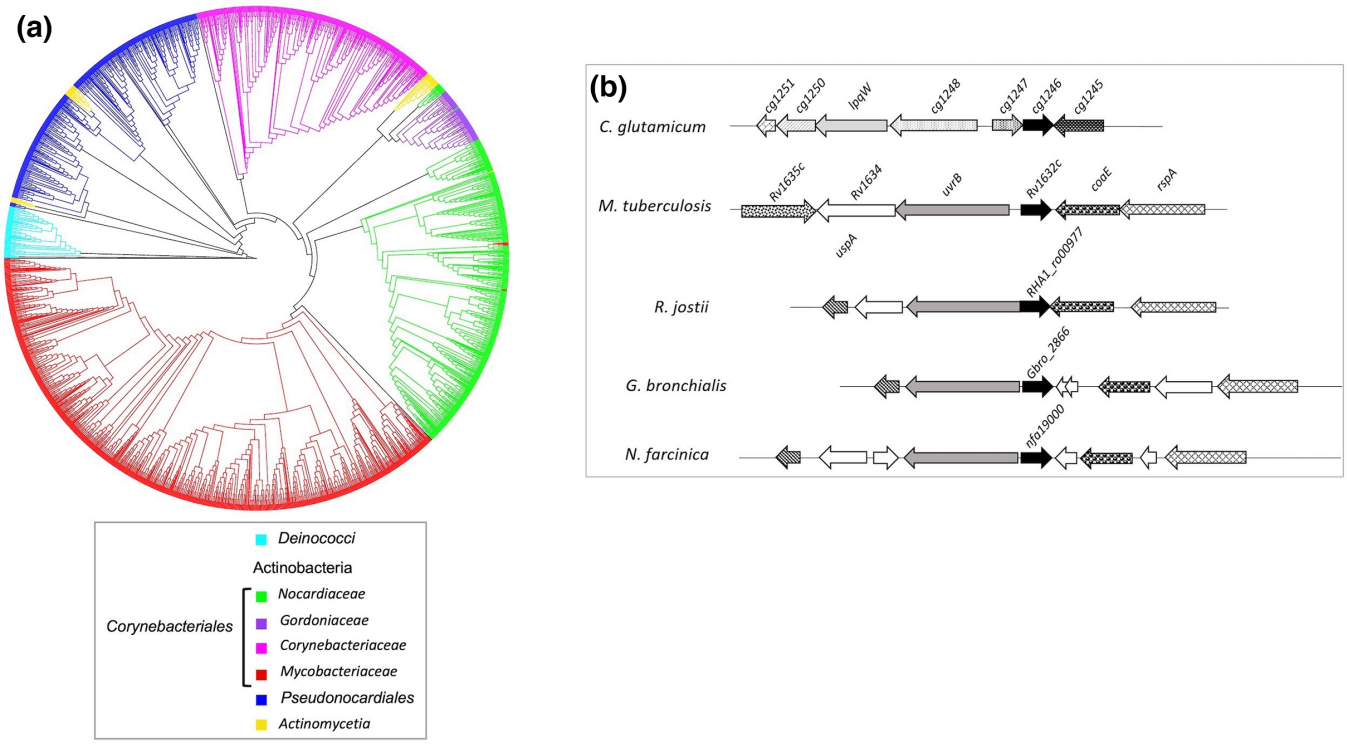


Fig. 1. Taxonomic repartition of Cg1246 orthologues and locus organization. (a) Circular neighbour-joining phylogenetic tree showing the taxonomic repartition of Cg1246. The tree was constructed using Geneious, from 2095 sequences obtained after a BLAST search using the protein sequence of Cg1246 from *C. glutamicum* ATCC13032, against NCBI databases. Branches are coloured according to the taxonomic annotation of the sequences. (b) Genetic organization of the *cg1246* (or its orthologous gene) locus in some representative *Corynebacteriales* (*Corynebacterium glutamicum* ATCC13032: *cg1246*, *Rhodococcus jostii* RHA1: *RHA1_ro00977*, *Mycobacterium tuberculosis* H37Rv: *Rv1632c*, *Gordonia bronchialis* 3410 DSM43247: *Gbro_2866*, *Nocardia farcinica* IFM 10152: *nfa19000*).

Model generation

To model the structure of Cg1246 protein, we have performed a BLAST (Basic Local Alignment Tool) search against the Protein Data Bank (PDB) to identify proteins of known three-dimensional structure and sharing significant sequence identity with Cg1246. This led to the identification of a protein encoded by the *RHA1_ro00977* gene from *Rhodococcus jostii* RHA1 strain (PDB code: 2P12), which shares 45% sequence identity with Cg1246 gene product. As Cg1246 protein is homodimeric in solution, we have modelled the structure of the Cg1246 homodimer with the ROBETTA server [42] using the coordinates of RHA1_ro00977 homodimer as template. The resulting model for the monomer exhibits a root-mean-square deviation (rmsd) value of 0.4 Å over 170 Cα atoms compared to the template structure. Such a value corresponds to classical ones obtained when comparing the crystal structures of proteins sharing 45–50% sequence identity.

RESULTS

Cg1246 is mostly found in bacteria belonging to the *Corynebacteriales* order

Cg1246 is a putative soluble protein of 179 amino acids. A BLAST search analysis, performed with Cg1246 sequence, indicated that this protein is well conserved and mostly associated with *Corynebacteriales* although some orthologues are also found in other *Actinobacteria*, principally the *Pseudonocardiales* and some are also associated with bacteria of the *Deinococcus-Thermus* phylum (Fig. 1a).

In corynebacteria, *cg1246* is always associated with another gene, which is *cg1247* in *C. glutamicum* ATCC13032. This gene, predicted to be organized in operon with *cg1246*, encodes a putative secreted protein of unknown function (Fig. 1b). Search for Cg1247 orthologous proteins, showed that it is conserved and strictly specific to the *Corynebacteriales*. In mycobacteria, as in other *Corynebacteriales* (*Rhodococcus* sp., *Gordonia* sp. or *Nocardia* sp.), *cg1246* and *cg1247* orthologues (*Rv1632c* and *Rv1157c*, respectively, in *M. tuberculosis*) are not localized at the same genetic locus (Fig. 1b).

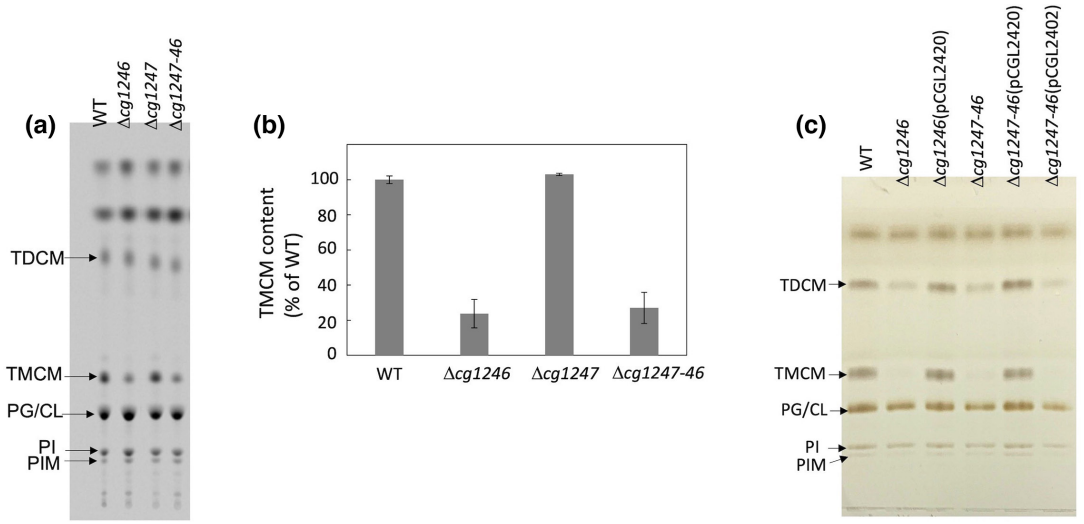


Fig. 2. Analyses of total lipids extracted from WT and mutant strains. (a) TLC-phosphorimaging of lipids extracted by $CHCl_3/CH_3OH$ from radiolabelled cells of RES167 strain (WT) and its derivative Δ*cg1246*, Δ*cg1247* and Δ*cg1247-46*, in exponentially growing phase. (b) Quantification of TMCM bands visualized in (a), expressed as % of the TMCM quantity of the control WT strain (absolute value corresponding to 100 %: 13,5 arbitrary unit). (c) TLC analysis of crude lipid extracts from exponentially growing cells of strains Res167 (WT), Δ*cg1246*, Δ*cg1246*(pCGL2420), Δ*cg1247-46*, Δ*cg1247-46*(pCGL2420) and Δ*cg1247-46*(pCGL2402). The values in (b) are the means±standard deviations of at least three independent experiments. Arrows indicate the position of phosphatidylinositol mannoside (PIM), phosphatidylinositol (PI), phosphatidylglycerol (PG), cardiolipine (CL), trehalose monocorynomycolate (TMCM) and trehalose dicorynomycolate (TDCM).

Loss of the *cg1246* but not *cg1247* gene affects the TMCM content in *C. glutamicum*

We previously showed that deletion of the *cg1246* gene led to an important decrease in the pool of TMCM, suggesting a role for the protein in mycolate metabolism [34]. Because of the synteny observed for *cg1247* and *cg1246* in *Corynebacteriaceae*, we wondered if Cg1247 is involved in the same metabolic pathway. To answer this question, we deleted the *cg1247* gene or the entire operon *cg1247-cg1246* in the Res167 strain, a derivative of the ATCC13032 reference strain [43]. For that purpose, we used the same two-step recombination method that we previously used for the construction of the Δ*cg1246* strain [34]. Mutant strains, corresponding to each of the desired deletion were easily isolated (strains: Δ*cg1247* and Δ*cg1247-46*). We thus performed quantification of TMCM, after radiolabelling with [1-^{14}C] palmitate of exponentially growing cells and TLC analysis of their extractable lipids. As shown in Fig. 2(a, b), while *cg1246* deletion led to an approximately 70–80% decrease in the TMCM pool compared to the wild-type (WT) strain, no difference could be detected for the Δ*cg1247* strain. As expected from these results, the Δ*cg1247-46* strain displayed the same profile as the Δ*cg1246* strain. The TMCM level was fully restored when Δ*cg1246* or Δ*cg1247-46* strains were complemented with a plasmid carrying the *cg1246* gene (pCGL2420) (Fig. 2c). In contrast, as expected, no complementary effect could be seen when a plasmid carrying the *cg1247* gene alone (pCGL2402) was introduced in the Δ*cg1247-46* mutant strain (Fig. 2c). From these data, we concluded that, while *cg1246* inactivation has a profound impact on the pool of TMCM, inactivation of *cg1247* has no effect on the trehalose mono- or di-mycolate content of the cell.

Loss of Cg1246 has an impact on all mycoloylated compounds

To better characterize the lipid deficiency of the Δ*cg1246* strain, we first quantified the total MA and fatty acid contents of the mutant *vs* WT strains. For that purpose, cells were labelled with [^{14}C] acetate and grown until the beginning of the stationary phase. Fatty acids and MAs were then released from the bacteria by saponification of whole cells and analysed as methyl esters by TLC and phosphorimaging. As shown in Fig. 3(a), compared to the WT strain, the total MA content was lower in the mutant strain by about 30%. This decrease is presumably not the result of a reduction in precursor production, since there is a concomitant increase in fatty acid content in the Δ*cg1246* strain as compared to the parental strain (Fig. 3b). We also compared the mycolate distribution between the cell-wall-bound ones, esterifying AG, and those esterifying trehalose, in the WT, Δ*cg1246*, and complemented strains. As shown in Fig. 3(c), both the level of trehalose and AG-linked MAs are reduced by approximately 50 and 20%, respectively, in the Δ*cg1246* as compared to the parental strain. Complementation abolished all the differences observed in the lipid content of the Δ*cg1246* mutant strain (Fig. 3a–c).

Regarding TDCM and because its amount depends on the growth phase (with a low level in early exponential phase increasing with time up to the stationary phase), we examined the dynamics of TMCM conversion to TDCM in the WT and Δ*cg1246* strains. For that purpose, we performed pulse-chase experiments using [1-^{14}C] palmitic acid as a precursor of mycolates. After

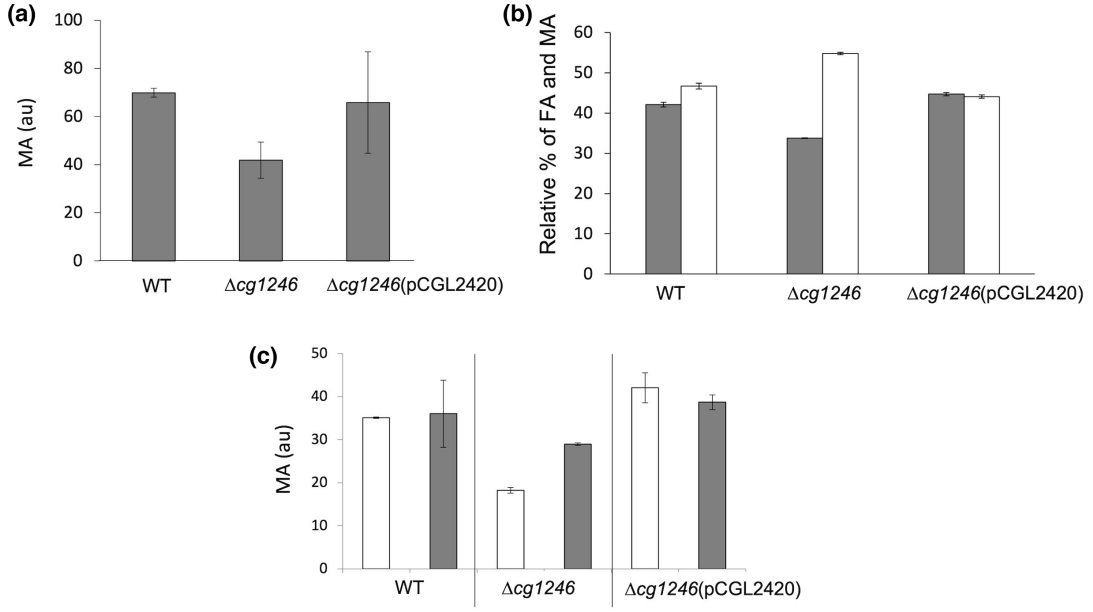


Fig. 3. Mycolic acid and fatty acid analysis of RES167 strain (WT) and its derivative Δ*cg1246* and Δ*cg1246*(pCGL2420). (a) Total mycolic acids (MA) content. (b) Relative quantification of fatty acid methyl esters (FA, white bars) and mycolic acids (MA, grey bars). (c) Mycolic acids' distribution between those bound to AG (grey bars) or to trehalose (white bars). In all experiments, lipids were extracted as described in Methods. Quantifications were obtained by measuring the radioactivity associated with the bands of interest on the TLCs and results were expressed in arbitrary units (a.u, a and c) or as percentages of total lipids obtained after saponification of whole bacteria (b). Values are means of three independent determinations±standard deviations.

15 min labelling, radioactive fatty acids were incorporated into TMCM and TDCM as well as in phospholipids in both strains (Fig. 4a). Although, labelling was lower for TMCM and, to a lesser extent for TDCM, in the mutant strain, the curve profiles for TDCM synthesis and TMCM consumption were quite similar between the Δ*cg1246* and the WT strains, with persistently lower levels of both lipid species in the mutant strain (Fig. 4b). This result showed that TDCM synthesis from TMCM follows the same kinetics in the Δ*cg1246* mutant and its parental strains. In this context, it should be noted that difference in the amount of TDCM between the WT and mutant strains is clearly visible in Fig. 2(c) but not in Fig. 2(a). This discrepancy is explained by the fact that the cells in Fig. 2(a) were in an earlier exponential phase than those in Fig. 2(c), during which the difference in the amount of TDCM between the two strains is much less marked (Fig. 4b).

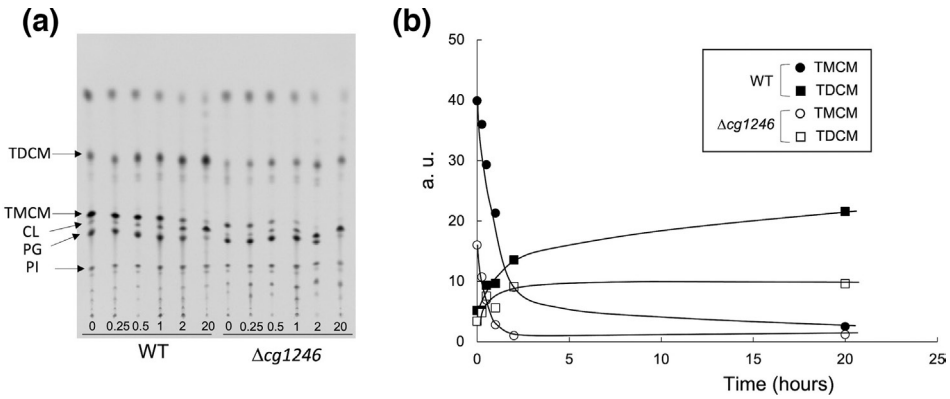


Fig. 4. Kinetics of TDCM synthesis and TMCM consumption in WT and Δ*cg1246* mutant strains. (a) TLC-phosphorimaging of lipids extracted from radiolabelled cells. Cells were pulse-labelled with [1-^{14}C] palmitic acid as described in Methods and aliquots were removed at the indicated times (0.25, 0.5, 1, 2 or 20 h). (b) Graph showing the kinetics of TDCM synthesis (squares) and TMCM consumption (circles), in WT (black symbols) or Δ*cg1246* mutant (white symbols), obtained after quantification of each of the corresponding bands of the TLC presented in (a) and expressed in arbitrary units (a.u.).

Because TMCM is also the precursor for protein mycoloylation, we examined the impact of *cg1246* inactivation on the modification of the small lipophilic porins PorA, PorH and ProtX, known to be O-mycoloylated [9]. Proteins were extracted with organic solvents and analysed directly by MALDI-TOF mass spectrometry (Fig. S2). As previously described [9, 10], the mass spectrum of the WT extract showed peaks corresponding mainly to mycoloylated proteins (m/z: 4408, 5213 and 6695 for ProtX, PorA and PorH, respectively) with traces of non-mycoloylated PorA and PorH (m/z: 4709 and 6190, respectively). A different profile was obtained for the mutant extract, with new peaks visible on the mass spectrum. In particular, a peak corresponding to non-mycoloylated form of ProtX was now present (m/z: 3903) and the fraction of non-mycoloylated PorA has increased relative to the mycoloylated form. Non-mycoloylated forms of ProtX and PorH disappeared in the complemented strain, while a very small fraction of non-mycoloylated PorA remained visible. Although there are more than three mycoloylated proteins in *C. glutamicum*, these analyses strongly suggest that the Δ*cg1246* strain has a protein mycoloylation deficiency.

All these data showed a decrease of all mycoloylated species in the mutant strain, a result that can be easily accounted for by a low level of the TMCM pool in the envelope. This TMCM depletion is clearly not the result of an increase in its consumption by one or more of the mycoloyltransferases, but rather of a deficiency in this precursor biosynthesis.

Cg1246 shares strong structural similarities with nucleoside diphosphatases

Cg1246 belongs to the large family of DUF402 proteins, the representative of which is FomD. The crystal structure of *Sc*FomD was first determined in 2008 and deposited in the protein data bank (PDB code: 3EXM). Although this structure was not associated with a functional characterization, crystallization of the protein with a non-hydrolyzable GDP analogue (GMPCP or Guanosine-5'-[(α,*β*)-methyleno]diphosphate) suggested a nucleoside phosphatase activity for this protein. This activity was confirmed for *Sw*FomD, another DUF402 protein, which hydrolyzes cytidylyl (S)−2-hydroxypropylphosphonate into CMP and (S)−2-hydroxypropylphosphonate [36]. Furthermore, the crystal structure of the orthologous *Sf*FomD protein has been solved in the absence, but also the presence of cytidine-5'-diphosphate (CDP) (PDB code: 5ZDN [36]). A nucleoside diphosphatase activity was also reported for the SA1684 protein of *Staphylococcus aureus*, another DUF402 domain-containing protein [37]. However, although these proteins belong to the DUF402 family, they share very limited sequence identity (less than 20% on a limited region) with Cg1246.

To get deeper insights into Cg1246 function, we have performed BLAST searches against the PDB to identify structures of closely related proteins. This led us to identify the RHA1_ro00977 protein from *Rhodococcus jostii* RHA1, which was previously determined by a structural genomics consortium (PDB code: 2P12). RHA1_ro00977 shares 45% sequence identity with Cg1246 (Fig. S3) and forms homodimer in the crystal, suggesting that Cg1246 could also be a homodimeric protein. To investigate this possibility, we expressed a recombinant C-terminally His-tagged protein (Cg1246-His) in *E. coli* and purified it by Ni^{2+} affinity chromatography. The purity of the protein was assessed by SDS-PAGE and showed only one band migrating at a molecular mass that is compatible with the theoretical mass of the monomer (20953 Da, Fig. 5a). We next analysed the oligomeric state of the Cg1246-His protein by size exclusion chromatography, yielding to a single homogeneous peak (Fig. 5b), with a retention volume that is compatible with that of a dimer (Fig. 5b inset). Thus, similarly to RHA1_ro00977, Cg1246 forms a homodimer in solution while *Sw*FomD was shown to be monomeric in solution [36].

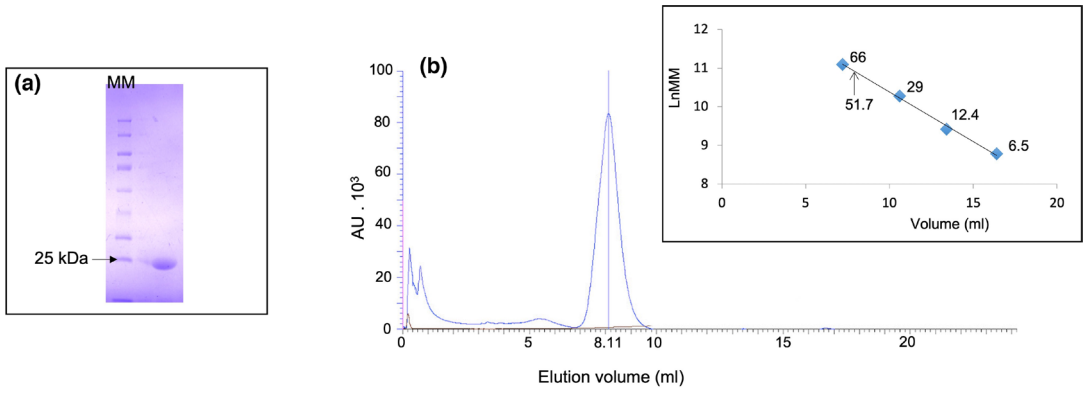


Fig. 5. Biochemical characterization of Cg1246-His. (a) SDS-PAGE of purified Cg1246-His. MM: markers of molecular mass. (b) Fast protein liquid chromatography (FPLC) chromatogram of purified Cg1246-His. AU: absorbance unit. The inset shows the calibration curve obtained by plotting the Ln of the molecular mass (MM) of four different proteins (albumin 66000 Da; carbonic anhydrase 29000 Da; cytochrome c 12400 Da and aprotinin 6500 Da) in function of their elution volumes. The arrow indicated the calculated molecular mass of Cg1246.

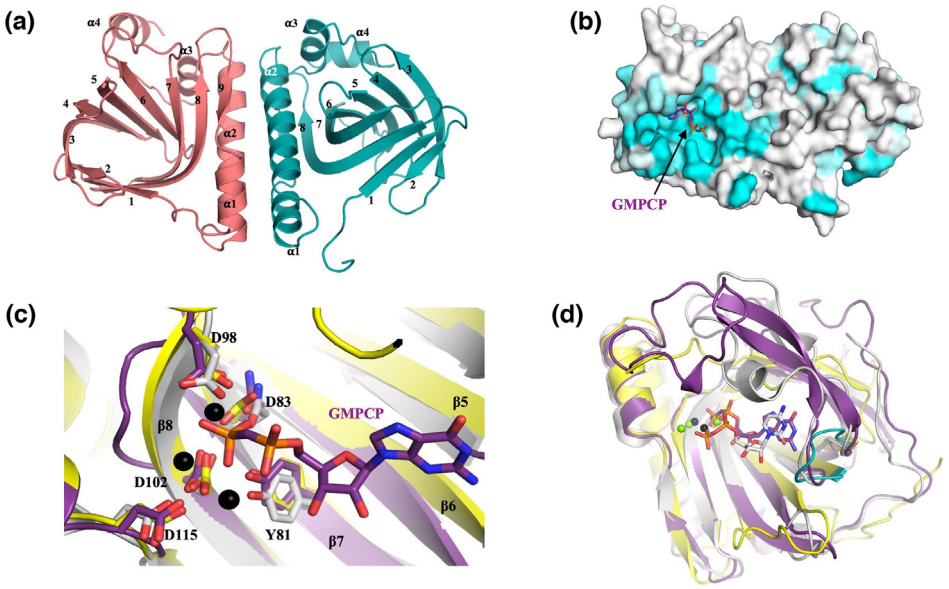


Fig. 6. Model of Cg1246 protein and comparison with other DUF402 members. (a) Sketch representation of the model of *C. glutamicum* Cg1246 homodimer complex. Secondary structure elements are labelled but for β-strands, only their number is indicated for the sake of clarity. (b) Mapping of the sequence conservation score at the surface of the homodimer. Colouring is from poorly conserved (white) to highly conserved (cyan). This score has been calculated from an alignment of 30 sequences using the Consurf server [52]. The same orientation as for (a) is used. The GMPCP molecule bound to SC4828 (*Sc*FomD) protein structure is shown as purple sticks. (c) The divalent metal ion binding sites from Cg1246 (yellow), SC4828 (purple) and *Sf*FomD (grey) are strongly conserved. Side chains from residues involved in metal coordination are shown as sticks. Calcium ions bound to SC4828 and GMPCP are shown as black spheres. For the sake of clarity, only Cg1246 amino acids numbering is shown. (d) Sketch representation of the superimposed structures of Cg1246 (yellow), SC4828 (purple) and *Sf*FomD (grey). GMPCP and CDP bound to SC4828 and *Sf*FomD, respectively, are shown as sticks. The loop connecting strands β4 and β5 in Cg1246 is highlighted in blue. Metal ions bound to SC4828 (Ca^{2+}) and *Sf*FomD (Mg^{2+}) are shown as black and green spheres, respectively.

Cg1246 exhibits a conserved enzyme active site

We next used the crystal structure of RHA1_ro00977 as a template to model the three-dimensional structure of Cg1246 homodimer with strong confidence. As shown in Fig. 6(a), Cg1246 protein is predicted to be made of an antiparallel 9-stranded β-barrel with four α-helices from the C-terminal end of this protein lying on one external face of the barrel. Two of these helices (α1 and α2) together with strand β9 from one monomer are contacting the corresponding regions from the second monomer to likely form a large homodimerization surface (Fig. 6a). Mapping of the sequence conservation score among Cg1246 orthologues at the surface of the homodimer reveals the presence of several extremely well-conserved residues, clustered at the flared extremity of the β-barrel, forming a large pocket (Fig. 6b). Interestingly, this highly conserved pocket matches with the active sites of the *Sf*FomD and *Sc*FomD enzymes (Fig. 6b; PDB code: 5ZDN [36] and 3EXM). Furthermore, the residues involved in the coordination of the divalent cations coordinating the phosphate groups of CDP bound to *Sf*FomD (N109, D125, D129 and D143) and of GMPCP in *Sc*FomD (N133, D149, D153 and D166) are strongly to strictly conserved in Cg1246 orthologues (D83, D98, D102 and D115 according to Cg1246 numbering; Figs 6(c) and S3). An additional tyrosine residue pointing towards the β phosphate in *Sf*FomD (Y107) and *Sc*FomD (Y131) is also present in Cg1246 (Y81) and is strictly conserved among Cg1246 orthologues (Fig. 6c, d). Interestingly, in the *S. aureus* SA1684 phosphatase, the substitution of the equivalent tyrosine by an alanine (Y88A mutant), completely inactivates its enzymatic activity [37], while its replacement by a phenylalanine (Y107F) in *Sf*FomD, does not influence its Km for the substrate but strongly affects its k_{cat} value (about 250-fold lower than that of the WT; [36]). Altogether, these observations strongly argue that Cg1246 has a divalent cation dependent enzymatic phosphatase activity. However, the comparison of the base binding pocket shows several important differences between Cg1246 and both *Sf*FomD and *Sc*FomD. First, the loop connecting strands β4 and β5 in Cg1246 is much shorter (seven residues) than in *Sc*FomD (39 residues including a long β-hairpin; Fig. 6(d)) or in *Sf*FomD (29 residues including an α-helix and a short β-strand; Fig. 6d). Second, the aromatic cage interacting with the nucleotide rings in *Sf*FomD and *Sc*FomD is not conserved in Cg1246. Hence, apart from the catalytic machinery, the substrate-binding pocket is strongly different, indicating that the substrate specificity of Cg1246 might be significantly different from that of the two FomD proteins.

Cg1246 has an alkaline phosphatase activity *in vitro*, that requires Mg^{2+}

Our analysis of the structural model of Cg1246 strongly suggests that the protein has a phosphatase activity that requires the presence of a divalent cation. To test this hypothesis and determine the optimal conditions for enzymatic activity, we first performed preliminary experiments using recombinant Cg1246-His with *p*-NitroPhenyl Phosphate as a substrate. Although the activity was low, as expected, the protein hydrolyzed *p*-NitroPhenyl Phosphate to pNitrophenol, but only in alkaline conditions and in the presence of $MgCl_2$. The best hydrolysis conditions were obtained at pH 9.5–10, with 10 mM $MgCl_2$ and 100 mM NaCl. We thus examined Cg1246-His activity, in these conditions, towards different nucleotides, including mono- (AMP, GMP, CMP), di- (GDP, ADP, UDP) or tri-phosphorylated (GTP, ATP, CTP) substrates. At 1 mM nucleotide concentration, a significant release of Pi was only observed for GTP and, to a much lesser extent, for ADP (approximately 10 and 0.4 µmol of Pi /mg enzyme min, respectively). At 50 µM, no hydrolysis of ADP could be detected, while a specific activity of 0.164 (±0.009) µmol of Pi released/mg enzyme min was determined for GTP. When $MgCl_2$ concentration was lowered to 1 mM, this activity decreased by a factor of 2.5 and no GTP hydrolysis could be detected when Mg^{2+} was replaced by Mn^{2+}, up to 10 mM. Thus, Cg1246 cation requirement is different from that of *Sf*FomD, *Sc*FomD and SA1684 that function with Mn^{2+} (or Co^{2+}), but not with Mg^{2+} for nucleotide hydrolysis [36, 37].

Given the differences that we observed between the substrate-binding pocket of Cg1246 and that of the FomD proteins (Fig. 6d), we searched for molecules that might be better substrates for Cg1246, than GTP. We therefore tested different phosphate-containing compounds that could be relevant in the context of envelope biosynthesis. This included nucleotide sugars (GDP-mannose and GDP-glucose), mono (Glucose-6P) or di-phosphorylated (Fructose 1-6diP) carbohydrates or lipid (phosphatidic acid) and other molecules that contain a pyrophosphate (PP) linkage [diacylglycerol-PP, and Farnesyl-PP (FPP)]. No significant hydrolysis could be detected for any of these molecules except FPP, for which a specific activity of 1.68 (±0.31) µmol of Pi released/mg enzyme min was measured, at 50 µM substrate concentration. This activity is 10 times higher than that measured for GTP at the same concentrations of substrate. Because FPP is a 15-carbon intermediate in the isoprenoid biosynthesis pathway that provides the lipid carrier decaprenyl-PP (C_{50}-PP), we also tested Geranyl-PP (GPP) and GeranylGeranyl-PP, the C_{10} and C_{20} intermediates, respectively. While a slightly lower activity was measured with Geranyl-PP [0.35 (±0.09) µmol of Pi released/mg enzyme min], no hydrolysis could be detected using GeranylGeranyl-PP as the substrate for Cg1246, indicating that isoprenoids longer than FPP are not recognized by the protein.

To estimate and compare the catalytic efficiency of Cg1246 for GTP and FPP, we performed reaction kinetics for the two substrates. Because the malachite green-based assay used to measure the Pi release, does not allow monitoring the kinetics continuously, initial rates could solely be determined at substrate concentrations for which the hydrolytic kinetics are not too fast (up to 100 and 350 µM for FPP and GTP, respectively). In these conditions, a Michaelis–Menten plot gave a straight line for both FPP and GTP, indicating that in each case, substrate concentrations are much lower than the K_m value (Fig. 7). Although neither K_m nor V_{max} values could be determined from this graph, the k_{cat}/K_m ratio could easily be calculated from the slope of the line. This ratio, which represents the catalytic efficiency of the enzyme, was 600 M^{-1} s^{-1} for GTP (Fig. 7a) and 12 200 M^{-1} s^{-1} for FPP (Fig. 7b), indicating that FPP is a much better substrate than GTP, which is apparently poorly recognized by Cg1246.

Finally, because FPP showed the best catalytic efficiency, we used this substrate to determine more precisely the pH-dependence of Cg1246 activity (Fig. S4). Cg1246 exhibited its highest activity (up to 65% of the maximum activity) between 9.5 and 11 with

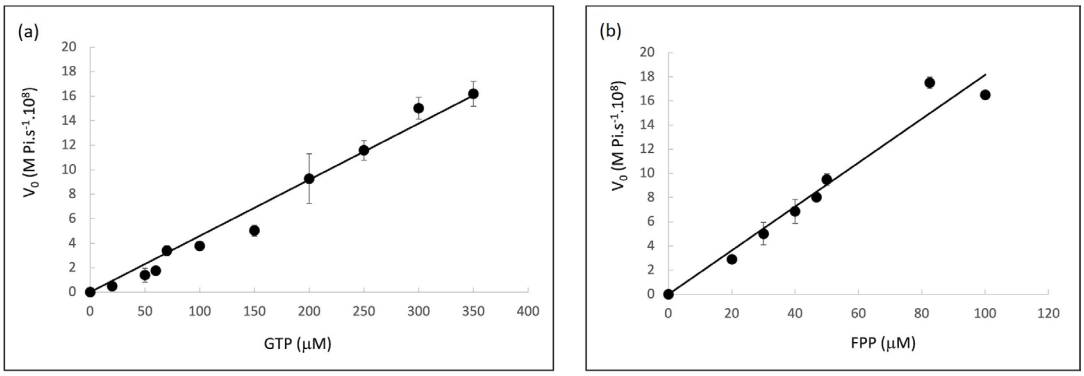
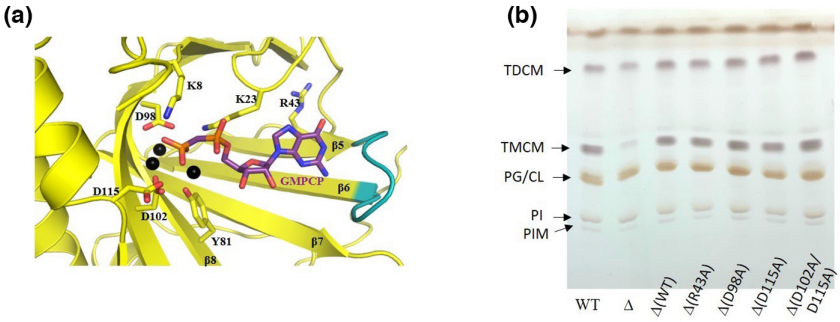


Fig. 7. Michaelis–Menten plots (initial rate V_0 in function of the substrate concentration) for GTP (a) and FPP (b). V_o were measured with 16 and 3 µg ml^{-1} of Cg1246-His for GTP and FPP, respectively, and expressed as Pi concentrations (M) released in the reaction medium, by second. Each value is the means±standard deviation of three independent experiments. For both GTP and FPP, values could not be fitted by an hyperbole and showed linear patterns. The slope of the lines (which represents k_{cat}/K_m) were calculated from simple linear regression analysis using excel.



(c)

Mutation	FPP hydrolysis (µmol /min.mg enzyme)	Mycolic acid complementation
WT	0.860 ± 0.062	Yes (line 3 Fig 8B)
R43A	0	Yes (line 4 Fig 8B)
D98A	0	Yes (line 5 Fig 8B)
D115A	0	Yes (line 6 Fig 8B)
D102A-D115A	0	Yes (line 7 Fig 8B)
K23A	0.885 ± 0.025	not determined
Y81A	0	not determined

Fig. 8. *In vitro* and *in vivo* activities of Cg1246 and its variants. (a) Hybrid model of GMPCP (purple) bound to Cg1246 obtained by superimposing the structure of SC4828 (*Sc*FomD) bound to GMPCP onto our Cg1246 model. Side chains mutated into alanine in this study are highlighted as sticks (D98, D115, D102, K23, R43 and Y81). Calcium ions bound to GMPCP are shown as green spheres. (b) TLC analyses of extractable lipids from the WT and the isogenic mutant strains. 1: WT; 2: Δ*cg1246*; 3: Δ*cg1246*(pCGL2420); 4: Δ*cg1246*(pCGL2420-R43A); 5: Δ*cg1246*(pCGL2420-D98A); 6: Δ*cg1246*(pCGL2420-D115A); 7: Δ*cg1246*(pCGL2420-D102A-D115A). (c) Summary of the *in vitro* and *in vivo* activities of Cg1246 and its variants. *In vitro* activities were measured using 3 µg ml^{-1} of the His-tagged proteins in the presence of 20 µM of FPP. *In vivo* activities of the variant proteins were estimated by comparison of the intensities of the TMCM and TDCM bands with those of the WT and Δ*cg1246* on the TLC showed in (b).

an optimum at pH 10.5 (100% of the activity). A small peak was also visible between pH 8 and 9 (around 45% of activity). No activity was detected for pH 7 or below.

All together, these results indicate that Cg1246 is a dimeric alkaline pyrophosphatase that requires Mg^{2+} to function. Although the protein shares many structural characteristics with known nucleoside phosphatases, its function seems to be different as evidenced by the substrate preferences of Cg1246 found in this study.

The phosphatase activity of Cg1246 is not involved in mycolic acid metabolism

To characterize more deeply the enzymatic activity of Cg1246, we mutated into alanine several residues strictly conserved in orthologous proteins and located within its putative active site (Figs S3, 6c and 8a) . First, we generated two single- and one double-point mutants affecting divalent cation binding (D98A and D115A and D102A-D115A). Next, we constructed the Y81A mutant, as the corresponding residue in *Sf*FomD (Y107) was proposed to activate a water molecule serving as a base for nucleophilic attack of the phosphonate group. Additionally, we generated the K23A and R43A single-point mutants to assess the role of these strictly conserved amino acids in Cg1246 function. All variants were first constructed in *E. coli*, from the plasmid pET9sn-*cg1246his*, which expresses the recombinant Cg1246-His (Table 1). The different variant proteins were purified by Ni^{2+} affinity chromatography and during purification, we did not observe any difference in the behaviour of these proteins compared to the WT one, indicating that the mutations are unlikely to affect protein folding. The phosphatase activity of all these proteins was measured *in vitro*, using 20 µM FPP as the substrate and 3 µg ml^{-1} of enzyme. Except for the K23A, which displayed the same phosphatase activity as the WT protein (Fig. 8c), no FPP hydrolysis could be observed for the other variant proteins. To ensure the loss of activity of these variants, enzyme concentration was increased by a factor of 10. In this condition, while a plateau was reached within 1 min for the WT protein, no hydrolysis could be detected, even after 30 min, for any of the variant of interest (D98A, D115A, D102A-D115A, Y81A and R43A). These results strongly support that as expected from the structural predictions, except for K23, these residues are essential for catalysis due to a role either in Mg^{2+} binding (D98 and D115), in the catalytic step (Y81, D98 and D115) or in substrate binding (R43).

To test the influence of these mutations on mycolic acid synthesis *in vivo*, mutations were also inserted into Cg1246 in the pCGL2420 vector. Plasmids encoding the D98A, D115A, D102A-D115A and R43A variants were constructed and introduced in the Δ*cg1246* strain (Table 1). For some unknown reason, we failed to construct the Y81A mutant in *C. glutamicum*. TLC analysis

of the extractable lipids of the different strains showed that, surprisingly, all the variants fully complemented the TMCM deficit of the Δ*cg1246* strain, as did the WT protein (Fig. 8b, c). These results indicate that the phosphatase activity of Cg1246 is not involved in mycolic acid metabolism and that the protein possesses two distinct and independent functions.

DISCUSSION

In the current study, we report biochemical and functional characterization of Cg1246, a small soluble protein that we previously identified by screening a transposon library of mutants exhibiting cell envelope alterations [34]. Because preliminary experiments have shown that the protein is very likely involved in MA metabolism, our first goal was to determine the importance of Cg1246 on the synthesis of all mycolate-containing compounds present in the cell envelope of *C. glutamicum*. Quantitative measurements of TMCM, TDCM, mycoloyl-AG and mycoloyl-proteins, showed that the absence of Cg1246, led to a severe decrease of all these compounds. As TMCM is the only known donor of mycolates in the cell envelope [2, 22], these results indicate that Cg1246 is involved in TMCM biosynthesis rather than in its utilization. Our results strongly suggest that the protein acts upstream of the periplasmic TMCM pool, allowing a sufficient level of precursors for the mycoloyltransferases to transfer MA onto their terminal acceptors. Therefore, the low TMCM level observed in the Δ*cg1246* mutant strain could reflect a deficiency in one (or more) of the cytoplasmic step(s) of the MA biosynthesis or in their transport across the inner membrane (Fig. S1). Because the lipid profile of the Δ*cg1246* strain did not match any of those observed for the previously characterized Δ*cmrA*, Δ*tmaT*, or Δ*cmpL1* or Δ*cmpL4* mutants [17–19], it seems unlikely that these steps (reduction, acetylation and transport) are impacted by the loss of the protein. In addition, we did not detect any accumulation of a new product that could correspond to a metabolic intermediate in the Δ*cg1246* mutant strain. We thus consider that our results are more consistent with a slow-down production of TMCM, i.e. fatty acid activation and/or condensation steps and/or attachment of the MA to trehalose (Fig. S1).

The mechanism by which Cg1246 might influence the TMCM level was investigated. The protein belongs to the DUF402 protein family that contains some previously characterized nucleoside phosphatases (FomD and the *Staphylococcal* protein SA1684 [36, 37],) and many orthologous proteins of uncharacterized function but one of which (RHA1_ro00977 in *R. jostii*), has been crystallized. Using all these data, we first performed a structural *in silico* analysis of Cg1246. This allowed us to highlight a highly conserved pocket in the protein, which most likely corresponds to the active site of the enzyme, arguing in favour of Cg1246 being a phosphatase with divalent cations requirement. Enzymatic assays performed using recombinant Cg1246 showed that, while the protein is actually a phosphatase *in vitro*, it is most likely not a nucleotide phosphatase. Indeed, no significant activity was found against all the canonical nucleotides tested here, except GTP, but for which a very low catalytic efficiency was determined. Although we cannot exclude non-canonical nucleotides as better substrates for Cg1246, the observed differences between the substrate-binding pockets of Cg1246 and FomD do not argue for a nucleotide as a specific substrate. Besides, we found two other notable differences between Cg1246 and the nucleoside phosphatases FomD and SA1684 [36, 37]: the metal ion dependency (Mg^{2+} *vs* Mn^{2+}) and the optimal working pH (alkaline *vs* neutral). These discrepancies indicate that not all hydrolases of the DUF402 family share the same enzymatic properties as previously suggested by Sato *et al.* [36]. Among all the other phosphate-containing molecules tested in this study, only the isoprenoids GPP and FPP were significantly hydrolysed by Cg1246. However, a detailed kinetic analysis performed for FPP, revealed that the k_{cat}/K_m value for this substrate is also very low. Extrapolation of these results to determine whether these molecules are related to the physiological substrate(s) of Cg1246 is not possible at this stage.

To link the phosphatase activity of Cg1246 to its function in TMCM metabolism, we performed site directed mutagenesis on target residues potentially involved in phosphatase activity. Unexpectedly, all the mutants compromised in their phosphatase function fully restored the TMCM level in the Δ*cg1246* mutant strain. These results indicate that Cg1246 protein participates in mycolate metabolism, by a mechanism that does not involve its phosphatase activity. Thus, if the phosphatase activity of Cg1246 exists *in vivo*, it is involved in another metabolic pathway that is not linked to TMCM biosynthesis. Interestingly, *cg1246* was also found in a study combining high-throughput screening of a transposon mutant library of *C. glutamicum*, with metabolome analysis [44]. In this study, Reimer *et al.* searched for mutants with modifications in metabolite profiles of the TCA cycle and/or the glutamate biosynthesis pathway. They showed that *cg1246* interruption leads to a mutant strongly affected in its TCA cycle (with decreased concentrations of the late cycle metabolites), that also produced very high level of proline (10 times more than the WT) and very low level of N-acetyl glutamate (0.07 times less than the WT). Although they could not identify the function of Cg1246, to explain the very low level of N-acetyl glutamate observed in the mutant strain, the authors proposed that the protein could be involved in the arginine biosynthesis pathway. In this context, it is tempting to assume that the metabolic disturbances observed by Reimer *et al.* in their *cg1246*-interrupted strain could be related to the loss of Cg1246 phosphatase activity. It should also be noted that Reimer *et al.* reported a 50% decreased in the growth of the Δ*cg1246* mutant strain in minimal medium, while we did not observe any growth defect for the Δ*cg1246* strain in rich medium conditions, a result that is consistent with Cg1246 involved in the biosynthesis of an amino acid.

Search for other omics studies citing *cg1246* in their dataset, retrieved three transcriptomic different results. The first one reports a two-fold downregulation under nitrogen limitation [45]. As pointing out by Reimer *et al.* [44], because nitrogen limitation leads to transcriptional repression of the arginine biosynthetic genes [45], a concomitant reduction in *cg1246* expression is consistent

with the hypothesis of this gene being involved in arginine metabolism. The second study also reports a slight downregulation of *cg1246* when the McbR repressor, involved in sulphur metabolism of *C. glutamicum*, was inactivated [46]. However, *cg1246* is not part of the large McbR regulon, and the downregulation observed in the absence of the regulator could result from a secondary regulation, as other global regulators were found deregulated by McbR inactivation. Finally, it was also shown that transcription of *cg1246* depends on the sigma factor SigD, an important regulator of mycomembrane biosynthesis [29, 30]. This data completely supports our finding that Cg1246 has a crucial role in MA biosynthesis.

Is Cg1246 a bifunctional protein with two distinct activities *in vivo*, or does the phosphatase activity that we detected *in vitro* is not physiological? As described in the results section, the protein possesses a highly conserved catalytic machinery with proteins that are phylogenetically very distant (FomD, SA1684). As shown in Fig. S3, the catalytic residues are conserved in actinobacterial homologues, and a sequence alignment of all orthologous sequences found in *Corynebacterineae* shows that the catalytic residues identified in Cg1246 are conserved in all corynebacterial orthologs. Although this is not a proof, it would be surprising that evolution has maintained a catalytic site that would no longer be useful to the bacterium. These observations argue for an authentic function of this phosphatase activity *in vivo*. It is thus plausible that Cg1246 is involved in two distinct activities. This possible bifunctionality is reminiscent of several bacterial proteins which are qualified as moonlighting proteins [47, 48]. These include many metabolic enzymes and molecular chaperones, displaying a secondary independent function, which is often a non-catalytic one [49]. There are several well-documented examples of this in mycobacteria, among which the *M. tuberculosis* aconitase, that both catalyses the isomerization of citrate to isocitrate in the TCA cycle in addition to its ability to bind RNA and regulate the expression of proteins involved in iron homeostasis [50]. Although Cg1246 involvement in two distinct biosynthetic pathways is strongly suggested by our results, the mechanism by which the protein performs its dual function must be elucidated. We are convinced that this protein has an enzymatic activity, with a catalytic site dedicated to phosphate hydrolysis, but the precise substrate(s) need to be identified. The mechanism by which Cg1246 acts on TMCM biosynthesis must also be determined. A second enzymatic activity should not be excluded, although, at this stage, no structural or biochemical evidence support this hypothesis. Alternatively, Cg1246 may be part of a larger enzymatic complex acting as a coupling or scaffolding protein within the complex. Finaly, Cg1246 could act as a regulator in MA biosynthesis, operating through interactions with another macromolecule (protein, DNA or RNA [51]). More investigations are clearly required to identify potential macromolecular interactant(s) of Cg1246 that will help to decipher its function in MA biosynthesis.

Funding information
This research was supported by the Centre National de la Recherche Scientifique (CNRS), the Université Paris-Saclay and the Ecole Polytechnique.

Acknowledgements
We are very grateful to C. Couteau for all its technical help. This work has benefited from the facilities and expertise of the I2BC proteomic platform (Proteomic-Gif, SICaPS) supported by IBiSA, Ile de France Region, Plan Cancer, CNRS and Paris-Sud University.

Author contributions
All the authors have made a substantial contribution to the work. Conceptualization and investigation: Cd.S., F.C., M.G., M.T., P.C. and C.H. Supervision and project administration: M.D., N.B., M.G. and C.H. Methodology and validation: Cd.S., F.C., M.G., M.T., P.C. and C.H. Original draft of the manuscript and visualisation: C.H. and M.G. All authors reviewed and edited the manuscript.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Daffé M, Marrakchi H. Unraveling the structure of the mycobacterial envelope. *Microbiol Spectr* 2019;7.
2. Houssin C, de Sousa d'Auria C, Constantinesco F, Dietrich C, Labarre C, *et al*. Architecture and biogenesis of the cell envelope of *Corynebacterium glutamicum*. In: Inui M and Toyoda K (eds). *Corynebacterium Glutamicum: Biology and Biotechnology*. Cham: Springer International Publishing; 2020. pp. 25–60.
3. Zuber B, Chami M, Houssin C, Dubochet J, Griffiths G, *et al*. Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. *J Bacteriol* 2008;190:5672–5680.
4. Hoffmann C, Leis A, Niederweis M, Plitzko JM, Engelhardt H. Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proc Natl Acad Sci U S A* 2008;105:3963–3967.
5. Marrakchi H, Lanéelle M-A, Daffé M. Mycolic acids: structures, biosynthesis, and beyond. *Chem Biol* 2014;21:67–85.
6. Quémard A. New insights into the mycolate-containing compound biosynthesis and transport in mycobacteria. *Trends Microbiol* 2016;24:725–738.
7. Lanéelle MA, Tropis M, Daffé M. Current knowledge on mycolic acids in *Corynebacterium glutamicum* and their relevance for biotechnological processes. *Appl Microbiol Biotechnol* 2013;97:9923–9930.
8. Huc E, Meniche X, Benz R, Bayan N, Ghazi A, *et al*. O-mycoloylated proteins from *Corynebacterium*: An unprecedented post-translational modification in bacteria. *J Biol Chem* 2010;285:21908–21912.
9. Huc E, de Sousa-D'Auria C, de la Sierra-Gallay IL, Salmeron C, van Tilbeurgh H, *et al*. Identification of a mycoloyl transferase selectively involved in o-acylation of polypeptides in *Corynebacteriales*. *J Bacteriol* 2013;195:4121–4128.
10. Issa H, Huc-Claustre E, Reddad T, Bonadé Bottino N, Tropis M, *et al*. Click-chemistry approach to study mycoloylated proteins: Evidence for PorB and PorC porins mycoloylation in *Corynebacterium glutamicum*. *PLoS One* 2017;12:e0171955.
11. Paweł Czyk J, Kremer L, Hatfull GF, Jacobs Jr. WR, editors. The Molecular Genetics of Mycolic Acid Biosynthesis. *Microbiol Spectr* 2014;2:MGM2-0003.
12. Radmacher E, Alderwick LJ, Besra GS, Brown AK, Gibson KJC, *et al*. Two functional FAS-I type fatty acid synthases in *Corynebacterium glutamicum*. *Microbiology (Reading)* 2005;151:2421–2427.

13. Portevin D, de Sousa-D'Auria C, Montrozier H, Houssin C, Stella A, *et al*. The acyl-AMP ligase FadD32 and AccD4-containing acyl-CoA carboxylase are required for the synthesis of mycolic acids and essential for mycobacterial growth: identification of the carboxylation product and determination of the acyl-CoA carboxylase componen. *J Biol Chem* 2005;280:8862–8874.
14. Portevin D, De Sousa-D'Auria C, Houssin C, Grimaldi C, Chami M, *et al*. A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms. *Proc Natl Acad Sci U S A* 2004;101:314–319.
15. Gande R, Gibson KJC, Brown AK, Krumbach K, Dover LG, *et al*. Acyl-CoA carboxylases (accD2 and accD3), together with a unique polyketide synthase (Cg-pks), are key to mycolic acid biosynthesis in *Corynebacterianeae* such as *Corynebacterium glutamicum* and *Mycobacterium tuberculosis*. *J Biol Chem* 2004;279:44847–44857.
16. Gavalda S, Bardou F, Laval F, Bon C, Malaga W, *et al*. The polyketide synthase Pks13 catalyzes a novel mechanism of lipid transfer in mycobacteria. *Chem Biol* 2014;21:1660–1669.
17. Lea-Smith DJ, Pyke JS, Tull D, McConville MJ, Coppel RL, *et al*. The reductase that catalyzes mycolic motif synthesis is required for efficient attachment of mycolic acids to arabinogalactan. *J Biol Chem* 2007;282:11000–11008.
18. Varela C, Rittmann D, Singh A, Krumbach K, Bhatt K, *et al*. MmpL genes are associated with mycolic acid metabolism in mycobacteria and corynebacteria. *Chem Biol* 2012;19:498–506.
19. Yamaryo-Botte Y, Rainczuk AK, Lea-Smith DJ, Brammananth R, van der Peet PL, *et al*. Acetylation of trehalose mycolates is required for efficient MmpL-mediated membrane transport in *Corynebacterineae*. *ACS Chem Biol* 2015;10:734–746.
20. Rainczuk AK, Klatt S, Yamaryo-Botté Y, Brammananth R, McConville MJ, *et al*. Mtrp, a putative methyltransferase in corynebacteria, is required for optimal membrane transport of trehalose mycolates. *J Biol Chem* 2020;295:6108–6119.
21. Cashmore TJ, Klatt S, Brammananth R, Rainczuk AK, Crellin PK, *et al*. MmpA, a conserved membrane protein required for efficient surface transport of trehalose lipids in *Corynebacterineae*. *Biomolecules* 2021;11:13417–13423.
22. Dautin N, de Sousa-d'Auria C, Constantinesco-Becker F, Labarre C, Oberto J, *et al*. Mycoloyltransferases: A large and major family of enzymes shaping the cell envelope of *Corynebacteriales*. *Biochim Biophys Acta Gen Subj* 2017;1861:3581–3592.
23. De Sousa-D'Auria C, Kacem R, Puech V, Tropis M, Leblon G, *et al*. New insights into the biogenesis of the cell envelope of corynebacteria: identification and functional characterization of five new mycoloyltransferase genes in *Corynebacterium glutamicum*. *FEMS Microbiol Lett* 2003;224:35–44.
24. Brand S, Niehaus K, Pühler A, Kalinowski J. Identification and functional analysis of six mycolyltransferase genes of *Corynebacterium glutamicum* ATCC 13032: the genes *cop1*, *cmt1*, and *cmt2* can replace each other in the synthesis of trehalose dicorynomycolate, a component of the mycolic acid l. *Arch Microbiol* 2003;180:33–44.
25. Nickel J, Irzik K, van Ooyen J, Eggeling L. The TetR-type transcriptional regulator FasR of *Corynebacterium glutamicum* controls genes of lipid synthesis during growth on acetate. *Mol Microbiol* 2010;78:253–265.
26. Irzik K, van Ooyen J, Gätgens J, Krumbach K, Bott M, *et al*. Acyl-CoA sensing by FasR to adjust fatty acid synthesis in *Corynebacterium glutamicum*. *J Biotechnol* 2014;192 Pt A:96–101.
27. Lee DS, Kim Y, Lee HS. The whcD gene of *Corynebacterium glutamicum* plays roles in cell division and envelope formation. *Microbiol (United Kingdom)* 2017;163:131–143.
28. Lee JH, Jeong H, Kim Y, Lee HS. *Corynebacterium glutamicum* whiA plays roles in cell division, cell envelope formation, and general cell physiology. *Antonie van Leeuwenhoek, Int J Gen Mol Microbiol* 2020;113:629–641.
29. Toyoda K, Inui M. Extracytoplasmic function sigma factor σD confers resistance to environmental stress by enhancing mycolate synthesis and modifying peptidoglycan structures in Corynebacterium glutamicum. *Mol Microbiol* 2018;107:312–329.
30. Taniguchi H, Busche T, Patschkowski T, Niehaus K, Pátek M, *et al*. Physiological roles of sigma factor SigD in *Corynebacterium glutamicum*. *BMC Microbiol* 2017;17:158.
31. Alsayed SSR, Beh CC, Foster NR, Payne AD, Yu Y, *et al*. Kinase targets for mycolic acid biosynthesis in *Mycobacterium tuberculosis*. *Curr Mol Pharmacol* 2019;12:27–49.
32. Le N-H, Locard-Paulet M, Stella A, Tomas N, Molle V, *et al*. The protein kinase PknB negatively regulates biosynthesis and trafficking of mycolic acids in mycobacteria. *J Lipid Res* 2020;61:1180–1191.
33. Meniche X, Labarre C, de Sousa-d'Auria C, Huc E, Laval F, *et al*. Identification of a stress-induced factor of *Corynebacterineae* that is involved in the regulation of the outer membrane lipid composition. *J Bacteriol* 2009;191:7323–7332.
34. de Sousa-d'Auria C, Constantinesco-Becker F, Constant P, Tropis M, Houssin C. Genome-wide identification of novel genes involved in *Corynebacteriales* cell envelope biogenesis using *Corynebacterium glutamicum* as a model. *PLoS One* 2020;15:e0240497.
35. Woodyer RD, Shao Z, Thomas PM, Kelleher NL, Blodgett JAV, *et al*. Heterologous production of fosfomycin and identification of the minimal biosynthetic gene cluster. *Chem Biol* 2006;13:1171–1182.
36. Sato S, Miyanaga A, Kim S-Y, Kuzuyama T, Kudo F, *et al*. Biochemical and structural analysis of FomD that catalyzes the hydrolysis of cytidylyl (S)-2-hydroxypropylphosphonate in fosfomycin biosynthesis. *Biochemistry* 2018;57:4858–4866.
37. Imae K, Saito Y, Kizaki H, Ryuno H, Mao H, *et al*. Novel nucleoside diphosphatase contributes to *Staphylococcus aureus* virulence. *J Biol Chem* 2016;291:18608–18619.
38. Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning: A Laboratory Manual*. 2nd ed. New York: Cold Spring Harbor Laboratory, Cold Spring Harbor; 1989.
39. Ausubel F, Brent R, Kingston R, Moore D, Seidman J, *et al*. *Current Protocols in Molecular Biology*. (Curr Protoc Mol Biol). New York: Wiley Interscience; 1987.
40. Schäfer A, Tauch A, Jäger W, Kalinowski J, Thierbach G, *et al*. Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene* 1994;145:69–73.
41. Peyret JL, Bayan N, Joliff G, Gulik-Krzywicki T, Mathieu L, *et al*. Characterization of the cspB gene encoding PS2, an ordered surface-layer protein in *Corynebacterium glutamicum*. *Mol Microbiol* 1993;9:97–109.
42. Song Y, DiMaio F, Wang RY-R, Kim D, Miles C, *et al*. High-resolution comparative modeling with RosettaCM. *Structure* 2013;21:1735–1742.
43. Dusch N, Pühler A, Kalinowski J. Expression of the *Corynebacterium glutamicum* panD gene encoding L-aspartate-alpha-decarboxylase leads to pantothenate overproduction in *Escherichia coli*. *Appl Environ Microbiol* 1999;65:1530–1539.
44. Reimer LC, Spura J, Schmidt-Hohagen K, Schomburg D. High-throughput screening of a *Corynebacterium glutamicum* mutant library on genomic and metabolic level. *PLoS One* 2014;9:e86799.
45. Silberbach M, Schäfer M, Hüser AT, Kalinowski J, Pühler A, *et al*. Adaptation of *Corynebacterium glutamicum* to ammonium limitation: a global analysis using transcriptome and proteome techniques. *Appl Environ Microbiol* 2005;71:2391–2402.
46. Rey DA, Nentwich SS, Koch DJ, Rückert C, Pühler A, *et al*. The McbR repressor modulated by the effector substance S- adenosylhomocysteine controls directly the transcription of a regulon involved in sulphur metabolism of *Corynebacterium glutamicum* ATCC 13032. *Mol Microbiol* 2005;56:871–887.
47. Jeffery CJ. Moonlighting proteins: Old proteins learning new tricks. *Trends Genet* 2003;19:415–417.

48. Jeffery CJ. What is Protein Moonlighting and Why is it Important? In: Henderson B (eds). *Moonlighting Proteins: Novel Virulence Factors in Bacterial Infections*. 2017. pp. 1–19.
49. Mani M, Chen C, Amblee V, Liu H, Mathur T, *et al*. MoonProt: a database for proteins that are known to moonlight. *Nucleic Acids Res* 2015;43:D277-82.
50. Banerjee S, Nandyala AK, Raviprasad P, Ahmed N, Hasnain SE. Iron-dependent RNA-binding activity of *Mycobacterium tuberculosis* aconitase. *J Bacteriol* 2007;189:4046–4052.
51. Commichau FM, Stülke J. Trigger enzymes: bifunctional proteins active in metabolism and in controlling gene expression. *Mol Microbiol* 2008;67:692–702.
52. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, *et al*. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* 2016;44:W344-50.

Edited by: I. Martin-Verstraete and D. JV Beste