The endogenous galactofuranosidase GlfH1 hydrolyzes mycobacterial arabinogalactan

Received for publication, November 10, 2019, and in revised form, February 11, 2020 Published, Papers in Press, February 27, 2020, DOI 10.1074/jbc.RA119.011817

Lin Shen‡, Albertus Viljoen§, Sydney Villameule†, Maju Joe‡, Iman Halloum§, Loïc Chêne‡, Alexandre Méry‡, Emeline Fabre§, Kaoru Takegawa**, Todd L. Lowary¶, Stéphane P. Vincent‡, Laurent Kremer‡‡, Yann Guérardel†, and Christophe Mariller‡‡‡

From the ‡Univ. Lille, CNRS, UMR8756 - UGSF - Unité de Glycobiologie Structurale et Fonctionnelle, F-59000 Lille, France, §Institut de Recherche en Infectiologie de Montpellier (IRIM), UMR9004 - CNRS/UM, 1919 route de Mende, 34293 Montpellier cedex 5, France, †Laboratoire de Chimie Bio-Organique (CBO), Université de Namur, rue de Bruxelles 61, 5000 Namur, Belgium, ¶Alberta Glycomics Centre and Department of Chemistry, University of Alberta, Edmonton AB T6G 2G2, Canada, **Department of Bioscience and Biotechnology, Faculty of Agriculture, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, 819-0395, Japan, and ‡‡‡INSERM, Institut de Recherche en Infectiologie de Montpellier, 34293 Montpellier, France

Edited by Gerald W. Hart

Despite impressive progress made over the past 20 years in our understanding of mycolylarabinogalactan-peptidoglycan (mAGP) biogenesis, the mechanisms by which the tubercle bacillus Mycobacterium tuberculosis adapts its cell wall structure and composition to various environmental conditions, especially during infection, remain poorly understood. Being the central portion of the mAGP complex, arabinogalactan (AG) is believed to be the constituent of the mycobacterial cell envelope that undergoes the least structural changes, but no reports exist supporting this assumption. Herein, using recombinantly expressed mycobacterial protein, bioinformatics analyses, and kinetic and biochemical assays, we demonstrate that the AG can be remodeled by a mycobacterial endogenous enzyme. In particular, we found that the mycobacterial GlfH1 (Rv3096) protein exhibits exo-β-D-galactofuranose hydrolase activity and is capable of hydrolyzing the galactan chain of AG by recurrent cleavage of the terminal β-(1,5) and β-(1,6)-Galf linkages. The characterization of this galactosidase represents a first step toward understanding the remodelling of mycobacterial AG.

The mycobacterial cell envelope exhibits remarkable complexity and provides a permeability barrier against drugs as well as the host’s immune attack. Despite being indexed as Gram-positive bacteria, the mycobacterial cell-wall presents characteristics reminiscent of Gram-negative bacteria, such as an outer membrane and, thus, differentiates the Mycobacterium genus from the majority of other prokaryotic organisms.

From the inside to the outside, the mycobacterial cell envelope is constituted of a plasma membrane enveloped by the mycolylarabinogalactan-peptidoglycan (mAGP) complex. The mAGP is composed of a central highly branched arabinogalactan (AG) polysaccharide linked to the peptidoglycan at the reducing end and to the long-chain mycolic acids at the nonreducing end. An outer membrane, also called the mycomembrane, analogous but dissimilar in its constitution to that of Gram-negative bacteria, contains in its inner leaflet mycolic acid residues covalently attached to arabinogalactan (1). In Mycobacterium tuberculosis, the outer membrane is further composed of numerous lipids and noncovalently bound glycoconjugates such as lipomannan, lipoarabinomannan, diacylglycerol, glycopeptidolipids, phenolic glycolipids, sulfolipids, trehalose dimycolate, phthiobiose dimycocerosate, di- and pentacyl trehaloses, and triacylglycerols, the relative amounts of which vary between species (2). mAGP represents an essential and major constituent of the envelope (3). In this composite molecule, AG is the polysaccharide providing the physical link between the PG and the mycolic acids producing a viscous hydrophilic region (4). Although uncertainties regarding some structural features remain, the overall architecture of the AG is now well-established, especially in the case of Mycobacterium tuberculosis (5, 6).

Unlike other bacterial polysaccharides, the primary structure of AG is not based on the repetition of a short unit but is rather composed of a linker to the PG and a galactan chain substituted by several arabinan chains. In mycobacteria, AG is linked from the reducing end to the PG via a disaccharide-phosphate unit α-L-Rhap-(1,3)-D-GlcpcNAc(1, P) leading to the presence of about 1.3 molecules AG/10 repeating units of PG and resulting in the substitution of 10–12% of the N-glycolylmuramic acid (MurNGlc) residues via a phosphodiester bond. From there

The abbreviations used are: mAGP, mycolylarabinogalactan-peptidoglycan; AG, arabinogalactan; HPAEC, high performance anion exchange chromatography; m.o.i., multiplicity of infection; hpi, hours post infection; OADC, oleic acid-albumin-dextrose-catalase; PG, peptidoglycan; equilibrate buffer A, phosphate buffer 20 mM, pH 7.4, NaCl 500 mM, imidazole 40 mM, Tween 20 0.05%; wash buffer A’, equilibrate buffer A, ATP 5 mM, MgCl₂ 10 mM, adjust the pH by NaOH to 8.5; elution buffer B, phosphate buffer 20 mM, pH 7.4, NaCl 500 mM, imidazole 250 mM, Tween 20 0.05% (v/v); buffer C, Tris 40 mM pH 7.4, DTT 0.1 mM, NaCl 150 mM, NaN₃ 0.1%, LDAO 0.2% (m/v).
extends the galactan moiety, a linear chain with alternating 5- and 6-linked \(\beta-d\)-Galf residues with a length estimated of around 25 residues (5). Arabinan chains are attached via the O-5 of the \(\beta-(1,6)\)-Galf residues at positions 8, 10, and 12 of the galactan chain although it has been proposed recently that only two of the three positions are substituted by arabinan chains; however, the exact positions of these two attached arabinan units remain unknown (5, 6). These arabinan chains consist of a backbone of \(\alpha-5\)-linked \(\alpha-d\)-Araf residues with several \(\alpha-(1,3)\)-linked residues forming 3,5-Araf/branching points. The nonreducing end forms a hexa-arabinoside motif, \((\beta-d\)-Galf-(1,2)\(-\alpha-d\)-Araf)\(2\)-3,5-\(\alpha-d\)-Araf(1,5)-\(\alpha-d\)-Araf, esterified with MAs on position 5 of the terminal \(\beta-d\)-Araf and the terminal 2-\(\alpha-d\)-Araf residues.

The structure of the mycobacterial cell envelope is expected to undergo numerous modifications, especially during cell growth, division, under antibiotic stress conditions, and also during infection of the host. It is very likely that survival of mycobacteria during infection is highly dependent on the plasticity of its cell envelope (7). Remodeling this envelope is a key mechanism that, for example, \(M.\) \(tuberculosis\) uses to adapt to the changing environment of the host by heavily modifying the PG, AG, and mycolic acids.

Although AG is often considered as a polysaccharide with a rigid structure, its structure is potentially amenable through the action of specific glycosidases, either by conventional hydrolysis or by transglycosylation. This hypothesis is supported by the observation of a yet unidentified endoarabinase activity able to cleave the arabinan chain in \(Mycobacterium \)\(smegmatis\) \((8)\) and by the discovery of the extractable lipid dimycyol diarabinanoglycerol (DMAG) in the cell envelope of \(Mycobacterium \)\(bovis\) BCG \((9)\). In this context, DMAG can also be viewed as the transglycosylation product of a mycolylated arabinan moiety onto glycerol by the action of a transglycosidase.

In this work, we sought to identify glycosidases specific to mycobacterial AG that could dynamically modify its structure. We present the first identification and characterization of a galactofuranohydrolase in \(M.\) \(tuberculosis\) (GlfH1), which acts processively on the galactan chain of AG. This enzyme has unusual characteristics among those with similar functions in its ability to cleave two types of glycosidic linkages and the absolute \(Ca^{2+}\) need for functional activity. GlfH1 is also an unusual member of the GH5_13 CAZy subfamily, which comprises almost exclusively glucohydrolases and mannosidases.

**Results**

**Identification of a novel galactosidase from \(M.\) \(tuberculosis\)**

The seminal observation of arabinan-degrading activity in \(M.\) \(smegmatis\) introduced the idea that mycobacteria may produce a set of specific glycosidases designed to remodel endogenous cell wall polysaccharides \((8)\). Hence, we initiated the search of such enzymes by looking for galactosidases capable of cleaving the galactan chain of AG in \(M.\) \(tuberculosis\). To initiate this work, a first search in the CAZy database (February 2019 update) failed to identify possible galactofuranohydrolases out of the 31 predicted glycosidases of \(M.\) \(tuberculosis\). Based on the known similarity between \(\beta\)-galactosidase and \(\alpha\)-arabinase activities \((10)\) and the structural similarity between \(\alpha\)-Araf and \(\beta\)-Galf, we have used \(\alpha\)-\(\alpha\)-arabinofuranosidases as a template to search for potential \(\beta\)-galactofuranosidases \((EC\ 3.2.1.146)\). To this date, \(\alpha\)-\(\alpha\)-arabinofuranosidases \((EC\ 3.2.1.55)\), endo-\(\beta\)-xylosidases \((EC\ 3.2.1.-)\), arabinanases \((EC\ 3.2.1.99)\), \(exo\)-\(\alpha\)-\(\alpha\)-1,5-arabinanases \((EC\ 3.2.1.-)\), and \(\beta\)-\(\beta\)-arabinofuranosidases \((EC\ 3.2.1.-)\) have been characterized in numerous CAZY GH families. Among all these families, only the GH family 3 possesses two members, \(Rv0186\) \((NP\_214700.1)\) and \(Rv0237/LpqI\) \((YP\_177702.1)\) \((11)\) within \(M.\) \(tuberculosis\) H37Rv annotated ORFs. A third potential galactofuranosidase candidate, \(Rv3096/GlfH1\) \((NP\_217612.1)\) was eventually found in the GH5 family, which comprises \(\beta\)-xylosidases \((EC\ 3.2.1.-)\) often sharing sequence similarities and activity with \(\alpha\)-\(\alpha\)-arabinofuranosidases \((12–14)\). Among these three candidates that were initially produced as recombinant proteins in \(E.\) \(coli\) and purified, \(Rv0186\) was not able to be expressed correctly \((Fig.\ 1)\) and \(Rv0237/LpqI\) did not show any activity using synthetic substrate \((Table\ S1)\) or \(M.\) \(bovis\) BCG AG. However, \(Rv3096\), subsequently designed GlfH1, showed a glycosidic activity toward the galactan moiety of \(M.\) \(bovis\) BCG AG by releasing free galactose \((Fig.\ S2)\). This protein was selected for further analysis.

The \(Rv3096\) gene is composed of 1140 bp encoding a 379–amino acid residue protein predicted to contain a 27–amino acid signal peptide by the SignalP-5.0 server \((15)\). The molecular weight of GlfH1 without the signal peptide was calculated at 40,382 Da. Sequence analysis by SSDB showed a Pfam Cellulase GH5 motif spanning residues 150 to 282 \((Fig.\ 1A)\). GlfH1 was classified in family GH5_13 in the CAZy database. Surprisingly, to this date, this family only contains glycosidases that use glucose or mannose-containing polysaccharides as substrates \(e.g.\) mannosidases, mannanases, cellulases, xylanases \((EC\ 3.2.1.-)\) and so on. All members of this CAZY family are \(\beta\)-retaining glycosidases.

Based on multiple sequence alignments of 13 different mycobacterial members from the GH5_13 family generated by the MUSCLE algorithm, all 13 sequences exhibited a high similarity with a maximum of 36% in the Hamming dissimilarity test \((56%\) of the residues are conserved between all the sequences in the alignment) \((Fig.\ S3)\). Among these sequences, 15 aspartic acid residues and 7 glutamic acid residues are conserved, suggesting that these amino acids are candidates for the pair of catalytic acid residues required for the glycosidase activity.

All 13 GH5_13 mycobacterial glycosidase sequences analyzed by PHYLINK Neighbor Joining phylogenetic analysis were placed into three major clades following the genome-based classification of \(Mycobacterium\) based on the concatenated sequence of 1941 core proteins from this genus \((16)\), which are the \(Mycobacterium\) \(abscessus-chelonae\) clade, the \(Mycobacterium\) \(fortuitum-vaccae\) clade, and the \(Mycobacterium\) \(tuberculosis-simiae\) clade \((Fig.\ 1B)\). A maximum-likelihood analysis for the same sequences confirms this observation \((Fig.\ S4)\) \((17)\).

In the subfamily GH5_13, the crystal structure of several glycosidases has been solved and the two catalytic amino acid residues, the first residue allowing the nucleophilic attack and the second operating in the acid/base transition, have been identified \((18–20)\). Of note, proteins of the GH5_13 family harbor two glutamic acid residues as catalytic amino acids in the CAZY
**GlH1 hydrolyzes mycobacterial galactan**

**A**

![Rv3096](image)

**B**

![Sequence Alignment](image)

**C**

![Tree](image)

**Figure 1. Bioinformatic analysis of GlfH1.** A, the annotation of the GlfH1 sequence shows the presence of a signal peptide (1–27) and a Pfam Cellulase motif (150–282). B, phylogenetic analysis of the GH5_13 subfamily mycobacterial glycosidases using the PHYLIP Neighbor Joining method with a bootstrapping of 100 replicates. C, multiple sequence alignment of GlfH1 with other members of the GH5_13 subfamily. Asterisks indicate experimentally assigned catalytic or essential residues for the activity of CmMan5A (18), CsMan5 (20), RmMan5B (19), and ManKB.

Production and purification of GlfH1

The protein GlfH1 of *M. tuberculosis* was produced in *E. coli* BL21 as a fusion protein using the pET32a expression vector (sequencing of vector pET32a-Rv3096 present in Fig. S3). Initial attempts at producing GlfH1 in *E. coli* showed that the fusion protein was mainly produced as inclusion bodies. Indeed, GlfH1 was found in large quantities in the insoluble fraction but barely observable in the soluble fraction (Fig. S6). As the main causes of inclusion body formation are the high rate of translation and folding of proteins in *E. coli* (10-fold higher than other observed eukaryotic cells) (22), dedicated Auto-Induction Media was thus used to minimize this phenomenon as much as possible, as described in “Experimental procedures.” AIM was formulated to allow the culture of iso-propyl 1-thio-β-D-galactopyranoside–inducible expression strains, initially without induction, and then to induce the production of the recombinant protein. The detergent N,N-dimethyldodecylamine N-oxide (LDAO) was used during the extraction step to stabilize the soluble fraction of GlfH1 and prevent it from aggregating once extracted. A three-step chromatography protocol was performed for purification (Fig. S7), the first one being the most important for GlfH1 isolation. The chaperone protein of *E. coli* associated with GlfH1 was finally washed away using ATP incubation. The second step served to concentrate the fractions obtained previously because it appeared that GlfH1 could not be concentrated by an ultrafiltration membrane. A final gel-filtration step was used to remove imidazole from the eluent and to recover GlfH1 directly into its storage buffer. The final purity of GlfH1 was confirmed...
GlfH1 hydrolyzes mycobacterial galactan

Of the 12 synthetic pNP-glycoside substrates tested, only pNP-β-D-Galp was hydrolyzed by GlfH1, confirming that it specifically hydrolyzes β-D-Galp. Importantly, GlfH1 showed no activity toward pNP-α-L-Araf although these activities are very often associated for other galactofuranosidas (23). Following initial characterization of the enzymatic activity, large quantities of pNP-β-D-Galp were synthesized through a new synthetic route to further characterize enzyme parameters. Synthesis was achieved in six steps from commercially available D-galactose (see supplemental experimental procedures). The analytical data (1H and 13C NMR spectroscopy and MS) corresponded to those reported in the literature (24–26).

To characterize the stereochemical specificity of a glycosidase, one must take into account the spontaneous interconversion between the two anomeric forms during the release of the monosaccharide. For a hexose like glucose, the rate of mutarotation is 0.015 min⁻¹ (27), which is incompatible with the usual structural analysis techniques. One approach consists of using methanol as a nucleophile instead of water to produce a methyl glycoside that does not undergo immediate mutarotation. GlfH1 retained 80% of its activity in the presence of 20% methanol (v/v) for short incubation times (e.g., 15 min, Fig. S9). This property was used to release 1-O-methyl Galp using pNP-Galp as substrate. The reaction products were then derivatized by per-trimethylsilylation and analyzed by GC/MS analysis (Fig. 2B, lower chromatogram) and compared with free galactose treated in the same way (Fig. 2B, upper chromatogram). For the galactose standard, peak assignment was performed by considering the literature data for the ratio of the four galactose isomers (αf, αβ, βf, and ββ) (28) and MS analysis to assign the ring size (Fig. S10). Hydrolysis of pNP-Galp by GlfH1 in the presence of 20% methanol and subsequent per-trimethylsilylation produces a compound with the same retention time and fragmentation profile as 1-O-methyl-β-Galp. This establishes that the hydrolysis of a β-linked galactose by GlfH1 retains the anomeric stereochic of the monosaccharide and that the enzyme is a β-retaining glycosidase.

Initial rate kinetics, enzymatic requirement, and inhibition

The effect of pH on GlfH1 activity was studied from pH 3.0 to 9.0 using two overlapping buffer ranges. The optimum pH for GlfH1 enzymatic activity was established to be 4.5 (Fig. 3A). The impact of several divalent cations, i.e., Mn²⁺, Mg²⁺, Ca²⁺, Ni²⁺, and Zn²⁺ on enzyme activity was assessed using their respective chloride salts. The activity of GlfH1 decreased in the presence of Ni²⁺ and Zn²⁺ by 76 and 86%, respectively (Fig. S11). However, with Ca²⁺, an increase of 35% in GlfH1 activity was observed. To better understand the role of Ca²⁺ on GlfH1 activity, the enzyme was incubated in the presence of EDTA or EGTA to deprive it of divalent metal salts. In both cases, the activity of GlfH1 was significantly diminished (Fig. 3B). On the other hand, the activity of GlfH1 in the presence of EDTA and EGTA was fully restored when Ca²⁺ was added thus confirming the requirement of this metal ion for enzymatic activity. Next, the kinetic constants were determined in the presence of 0–5 mM pNP-β-D-Galp under the standard activity assay con-

by SDS-PAGE (Fig. S8). This optimized purification pipeline enabled to reach a maximum yield of 2 mg/liter of active enzyme, and the minimum yield is 0.2 mg/liter of active enzyme.

All the enzymatic parameters (initial rate kinetics, optimum pH, and specificity) of GlfH1 were measured on three forms of the protein: The total form of the expressed protein (with signal peptide and tags), without the signal peptide, and without the tags. For the three forms, the results obtained are similar (Table S2). The results presented in this work were obtained with the total form of the GlfH1.

GlfH1 is a β-retaining exo-β-D-galactofuranohydrolase

To determine the substrate specificity of GlfH1, its hyrolytic activity was measured against synthetic pNP-glycoside substrates differing by the nature of their monosaccharide but also by their anomeric stereochic and ring size (Fig. 2A).
**GlfH1 hydrolyzes mycobacterial galactan**

**GlfH1 cleaves alternately β-(1,5) and β-(1,6) linkages**

The enzymatic activity of GlfH1 on a synthetic substrate that mimics the structure of its natural substrate, the galactan moiety of AG, was next assayed. To do so, a homologous series of oligosaccharides of 5 to 12 Gal residues linked to a α-(Rhap-(1,3)-α-GlcNAc-C8 motif (Lin-Galf), with alternating β-(1,5) and β-(1,6) glycosidic linkages (Fig. 4A), synthesized previously (29), were used as substrates for GlfH1. The Lin-Galf, nomenclature is used with x as the total number of Gal residues and Lin for the presence of the disaccharide linker α-1-Rhap-(1,3)-α-1-GlcNAc-C8. Enzymatic hydrolysis of these oligosaccharides was carried out using standard conditions, with incubation times ranging from 2 to 120 h. GlfH1 was added regularly to maintain a significant hydrolytic activity. As shown in Figs. S12–S16, partial hydrolysis of oligosaccharides can be detected from 2 h of incubation with all tested compounds. Within 24 h of incubation, the initial oligosaccharides almost completely disappeared to give oligosaccharides of lower degree of polymerization. This is exemplified for Lin-Galfx (Fig. 4B), which finally generated low DP oligosaccharides (Lin-Galfx-4) as the major products together with free galactose after 120 h of incubation. All oligosaccharides were degraded, irrespective of the nature of the Gal residue at the nonreducing end, either β-(1,5) or β-(1,6). The nature of released oligosaccharides was then confirmed by MALDI-TOF, as shown in Fig. 4C for Lin-Galf12. Overall, these results confirm the nature of GlfH1 as an exo-galactofuranohydrolase that degrades mycobacterial galactan through the sequential cleavage of alternating β-(1,5) or β-(1,6) residues from the nonreducing end of the molecule.

**Mycobacterial AG is the endogenous substrate of GlfH1**

The exquisite substrate specificity of GlfH1 toward β-α-Gal strongly suggests that this enzyme may be involved in the remodeling of endogenous cell wall arabinogalactan. To define GlfH1’s activity on this substrate (Fig. 5A), soluble AG, AGP, and mAGP purified from *M. bovis* BCG were used under standard conditions. GlfH1 was added regularly to maintain a significant hydrolytic activity. The enzymatic reaction was analyzed by high performance anion exchange chromatography (HPAEC) and GC/MS. As shown on the HPAEC chromatogram (Fig. 5B), the release of free Gal residues following the GlfH1 treatment of AG, AGP, and mAGP was clearly observed between 5 and 7 min. However, GlfH1 preferentially uses AGP compared with AG and mAGP (Fig. 5B). The release of Gal from AGP was faster than from AG or mAGP under standard conditions (release efficiency AGP/AG/mAGP: 5.6:1:0.13). The nature of released monosaccharide as Gal was confirmed for all conditions by GC/MS analysis (Figs. S17–S19).

To further define the activity of GlfH1 on a mycobacterial endogenous substrate, soluble AG was incubated with and without GlfH1 for 48 h using standard conditions. The reaction mixtures were lyophilized, exchanged twice in deuterium oxide and subjected to high field 1H and 13C NMR spectroscopy analyses. As seen in the 1D 1H-NMR spectrum (Fig. 5C), the major residues in the nontreated AG include 2-, 3-, and 3,5-linked α-Araf, terminal β-Araf, 5- and 6-linked β-Galf residues (according to literature and further confirmed by 1H/1H-

---

**Figure 3. Enzymatic characteristics of GlfH1 Galf-ase.** A, the optimum pH of GlfH1 Galf-ase activity was determined using buffers of sodium citrate (3–6, open circle) and tris malate (6–9, open triangle). B, effect of calcium. + EDTA and + EGTA. GlfH1 was pre-incubated with 10 μmole EDTA or EGTA for 10 min at 37 °C before measuring the residual activity of GlfH1 according to the standard protocol. EDTA → Ca2+ and EDTA → Ca2+. GlfH1 was pre-incubated with 10 μmole EDTA or EGTA for 10 min at 37 °C, then 50 μm of CaCl2 were added. Residual activity was measured according to the standard protocol. C, Galf-ase activity was quantified using pNP-β-α-Gal as a substrate. Enzyme assays were carried out at 37 °C with 0.5 μg of enzyme in 20 μl citrate, pH 4.5, citrate 20 mM buffer in a total volume of 100 μl.
TOCSY and $^1$H/$^{13}$C-HSQC analyses (6) NMR analysis of GlfH1-treated AG showed a similar pattern with two significant differences. First, a clear H1 signal appeared at $\delta$ 4.57 (1,2 $J_{H,H}$, 7.8 Hz) that was associated with a C1 signal at $\delta$ 97.8. This species was subsequently identified as free -Galp, according to H/C correlations (H2/C2, H3/C3, and H4/C4 at $\delta$ 3.48/73.2, 3.64/74.1, and 3.93/71.7, respectively, (Fig. 5C) and coupling constants (2,3 $J_{H,H}$, 9.9 Hz, 3,4 $J_{H,H}$, 3.5 Hz, and 4,5 $J_{H,H}$, 1 Hz). A more discreet $^1$H/$^{13}$C signal was also observed at 5.25/93.6 ppm, which was assigned as a -Galp residue, although its spin system could not be entirely established because of lower quantity and overlapping signals (Fig. 5D). The occurrence of these signals confirmed the liberation of free Galp following the GlfH1 treatment. The observation of a -Galp as the major species and -Galp as the minor species and no free Galp is in agreement with the known equilibrium of Gal in solution (30/64/2.5/3.5 for ap/βp/αf/βf at 31 °C) (28). Second, the appearance of free Gal signals was concomitant with the specific decrease of the signal intensities of both -5-β-Galp and -6-β-Galp anomer signals at $\delta$ 5.02 and 5.23, respectively in the 1D $^1$H-NMR spectrum (Fig. 5C). Relative quantification of these two signals versus all other anomer signals showed an overall 25% reduction of both signals compared with nontreated AG. It is noteworthy that the signal intensities of all Ara residues did not show any variation between each other, further confirming that the arabinan moiety of AG remains unaffected by GlfH1.

Deletion in Mycobacterium smegmatis of glfH1 (MSMEG_5877) leads to an attenuated growth phenotype in an amoebal infection model

To investigate the role of GlfH1 in bacterial physiology and infection, we aimed to construct a deletion mutant of the M. smegmatis GlfH1–closest orthologue, which was identified as MSMEG_5877 by a BLAST search limited to M. smegmatis mc2155 (taxid: 246196). The global alignment of MSMEG_5877 sequence with the sequence of GlfH1 exhibits identity and a similarity score of 73.3 and 82.1%, respectively (Fig. 6A) and coupling constants (2,3 $J_{H,H}$ = 9.9 Hz, 3,4 $J_{H,H}$ = 3.5 Hz, and 4,5 $J_{H,H}$ < 1 Hz). A more discreet $^1$H/$^{13}$C signal was also observed at 5.25/93.6 ppm, which was assigned as a -Galp residue, although its spin system could not be entirely established because of lower quantity and overlapping signals (Fig. 5D). The occurrence of these signals confirmed the liberation of free Gal following the GlfH1 treatment. The observation of a -Galp as the major species and -Galp as the minor species and no free Galp is in agreement with the known equilibrium of Gal in solution (30/64/2.5/3.5 for ap/βp/αf/βf at 31 °C) (28). Second, the appearance of free Gal signals was concomitant with the specific decrease of the signal intensities of both -5-β-Galp and -6-β-Galp anomer signals at $\delta$ 5.02 and 5.23, respectively in the 1D $^1$H-NMR spectrum (Fig. 5C). Relative quantification of these two signals versus all other anomer signals showed an overall 25% reduction of both signals compared with nontreated AG. It is noteworthy that the signal intensities of all Ara residues did not show any variation between each other, further confirming that the arabinan moiety of AG remains unaffected by GlfH1.

Deletion in Mycobacterium smegmatis of glfH1 (MSMEG_5877) leads to an attenuated growth phenotype in an amoebal infection model

To investigate the role of GlfH1 in bacterial physiology and infection, we aimed to construct a deletion mutant of the M. smegmatis GlfH1–closest orthologue, which was identified as MSMEG_5877 by a BLAST search limited to M. smegmatis mc2155 (taxid: 246196). The global alignment of MSMEG_5877 sequence with the sequence of GlfH1 exhibits identity and a similarity score of 73.3 and 82.1%, respectively (Fig. 6A). The three amino acid residues essential for GlfH1’s activity (Glu-170, His-242 and Glu-268) are thoroughly conserved.

To this effect, we first cloned MSMEG_5877 into pET32a to purify the protein from E. coli. The enzymatic parameters of GlfH1 and MSMEG_5877, including $K_m$ and $k_{cat}$, optimum pH and Ca$^{2+}$ requirement were similar (Fig. 6B), strongly supporting the fact that both enzymes share similar functions in M. smegmatis and M. tuberculosis.
A deletion mutant of \( \text{glfH1} \) was therefore constructed in \( M. \text{smegmatis} \) using mycobacterial recombineering which resulted in the replacement of MSMEG\_5877 with a hygromycin-resistant cassette (Fig. 6, upper panel). Putative \( \text{glfH1::hyg} \) colonies were genotypically analyzed using a combination of a PCR and Sanger sequencing analysis in which the junctions between the native \( M. \text{smegmatis} \) chromosome, cloned sequences left and right of MSMEG\_5877 and the hygromycin cassette were confirmed (Fig. 6, lower panel).

In vitro growth of \( \text{glfH1::hyg} \) was not different from that of WT or complemented strains in liquid or in solid media (Fig. S20a). However, we observed a marginal increase in sensitivity of the mutant to rifampicin, but not to imipenem (Fig. S20b). Considering the hydrophobicity of rifampicin, we speculated that the slight increase in sensitivity to this antibiotic indicated a defect in the cell wall of the bacteria permitting more efficient access of rifampicin to its cytosolic target. However, analysis of AG extracted from WT, \( \Delta \text{glfH1} \), and complemented \( \Delta \text{glfH1} \) strain showed that \( \text{glfH1} \) deletion did not have an observable impact on AG structure (Fig. S21), suggesting that modifications are confined to a subclass of AG and cannot be assessed by a global analysis. We thus addressed whether this moderate cell wall defect would translate into an attenuated survival of the mycobacterium in the harsh environment it would encounter inside a phagocytic cell.

\( M. \text{smegmatis} \), which does not survive infection in macrophages, was shown to survive and proliferate in \( \text{Acanthamoeba polyphaga} \) (31). We therefore infected \( \text{Acanthamoeba polyphaga} \) with \( M. \text{smegmatis} \) WT, \( \Delta \text{glfH1} \), and its complemented strain at an m.o.i. of 10 and monitored the growth of the bacteria 2 h post infection (hpi), 24 hpi, and 48 hpi by first...

---

**Figure 5.** The \( \text{GlfH1} \) protein is an \( \text{exo-}\beta\text{-D-Gal} \)-ase able to remodel the galactan chain of mycobacterial AG. A, schematic structure of the \( M. \text{tuberculosis} \) AG. B, digestion of \( \text{GlfH1} \) endogenous substrates AG, AGP, and mAGP by \( \text{GlfH1} \) releases free galactose. Hydrolysis of the \( M. \text{bovis} \) BCG AG, AGP, and mAGP was performed at 37 °C for 48 h with 200–400 \( \mu \)g of substrate, 20 \( \mu \)g of \( \text{GlfH1} \) in 1 ml of citrate 20 mM, pH 4.5 buffer. After incubation, the samples were heated at 100 °C for 5 min and centrifuged at 10,000 × g for 2 min. Supernatants were collected for HPAEC-PAD analysis. The retention time of signals in the monosaccharide elution region were compared with free galactose. C, NMR analysis of the \( M. \text{tuberculosis} \) AG before and after the hydrolysis of \( \text{GlfH1} \). 1 mg of \( M. \text{tuberculosis} \) AG was incubated with 100 \( \mu \)g of \( \text{GlfH1} \) in 1 ml of 20 mM citrate, pH 4.5, \( \text{CaCl}_2 \), 20 mM buffer for 48 h. Following incubation with \( \text{GlfH1} \), 1D \( ^1\text{H}-\text{NMR} \) spectra show the concomitant appearance of \( \alpha/\beta\text{-Galp} \) signals and the decrease of two signals (blue and red triangles) identified as -5-\( \beta\text{-Galp} \) and -6-\( \beta\text{-Galp} \), based on \( ^1\text{H}\text{-}^1\text{H} \) TOCSY and \( ^1\text{H}\text{-}^{13}\text{C} \) HSQC NMR spectra (D).
lysing the amoebal cells and then plating the lysates on 7H10-OADC to monitor the colony-forming unit(s) per milliliter. Massive extracellular growth of the bacteria, which could not be controlled by washing the amoebal monolayers or by addition of antibiotics, impeded extension of the infection period to time points beyond 48 hpi. As clearly shown in Fig. 6, H9004 glfH1 suffered of a significant growth impediment in amoebae with a bacillary load of approximately half a log less than that of the WT strain at 48 hpi. Genetic complementation of H9004 glfH1, on the other hand, led to a partial restoration of the WT phenotype, supporting a role of GlfH1 in intra-amoebal growth.

Discussion

Significant attention has been paid to glycosidases at the therapeutic level in the past two decades: several glycosidase inhibitors have been discovered for potential use in the treatment of cancer, viral infections, neurodegenerative diseases, autoimmune diseases, and diabetes (32). In contrast, very little interest has been given to mycobacterial glycosidases, despite the fact that they may play a crucial role in the virulence and persistence of M. tuberculosis. The absence of Galβ in mammalian glycoconjugates further strengthens its biosynthetic pathway as an attractive therapeutic target (33).

Herein, we demonstrated the existence of the first endogenous galactosidase that targets mycobacterial AG. The behavior of GlfH1 is surprising in several ways and shows that it is an unusual glycosidase. First, the failure of GlfH1 to hydrolyze pNP-β-D-galactopyranoside suggests that its enzymatic activity is specific to the furanose form of the monosaccharide. We also demonstrated that GlfH1 is an exo-Galβ-ase that processes

Figure 6. glfH1 knock-out in M. smegmatis results in attenuation of growth in amoebae. A, pairwise sequence alignment of GlfH1 with MSMEG_5877. Asterisks indicate experimentally assigned catalytic or essential residues for the activity of GlfH1. B, comparison of the functional characteristics of GlfH1 and MSMEG_5877, its orthologue in M. smegmatis. C, deletion of glfH1. Upper panel, scheme of the recombineering strategy followed to generate the knockout. Blue and green dotted lines flanking the hygR cassette that replaced glfH1 in H9004 indicating amplicons that can only be generated from the genome of this strain. Lower panel, confirmation of a H9004 glfH1 mutant by PCR. D, growth of the H9004 glfH1 mutant is marginally compromised in amoeba. WT M. smegmatis, H9004 glfH1, or its complement strain were co-cultured with amoebae for 2, 24, and 48 h using an m.o.i. of 10:1. Histograms and error bars represent mean ± S.E., respectively, calculated from five independent experiments (Kruskal-Wallis test with Dunn’s multiple comparisons post-test; ***, p < 0.001).
alternating $\beta$-(1,5) and $\beta$-(1,6)-linkages. For instance, none of the $\beta$-galactofuranosidas reported so far in Aspergillus or Penicillium showed a similar activity on both $\beta$-(1,5) and $\beta$-(1,6) linkages. Moreover, because of the structural similarities of $\alpha$-l-arabinofuranosides and $\beta$-d-galactofuranosides, numerous $\alpha$-l-arabinofuranosidas also show a concomitant $\beta$-d-Galf-ase activity (34). For example, the $\beta$-galactofuranosisdases found in Streptomyces display both $\alpha$-l-arabinofuranosidase and $\beta$-d-Galf-ase activities (23); however, unlike these enzymes, GlfH1 shows no activity toward arabinofuranosides. In addition, this enzyme shows an absolute requirement for Ca$^{2+}$ to function, which is a very rare feature among glycosidases as typically only a pair of carboxylate residues (Asp or Glu) is required in the catalytic site. The need of Ca$^{2+}$ for activity was observed for an endo-$\alpha$-l-arabinanase of GH43 in Bacillus subtilis, several exo-mannosidas of GH38, GH47, GH92, as well as an $\alpha$-glucosidase and an $\alpha$-galactosidase of GH97 (35, 36). However, there is no observation of such a requirement in the $\beta$-Galf-ases reported so far. Interestingly, three histidine residues are conserved in several glycosidas from the GH5_13 family of which their structure has been solved (Fig. 1C). They may be implicated in the stabilization of a Ca$^{2+}$ ion that is required for the hydrolytic activity of the enzyme.

Besides, it should be noted, after the research that led to the discovery of GlfH1, a $\beta$-d-galactofuranosidase (ALJ47066.1) is now classified in the GH5_13 family (update 2019–10–16). The gene coding for this protein has been identified in Bacteroides ovatus ATCC 8483 (37), nevertheless its function has only been inferred for the moment. An alignment with the sequence of GlfH1 shows 41.1% identity and strongly suggests a shared ancestry between them.

The analysis of the stereochemical specificity of GlfH1 done by GC/MS suggests clearly that GlfH1 acts by retention of the anomeric configuration. We can also estimate the distance between two catalytic residues to be around 5.5 Å, as the fact that two catalytic residues in retaining glycosidases are 5.5 Å apart, instead of 10.5 Å in inverting glycosidas (38). This result will be very helpful for the 3D structure determination and the identification of its catalytic residues. Hydrolysis of the synthetic substrate that mimics the structure of AG by GlfH1 suggests that this enzyme hydrolyzes the oligosaccharides regardless of the nature of the galactofuranose residue at the nonreducing end. Hence, it would be instructive to investigate if GlfH1 has the same rate of hydrolysis for the $\beta$-Galf-(1,5) and $\beta$-Galf-(1,6) residues. Different release efficiencies by GlfH1 were seen between AG, AGP and mAGP, with AGP being the best substrate. Our hypothesis is that AG, because of its unusual branching pattern, has a structure compacted by hydrogen bonds (39), especially when it is no longer bound to PG and mycic acid. When linked to PG (AGP), fewer hydrogen bonds are formed, and AG becomes more accessible to GlfH1. In contrast, when simultaneously bound to PG and mycic acids (mAGP), accessibility of AG to GlfH1 is compromised because of the presence of the mycic acid layer. That the deletion of glfH1 (MSMEG_5877) shows an attenuated growth phenotype inside amoeba strongly suggests the importance of GlfH1 during infection. However, in-depth study of the role of GlfH1 in bacterial growth and infection using M. tuberculosis mouse models remains to be accomplished. Also, the regulation of GlfH1-mediated AG remodeling and possible presence of additional yet unidentified enzymes that may have partially overlapping function to GlfH1 will need to be taken into account in further detailed *in vivo* studies of AG remodeling and its impact of bacterial physiology and pathogenicity.

The identification of the first glycosidase involved in the degradation of the AG allows us to speculate on the presence of a functional complex of polysaccharide remodeling enzymes and recycling of degradation products. According to the KEGG pathway of galactose metabolism in M. tuberculosis (RRID: SCR_018145 (identifier map00520)), several enzymes are required for the Leloir pathway to recycle galactose, including galactose mutarotase, galactokinase Rv0620/galK (NP_215134.1), galactose-1-phosphate uridyltransferase Rv0618/galTa (WP_003900189.1) and Rv0619/galTb (CCP43360.1), UDP-galactopropyanose mutase Rv3809c/glf (NP_218326.1), and galactosidase GlfH1/Rv3096 (NP_217612.1). For instance, the galactose mutarotase remains unidentified. Therefore, it would be essential to characterize this enzyme, as well as the means used by mycobacteria to transfer galactose back to the cytoplasm, so that the identification of the enzymes implicated in the Leloir pathway will be completed. On the other hand, the recycling of PG in *M. tuberculosis* and in *M. bovis* BCG has been reported recently. The lipoprotein LpqI expresses an *exo*-N-acetylglucosaminidase that can cleave PG fragments *in vitro*. LpqI cleaves GlcpNAc-MurfNAc disaccharides, once its stem peptide is removed, allowing the release of free MurpNAc. The MurpNAc is then followed by $\beta$-lactyl-ether cleavage to release lactate, which can be used by the cell under aerobic conditions (11). The highlight of these two mycobacterial enzymes, LpqI and GlfH1, allows us to progress in the understanding of the remodeling and the recycling of mycobacterial major cell envelope constituents.

Finally, the importance of GlfH1 in modifying the structure of mycobacterial arabinogalactan may be associated with the growth of mycobacteria during infection, as shown by the infection of *A. polyphaga* with the *M. smegmatis* mutants. Interestingly, a previous study reported that GlfH1 is among the 41 proteins overexpressed in a *M. tuberculosis* mutant strain deficient in the expression of *perM*, when cultivated at pH 4.5 (40). *PerM* is an essential protein for the persistence of *M. tuberculosis* within IFN-$\gamma$–activated macrophages following the acidification of phagosomes at pH 4.5 (40). Among the remaining 40 proteins of known functions, most of them are involved in the remodeling of elements constituting the envelope, such as PG or mycic acids. In this biological context, the optimum activity pH of GlfH1 at 4.5 is completely adapted to the environment of *M. tuberculosis* during chronic infection of the macrophages. That GlfH1 appears necessary for optimal growth in *A. polyphaga* is also compatible with this hypothesis when considering that mycobacteria reside also in acidic organelles in a similar way to what happens in macrophages (41). The importance of GlfH1 in cell wall remodeling is further emphasized by its genomic environment, as it is adjacent to Rv3097c, also known as lipY, that encodes a well-characterized PE protein expressing triacylglycerol hydrolase activity. LipY participates to consumption and reprocessing of both mycobacterial
lipids and host lipids, thereby contributing to persistence of the tubercle bacillus inside foamy macrophages (42–44).

Based on these observations, the detailed expression of GlfH1 should be carried out to uncover the molecular mechanisms of resistance within the phagolysosome. Likewise, future studies will be necessary to investigate the putative contribution of GlfH1 during mycobacterial cell division, virulence, and persistence.

**Experimental procedures**

**General**

For all the preparation of buffers, substrates, reagents, enzyme reactions, ultrapure water with an electrical resistivity > 18 MΩ cm (ELGA, Veolia Water STI, UK) was used. All chemical substances used were purchased from Sigma-Aldrich or Acros Organics (France) depending on the availability in stock. HisTrap™ HP and HiTrap™ Desalting prepacked columns were purchased from GE Healthcare. Micro BCA™ Protein Assay Kit was purchased from Thermo Fisher Scientific for determination of protein concentration. Mass spectrometry grade trypsin (porcine) was purchased from G-BioSciences (St. Louis, MO).

**Bacterial strains, plasmids, and growth conditions**

For recombinant protein production, *E. coli* strains DH5α (New England Biolabs) and BL21 (DE3) (Invitrogen) were used for cloning and protein expression. pET32a (Invitrogen) was used as an expression vector. The *E. coli* strains were grown in Auto-Induction Media (recipe Formedium) or Luria broth (LB) agar with the appropriate concentrations of antibiotics (ampicillin 100 μg/ml) at 25 °C overnight with agitation.

For deletion of *glfH1* in *M. smegmatis*, *Escherichia coli* HB101 was cultured in lysogeny broth (LB) or LB agar and *Mycobacterium smegmatis* mc²155 was cultured in 7H9 broth containing 10% oleic acid-albumin-dextrose-catalase (OADC) supplement and 0.05% v/v Tween 80 (7H9-OADC) or 7H10 containing 10% OADC supplement (7H10-OADC), except where otherwise stated. To maintain or select for plasmids in *E. coli*, the following antibiotic concentrations were used: Hygromycin 75 μg/ml and kanamycin 50 μg/ml. For selection in *M. smegmatis*, the following concentrations were used: Hygromycin 100 μg/ml and kanamycin 50 μg/ml.

**Plasmid generation**

Cloning of GlfH1—The GlfH1 open reading frame (ORF) was PCR amplified with the left primer 5’-CACCGTCAAGAGCCGCGC-3’ and right primer 5’-ACGGAAAGTCTTACACGAGCCG-3’. The PCR amplicon was restricted with HindIII only and ligated to EcoRV-HindIII-linearized pET32a.

**MSMEG_5877 expression in E. coli**—Briefly, the *MSMEG_5877* ORF plus an N-terminal tobacco etch virus cleavage site was PCR amplified using the primer set 5’-GAGAAATCTGTA-CTTCAAGGGAGTGACACACCCCGGC-3’ and 5’-ACGAGGAACTTCTAACGAGCCGTCAGGTT-3’. The PCR amplicon was restricted with HindIII only and ligated to EcoRV-HindIII-linearized pET32a.

**MSMEG_5877 knock-out and complementation**—To accomplish targeted deletion of *MSMEG_5877*, a construct containing an allelic exchange substrate for the replacement of *MSMEG_5877* with a hygromycin cassette was first generated. Fragments of DNA flanking *MSMEG_5877* on the left and right arms were first PCR amplified using the primer sets 5’-AGCTGAACTTCTTACGAGGAGTGTTG-3’ and 5’-ACGGAACTTCTTACGAGGAGTGTTG-3’ (left arm) and 5’-ACGGAGATCTTACATGCGGAACACGGCC-3’ and 5’-ACGGAGATCTTACATGCGGAACACGGCC-3’ (right arm). The left arm was first cloned into the Spel-HindIII restricted pJSC347 (45), followed by the right arm using the Xbal and KpnI restriction sites leading to pJSC347-glfH1. For genetic complementation of the Δ*glfH1* mutant, a fragment of DNA containing the *MSMEG_5877* ORF, its putative 5’-UTR and promoter as well as a C-terminal Human influenza hemagglutinin (HA) tag was PCR amplified with the primer set 5’-GAGAGATCTAGAACACTGACACTGACACTGACACT-3’ and 5’-CGAGCTTCGGTGCACCACCAAAC-3’. The PCR amplicon was restricted with HindIII only and ligated to EcoRV-HindIII-linearized pET32a.

**Protein production and purification**

The sequence-confirmed recombinant plasmid pET32a-Rv3096 was transformed into *E. coli* BL21 (DE3) using the heat shock method. Transformants were cultured on LB agar plates containing 100 μg/ml ampicillin overnight at 37 °C. One single colony was selected and grown overnight in 10 ml of LB pre-culture with 100 μg/ml ampicillin at 37 °C with agitation, then enlarged into Auto-Induction Media containing 100 μg/ml ampicillin overnight at 25 °C with agitation. Cells were harvested by centrifugation for 15 min at 2450 × g, resuspended in lysis buffer (Tris 25 mM, pH 7.5, NaCl 300 mM, imidazole 20 mM) with Halt™ Protease inhibitor EDTA-free 1× (Thermo Scientific) and lysed by French press (SimoAminco) at 1650 psi. The soluble fraction and the pellet were separated by centrifugation for 30 min at 13,000 × g. Immobilized metal ion affinity chromatography was used for purification of GlfH1. The soluble fraction was purified by using ÄKTA star system (GE Healthcare Life Sciences) with a 5 ml HisTrap HP column (GE Healthcare Life Sciences). The three buffers used for chromatography were equilibrated buffer A, wash buffer A⁻, and elution buffer B. The purification contained a first step of loading the crude protein onto the column, which was realized at 1 ml/min, a second step of incubation by using the wash buffer A⁻ at room temperature for 1 h, followed by a third step to elute the protein with 100% of buffer B at 1 ml/min. Fractions containing GlfH1 were collected and diluted 10 times by buffer A before concentration. The concentration of the protein was performed as described before with a 1 ml HisTrap HP column (GE Healthcare Life Sciences). Finally, the fractions containing GlfH1 were collected and desalted by using ÄKTA star system (GE Health-
GlfH1 hydrolyzes mycobacterial galactan

care Life Sciences) with a 5 ml HiTrap Desalting column (GE Healthcare Life Sciences). The buffer used for gel filtration chromatography and the storage of GlfH1 was buffer C. The GlfH1 protein was analyzed by SDS-PAGE, and its concentration was determined by the BCA method according to the protocol provided by the manufacturer.

Bioinformatics

Sequences for the GH5_13 subfamily CAZy glycosidases from mycobacteria were extracted from their NCBI reference genome. For all the protein sequences used, the presence of a signal peptide was sought with the SignalP 5.0 server (15) and the eventual signal peptide was removed from the sequence. Subsequently, the protein sequences were numbered without the signal peptide. The results obtained from the consultation of databases were respectively obtained with the release 230 (Feb 2019) for the NCBI (RRID:SCR_006472), release 2019_02 (13 Feb 2019) for UniProt (RRID:SCR_002380), 20 February release for RCSB PDB (RRID:SCR_012820) and February 2019 release for CAZy (RRID:SCR_012909).

The Unipro UGENE v1.32 (Feb 2019) bioinformatic software was used for data sequence management, multiple sequence alignments, and phylogenic analysis (47). Multiple alignments were realized using the MUSCLE algorithm (48) with default parameters for best accuracy. Phylogeny analysis was performed using the PHYLIP Neighbor Joining method (49) with the Henikoff/Tillier PMB as distance matrix model, gamma distributed rates across sites (coefficient of variation of substitution rate among sites of 0.50) and a bootstrapping of 100 replicates with majority rule (extended) as the consensus type. The PhyML Maximum Likelihood method (50) was also used for confirmation with the Dayhoff substitution model and a bootstrapping of 100 replicates.

Initial rate kinetics, pH profile, inhibition, and cation requirement

The activity of GlfH1 was determined using a substrate synthetic, p-nitrophenol-α-galactofuranose (pNPGal). Enzyme assays were carried out with 0.5 µg of enzyme with 1 mM of pNPGal and 20 mM of different buffer (variable) with a total volume of 100 µl at 37 °C for 15 min. The incubation mixtures were pre-incubated at 37 °C with the enzyme for 15 min then the reaction was completed by adding 50 mM of CaCl2 to the standard enzyme assay. Residual activity was measured according to the standard protocol.

Mechanism type analysis

The stereochemistry of enzyme GlfH1 was determined using GC. 2 mM of pNPGal was hydrolyzed by GlfH1 in presence of 20% (v/v) methanol at 37 °C during 24 h then lyophilized prior to derivatization by per-trimethylsilylation as described below.

Enzyme product analyses

Hydrolysis of synthetic oligosaccharides by GlfH1—Hydrolysis of the synthetic C8-Linker-Galf, oligosaccharides was performed at 37 °C for a duration of 0–120 h with 15 µg of substrate, 10 µg of GlfH1, and 20 mM citrate pH 4.5 buffer with 20 mM of CaCl2 in a total volume of 100 µl; 10 µg of GlfH1 was added every 2 h during 8 h then every 24 h to compensate for inactivation. After incubation, the samples were heated at 100 °C for 5 min then centrifuged at 10,000 × g for 2 min. Supernatants were collected for HPAEC-PAD and MALDI-TOF analysis.

MALDI-TOF analysis—Oligosaccharides hydrolyzed by GlfH1 using synthetic C8-Linker-Galf, oligosaccharides as substrates were permethylated according to the method of Cicianu and Kerek (51) for MS analysis. Permethylated samples were dissolved in 50 µl of acetonitrile and 1 µl of the sample was mixed directly on the MALDI plate with 1 µl of superDHB (9:1 mixture of 2,5 dihydroxybenzoic acid and 2-hydroxy-5-methoxybenzoic acid) matrix solution (10 mg/ml in acetonitrile 50% v/v, 1 mM NaOH). MALDI-TOF spectra were performed by using 4800 Proteomics Analyzer mass spectrometer (Applied Biosystems, Framingham, MA) with reflection positive mode by delayed extraction using an acceleration mode of 20 kV, a pulse delay of 200 ns, and grid voltage of 66%. For each spectrum, 2000 laser shots were accumulated.

Hydrolysis of AG/AGP/mAGP by GlfH1—M. bovis BCG strain Pasteur was grown in flask in Sauton medium at 37 °C then harvested in late log phase. The preparation of AG from M. bovis BCG was performed as reported previously (52). Hydrolysis of the M. bovis BCG AG, AGP, and mAGP was performed with 200, 300, and 400 µg of substrate, respectively, 20 µg of GlfH1 in 1 ml of citrate 20 mM, pH 4.5, buffer with 20 mM of CaCl2 at 37 °C for 48 h with shaking. An additional 20 µg of GlfH1 was added every 24 h to maximize the hydrolysis. After incubation, samples were heated at 100 °C for 5 min and centrifuged at 10,000 × g for 2 min. The monosaccharides released by GlfH1 were separated by ultrafiltration with a 1 kDa filter, fragments < 1 kDa were collected for HPAEC-PAD and GC/MS analysis.

HPAEC-PAD analysis—To analyze the monosaccharides or oligosaccharides released by GlfH1, high performance anion exchange chromatography was done. A Dionex ICS5000 HPAEC system controlled by Chromelone software version 7.1 with pulse amperometric detection (PAD) was used. A CarboPac PA1 guard (2 mm × 50 mm) (Thermo Scientific) and analytical column (250 mm × 2 mm) (Thermo Scientific) for monosaccharides or a CarboPac PA-200 anion exchange analytical column (3 mm × 250 mm), equipped with a guard column PA-200 (3 mm × 50 mm) for oligosaccharides were employed. All separations were performed using 100 µM NaOH as the eluent and an elution gradient with sodium ace-
Progenitor. The 

0.25

\text{/H9262}

min. The temperature program used was

in deuterium oxide, and subjected to high field 1\text{H} and 1\text{3C} NMR spectroscopy analyses. The reaction mixtures were lyophilized, exchanged twice

bated with and without GlfH1 for 48 h using standard condi-

tions. Pulsed amperometric detector was used with standard quad wave potential provided by the manufacturer.

**Per-trimethylsilylation analysis**—The monosaccharides hydrolyzed by GlfH1, after lyophilization, were derivatized by 20 \mu l of pyridine and 20 \mu l of Bis-Silyl-TriFluoro-Acetamide at room temperature for 2 h, dried under nitrogen, and finally taken up in 100 \mu l of heptane. Derivatives were analyzed by GC/MS, a gas chromatograph (Thermo Scientific Trace GC Ultra) with a SolGel-1ms\textsuperscript{TM} Capillary column (0.25 mm 30 m, 0.25 \mu m film, SGE) was interfaced with a MS (TSQ Quantum GC, Thermo Scientific). The temperature program used was 120 °C to 240 °C with a slope of 2 °C/min; hold at 240 °C for 10 min.

**NMR analysis**—The soluble AG of \textit{M. tuberculosis} was incubated with and without GlfH1 for 48 h using standard conditions. The reaction mixtures were lyophilized, exchanged twice in deuterium oxide, and subjected to high field 1\text{H} and 1\text{3C} NMR spectroscopy analyses. Spectra were recorded on 21.4-T spectrometers equipped with 5-mm triple resonance cryoprobe inverse (TCl) cryoprobe with 1\text{H}, 1\text{3C}, 1\text{4C} cooled channels and a 1\text{1N} channel with a z-gradient. (Unité de Glycobiologie Structurale et Fonctionnelle, Infrastructure de Recherche- Très Hauts Champs- Résonance Magnétique Nucléaire, CNRS) where pro-

tons resonate at 900 MHz and 1\text{3C} resonates at 250 MHz, respectively. All samples were put in 5-mm tubes matched for D\text{2}O. Acetone was added as an internal standard, starting from a solution of 2.5 \mu l of acetone in 10 ml of D\text{2}O. All pulse sequences were taken from the Bruker library of pulse pro-

grams and then optimized for each sample.

**Generation of a MSMEG_5877 deletion mutant (\Delta glfH1)**

A deletion mutant of MSMEG_5877 was generated using the mycobacterial recombineering strategy (30). Briefly, an exponential phase culture of \textit{M. smegmatis} mc\textsuperscript{2}155 carrying pJV53 was washed with PBS and used to inoculate to an \textit{A}\text{\textsubscript{600}} = 0.02

79H supplemented with 0.2% succinate, 0.05% Tween 80 and 50 \mu g/ml kanamycin. This culture was incubated overnight at 37 °C with agitation after which acetamide was added (0.2% w/v final concentration) and the culture was incubated for an additional 3 h at 37 °C with shaking. The recombineering-proficient bacteria were then made electrocompetent through several washes with 10% glycerol– 0.05% Tween 80 at 4 °C and electro-

transformed with 100 ng of an allelic exchange substrate, which was prepared by restriction digest of pJSC347-glfH1 with NheI and KpnI and gel-purification of the 3.8 kb DNA band. Follow-

ing transformation, cells were recovered in 79H-OADC for 4 h at 37 °C before being spread out on 79H10-OADC plates containing 100 \mu g/ml hygromycin. After 5 days of incubation at 37 °C colonies were transferred to liquid media and PCR-genotyped with the primer sets 5'-GACGACCCCTAGCTCCTG-TCC-3' and 5'-CGGGTTGCGACGACTTCAC-3' (left arm) and 5'-GACTGTCGACACCCCTCAC-3' and 5'-GACACCGCC-

CCGGCGCCTGTA-3' (right arm), which only produced specific PCR amplicons in the \Delta glfH1 mutant and not in the WT progenitor. The \textit{A}\textit{glfH1} mutant was subsequently, electrotrans-

formed with the complementation construct, pMV306-MS-

MEG_5877 (see “Plasmid generation”) and the complemented colonies were selected on LB plates supplemented with kana-

mycin and confirmed by PCR.

**Co-culture of \textit{M. smegmatis} strains and \textit{A. polyphaga}**

\textit{A. polyphaga} cultures were grown and maintained in PYG (peptone yeast glucose) medium at 28 °C. For amoeba infection assays, cells were harvested and washed three times with Page’s modified Neff’s amoebae saline (PAS), resuspended in fresh PYG media, and adjusted to a final concentration of 5 × 10\textsuperscript{3} cells ml\textsuperscript{-1} (10\textsuperscript{7} cells ml\textsuperscript{-1} for microscopy experiments). \textit{M. smegmatis} cultures were pelleted, dispersed by 10 passages of the bacterial suspension through a 26-gauge needle, and then used to infect \textit{A. polyphaga} at an m.o.i. of 10:1. Following 6 h of incubation at 32 °C, three thorough washes in PAS buffer were performed, followed by 2 h of incubation in the presence of 100 \mu g/ml of amikacin, to kill all extracellular mycobacteria. The number of colony-forming unit(s) per milliliter was determined for each \textit{M. smegmatis} strain after lysing the \textit{A. polyphaga} monolayer with 0.5% SDS for 10 min at 32 °C, after 2, 24, and 48 h of co-culture.

**Author contributions**—L. S., A. V., K. T., T. L. L., S. P. V., L. K., Y. G., and C. M. conceptualization; L. S., A. V., S. V., M. J., I. H., L. P. C., A. M., Y. G., and C. M. investigation; L. S., A. V., and C. M. writing-original draft; L. S., A. V., T. L. L., S. P. V., L. K., Y. G., and C. M. writing-review and editing; E. F., S. P. V., and C. M. methodology; S. P. V., L. K., and Y. G. funding acquisition; Y. G. project adminis-

tration; C. M. supervision.

**Acknowledgments**—We are indebted to the Plateforme d’Analyse des Glycoconjugués and to the Research Federation FRABio (Univ.Lille, CNRS, FR 3688, FRABio, Biochimie Structurale et Fonctionnelle des assemblages Biomoléculaires) for providing the scientific and technical environment conducive to achieving this work. We thank Anne Harduin-Lepers, Olivier Vidal, Isabelle Huvent, Emiko Matsunaga, Shin-Yi Yu, Bernardette Coddeville, and Emmanuel Maes for scientific and technical advice.

**References**

1. Daffé, M. (2015) The cell envelope of tubercle bacilli. \textit{Tuberculosis} \textbf{95}, S155–S158 \textit{CrossRef Medline}

2. Puffal, J., Garcia–Heredia, A., Rahlves, K. C., Siegist, M. S., and Morita, Y. S. (2018) Spatial control of cell envelope biosynthesis in mycobacteria. \textit{Pathog. Dis.} \textbf{76}, hty027 \textit{CrossRef Medline}

3. Alderwick, L. J., Harrison, J., Lloyd, G. S., and Birch, H. L. (2015) The mycobacterial cell wall—peptidoglycan and arabinogalactan. \textit{Cold Spring Harb. Perspect. Med.} \textbf{5}, a021113–16 \textit{CrossRef Medline}

4. Kaur, D., Guerin, M. E., Skovierova, H., Brennan, P. J., and Jackson, M. (2009) Chapter 2: Biogenesis of the cell wall and other glycoconjugates of \textit{Mycobacterium tuberculosis}. \textit{Adv. Appl. Microbiol.} \textbf{69}, 23–78 \textit{CrossRef Medline}

5. Bhamidi, S., Scherman, M. S., Jones, V., Crick, D. C., Belisle, J. T., Brennan, P. J., and McNeil, M. R. (2011) Detailed structural and quantitative analysis reveals the spatial organization of the cell walls of \textit{in vivo} grown \textit{Mycobacterium leprae} and \textit{in vitro} grown \textit{Mycobacterium tuberculosis}. \textit{J. Biol. Chem.} \textbf{286}, 23168–23177 \textit{CrossRef Medline}

6. Lee, R. E. B., Li, W., Chatterjee, D., and Lee, R. E. (2005) Rapid structural characterization of the arabinogalactan and lipoarabinomannan in live mycobacterial cells using 2D and 3D HR-MAS NMR: Structural changes
in the arabian due to ethambutol treatment and gene mutation are observed. *Glycobiology* **15**, 139–151 CrossRef Medline
7. Kieser, K. J., and Rubin, E. J. (2014) How sisters grow apart: Mycobacterial growth and division. *Nat. Rev. Micro.* **12**, 550–562 CrossRef Medline
8. Xin, Y., Huang, Y., and McNeil, M. (1999) The presence of an endogenous endo-β-arabinofuranosidase in *Mycobacterium smegmatis* and characterization of its oligoarabinoside product. *Biochim. Biophys. Acta* **1473**, 267–271 CrossRef Medline
9. Rombouts, Y., Brust, B., Ohja, A. K., Maes, E., Coddeville, B., Elass, E., Kremer, L., and Guerardel, Y. (2012) Exposure of mycobacteria to cell wall-inhibitory drugs decreases production of arabinogalactanolipid related to mycolyl-arabinogalactan-peptidoglycan metabolism. *J. Biol. Chem.* **287**, 11060–11069 CrossRef Medline
10. Saha, B. C. (2000) α-L-arabinofuranosidases: Biochemistry, molecular biology and application in biotechnology. *Biotechnol. Adv.* **18**, 403–423 CrossRef Medline
11. Moynihan, P. J., Cadby, I. T., Veerapen, N., Jankute, M., Crosatti, M., Mulkamolova, G. V., Lovering, A. L., and Besra, G. S. (2019) The hydrolysis LpqL primes mycobacterial peptidoglycan recycling. *Nat. Commun.* **10**, 2647 CrossRef Medline
12. Margoles-Clark, E., Tenkanen, M., Nakari-Setälä, T., and Penttilä, M. (1996) Cloning of genes encoding α-L-arabinofuranosidase and beta-xyllosidase from *Trichoderma reesi* celsi by expression in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **62**, 3840–3846 Medline
13. Ferrer, M., Ghazi, A., Beloqui, A., Vieites, J. M., López-Cortés, N., Marín-Navarro, I., Nechitaylo, T. Y., Guazzaroni, M. E., Polaina, J., Waliczek, A., Chernikova, T. N., Reva, O. N., Golyschina, O. V., and Golyschin, P. N. (2012) Functional metagenomics unveils a multifunctional glycosyl hydrolase from the family 43 catalysing the breakdown of plant polymers in the calf rumen. *PLoS One* **7**, e38134–11 CrossRef Medline
14. Naumoff, D. G. (1999) Conserved sequence motifs in levansucrases and bifunctional beta-xyllosidases and alpha-L-arabinases. *FEBS Lett.* **448**, 177–179 CrossRef Medline
15. Almagro Armenteros, I. J., Tsirigos, K. D., Sonderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H. (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 CrossRef Medline
16. Gupta, R. S., Lo, B., and Son, J. (2018) Phylogenomics and comparative genomic studies robustly support division of the genus *Mycobacterium* into an emended genus *Mycobacterium* and four novel genera. *Front. Microbiol.* **9**, 67 CrossRef Medline
17. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDrscipt server. *Nucleic Acids Res.* **42**, W320–W324 CrossRef Medline
18. Dias, F. M. V., Vincent, P., Pell, G., Prates, I. A. M., Centeno, M. S. J., Tailford, L. E., Ferreira, L. M. A., Fontes, C. M. G. A., Davies, G. J., and Gilbert, H. J. (2004) Insights into the molecular determinants of substrate specificity in glycoside hydrolase family 5 revealed by the crystal structure and kinetics of *Celtovibrio manansian* 5A. *J. Biol. Chem.* **279**, 25517–25526 CrossRef Medline
19. Zhou, P., Liu, Y., Yan, Q., Chen, Z., Qin, Z., and Jiang, Z. (2014) Structural insights into the substrate specificity and transglycosylation activity of a fungal glycoside hydrolase family 5 β-mannanase. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 2970–2982 CrossRef Medline
20. Gonçalves, A. M. D., Silva, C. S., Madeira, T. I., Coelho, R., de Santis, D., San Romão, M. V., and Bento, I. (2012) Endo-β-D-1,4-mannanase from Chrysoporla siitopha displays a novel loop arrangement for substrate selectivity. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 1468–1478 CrossRef Medline
21. Hilge, M., Gloor, S. M., Rypniewski, W., Sauer, O., Heightman, T. D., Zimmermann, W., Winterhalter, K., and Piontek, K. (1998) High-resolution native and complex structures of thermostable β-mannanase from *Thermomonospora fusca*–substrate specificity in glycosyl hydrolase family 5. *Structure* **6**, 1433–1444 CrossRef Medline
22. Burgess, R. R., and Deutscher, M. P. (eds) (2009) Methods in Enzymology: Guide to Protein Purification, Vol. 463. Academic Press, Burlington, MA
son, M. (2014) Long-term survival and virulence of *Mycobacterium leprae* in amoebal cysts. *PLoS Negl. Trop. Dis.* 8, e3405 CrossRef Medline

42. Santucci, P., Diomandé, S., Poncin, I., Alibaud, L., Viljoen, A., Kremer, L., de Chastellier, C., and Canaan, S. (2018) Delineating the physiological roles of the PE and catalytic domains of LipY in lipid consumption in Mycobacterium-infected foamy macrophages. *Infect. Immun.* 86, e00394–18 CrossRef Medline

43. Santucci, P., Smichi, N., Diomandé, S., Poncin, I., Point, V., Gaussier, H., Cavalier, J. F., Kremer, L., and Canaan, S. (2019) Dissecting the membrane lipid binding properties and lipase activity of *Mycobacterium tuberculosis* LipY domains. *FEBS J.* 286, 3164–3181 CrossRef Medline

44. Mishra, K. C., de Chastellier, C., Narayana, Y., Bifani, P., Brown, A. K., Besra, G. S., Katoch, V. M., Joshi, B., Balaji, K. N., and Kremer, L. (2008) Functional role of the PE domain and immunogenicity of the Mycobacterium tuberculosis triacylglycerol hydrolase LipY. *Infect. Immun.* 76, 127–140 CrossRef Medline

45. Sambandamurthy, V. K., Wang, X., Chen, B., Russell, R. G., Derrick, S., Collins, F. M., Morris, S. L., and Jacobs, W. R. (2002) A pantothenate auxotroph of *Mycobacterium tuberculosis* is highly attenuated and protects mice against tuberculosis. *Nat. Med.* 8, 1171–1174 CrossRef Medline

46. Stover, C. K., de la Cruz, V. F., Fuerst, T. R., Burlein, J. E., Benson, L. A., Bennett, L. T., Bansal, G. P., Young, J. F., Lee, M. H., Hatfull, G. F., Snapper, S. B., Barletta, R. G., Jacobs, W. R., and Bloom, B. R. (1991) New use of BCG for recombinant vaccines. *Nature* 351, 456–460 CrossRef Medline

47. Okonechnikov, K., Golosova, O., Fursov, M., and UGENE team (2012) Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* 28, 1166–1167 CrossRef Medline

48. Edgar, R. C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797 CrossRef Medline

49. Saitou, N., and Nei, M. (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425 CrossRef Medline

50. Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biol.* 52, 696–704 CrossRef Medline

51. Ciucanu, I., and Kerek, F. (1984) A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* 131, 209–217 CrossRef

52. Besra, G. S. (1998) Preparation of cell-wall fractions from mycobacteria. *Methods Mol. Biol.* 101, 91–107 CrossRef Medline