



# The structural characterization of a glucosylglycerate hydrolase provides insights into the molecular mechanism of mycobacterial recovery from nitrogen starvation

Tatiana Barros Cereija,<sup>a,b</sup> Susana Alarico,<sup>c,d</sup> Eva C. Lourenço,<sup>e</sup> José António Manso,<sup>a,b</sup> M. Rita Ventura,<sup>e</sup> Nuno Empadinhas,<sup>c,d</sup> Sandra Macedo-Ribeiro<sup>a,b</sup> and Pedro José Barbosa Pereira<sup>a,b\*</sup>

Received 13 February 2019

Accepted 18 April 2019

Edited by J. L. Smith, University of Michigan, USA

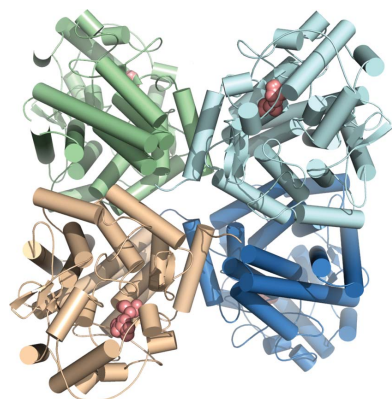
**Keywords:** *MhGgH*; GH63; glycoside hydrolase; *Mycobacterium hassiacum*; protein structure; molecular recognition; X-ray crystallography; enzyme mechanism; solution scattering.

**PDB references:** *MhGgH*, apo, 6q5t; without serine, 5ohc; SeMet, 5ohz; *MhGgH*-Ser-GOL, 5oi0; D182A-GG, 5oiw; D43A-Ser-GOL, 5oiv; E419A-GG, 5oju; D182A-MG, 5oj4; E419A-Ser-GOL, 5oie; D182A-Ser-GOL, 5oi1; E419A-MG, 5ojv; E419A-GGlycerol, 5ont; D182A-GGlycolate, 5onz; E419A-GGlycolate, 5oo2

**Supporting information:** this article has supporting information at [www.iucrj.org](http://www.iucrj.org)

<sup>a</sup>IBMC – Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal, <sup>b</sup>Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal, <sup>c</sup>CNC – Centro de Neurociências e Biologia Celular, Universidade de Coimbra, Coimbra, Portugal, <sup>d</sup>IIUC – Instituto de Investigação Interdisciplinar, Universidade de Coimbra, Coimbra, Portugal, and <sup>e</sup>ITQB – Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal. \*Correspondence e-mail: [ppereira@ibmc.up.pt](mailto:ppereira@ibmc.up.pt)

Bacteria are challenged to adapt to environmental variations in order to survive. Under nutritional stress, several bacteria are able to slow down their metabolism into a nonreplicating state and wait for favourable conditions. It is almost universal that bacteria accumulate carbon stores to survive during this non-replicating state and to fuel rapid proliferation when the growth-limiting stress disappears. Mycobacteria are exceedingly successful in their ability to become dormant under harsh circumstances and to be able to resume growth when conditions are favourable. Rapidly growing mycobacteria accumulate glucosylglycerate under nitrogen-limiting conditions and quickly mobilize it when nitrogen availability is restored. The depletion of intracellular glucosylglycerate levels in *Mycobacterium hassiacum* (basonym *Mycobacterium hassiacum*) was associated with the up-regulation of the gene coding for glucosylglycerate hydrolase (GgH), an enzyme that is able to hydrolyse glucosylglycerate to glycerate and glucose, a source of readily available energy. Highly conserved among unrelated phyla, GgH is likely to be involved in bacterial reactivation following nitrogen starvation, which in addition to other factors driving mycobacterial recovery may also provide an opportunity for therapeutic intervention, especially in the serious infections caused by some emerging opportunistic pathogens of this group, such as *Mycobacteroides abscessus* (basonym *Mycobacterium abscessus*). Using a combination of biochemical methods and hybrid structural approaches, the oligomeric organization of *M. hassiacum* GgH was determined and molecular determinants of its substrate binding and specificity were unveiled.



## 1. Introduction

In a changing environment, the basic requirements for bacterial growth are not always available. In order to accomplish one single goal, survival, bacteria have evolved different strategies (Rittershaus *et al.*, 2013). When exposed to a growth-limiting stress, such as desiccation, temperature and pH variations, oxidative stress, hypoxia, antibiotics or nutrient limitation, bacterial populations balance between cell death and decreased growth rates (Finkel, 2006; Lipworth *et al.*, 2016; Eoh *et al.*, 2017). Some of the surviving cells can slow down or suspend their growth to a viable nonreplicating state and persist for months or years (Lewis, 2007). This process,

which is known as dormancy, allows a viable population size to be maintained during the period of stress (Jones & Lennon, 2010). Despite being genetically identical to replicating bacteria, dormant cells are more tolerant to external stress (Balázsi *et al.*, 2011; Eldar & Elowitz, 2010; Dhar & McKinney, 2007). The dormancy state requires several structural modifications to maintain cell viability. Dormant cells seem to display a more compact and stable chromosome (Summers *et al.*, 2012; Nair & Finkel, 2004), a lower transcription rate and more stable messenger RNA (Rustad *et al.*, 2013), a lower but steady level of ATP (Gengenbacher *et al.*, 2010; Rao *et al.*, 2008), increased peptidoglycan mass in the cell wall accompanied by changes in the number and type of cross-links (Lavollay *et al.*, 2008; Zhou & Cegelski, 2012), and an accumulation of carbon stores (Bourassa & Camilli, 2009). Although the type of carbon store varies, the main common goal seems to be guaranteeing a rapidly mobilizable energy source that is able to promote fast cell proliferation when the environmental conditions improve (Shi *et al.*, 2010), which is an advantage at this moment in outcompeting neighbouring organisms.

In mycobacteria, the dormant state seems to be associated with the accumulation of triacylglycerol (Daniel *et al.*, 2004) and wax esters (Sirakova *et al.*, 2012). However, it has been demonstrated that rapidly growing mycobacteria accumulate glucosylglycerate (GG) under severe nitrogen-limiting conditions (Alarico *et al.*, 2014; Behrends *et al.*, 2012), which are able to induce dormancy (Anuchin *et al.*, 2009; Shleeva *et al.*, 2004). *In vitro*, GG prevented loss of activity in a number of enzymes (Sawangwan *et al.*, 2010), suggesting that it is also likely to contribute to protein stability *in vivo* during a slowly growing or nonreplicating phase. In *Mycobacterium smegmatis* and *M. hassiacum*, intracellular GG accumulated during nitrogen starvation is quickly depleted when nitrogen availability is restored (Alarico *et al.*, 2014; Behrends *et al.*, 2012). In *M. hassiacum*, GG depletion was associated with the up-regulation of a gene (*ggH*) coding for a glucosylglycerate hydrolase (GgH; Alarico *et al.*, 2014). Data on the role of GgH in bacterial survival, dormancy and infection are still lacking. However, the high degree of conservation of GgH among rapidly growing mycobacteria, which based on comparative genomic analyses were recently included in four new genera and separated from the *Mycobacterium* genus, where only the major human pathogens remain (Gupta *et al.*, 2018), and in other unrelated phyla suggests that the ability to produce this hydrolase is an evolutionary advantage. Moreover, since *ggH* expression is up-regulated upon relief of the growth-limiting stress, GgH is likely to participate in the reactivation of growth by hydrolysing GG to glycerate and glucose (Alarico *et al.*, 2014), which can be quickly used for energy production and for the synthesis of structural molecules (Mendes *et al.*, 2012) that are necessary for rapid bacterial proliferation. While most mycobacteria are environmental saprophytes, some species have increasingly been detected in oligotrophic drinking-water distribution systems, to which mycobacteria adapt and where they survive for long periods, namely in showerhead biofilms, from which they may access and infect humans

(Gebert *et al.*, 2018). Understanding the mechanisms that allow the reactivation of growth following long periods under nutrient limitation may therefore be important in trying to halt the reactivation process and prevent infection.

*M. hassiacum* GgH (*MhGgH*) belongs to the large CAZy family GH63 of glycoside hydrolases (<http://www.cazy.org>). Only four of the more than 2000 assigned members of family GH63, displaying  $\alpha$ -glucosidase (EC 3.2.1.20),  $\alpha$ -1,3-glucosidase (EC 3.2.1.84), processing  $\alpha$ -glucosidase (EC 3.2.1.106) and mannosylglycerate hydrolase (EC 3.2.1.170) activities, have been structurally characterized (Kurakata *et al.*, 2008; Barker & Rose, 2013; Miyazaki *et al.*, 2015). In order to unveil its molecular mechanism of action, a thorough structural and functional characterization of *MhGgH* was performed, elucidating its quaternary architecture and providing an atomic detail view of the determinants of substrate specificity, thus providing new insights into a potentially crucial enzyme underlying the metabolic reactivation of rapidly growing mycobacteria following severe nutrient starvation and expanding the options for therapeutic intervention.

## 2. Experimental procedures

### 2.1. Site-directed mutagenesis

The point mutations in *M. hassiacum ggH* were obtained by site-directed mutagenesis using pETM11-*MhGgH* (Cereija *et al.*, 2017) as a template and the primers 5'-CACATGTGGA GTTGGGCCGCCGCTTC-3' and 5'-GAACGCGGCGGC CCAACTCCACATGTG-3' (producing pETM11-D43A), 5'-GAGTCCGGGATGGCCAACTCG-3' and 5'-CGAGTTGG CCATCCCGGACTC-3' (producing pETM11-D182A), and 5'-TCGTTCCGCCGCTACTACGAA-3' and 5'-TTCGTAGT ACGCGGCGAACGA-3' (producing pETM11-E419A).

### 2.2. Expression and purification of *MhGgH* variants

*Escherichia coli* BL21 (DE3) cells transformed with the pET30a-*MhGgH* (Alarico *et al.*, 2014), pETM11-*MhGgH* (Cereija *et al.*, 2017), pETM11-D43A, pETM11-D182A or pETM11-E419A plasmids were used for protein expression as described previously (Cereija *et al.*, 2017; Alarico *et al.*, 2014). All proteins were purified using a combination of immobilized metal-affinity and size-exclusion chromatography (Cereija *et al.*, 2017). The affinity tag of those proteins obtained from pETM11-based constructs was removed by cleavage with TEV protease. The concentration of the purified proteins (in 20 mM Tris-HCl pH 8.0, 400 mM NaCl; storage buffer) was estimated by measuring their absorbance at 280 nm prior to flash-freezing in liquid nitrogen and storage at  $-80^{\circ}\text{C}$  until needed.

### 2.3. Analytical size-exclusion chromatography

Analytical size-exclusion chromatography was performed on a Superdex 200 Increase 5/150 GL column (GE Healthcare) pre-equilibrated with storage buffer. Blue dextran (2000 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa) and ribonuclease A (13.7 kDa) were used as standards for column calibration. The

$K_{av}$  versus log molecular weight was calculated using the equation  $K_{av} = (V_e - V_0)/(V_t - V_0)$ , where  $V_e$  is the elution volume of the protein,  $V_0$  is the void volume of the column and  $V_t$  is the column bed volume.

#### 2.4. Dynamic light-scattering analysis

Dynamic light-scattering (DLS) analysis was performed on a Zetasizer Nano ZS DLS system (Malvern Instruments). Protein samples were centrifuged at 13 000g and 4°C for 20 min, loaded onto a ZEN2112 cuvette and three independent measurements were recorded at 20°C. All data were analysed using the *Zetasizer* software v.7.11 (Malvern Instruments).

#### 2.5. Differential scanning fluorimetry

The melting temperatures of the *MhGgH* variants were determined using a thermal shift (Thermofluor) assay. Each protein sample (0.5 mg ml<sup>-1</sup> final concentration) was centrifuged at 13 000g and 4°C for 15 min, mixed with 5× SYPRO Orange (Life Technologies) in storage buffer and loaded into white 96-well PCR plates (Bio-Rad) sealed with Optical Quality Sealing Tape (Bio-Rad). The plate was heated from 25 to 95°C in 0.5°C steps with 30 s hold time per step on an iCycler iQ5 Multicolor Real-Time PCR Detection System (Bio-Rad) and the fluorescence was followed using a Cy3 dye filter (545 nm excitation/585 nm emission). Each experiment was performed in triplicate. The melting curves were analysed using the *CFX Manager* software (Bio-Rad) and the melting temperature was determined as the inflection point of the melting curve.

#### 2.6. Chemical syntheses of glucosylglycerate, mannosylglycerate and glucosylglycolate

The chemical syntheses of glucosylglycerate [GG; 2-*O*-( $\alpha$ -D-glucopyranosyl)-D-glycerate], mannosylglycerate [MG; 2-*O*-( $\alpha$ -D-mannopyranosyl)-D-glycerate] and glucosylglycolate [GGlycolate; 2-(1-*O*- $\alpha$ -D-glucopyranosyl)acetic acid] were performed as described previously (Faria *et al.*, 2008; Lourenço *et al.*, 2009; Lourenço & Ventura, 2011).

#### 2.7. Substrate specificity of *MhGgH*

The catalytic activity of *MhGgH* was evaluated in a 50  $\mu$ l reaction mixture consisting of 2.75  $\mu$ M enzyme, 20 mM GG, 25 mM sodium phosphate pH 6.0, 5 mM MgCl<sub>2</sub>, 100 mM KCl (standard reaction). The hydrolysis of GG was detected by thin-layer chromatography (TLC; Silica Gel 60, Merck) using a solvent system composed of chloroform/methanol/acetic acid/water [30:50:8:4(v:v:v:v)] and the products were stained as described previously (Alarico *et al.*, 2014). Enzyme specificity was also probed with 5 and 20 mM MG, GGlycolate, glucosylglycerol [GGlycerol; 2-*O*-( $\alpha$ -D-glucopyranosyl)-D-glycerol; Bitop] or  $\alpha$ -1,4-mannobiose (Carbosynth) as potential substrates. All reactions were performed at 42 and 50°C for 1 h, 3 h and overnight. Control reactions without enzyme were also performed. The reaction products were analysed by TLC as described above for GG, except for those from reactions

containing GGlycerol, for which the solvent system used was chloroform/methanol/25% ammonia [30:50:25(v:v:v)].

#### 2.8. Biochemical analysis and kinetic parameters of *MhGgH*

The effect of temperature and pH on the catalytic activity of *MhGgH* was evaluated by quantifying the glucose released upon hydrolysis of GG using the Glucose Oxidase Assay Kit (Sigma–Aldrich) as described previously (Alarico *et al.*, 2014). The temperature profile of *MhGgH* was traced from 20 to 60°C using the standard reaction conditions (see Section 2.7). The effect of pH on the activity of *MhGgH* was determined at 55°C using 20 mM sodium acetate (pH 4.0–5.5) or 20 mM sodium phosphate (pH 5.8–7.0) buffer.

Kinetic parameters for the *MhGgH*-catalysed hydrolysis of GG and MG were determined by quantifying the release of glucose (as described above) or mannose (using the K-MANGL 01/05 assay kit; Megazyme), respectively. A constant enzyme concentration (2.75  $\mu$ M) was incubated with increasing concentrations of GG (0–35 mM) or MG (0–150 mM) in 20 mM sodium phosphate pH 6.0, 100 mM KCl, 5 mM MgCl<sub>2</sub> at 50°C. The maximum velocity ( $V_{max}$ ) and half constant ( $K_{0.5}$ ) were calculated with *Prism* 5.0 (GraphPad Software) using the allosteric sigmoidal equation. Experiments using GG were performed in triplicate, while those using MG as substrate were performed in duplicate.

#### 2.9. Catalytic activity of *MhGgH* variants

The catalytic activity of *MhGgH* variants was evaluated in 50  $\mu$ l reactions consisting of 2.75  $\mu$ M enzyme and 10 or 20 mM GG or MG in 25 mM buffer (sodium phosphate pH 6.0, sodium acetate pH 4.5 or 5.0, bis-Tris propane pH 7.0 or Tris–HCl pH 8.0) with 5 or 10 mM MgCl<sub>2</sub> in the presence or absence of 100 mM KCl. Reactions were performed at 37, 42 and 50°C for 1 h and overnight. Reaction mixtures with and without wild-type *MhGgH* were used as controls. The hydrolysis of GG was evaluated by TLC as described in Section 2.7.

#### 2.10. Crystallization of *MhGgH* variants

Crystals of *MhGgH* were obtained as described previously (Cereija *et al.*, 2017). *MhGgH* crystals growing from less than 30% GOL\_P4K (glycerol, PEG 4000) were transferred into a solution containing at least 30% precipitant prior to flash-cooling in liquid nitrogen. The D43A, D182A and E419A *MhGgH* variants were crystallized in the same conditions, although wild-type *MhGgH* macro-seeds were employed to promote crystal growth. Complexes of the *MhGgH* D182A variant with GG, MG and GGlycolate were obtained by soaking the crystals in mother liquor supplemented with 100 mM ligand for 2 h (GG and MG) or 25 min (GGlycolate) before flash-cooling in liquid nitrogen. Complexes of the *MhGgH* E419A variant with GG, MG, GGlycolate and GGlycerol were obtained by soaking the crystals in mother liquor supplemented with 100 mM ligand for 5 min (GG), 50 min (MG), 40 min (GGlycolate) or 10 min (GGlycerol). An additional crystallization condition was identified in-house at 293 K, yielding hexagonal crystals within a day from drops

**Table 1**  
Data-collection and refinement statistics.

Values in parentheses are for the outermost shell.

Crystal	SeMet <i>MhGgH</i>	<i>MhGgH</i> – Ser–GOL	<i>MhGgH</i> without serine	Apo <i>MhGgH</i>	D43A– Ser–GOL	D182A– Ser–GOL	E419A– Ser–GOL
<b>Data collection</b>							
Synchrotron-radiation facility	ESRF	ESRF	ESRF	ESRF	ESRF	ESRF	ESRF
Beamline	ID29	ID30B	ID30A-3	ID29	ID30A-1	ID30A-3	ID30A-1
Detector	PILATUS3 6M, Dectris	PILATUS 6M, Dectris	EIGER X 4M, Dectris	PILATUS3 6M, Dectris	PILATUS3 2M, Dectris	EIGER X 4M, Dectris	PILATUS3 2M, Dectris
Wavelength (Å)	0.97909	0.97265	0.96770	0.96863	0.96598	0.96770	0.96600
Reflections (measured/unique)	428683/151619	595791/135623	390265/84736	474316/66272	460425/114314	581967/119738	338529/73877
Space group	$P2_1$	$P2_12_12$	$P2_12_12$	$P6_22$	$P2_12_12$	$P2_12_12$	$P2_12_12$
$a, b, c$ (Å)	90.8, 86.1, 159.7	86.0, 158.8, 87.8	85.9, 159.3, 91.2	167.0, 167.0, 243.3	85.9, 159.1, 88.2	86.3, 158.1, 87.7	86.2, 159.4, 88.4
$\alpha, \beta, \gamma$ (°)	90.0, 93.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 120.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	57.4–2.04 (2.07–2.04)	48.6–1.68 (1.74–1.68)	40.3–2.00 (2.04–2.00)	49.9–2.54 (2.63–2.54)	48.7–1.78 (1.85–1.78)	42.2–1.75 (1.78–1.75)	48.8–2.07 (2.14–2.07)
$R_{\text{merge}}$ ( $I/\sigma(I)$ )	0.105 (0.637) 7.6 (1.6)	0.047 (0.860) 16.4 (1.7)	0.084 (1.276) 12.1 (1.2)	0.109 (1.165) 11.9 (1.8)	0.057 (0.824) 13.0 (1.7)	0.070 (0.987) 11.6 (1.6)	0.094 (1.032) 10.0 (1.3)
Completeness (%)	96.5 (82.1)	99.0 (98.4)	99.6 (99.8)	99.8 (99.7)	98.9 (93.5)	98.8 (99.3)	98.9 (92.6)
Multiplicity	2.8 (2.8)	4.4 (4.5)	4.6 (4.8)	7.2 (7.3)	4.0 (4.0)	4.9 (5.0)	4.6 (4.0)
<b>Refinement</b>							
Resolution (Å)	57.4–2.04	48.6–1.68	40.3–2.00	49.9–2.54	45.4–1.78	40.0–1.75	48.8–2.07
$R_{\text{work}}/R_{\text{free}}$ (%)	19.9/24.5	14.6/17.2	16.7/21.0	16.6/20.4	14.9/17.9	14.6/17.2	16.7/20.8
No. of reflections							
Working set	255146	135565	84685	66183	114235	119673	73714
Test set	12657	6808	4330	3359	5735	5953	3718
Total No. of atoms	15955	8595	8115	7635	8365	8421	7767
Ligands at active site	SER, GOL	SER, GOL	GOL		SER, GOL	SER, GOL	SER, GOL
No. of water molecules	1289	838	705	379	761	776	420
Wilson $B$ factor (Å <sup>2</sup> )	28.4	25.1	33.4	53.0	29.3	26.0	39.9
R.m.s. deviations							
Bond lengths (Å)	0.010	0.010	0.010	0.009	0.010	0.010	0.010
Bond angles (°)	1.099	1.031	0.993	1.058	1.009	1.019	1.004
Ramachandran plot							
Favoured (%)	95.6	96.8	96.7	96.0	97.0	97.4	96.3
Allowed (%)	4.3	3.2	3.3	3.6	3.0	2.6	3.6
Outliers (%)	0.1	0.0	0.0	0.4	0.0	0.0	0.1
Molecules in asymmetric unit	4	2	2	2	2	2	2
PDB code	5ohz	5oi0	5ohc	6q5t	5oiv	5oi1	5oie
SBGrid code	465	467	464	641	471	468	470

consisting of equal volumes (1  $\mu\text{l}$ ) of protein (9.5 mg ml<sup>-1</sup> in storage buffer) and precipitant solution equilibrated against solution No. 22 (0.1 M ADA pH 6.5, 1.0 M ammonium sulfate) from the MembFac sparse-matrix crystallization screen (Hampton Research). These crystals were cryoprotected with Perfluoropolyether PFO-X175/08 (Hampton Research) prior to flash-cooling in liquid nitrogen.

### 2.11. Data collection and processing

Diffraction data were collected from cryocooled (100 K) single crystals on beamlines ID23-2 (Flot *et al.*, 2010), ID29 (de Sanctis *et al.*, 2012), ID30A-1 (Bowler *et al.*, 2015; Svensson *et al.*, 2015), ID30A-3 (Theveneau *et al.*, 2013) and ID30B (Mueller-Dieckmann *et al.*, 2015) of the European Synchrotron Radiation Facility (ESRF), Grenoble, France and PROXIMA-2A of the French National Synchrotron Source (SOLEIL), Gif-sur-Yvette, France. All data sets were automatically processed using the *GrenADES* (Grenoble Automatic Data processing System) pipeline (Monaco *et al.*, 2013), except for those collected on the ID30A-3 and PROXIMA-2A beamlines, which were processed with *XDS* (Kabsch, 2010) and reduced with utilities from the *CCP4* program suite (Winn

*et al.*, 2011). X-ray diffraction data-collection and processing statistics are summarized in Table 1. The X-ray diffraction images have been deposited in the SBGrid Data Bank (Meyer *et al.*, 2016).

### 2.12. Structure determination, model building and refinement

The structure of *MhGgH* was solved by multi-wavelength anomalous diffraction as reported previously (Cereija *et al.*, 2017). The refined coordinates were used as a search model to solve the structures of all of the other *MhGgH* variants and complexes by molecular replacement using *Phaser* (McCoy *et al.*, 2007). Alternating cycles of model building with *Coot* (Emsley *et al.*, 2010) and refinement with *PHENIX* (Adams *et al.*, 2010) were performed until model completion. All crystallographic software was supported by SBGrid (Morin *et al.*, 2013). Refined coordinates and structure factors were deposited in the Protein Data Bank (Berman *et al.*, 2000). Refinement statistics are summarized in Table 1.

Table 1 (continued)

Crystal	D182A–GG	D182A–MG	D182A–GGlycolate	E419A–GG	E419A–MG	E419A–GGlycolate	E419A–GGlycerol
<b>Data collection</b>							
Synchrotron-radiation facility	ESRF	ESRF	ESRF	ESRF	ESRF	SOLEIL	ESRF
Beamline	ID30A-1	ID30A-1	ID23-2	ID30A-1	ID30B	PROXIMA-2A	ID30A-3
Detector	PILATUS3 2M, Dectris	PILATUS3 2M, Dectris	PILATUS3 2M, Dectris	PILATUS3 2M, Dectris	PILATUS 6M, Dectris	EIGER X 9M, Dectris	EIGER X 4M, Dectris
Wavelength (Å)	0.96600	0.96600	0.87290	0.96599	0.97625	0.98011	0.96770
Reflections (measured/unique)	801298/133666	780685/117039	547439/91224	301215/64817	475955/74395	500266/75133	322372/76309
Space group	<i>P</i> 2 <sub>1</sub> 2	<i>P</i> 2 <sub>1</sub> 2	<i>P</i> 2 <sub>1</sub> 2	<i>P</i> 2 <sub>1</sub> 2	<i>P</i> 2 <sub>1</sub> 2	<i>P</i> 2 <sub>1</sub> 2	<i>P</i> 2 <sub>1</sub> 2
<i>a</i> , <i>b</i> , <i>c</i> (Å)	85.3, 159.6, 91.0	85.3, 159.6, 91.2	86.2, 158.9, 88.0	86.1, 159.1, 88.4	87.8, 158.2, 87.6	86.9, 157.7, 87.6	86.9, 158.8, 87.7
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	49.1–1.71 (1.77–1.71)	49.1–1.79 (1.85–1.79)	48.7–1.93 (2.00–1.93)	48.7–2.17 (2.25–2.17)	48.8–2.06 (2.13–2.06)	45.1–2.06 (2.10–2.06)	45.3–2.05 (2.09–2.05)
<i>R</i> <sub>merge</sub> ( <i>I</i> $\sigma$ ( <i>I</i> ))	0.070 (1.081)	0.074 (1.052)	0.126 (1.227)	0.091 (0.790)	0.086 (0.979)	0.124 (1.410)	0.091 (0.901)
Completeness (%)	99.5 (99.0)	99.4 (95.9)	99.8 (98.7)	99.8 (99.5)	98.0 (89.3)	100.0 (100.0)	99.4 (100.0)
Multiplicity	6.0 (6.1)	6.7 (6.5)	6.0 (6.2)	4.6 (4.5)	6.4 (6.0)	6.7 (6.9)	4.2 (4.4)
<b>Refinement</b>							
Resolution (Å)	45.9–1.71	49.1–1.79	48.7–1.93	45.5–2.17	48.8–2.06	41.9–2.06	45.3–2.05
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	14.6/16.8	14.4/17.3	14.9/18.5	15.6/19.8	15.5/19.4	16.2/20.2	15.4/19.5
No. of reflections							
Working set	133615	116974	91157	64760	74327	75046	76246
Test set	6717	5861	4566	3281	3733	3625	3909
Total No. of atoms	8376	8421	8283	7817	7840	7868	7944
Ligands at active site	9WN	2M8	GOL, SER, 9YW	9WN	GOL, SER, 2M8	GOL, SER, 9YW	A0K
No. of water molecules	847	846	677	419	463	426	566
Wilson <i>B</i> factor (Å <sup>2</sup> )	24.9	25.6	28.3	35.7	39.5	39.7	33.7
R.m.s. deviations							
Bond lengths (Å)	0.010	0.010	0.010	0.009	0.009	0.010	0.010
Bond angles (°)	1.025	0.986	0.984	1.005	0.985	1.030	1.011
Ramachandran plot							
Favoured (%)	97.2	96.5	97.0	96.6	97.5	96.5	96.4
Allowed (%)	2.8	3.5	3.0	3.3	2.4	3.4	3.1
Outliers (%)	0.0	0.0	0.0	0.1	0.0	0.1	0.5
Molecules in asymmetric unit	2	2	2	2	2	2	2
PDB code	5oiw	5oj4	5onz	5oju	5ojv	5oo2	5ont
SBGrid code	472	473	482	474	475	483	481

### 2.13. Analysis of crystallographic structures

The crystallographic models were superposed with *SUPERPOSE* (Krissinel & Henrick, 2004) and the secondary-structure elements were identified with *DSSP* (Kabsch & Sander, 1983; Touw *et al.*, 2015). The interface area between the monomers was determined using *PISA* (Krissinel & Henrick, 2007). The surface electrostatic potential was calculated with *APBS* (Baker *et al.*, 2001) using the AMBER force field (Cornell *et al.*, 1995). Figures depicting molecular models were created with *PyMOL* (Schrödinger).

### 2.14. Small-angle X-ray scattering measurements and analysis

Small-angle X-ray scattering (SAXS) measurements were recorded on beamline BM29 (Pernot *et al.*, 2013) of the ESRF, Grenoble, France with radiation of 0.9919 Å wavelength using a PILATUS 1M detector (Dectris). The protein sample was loaded onto a Superdex 200 3.2/300 GL column (GE Healthcare) and eluted with storage buffer. Measurements (1 Hz data-collection rate) were performed on the column eluate at 4°C over a scattering-vector ( $s = 4\pi \sin\theta/\lambda$ ) range of 0.033–4.933 nm<sup>-1</sup>. Data were processed and analysed with the *ATSAS* package (Petoukhov *et al.*, 2012; Franke *et al.*, 2017). A Guinier plot was calculated using *PRIMUS/qt* (Konarev *et al.*, 2003). The theoretical scattering curve from the crystallo-

graphic model was fitted to the experimental scattering curve with *CRY SOL* (Svergun *et al.*, 1995).

## 3. Results

### 3.1. Catalytic activity of *MhGgH*

Recombinant *M. hassiacum* GgH (*MhGgH*) containing only an additional Gly-Ala dipeptide at the N-terminus (Cereija *et al.*, 2017) was produced in *E. coli* and purified to homogeneity. *In vitro*, *MhGgH* was able to hydrolyse both GG and MG, although with higher efficiency for the former. Under the conditions tested, this tag-less *MhGgH* variant displayed maximum activity between 50 and 55°C, which was significantly higher than that reported for the C-terminally tagged variant (*MhGgH*-His<sub>6</sub>, 42°C; Alarico *et al.*, 2014) and was in agreement with the optimal temperature of growth of *M. hassiacum* (50°C; Tiago *et al.*, 2012) [Fig. 1(a)].

At its optimal temperature, tag-less *MhGgH* displayed maximum activity at pH 6.0 [Fig. 1(b)].

The kinetic parameters for *MhGgH* were determined at 50°C. The experimental data were best fitted to an allosteric sigmoidal curve for both tag-less and tagged *MhGgH*, with a Hill coefficient above 1, which suggests positive cooperativity [Fig. 1(c), Table 2]. The determined kinetic values for the

hydrolysis of GG and MG by tag-less *MhGgH* reflect an almost tenfold higher hydrolysis efficiency for GG (Table 2). Under the experimental conditions used it was not possible to obtain a complete kinetic curve for MG, resulting in a rough estimation of the value of  $K_{0.5}$  that nevertheless suggests higher affinity for GG.

### 3.2. Overall structure of *MhGgH*

Although both tagged and tag-less *MhGgH* variants were used in crystallization experiments, only the tag-less variant yielded three-dimensional crystals. Orthorhombic crystals belonging to space group  $P2_12_12$  that diffracted X-rays to beyond 1.7 Å resolution at a synchrotron source were obtained. Despite crystallizing in the same conditions, selenomethionine-substituted *MhGgH* produced monoclinic crystals (space group  $P2_1$ ) that were used for structure solution by multi-wavelength anomalous diffraction at the  $K$  absorption edge of selenium as described previously (Cereija *et al.*, 2017). The orthorhombic crystals contained two *MhGgH* molecules (here termed *A* and *B*) in the asymmetric unit, which were modelled from residues Pro2 to Gly446.

The overall globular *MhGgH* monomer [Fig. 2(a)] is composed of an  $(\alpha/\alpha)_6$ -barrel domain that encompasses helices  $\alpha 2$ ,  $\alpha 4$ ,  $\alpha 6$ ,  $\alpha 8$ ,  $\alpha 10$  and  $\alpha 12$  in the inner layer and  $\alpha 1$ ,  $\alpha 3$ ,  $\alpha 5$ ,  $\alpha 7$ ,  $\alpha 9$  and  $\alpha 11$  in the outer layer, and a more flexible cap domain that constrains access to the active site of the enzyme and can in turn be divided into two subdomains termed the *A'*-region (residues 163–252) and the *B'*-region (residues 68–118). Five mobile loops are also noteworthy: loop A (between  $\alpha 1$  and  $\beta 1$ ; residues 23–38), loop B (between  $B'\beta 1$  and  $B'\alpha 1a$ ; residues 81–91), loop C (between  $A'\alpha 2b$  and  $A'\alpha 3$ ; residues 193–205), loop D (between  $\alpha 9$  and  $\alpha 10$ ; residues 346–381) and loop E (between  $\beta 6$  and  $\alpha 12$ ; residues 430–434) [Fig. 2(a), Supplementary Fig. S1].

*M. hassiacum* GgH displays 68% secondary-structure identity to the single structurally characterized mannosylglycerate hydrolase *Thermus thermophilus* HB8 MgH (*Tt8MgH*; PDB entry 4wva; Miyazaki *et al.*, 2015), despite the much lower amino-acid sequence identity of 36% (calculated with *PDBFold*; Krissinel & Henrick, 2004). To date, only three other MgH orthologues have been characterized biochemically: the MgH enzymes from *Selaginella moellendorffii*

**Table 2**

Kinetic parameters for hydrolysis of GG and MG.

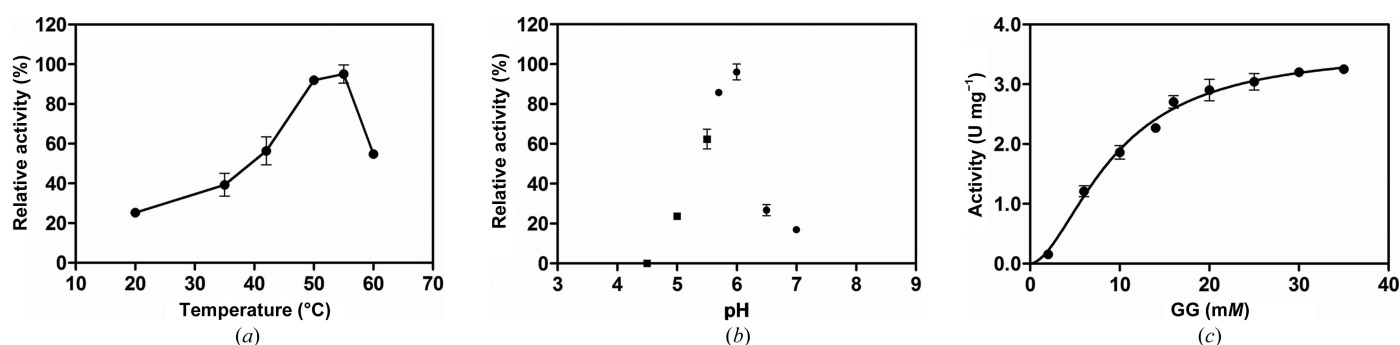
Experimental data were analysed using the allosteric kinetic model. A lower affinity for MG is expected owing to the higher estimated  $K_{0.5}$  value.

Kinetic parameters	GG	MG
$V_{\max}$ ( $\mu\text{mol min}^{-1}$ per milligram of protein)	$3.60 \pm 0.18$	$3.09 \pm 0.66$
$K_{0.5}$ (mM)	$9.36 \pm 0.69$	$84.18 \pm 30.27$
$h$	$1.77 \pm 0.20$	$1.29 \pm 0.23$
$R^2$	0.974	0.990

(Nobre *et al.*, 2013), *T. thermophilus* HB27 (99% amino-acid sequence identity to *Tt8MgH*) and *Rubrobacter radiotolerans* (Alarico *et al.*, 2013). Despite relatively low overall amino-acid sequence conservation, some regions are strictly conserved in *MhGgH* and in all characterized MgH enzymes, including the putative catalytic Asp182 and Glu419 and most of the substrate-interacting residues (Tyr36, Trp40, Trp42, Asp43, Tyr88, Gln115, Gly180, Arg216, Tyr222, Tyr375 and Trp376) (Miyazaki *et al.*, 2015; Supplementary Fig. S2).

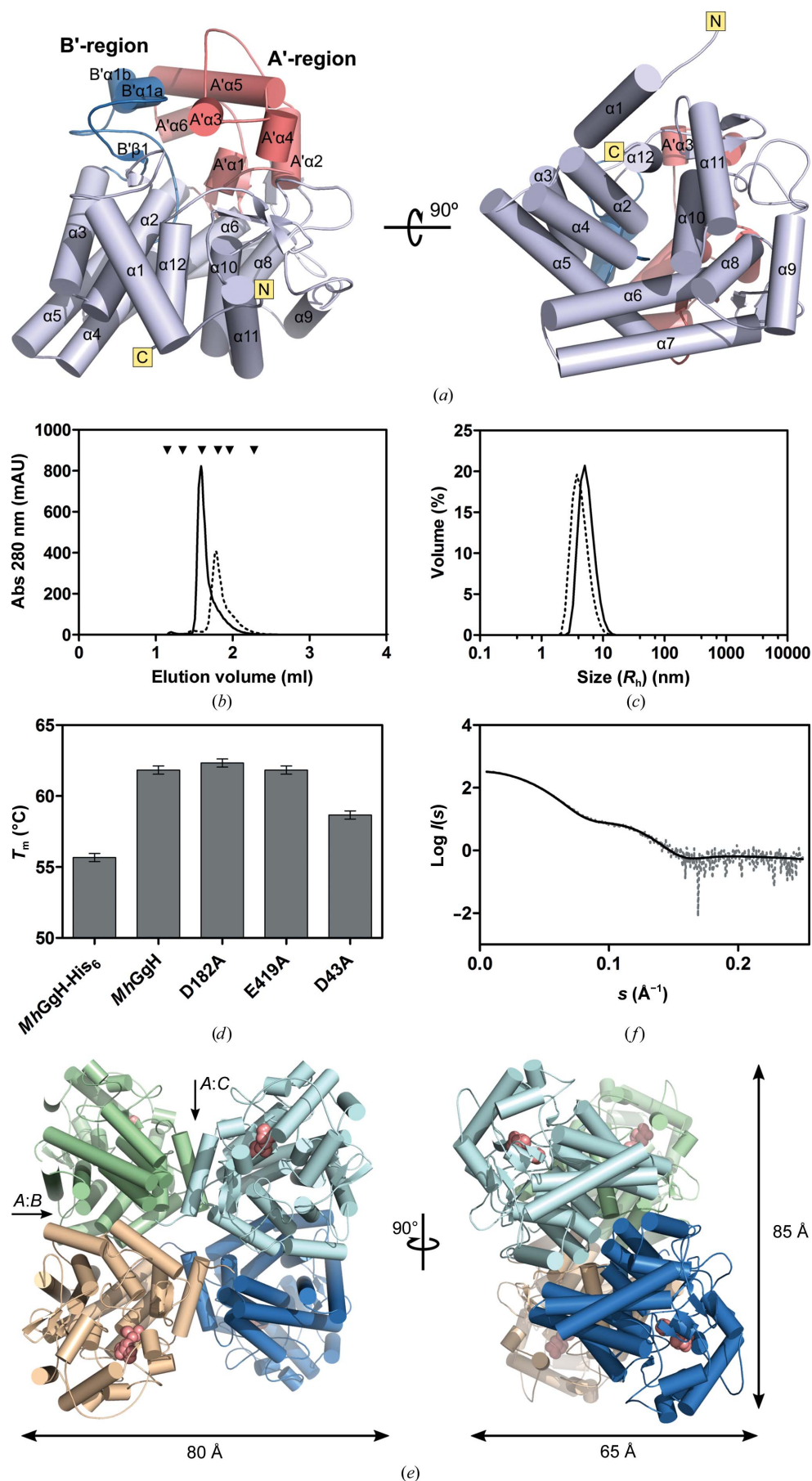
### 3.3. Quaternary structure of *MhGgH*

The apparent molecular weights of the *MhGgH* variants in solution were evaluated by size-exclusion chromatography and DLS [Figs. 2(b) and 2(c)]. Both tagged and tag-less *MhGgH* variants were analysed. While the *MhGgH*-His<sub>6</sub> variant displayed an apparent molecular weight of 92.6 kDa, corresponding to 1.8 times the expected mass of the monomer (50.9 kDa) and compatible with a dimeric arrangement, the apparent molecular weight of tag-less *MhGgH* was 178.6 kDa, which is 3.5 times greater than the molecular weight of the monomer and is compatible with a trimeric or a tetrameric organization [Fig. 2(b)]. DLS analysis of the tagged and tag-less *MhGgH* variants revealed an increase in the hydration radius (from 5.75 nm for tagged *MhGgH* to 7.34 nm for the tag-less variant), which is in agreement with a higher oligomeric arrangement for tag-less *MhGgH*, accompanied by a lower polydispersity index, which is indicative of higher homogeneity [Fig. 2(c)]. In agreement, tag-less *MhGgH* displayed a significantly higher thermal stability ( $T_m = 62^\circ\text{C}$ ) than the tagged variant ( $T_m = 56^\circ\text{C}$ ) [Fig. 2(d)], explaining its higher optimal temperature of activity, and suggesting that the introduction of a hexahistidine tag at the C-terminus of



**Figure 1**

Biochemical characterization of *MhGgH*. (a) Temperature profile of *MhGgH*, highlighting its maximal activity at 50–55°C. (b) The effect of pH on the activity of *MhGgH* assessed in 20 mM sodium acetate (squares) or 20 mM sodium phosphate (circles). (c) Kinetic curve using GG as the substrate. The sigmoidal shape of the experimental curve suggests the existence of a cooperative effect. Error bars correspond to standard deviations.



*MhGgH* affected protein stability by impairing quaternary-structure formation.

In the crystals, *MhGgH* is arranged as a dimer of dimers with approximate dimensions of  $85 \times 80 \times 65 \text{ \AA}$  [Fig. 2(e)]. The total surface area of each monomer is  $\sim 17\,300 \text{ \AA}^2$ , of which  $\sim 1600 \text{ \AA}^2$  is buried in intermonomer contacts. The largest interface area ( $\sim 900 \text{ \AA}^2$ ) occurs between molecules *A* and *C* (interface *A:C*) and molecules *B* and *D*, and involves 14 (*A:C*) or 12 (*B:D*) hydrogen bonds. With approximately half of the size ( $\sim 480 \text{ \AA}$ ), the interface between dimers *A:B* and *C:D* is stabilized

**Figure 2** Structural and biophysical characterization of *MhGgH*. (a) Cartoon representation of the overall structure of the *MhGgH* monomer. The  $(\alpha/\alpha)_6$  domain is coloured mauve, and the A'- and B'-regions are coloured salmon and blue, respectively. The N- and C-termini are indicated in yellow boxes. The views in the left and right panels are related by a  $90^\circ$  rotation around *x*. (b) Analytical size-exclusion chromatogram of tagged (dotted line) and tag-less (solid line) *MhGgH* variants. The standards used for column calibration (see Section 2) are indicated as inverted black triangles. (c) Analysis of tagged (dotted line) and tag-less (solid line) *MhGgH* variants by DLS. The tag-less variant displayed a larger hydrodynamic radius ( $R_h = 7.34 \text{ nm}$ ) and a lower polydispersity index ( $\text{PdI} = 0.092$ ) than the tagged *MhGgH* variant ( $R_h = 5.75 \text{ nm}$ ;  $\text{PdI} = 0.201$ ). (d) Melting temperatures of *MhGgH* variants determined by differential scanning fluorimetry, highlighting the lower stability of the tagged *MhGgH* variant. Error bars correspond to standard deviations. (e) Quaternary structure of *MhGgH*. Monomers are coloured green (molecule *A*), wheat (molecule *B*), cyan (molecule *C*) and blue (molecule *D*). The *A:B* and *A:C* interfaces are indicated. The glycerol and serine molecules found in the active-site region are represented by salmon spheres. The approximate dimensions of the homotetramer are indicated. The views on the left and right are related by a  $90^\circ$  rotation around *y*. (f) Superposition of the experimental SAXS data (dotted grey line) and the theoretical SAXS curve calculated from the tetrameric crystallographic model of *MhGgH* (solid black line).

by four salt bridges. The smallest interface occurs between molecules *A* and *D* (and molecules *B* and *C*), with a buried surface of  $\sim 260 \text{ \AA}^2$  and a single hydrogen bond (Supplementary Table S1). The C-terminus of each *MhGgH* monomer is in the close vicinity of the *A:C* (or *B:D*) interface and the addition of the C-terminal affinity tag is likely to disrupt dimer–dimer association and impact the quaternary organization of the enzyme, which is in line with the observed lower maximum temperature of activity and decreased stability of the *MhGgH*-His<sub>6</sub> variant.

The oligomeric arrangement of *MhGgH* in solution was also assessed by small-angle X-ray scattering (SAXS). The SAXS data are compatible with a tetrameric arrangement of the enzyme, and superposition of the experimental SAXS curve with that calculated from the crystallographic tetrameric

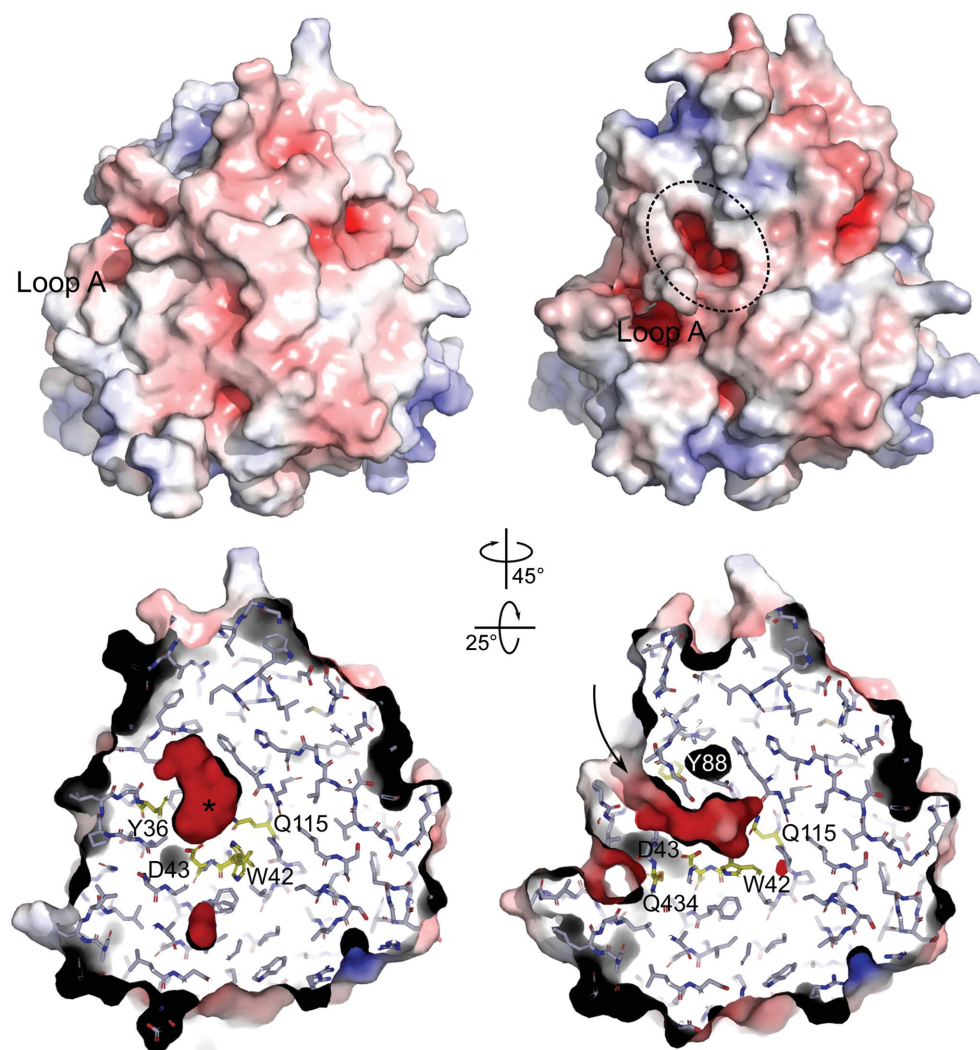
model of *MhGgH* reveals good agreement, further supporting that the crystallographic oligomer represents the quaternary architecture of the enzyme in solution [Fig. 2(*f*)].

#### 3.4. Open and closed: mobility as an essential feature for substrate binding and hydrolysis

In the orthorhombic crystals, the *MhGgH* molecules adopt a closed conformation concomitant with the presence of two ligands, a molecule of glycerol and a molecule of serine, which are components of the crystallization buffer, at the active site (*MhGgH*-Ser-GOL). The glycerol molecule occupies subsite -1 and serine is found at subsite +1 of the active site, inducing a closed state of *MhGgH* that renders them inaccessible to the solvent (Fig. 3). These ligands are stabilized mainly by polar contacts, and the putative catalytic residues, Asp182 and Glu419, are facing the lumen of the active-site cavity.

In an alternative crystallization condition (space group *P*<sub>6</sub>22), two *MhGgH* protomers are present in the asymmetric unit, corresponding to molecules *A* and *C* of the tetramer observed in the orthorhombic crystals, which were modelled from Pro5 (molecule *A*) or Ala0 (molecule *C*) to Gly446. The active sites of both molecules contain only solvent (apo *MhGgH*) and adopt an open conformation (Fig. 3). In the open conformation, the active site is accessible to the exterior through a negatively charged tunnel lined by the side chains of Trp40, Asp43, Tyr88, Gln115, Asp212, Ser214, Gln215, Met432, Gln433 and the carbonyl groups of Trp177 and Gly180 (Fig. 3). In contrast to the closed conformation, the putative catalytic residues point away from the active-site cavity: Asp182 is stabilized by polar contacts with the side chains of Tyr191 and Tyr225 and with the carbonyl group of Arg216 through a water molecule, whereas the side chain of Glu419 is hydrogen-bonded to the side chain of Ser435 and the amide N atoms of Met432 and Thr437 (Supplementary Fig. S3).

The two structures of *MhGgH*, apo *MhGgH* and *MhGgH*-Ser-GOL, corresponding to its open and closed conformations, reveal



**Figure 3**

Closed and open conformations of *MhGgH*. Solid-surface representation coloured according to electrostatic potential [contoured from  $-8 \text{ kT/e}$  (red) to  $8 \text{ kT/e}$  (blue)] (upper panel) and cross-section (lower panel) of *MhGgH* in closed (left) and open (right) conformations. In the closed state (left), the active-site cavity (marked with an asterisk) becomes inaccessible to the solvent. In the open state (right), an opening leading to an acidic cavity is observed (dashed ellipse; upper panel); a negatively charged tunnel (arrow) connects the active-site cavity to the exterior of the molecule (lower panel). Substrate-binding residues are highlighted in yellow. The left and right poses in each panel are related by  $25^\circ$  and  $45^\circ$  rotation around *x* and *y*, respectively.



the structural modifications that occur upon substrate binding. The active site of *MhGgH* is surrounded by mobile loops that disclose the active site, exposing a polar surface for substrate binding. Indeed, several residues involved in substrate binding are present in these loops, including Tyr36 (loop A), His78, Tyr88 (loop B), Tyr375, Trp376 (loop D) and Gln434 (loop E). Access to the active site is additionally restricted by a flexible cap that also contains important residues for substrate binding (Gly180, Arg216 and Tyr222), as well as the putative catalytic residue Asp182. Substrate binding contributes to the formation of an additional helix in this cap (residues 206–209) that is absent in the open conformation (Supplementary Fig. S4).

A particularly significant modification is observed in the segment Arg21–Ala31 (loop A). In the closed conformation, this loop interacts with the segment Ser431–Ser435 (with polar interactions between the side chain of Asn23 and Gln433 and Ser435) that contains the substrate-interacting residue Gln434, which faces the active site and binds to the substrate. Moreover, upon substrate binding loop A is stabilized by polar contacts with the substrate via Tyr36 and with helices  $\alpha 3$  and  $\alpha 12$ . In the open conformation, Gln433 and Gln434 move outwards (with displacements of their side chains of  $\sim 11$  and  $\sim 8$  Å, respectively), leading to a concerted rearrangement of loop E and loop A. Loop A (Arg21–Ala31) becomes mostly disordered, with difficult-to-interpret density that only allowed the modelling of two conformations for the segment Leu20–Leu25. In this short stretch, the contribution of Asp24-mediated interactions (with Arg21 or Arg58 of the neighbouring molecule) appears to be central to the overall conformation of this region. Moreover, given their spatial proximity, the conformation adopted by one subunit determines the position of the equivalent region of the neighbouring molecule, potentially impacting enzyme activity, which is in good agreement with the cooperative behaviour observed in the kinetics experiments [Fig. 1(c)].

### 3.5. Binding of substrates and substrate analogues to *MhGgH*

**3.5.1. Inactive variants of *MhGgH*.** In order to understand the molecular determinants of substrate binding and specificity, *MhGgH* was also crystallized in the presence of its substrates (GG and MG), substrate analogues (GGlycerol and GGlycolate) and reaction products (glucose, mannose and glycerate), and wild-type crystals were soaked in buffers containing these compounds. However, none of these approaches yielded crystals of the intended complexes.

To avoid substrate hydrolysis during crystallization or soaking, three catalytically inactive variants of *MhGgH* were produced. The sequence variations were identified by homology to other characterized MgH enzymes (Supplementary Fig. S2) and analysis of the *MhGgH*–Ser–GOL ternary-complex structure. Two putative catalytic residues (Asp182 and Glu419) and one substrate-interacting residue (Asp43) were identified and replaced by alanine to produce the D43A, D182A and E419A variants. The thermal stability of the D182A and E419A variants was comparable to that of wild-type *MhGgH*, while that of the D43A variant was slightly

lowered (3°C) [Fig. 2(d)]. None of the three variants displayed a detectable catalytic activity towards GG or MG (Supplementary Fig. S5).

**3.5.2. Inactive variants in complex with substrates.** Binary complexes of *MhGgH* with GG or MG were obtained by soaking orthorhombic crystals of the inactive variants in mother liquor containing these compounds. There is residual electron density compatible with presence of the substrates in the structures of the D182A and E419A variants, but not in that of the D43A variant, independent of the soaking time, highlighting the contribution of Asp43 to substrate binding. For the D182A and E419A *MhGgH* variants, the glucose/mannose moiety of each substrate occupies subsite –1 and the glycerate moiety occupies subsite +1 of the active site [Figs. 4(a) and 4(c)].

The active sites of *MhGgH* in the D182A–GG and E419A–GG complexes are virtually identical (r.m.s.d. of 0.24 Å for 14 C $^{\alpha}$  pairs). The glucose moiety adopts a  ${}^4C_1$  chair conformation with an  $\alpha$ -anomeric configuration stabilized mainly by hydrogen bonds to the side chains of residues Trp42, Asp43, Gln115, Asp182, Tyr375, Trp376 and Gln434 and to the carbonyl group of Gly180, as well as by a water-mediated contact with Glu419 [Fig. 4(a)]. The glucose moiety is also oriented by hydrophobic contacts with Tyr36, Trp40, Trp376, Trp381 and Trp436. The glycerate moiety of the substrate is stabilized by polar contacts with Tyr36 (elongated in the D182A variant), Trp40, Tyr88, Arg216 and Tyr375. Water-mediated contacts with Phe89, Gln115, Trp177, Asp182 and Tyr222 also contribute to substrate binding, as do hydrophobic contacts with Trp177. One (D182A) or two (E419A) solvent molecules occupy the space and mimic the interactions of the missing side chains.

The substrate-interacting residues in the complexes between *MhGgH* variants and GG are organized in a very similar way to that of *MhGgH*–Ser–GOL [r.m.s.d.s of 0.19 Å (D182A) and 0.22 Å (E419A) for 14 C $^{\alpha}$  pairs]. The serine and glycerol molecules present in the active site of the *MhGgH*–Ser–GOL structure partially mimic the substrate, with serine superposing nicely with the glycerate moiety and with glycerol establishing contacts with Asp43, a substrate placeholder that stabilizes the sugar moiety [Fig. 4(b)]. Taken together, these results suggest that the *MhGgH*–Ser–GOL and variant–GG complexes are likely to reflect the substrate-binding mode of GG to wild-type *MhGgH*.

The active site of both variants is also very similar when MG is bound (r.m.s.d. of 0.47 Å for 14 aligned C $^{\alpha}$  atoms) [Fig. 4(c)]. A similar hydrogen-bonding network stabilizes both substrates, with the most significant difference being the hydroxyl group at position C2 of the glucose/mannose moiety [Fig. 4(d)]. While the substrate O15 is hydrogen-bonded to Asp182 and Trp376 in the E419A–GG complex, in the E419A–MG complex only an elongated (3.5 Å) hydrogen bond to Trp376 is preserved, together with a novel water-mediated contact with Gln434.

**3.5.3. Mechanism of reaction.** At the active site of *MhGgH*, the glycosidic O atom of both substrates is within hydrogen-bonding distance of the catalytic residue Asp182, while

Glu419 establishes a water-mediated contact with both ligands. The distance between the catalytic residues and the substrate, as well as the presence of a single water molecule mediating the Glu419–substrate interaction, suggest that *MhGgH* hydrolyses GG and MG through a classic inverting mechanism (Supplementary Fig. S6), with Asp182, Glu419 and a water molecule acting as the acid, base and nucleophile, respectively. During hydrolysis, the negatively charged Glu419 is likely to activate the water molecule that performs a nucleophilic attack on the anomeric C atom, while Asp182

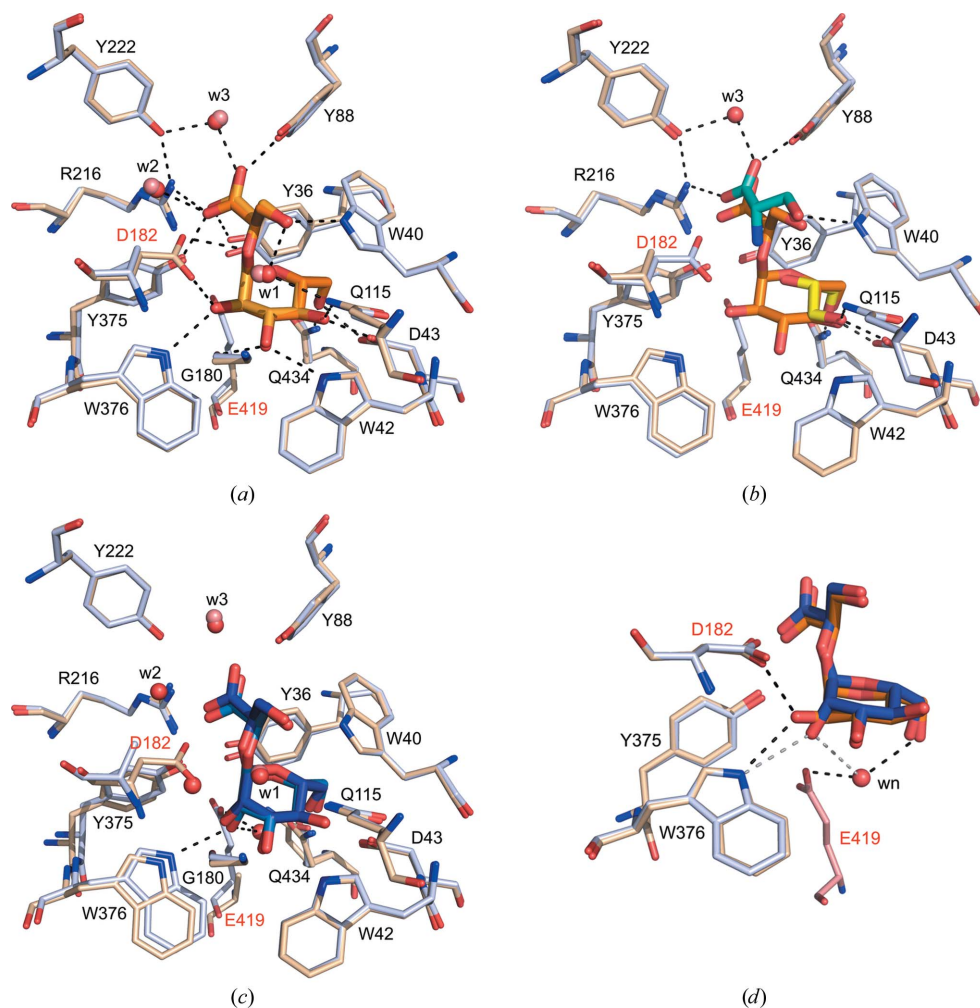
donates a proton to the leaving glycerate. Owing to the positions of Glu419 and the water molecule, inversion of the anomeric configuration of the glucose/mannose upon hydrolysis is expected, in agreement with that previously observed for the MgH enzymes from *R. radiotolerans* and *T. thermophilus* HB27 (Alarico *et al.*, 2013).

**3.5.4. Inactive variants in complex with substrate analogues.** *MhGgH* was unable to hydrolyse GGlycerol and GGlycolate, despite their considerable similarity to GG (Supplementary Fig. S7). The structures of the E419A variant

in complex with both compounds and that of the D182A variant in complex with GGlycolate were determined and help in understanding this behaviour. The most significant differences from the complexes with GG/MG are observed at subsite +1. In the E419A–GGlycerol complex the glycerol moiety of the ligand is stabilized by a direct polar contact with Trp40 and by water-mediated contacts with Tyr88, Gln115, Trp177, Asp182 and Tyr222.

In the case of GGlycolate, the glycolate moiety is found in a single position in the E419A variant but adopts two different conformations in the D182A variant. In the E419A variant, this portion of the substrate analogue is stabilized by direct contacts with Trp40 and Tyr88 and by water-mediated interactions with Gln115, Trp177, His78 and Asp182. In the D182A variant, one of the glycolate conformations is close to that found in the E419A variant (direct contacts with Trp40 and Tyr88 and water-mediated interactions with Gln115 and Trp177), while the second conformation of the ligand is stabilized by hydrogen bonds to Tyr88, Arg216 and Tyr375 and by solvent-mediated contacts with Phe89, Trp177 and Tyr222.

Both substrate analogues establish fewer interactions with the active-site region of *MhGgH* than bona fide substrates, which certainly impacts on the affinity of the enzyme for these compounds. In particular, interactions with Tyr36, a key residue



**Figure 4**

The active site of *MhGgH* variants in complex with substrates. (a) Superposition of the active-site region of *MhGgH* D182A and E419A variants in complex with GG. The position of GG (dark or light orange for the D182A or E419A variants, respectively) in the active site of the D182A (light blue) and E419A (wheat) variants is stabilized mainly by direct hydrogen bonds or by water (w)-mediated contacts (dashed lines) with the labelled residues. Water molecules in the D182A and E419A variants are represented by red and salmon spheres, respectively. The catalytic residues (Asp182 and Glu419) are highlighted in red. (b) Superposition of the active-site region of the D182A–GG complex (light blue with the ligand in orange) with that of the *MhGgH*–Ser–GOL ternary complex (wheat). Hydrogen bonds between serine (cyan) or glycerol (yellow) and the residues of the active site are represented by dashed lines. (c) Superposition of the active-site regions of the *MhGgH* D182A and E419A variants in complex with MG. The hydrogen-bonding network stabilizing MG at the active site is similar to that observed for GG [interacting residues are shown as in (a)]. The newly established contacts are represented by dashed lines. Water molecules (w) are coloured as in (a). (d) Superposition of the active-site region of the E419A variant of *MhGgH* in complex with GG and MG and of the D182A variant in complex with MG (Glu419 in salmon). The hydrogen bonds between *MhGgH* and GG (orange) or MG (blue) are represented by black or grey dashed lines, respectively. The nucleophilic water (wn) is also indicated.

for active-site closure and organization, are always absent, while different subsets of interactions are observed for GGlycerol and GGlycolate. It is therefore clear that substrate binding in *MhGgH* is a well coordinated and fine-tuned event involving the concerted movement of flexible loops and interactions with highly conserved residues. Any deviation from these strict interaction patterns will result in a decreased affinity for and highly reduced activity towards the compound, as observed for GGlycerol and GGlycolate.

#### 4. Discussion

Mycobacteria encompass a large number of species, from the well known pathogen *M. tuberculosis*, which is able to cause tuberculosis in humans and in animals and is still the leading cause of death from a single infectious agent worldwide (World Health Organization, 2018), to the ubiquitous and opportunistic *M. abscessus* and the environmental and thermophilic *M. hassiacum*, which were recently included in the newly created genera *Mycobacteroides* and *Mycolicibacterium*, respectively (Gupta *et al.*, 2018). Although most of the known mycobacteria are considered to be nonpathogenic, an increasing number of infections by opportunistic nontuberculous mycobacteria (NTM) have been reported over the last decade, which is likely to be a result of improved imaging techniques and molecular-sequencing methods that facilitate their identification (Alcaide *et al.*, 2017). On the other hand, ineffective sanitary control of water-distribution systems, as well as a number of host susceptibility factors, including ageing populations and an increased incidence of chronic diseases, may also be contributing factors to the increased rate of NTM infections detected worldwide (López-Varela *et al.*, 2015). Nontuberculous mycobacteria display high resilience against stress conditions, including an intrinsically high resistance to disinfectants and antibiotics; for this reason, NTM infections are a considerable clinical challenge for which therapeutic solutions are scarce (Falkinham, 2010). No significant advances in the treatment of NTM infections in general have recently been achieved, and the lengthy and toxic therapeutic plans in current use are often ineffective, which reinforces the need for more active drug development (Nessar *et al.*, 2012).

Although scarce, there are reports pointing to the accumulation of GG by environmental mycobacteria during nitrogen-limiting growth (Behrends *et al.*, 2012; Alarico *et al.*, 2014), a condition that is able to induce dormancy (Shleevea *et al.*, 2004; Anuchin *et al.*, 2009). As a compatible solute, GG can be accumulated intracellularly to high concentrations, and is a potential source of carbon and energy (Nunes-Costa *et al.*, 2017). Indeed, accumulated GG is quickly depleted upon exposure to an assimilable source of nitrogen, potentially fuelling bacterial growth. A glucosylglycerate hydrolase (GgH) identified in *M. hassiacum* and found to be highly conserved among rapidly growing mycobacteria is likely to be responsible for the rapid mobilization of GG accumulated during nitrogen starvation by hydrolysing it to glucose and glycerate (Alarico *et al.*, 2014). A recombinant form of this

enzyme containing a C-terminal hexahistidine tag has been characterized biochemically (Alarico *et al.*, 2014), but failed to form three-dimensional crystals suitable for structural studies. An alternative construct containing a cleavable N-terminal hexahistidine tag was recently generated (Cereija *et al.*, 2017) and removal of the affinity tag yielded an *MhGgH* variant with an additional N-terminal Gly-Ala dipeptide, which readily crystallized in two different conditions and diffracted X-rays to 1.7 Å resolution. The crystallographic structure of *MhGgH* revealed a homotetrameric architecture, which is compatible with the oligomeric organization of the enzyme in solution as assessed by SAXS. The C-terminus of *MhGgH* was found to be involved in monomer–monomer association, which was likely to be impaired by the C-terminal placement of the affinity tag in the original construct, also explaining the lower stability of this variant.

The *MhGgH* monomer displays an ( $\alpha/\alpha$ )<sub>6</sub>-barrel domain typical of glycoside hydrolase family 63 (GH63), to which *MhGgH* belongs. While the active site of GH63 members acting on larger substrates is located in an open, solvent-accessible cleft, those of *MhGgH* and of MgH from *T. thermophilus* are covered by a cap domain (subdivided into A' and B'-regions) with constrained access through a narrow negatively charged tunnel. Upon substrate binding, the cap domain closes, establishing contacts necessary to stabilize and orient the small substrate and to prevent access of bulk solvent to the active site. The open and closed states of *MhGgH* are determined by the well coordinated movement of several mobile loops that contain some of the substrate-interacting residues.

A kinetic study of *MhGgH* revealed a cooperative effect between the units of the tetramer, which may result from intersubunit interactions mediated by the mobile loop A. Indeed, in the open conformation loop A regions from adjacent monomers interact, potentially impacting on the enzymatic activity. Substrate binding by one subunit leads to the stabilization of its loop A, which is likely to facilitate access to the active site of the neighbouring subunit.

*MhGgH* was able to hydrolyse GG more efficiently than MG *in vitro*, in contrast to the similar efficiency for both substrates displayed by the MgH orthologues from *T. thermophilus* and *S. moellendorffii* (Nobre *et al.*, 2013; Alarico *et al.*, 2013). This behaviour of *MhGgH* could be explained by its distinct binding affinities for the two compounds. The  $\alpha$ -D-glucose and  $\alpha$ -D-mannose moieties of GG and MG, respectively, differ in the orientation of the C2 hydroxyl group, which is equatorial in  $\alpha$ -D-glucose and axial in  $\alpha$ -D-mannose. As a consequence, the C2 hydroxyl group of the glucose moiety of GG establishes polar contacts with Asp182 and Trp376, while that of the mannose moiety of MG establishes a single contact with Trp376. The larger number of interactions between GG and *MhGgH* are likely to translate into a higher affinity of binding and to explain the preference of the enzyme for this substrate.

The contribution of subsite +1 to substrate recognition was also evaluated using the substrate analogues GGlycerol and GGlycolate, which differ from GG in the aglycone moiety. *In*

*vitro*, MhGgH was unable to hydrolyse either compound. The structures of catalytically inactive variants of MhGgH in complex with these substrate analogues were determined, revealing the important contribution of the reducing end of the substrate to active-site closure and the ensuing enzymatic processing.

Substrate binding to MhGgH is a well coordinated event involving highly conserved residues and a complex network of polar contacts. During evolution, the active site of MhGgH has been optimized to harbour specific substrates, such as GG. The presence of MG in *M. hassiacum* cells has not so far been reported. Assuming that *M. hassiacum* is unable to produce MG, the ability of MhGgH to hydrolyse this compound may reflect a vestigial function from an ancestor enzyme. On the other hand, since cells are able to exchange molecules with the environment using different strategies, from passive diffusion to active transport, it is also possible that *M. hassiacum* possesses adequate machinery for scavenging MG released by other organisms from the environment through active export or cell death (Sampaio *et al.*, 2004) as a potential source of carbon and energy mobilized through hydrolysis by MhGgH.

The antibiotics currently in use mainly counter DNA replication or RNA, protein or cell-wall synthesis, which are indispensable functions for cell growth (Kohanski *et al.*, 2010). Since these processes are almost suppressed in dormant cells, they are more likely to survive treatment with antibiotics. Indeed, dormant cells have been associated with post-treatment relapse and the development of genetic resistance (Levin & Rozen, 2006; Gomez & McKinney, 2004). As a likely intervenient in cell recovery upon nitrogen stress relief, GgH may be viewed as a potential target for the development of new more efficient antimycobacterial drugs. Its comprehensive structural characterization contributes to clarification of the molecular determinants of substrate binding and specificity and provides a detailed molecular scaffold for the rational design of specific inhibitors.

## 5. Related literature

The following references are cited in the Supporting Information for this article: Bond & Schüttelkopf (2009) and Sievers *et al.* (2011).

## Acknowledgements

We acknowledge the European Synchrotron Radiation Facility and the French National Synchrotron Source for the provision of synchrotron-radiation facilities and thank their staff for help with data collection. The support of the X-ray Crystallography, Biointerfaces and Nanotechnology and Cell Culture and Genotyping scientific platforms of i3S (Porto, Portugal) are also acknowledged. We also acknowledge CERMAX of ITQB-NOVA, Oeiras, Portugal for the collection of NMR data.

## Funding information

The following funding is acknowledged: Norte Portugal Regional Operational Programme (NORTE 2020) (grant No. Norte-01-0145-FEDER-000012); COMPETE 2020 – Operational Programme for Competitiveness and Internationalization (POCI), PORTUGAL 2020 [grant No. POCI-01-0145-FEDER-007274; grant No. POCI-01-0145-FEDER-029221; grant No. UID/NEU/04539/2019; grant No. LISBOA-01-0145-FEDER-007660]; Fundação para a Ciência e a Tecnologia (scholarship No. SFRH/BD/92955/2013 to Tatiana Barros Cereija; scholarship No. SFRH/BPD/108299/2015 to Susana Alarico).

## References

- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Cryst.* **D66**, 213–221.
- Alarico, S., Costa, M., Sousa, M. S., Maranhã, A., Lourenço, E. C., Faria, T. Q., Ventura, M. R. & Empadinhas, N. (2014). *Sci. Rep.* **4**, 6766.
- Alarico, S., Empadinhas, N. & da Costa, M. S. (2013). *Enzyme Microb. Technol.* **52**, 77–83.
- Alcaide, F., Peña, M. J., Pérez-Risco, D., Camprubi, D., Gonzalez-Luquero, L., Grijoa-Camino, M. D., Dorca, J. & Santin, M. (2017). *Eur. J. Clin. Microbiol. Infect. Dis.* **36**, 1425–1432.
- Anuchin, A. M., Mulyukin, A. L., Suzina, N. E., Duda, V. I., El-Registan, G. I. & Kaprelyants, A. S. (2009). *Microbiology*, **155**, 1071–1079.
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
- Balázs, G., van Oudenaarden, A. & Collins, J. J. (2011). *Cell*, **144**, 910–925.
- Barker, M. K. & Rose, D. R. (2013). *J. Biol. Chem.* **288**, 13563–13574.
- Behrends, V., Williams, K. J., Jenkins, V. A., Robertson, B. D. & Bundy, J. G. (2012). *J. Proteome Res.* **11**, 3888–3896.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
- Bond, C. S. & Schüttelkopf, A. W. (2009). *Acta Cryst.* **D65**, 510–512.
- Bourassa, L. & Camilli, A. (2009). *Mol. Microbiol.* **72**, 124–138.
- Bowler, M. W., Nurizzo, D., Barrett, R., Beteva, A., Bodin, M., Caserotto, H., Delagenière, S., Dobias, F., Flot, D., Giraud, T., Guichard, N., Guijarro, M., Lentini, M., Leonard, G. A., McSweeney, S., Oskarsson, M., Schmidt, W., Snigirev, A., von Stetten, D., Surr, J., Svensson, O., Theveneau, P. & Mueller-Dieckmann, C. (2015). *J. Synchrotron Rad.* **22**, 1540–1547.
- Cereija, T. B., Alarico, S., Empadinhas, N. & Pereira, P. J. B. (2017). *Acta Cryst.* **F73**, 536–540.
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). *J. Am. Chem. Soc.* **117**, 5179–5197.
- Daniel, J., Deb, C., Dubey, V. S., Sirakova, T. D., Abomoelak, B., Morbidoni, H. R. & Kolattukudy, P. E. (2004). *J. Bacteriol.* **186**, 5017–5030.
- Dhar, N. & McKinney, J. D. (2007). *Curr. Opin. Microbiol.* **10**, 30–38.
- Eldar, A. & Elowitz, M. B. (2010). *Nature (London)*, **467**, 167–173.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Eoh, H., Wang, Z., Layre, E., Rath, P., Morris, R., Moody, D. B. & Rhee, K. Y. (2017). *Nat. Microbiol.* **2**, 17084.
- Falkingham, J. O. (2010). *Future Microbiol.* **5**, 951–960.

- Faria, T. Q., Mingote, A., Siopa, F., Ventura, R., Maycock, C. & Santos, H. (2008). *Carbohydr. Res.* **343**, 3025–3033.
- Finkel, S. E. (2006). *Nat. Rev. Microbiol.* **4**, 113–120.
- Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., van Brussel, D., Morawe, C., Borel, C., Hignette, O., Chavanne, J., Nurizzo, D., McSweeney, S. & Mitchell, E. (2010). *J. Synchrotron Rad.* **17**, 107–118.
- Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., Kikhney, A. G., Hajizadeh, N. R., Franklin, J. M., Jeffries, C. M. & Svergun, D. I. (2017). *J. Appl. Cryst.* **50**, 1212–1225.
- Gebert, M. J., Delgado-Baquerizo, M., Oliverio, A. M., Webster, T. M., Nichols, L. M., Honda, J. R., Chan, E. D., Adjemian, J., Dunn, R. R. & Fierer, N. (2018). *MBio*, **9**, e01614-18.
- Gengenbacher, M., Rao, S. P., Pethe, K. & Dick, T. (2010). *Microbiology*, **156**, 81–87.
- Gomez, J. E. & McKinney, J. D. (2004). *Tuberculosis*, **84**, 29–44.
- Gupta, R. S., Lo, B. & Son, J. (2018). *Front. Microbiol.* **9**, 67.
- Jones, S. E. & Lennon, J. T. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 5881–5886.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
- Kohanski, M. A., Dwyer, D. J. & Collins, J. J. (2010). *Nat. Rev. Microbiol.* **8**, 423–435.
- Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 1277–1282.
- Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* **D60**, 2256–2268.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Kurakata, Y., Uechi, A., Yoshida, H., Kamitori, S., Sakano, Y., Nishikawa, A. & Tonozuka, T. (2008). *J. Mol. Biol.* **381**, 116–128.
- Lavollay, M., Arthur, M., Fourgeaud, M., Dubost, L., Marie, A., Vezeris, N., Blanot, D., Gutmann, L. & Mainardi, J. L. (2008). *J. Bacteriol.* **190**, 4360–4366.
- Levin, B. R. & Rozen, D. E. (2006). *Nat. Rev. Microbiol.* **4**, 556–562.
- Lewis, K. (2007). *Nat. Rev. Microbiol.* **5**, 48–56.
- Lipworth, S., Hammond, R. J., Baron, V. O., Hu, Y., Coates, A. & Gillespie, S. H. (2016). *Tuberculosis*, **99**, 131–142.
- López-Varela, E., García-Basteiro, A. L., Santiago, B., Wagner, D., van Ingen, J. & Kampmann, B. (2015). *Lancet Respir. Med.* **3**, 244–256.
- Lourenço, E. C., Maycock, C. D. & Ventura, M. R. (2009). *Carbohydr. Res.* **344**, 2073–2078.
- Lourenço, E. C. & Ventura, M. R. (2011). *Carbohydr. Res.* **346**, 163–168.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Mendes, V., Maranhã, A., Alarico, S. & Empadinhas, N. (2012). *Nat. Prod. Rep.* **29**, 834–844.
- Meyer, P. A., Socias, S., Key, J., Ransey, E., Tjon, E. C., Buschiazzo, A., Lei, M., Botka, C., Withrow, J., Neau, D., Rajashankar, K., Anderson, K. S., Baxter, R. H., Blacklow, S. C., Boggon, T. J., Bonvin, A. M. J. J., Borek, D., Brett, T. J., Cafisch, A., Chang, C.-I., Chazin, W. J., Corbett, K. D., Cosgrove, M. S., Crosson, S., Dhe-Paganon, S., Di Cera, E., Drennan, C. L., Eck, M. J., Eichman, B. F., Fan, Q. R., Ferré-D’Amaré, A. R., Christopher Fromme, J., Garcia, K. C., Gaudet, R., Gong, P., Harrison, S. C., Heldwein, E. E., Jia, Z., Keenan, R. J., Kruse, A. C., Kvangsakul, M., McLellan, J. S., Modis, Y., Nam, Y., Otwinowski, Z., Pai, E. F., Pereira, P. J. B., Petosa, C., Raman, C. S., Rapoport, T. A., Roll-Mecak, A., Rosen, M. K., Rudenko, G., Schlessinger, J., Schwartz, T. U., Shamoo, Y., Sondermann, H., Tao, Y. J., Tolia, N. H., Tsodikov, O. V., Westover, K. D., Wu, H., Foster, I., Fraser, J. S., Maia, F. R. N. C., Gonen, T., Kirchhausen, T., Diederichs, K., Crosas, M. & Sliz, P. (2016). *Nat. Commun.* **7**, 10882.
- Miyazaki, T., Ichikawa, M., Iino, H., Nishikawa, A. & Tonozuka, T. (2015). *J. Struct. Biol.* **190**, 21–30.
- Monaco, S., Gordon, E., Bowler, M. W., Delagenière, S., Guijarro, M., Spruce, D., Svensson, O., McSweeney, S. M., McCarthy, A. A., Leonard, G. & Nanao, M. H. (2013). *J. Appl. Cryst.* **46**, 804–810.
- Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P. C., Timony, M. A., Ottaviano, M. & Sliz, P. (2013). *Elife*, **2**, e01456.
- Mueller-Dieckmann, C., Bowler, M. W., Carpentier, P., Flot, D., McCarthy, A. A., Nanao, M. H., Nurizzo, D., Pernot, P., Popov, A., Round, A., Royant, A., de Sanctis, D., von Stetten, D. & Leonard, G. A. (2015). *Eur. Phys. J. Plus*, **130**, 70.
- Nair, S. & Finkel, S. E. (2004). *J. Bacteriol.* **186**, 4192–4198.
- Nessar, R., Cambau, E., Reytrat, J. M., Murray, A. & Gicquel, B. (2012). *J. Antimicrob. Chemother.* **67**, 810–818.
- Nobre, A., Empadinhas, N., Nobre, M. F., Lourenço, E. C., Maycock, C., Ventura, M. R., Mingote, A. & da Costa, M. S. (2013). *Planta*, **237**, 891–901.
- Nunes-Costa, D., Maranhã, A., Costa, M., Alarico, S. & Empadinhas, N. (2017). *Glycobiology*, **27**, 213–227.
- Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., Morawe, C., Mueller-Dieckmann, C., Ohlsson, S., Schmid, W., Surr, J., Thevenneau, P., Zerrad, L. & McSweeney, S. (2013). *J. Synchrotron Rad.* **20**, 660–664.
- Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V. & Svergun, D. I. (2012). *J. Appl. Cryst.* **45**, 342–350.
- Rao, S. P., Alonso, S., Rand, L., Dick, T. & Pethe, K. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 11945–11950.
- Rittershaus, E. S., Baek, S. H. & Sasseti, C. M. (2013). *Cell Host Microbe*, **13**, 643–651.
- Rustad, T. R., Minch, K. J., Brabant, W., Winkler, J. K., Reiss, D. J., Baliga, N. S. & Sherman, D. R. (2013). *Nucleic Acids Res.* **41**, 509–517.
- Sampaio, M. M., Chevance, F., Dippel, R., Eppler, T., Schlegel, A., Boos, W., Lu, Y.-J. & Rock, C. O. (2004). *J. Biol. Chem.* **279**, 5537–5548.
- Sanctis, D. de, Beteva, A., Caserotto, H., Dobias, F., Gabadinho, J., Giraud, T., Gobbo, A., Guijarro, M., Lentini, M., Lavault, B., Mairs, T., McSweeney, S., Petitdemange, S., Rey-Bakaikoa, V., Surr, J., Thevenneau, P., Leonard, G. A. & Mueller-Dieckmann, C. (2012). *J. Synchrotron Rad.* **19**, 455–461.
- Sawangwan, T., Goedel, C. & Nidetzky, B. (2010). *Biotechnol. J.* **5**, 187–191.
- Shi, L., Sutter, B. M., Ye, X. & Tu, B. P. (2010). *Mol. Biol. Cell*, **21**, 1982–1990.
- Shleeva, M., Mukamolova, G. V., Young, M., Williams, H. D. & Kaprelyants, A. S. (2004). *Microbiology*, **150**, 1687–1697.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. (2011). *Mol. Syst. Biol.* **7**, 539.
- Sirakova, T. D., Deb, C., Daniel, J., Singh, H. D., Maamar, H., Dubey, V. S. & Kolattukudy, P. E. (2012). *PLoS One*, **7**, e51641.
- Summers, E. L., Meindl, K., Usón, I., Mitra, A. K., Radjainia, M., Colangeli, R., Alland, D. & Arcus, V. L. (2012). *PLoS One*, **7**, e38542.
- Svensson, O., Malbet-Monaco, S., Popov, A., Nurizzo, D. & Bowler, M. W. (2015). *Acta Cryst.* **D71**, 1757–1767.
- Svergun, D., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.
- Thevenneau, P., Baker, R., Barrett, R., Beteva, A., Bowler, M. W., Carpentier, P., Caserotto, H., Sanctis, D. de, Dobias, F., Flot, D., Guijarro, M., Giraud, T., Lentini, M., Leonard, G. A., Mattenet, M., McCarthy, A. A., McSweeney, S. M., Morawe, C., Nanao, M., Nurizzo, D., Ohlsson, S., Pernot, P., Popov, A. N., Round, A., Royant, A., Schmid, W., Snigirev, A., Surr, J. & Mueller-Dieckmann, C. (2013). *J. Phys. Conf. Ser.* **425**, 012001.
- Tiago, I., Maranhã, A., Mendes, V., Alarico, S., Moynihan, P. J., Clarke, A. J., Macedo-Ribeiro, S., Pereira, P. J. & Empadinhas, N. (2012). *J. Bacteriol.* **194**, 7010–7011.

- Touw, W. G., Baakman, C., Black, J., te Beek, T. A., Krieger, E., Joosten, R. P. & Vriend, G. (2015). *Nucleic Acids Res.* **43**, D364–D368.
- World Health Organization (2018). *Global Tuberculosis Report 2018*. Geneva: World Health Organization. [https://www.who.int/tb/publications/global\\_report/en/](https://www.who.int/tb/publications/global_report/en/)
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* **D67**, 235–242.
- Zhou, X. & Cegelski, L. (2012). *Biochemistry*, **51**, 8143–8153.