TATIANA BARROS REIS CEREIJA

BIOCHEMICAL AND STRUCTURAL CHARACTERIZATION OF NOVEL MYCOBACTERIAL DRUG TARGETS

Tese de Candidatura ao grau de Doutor em Biologia Molecular e Celular;
Programa Doutoral da Universidade do Porto
(Instituto de Ciências Biomédicas Abel Salazar e Faculdade de Ciências)

Orientador – Doutor Pedro José Barbosa Pereira

Categoria – Investigador Principal

Afiliação – Instituto de Biologia Molecular e Celular da Universidade do Porto.

Coorientador – Doutor Nuno Miguel da Silva Empadinhas

Categoria – Investigador Auxiliar

Afiliação – Centro de Neurociências e Biologia Celular da Universidade de Coimbra.

Este trabalho foi financiado pela

Fundação para a Ciência e a Tecnologia

SFRH/BD/92955/2013

This work was funded in part by grant Norte-01-0145-FEDER-000012 - Structured program on bioengineered therapies for infectious diseases and tissue regeneration, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (FEDER), and by grant FCOMP-01-0124-FEDER-014321 [PTDC/BIA-PRO/110523/2009] supported by national funds through Fundação para a Ciência e a Tecnologia and by EU-FEDER funding through the Operational Competitiveness Programme – COMPETE.

Aos meus pais e avó

Agradecimentos/Acknowledgements

Em primeiro lugar, agradeço ao meu orientador Pedro Pereira por me ter recebido, uma vez mais, no seu laboratório e me ter permitido continuar com a minha aprendizagem. A ambos, Pedro Pereira e Nuno Empadinhas, agradeço a oportunidade de participar neste projecto, assim como o apoio e ajuda ao longo destes últimos quatro anos.

Agradeço também à Sandra Macedo Ribeiro pelas suas sugestões e constante disponibilidade.

I acknowledge ESRF (Grenoble, France) and SOLEIL (Gif-sur-Yvette, France) for the provision of synchrotron-radiation facilities and thank the staff for their help with data collection. Transnational Access to the High Throughput Crystallization Laboratory of the European Molecular Biology Laboratory Grenoble Outstation (HTX lab, EMBL, Grenoble, France) was supported by the European Community Seventh Framework Program (FP7/2007–2013) Grant Protein Production Platform (PCUBE Agreement No. 227764). I also thank Instruct and Institut de Génétique et Biologie Moléculaire et Cellulaire (IGBMC, Strasbourg, France) for the access to microscale thermophoresis technology and the IGBMC staff for their help with data collection and analysis.

Agradeço à FCT pelo apoio financeiro através da bolsa de doutoramento SFRH/BD/92955/2013.

Aos meus colegas, em especial ao Jorge, José, Zsuzsa, Xana, Joana, Aninhas, Emma, Susana, Mafalda, Ana Maranha, Carol, Fátima, Rita, Celso, Andreia, Johnny, Zé Pedro, Mariana, São, Sara, Diana e João, agradeço as sugestões, disponibilidade, ajuda, motivação, companheirismo e amizade.

Finalmente, agradeço à minha família e aos meus amigos pelo inabalável apoio e contínuo encorajamento.

Contents

Publication list 13

Abbreviation list 15

Abstract 19

Resumo 21

Chapter 1 23

1.1. Mycobacteria 25

1.1.1. *Mycobacterium tuberculosis* complex and tuberculosis 26

1.1.2. Nontuberculous mycobacteria 28

1.1.3. Mycobacterial infections and the need for drug development 29

1.1.4. Mycobacterial cell envelope 30

1.2. Polymethylated polysaccharides 34

1.2.1. Physiological role 36

1.2.2. MMP biosynthesis 39

1.2.3. MGLP biosynthesis 40

1.3. Histidine phosphatases 52

1.3.1. Mechanism of reaction 52

1.3.2. Vanadate as transition-state mimic 55

1.3.3. Histidine phosphate mutases 56

1.4. Glycoside hydrolases 58

1.4.1. Mechanisms of reaction 58

Chapter 2 63

2.1. Introduction 65

2.2. Materials and methods 67

2.2.1. Expression and purification of recombinant *Mtu*GpgP 67

2.2.2. Analytical size-exclusion chromatography 67

2.2.3. Differential scanning fluorimetry 68

2.2.4. Chemical synthesis of glucosylglycerate 68
2.2.5. Crystallization of *Mtu*GpgP 68
2.2.6. Data collection and processing 69
2.2.7. Structure determination, model building and refinement 69
2.2.8. Analysis of crystallographic structures 70
2.3. Results and discussion 71
2.3.1. Biophysical analysis and crystallization of *Mtu*GpgP 71
2.3.2. Overall structure of *Mtu*GpgP 73
2.3.3. Quaternary structure of *Mtu*GpgP 78
2.3.4. *Mtu*GpgP in complex with vanadate 80
2.3.5. *Mtu*GpgP in complex with vanadate and GG 86
2.4. Supplemental information 89
Chapter 3 95
3.1. Introduction 99
3.2. Materials and methods 101
3.2.1. Cloning and site-directed mutagenesis of the ga*Mha*GgH variants 101
3.2.2. Expression and purification of *Mha*GgH-His_6 and ga*Mha*GgH variants 101
3.2.3. Expression and purification of selenomethionine-containing ga*Mha*GgH 102
3.2.4. Analytical size-exclusion chromatography 103
3.2.5. Dynamic light scattering analysis 103
3.2.6. Differential scanning fluorimetry 103
3.2.7. Chemical synthesis of GG, MG and GGlycolate 104
3.2.8. Substrate specificity of ga*Mha*GgH 104
3.2.9. Catalytic activity analysis of ga*Mha*GgH mutants 105
3.2.10. Biochemical and kinetic parameters of ga*Mha*GgH 105
3.2.11. Crystallization of ga*Mha*GgH variants 106
3.2.12. Data collection and processing 107
3.2.13. Structure determination, model building and refinement 107
3.2.14. Analysis of crystallographic structures 108

3.2.15. SAXS measurement and analysis 108
3.2.16. MST analysis 108
3.3. Results and discussion 110
3.3.1. *Mha*GgH production and crystallization 110
3.3.2. Overall structure of ga*Mha*GgH 112
3.3.3. Quaternary structure of ga*Mha*GgH 120
3.3.4. Biochemical characterization of *Mha*GgH variants 125
3.3.5. Open and closed: mobility as an essential feature for substrate binding and hydrolysis 127
3.3.6. Molecular details of substrate and substrate-analogue binding 131
3.3.7. Modulation of *Mha*GgH substrate specificity: an attempt 146
3.4. Supplemental information 149
Chapter 4 163
References 173

Publication list

Cereija TB, Alarico S, Empadinhas N, Pereira PJB (2017) Production, crystallization and structure determination of a mycobacterial glucosylglycerate hydrolase*. Acta Cryst*. F73, 536-540

Abbreviation list

3-PGA	3-phosphoglyceric acid
α-(1→6)-DGG	(1→6)-diglucosylglycerate; α-D-glucopyranosyl-(1→6)-α-D-glucopyranosyl-(1→2)-D-glycerate
β-ME	β-Mercaptoethanol
$AcPIM_2$	Tri-acylated phosphatidyl-*myo*-inositol dimannoside
$AcPIM_6$	Tri-acylated phosphatidyl-*myo*-inositol hexamannoside
Ac_2PIM_2	Tetra-acylated phosphatidyl-*myo*-inositol dimannoside
Ac_2PIM_6	Tetra-acylated phosphatidyl-*myo*-inositol hexamannoside
Acyl-CoA	Acyl-coenzyme A
ADA	*N*-(2-Acetamido)iminodiacetic acid
ADP-glucose	Adenosine-5'-diphosphoglucose
ATP	Adenosine triphosphate
BCG	Bacillus Calmette-Guérin
bPGM	Bisphosphoglycerate mutase
BTP	Bis-tris propane
CAPSO	3-(Cyclohexylamino)-2-hydroxy-1-propanesulfonic acid
CAZy	Carbohydrate-active enzymes
CCP4	Collaborative Computational Project No. 4
CHES	2-(Cyclohexylamino)ethanesulfonic acid
DAT	Diacyltrehaloses
DLS	Dynamic light scattering
DNAse I	Deoxyribonuclease I
DPG	Diphosphatidylglycerol
dPGM	2,3-bisphosphoglycerate dependent phosphatase mutase
DPI	Diffraction-data Precision Indicator
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FAS-I	Fatty acid synthase I
$GalNH_2$	Galactosamine
ga*Mha*GgH	Recombinant *M. hassiacum* glucosylglycerate hydrolase containing a N-terminal Gly-Ala dipeptide
GDP-mannose	Guanosine 5′-diphospho-D-mannose
GG	Glucosylglycerate; 2-*O*-(α-D-glucopyranosyl)-D-glycerate
GgH	Glucosylglycerate hydrolase

GGlycerol	Glucosylglycerol; 2-*O*-(α-D-glucopyranosyl)-D-glycerol
GGlycolate	Glucosylglycolate; 2-*O*-(α-D-glucopyranosyl)-D-glycolate
GH	Glycoside hydrolase
Glucose-1P	glucose-1-phosphate
GOL	Glycerol (Identifier code)
GOL_P4K	2:1 of (v/v) glycerol to (w/v) PEG 4,000 solution
GPG	Glucosyl-3-phosphoglycerate
GpgP	Glucosyl-3-phosphoglycerate phosphatase
GpgS	Glucosyl-3-phosphoglycerate synthase
h	Hill slope
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HIV	Human immunodeficiency virus
iPGM	2,3-bisphosphoglycerate independent phosphatase mutase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K_D	Dissociation constant
K_{half}	Substrate concentration producing a half-maximal enzyme velocity
K_m	Michaelis constant
LAM	Lipoarabinomannan
LB	Lysogeny broth
LM	Lipomannan
LPSN	List of Prokaryotic names with Standing in Nomenclature
mAGP	mycolyl-arabinogalactan-peptidoglycan
Mak	Maltokinase
Maltose-1P	Maltose-1-phosphate
ManT	α-(1,4)-mannosyltransferase
MDR-TB	Multidrug-resistant tuberculosis
MES	2-(*N*-morpholino) ethanesulfonic acid
MeT	3-*O*-methyltransferase
MG	Mannosylglycerate; 2-*O*-(α-D-mannopyranosyl)-D-glycerate
MGG	mannosylglucosylglycerate
MgH	Mannosylglycerate hydrolase
MGLP	6-*O*-methylglucose lipopolysaccharide
*Mha*GgH	*M. hassiacum* glucosylglycerate hydrolase
*Mha*GgH-His_6	Recombinant *M. hassiacum* glucosylglycerate hydrolase containing a C-terminal hexahistidine tag
MMP	3-*O*-methylmannose polysaccharide

MOTT	*Mycobacterium* other than tuberculosis
mPGM	Monophosphoglycerate mutase
MST	Microscale thermophoresis
*Mtb*GpgP	Recombinant *M. tuberculosis* glucosyl-3-phosphoglycerate phosphatase described elsewhere (Zheng, Jiang et al. 2014)
*Mtu*GpgP	Recombinant *M. tuberculosis* glucosyl-3-phosphoglycerate phosphatase described in this thesis
MW	Molecular weight
NCS	Non-crystallographic symmetry
NHS	N-hydroxysuccinimide
NMR	Nuclear magnetic resonance
NTM	Nontuberculous mycobacteria
OctT	Octanoyltransferase
PAT	Polyacyltrehaloses
PDB	Protein data bank
PdI	Polydispersity index
PDIM	Phthiocerol dimycocerosate
PE	Phosphatidylethanolamine
PEG	Polyethylene glycol
PEG MME	Polyethylene glycol monomethyl ether
PGM	Phosphoglycerate mutase
PhoE	*Bacillus stearothermophilus* phosphatase mutase
PI	Phosphotidyl-*myo*-inositol
PMPS	Polymethylated polysaccharide
PMSF	Phenylmethylsulfonyl fluoride
PNPP	p-nitrophenyl phosphate
RGM	Rapid-growing mycobacteria
R_H	Hydrodynamic radius
RMSD	Root-mean square deviation
SAD	Single-wavelength anomalous diffraction
SAH	*S*-adenosyl-homocysteine
SAM	*S*-adenosylmethionine
SAXS	Small-angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SeMet-ga*Mha*GgH	Selenomethionine-labelled ga*Mha*GgH
SER	Serine (Identifier code)

SGL	Sulfoglycolipids
SGM	Slow-growing mycobacteria
TB	Tuberculosis
TEV	Tobacco etch virus
TLC	Thin layer chromatography
T_m	Melting temperature
TreS	Trehalose synthase
Tt27MGH	*T. thermophilus* HB27 mannosylglycerate hydrolase
Tt8MGH	*T. thermophilus* HB8 mannosylglycerate hydrolase
TVG1/2	Trivanadate glycerol ester 1/2
UDP-glucose	Uridine diphosphate glucose
US-FDA	United States Food and Drug Administration
V_{max}	Maximum enzyme velocity
VO4	Vanadate ion
WHO	World Health Organization
XDR-TB	Extensively drug-resistant tuberculosis

Abstract

The *Mycobacterium* genus contains more than 180 species, including the causative agents of tuberculosis (*Mycobacterium tuberculosis* complex), one of the top 10 causes of death worldwide. Furthermore, many of the ubiquitous environmental nontuberculous mycobacteria are opportunistic pathogens able to cause local to disseminated infections in humans, particularly in susceptible individuals. The clinical impact of mycobacterial infections relies mostly on their increased resistance to the therapies currently in use, which reinforces the need of more active drug development. Mycobacteria produce unique intracellular carbohydrates, named polymethylated polysaccharides (PMPSs), which comprehend 3-*O*-methylmannose polysaccharides (MMPs) and 6-*O*-methylglucose lipopolysaccharides (MGLPs). MGLPs seem to be the only PMPSs common to all mycobacteria, revealing their importance to mycobacterial survival and the value of their biosynthetic pathway as source of new drug targets. Although MGLPs are known for about 50 years, their biosynthetic pathway is still not fully understood. Here, two mycobacterial enzymes – glucosyl-3-phosphoglycerate phosphatase (GpgP) and glucosylglycerate hydrolase (GgH) – proposed to be involved in the MGLPs biosynthetic pathway, particularly in glucosylglycerate (GG) metabolism, are structurally characterized.

GpgP was previously proposed to participate in the second step of MGLPs production by dephosphorylating glucosyl-3-phosphoglycerate to GG. The crystallographic structures of a recombinant GpgP and of its complex with the phosphate analogue vanadate and the reaction product GG enable the identification of the amino acids involved in substrate binding and catalysis. Also, the high resolution X-ray diffraction data allowed identifying novel trivanadate glycerol ester molecules. GgH is highly conserved among rapid-growing mycobacteria and was proposed to be involved in mycobacterial recovery from nitrogen starvation by promoting the rapid mobilization of GG accumulated under these conditions. *In vitro*, GgH catalyzes the hydrolysis of GG to glucose and glycerate. Two catalytically inactive point mutants were produced and characterized in complex with substrates and substrate analogues. The structures of the complexes, as well as the apo-structure of GgH, allowed the clarification of the molecular details of substrate specificity, binding and hydrolysis.

The present work contributes for elucidating the structure and reaction mechanism of the mycobacterial GpgP and GgH, both involved in the metabolism of GG, one of the earliest intermediates of the vital MGLP biosynthetic pathway.

Resumo

O género *Mycobacterium* contém mais de 180 espécies, incluindo os agentes etiológicos da tuberculose (complexo *Mycobacterium tuberculosis*), a qual se encontra entre as dez principais causas de morte a nível mundial. Além disso, muitas das micobactérias ambientais não tuberculosas são agentes patogénicos oportunistas capazes de causar infeções locais a disseminadas em seres humanos, particularmente em indivíduos susceptíveis. O impacto clínico das infeções micobacterianas está principalmente relacionado com a crescente resistência das micobactérias às terapias actualmente em uso, o que torna urgente o desenvolvimento de fármacos mais eficazes. As micobactérias produzem hidratos de carbono intracelulares únicos, designados polissacáridos polimetilados (PMPSs), os quais incluem os polissacáridos de 3-*O*-metilmanose (MMPs) e lipopolisacáridos de 6-*O*-metilglucose (MGLPs). Aparentemente, os MGLPs são os únicos PMPSs comuns a todas as micobactérias, o que sugere o seu carácter essencial e o valor da sua via biossintética como fonte de novos alvos terapêuticos. Embora os MGLPs tenham sido descritos há mais de 50 anos, a via biossintética que leva à sua produção ainda não é totalmente conhecida. Aqui são caracterizadas estruturalmente duas enzimas micobacterianas, a fosfatase de glucosil-3-fosfoglicerato (GpgP) e a hidrolase de glucosilglicerato (GgH), propostas como intervenientes na via de biossíntese dos MGLPs, especificamente no metabolismo de glucosilglicerato (GG).

A GpgP, considerada responsável pelo segundo passo de produção dos MGLPs, catalisa *in vitro* a desfosforilação de glucosil-3-fosfoglicerato a GG. A estrutura cristalográfica de uma GpgP recombinante e do seu complexo com o análogo do fosfato, vanadato, e o seu produto de reação, GG, permitiu identificar os aminoácidos envolvidos na ligação e na catálise do substrato. A elevada resolução dos dados de difracção de raios-X permitiu também descrever novos ésteres de trivanadato-glicerol. A GgH, altamente conservada entre as micobactérias de crescimento rápido, está implicada na recuperação das micobactérias submetidas a privação de azoto através da rápida mobilização do GG acumulado nestas condições. *In vitro*, a GgH é capaz de catalisar a hidrólise de GG a glucose e glicerato. Dois mutantes de GgH cataliticamente inactivos foram produzidos e caracterizados em complexo com os substratos da GgH e análogos dos mesmos. A estrutura dos complexos, assim como a estrutura da apo GgH, permitiu clarificar os detalhes moleculares envolvidos na especificidade, ligação e hidrólise dos substratos.

O presente trabalho contribui para elucidar a estrutura e o mecanismo de reação das GpgP e GgH micobacterianas, ambas envolvidas no metabolismo do GG, um dos primeiros intermediários na via vital de biossíntese dos MGLPs.

Chapter 1

General introduction

1.1. Mycobacteria

Mycobacteria are acid-fast rod-shaped bacilli, with a high G+C-content genome (65.6%) (Cole, Brosch et al. 1998). The genus *Mycobacterium* contains more than 180 species according to the *List of prokaryotic names with standing in nomenclature* (LPSN, www.bacterio.net), and to which the addition of two new species was recently proposed (Nouioui, Sangal et al. 2017).

Traditionally, mycobacteria are divided into slow-growing mycobacteria (SGM), when they require at least 7 days to form visible colonies on agar medium, and rapid-growing mycobacteria (RGM), when less than 7 days are sufficient (Primm, Lucero et al. 2004). Although most mycobacteria are environmental bacteria with saprophytic or symbiotic lifestyle, some species are aggressive pathogens such as the well-known *Mycobacterium tuberculosis* and *Mycobacterium leprae* (Primm, Lucero et al. 2004). Due to their impact in human health and for their ability to cause tuberculosis in humans or other mammals, mycobacteria are also commonly divided into *M. tuberculosis* complex (see section 1.1.1) and nontuberculous mycobacteria (NTM) (see section 1.1.2), with *M. leprae* and *Mycobacterium lepromatosis* (Han, Seo et al. 2008, Singh, Benjak et al. 2015), the causative agents of leprosy (Smith, Aerts et al. 2017, Steinmann, Reed et al. 2017), forming a third distinct group. Although nontuberculous mycobacteria are environmental organisms, they may also cause opportunistic infections in humans (Primm, Lucero et al. 2004).

Mycobacteria in general display high drug resistance, which makes mycobacterial infections a clinical challenge (see section 1.1.3). Despite their differences, mycobacterial species share important characteristics. One of the most dominant features of mycobacteria is their cell wall, composed of unique carbohydrates and lipids (see section 1.1.4). Mycobacteria possess the thickest outer membrane known, measuring 9-10 nm based on electron microscopy examination (Brennan and Nikaido 1995), that shields the cells from the environment, including drugs and the human immune system. Due to its contribution for mycobacteria success as pathogens, the pathways involved in the mycobacterial cell envelop synthesis are of high interest, especially for the identification of targets against which new drugs can be designed.

1.1.1. *Mycobacterium tuberculosis* complex and tuberculosis

The tubercle bacillus was discovered in 1882 by Robert Koch. Since its discovery, tuberculosis (TB) has been associated to several mycobacterial species, including *M. tuberculosis*, *M. bovis*, *M. africanum*, *M. caprae*, *M. canettii*, *M. microti*, *M. orygis*, *M. pinnipedii*, *M. suricattae*, and *M. mungi* (Esteban and Muñoz-Egea 2016), altogether known as the *Mycobacterium tuberculosis* complex.

Although the number of TB deaths fell by ~26% in 16 years (from 2000 to 2016), TB remains one of the top 10 causes of death worldwide and the leading cause from a single infectious agent (WHO 2017). The World Health Organization (WHO) estimated that in 2016 there were 10.4 million new TB cases worldwide (1.5% lower than in 2014) and 1.7 million TB deaths, 0.37 million of which among HIV-infected people (WHO 2017). In 2016, TB was also the leading cause of death among HIV-positive people, with 40% deaths due to TB (WHO 2017). It is estimated that about one-fourth of the world's population has latent TB infection with approximately 5 to 15% lifetime risk of developing the disease (WHO 2017).

Nowadays, active TB requires a minimum of 6 months of multidrug therapy in two phases: the intensive phase, consisting in 2 months of isoniazid (Bernstein, Lott et al. 1952), rifampicin (Maggi, Pasqualucci et al. 1966), pyrazinamide (Yeager, Munroe et al. 1952), and ethambutol (Thomas, Baughn et al. 1961) administration, followed by the continuation stage for 4 months with isoniazid and rifampicin (see Table 1.1 for details) (Zumla, Nahid et al. 2013). This treatment was introduced in 1970s and since then no new drugs were developed or approved as first-line treatment for TB (Zumla, Nahid et al. 2013). The poor adherence to the long-lasting TB treatment is one of the causes contributing for the emergence of multidrug-resistant TB (MDR-TB) strains and even extensively drug-resistant TB (XDR-TB) strains. While MDR-TB strains are resistant to at least isoniazid and rifampicin, XDR-TB is defined as MDR-TB plus resistance to at least one fluoroquinolone (such as levofloxacin or moxifloxacin) and a second-line injectable agent (amikacin, capreomycin or kanamycin), the two most important classes of drugs for MDR-TB treatment (Zumla, Nahid et al. 2013, WHO 2017). MDR-TB is a public health crisis and a health security threat. In 2016, the WHO estimated that there were 490,000 new cases of MDR-TB worldwide and an average proportion of MDR-TB cases with XDR-TB of 6.2% (WHO 2017).

In contrast to the standardized first-line anti-TB treatment, MDR-TB requires an individualized treatment preferentially planed according to several criteria, including the

drug resistance pattern in the local geographic region, drugs previously administered to the patient, underlying medical condition of the patient and the drug-associated adverse effects (Zumla, Nahid et al. 2013). The treatment of MDR-TB is performed with at least four second-line anti-TB drugs, most of them injectable, (see Table 1.1) and requires at least a 8 month intense phase and a total therapy duration of at least 20 months.

The XDR-TB treatment is significantly longer than that of MDR-TB and requires the use of third-line anti-TB drugs (see Table 1.1) (Zumla, Nahid et al. 2013). These drugs are expensive and often have more side effects than the first- and second-line anti-TB drugs. In HIV-infected patients, XDR-TB is associated with highly mortality rates. Worldwide, only 54% of MDR-TB and 30% of XDR-TB patients are currently successfully treated (WHO 2017).

Encouragingly, the development of anti-TB drugs has evolved dramatically over the past ten years, and novel drugs are currently in Phase III trials for the treatment of MDR-TB (Cole 2016). Among these, a novel member of the diarylquinoline class, named bedaquiline (also known as TMC207 and R207910), was the first drug with a new mechanism of action to be approved in over 40 years. This drug displays bactericidal activity *in vitro* against *M. tuberculosis* and other mycobacterial species, by inhibiting the c-subunit of ATP-synthase of mycobacteria with high specificity (Andries, Verhasselt et al. 2005, de Jonge, Koymans et al. 2007, Haagsma, Podasca et al. 2011). Although the Phase III trial has not been completed, bedaquiline was approved in 2012 by the United States Food and Drug Administration (US-FDA) based on the beneficial effect observed in two Phase IIb trials. The accelerated approval of this new drug aims to ensure that some eligible patients could benefit from this new drug. However, due to limited experience on bedaquiline use, the treatment with this drug should be performed under defined conditions and patients should be closely monitored (WHO 2013, WHO 2017). Up to now, bedaquiline was only approved for adult patients with pulmonary MDR-TB unresponsive to the existing drugs and should be given in combination with the multidrug treatment regimen recommended by WHO.

Table 1.1 – Main tuberculosis drugs in clinical use and their targets. Adapted from (Zumla, Nahid et al. 2013, Wishart, Feunang et al. 2017).

Drug (year of discovery)	Effect	Target
First-line drugs		
Isoniazid (1952)	Inhibits mycolic acid biosynthesis	Enoyl-[acyl-carrier-protein] reductase
Rifampicin (1966)	Inhibits transcription	β-subunit of RNA polymerase
Pyrazinamide (1952)	Inhibits translation and trans-translation; acidifies cytoplasm	S1 component of 30S ribosomal subunit
Ethambutol (1961)	Inhibits arabinogalactan biosynthesis	Arabinosyl transferases
Second-line drugs		
p-amino salicylic acid (1948)	Inhibits folate biosynthesis	Dihydropteroate synthase
Streptomycin[a] (1944)	Inhibits protein synthesis	S12 and 16S rRNA components of 30S ribosomal subunit
Ethionamide (1961)	Inhibits mycolic acid biosynthesis	Enoyl-[acyl-carrier-protein] reductase
Ofloxacin (1980)	Inhibits DNA supercoiling	DNA gyrase and DNA topoisomerase
Capreomycin[a] (1963)	Inhibits protein synthesis	Interbridge B2a between 30S and 50S ribosomal subunits
Kanamycin[a] (1957)	Inhibits protein synthesis	30S ribosomal subunit
Amikacin[a] (1972)	Inhibits protein synthesis	30S ribosomal subunit
Cycloserine (1955)	Inhibits peptidoglycan synthesis	D-alanine racemase and ligase
Third-line drugs		
Clofazimine (1954)	Cell cycle disruption; inhibition of potassium transporters function	DNA; potassium transporters
Linezolid (1990s)	Inhibits protein synthesis	S23 rRNA component of 50S ribosomal subunit
Amoxicillin (1985) *plus* clavulanate (1974)[b]	Inhibition of cell wall synthesis	Penicillin-binding protein
Imipenem (1970s)[a] *plus* cilastatin (1995)[a,c]	Inhibition of cell wall synthesis	Penicillin-binding protein
Clarithromycin (1980s)	Inhibits protein synthesis	S23 rRNA component of 50S ribosomal subunit

[a]Belongs to the group of injectable anti-TB drugs.
[b]Clavulanate inhibits a wide variety of β-lactamases commonly found in penicillin-resistant microorganisms, improving the action of amoxicillin.
[c]Cilastatin is used to prevent the hydrolysis of imipenem by renal dehydropeptidase-I, through its specific and reversible inhibition.

1.1.2. Nontuberculous mycobacteria

Nontuberculous mycobacteria, also called atypical mycobacteria, *Mycobacterium* other than tuberculosis (MOTT) or environmental mycobacteria (Dawson 2000), include species other than the *M. tuberculosis* complex and *M. leprae* and accounts for more than 170 species from the *Mycobacterium* genus (LPSN, www.bacterio.net). Many of these ubiquitous environmental bacteria are opportunistic pathogens, capable of causing lymphadenopathy in children, pulmonary, skin and soft tissue infections or even disseminated infections, particularly in susceptible individuals (Griffith, Aksamit et al. 2007, Piersimoni and Scarparo 2008, Henkle and Winthrop 2015). Moreover, the number of reported NTM infections seems to be increasing over the last decades likely as a result

of improved imaging techniques and new molecular sequencing methods that allow better identification of different mycobacterial species (Alcaide, Pena et al. 2017). On the other hand, several environmental causes as well as host susceptibility factors may also contribute for an increase in NTM infections, including population aging, increased incidence of chronic diseases and deficient sanitary control of water-distribution systems (Lopez-Varela, Garcia-Basteiro et al. 2015).

Since NTM are exposed to several environmental challenges, they display high resilience against stress conditions including nutrient starvation, desiccation, pH and temperature variations, as well as antibiotic and disinfectant action (Falkinham 2010). These characteristics allowed them to colonize artificial environments such as domestic and hospital water distribution systems, from where they easily access susceptible hosts. Moreover, NTM are naturally resistant to common antibiotics and some infections do not even respond to aggressive anti-TB treatments (Griffith, Brown-Elliott et al. 2015, Shahraki, Heidarieh et al. 2015, Nunes-Costa, Alarico et al. 2016). NTM infections are thus becoming a considerable clinical challenge for which therapeutic solutions are scarce. No significant treatment advances for NTM infection in general have been recently accomplished and the increasing resistance of some mycobacterial species to the therapeutics currently in use reinforces the need of more active drug development (Nessar, Cambau et al. 2012).

1.1.3. Mycobacterial infections and the need for drug development

One of the major barriers in mycobacterial infection treatment is the long-duration of the therapy that is prone to poor adherence and ultimately promotes drug resistance. Shorter periods of treatment have been associated to increased relapse rates, due to the presence of a bacterial subpopulation that is able to survive the action of antibiotics (Connolly, Edelstein et al. 2007, Dhar and McKinney 2007). The surviving bacteria are named persisters (Lewis 2010). One of the features that lead to persistence is the ability of some bacteria to slow down or suspend their growth to a viable non-replicating state and persist for months or years in a dormant state (Corper and Cohn 1933, Lewis 2007). Dormancy can be induced by different stresses, such as hypoxia, nutrient starvation and acidification (Lipworth, Hammond et al. 2016). Unfortunately, most of the studies on mycobacterial dormancy were so far restricted to *M. tuberculosis* and *M. smegmatis*.

The antibiotics currently in use mainly counter DNA replication, or RNA, protein or cell wall synthesis, indispensable functions for cell growth (Kohanski, Dwyer et al. 2010). Since these processes are almost suppressed in dormant cells, they are more likely to survive even to long-term therapy. Indeed, dormant cells have been associated to post-treatment relapse and development of genetic resistance (Gomez and McKinney 2004, Levin and Rozen 2006). Targeting essential molecules for mycobacterial dormancy and resuscitation may be a viable path for an effective treatment of mycobacterial infections (Downing, Mischenko et al. 2005, Lewis 2007, Kana, Gordhan et al. 2008, Rittershaus, Baek et al. 2013, Cano-Muniz, Anthony et al. 2018).

Understanding more in depth the biology of mycobacteria is important to find new, shorter and more effective anti-mycobacterial therapies. The cell envelope of mycobacteria is highly complex and one of the main reasons for their success as pathogens. One of the most characteristic components of the cell wall are the mycolic acids. The biosynthesis of mycolic acids precursors seems to be somehow modulated by unique carbohydrates (see section 1.2), whose structural uniqueness has been seen as a source of potential antigens and targets for new vaccines, diagnostics and drug development. Although they were discovered in 1960s, their function and biosynthetic pathway are still not fully understood (Mendes, Maranha et al. 2012). The development of new molecular tools allowed identifying the genes implicated in their biosynthesis. Since some of these genes were shown to be essential for mycobacterial *in vitro* growth, they represent promising drug targets for new anti-mycobacterial drug development (Sassetti, Boyd et al. 2003, Griffin, Gawronski et al. 2011, DeJesus, Gerrick et al. 2017).

1.1.4. Mycobacterial cell envelope

The cell envelope of mycobacteria comprises four main layers, named inner membrane, periplasm, peptidoglycan-arabinogalactan layer and outer membrane (Figure 1.1) (Jankute, Cox et al. 2015). The inner membrane, also termed cytoplasmatic membrane, is composed of a phospholipid bilayer similar to that found in other bacteria. The cytoplasmatic membrane contains glycolipids, such as diphosphatidylglycerol (DPG) and phosphatidylethanolamine (PE), as well as phosphotidyl-*myo*-inositol (PI), which in contrast to the former that are found in all living organisms, is an essential phospholipid of eukaryotic cells (Antonsson 1997, Nikawa and Yamashita 1997) rarely found in prokaryotes (Kataoka and Nojima 1967, Brennan and Ballou 1968, Yano, Furukawa et al. 1969, Brennan and Lehane 1971). The glycolipid PI is a precursor of the phosphotidyl-

myo-inositol mannosides, from which the most abundant are the tri- and tetra-acylated phosphatidyl-*myo*-inositol dimannosides ($AcPIM_2$ and Ac_2PIM_2, respectively), tri- and tetra-acylated phosphatidyl-*myo*-inositol hexamannosides ($AcPIM_6$ and Ac_2PIM_6, respectively), lipomannan (LM) and lipoarabinomannan (LAM) (Jankute, Cox et al. 2015).

Outside the cytoplasmatic membrane an electron-transparent periplasmatic space is present, displaying a typical bilayer aspect (Zuber, Chami et al. 2008). The periplasm contains the lipoglycans LM and LAM attached to the plasma membrane *via* their phosphatidyl-*myo*-inositol anchor (Chatterjee, Hunter et al. 1992, Khoo, Dell et al. 1995, Jankute, Grover et al. 2014). Although not represented in Figure 1.1, LM and LAM are also abundant in the outer membrane of mycobacteria and are key immunomodulatory molecules (Ortalo-Magné, Lemassu et al. 1996, Pitarque, Larrouy-Maumus et al. 2008).

The peptidoglycan-arabinogalactan layer is composed of peptidoglycan covalently linked to the heteropolysaccharide arabinogalactan. The highly cross-linked network of peptidoglycan is one of the main structural elements of the mycobacterial cell wall, conferring a structural integrity to the bacterium and withstanding the osmotic pressure (Kaur, Guerin et al. 2009). The peptidoglycan is covalently linked to the complex polysaccharide arabinogalactan by a unique α-L-rhamnopyranose–(1→3)-α-D-*N*-acetyl-α-D-glucosamine-(1→P) linker unit (McNeil, Daffe et al. 1990). Arabinogalactan is esterified at its non-reducing end to long-chains (C_{70}–C_{90}) of mycolic acids (McNeil, Daffe et al. 1991). The mycobacterial mycolyl-arabinogalactan also contains galactosamine ($GalNH_2$) attached to the arabinan chains of the slow-growing mycobacteria (Draper, Khoo et al. 1997), or succinyl groups attached to the non-mycolylated arabinogalactans of rapid and slow-growing mycobacteria (Bhamidi, Scherman et al. 2008). Altogether, the complex of peptidoglycan, arabinogalactan and mycolic acids, known as mycolyl-arabinogalactan-peptidoglycan (mAGP) complex, is usually referred as the cell wall core of mycobacteria (Jackson 2014, Jankute, Cox et al. 2015).

Mycolic acids are long chain (C_{60}-C_{90}) α-alkyl-β-hydroxy fatty acids and components of the electron-dense outer membrane of mycobacteria. These unique fatty acids have three distinct subclasses, termed α-, methoxy- and keto- mycolic acids (Watanabe, Aoyagi et al. 2001, Watanabe, Aoyagi et al. 2002). Although unique, mycolic acids are not exclusive to mycobacteria as they were also found in other related genera such as *Nocardia*, *Rhodococcus* and *Corynebacterium* (Marrakchi, Laneelle et al. 2014). Beside mycolic acids, the outer membrane of mycobacteria also contains extractable lipids, including trehalose monomycolates, trehalose dimycolates, phthiocerol dimycocerosate (PDIM), sulfoglycolipids (SGL), diacyltrehaloses (DAT), polyacyltrehaloses (PAT), phenolic

glycolipids and lipooligosaccharides (Jackson 2014). Some of these lipids are almost limited to pathogenic mycobacteria, such as PDIM (Jackson, Stadthagen et al. 2007). Also, DAT and PAT seem to be limited to the virulent isolates of *M. tuberculosis* complex and SGL were exclusively found in *M. tuberculosis* (Jackson, Stadthagen et al. 2007, Layre, Paepe et al. 2011, Nobre, Alarico et al. 2014). The components of the mycobacterial outer membrane were proposed to be organized in two leaflets (Minnikin 1982, Rastogi, Hellio et al. 1991, Hoffmann, Leis et al. 2008, Zuber, Chami et al. 2008), with the mycolic acids packed in a monolayer and oriented perpendicular to the plasma membrane (Nikaido, Kim et al. 1993). While the mycolic acids form the inner leaflet, the other lipids form the outer leaflet with their fatty acid chains intercalating into the mycolic acid layer (Minnikin 1982, Hoffmann, Leis et al. 2008, Zuber, Chami et al. 2008).

An additional layer at the outermost compartment of the cell envelope, termed capsule, seems to be present in the pathogenic mycobacterial species (Chapman, Hanks et al. 1959, Daffe and Draper 1998). The components of this capsule-like structure are loosely attached and include polysaccharides, proteins and a small fraction of lipids (Seibert 1949, Lemassu and Daffé 1994, Ortalo-Magne, Dupont et al. 1995, Sani, Houben et al. 2010).

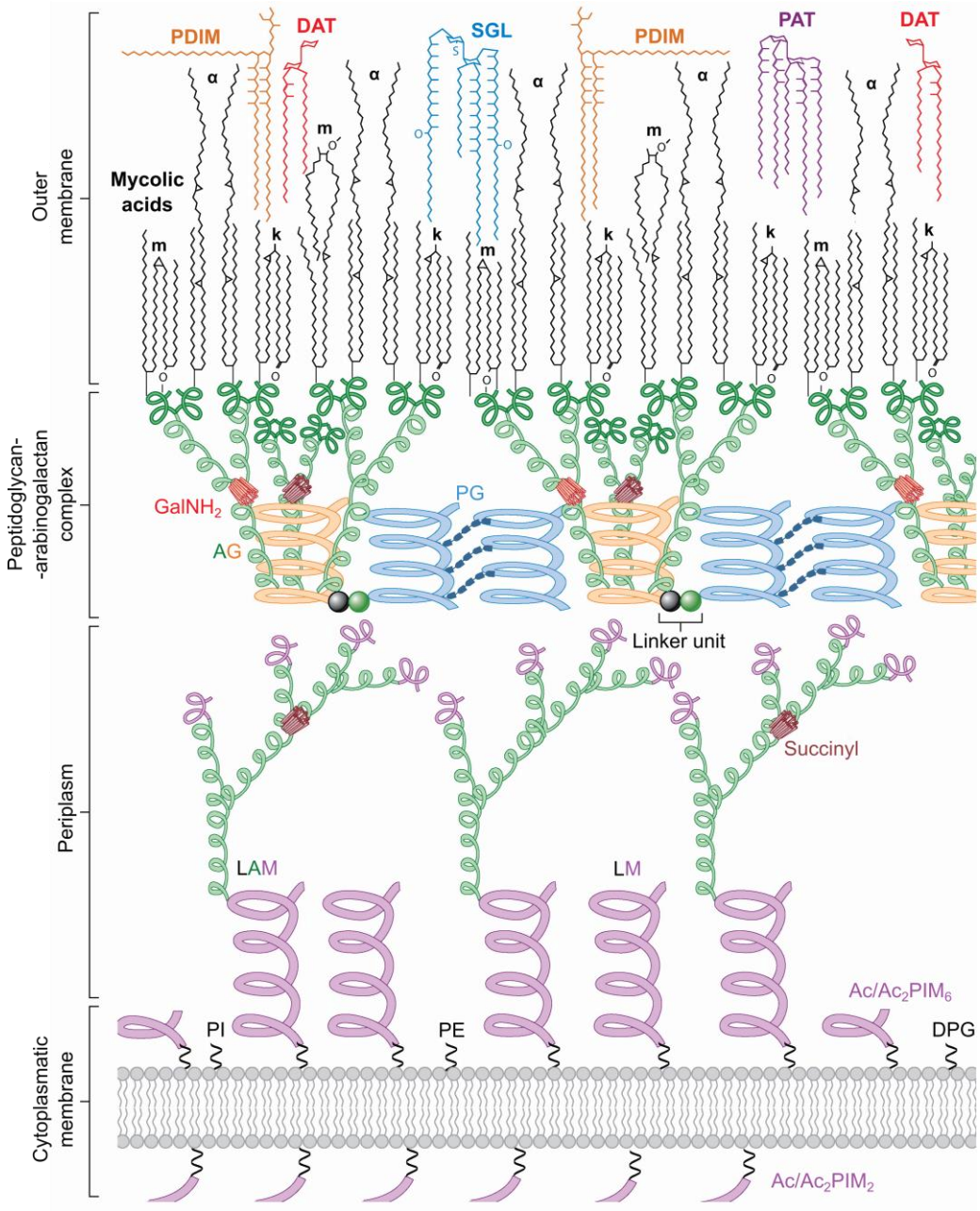


Figure 1.1 – Schematic representation of the mycobacterial cell envelope. The abbreviations are: Ac/Ac_2PIM_2, tri-/tetra-acylated phosphatidyl-*myo*-inositol-dimannosides; Ac/Ac_2PIM_6, tri-/tetra-acylated phosphatidyl-*myo*-inositol-hexamannosides; AG, arabinogalactan; DAT, diacyltrehalose; DPG, diphosphatidylglycerol; $GalNH_2$, galactosamine; k, keto; LAM, lipoarabinomannan; LM, lipomannan; m, methoxyl; PDIM, phthiocerol dimycocerosate; PE, phosphatidylethanolamine; PG, peptidoglycan; PI, phosphatidyl-*myo*-inositol; PAT, polyacyltrehalose; SGL, sulfoglycolipid. The α-L-rhamnopyranose–(1→3)-α-D-*N*-acetyl-α-D-glucosamine-(1→P) linker unit is indicated. Adapted from (Jankute, Cox et al. 2015).

1.2. Polymethylated polysaccharides

Polymethylated polysaccharides (PMPSs) are a family of unusual carbohydrates unique to the order Actinomycetales (Jackson and Brennan 2009). These cytoplasmatic (lipo)polysaccharides are composed of 10 to 20 sugar units, many of which partially *O*-methylated, which contributes to their hydrophobicity. In mycobacteria two classes of PMPSs were found, named 3-*O*-methylmannose polysaccharides (MMPs) (Gray and Ballou 1971, Maitra and Ballou 1977) and 6-*O*-methylglucose lipopolysaccharides (MGLPs) (Lee and Ballou 1964, Lee 1966) (Figure 1.2).

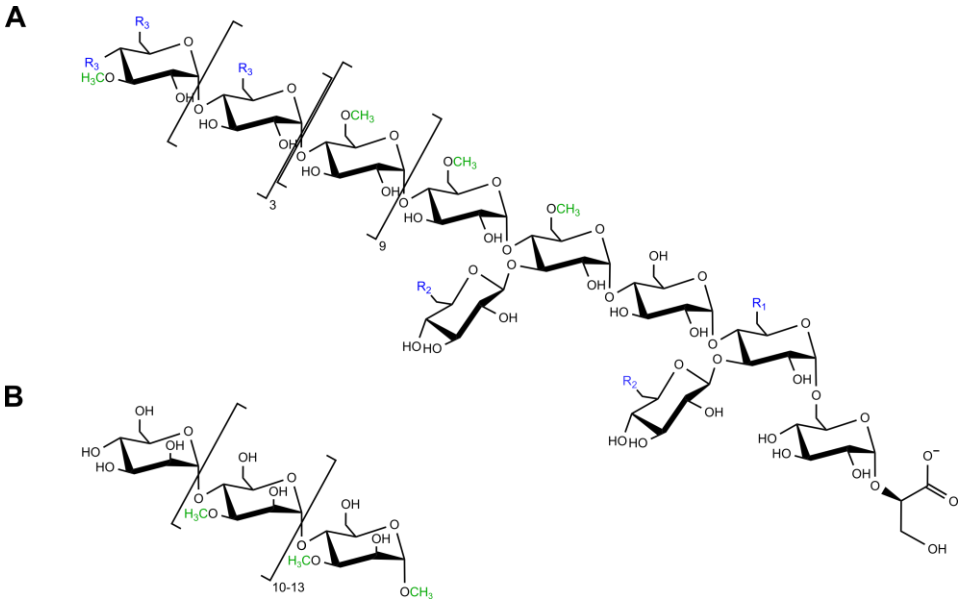


Figure 1.2 – Structures of mycobacterial PMPSs. A – MGLP from *M. bovis* BCG. B – MMP from *M. smegmatis*. R1, R2 and R3 represent octanoate, succinate and acetate groups, respectively (blue). Methyl groups are coloured green. Adapted from (Mendes, Maranha et al. 2012, Maranha, Moynihan et al. 2015).

MMPs are linear chains of 10 to 13 α-(1→4)-linked 3-*O*-methyl-D-mannoses, containing a single α-(1→4)-linked unmethylated D-mannose at the non-reducing end and an α-methyl aglycon at the reducing end (Figure 1.2B) (Maitra and Ballou 1977). Structurally, MGLPs display higher complexity compared to MMPs. MGLPs are composed of α-(1→4)-linked glucose and 6-*O*-methylglucose units, terminated at the non-reducing end by a 3-*O*-methylglucose unit and at the reducing end by an α-(1→6)-linked α-D-glucopyranosyl-(1, 2)-D-glycerate (Figure 1.2A) (Tuffal, Albigot et al. 1998). Also, two β-(1→3)-linked branching glucoses are present at the first and third glucoses of the MGLPs main chain (Tuffal, Albigot et al. 1998). MGLPs are composed of 15 to 20 sugar units that may possess variable chemical modifications with acyl groups, including acetate, propionate and isobutyrate, at the glucose units close to the non-reducing end, and succinate, esterified to the branching glucose units (Keller and Ballou 1968, Smith and Ballou 1973). An octanoate modification is also present close to the reducing end of MGLPs, likely at the atom O6 of the second glucose unit (Smith and Ballou 1973, Maranha, Moynihan et al. 2015). An unusual monosaccharide, named 2-*N*-acetyl-2,6-dideoxy-β-glucopyranose, was so far identified only in the MGLP of *M. bovis* BCG (Tuffal, Albigot et al. 1998).

While MMPs have been isolated from multiple non-pathogenic RGM species, such as *M. smegmatis* and *M. phlei*, MGLPs have been found in both RGM and SGM species (Table 1.2). Interestingly, MMPs were never detected in SGM, raising the hypothesis that this group of organisms might be unable to produce mannosylated PMPSs. Moreover, the ability of both RGM and SGM, as well as of other members of the *Actinobacteria* phylum, to produce MGLPs, indicates that their biosynthesis must be fundamental for mycobacterial survival. Although PMPSs were first isolated in 1960s (Lee and Ballou 1964, Lee 1966, Keller and Ballou 1968), the knowledge about PMPS function (see section 1.2.1) and biogenesis (see sections 1.2.2 and 1.2.3) is still limited.

Table 1.2 – Distribution of PMPSs among mycobacteria and other related genera.

Species	MGLPs	MMPs	Growth type	References
Mycobacterium				
M. smegmatis	✓	✓	Rapid-growing	(Keller and Ballou 1968, Bergeron, Machida et al. 1975, Weisman and Ballou 1984, Kamisango, Dell et al. 1987)
M. phlei	✓	✓	Rapid-growing	(Lee 1966, Saier and Ballou 1968, Gray and Ballou 1971, Weisman and Ballou 1984)
M. parafortuitum	✓	✓	Rapid-growing	(Weisman and Ballou 1984)
M. aurum (*M. cuneactum, M. petrophilum)*	✓	✓	Rapid-growing	(Weisman and Ballou 1984)
M. chitae	✓	✓	Rapid-growing	(Weisman and Ballou 1984)
M. vaccae		✓	Rapid-growing	(Tian, Li et al. 2000)
M. tuberculosis	✓		Slow-growing	(Lee 1966)
M. bovis	✓		Slow-growing	(Tuffal, Albigot et al. 1998)
M. leprae	✓		Slow-growing	(Hunter, Gaylord et al. 1986)
M. xenopi	✓		Slow-growing	(Tuffal, Ponthus et al. 1995)
Nocardia				
N. otitidis-caviarum	✓			(Pommier and Michel 1986)
N. brasiliense	✓			(Pommier and Michel 1986)
N. farcinica	✓			(Pommier and Michel 1986)
N. kirovani	✓			(Pommier and Michel 1986)
N. corallina	n.d.			(Saier and Ballou 1968)
Streptomyces				
S. griseus		✓		(Candy and Baddiley 1966, Harris and Gray 1977)
Corynebacterium				
"*C. adiphtheriae*"	n.d.			(Pommier and Michel 1986)
C. pseudodiphtheriticum	n.d.			(Pommier and Michel 1986)
C. xerosis	n.d.			(Saier and Ballou 1968)
Propionibacterium				
P. shermanii	n.d.			(Saier and Ballou 1968)

n.d. not demonstrated.

1.2.1. Physiological role

The PMPSs were proposed to play a role in fatty acid biosynthesis, with MMPs and MGLPs sharing similar properties. *In vitro*, PMPSs are able to form stable 1:1 ratio complexes with fatty acyl-coenzyme A (acyl-CoA) derivatives, preferentially with chain length ranging from C_{16} to C_{22} (Machida and Bloch 1973). Also, PMPSs seem to increase the overall rate of fatty acid synthesis by fatty acid synthase I (FAS-I) and shift the length of synthesized chains from long (C_{20-24}) to short (C_{14-18}) (Ilton, Jevans et al. 1971, Flick and Bloch 1974, Wood, Peterson et al. 1977). In agreement, PMPSs markedly lower the K_m of FAS-I for acyl-CoA substrates, which enable fatty acids biosynthesis to proceed at

low substrate concentrations, as well as may shift product distribution in favour of more rapidly diffusing short-chain fatty acid products (Ilton, Jevans et al. 1971, Vance, Mitsuhashi et al. 1973, Banis, Peterson et al. 1977). On the other hand, PMPSs seem to facilitate the release of *de novo* synthetized fatty acid chains by sequestering them from FAS-I, contributing for enzyme turnover and earlier termination of fatty acid chain elongation (Banis, Peterson et al. 1977).

PMPSs were also proposed to act as fatty acyl carriers, regulating their processing for the synthesis of longer, and more complex and insoluble fatty acids, including mycolic acids, by increasing the mycobacteria tolerance to high cytoplasmatic concentrations of these compounds (Kawaguchi and Bloch 1976, Yabusaki and Ballou 1979, Mita and Yasumasu 1981), as well as protecting them from enzymatic degradation (Yabusaki and Ballou 1979). These findings are also in agreement with the proposed PMPS structure (Figure 1.3). Both MMPs and MGLPs were proposed to adopt a coiled conformation in solution, with a hydrophobic interior determined by the inward projection of the methyl groups, in which PMPSs can accommodate their hydrophobic ligands. (Bergeron, Machida et al. 1975, Yabusaki and Ballou 1978, Yabusaki, Cohen et al. 1979, Maggio 1980, Tuffal, Ponthus et al. 1995). Thus, PMPS may physically avoid the access of enzymes to the fatty acids, while coating fatty acids with a hydrophilic surface that increases their solubility. More recently, molecular dynamic simulations revealed that the MMP helix with a hydrophobic channel lined by methyl groups immediately unfolds into disordered conformations and results in unrealistic carbohydrate ring conformations (Liu, Siuda et al. 2016). A lower energy conformation was found for helical MMP with 3-*O*-methyl groups along the outside of the helical axis. When bound to fatty acid chain, hydrophilic hydroxyl and 3-*O*-methyl groups of MMP are solvent exposed and shield the fatty acid chain from water molecules (Figure 1.3A). Molecular dynamic simulations also pointed for the lower energy required for MMP-fatty acid complex when the carboxyl group of the fatty acid is oriented towards the reducing end of MMP (Liu, Siuda et al. 2016).

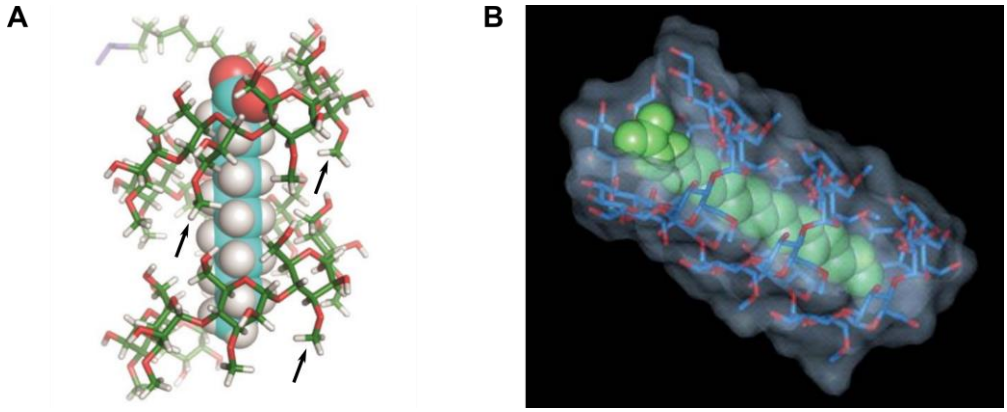


Figure 1.3 – Three-dimensional representation of PMPSs. A – Structural model of a lauric acid molecule (spheres) bound to a MMP chain (sticks), whose methyl groups are indicated by black arrows. Adapted from (Liu, Siuda et al. 2016). B – Structural model of the stearic acid (green spheres) bound to a 6-*O*-methylglucose polysaccharide chain containing 20 glucose units, of which 12 are *O*-methylated (sticks). Adapted from (Jackson and Brennan 2009).

However, physiological evidence for the role of PMPSs in lipid metabolism regulation is still limited. A *M. smegmatis* spontaneous mutant with half of the wild-type production of MGLPs and only residual production of MMPs displayed no significant change in fatty acids synthesis, and surprisingly showed higher accumulation of short-chain fatty acids instead of the expected long-chain fatty acids (Maloney and Ballou 1980). Since the production of fatty acids still occurred in near absence of MMPs, MGLPs are likely to display a more important role in fatty acids metabolism than MMPs.

PMPSs may also play a role in mycobacterial thermal adaptation. Indeed, it was observed that mutants with limited production of MGLPs were more sensitive to high temperatures, displaying a lower growth rate or inability to grow at temperatures usually well tolerated by the wild-type strain (Stadthagen, Sambou et al. 2007). The growth temperature seems also to influence the production of mycolic acids. During growth at high temperatures, increased synthesis of long-chain mycolic acids and a decrease in unsaturated fatty acids were observed for *M. phlei* (Toriyama, Yano et al. 1980). In agreement, an increase in short-chain mycolic acids synthesis, with a concomitant decrease in long-chain mycolic acids production, were observed for *M. smegmatis* growing at low temperatures (Baba, Kaneda et al. 1989). Since PMPS-deficient strains are more sensitive to temperature and the proportion of short to long fatty acid chains is temperature-dependent, PMPSs are likely involved in thermal adaptation by modulating fatty acid chain length.

Therefore, the physiological role of these unique PMPSs in fatty acids metabolism still needs to be clarified. Also, the existence of two types of PMPSs apparently with redundant functions in some mycobacteria is puzzling. For a better understanding of PMPSs physiological role it is necessary to identify the genes involved in PMPSs biosynthesis, to evaluate their relevance in bacterial survival and virulence, and to characterize functional and structurally their coding proteins.

1.2.2. MMP biosynthesis

The currently knowledge on the biosynthesis of MMPs is limited to the work performed by Ballou's research group from 1970s to 1980s and more recently by the Lowary laboratory (Xia, Zheng et al. 2012, Xia 2013). A model for MMP biosynthesis was proposed after the isolation of MMP precursors (Maitra and Ballou 1977, Yamada, Cohen et al. 1979) and the characterization of an α-(1,4)-mannosyltransferase (ManT) (Weisman and Ballou 1984) and a 3-*O*-methyltransferase (MeT) (Weisman and Ballou 1984) from *M. smegmatis* cell-free extracts. According to the proposed model, MMPs are likely to be elongated by a linear alternating process of mannosylation and *O*-methylation, using GDP-mannose as sugar donor and S-adenosylmethionine (SAM) as source of methyl groups (Figure 1.4) (Weisman and Ballou 1984).

When the MMP chain reaches 10 to 13 units of 3-*O*-methylmannose the elongation process ends (Maitra and Ballou 1977). At this length, the polysaccharide displays good fatty acid-binding properties and the MMP-fatty acid complex is no longer available as substrate for MeT, resulting in the termination of the elongation reaction (Weisman and Ballou 1984). Also, since the K_m of MeT for the polysaccharide-fatty acid complex is significantly higher than that of ManT, the MMP chain ends with an unmethylated mannose (Weisman and Ballou 1984).

The MMP biosynthetic gene cluster of *M. smegmatis* mc^2155 was recently proposed (Xia 2013). This cluster includes three genes predicted to encode for an α-(1,4)-mannosyltransferase (*MSMEG_6484*), a 3-*O*-methyltransferase (*MSMEG_6483*) and another methyltransferase (*MSMEG_6482*). The function of a recombinant *MSMEG_6484* encoding protein was evaluated *in vitro* and confirmed to display α-(1,4)-mannosyltransferase activity (Xia 2013).

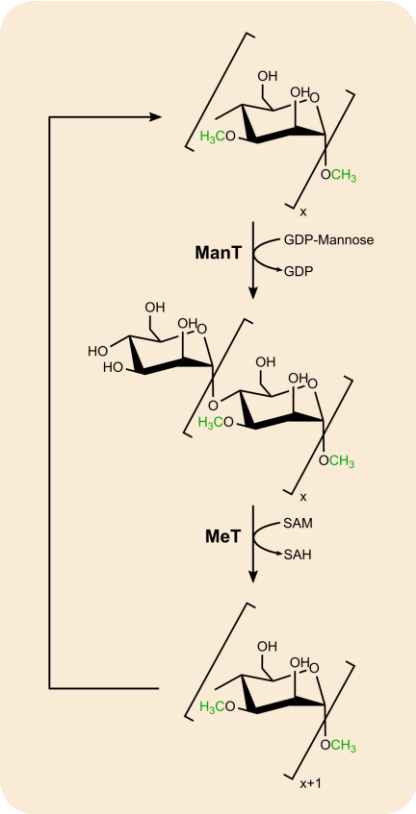


Figure 1.4 – Proposed MMP biosynthetic pathway. The proposed pathway for MMP biosynthesis involves the sequential mannosylation and methylation reactions performed by a α-(1,4)-mannosyltransferase (ManT) and a 3-*O*-methyltransferase (MeT), respectively. The ManT uses GDP-mannose as mannose donor, whereas MeT uses *S*-adenosylmethionine (SAM) as source of methyl groups, releasing S-adenosyl-homocysteine (SAH). The elongation ends when the MMP is long enough (10 to13 mannose units) to bind fatty acids and is no longer available as substrate for MeT. Adapted from (Stadthagen, Sambou et al. 2007).

1.2.3. MGLP biosynthesis

The steps that led to MGLP formation started to be defined after the partial purification and characterization of a methyltransferase able to transfer methyl groups from SAM to glucose units of the polysaccharide (Ferguson and Ballou 1970, Grellert and Ballou 1972), and of an acyltransferase that catalysed the acylation of MLGP with acetyl, propionyl, isobutyl, succinyl and octanoyl groups from acyl-CoA derivatives (Tung and Ballou 1973). The acylation and methylation steps were proposed to occur together during the synthesis of MGLPs, with one process affecting the extent and direction of the other (Grellert and Ballou 1972). Approximately 14 years later, free 2-*O*-(α-D-glucopyranosyl)-D-glycerate (glucosylglycerate, GG) and α-D-glucopyranosyl-(1→6)-α-D-glucopyranosyl-(1→2)-D-

glycerate [α-(1→6)-diglucosylglycerate, α-(1→6)-DGG] were detected in cell extracts of *M. smegmatis* and were considered the precursors of MGLPs (Kamisango, Dell et al. 1987). A new biosynthetic mechanism was proposed, in which MGLP elongation would occur stepwise through sequential glucosylation and methylation reactions from the reducing end to the non-reducing end (Kamisango, Dell et al. 1987).

Although the pathway leading to the production of MGLPs is still not fully understood, the availability of the whole-genome sequence of *M. tuberculosis* and of closely related microorganisms, as well as the development of tools for genetic manipulation of mycobacteria, have significantly contributed to the identification and functional characterization of the genes potentially involved in MGLP biosynthesis.

1.2.3.1. Initiation of the pathway

It is currently accepted that the MGLP pathway is initiated by the production of glucosyl-3-phosphoglycerate (GPG) by the glucosyl-3-phosphoglycerate synthase (GpgS), using UDP-glucose as donor substrate and 3-phosphoglyceric acid (3-PGA) as acceptor substrate (Figure 1.5). This GpgS activity was initially identified in methanogenic archaea and in hyperthermophilic bacteria (Costa, Empadinhas et al. 2006, Costa, Empadinhas et al. 2007). The non-homologous mycobacterial GpgS was identified 1 year later: recombinant enzymes from *M. smegmatis* and *M. bovis* BCG were biochemically characterized (Empadinhas, Albuquerque et al. 2008), and the crystallographic structure of a recombinant *M. tuberculosis* GpgS was determined (Figure 1.6A) (Pereira, Empadinhas et al. 2008), being the first enzyme of the MGLP pathway to have its three-dimensional structure unveiled. In *M. tuberculosis* H37Rv, this enzyme is encoded by the *Rv1208* gene that was considered essential for mycobacterial *in vitro* growth (Sassetti, Boyd et al. 2003, Griffin, Gawronski et al. 2011, DeJesus, Gerrick et al. 2017). Also, the inactivation of the *Rv1208* ortholog in *M. smegmatis* dramatically decreased the MGLP levels, suggesting the involvement of this gene in MGLP biosynthesis (Kaur, Pham et al. 2009). Other studies followed to better characterize this enzyme, including site-directed mutagenesis (Kumar, Guan et al. 2014) and co-crystallization with substrates (Albesa-Jove, Mendoza et al. 2015, Albesa-Jove, Romero-Garcia et al. 2017), which allowed a more clear understanding of the molecular details of the reaction.

The reaction product of GpgS is subsequently dephosphorylated by the glucosyl-3-phosphoglycerate phosphatase (GpgP), producing GG. In *M. tuberculosis*, the GpgP

activity was associated to the *Rv2419c* gene (Mendes, Maranha et al. 2011), which was considered non-essential for mycobacterial growth since mycobacteria lacking this gene were still able to grow *in vitro* (Sassetti, Boyd et al. 2003, Griffin, Gawronski et al. 2011, DeJesus, Gerrick et al. 2017). Nevertheless, no studies were performed to evaluate the impact of *Rv2419c*-disrupted gene in bacterial virulence and survival in the host, as well as in MGLP production. Although *Rv2419c* was first annotated as coding for a putative phosphoglycerate mutase, GpgP was proved recently to display only residual mutase activity (Mendes, Maranha et al. 2011). The recombinant GpgP from *M. tuberculosis* displays dephosphorylating activity against mannosyl-3-phosphoglycerate, mannosylglucosyl-3-phosphoglycerate and, with much higher specificity, against GPG (Mendes, Maranha et al. 2011). Highlighting its function, it was reclassified as glucosyl-3-phosphoglycerate phosphatase (EC 3.1.3.85). In this thesis, the structure of GpgP is elucidated, as well as the molecular details of substrate binding and catalysis (Chapter 2). For this reason, the histidine phosphatase superfamily to which GpgP belongs is described in more detail in section 1.3.

1.2.3.2. MGLP Elongation

As initially proposed, GG is considered the precursor of α-(1→6)-DGG that in turn is the starting moiety for the α-(1→4)-glycosidic elongation of MGLP (Kamisango, Dell et al. 1987). The gene *Rv3032*, coding for a α-(1→4)-glycosyltransferase, was proposed to be involved in MGLP elongation (Stadthagen, Sambou et al. 2007). This gene displays a limited distribution within prokaryotes, with orthologs only among mycobacteria species. Moreover, the disruption of gene *Rv3032* led to impairment of mycobacterial *in vitro* growth (Sassetti, Boyd et al. 2003, Griffin, Gawronski et al. 2011, DeJesus, Gerrick et al. 2017). The mutation of this gene in *M. smegmatis* conduced to a significant reduction of MGLP production, concomitant with a reduction of MGLP chain length (Stadthagen, Sambou et al. 2007). In agreement, the introduction of an exogenous gene to the *Rv3032*-deficient strain restored the production of MGLP and shifted the MGLP length towards the mature forms (Stadthagen, Sambou et al. 2007). Although *Rv3032* seems to be clearly involved in MGLP biosynthesis, the low levels of MGLP that were found in the *RV3032*-deficient strain suggest that a compensatory mechanism may exist, and that another gene may partially replace *Rv3032* function. Indeed, a second gene (*Rv1212c*) encoding an α-(1→4)-glycosyltransferase (GlgA) and known to be involved in glycogen and capsular glucan synthesis was shown to compensate the absence of gene *Rv3032* (Sambou, Dinadayala et al. 2008). However, unlike the *Rv3032* mutant, the *Rv1212c*-disrupted

strain was still able to produce wild-type amounts of MGLP. Curiously, overexpression of gene *Rv3032* fully restored the production of glucan synthesis in the *Rv1212c* mutant (Sambou, Dinadayala et al. 2008). Also, the *Rv3032* mutant revealed an important decrease in glycogen content compared to the wild-type strain (Stadthagen, Sambou et al. 2007). Therefore, these two genes seem to have partially redundant functions, although *Rv1212c* seems to be preferentially used for glucan synthesis, whereas *Rv3032* is preferred to synthetize glycogen and MGLP. A double knockout mutant for genes *Rv1212c* and *Rv3032* was never recovered, suggesting that at least one of these genes is strictly essential for mycobacterial survival (Sambou, Dinadayala et al. 2008).

Rv3032 is separated by only 32 bp from gene *Rv3031*, suggesting that they might be expressed together. Gene *Rv3031* was reported as essential for mycobacterial *in vitro* growth (Sassetti, Boyd et al. 2003, Griffin, Gawronski et al. 2011, DeJesus, Gerrick et al. 2017). Up to date, only a single study highlighted the contribution of *Rv3031* for MGLP biosynthesis. Disruption of the *M. smegmatis MSMEG_2349* gene induced a significant reduction in MGLP cell content and shifted the chain type production from mature to short polysaccharide chains (Stadthagen, Sambou et al. 2007). *MSMEG_2349* is an ortholog of *Rv3031*, with *MSMEG_2349* and *Rv3031* encoding proteins sharing 80% sequence identity for 99% sequence coverage (Kapopoulou, Lew et al. 2011). This suggested that *Rv3031* might be implicated in MGLP production, however unequivocal functional evidence is still needed. The *Rv3031* gene displays sequence similarity with enzymes from the glycoside hydrolase family 57 (GH57) (CAZy database, http://www.cazy.org/Glycoside-Hydrolases.html). Within GH57 there are two structurally characterized enzymes that catalyse α-(1→6)-branching points in polysaccharides (EC 2.4.1.18). These proteins are produced by *Thermococcus kodakaraensis* (Murakami, Kanai et al. 2006, Santos, Tonoli et al. 2011) and *Thermus thermophilus* (Palomo, Pijning et al. 2011) and with which *Rv3031*-encoding protein displays 30 and 34% sequence identity, respectively. Although the sequence identity is relatively low, the overall structure and function are usually better preserved. Therefore, *Rv3031* is a potential candidate for α-(1→6)-DGG synthesis. The proposed mechanism for the *T. thermophilus* branching enzyme suggests that this enzyme acts on α-(1→4)-linked polysaccharide, hydrolysing the α-(1→4)-linkage and promoting a new α-(1→6)-linkage between the same sugar units (Palomo, Pijning et al. 2011). If the *Rv3031*-encoding enzyme catalyses the α-(1→6)-DGG formation through an identical mechanism, it is likely that a α-D-glucopyranosyl-(1,4)-α-D-glucopyranosyl-(1,2)-D-glycerate (α-(1→4)-DGG) molecule precedes the reported α-(1→6)-DGG.

In close vicinity to *Rv3031* and *Rv3032*, there are three additional genes coding for two putative *S*-adenosylmethionine-dependent methyltransferases (*Rv3030* and *Rv3037c*) and one putative acyltransferase (*Rv3034c*), which are also likely to play a role in MGLP biosynthesis (Stadthagen, Sambou et al. 2007). Analysis of gene essentiality by *Himar*1-based transposon mutagenesis indicated that *Rv3030* (Sassetti, Boyd et al. 2003, Griffin, Gawronski et al. 2011, DeJesus, Gerrick et al. 2017) and *Rv3034c* (Griffin, Gawronski et al. 2011, DeJesus, Gerrick et al. 2017) are essential for mycobacterial *in vitro* growth. Although *Rv3037c* was considered non-essential for mycobacterial *in vitro* growth (Sassetti, Boyd et al. 2003, DeJesus, Gerrick et al. 2017), this gene was predicted to be required for mycobacterial survival in macrophages (Rengarajan, Bloom et al. 2005).

Recently, the function of gene *Rv2418c* (adjacent to the GpgP-encoding gene) was elucidated by *in vitro* analysis of the activity of recombinant protein orthologs from *M. smegmatis* and *M. hassiacum* (Maranha, Moynihan et al. 2015). This gene encodes an acyltransferase, named octanoyltransferase (OctT), that transfers *in vitro* the ester groups from octanoyl- and hexanoyl-CoA preferentially to the α-(1→6)-DGG (Maranha, Moynihan et al. 2015). Since methylation of MGLP was only observed for partially acylated gluco-oligosaccharides (Grellert and Ballou 1972), OctT may be responsible for the initial acylation of the polysaccharide required for 6-*O*-methylation to occur. Although gene *Rv2418c* was considered essential for mycobacterial *in vitro* growth (Sassetti, Boyd et al. 2003, Griffin, Gawronski et al. 2011), a more recent study reported only an *in vitro* growth defect for a *Rv2418c*-disrupted strain (DeJesus, Gerrick et al. 2017).

An octanoyl group was previously suggested to be present at the glycerate moiety of MGLP (Smith and Ballou 1973), however the analysis of the reaction product obtained *in vitro* by OctT showed that octanoate was added to the O6 atom of the second glucose unit of α-(1→6)-DGG (Maranha, Moynihan et al. 2015). The presence of octanoate at atom O6 may impair the addition of a new α-(1→6)-linked glucose, directing to α-(1→4)-glucan elongation. Distinct acyl groups were initially proposed to be added to MGLP by a single enzyme, however the same report showed that octanoyl-CoA is a poor substrate for the acetyltransferase enzyme preparation analysed (Tung and Ballou 1973). In contrast, OctT displays *in vitro* significantly lower efficiency for shorter chain fatty acids (Maranha, Moynihan et al. 2015). Altogether, these studies suggest that more than one acyltransferase may be required for MGLP maturation. As previously mentioned, the *Rv3034c* gene is likely to encode an acyltransferase, and due to its localization in the mycobacterial genome its transcription may occur simultaneously to other MGLP

biosynthetic enzymes. Therefore, *Rv3034c* is a promising candidate for catalysing an acylation reaction other than octanylation using MGLP as acceptor substrate.

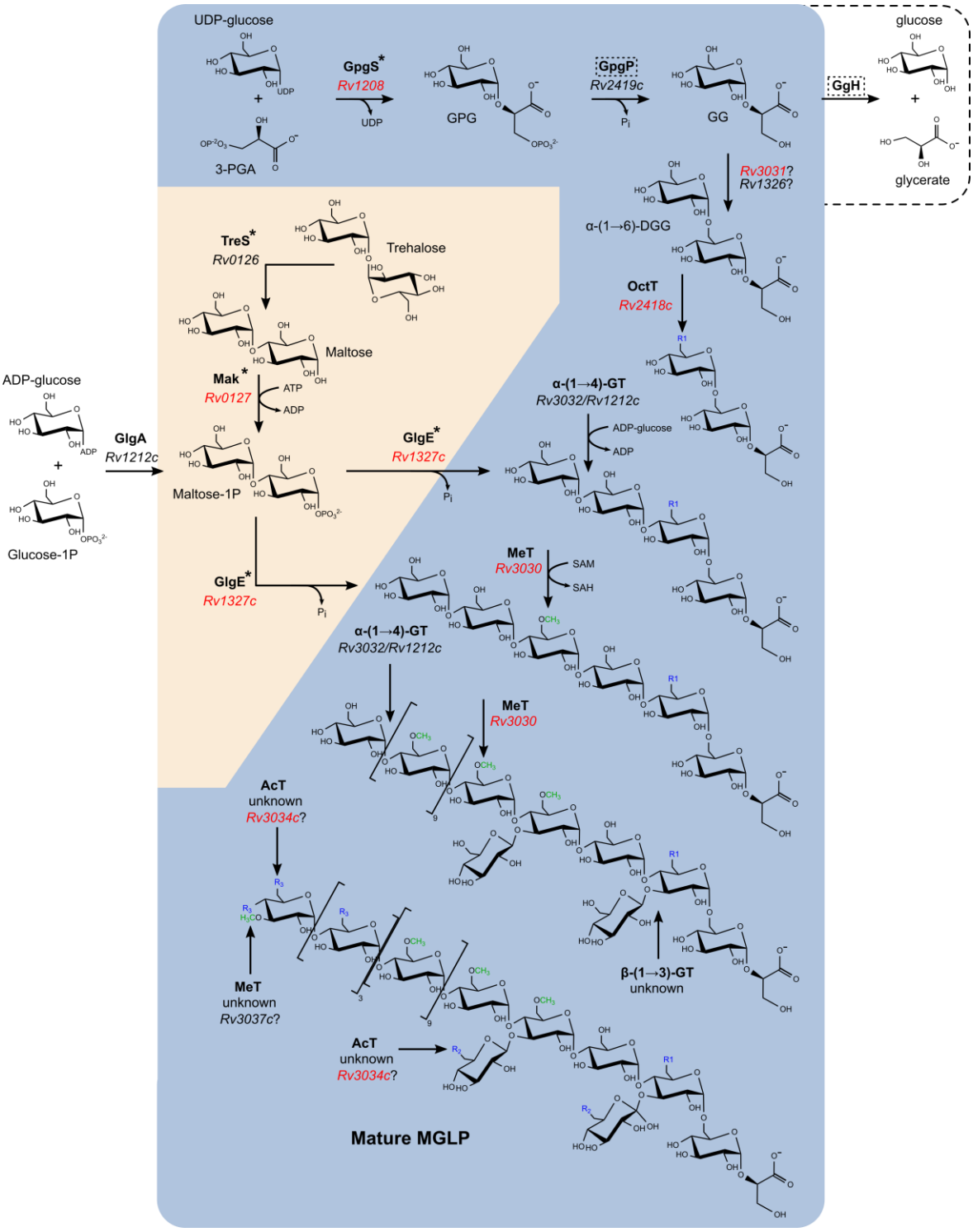


Figure 1.5 – Proposed pathways for MGLP biosynthesis. The GpgS-GpgP and TreS-Mak-GlgE pathways are highlighted in blue and wheat, respectively. The enzymes catalysing the indicated reactions and their encoding genes are in bold and italic, respectively. The genes with question marks were proposed to be involved in the corresponding reaction. Essential genes are coloured red. Enzymes marked with an asterisk were characterized structurally. The glycosylation, methylation and most of the acylation reactions occur at unknown stages of MGLP biosynthesis. Methyl and acyl groups [octanoate (R1), succinate (R2), and acetate, propionate and isobutyrate (R3)] are coloured in green and blue, respectively. A GgH, apparently restricted to rapid-growing

mycobacteria, may be responsible for GG recycling (dashed line). The enzymes structurally characterized in this thesis are in dotted squares. The abbreviations are: 3-PGA, 3-phosphoglyceric acid; α-(1→4)-GT, α-(1→4)-glucosyltransferase; α-(1→6)-DGG, α-(1→6)-diglucosylglycerate; β-(1→3)-GT, β-(1→3)-glucosyltransferase; AcT, acetyltransferase; GG, glucosylglycerate; GgH, glucosylglycerate hydrolase; glucose-1P, glucose-1-phosphate; GPG, glucosyl-3-phosphoglycerate; GpgP, glucosyl-3-phosphoglycerate phosphatase; GpgS, glucosyl-3-phosphoglycerate synthase; maltose-1P, maltose-1-phosphate; Mak, maltokinase; MeT, methyltransferase; OctT, octanoyltransferase; TreS, trehalose synthase; SAM, *S*-adenosylmethionine; SAH, *S*-adenosyl-homocysteine. Adapted from (Mendes, Maranha et al. 2012, Alarico, Costa et al. 2014, Maranha, Moynihan et al. 2015).

1.2.3.3. Alternative pathway

An alternative pathway for MGLP biosynthesis has also been hypothesized, starting with the disaccharide trehalose (Figure 1.5) (Kalscheuer, Syson et al. 2010). This pathway seems to be controlled by three genes: *Rv0126*, *Rv0127*, *and Rv1327c*, coding for a trehalose synthase (TreS), a maltokinase (Mak), and a maltosyltransferase (GlgE), respectively. Three studies have evaluated the essentiality of these genes by *Himar1*-based transposon mutagenesis (Sassetti, Boyd et al. 2003, Griffin, Gawronski et al. 2011, DeJesus, Gerrick et al. 2017). While *Rv1327c* was considered an essential gene for mycobacterial *in vitro* growth in all these studies, the essentiality of *Rv0126* and *Rv0127* was only reported by two (Sassetti, Boyd et al. 2003, Griffin, Gawronski et al. 2011) and one (Griffin, Gawronski et al. 2011) of them, respectively. An *in vivo* study linked the *Rv1327c* essentiality to the accumulation of toxic levels of maltose-1-phosphate (maltose-1P) in a *Rv1327c*-deleted mutant (Kalscheuer, Syson et al. 2010).

This alternative pathway involves three enzymatic steps, including the conversion of trehalose into maltose by TreS (Figure 1.6B) (Caner, Nguyen et al. 2013, Miah, Koliwer-Brandl et al. 2013), followed by maltose phosphorylation by Mak (Figure 1.6C) (Mendes, Maranha et al. 2010, Roy, Usha et al. 2013, Li, Guan et al. 2014, Fraga, Maranha et al. 2015) and finally the transference of the maltose moiety from maltose-1P to the non-reducing end of α-1,4-glucans by GlgE (Figure 1.6D) through an α-retaining catalytic mechanism (Kalscheuer, Syson et al. 2010, Syson, Stevenson et al. 2011, Lindenberger, Veleti et al. 2015). This alternative pathway is dispensable for viability and virulence as long as the GpgS-GpgP pathway is functioning. Indeed, the simultaneous inhibition of both pathways proved to be lethal for *M. tuberculosis* H37Rv (Kalscheuer, Syson et al. 2010). GlgA, one of the candidates for MGLP elongation (section 1.2.3.2), was recently

demonstrated to catalyse the formation of maltose-1P using ADP-glucose and glucose-1-phosphate (glucose-1P) as donor and acceptor substrates, respectively (Koliwer-Brandl, Syson et al. 2016). Furthermore, GlgA was clearly more efficient in catalysing the formation of the maltose-1P than in elongating the non-reducing ends of glycogen using ADP-glucose as a donor substrate, suggesting that mycobacterial GlgA may play only a minor direct role in α-glucan elongation (Koliwer-Brandl, Syson et al. 2016).

The gene coding for GlgE is adjacent to a gene encoding an α-1,6-branching enzyme (*Rv1327c*) that is also suspected to be implicated in MGLP pathway. This glycogen branching enzyme, named GlgB (EC 2.4.1.18), catalyses the introduction of α-(1→6)-linked branches into linear α-1,4-glucans. Its coding gene was found essential for mycobacterial survival (Sambou, Dinadayala et al. 2008) and later on, its essentiality was linked to the intracellular accumulation of maltose-1P (Kalscheuer, Syson et al. 2010, Miah, Bibb et al. 2016). It was hypothesized that in the absence of GlgB, the activity of GlgE is limited by the availability of acceptor substrate, leading to toxic build-up of maltose-1P (Kalscheuer, Syson et al. 2010).

The assembly mechanism of α-glucans by GlgE and GlgB was proposed recently (Rashid, Batey et al. 2016). According to this mechanism, the polysaccharide synthesis is initiated by GlgE and its donor substrate maltose-1P. GlgE yields a linear polysaccharide sufficiently long (~16 sugar units) for GlgB to introduce a branch. The branching by *M. tuberculosis* GlgB seems to be strictly intrachain. After branching, GlgE preferentially extends the new chain until it is long enough to undergo a new GlgB branching. These simultaneous events generate the arboreal structure of the α-glucans. The length of α-glucans is likely controlled by the ability of GlgE to catalyse a disproportionation reaction through maltosyl transference between malto-oligosaccharides. The *M. tuberculosis* GlgB transfers maltosyl units from a donor to an acceptor polysaccharide, both at least four sugar units long (Kalscheuer, Syson et al. 2010). During MGLP biosynthesis, the activity of GlgB may be limited by the methyl and acyl groups introduced at atom O6 of the glucose units.

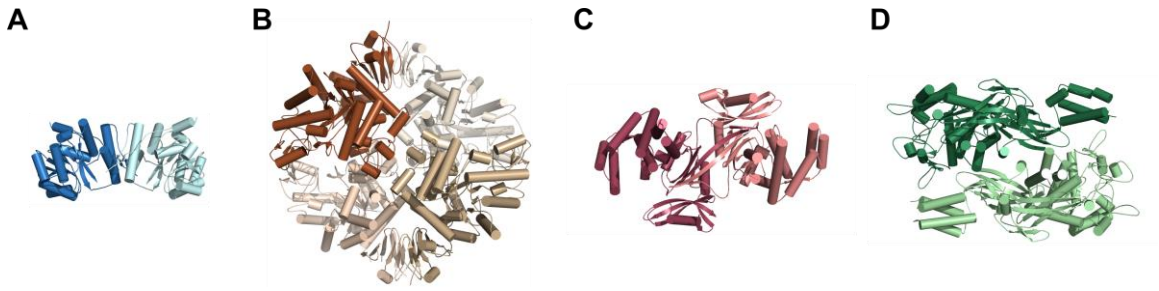


Figure 1.6 – Three-dimensional structure of mycobacterial GpgS, TreS, Mak and GlgE. The α-helices and β-strands of *M. tuberculosis* GpgS (A, PDB accession code 3E26), *M. smegmatis* TreS (B, PDB accession code 3ZO9), *M. tuberculosis* Mak (C, PDB accession code 4O7O) and *M. tuberculosis* GlgE (D, PDB accession code 4U33) are represented by cylinders and flat arrows, respectively. Oligomerization is highlighted in different colour shades.

1.2.3.4. GG bioversatility and potential roles

Glucosylglycerate, one of the earlier intermediates in MGLP biosynthesis, is widespread among microorganisms. At least four distinct GG-containing molecules are known, including MGLPs, α-(1→6)-DGG, mannosylglucosylglycerate (MGG) and GG-containing glycolipids (Nunes-Costa, Maranha et al. 2017). The first GG-containing molecules to be described were the MGLPs (Lee 1966), which have been isolated from several species of *Mycobacterium*, *Nocardia* and *Streptomyces* genera (Table 1.2). α-(1→6)-DGG, also a precursor of the mycobacterial MGLPs, was found in the hyperthermophilic bacterium *Persephonella marina*, which seems to accumulate α-(1→6)-DGG as a response to salt stress (Lamosa, Rodrigues et al. 2013). Up to now, GG-containing glycolipids were only isolated from *Nocardia caviae* (Pommier and Michel 1981). The most recently discovered GG-containing molecule, MGG, was detected for the first time in thermophilic anaerobic bacterium *Petrotoga miotherma* (Jorge, Lamosa et al. 2007) and later in *Petrotoga mobilis* and mesophilic *Rhodopirellula baltica* (Fernandes, Mendes et al. 2010, d'Avo, Cunha et al. 2013). Two pathways for MGG production in *P. mobilis* was proposed and one of which uses GG as a precursor molecule (Fernandes, Mendes et al. 2010).

GG is also considered a compatible solute that can be accumulated intracellularly even at very high levels without affecting the cell functions (Empadinhas and da Costa 2011). Free GG was found in diverse organisms and its accumulation proposed as a strategy to cope with salt stress and nitrogen starvation. In the cyanobacterium *Synechococcus* PCC7002 (formaly known as *Agmenellum quadruplicatum*), the intracellular GG levels were shown to increase during growth in nitrogen poor medium, pointing for the novel role

of GG during nitrogen starvation (Kollman, Hanners et al. 1979). Later, the accumulation of GG in this species was also associated with salt stress and shown to be considerable stimulated by salt stress and nitrogen starvation combined (Engelbrecht, Marin et al. 1999, Klahn, Steglich et al. 2010). Similarly, in the marine cyanobacterium *Prochlorococcus marinus* the accumulation of GG during growth in high salinity medium was exacerbated by low nitrogen availability (Klahn, Steglich et al. 2010). Interestingly, both *Synechococcus* spp. and *P. marinus* accumulate glutamate when nitrogen is available, however when nitrogen is limited the intracellular glutamate is reduced and replaced by the negatively charged GG (Klahn, Steglich et al. 2010). The planctomycete *Gimesia maris* also replaces the glutamate with GG as main anionic solute when growing under high salt and low nitrogen conditions (Ferreira, Soares et al. 2016). The accumulation of GG instead of glutamate in hypertonic and nitrogen-deficient medium was also observed for the marine α-proteobacterium *Dinoroseobacter shibae* (Kleist, Ulbrich et al. 2017).

Although in the previous examples both nitrogen scarcity and osmotic pressure were necessary to induce significant GG accumulation, in *Hydrogenothermus marinus* and *Mycobacterium* spp. that was not observed. In *H. marinus* the salt stress was shown to be sufficient to trigger the accumulation of GG (Beblo-Vranesevic, Galinski et al. 2017). On the other hand, in *M. smegmatis* and *M. hassiacum* the accumulation of GG occurred under nitrogen starvation and was shown to be independent on osmotic pressure (Behrends, Williams et al. 2012, Alarico, Costa et al. 2014). In *M. smegmatis*, the accumulation of GG was linked to the up-regulation of GpgS encoding gene (*MSMEG_5084*) that catalyzes the first step of GG synthesis (Behrends, Williams et al. 2012). In agreement, a *gpgS* deleted mutant, unable to produce detectable amounts of GG, showed a slight reduction in growth rate under nitrogen-limiting conditions compared to wild-type (Behrends, Williams et al. 2012). Although GG seems not essential for mycobacterial survival under low nitrogen, the phenotypic effect observed suggest that it may contribute for bacterial fitness in diverse environmental situations.

In vitro, GG was shown to increase the thermostability of proteins and prevent the loss of activity for a number of enzymes (Faria, Mingote et al. 2008, Sawangwan, Goedl et al. 2010, Sato, Kitamoto et al. 2014). Additionally, GG also displays a protective effect against DNA scission (Sato, Kitamoto et al. 2014). Thus, the accumulation of intracellular GG may increase the stability of macromolecules, such as nucleic acids and proteins, fundamental for keeping cell viability until improvement of the environmental conditions.

1.2.3.5. GG recycling

The GG stored during mycobacterial growth under nitrogen limited conditions was shown to decrease when the nitrogen availability was restored (Alarico, Costa et al. 2014). The reduction of intracellular GG in *M. hassiacum* was linked to the up-regulation of a gene encoding for a glucosylglycerate hydrolyse (GgH, EC 3.2.1.-) that catalyses *in vitro* the hydrolysis of GG to glucose and glycerate (Alarico, Costa et al. 2014). Therefore, GgH seems to contribute for recycling the accumulated GG, allowing the assimilation of the metabolite when nitrogen availability is restored, and the rapid energy mobilisation required for cell growth and division.

Although no information is available regarding the importance of GgH for bacterial survival and pathogenicity, this enzyme is highly conserved among rapid-growing mycobacteria and members of related genera, reflecting its evolutive importance. On the basis of its amino acid sequence, GgH is grouped into the glycoside hydrolase family 63 (GH63). In Chapter 3, the three-dimensional structure of recombinant *M. hassiacum* GgH (*Mha*GgH) is presented, highlighting its molecular determinants of substrate specificity. An overview of glycoside hydrolases classification, structural features and mechanism of reaction is given in section 1.4.

1.3. Histidine phosphatases

The histidine phosphatases are found in both prokaryotic and eukaryotic organisms, forming a group with an impressive range of activities, including regulatory and signalling functions (Rigden 2008). These proteins share a small RHG motif at the beginning of their amino-acid sequence, whose histidine is transiently phosphorylated during the catalytic reaction (see section 1.3.1) (Rigden 2008).

The histidine phosphatase superfamily (clan ID CL0071) of the Pfam database (Finn, Coggill et al. 2016) hosts all histidine phosphatases. However, due to their sequence divergence, histidine phosphatases are subdivided into two branches, named branch 1 (Pfam family ID PF00300) and branch 2 (Pfam family ID PF00328). In the present version of the Pfam database (Pfam 31.0, March 2017, http://pfam.xfam.org/), branch 1 is composed mainly of prokaryotic proteins (17027 prokaryotic sequences against 9272 eukaryotic sequences), whereas branch 2 contains a higher number of eukaryotic members (5975 eukaryotic sequences against 305 prokaryotic sequences).

The development of bioinformatics tools allowed the fast annotation of the genomic data produced over the years. However, automated genome annotation tools are also prone to mis- and over-annotation (Friedberg 2006). The sequence and functional diversity of the histidine phosphatase superfamily represent a particular challenge. Also, since the histidine phosphatase superfamily has a mutase-dominated history, many members found in the genomes have been annotated as mutases (see section 3.3), when phosphatase activity is much more likely (Rigden 2008). Thus, the functional and structural characterization of these enzymes is essential to validate the annotated function and optimize future annotations by increasing knowledge regarding structure-function relationships.

1.3.1. Mechanism of reaction

The catalytic activity of histidine phosphatases is centred in the phosphorylation and dephosphorylation of a highly conserved histidine (Figure 1.7) usually localized closer to the N-terminus, immediately after the first β-sheet.

The transference of the phosphoryl group from the substrate to the enzyme is aided by several surrounding residues, which are responsible for holding the phosphoryl group during the course of reaction (Wang, Liu et al. 2006). These residues include a pair of

arginines and an additional histidine that are highly conserved among the members of the histidine phosphatase superfamily (Rigden 2008). Additionally, a highly conserved glycine, occasionally replaced by an alanine, is part of the well-known and characteristic RHG motif. This residue forms a hydrogen bond *via* its carbonyl group with the catalytic histidine side chain, important for keeping it in an appropriate position for catalysis. The phosphate pocket also contains other less conserved residues (represented as PP in Figure 1.7) involved in additional phosphoryl group interactions, including an asparagine (Hamada, Kato et al. 2005) and a glutamine (Rigden, Littlejohn et al. 2003).

Other important component of the catalytic machinery is the proton donor, required to donate a proton to the leaving group of the substrate upon the transference of the phosphoryl group to the catalytic histidine. Most of the branch 1 members have a glutamate as proton donor (Lin, Li et al. 1992), however is was predicted that in other members an aspartate is likely to perform this function (Hamada, Kato et al. 2005). Although for the branch 2 members it is consensus that the proton donor is an aspartate (Ostanin and Van Etten 1993), an histidine residue was proposed to act as secondary proton donor (Lee, Cottrill et al. 2003). The proton donor is usually localized shortly after strands β1 or β3 in branch 1, or β4 in branch 2 members.

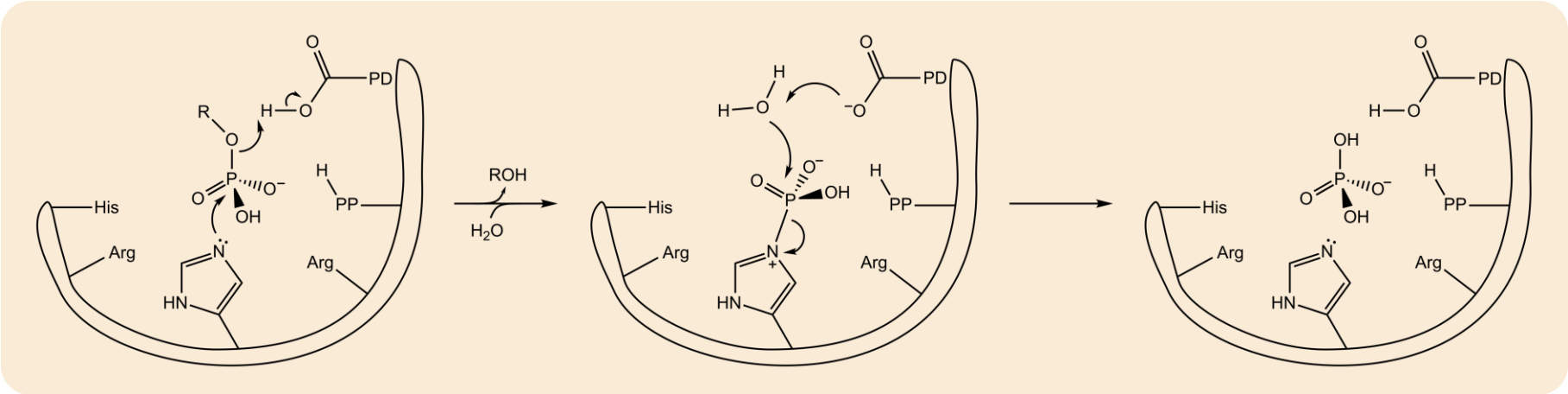


Figure 1.7 – Schematic representation of the catalytic mechanism of histidine phosphate superfamily members. The catalytic histidine residue becomes phosphorylated during the course of the reaction, while the surrounding residues interact electrostatically with the phosphoryl group. Neutral or positively charged residues (PP) may also contribute for phosphoryl group stabilization by hydrogen bonding. The proton donor (PD), an aspartate or glutamate, is indicated. Adapted from (Rigden 2008).

1.3.2. Vanadate as transition-state mimic

The identification of the residues involved in substrate binding and catalysis is usually performed by engineered site-specific amino acid substitutions and/or by structure determination in complex with substrates, substrate analogues or inhibitors.

In general, phosphatases are inhibited by oxometalate anions such as arsenate, molybdate, tungstate and vanadate, which are often described as phosphate analogues (Dennis Chasteen 1983, Zhang, Zhou et al. 1997, Foster, Young et al. 1998, Reiter, White et al. 2002, Hamada, Kato et al. 2005, Peck, Sunden et al. 2016). Among these compounds, vanadate is the most commonly used due to its structural and chemical similarities to phosphate and to its high binding affinity (Crans, Smee et al. 2004).

Vanadate and phosphate are structural analogues of similar size, displaying circumscribing spheres of 125 and 105 $Å^3$, respectively (Rehder 2013). The most significant difference between vanadate and phosphate is the ability of vanadate to easily reach a stable penta-coordination, commonly in a trigonal bi-pyramidal geometry, while this type of coordination is just a transition state in the case of phosphate (Crans, Tarlton et al. 2014, Rehder 2015). Thus, vanadate can mimic the transition-state of the phosphoryl group transfer, by forming a stable complex with the catalytic residue (Figure 1.8). Moreover, the formation of this complex leads to enzyme inhibition since the active site is no longer accessible to the substrate.

Vanadate is thus a powerful tool for the structural study of phosphatases in general (Davies and Hol 2004). In one hand, vanadate is able to form a highly stable complex with the enzyme and mimic the transition-state of phosphoryl group transfer, allowing the identification of the residues involved in substrate binding and providing information about the enzyme mechanism. On the other hand, its high affinity of binding is important to promote the fast formation of stable complexes, increasing the molecular homogeneity necessary to produce high quality and easily interpretable structural data.

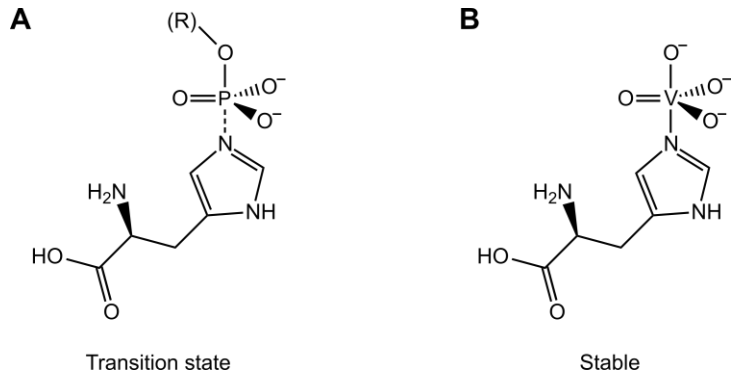


Figure 1.8 – Analogy between phosphate and vanadate. Penta-coordinated species produced by the interaction of phosphate (A) and vanadate (B) with a histidine residue. Phosphate forms a labile transition state, symbolized by a dashed bond between phosphorus and nitrogen atoms, whereas vanadate forms a stable complex. Adapted from (Rehder 2015).

1.3.3. Histidine phosphate mutases

Although most histidine phosphatases are responsible for the simple removal of a phosphoryl group from the substrate (EC 3.1.3.-), other members can act as mutases (EC 5.4.2.-) (Rigden 2008). The phosphatase mutases (PGMs) are divided in two groups, named monophosphoglycerate mutases (mPGMs) and bisphosphoglycerate mutases (bPGMs), based on the number of phosphoryl groups in the reaction product glycerate. Since mPGMs may or not require 2,3-bisphosphoglycerate as initiator of the reaction, they are also subdivided into 2,3-bisphosphoglycerate dependent (dPGM) and 2,3-bisphosphoglycerate independent PGMs (iPGMs), respectively.

While mPGMs catalyse the interconversion between 2-phosphoglycerate and 3-phosphoglycerate (EC 5.4.2.11, dPGMs; EC 5.4.2.12, iPGMs), bPGMs are involved in the production of 2,3-bisphosphoglycerate from 1,3-bisphosphoglycerate (EC 5.4.2.4). The activity of dPGMs and bPGMs is interconvertible, although at a much lower reaction rate (Fothergill-Gilmore and Watson 1989).

The difference between mutases and phosphatases is the ability of the former to keep the intermediates in the active site, allowing their reorientation and protecting the phosphorylated enzyme from hydrolysis (Rigden 2008). This protection is likely promoted by the C-terminal tail. Indeed, its removal was associated to a reduction of the mutase activity (Walter, Nairn et al. 1999). Moreover, the structure of the C-terminal tail was only

determined for the phosphorylated form of dPGMs and found to interact with the residues implicated in substrate binding (Bond, White et al. 2001, Rigden, Littlejohn et al. 2003).

1.4. Glycoside hydrolases

Glycoside hydrolases (GHs), also known as glycosidases or glycosyl hydrolases (EC 3.2.1.-), are enzymes that catalyse the hydrolysis of the glycosidic linkage of glycosides (Figure 1.9), producing a hemiacetal or hemiketal sugar and the corresponding free aglycon (Withers and Williams 2017).

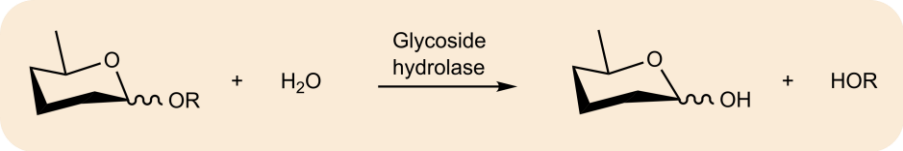


Figure 1.9 – General reaction catalysed by a glycoside hydrolase. Adapted from (Withers and Williams 2017).

These enzymes act on a variety of carbohydrates, and due to their ubiquity in nature, GHs are of high interest in human health and disease, as well as in biotechnology industries. GHs form a diverse group of enzymes that are classified based on their amino acid sequence and according to the Carbohydrate-Active Enzymes (CAZy) database system (http://www.cazy.org/Glycoside-Hydrolases.html) (Lombard, Golaconda Ramulu et al. 2014). Up to date, GHs are divided into 149 families (numbered from 1 to 149), and approximately one-third of the total number of families are grouped into 17 clans (named from GH-A to GH-Q), highlighting their folding similarities.

1.4.1. Mechanisms of reaction

The gross mechanisms of GHs were first proposed more than 60 years ago (Koshland 1953), remaining largely true today. The hydrolysis of the glycosidic bond occurs *via* a general acid catalysis, involving two critical residues, an acid and a nucleophile/base (Koshland 1953). The reaction consists in the nucleophilic substitution at the anomeric carbon by two possible mechanisms, named retaining and inverting mechanism according to the stereochemical outcomes for the hydrolysis of the glycosidic bond (Koshland 1953). While the retaining mechanism leads to the overall conservation of the anomeric configuration, the inversion mechanism induces its inversion (see sections 1.4.1.1. and

1.4.1.2. for details). Therefore, GHs are also classified into retaining and inverting GHs according to the type of hydrolysis mechanism used.

The increasing knowledge on GHs reaction mechanism has been promoted by the structural analysis of GHs and of their complexes with ligands. In addition to the vast number of GHs that operate by the classical Koshland retaining and inverting mechanisms, a growing number of enzymes acting on glycosides by slightly different to new mechanisms are coming to light (see section 1.4.1.3. for more details).

1.4.1.1. Classical inverting mechanism

The inverting mechanism, thus named due to the inversion of the anomeric configuration of the glycosyl group, comprises a single step reaction involving the formation of an oxocarbenium ion-like transition state (Figure 1.10A) (Koshland 1953). The reaction occurs *via* the nucleophilic attack by a water molecule, previously activated by a negatively charged base, and concomitantly to proton transfer from an acidic residue. This reaction involves typically glutamates or aspartates, which are usually separated by ~9.0 and 9.5 Å in α-glycosidases and β-glycosidases, respectively (McCarter and Withers 1994).

1.4.1.2. Classical retaining mechanism

The glycoside hydrolysis *via* a retaining mechanism occurs by a two-step reaction marked by the formation of two transition states and the preservation of the anomeric stereochemistry (Figure 1.10B) (Koshland 1953). During the first step, the nucleophilic attack is performed directly by an amino acid residue (nucleophile) to the anomeric carbon, concomitantly to proton transfer from the acidic residue to the glycosidic oxygen. The first oxocarbenium ion-like transition state is formed. As soon as the glycosidic oxygen is protonated, the aglycon is displaced whereas the glycosyl group remains covalently bound to the nucleophile. Due to the formation of this glycosyl-enzyme intermediate, this step is often referred as glycosylation step.

In a second reaction step, the residue previously acting as acid, is now negatively charged and acts as a base, activating a water molecule. The activated water performs a nucleophilic attack against the anomeric carbon of the glycosyl group and the second

oxocarbenium ion-like transition state forms. Once the glycosydic oxygen becomes protonated, the glycosyl group is released. At the end of the reaction, the charge of the nucleophile and the protonation state of the acid are re-established. The retaining mechanism usually involves glutamates and aspartates as catalytic residue, which are closer (~4.8 and 5.3 Å for α-glycosidases and β-glycosidases, respectively) than those of inverting GHs (McCarter and Withers 1994).

1.4.1.3. Variations of the classical retaining mechanism

A variation of the classical retaining mechanism was found among myrosinases, S-glycosidases responsible for hydrolysing the plant anionic 1-thio-β-D-glucosides (glucosinolates) and whose products of reaction were linked to plant defence (Bones and Rossiter 1996, Cottaz, Henrissat et al. 1996). Although belonging to the GH family 1 (GH1), myrosinases are unusual members in this family in that they lack the classical acid/base residue (glutamate) at their active site. The catalytic glutamate residue of the classical β-glucosidases is replaced by a glutamine in myrosinases that contributes for the correct positioning of the nucleophilic water (Burmeister, Cottaz et al. 1997). The nucleophilic water seems to be activated by the co-factor ascorbic acid, used by myrosinases as an alternative base (Burmeister, Cottaz et al. 2000).

Another variation was observed on sialidases and trans-sialidases. Sialidases, also known as neuraminidases, and trans-sialidases hydrolyse or transfer sialic acids with retention of the anomeric configuration (Chong, Pegg et al. 1992, Kao, Lerner et al. 1997, Todeschini, Mendonca-Previato et al. 2000). While sialidases are separated into three distinct families based on their amino acid sequence similarities (GH33, GH34 and GH83), trans-sialidases are found in a single GH family (GH33). Despite their amino acid sequence differences, they seem to share a common active-site machinery. These enzymes use a tyrosine as catalytic nucleophile, instead of the usual acidic residue. Since tyrosine is a neutral amino acid, an adjacent base residue (glutamate) is required to enhance its nucleophilicity (Watson, Dookhun et al. 2003, Watts, Damager et al. 2003, Amaya, Watts et al. 2004).

Some enzymes, belonging to GH families 18, 20, 25, 56, 84 and 85, hydrolyse substrates containing an *N*-acetyl group at the 2-position. These enzymes lack the catalytic nucleophile and use the 2-acetyl group as intramolecular nucleophile, leading to the formation of an oxazolinium ion intermediate (Terwisscha van Scheltinga, Armand et al.

1995, Drouillard, Armand et al. 1997, Mark, Vocadlo et al. 2001, Vocadlo and Withers 2005).

Finally, the two GH families that stand out mechanistically are GH4 and GH109. The enzymes from these families are unique among GHs because they require NAD^+ for catalysis. The mechanism involves elimination and redox reactions, leading to the formation of anionic intermediates rather than the oxocarbenium ion-like transition states of the classical Koshland mechanisms. In GH4, NAD^+ effects a transient oxidation of the substrate, whereas a divalent cation, usually Mn^{2+}, stabilizes the negative charges that develop during the reaction (Rajan, Yang et al. 2004, Yip, Varrot et al. 2004). Although GH109 seems to follow a mechanism very similar to that of the GH4 family, GH109 does not require a metal cofactor for its activity (Liu, Sulzenbacher et al. 2007).

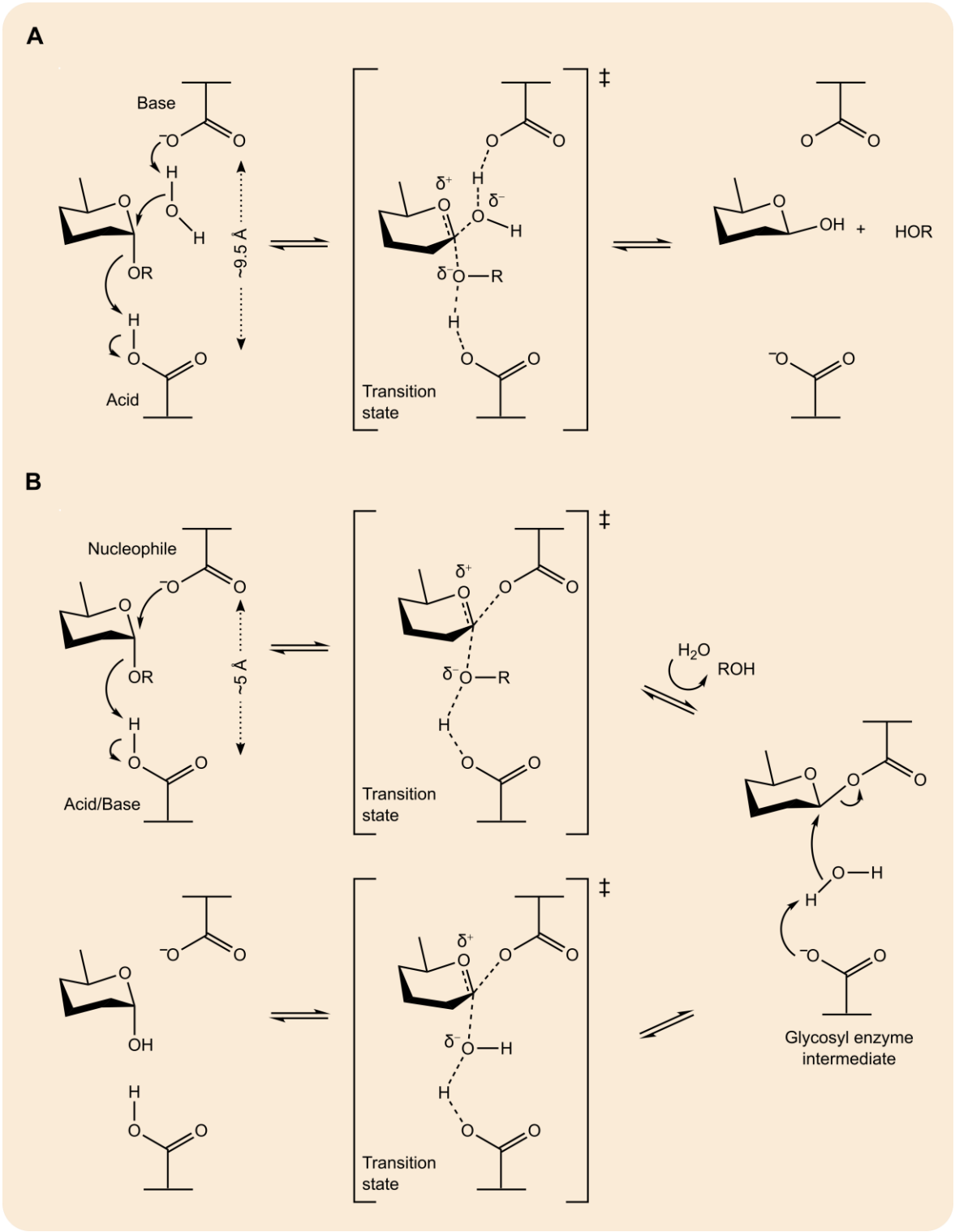


Figure 1.10 – General glycosidase mechanism for inverting α-glycosidase (A) and retaining α-glycosidase (B). Adapted from (Withers and Williams 2017).

Chapter 2

Structural characterization and substrate binding mode of *M. tuberculosis* GpgP

2.1. Introduction

The genus *Mycobacterium* contains more than 180 species including those belonging to the *Mycobacterium tuberculosis* complex, the causative agents of tuberculosis. Although the number of TB deaths has fallen by ~26% from 2000 to 2016, TB remains one of the top 10 causes of death worldwide (WHO 2017). It is estimated that about one-fourth of the world's population has latent TB infection with 5 to 15% lifetime risk of developing the disease (WHO 2017). The increasing number of multidrug resistant and even extensively drug-resistant TB strains urges for the identification of novel targets against which efficient drugs can be designed.

Mycobacteria have one of the thickest biological membranes known, rich in complex carbohydrates and unique lipids, namely mycolic acids, to which mycobacteria owe much of their pathogenicity (Jackson 2014, Jankute, Cox et al. 2015). It has been proposed that the biosynthesis of fatty acids, such as mycolic acids, is regulated by intracellular polymethylated polysaccharides (Ilton, Jevans et al. 1971, Machida and Bloch 1973, Flick and Bloch 1975, Banis, Peterson et al. 1977). Two types of PMPSs are known: the methylmannose polysaccharides and the methylglucose lipopolysaccharides (Lee 1966, Gray and Ballou 1971). Interestingly, while the two types of PMPSs were found in the environmental mycobacteria, MGLPs were the only ones found in slow-growing pathogenic mycobacteria, such as *M. tuberculosis* (Mendes, Maranha et al. 2012).

The biosynthesis of MGLPs is a complex process that involves different classes of enzymes (Mendes, Maranha et al. 2012). The first intervenient enzyme in the MGLP biosynthetic pathway is a glucosyl-3-phosphoglycerate synthase (GpgS), and it was also the first enzyme to be biochemical and structurally characterized (Empadinhas, Albuquerque et al. 2008, Pereira, Empadinhas et al. 2008). GpgS synthesizes *in vitro* glucosyl-3-phosphoglycerate from UDP-glucose and 3-phosphoglyceric acid (Empadinhas, Albuquerque et al. 2008). The reaction product of GpgS is the substrate of the second intervenient enzyme, glucosyl-3-phosphoglycerate phosphatase. *In vitro*, this enzyme dephosphorylates GPG producing glucosylglycerate (Mendes, Maranha et al. 2011).

The gene coding for GpgP is highly conserved among mycobacteria and related genera. Although it was reported as non-essential for mycobacterial *in vitro* growth (Sassetti, Boyd et al. 2003), the high conservation of this gene suggests that it may confer an evolutive advantage. Indeed, its reaction product, GG, is a compatible solute involved in osmoadaptation and a precursor of other molecules such as glycolipids and

polysaccharides (Pommier and Michel 1981, Kamisango, Dell et al. 1987, Jorge, Lamosa et al. 2007, Empadinhas and da Costa 2011, Behrends, Williams et al. 2012, Alarico, Costa et al. 2014, Nunes-Costa, Maranha et al. 2017). The impairment of GG production by targeting GpgP activity may reduce the ability of mycobacteria to adapt to a fluctuating environment, which is highly important during infection. Therefore, GpgP appears as potential target for drug development against mycobacteria infections, including TB.

Here, the crystallographic structure of a recombinant *M. tuberculosis* GpgP (*Mtu*GpgP) is presented. Moreover, the substrate binding mode of *Mtu*GpgP is also proposed based on the structure of its complex with the phosphate mimic vanadate and the product of reaction GG. The high resolution of the data allowed the identification of the amino-acid residues involved in substrate binding, orientation and catalysis.

2.2. Materials and methods

2.2.1. Expression and purification of recombinant *Mtu*GpgP

Escherichia coli BL21 (DE3) cells transformed with a pET30a-derived plasmid harbouring the optimized *Rv2419c* gene sequence (Mendes, Maranha et al. 2011) were grown in LB medium supplemented with 50 µg mL^{-1} kanamycin at 37°C to an OD_{600} of 0.6-0.8. At this point, the temperature was decreased to 20°C and expression induced with 1 mM IPTG. After 4 h growth, cells were harvested, resuspended in 20 mM sodium phosphate 7.5, 500 mM NaCl, 20 mM imidazole, 5 mM *β*-ME (buffer A) and stored at -20°C. Thawed cells were stirred gently on ice for 1 h after adding 5 µg mL^{-1} lysozyme, 5 µg mL^{-1} DNAse I, 5 mM $MgCl_2$, 1 mM PMSF and one tablet of cOmplete™ EDTA-free protease inhibitor cocktail (Roche). The protein extract was clarified by centrifugation and loaded onto an immobilized metal affinity column (GE Healthcare) loaded with nickel and pre-equilibrated with buffer A. Bound *Mtu*GpgP was eluted with 175 mM imidazole in buffer A. Fractions containing *Mtu*GpgP protein were pooled and loaded onto a HiPrep 26/60 Sephacryl 200 (GE Healthcare) size exclusion column using 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 10 mM DTT as mobile phase. The purest *Mtu*GpgP-containing fractions were dialyzed against 50 mM BTP pH 7.0, 50 mM NaCl, 10 mM DTT and concentrated on a 3 kDa cutoff ultrafiltration device (Millipore). The concentration of the purified *Mtu*GpgP was estimated by measuring the absorbance at 280 nm on a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies). The protein was flash-frozen in liquid nitrogen and kept at -80°C until needed.

2.2.2. Analytical size-exclusion chromatography

*Mtu*GpgP was analysed by size exclusion chromatography on a Superdex 200 increase 5/150 GL column (GE Healthcare) pre-equilibrated with 50 mM BTP pH 7.0, 50 mM NaCl, 10 mM DTT. The column was calibrated using ferritin (400 kDa), aldolase (158 kDa), ovalbumin (43 kDa) and ribonuclease A (13.7 kDa) as standards. The K_{av} *vs* log MW was calculated based on the equation $K_{av} = \frac{V_e - V_0}{V_t - V_0}$, 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.

2.2.3. Differential scanning fluorimetry

The melting temperature of *Mtu*GpgP was determined by a thermal shift (Thermofluor) assay. After centrifugation at 13,000x*g* and 4°C for 15 min, *Mtu*GpgP (0.2 mg mL^{-1}) was mixed with 5x SYPRO Orange (Life Technologies) in 50 mM BTP pH 7.0, 50 mM NaCl, 10 mM DTT, and loaded in 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 over time using a Cy3 dye filter (545 nm excitation, 585 nm emission). The experiment was performed in triplicate. The melting curves were analysed using the CFX Manager software (Bio-Rad), and the melting temperature determined as the inflexion point of the melting curve.

2.2.4. Chemical synthesis of glucosylglycerate

The chemical synthesis of glucosylglycerate [GG, 2-*O*-(α-D-glucopyranosyl)-D-glycerate)], was performed by Eva Lourenço and Rita Ventura (Bioorganic Chemistry Group, ITQB – Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Portugal) as previously described (Lourenco, Maycock et al. 2009, Lourenco and Ventura 2011).

2.2.5. Crystallization of *Mtu*GpgP

Initial crystallization conditions were screened at 293 K using the sitting-drop vapour-diffusion method with commercial sparse-matrix crystallization screens (Hampton Research). The drops consisted of equal volumes (1 µL) of protein (at 13 mg mL^{-1} in 50 mM BTP pH 7.0, 50 mM NaCl, 10 mM DTT) and precipitant solution and were equilibrated against 300 µL of precipitant. Crystals were obtained after 6 days using solution number 7 [0.2 M calcium chloride dihydrate, 20% (w/v) PEG 3,350] of PEG/Ion Crystallization Screen (Hampton Research) as precipitant. The condition was refined and improved crystals were obtained from 15 to 30% (w/v) PEG 3,350 both in presence and absence of 0.1 M MES pH 6.5 or 6.7. The *Mtu*GpgP-vanadate and *Mtu*GpgP-vanadate-GG complexes were obtained by co-crystallization of *Mtu*GpgP with approximately 50-fold molar excess of activated sodium orthovanadate (Sigma) and from drops composed of 1 µL each of protein and precipitant solutions and 2 µL of 10 mM sodium orthovanadate. For the *Mtu*GpgP-vanadate-GG complex, protein was previously incubated with a 10-fold

molar excess of GG for 1h on ice. Crystals appeared after approximately 15 days, were cryo-protected in mother liquor solution supplemented with 15% (v/v) glycerol and were flash-cooled in liquid nitrogen. Iodide-containing crystals were obtained from a second crystallization condition consisting of 0.2 M sodium iodide, 0.12 M MES pH 6.0, 20% (w/v) PEG 6,000. Crystals were cryo-protected in mother liquor solution containing 10% (v/v) glycerol and were flash-cooled in liquid nitrogen. The initial *Mtu*GpgP crystallization conditions were determined by Fernando Vilela and Sara Fernandes (IBMC, Porto, Portugal).

2.2.6. Data collection and processing

High-resolution data sets were collected from cryo-cooled (100 K) single crystals on a PILATUS 6M detector (Dectris) at beamlines ID29 (de Sanctis, Beteva et al. 2012) and ID23-1 (Nurizzo, Mairs et al. 2006) of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) and PROXIMA-1 of the French National Synchrotron Source (SOLEIL, Gif-sur-Yvette, France). All data sets were processed with XDS (Kabsch 2010) and reduced with utilities from the CCP4 software suite (Winn, Ballard et al. 2011), except for *Mtu*GpgP-vanadate-GG complex data, which were processed using the Grenoble Automatic Data ProcEssing pipeline (Monaco, Gordon et al. 2013). X-ray diffraction data collection and processing statistics are summarized in Table 2.1.

2.2.7. Structure determination, model building and refinement

The structure of *Mtu*GpgP was solved by SAD using the anomalous signal from an iodide-derivative crystal with the SAS protocol of Auto-Rickshaw pipeline (Panjikar, Parthasarathy et al. 2005, Panjikar, Parthasarathy et al. 2009). The refined coordinates were used as search model to solve the structure of all other *Mtu*GpgP complexes by molecular replacement with PHASER (McCoy, Grosse-Kunstleve et al. 2007). Alternating cycles of model building with Coot (Emsley, Lohkamp et al. 2010) and refinement with PHENIX (Adams, Afonine et al. 2010) were performed until model completion. Refinement statistics are summarized in Table 2.2.

2.2.8. Analysis of crystallographic structures

The crystallographic models were superposed with Coot (Emsley, Lohkamp et al. 2010) and the secondary structure elements were identified with DSSP (Kabsch and Sander 1983, Touw, Baakman et al. 2015). The interface area between the monomers was determined using PISA (Krissinel and Henrick 2007). The electrostatic potential surface was calculated with APBS (Baker, Sept et al. 2001) using the AMBER force field (Cornell, Cieplak et al. 1995). Figures depicting molecular models were created using PyMOL (Schrödinger).

2.3. Results and discussion

2.3.1. Biophysical analysis and crystallization of *Mtu*GpgP

Recombinant *Mycobacterium tuberculosis* GpgP (*Mtu*GpgP) containing a C-terminal hexahistidine tag was produced in *E. coli* and purified to homogeneity by a combination of immobilized metal-affinity and size-exclusion chromatography. The purity of *Mtu*GpgP was assessed by analytical size-exclusion chromatography and SDS-PAGE (Figure 2.1A and B). In SDS-PAGE *Mtu*GpgP migrated with an apparent MW of 25 kDa, which is consistent with the theoretical MW of 25.7 kDa. In the analytic size-exclusion chromatography, *Mtu*GpgP eluted with an apparent MW of 53.2 kDa, almost twice the theoretical MW of monomeric *Mtu*GpgP, suggesting that a dimer may form in solution.

The thermostability of *Mtu*GpgP was evaluated by differential scanning fluorimetry (Figure 2.1C). *Mtu*GpgP displays a typical denaturation curve with a clear and sharp temperature transition, allowing the straightforward estimation of a melting temperature (T_m) of 44ºC in 50 mM BTP pH 7.0, 50 mM NaCl, 10 mM DTT. As previously proposed, the T_m of a molecule can be used to estimate its crystallization propensity (Dupeux, Rower et al. 2011). In agreement, the T_m of *Mtu*GpgP suggests that it is likely to crystallize at 293K. Thus, initial crystallization experiments were performed at this temperature and high quality crystals were obtained from two different crystallization conditions. The first crystallization condition yielding crystals used PEG 3,350 as precipitant (Figure 2.1D). The second crystallization condition contained sodium iodide and used PEG 6,000 as precipitant.

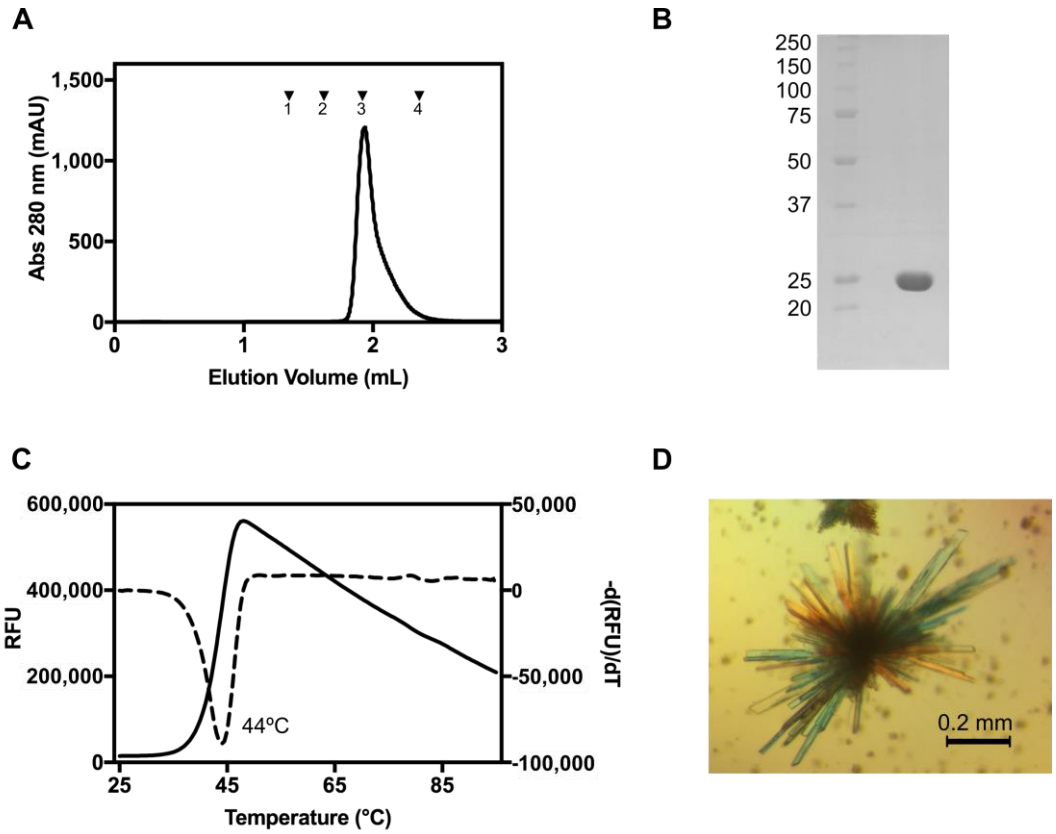


Figure 2.1 – Biophysical characterization and crystallization of recombinant *Mtu*GpgP. A – Analytical size-exclusion chromatogram. The standards (1, Ferritin; 2, Aldolase; 3, Ovalbumin; 4, Ribonuclease A) used for column calibration are indicated by downward black triangles. B – SDS-PAGE analysis of purified *Mtu*GpgP. The MW of the standards (lane 1) is indicated in kDa to the left of the gel. *Mtu*GpgP (lane 2) displays an apparent MW of 25 kDa. C – Melting curve of *Mtu*GpgP (solid line) obtained by differential scanning fluorimetry and its first derivative curve (dashed line). *Mtu*GpgP displays a melting temperature of 44°C in the assay conditions. D – Monoclinic crystals of recombinant *Mtu*GpgP belonging to space group C2.

2.3.2. Overall structure of *Mtu*GpgP

Diffraction data were collected from recombinant *Mtu*GpgP crystals at synchrotron radiation sources (Supplemental figure S2.1). Crystals obtained from the iodide-containing condition and belonging to the monoclinic space group $P2_1$ (Supplemental figure S2.2A) were measured with 1.5498 Å radiation to maximize the anomalous signal of iodide (Supplemental figure S2.3). Two data sets were collected to increase the completeness and multiplicity of the data. At this wavelength, the crystals diffracted X-rays to beyond 2.5 Å resolution (Table 2.1). The phase problem was subsequently solved by single-wavelength anomalous diffraction (SAD), using the anomalous signal of iodide with the SAS protocol of the Auto-Rickshaw pipeline (http://www.embl-hamburg.de/Auto-Rickshaw/) (Panjikar, Parthasarathy et al. 2005, Panjikar, Parthasarathy et al. 2009). The input diffraction data was first prepared and converted for use in the Auto-Rickshaw pipeline using programs of the CCP4 software suite (Winn, Ballard et al. 2011). Fifty six heavy atom sites were identified using SHELXD (Sheldrick 2010) with CC_{all} and CC_{weak} of 40.5 and 19.9%, respectively. The correct hand for the substructure was determined using ABS (Hao 2004) and SHELXE (Sheldrick 2010), followed by refinement of the substructure atom occupancy using BP3 (Winn, Ballard et al. 2011). The twofold NCS operator was found using RESOLVE (Terwilliger 2000). Density modification, phase extension and NCS-averaging were performed with DM (Cowtan 1994), resulting in a FOM of 0.713 and real-space-free-residual of 0.131. A partial α-helical model containing 49 α-carbons was produced using HELICAP, an ARP/wARP module for tracing helices and strands (Langer, Cohen et al. 2008). A partial model, containing 279 out of the total 944 residues in the asymmetric unit, was finally obtained with BUCCANEER (Cowtan 2006).

The partial model was subsequently improved with the MRSAD protocol, which combines molecular replacement and SAD phasing. This was performed with PHASER (McCoy, Grosse-Kunstleve et al. 2007), which analysed the heavy atom sites based on the starting model. Heavy atom occupancy was refined with MLPHARE (Otwinowski 1991) and density modification was performed with PIRATE (Cowtan 2000) followed by NCS detection and density modification with RESOLVE (Terwilliger 2000). A step of model building with SHELXE (Sheldrick 2010) was followed by amino acid sequence docking in the SHELXE model using RESOLVE (Terwilliger 2003). The resulting model was refined with REFMAC5 (Murshudov, Vagin et al. 1997) and PHENIX (Adams, Afonine et al. 2010). A final step of model building was performed with BUCCANEER (Cowtan 2006) and the model obtained was refined in two steps with REFMAC5 (Murshudov, Vagin et al.

1997) and PHENIX (Adams, Afonine et al. 2010). These steps of phasing, model building and refinement were performed twice. The resulting model, containing 872 out of the 944 total amino acid residues, displayed a R_{work} of 25.2% and R_{free} of 31.6%, with 86.5% of the residues in most favoured regions and only 0.4% in disallowed regions.

The automatically generated model was subsequently improved with alternating cycles of manual building with Coot (Emsley, Lohkamp et al. 2010) and refinement with PHENIX (Adams, Afonine et al. 2010). The asymmetric unit contained four *Mtu*GpgP monomers (named A, B, C and D), corresponding to a Mathews coefficient of 2.18 $Å^3Da^{-1}$ and a solvent content of 43.6% (Table 2.1). These four molecules were modelled from Ala3 to Ser219 (molecule A) or Ala 216 (molecule D), and from Arg2 to Ala 216 (molecule B) or Val218 (molecule C). The polypeptide segment from residues 196 to 202 was not entirely modelled for any of the molecules due to the absence of interpretable electron density (see Table 2.2 for details).

Higher resolution data (1.45 Å) were collected from a single cryo-protected crystal, belonging to space group C2 (Supplemental figure S2.2B). The phase problem was solved by molecular replacement with PHENIX using the coordinates of the lower-resolution structure as search model. Two molecules were located in the asymmetric unit, corresponding to a Matthews coefficient of 2.13 $Å^3Da^{-1}$ and a solvent content of 42.4% (Table 2.1). The high-resolution structure was improved with alternating cycles of manual building with Coot (Emsley, Lohkamp et al. 2010) and refinement with PHENIX (Adams, Afonine et al. 2010). The molecules present in the asymmetric unit, named molecule A and molecule B, were modelled from Met1 to His234, and from Met1 to His233, respectively (see Table 2.2 for details).

The *Mtb*GpgP monomer displays an elongated shape and is composed of 6 β-strands (β1 to β6) arranged into a single twisted β-sheet flanked by 11 α-helices (α1 to α11) (Figure 2.2). The recombinant *Mtu*GpgP contains an extra helix (α12) induced by the engineered C-terminal KLAAALEHHHHHH sequence, which includes a hexahistidine tag preceded by a linker to the native *Mtu*GpgP sequence. The strands β1, β2, β3, β4 and β6 run parallel, whereas β5 runs antiparallel. An uncommon ψ-loop from amino acid residue 158 to residue 189, involving strands β1, β4 and β5, is also present.

Table 2.1. Data collection and processing statistics[a].

Crystal	*Mtu*GpgP iodide	*Mtu*GpgP	*Mtu*GpgP vanadate	*Mtu*GpgP vanadate-GG
Synchrotron Radiation Facility	ESRF	ESRF	SOLEIL	ESRF
Beamline	ID29	ID29	PROXIMA 1	ID23-1
Wavelength (Å)	1.54980	0.97718	0.97934	0.97934
Frames	3,600 + 3,600	2,400	1,200	2,600
Rotation (°)	0.1	0.1	0.2	0.1
Space Group	$P2_1$	C2	C2	C2
Unit Cell Dimensions	a=97.0 Å; b=46.1 Å; c=103.2 Å; β=104.3°	a=99.0 Å; b=48.8 Å; c=93.48 Å; β=104.3°	a=98.7 Å; b=49.0 Å; c=93.6 Å; β=104.0°	a=98.8 Å, b=48.9 Å; c= 93.4 Å; β=103.8°
Resolution range (Å)	50.0 – 2.50 (2.60 – 2.50)	47.5 – 1.45 (1.47 – 1.45)	47.9 – 1.49 (1.58 – 1.49)	47.3 – 1.54 (1.59 – 1.54)
Reflections (measured/unique)	336,664/27,533	332,291/75,999	300,810/69,647	306,935/63,815
Completeness (%)	88.7 (49.8)	99.0 (98.9)	98.8 (94.1)	99.3 (98.0)
Multiplicity	12.2 (7.8)	4.4 (4.4)	4.3 (4.2)	4.8 (4.8)
Anomalous completeness	87.7 (46.3)	-	-	-
Anomalous multiplicity	6.3 (4.1)	-	-	-
R_{merge}[b]	0.085 (0.433)	0.082 (1.427)	0.068 (0.900)	0.084 (0.847)
R_{meas}[c]	0.089 (0.464)	0.093 (1.614)	0.077 (1.026)	0.094 (0.947)
$R_{p.i.m.}$[d]	0.025 (0.162)	0.043 (0.739)	0.036 (0.484)	0.041 (0.414)
Half-set correlation $CC_{1/2}$	0.999 (0.911)	0.998 (0.506)	0.998 (0.543)	0.998 (0.624)
Half-set anomalous correlation CC_{Anom}	0.605 (-0.007)			
Mean (I)/σ (I)	26.1 (4.3)	8.4 (1.1)	10.2 (1.3)	11.9 (1.8)
Protomers per asymmetric unit	4	2	2	2
Matthews coefficient ($Å^3Da^{-1}$)	2.18	2.13	2.14	2.13
Solvent content (%)	43.6	42.4	42.6	42.4
Average mosaicity (º)	0.10	0.13	0.11	0.15

[a] Values in parenthesis correspond to the outermost resolution shell.

[b] $R_{merge} = \sum_{hkl} \sum_i \left| I_{i(hkl)} - \langle I_{hkl} \rangle \right| / \sum_{hkl} \sum_i I_{i(hkl)}$, where $I_{i(hkl)}$ is the observed intensity and $\langle I_{hkl} \rangle$ is the average intensity of multiple observations of symmetry-related reflections.

[c] $R_{meas} = \sum_{hkl} \sqrt{\frac{n}{n-1}} \sum_{i=1}^{n} \left| I_{i(hkl)} - \langle I_{hkl} \rangle \right| / \sum_{hkl} \sum_i I_{i(hkl)}$, where $I_{i(hkl)}$ is the observed intensity and $\langle I_{hkl} \rangle$ is the average intensity of multiple observations of symmetry-related reflections.

[d] $R_{p.i.m.} = \sum_{hkl} \sqrt{\frac{1}{n-1}} \sum_{i=1}^{n} \left| I_{i(hkl)} - \langle I_{hkl} \rangle \right| / \sum_{hkl} \sum_i I_{i(hkl)}$, where $I_{i(hkl)}$ is the observed intensity and $\langle I_{hkl} \rangle$ is the average intensity of multiple observations of symmetry-related reflections.

Table 2.2. Refinement statistics and structure solution.

Crystal	*Mtu*GpgP iodide	*Mtu*GpgP	*Mtu*GpgP vanadate	*Mtu*GpgP vanadate-GG
Resolution range (Å)	50.0 – 2.50	47.5 – 1.45	47.9 – 1.50	47.3 – 1.54
R_{factor}/Free R_{factor} (%)	19.0/23.2	17.5/20.0	16.8/20.0	16.7/19.9
Unique reflections (working/test set)	52,752/2,532	75,971/3,676	69,633/3,430	63,810/3,223
Completeness (working/test) (%)	88.0/4.8	98.9/4.8	98.8/4.9	99.3/5.1
Wilson B-factor (Å^2)	38.2	19.6	20.9	16.8
Total number of atoms	6,761	4,179	4,212	4,338
B-factor (Å^2)				
Overall	48.2	32.6	33.0	26.6
Protein	48.3	31.4	32.2	25.0
Ligand	75.8	52.3	22.3	28.2
Water	42.3	41.4	41.7	36.6
RMSD				
Bond lengths (Å)	0.009	0.009	0.007	0.009
Bonds angles (°)	1.244	1.029	0.845	0.876
DPI	0.312	0.068	0.074	0.080
Ramachandran				
Favoured (%)	94.9	97.0	97.6	97.4
Outliers (%)	1.57	0.86	0.65	0.43
Polypeptide length	A: 234 (Ala3 – Ser219) B: 232 (Arg2 – Ala216) C: 231 (Arg2 – Val218) D: 229 (Ala3 – Ala216)	A: 234 (Met1 – His234) B: 233 (Met1 – His233)	A: 231 (Met1 – His231) B: 235 (Met1 – His235)	A: 231 (Met1 – His231) B: 236 (Met1 – His236)
Polypeptide gaps	4 (A, B: Trp196 – Ala198; C: Ala197 – Phe202; D: Trp196 – Ser200)	0	0	0
Water molecules	175	472	438	586
Ligands[a]	Iodide ion		VO4, TVG1, TVG2	VO4, TVG1, TVG2, 9WN

[a] VO4, 9WN, TVG1 and TVG2 correspond to orthovanadate, GG, and trivanadate glycerol 1 and 2, respectively.

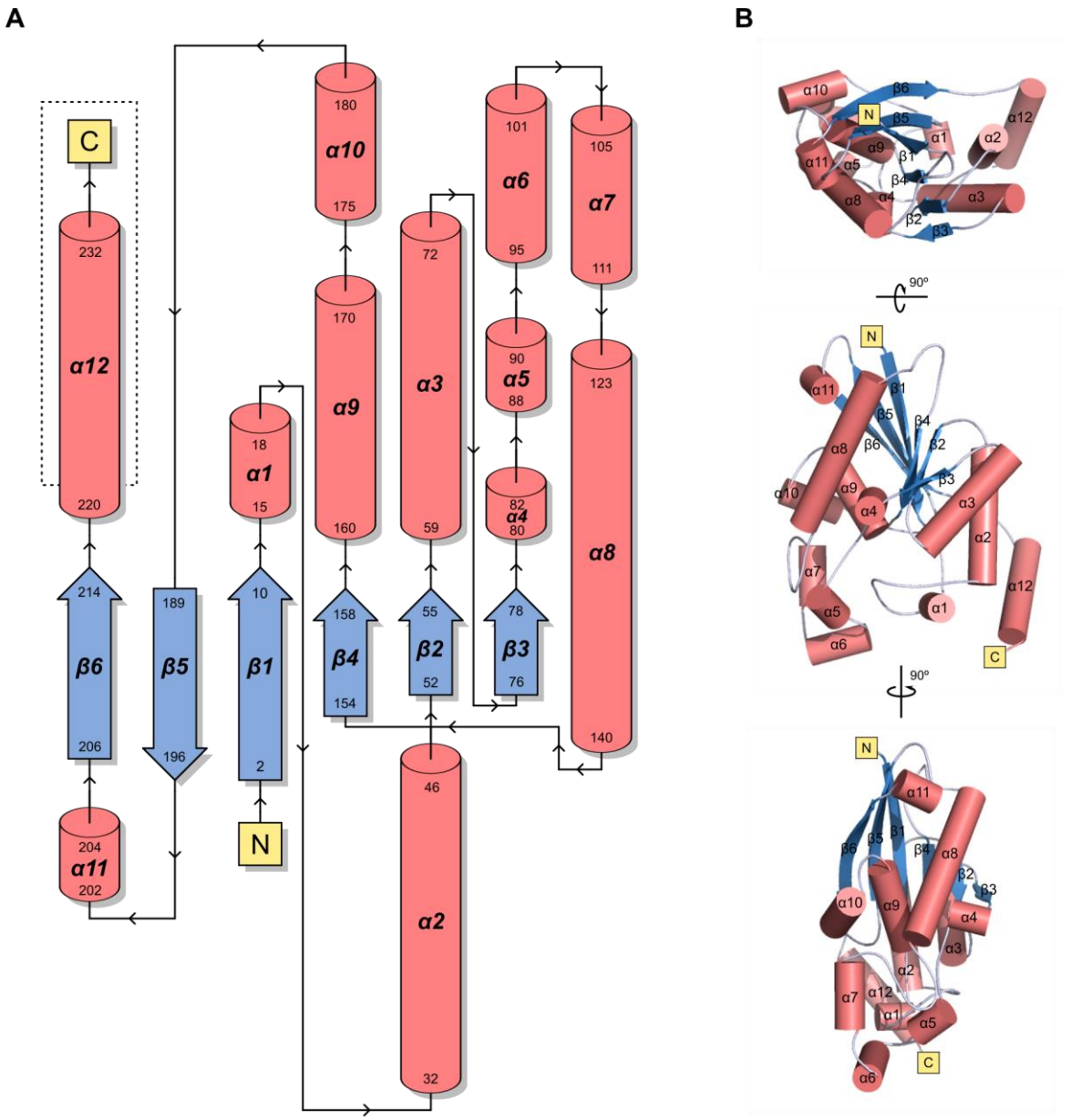


Figure 2.2 – Structural arrangement of *Mtu*GpgP. A – Topology diagram of recombinant *Mtu*GpgP. Secondary structural elements are coloured salmon (α-helices, from α1 to α12) and blue (β-strands, from β1 to β6) and their boundaries are indicated. C and N-termini are highlighted in yellow. The dashed box indicates the secondary structure induced by the engineered C-terminal linker and hexahistidine tag. B – Tertiary structure of *Mtu*GpgP coloured as in A.

2.3.3. Quaternary structure of *Mtu*GpgP

The quaternary structure of *Mtu*GpgP was initially evaluated by analytical size-exclusion chromatography. Since *Mtu*GpgP eluted with an apparent MW twice the theoretical value, a dimer was expected to form in solution (see section 2.3.1).

In agreement, the crystallographic structure of *Mtu*GpgP was analysed with PISA (Krissinel and Henrick 2007), which allows the evaluation of the molecular interfaces occurring in the asymmetric unit and between the symmetry-related molecules in the crystal. The largest molecular interface area was found for the two molecules present in the asymmetric unit. Molecules A and B share a total interface area of 904 Å^2, which corresponds to ~8% of the total surface area of each monomer (Table 2.3). This interface contains mainly β6-β6 and β6-α10 interactions (Figure 2.3), predominantly of hydrophobic character (Table 2.3). This dimer composed of molecules A and B is probably that formed in solution (Figure 2.3).

Table 2.3 – Analysis of the quaternary structure of *Mtu*GpgP using PISA (Krissinel and Henrick 2007).

		Total surface area (Å^2)	Total interface area (Å^2)	Interaction (No.)		
				Hydrogen bonds	Salt bridges	Non-bonded contacts
Monomer	A	10,996.2				
	B	10,974.9				
Interface	A:B		904.0	13	1	103

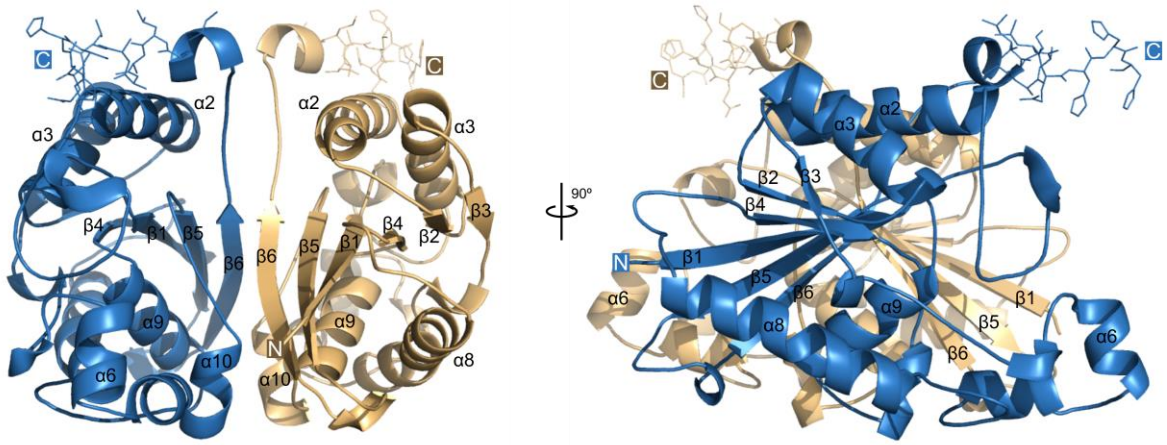


Figure 2.3 – Three-dimensional structure of recombinant *Mtu*GpgP. Homodimeric *Mtu*GpgP is shown in two different orientations. Monomers are coloured wheat and blue. The α-helices and β-sheets are labelled. The engineered C-terminal peptide linker and partially modelled tag are represented as sticks. Strand β6 and helix α10 are localized at the interface between the monomers.

Homodimeric *Mtu*GpgP displays twofold symmetry with the active sites localized at opposite positively charged clefts (Figure 2.4). Each cleft is shaped by both monomers and lined by the side chain of arginine and histidine amino acid residues, resulting in an overall basic character. The complementary contribution of the monomers for the active site clefts suggests that the dimer might be the minimal functional unit of *Mtu*GpgP. Indeed, the inability of the individual monomers to perform the dephosphorylation of the synthetic substrate *p*-nitrophenyl phosphate (PNPP) was reported previously (Zheng, Jiang et al. 2014).

The active site cleft of molecule B is larger than that of molecule A, which probably results from crystal packing constrains. Indeed, the segment Trp196-Ile205 (Figure 2.4B) of molecule A is stabilised by a symmetry-related molecule, forcing a superopen state of the active site cleft of molecule B, whereas in molecule B this segment displays higher mobility (Supplemental figure S2.4). The superopen state of the active site cleft of molecule B is also accompanied by the loss of an hydrogen bond between the carbonyl group of Leu181 and the side chain of Arg123, the position of which may be important for substrate binding (see section 2.3.4).

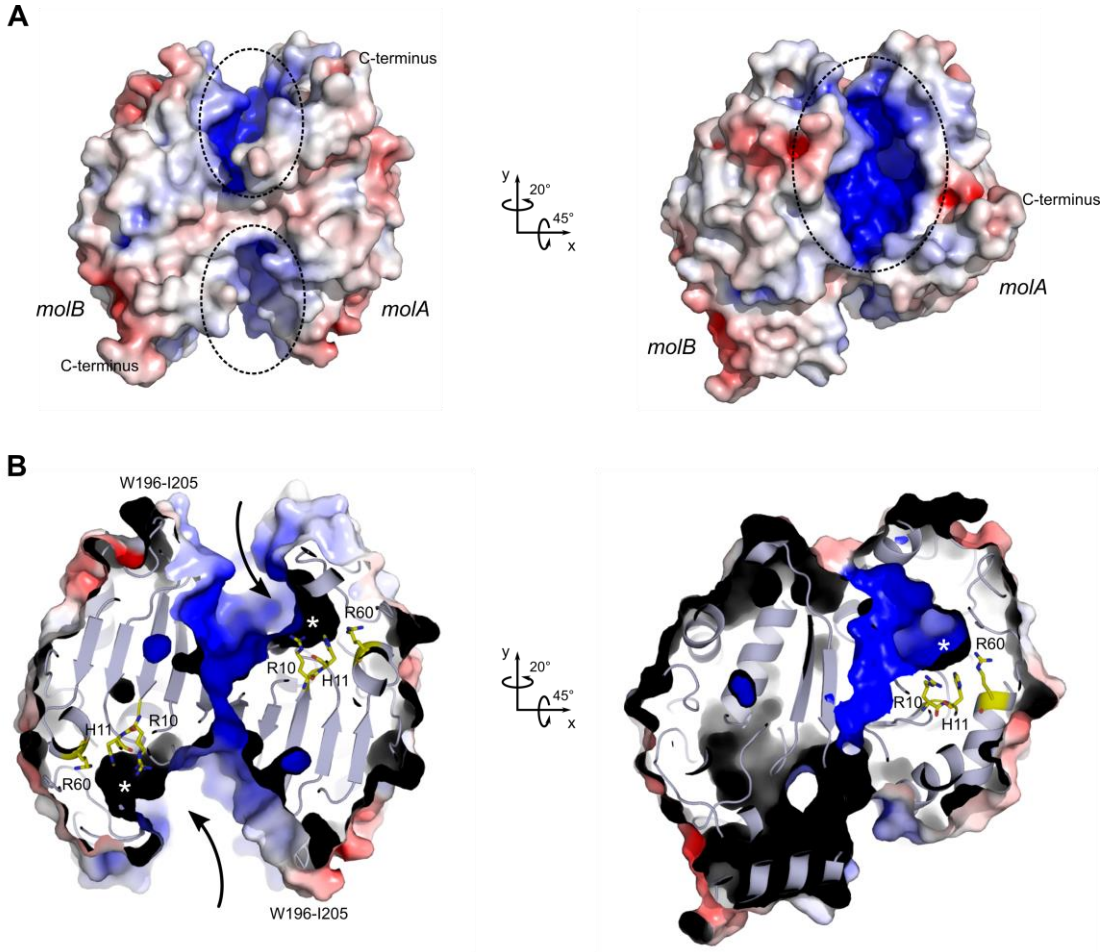


Figure 2.4 – Electrostatic potential of *Mtu*GpgP surface. A – Solid surface representation of dimeric *Mtu*GpgP coloured according to the electrostatic surface potential [contoured from -8kT/e (red) to 8kT/e (blue)], highlighting the positively charged active sites (dashed ellipses) and the position of the engineered C-terminus. B – Cross section of the *Mtu*GpgP homodimer oriented as in A. The catalytic His11 (yellow sticks) is localized at the bottom of the active site cleft (marked with an asterisk) flanked by Arg10 and Arg60 (yellow sticks). The entrance to the active sites is indicated by the arrows. The molecular surface was coloured by electrostatic potential as in A.

2.3.4. *Mtu*GpgP in complex with vanadate

In order to understand the molecular details of substrate binding, recombinant *Mtu*GpgP was crystallized in the presence of sodium orthovanadate. Vanadate is a phosphate analogue that mimics the transition-state of the phosphoryl group transfer, by forming a stable complex with the catalytic residue. Here, sodium orthovanadate was used to induce the formation of a stable transition-state that could allow the identification of the residues involved in phosphoryl group stabilisation and transfer.

Several crystals grown in the presence of sodium orthovanadate were measured and analysed. The active site of *Mtu*GpgP crystallized with orthovanadate is occupied by a large residual electron density, over 6 Å long. The anomalous difference electron density map, calculated from data collected at ~0.98 Å, reveals three positive difference peaks compatible with vanadium atoms (Figure 2.5A). Two peaks are localized close to the side chains of His11 and Glu84, suggesting the presence of covalent bonds to these residues (see distances in Figure 2.6D). The third peak is equally distant from the other two (~3Å), suggesting the existence of bonds between them. The superposition of the anomalous difference map with the 2Fo-Fc electron density map shows that the vanadium atoms are flanked by other atoms, which are likely the oxygens of orthovanadate (Figure 2.5A).

The anomalous difference peaks assigned to vanadium atoms display distinct heights: the histidine-bound vanadium corresponds to the highest peak, with the other two having identically lower heights (Figure 2.5A). This may arise from different occupancies of the vanadium atoms, suggesting that different species of vanadate may simultaneously exist in the active site of different subpopulations of molecules.

One of the chemical species compatible with the electron density map is a penta-coordinated vanadate (VO4) bound to His11 Nε2 (Figure 2.5B). The interatomic distances within VO4 suggest that a double bond may exist between V and O1 or O2 (Figure 2.6A and D). When bound to histidine, this oxometalate displays trigonal bipyramidal geometry and is a well-known mimic of the much less stable phophohistidine intermediate (Davies, Staker et al. 2011, Crans, Tarlton et al. 2014). The VO4-histidine complex is stabilised by a complex network of hydrogen bonds with the side chains of Arg10, Asn17, Gln23, Arg60, Glu84 and His159, and backbone nitrogen of Gly160 (Figure 2.5B). These hydrogen bonds are likely involved in orientation and stabilisation of the substrate phosphoryl group during the phosphorylation and dephosphorylation of the catalytic histidine.

The remaining electron density at the active site of *Mtu*GpgP can be explained by the presence of a trivanadate glycerol ester. Vanadate may occur as a polymeric structure and its cleavage into monomeric species is usually referred as activation (Gordon 1991) since it increases vanadate ability to inhibit phosphatases. However, vanadate activation is not an irreversible reaction and vanadate can shift back to oligomeric forms, a process especially favoured at neutral to acidic pH (Davies and Hol 2004). The presence of a trimeric form of vanadate at the active site reveals its tendency to oligomerize in the slightly acidic crystallization condition. Also, trivanadate displays higher occupancy compared to VO4 (~70%), which suggests a relatively higher ability to outcompete for the

active site compared to all the other vanadate species that may be present in the solution, including the monomeric species VO4.

Moreover, vanadate esters can form readily in aqueous solution (Crans, Smee et al. 2004). The trivanadate glycerol ester species proposed to be occupying the active site resulted from the reaction between vanadate and glycerol from the crystal cryo-protectant. Indeed, the derivatization of vanadate by glycerol is not without precedent and similar reactions were observed previously (Rath, Rajak et al. 1998, Davies, Interthal et al. 2002, Davies, Staker et al. 2011, Moise, Gallup et al. 2015). In the trivanadate glycerol ester (TVG1) molecule, the O1 and O3 atoms of glycerol (GOL) are bound to the V3 and V2 atoms of trivanadate, respectively (Figure 2.6B). The interatomic distances suggest that double bonding is likely to occur between V1 and O1 or O2, V2 and O6, and V3 and O7 (Figure 2.6B and D). In the active site, TVG1 V1 and V3 are covalently bound to His11 Nε2 and Glu84 Oε, respectively (Figure 2.5C). All vanadium atoms are penta-coordinated in a trigonal bipyramidal geometry. The interactions with other active site residues are comparable to those observed for VO4, except for new water-mediated contacts with the side chain of Asn17 and the carbonyl group of Arg21.

Finally, there is an electron density feature in the vicinity of the TVG1 that suggests the presence of a third molecular species at the active site (Figure 2.5C). This third species is identical to the previously described trivanadate glycerol ester (TVG1), except for that GOL O3 is not bound to V2, being replaced by an additional oxygen atom (Figure 2.5D and Figure 2.6C). This third species is here referred as TVG2.

The glycerol moiety of TVG1 and TVG2 establishes additionally hydrophobic contacts with Leu87, Met22, and Trp109. During substrate dephosphorylation, these residues are likely to be involved in substrate orientation by creating a hydrophobic surface where the glucose ring may lay and be further stabilised by polar contacts with Arg123, His95 and Arg110 (Figure 2.5D).

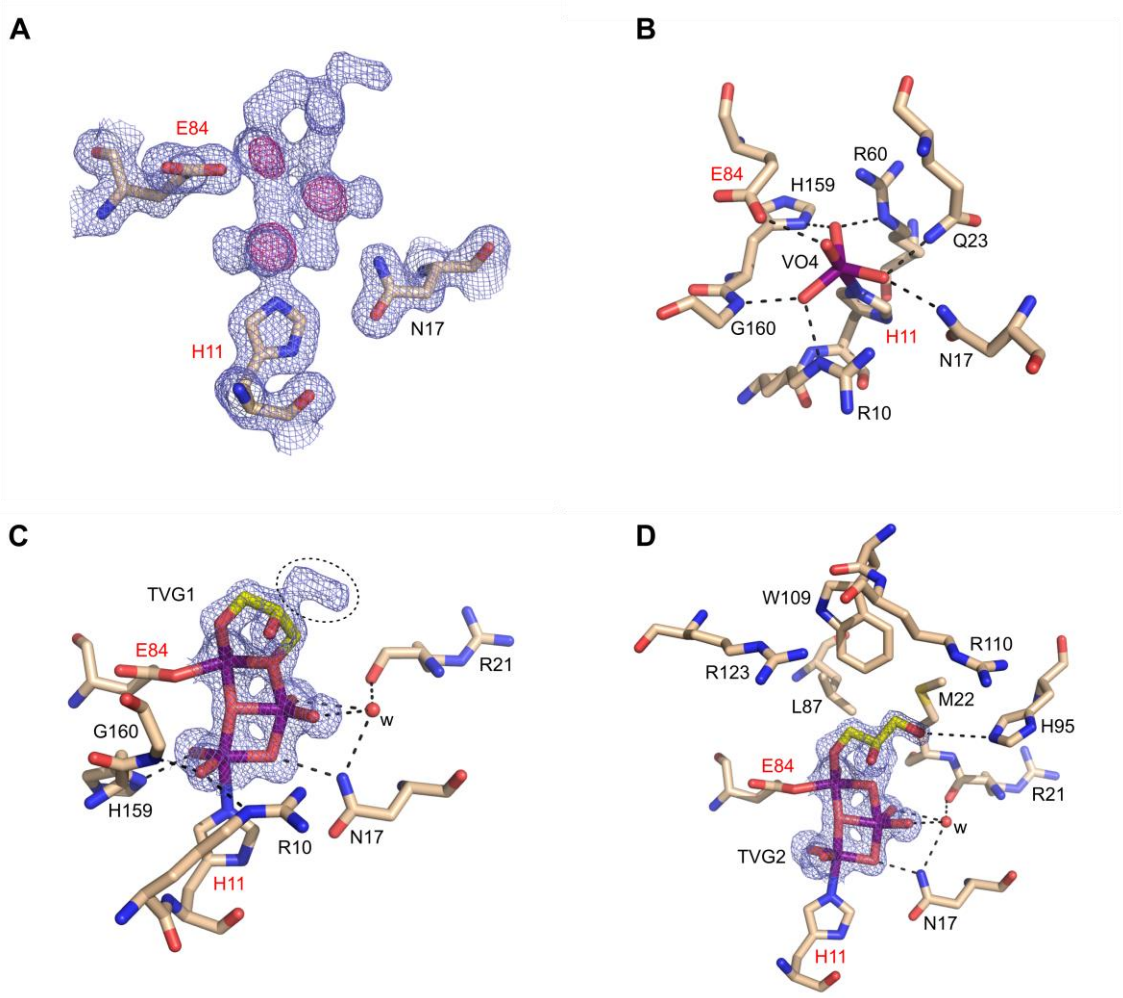
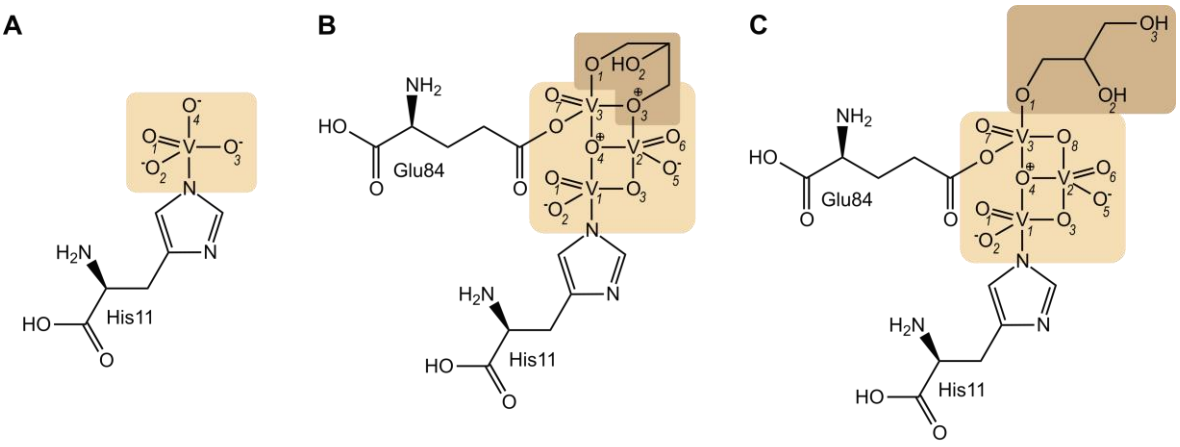


Figure 2.5 – Active site content of *Mtu*GpgP crystallized in the presence of sodium orthovanadate. A – The 2Fo-Fc electron density map (blue) contoured at 1 σ and the anomalous difference map (pink) contoured at 5 σ highlight the three anomalous difference peaks assigned to vanadium atoms. Catalytic residues are labelled in red. B – Penta-coordinated orthovanadate (VO4) covalently bound to His11 and stabilised by a complex network of hydrogen bonds (dashed lines). C – Trivanadate glycerol (TVG1) covalently bound to His11 and Glu84 (labelled in red). Vanadium, oxygen and carbon atoms of TVG1 are in purple, red and yellow, respectively. Hydrogen bonds are represented by dashed lines. The 2Fo-Fc electron density map is contoured as in A. The dotted ellipse highlights the electron density feature that suggests the presence of the species in D. D – Trivanadate glycerol (TVG2) covalently bound to His11 and Glu84 (labelled in red). The glycerol moiety is oriented by hydrophobic contacts with Met22, Leu87 and Trp109, and stabilised by a polar contact with His95. The atoms of TVG2 are coloured as in C. Catalytic residues are labelled as in A. The 2Fo-Fc electron density map is contoured as in A.



D

Atom 1	Atom 2	Distance (Å)	Bond type[a]
VO4			
V	O1	1.87	Single or double[b]
V	O2	1.87	Single or double[b]
V	O3	1.94	Single
V	O4	1.93	Single
V	Nε2 (His11)	2.22	Single
TVG1 or TVG2			
V1	O1	1.87	Single or double[b]
V1	O2	1.87	Single or double[b]
V1	O3	1.94	Single
V1	O4	1.93	Single
V1	Nε2 (His11)	2.22	Single
V2	O3	1.95	Single
V2	O4	2.08	Single
V2	O5	1.82	Single
V2	O6	1.74	Double
V2	O3 (GOL)	2.14	Single
V3	O4	1.83	Single

Figure 2.6 – Vanadate species present in the active site of *Mtu*GpgP. A – Orthovanadate (VO4) covalently bound to His11 and displaying a trigonal bipyramidal geometry. B and C – Trivanadate glycerol ester species (B – TVG1; C – TVG2), obtained from the reaction between vanadate (wheat) and glycerol (brown), and covalently bound to His11 and Glu84. All vanadium atoms are penta-coordinated with trigonal bipyramidal geometry. D – Interatomic distances in VO4 and TVG molecules. The bond type was assigned based on interatomic distances ([a]). The vanadium-oxygen double bond in histidine-bonded vanadate can occur with atom O1 or O2 ([b]).

In a previous work, the crystallographic structure of *Bacillus stearothermophilus* phosphatase (PhoE) in complex with trivanadate was reported (Rigden, Littlejohn et al. 2003). This phosphatase displays 34% amino acid sequence identity with GpgP from *M. tuberculosis* and they share a conserved active site. As for *Mtu*GpgP, the trivanadate present in the active site of PhoE is covalently bound to the catalytic residues His10 and

Glu83 (PhoE sequence numbering). However, trivanadate was modelled as V_3O_8 (ligand ID VN3), instead of $V_3O_9^{2-}$, with tetra-coordinated vanadium atoms.

More recently, the structure of a recombinant *Mtb*GpgP in complex with vanadate was reported (Zheng, Jiang et al. 2014). Bound vanadate was modelled as metavanadate (VO_3^{2-}, ligand ID VO3) covalently bound to the catalytic histidine, and displaying a pyramidal geometry. There is also a coordinating water molecule opposed to the histidine, suggesting penta-coordination of the vanadium. Moreover, the V-O interatomic distances are significantly shorter (~1.60 Å) than the distances to His Nε2 (~2.00 Å) or the water molecule (~2.25 Å). Given the covalent link to histidine, smaller differences in the interatomic distances would be expected.

The structures of PhoE and *Mtb*GpgP were determined at 2.0 and 2.3 Å resolution, respectively (Rigden, Littlejohn et al. 2003, Zheng, Jiang et al. 2014). The electron density maps (2Fo-Fc and Fo-Fc) of PhoE generated automatically in Coot (Emsley, Lohkamp et al. 2010) from PDB entry 1H2F and those of the *Mtb*GpgP in complex with vanadate (PDB entry 4QIH) calculated with PHENIX (Adams, Afonine et al. 2010) after conversion of the structure factor file format from the PDB database (Berman, Westbrook et al. 2000) with the online SF-tool (http://sf-tool.wwpdb.org/), revealed difficult to interpret residual electron density, which may have impaired the correct identification of the ligands. The considerably higher resolution (1.49Å) of *Mtu*GpgP-vanadate complex presented here allowed a better interpretation of the active site contents. In consequence, the vanadate-enzyme interaction network can be described in more detail, helping to understand the inhibition of phosphatases by vanadate and also elucidating the molecular details of phosphoryl group binding and transfer through a reaction intermediate mimic.

The currently accepted catalytic mechanism for histidine phosphatases involves primarily the phosphorylation of a histidine residue simultaneously to proton transfer from a glutamate residue to the substrate, followed by histidine dephosphorylation concomitantly to the nucleophilic attack of the phosphoryl group by a water molecule previously activated by the negatively charged glutamate (see chapter 1 section 1.3.1). In the *Mtu*GpgP-vanadate complex, Glu84 is located at a hydrogen-bond distance to the negatively charged O4 of VO4, which suggests that it is likely to be protonated and hydrogen bound to VO4. During the phosphorylation of His11, Glu84 is at an adequate distance to donate a proton directly to the leaving GG. Moreover, the presence of trivanadate glycerol molecules covalently bound to Glu84 suggests that this amino acid residue may have performed a nucleophilic attack directly to the trivanadate glycerol. The presence of orthovanadate and trivanadate glycerol molecules in the active site of

*Mtu*GpgP may highlight the ability of Glu84 to shift between protonated and ionized forms, suggesting that Glu84 may have a fundamental role in the dephosphorylation reaction.

2.3.5. *Mtu*GpgP in complex with vanadate and GG

*Mtu*GpgP was also crystallized in the presence of orthovanadate and of the reaction product, GG. The main goal was to obtain a ternary complex mimicking the enzyme-substrate complex. This approach was inspired on a previous work (Davies, Staker et al. 2011) in which the spontaneous generation of vanadate-3-phosphoglycerate from vanadate and 3-phosphoglycerate was observed. Vanadate-3-phosphoglycerate is an analogue of 2,3-bisphosphoglycerate, the intermediate of the reaction catalysed by the 2,3-bisphosphoglycerate-depent PGM from *Burkholderia pseudomallei*. The vanadate-3-phosphoglycerate was found covalently bound to the catalytic residue *via* vanadium atom, mimicking the 2,3-bisphosphoglycerate-enzyme intermediate. This transition-state mimic allowed the identification of key amino acid residues involves in substrate binding and catalysis.

Although a similar result was expected for *Mtu*GpgP, the substrate analogue glucosyl-3-vanadate-glycerate was not found in the active site of *Mtu*GpgP. Indeed, the active site of *Mtu*GpgP co-crystallized with sodium orthovanadate and GG is occupied by VO4 alternating with TVG molecules as observed for the *Mtu*GpgP-vanadate complex. However, an additional electron density blob compatible with GG at an arginine and histidine rich region is present in each active site cleft (Figure 2.7A and B).

The large electron density blob present at the entrance of the active site of molecule A (Figure 2.7A) is difficult to interpret, and is likely to result from different poses of GG. In molecule B the electron density blob is smaller and somewhat discontinuous suggesting that GG may have lower occupancy in this active site (Figure 2.7B). The different occupancy displayed by GG in the active site of molecule A and B is probably due to crystal packing constrains. Since the active site of molecule B is in a superopen state, the bonding network is probably weaker and thus GG is less frequently observed in this active site. In both cases, the electron density present in each cleft is within hydrogen bond distance of residues from molecules A and B, which highlights the interdependence of the two monomers of *Mtu*GpgP for substrate binding.

A GG molecule was modelled in the active site of molecule B (Figure 2.7B), establishing polar contacts with Arg106 and Arg110 of molecule B, and with Met1, Arg5 and Arg208 of

molecule A. Water mediated contacts are also present, involving the residues His95 of molecule B, and His195 and Asp210 of molecule A (Figure 2.7C).

The electron density of GG and surrounding residues suggests a transient interaction, compatible with the inward movement of the substrate and the outward shuttling of the product from the catalytic pore. Since the glucose moiety seems to be poorly stabilised, the negatively charged phosphoryl group of the substrate likely plays an essential role in substrate binding and correct orientation. The phosphoryl group-containing substrate is likely to be first electrostatically attracted to the positively charged active site cleft. Then, transient polar contacts with the glucose moiety of the substrate may be established, holding the substrate and allowing its movement to the active site. In the active site, a complex network of hydrogen bonds with the phosphoryl group may be finally established, ensuring the correct orientation and stabilisation of the substrate necessary for the reaction.

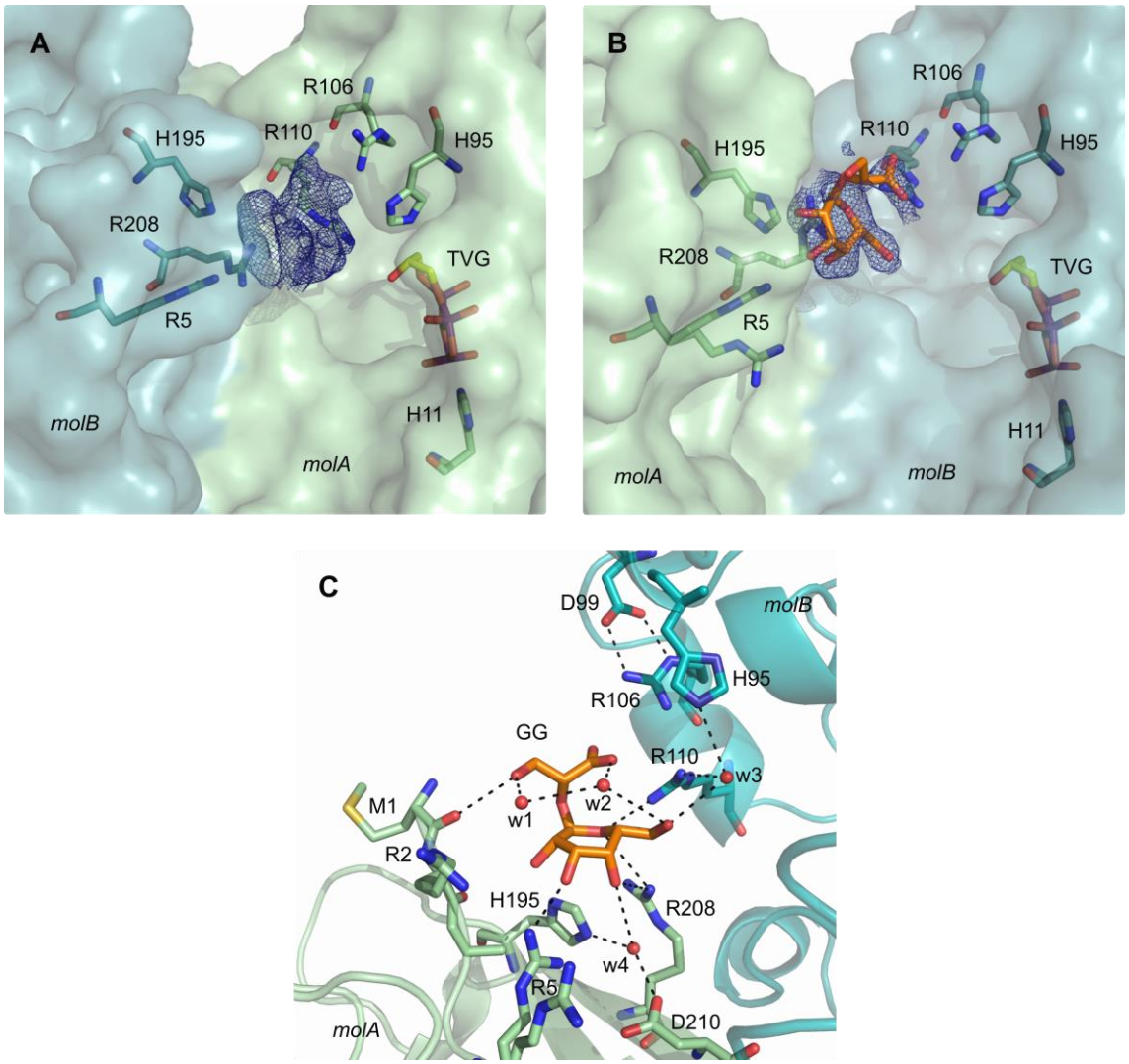
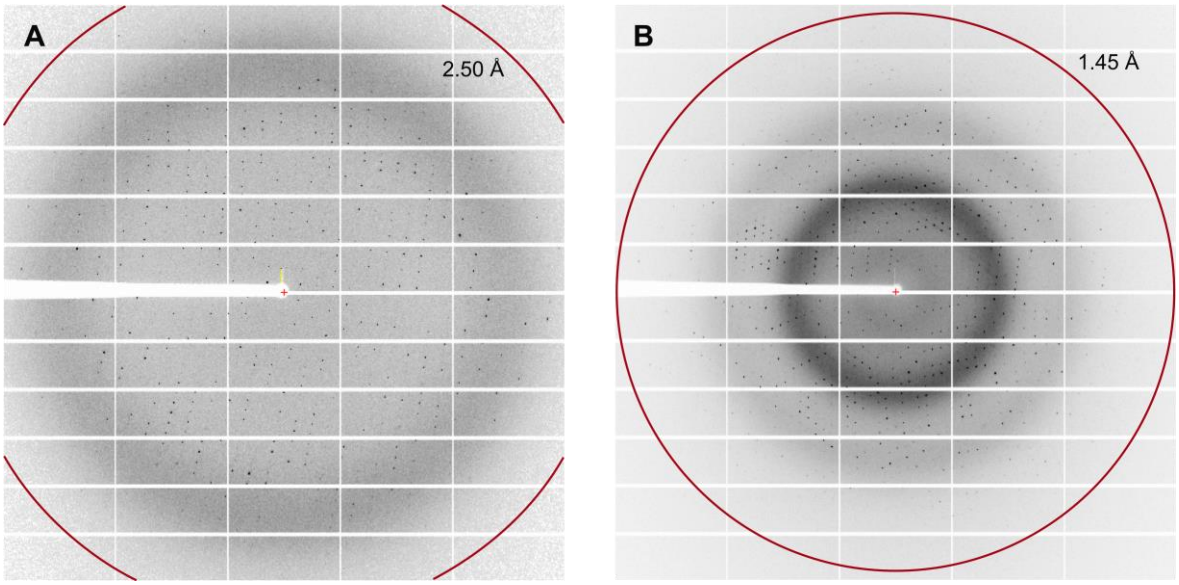
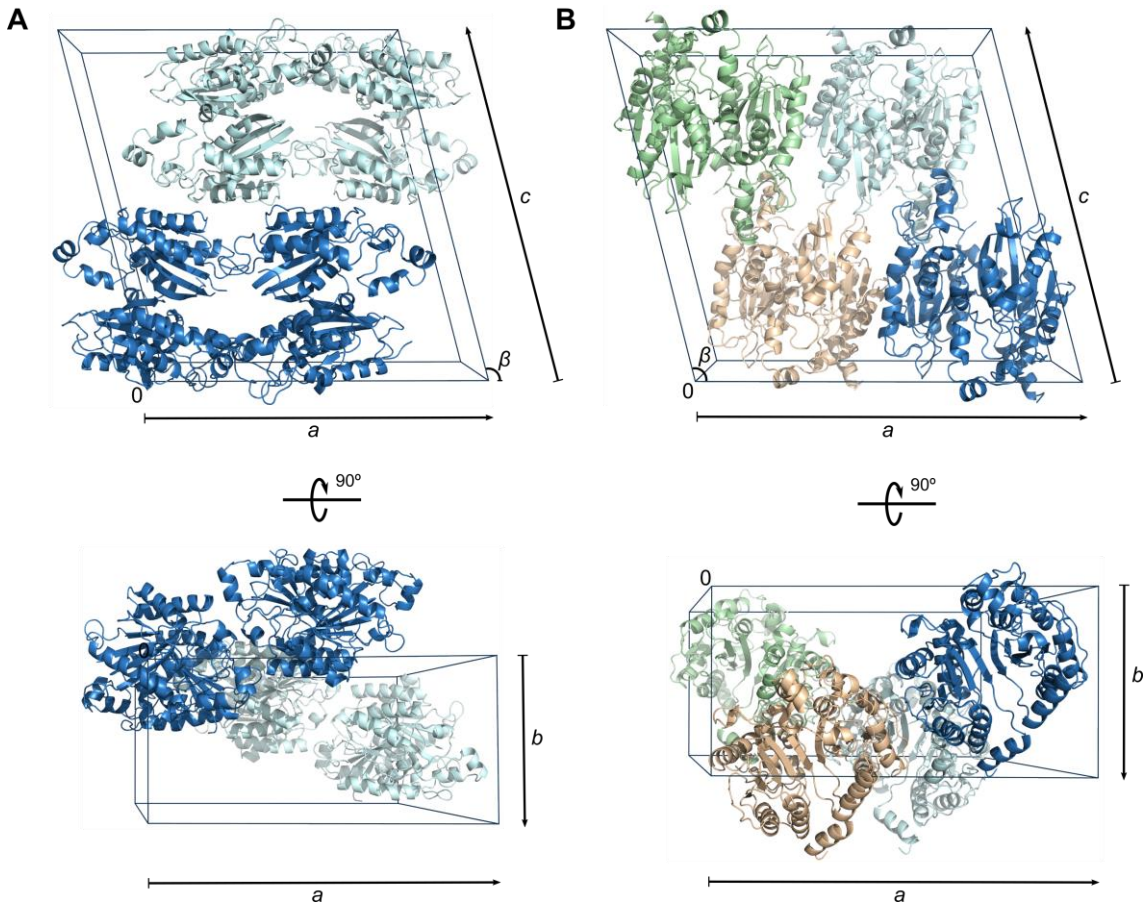


Figure 2.7 – Structure of *Mtu*GpgP in complex with GG. A and B – Active site clefts of the *Mtu*GpgP dimer, highlighting the position of GG (orange) in the active site of molecule B. Molecule A (molA) and molecule B (molB) are coloured pale green and teal, respectively. The electron density 2Fo-Fc map (blue) was contoured at 1 σ. C – Polar contacts established between GG and *Mtu*GpgP residues, revealing the contribution of both monomers for substrate binding. Polar contacts are represented as dashed lines. Water molecules (w) are labelled. Molecules A (molA) and B (molB) are coloured as in A and B.

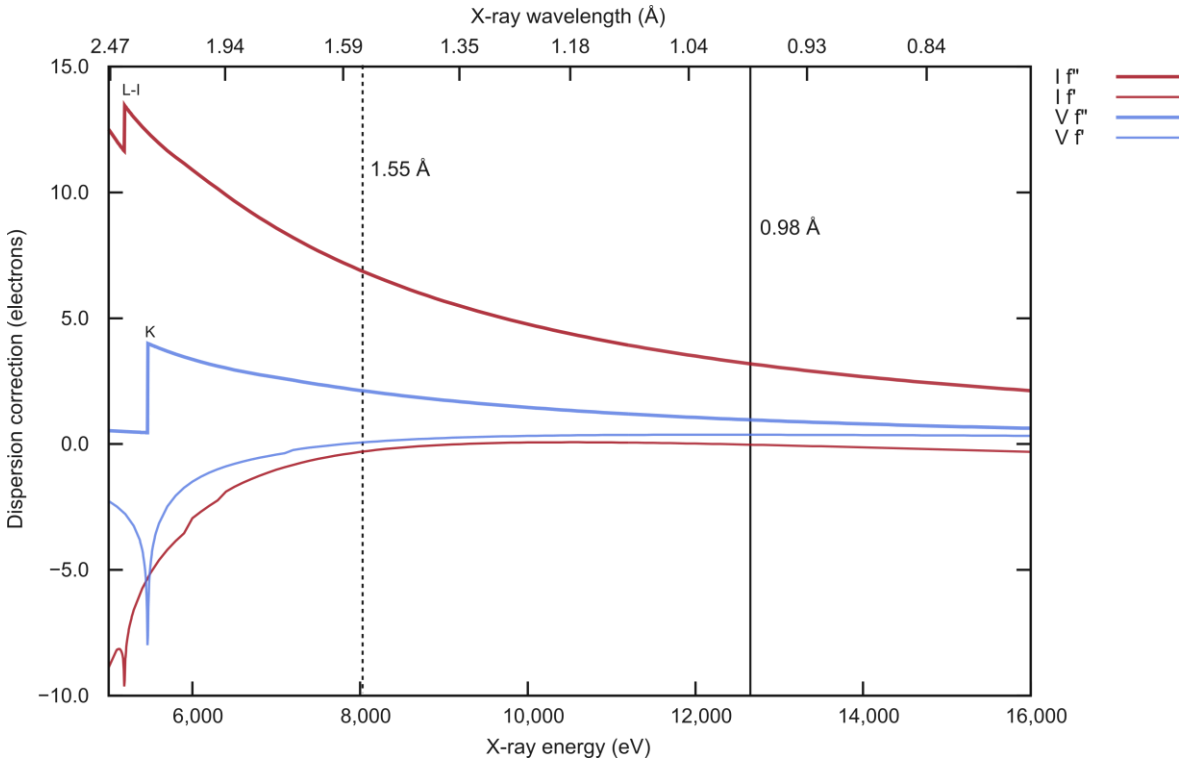
2.4. Supplemental information



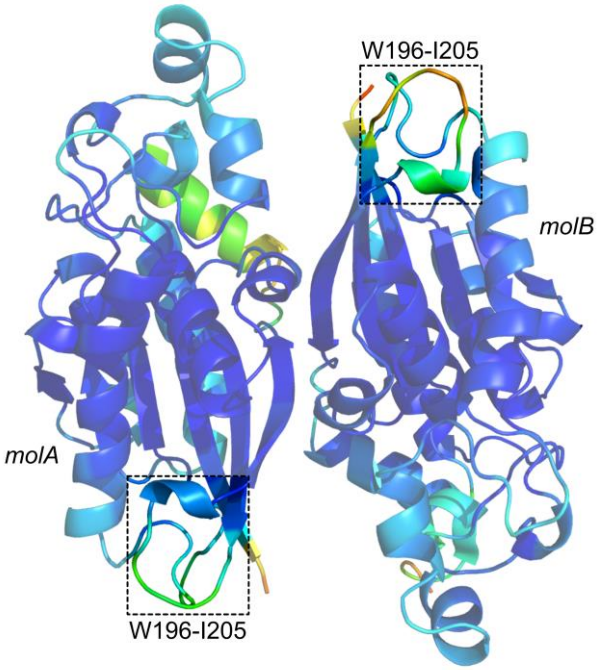
Supplemental figure S2.1 – X-ray diffraction patterns from *Mtu*GpgP crystals. The data from the iodide-containing crystal (A) and from the higher-resolution diffracting crystal (B) were collected at a synchrotron radiation source using 1.55 or 0.98 Å wavelength X-rays, respectively. The diffraction patterns in A and B are the result of merging 20 and 5 consecutive 0.1º rotation images, respectively. The red circles correspond to the resolution limits indicated.



Supplemental figure S2.2 – Unit cell arrangement of *Mtu*GpgP crystals. A – Unit cell content of *Mtu*GpgP crystals belonging to space group $P2_1$ in two different orientations. The molecules of the two asymmetric units are coloured in different shades of blue. The origin of the Cartesian coordinate system is indicated as 0. The unit cell *a*, *b* and *c*-axes are indicated and are ~97, ~46, and ~103 Å in length, respectively. The angle *β* (~104º) is indicated. B – Unit cell content of *Mtu*GpgP belonging to space group C2 in two different orientations. Molecules in the four asymmetric units are coloured blue, pale cyan, pale green and wheat. The origin of the Cartesian coordinate system is indicated as in A. The unit cell *a*, *b* and *c*-axes are indicated and are ~99, ~49, and ~93 Å in length, respectively. The angle *β* (~104º) is indicated.



Supplemental figure S2.3 – Theoretical scattering coefficients of iodide and vanadium as a function of X-ray energy. The theoretical values of f'' (thick line) and f' (thin line) for iodide (red) and vanadium (blue) were traced over a range of energies from 5,004 eV (2.47 Å) to 16,000 eV (0.77 Å). The plot was generated using the online tool X-ray Anomalous Scattering (http://skuld.bmsc.washington.edu/scatter/AS_index.html). The X-ray wavelengths used for data collection from the iodide-containing crystal (dashed line) and from all other *Mtu*GpgP crystals (solid line) are indicated. The theoretical L-I absorption edge of iodide (~5,188 eV, 2.39 Å) and K absorption edge of vanadium (~5,465 eV, 2.27Å) are labelled.



Supplemental figure S2.4 – *Mtu*GpgP three-dimensional structure coloured according to the atomic B-factor value. The segment Trp196-Ile205 of molecule A (molA) and molecule B (molB) displays different values of B-factor due to crystal packing constrains. In molecule A, the segment Trp196-Ile205 is stabilised by the neighbour molecule (symmetry operator -x, y, -z) displaying lower B-factor values, whereas in molecule B higher B-values are observed due to high mobility. Atoms were coloured from minimum to maximum B-factor values with a spectrum of colours from dark blue to red, respectively.

Chapter 3

Structural and biochemical characterization of *M. hassiacum* GgH, a key enzyme in nitrogen starvation recovery

Part of this work was recently published in

Cereija TB, Alarico S, Empadinhas N, Pereira PJB (2017) Production, crystallization and structure determination of a mycobacterial glucosylglycerate hydrolase. *Acta Cryst*. F73, 536-540

3.1. Introduction

In a changing environment the basic requirements for bacterial growth are not always available. When exposed to growth-limiting stress, such as desiccation, temperature and pH variations, oxidative stress, antibiotics, and nutrient limitations, bacterial populations balance between cell death and decreased growth rates (Finkel 2006, Lipworth, Hammond et al. 2016). Some of the surviving cells can slow down or suspend their growth to a viable non-replicating state that can persist for months or years (Corper and Cohn 1933). This process, known as dormancy, allows the viable population size to be maintained during the period of stress (Jones and Lennon 2010). Despite genetically identical to replicating bacteria, dormant cells are more tolerant to external stress, including heat and antibiotics (Dhar and McKinney 2007, Eldar and Elowitz 2010, Balazsi, van Oudenaarden et al. 2011).

The dormancy state requires several structural modifications to maintain cell viability, including the accumulation of carbon stores (Hosaka and Yamashita 1984, Sillje, Paalman et al. 1999, Bourassa and Camilli 2009). Although the type of carbon source varies, the main goal seems to be common: to guarantee a rapidly mobilisable energy source to promote fast cell proliferation when the environmental conditions improve (Shi, Sutter et al. 2010), which is an advantage at the moment of outcompeting neighbouring organisms.

In mycobacteria, the dormancy state seems to be associated to the accumulation of triacylglycerol (Daniel, Deb et al. 2004) and wax esters (Sirakova, Deb et al. 2012). However, it was demonstrated that mycobacteria accumulate GG under nitrogen-limiting conditions (Behrends, Williams et al. 2012, Alarico, Costa et al. 2014), which are able to induce dormancy in *M. smegmatis* (Shleeva, Mukamolova et al. 2004, Anuchin, Mulyukin et al. 2009). *In vitro*, GG prevented the loss of activity for a number of enzymes (Sawangwan, Goedl et al. 2010), suggesting that GG is also likely to contribute for protein stability *in vivo* during a slow-growing or non-replicating phase. In *M. smegmatis* and *M. hassiacum*, intracellular GG accumulated during nitrogen starvation is quickly depleted when nitrogen availability is restored, a reduction associated to the up-regulation of a gene coding for a glucosylglycerate hydrolase (GgH) (Alarico, Costa et al. 2014). Although no data are available on the role of GgH in bacterial survival, dormancy and infection, the high degree of conservation of GgH among rapid-growing mycobacteria and other unrelated phyla strongly suggests that the ability to produce this hydrolase is an evolutive advantage. Moreover, since it is up-regulated upon relief of the growing-limiting stress, GgH is likely to participate in the resuscitation phase of mycobacteria by hydrolysing GG into glucose and glycerate (Alarico, Costa et al. 2014) that can be quickly used for energy

production and for the synthesis of structural molecules necessary for fast bacterial proliferation.

The antibiotics currently in use for fighting mycobacterial infections mainly counter DNA replication or RNA, proteins or cell wall synthesis, indispensable functions for cell growth (Kohanski, Dwyer et al. 2010). Since these processes are almost suppressed in dormant cells, these are more likely to survive treatment with antibiotics. Indeed, dormant cells have been associated to post-treatment relapse and development of genetic resistance (Gomez and McKinney 2004, Levin and Rozen 2006). Targeting essential molecules for bacterial dormancy and resuscitation, such as potentially GgH, may be a viable path for an effective treatment (Lewis 2007, Rittershaus, Baek et al. 2013).

M. hassiacum GgH (*Mha*GgH) belongs to the CAZy GH63 family of glycoside hydrolases (CAZy database, http://www.cazy.org) (Lombard, Golaconda Ramulu et al. 2014), for which only four of more than 1900 assigned members have been structurally characterized (Kurakata, Uechi et al. 2008, Barker and Rose 2013, Miyazaki, Ichikawa et al. 2015). Although *Mha*GgH displays only 36% sequence identity to the recently characterized MgH from *Thermus thermophilus* (Tt8MGH) (Miyazaki, Ichikawa et al. 2015), it catalyses a similar reaction (Alarico, Empadinhas et al. 2013).

In order to unveil its molecular mechanism of action, a thorough structural and functional characterization of *Mha*GgH was performed, elucidating the quaternary architecture of the enzyme and providing an atomic detail view of the determinants of substrate specificity.

3.2. Materials and methods

3.2.1. Cloning and site-directed mutagenesis of the ga*Mha*GgH variants

The gene encoding *Mha*GgH was amplified by PCR from a pET-30a-based construct (Alarico, Costa et al. 2014) with primers 5'-TATACGTCTCACATGCCGCACGACCCGAGTTTCAC-3' and 5'-CGAATTCTTAGCCCAGCCAGTCGAGCACC-3' containing recognition sequences for BsmBI and EcoRI enzymes (underlined), respectively, and cloned into the NcoI/EcoRI restriction sites of pETM11, originating the pETM11-*Mha*GgH expression vector (see Table 3.1 for details).

The point mutants of *Mha*GgH were obtained by site-directed mutagenesis using pETM11-*Mha*GgH as template and primers 5'-CACATGTGGAGTTGGGCCGCCGCGTTC-3' and 5´-GAACGCGGCGGCCCAACTCCACATGTG-3' (originating pETM11-D43A), 5'-GAGTCCGGGATGGCCAACTCG-3' and 5'-CGAGTTGGCCATCCCGGACTC-3' (originating pETM11-D182A), 5'-TCGTTCGCCGCGTACTACGAA-3' and 5'-TTCGTAGTACGCGGCGAACGA-3' (originating pETM11-E419A), and 5'-CGTTGGGCAGCATGCAGTTCTCCTGGACC-3' and 5'-GGTCCAGGAGAACTGCATGCTGCCCAACG-3' (originating pETM11-Q434F).

3.2.2. Expression and purification of *Mha*GgH-His$_6$ and ga*Mha*GgH variants

E. coli BL21 (DE3) cells transformed with pET30a-*Mha*GgH (Alarico, Costa et al. 2014), pETM11-*Mha*GgH, pETM11-D43A, pETM11-D182A, pETM11-E419A or pETM11-Q434F plasmids were grown in LB medium supplemented with 50 μg mL^{-1} kanamycin at 28°C to an OD_{600} of 0.7-1.0. At this point, the temperature was decreased to 25°C and the expression induced with 0.5 mM IPTG. After overnight growth, the cells were harvested, resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM $MgCl_2$, 1 μg mL^{-1} DNAse, 0.3 mg mL^{-1} lysozyme, 1 mM PMSF), and stirred gently on ice for 1 h. The protein extracts were clarified by centrifugation, supplemented with 20 mM imidazole, and loaded onto immobilized metal affinity columns (Agarose Bead Technologies) loaded with nickel and pre-equilibrated in buffer B (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM

imidazole). Bound *Mha*GgH variants were eluted with 100 mM imidazole in buffer B. For histidine tagless *Mha*GgH variants (named ga*Mha*GgH due to the presence of a Gly-Ala dipeptide at the N-terminus of the proteins at the end of the purification procedure), *Mha*GgH-containing fractions were pooled and the hexahistidine tag removed by digestion with tobacco etch virus (TEV) protease (1:5 molar ratio) at 4°C concomitantly to an overnight dialysis against dialysis buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.5 mM EDTA, 1 mM DTT). Untagged ga*Mha*GgH was separated from the hexahistidine tag, tagged *Mha*GgH and TEV protease by a second immobilized-metal affinity chromatography step (see above for experimental conditions). The flowthrough containing ga*Mha*GgH was concentrated on 10 kDa cut-off ultrafiltration devices (Millipore) and loaded onto a HiPrep 26/60 Sephacryl-200 HR (GE Healthcare) size exclusion column using 20 mM Tris-HCl pH 8.0, 400 mM NaCl (storage buffer) as mobile phase. Fractions containing ga*Mha*GgH were pooled and concentrated on a 10 kDa cut-off ultrafiltration device (Millipore). For the hexahistidine-tagged *Mha*GgH variant (*Mha*GgH-His_6), *Mha*GgH-containing fractions from the immobilized-metal affinity chromatography were pooled, concentrated and loaded onto the HiPrep 26/60 Sephacryl-200 HR column using the same procedure as for ga*Mha*GgH variants. The size-exclusion chromatography fractions containing *Mha*GgH-His_6 were pooled and concentrated as described for ga*Mha*GgH variants. The concentration of the purified proteins was estimated by measuring their absorbance at 280 nm. The protein solutions were flash-frozen in liquid nitrogen and kept at -80°C until needed.

3.2.3. Expression and purification of selenomethionine-containing ga*Mha*GgH

E. coli B834 (DE3) cells transformed with pETM11-*Mha*GgH were grown in 25 mL LB medium overnight at 37°C. Cells were collected and washed thrice with sterile deionized water and used to inoculate 1 L SelenoMethionine Medium (Molecular Dimensions). The culture was grown at 30°C until an OD_{600} of 0.7-0.8 before inducing expression with 0.5 mM IPTG. After overnight growth, the cells were harvested, resuspended in lysis buffer, and incubated on ice with gentle agitation for 1 h. The protein extract was clarified by centrifugation, supplemented with 20 mM imidazole and 5 mM β-ME, and loaded onto an immobilized metal affinity column (Agarose Beads Technologies) loaded with nickel and pre-equilibrated with buffer B supplemented with 5 mM β-ME. Bound protein was eluted with 100 mM imidazole and 5 mM β-ME in buffer B. Fractions containing

selenomethionine-labelled ga*Mha*GgH (SeMet-ga*Mha*GgH) were pooled and dialysed in the presence of TEV protease (1:5 molar ratio) against 50 volumes of dialysis buffer for 2 h at room temperature followed by an overnight incubation at 4°C. Untagged SeMet-ga*Mha*GgH was separated from non-cleaved material and TEV protease by a second immobilized-metal affinity step under the same experimental conditions. The flowthrough containing the untagged protein was concentrated to 2.3 mg mL^{-1} in a 30 kDa cutoff ultrafiltration device (Millipore) with concomitant buffer exchange to storage buffer supplemented with 0.5 mM EDTA and 1 mM DTT.

3.2.4. 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. The column was calibrated with blue dextran (2,000 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa) and ribonuclease A (13.7 kDa) standards. The K_{av} *vs* log MW was calculated with the equation $K_{av} = \frac{V_e - V_0}{V_t - V_0}$, 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.

3.2.5. 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 13,000x*g* and 4°C for 20 min, loaded on a ZEN2112 cuvette, and three independent measurements recorded at 20°C. All data were analysed using Zetasizer software 7.11 (Malvern Instruments).

3.2.6. Differential scanning fluorimetry

The melting temperature of all *Mha*GgH variants was determined by a thermal shift (Thermofluor) assay. Each protein sample (0.5 mg mL^{-1} final concentration) was centrifuged at 13,000x*g* and 4°C for 15 min, mixed with 5x SYPRO Orange (Life Technologies) in storage buffer, and loaded in 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). The experiment was performed in triplicate. The melting curves were analysed using the CFX Manager software (Bio-Rad), and the melting temperature determined as the inflexion point of the melting curve.

3.2.7. Chemical synthesis of GG, MG and GGlycolate

The chemical synthesis of glucosylglycerate [GG, 2-*O*-(α-D-glucopyranosyl)-D-glycerate], mannosylglycerate [MG, 2 -*O*-(α-D-mannopyranosyl)-D-glycerate] and glucosylglycolate [GGlycolate, 2-*O*-(α-D-glucopyranosyl)-D-glycolate) were performed by Eva Lourenço and Rita Ventura (Bioorganic Chemistry Group, ITQB, Universidade Nova de Lisboa, Portugal) as previously described (Lourenço et al. 2009; Lourenço and Ventura, 2011).

3.2.8. Substrate specificity of ga*Mha*GgH

The activity of ga*Mha*GgH was evaluated in a 50 µL reaction mixture containing 2.75 µM enzyme, 20 mM GG, 25 mM sodium phosphate pH 6.0, 5 mM $MgCl_2$, 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 stained as previously described (Alarico, Costa et al. 2014). Enzyme specificity was probed with 5 mM and 20 mM MG, GGlycolate or glucosylglycerol [GGlycerol, 2-*O*-(α-D-glucopyranosyl-D-glycerol; Bitop, Germany] as substrates (experimental conditions described above). All reactions were performed at 42°C and 50°C for 1h, 3h and overnight. Control reactions without enzyme were also performed. The reaction products were analysed by TLC as described above, except for the reactions containing GGlycerol in which the solvent system used was a chloroform/methanol/ammonia 25% (30:50:25, v/v/v) mixture. These experiments were performed by Susana Alarico (Molecular Mycobacteriology Group, CNC – Center for Neuroscience and Cell Biology, University of Coimbra, Portugal).

3.2.9. Catalytic activity analysis of ga*Mha*GgH mutants

The activity of ga*Mha*GgH point mutants (D43A, D182A, E419A, Q434F) was evaluated with GG and MG in 50 µL reactions containing 2.75 µM enzyme and 10 or 20 mM GG or MG in 25 mM buffer (sodium phosphate pH 6.0, or sodium acetate pH 4.5 or pH 5.0, or Bis-Tris Propane (BTP) pH 7.0, or Tris-HCl pH 8.0), 5 or 10 mM $MgCl_2$, in presence or absence of 100 mM KCl. Reactions were performed at 37°C, 42°C and 50°C for 1 h and overnight. Reaction mixtures with and without wild-type ga*Mha*GgH were used as controls. Substrate hydrolysis was evaluated by TLC, as described above. The experiments for the evaluation of Q434F mutant activity were performed by Susana Alarico (Molecular Mycobacteriology Group, CNC, University of Coimbra, Portugal).

3.2.10. Biochemical and kinetic parameters of ga*Mha*GgH

The effect of temperature, pH, metallic cations and salt concentration on ga*Mha*GgH activity and its half-life were evaluated by quantifying the glucose released upon GG hydrolysis by ga*Mha*GgH using the Glucose Oxidase Assay Kit (Sigma-Aldrich), as previously described (Alarico, Costa et al. 2014). The temperature profile of ga*Mha*GgH and *Mha*GgH-His_6 variants was traced from 20 to 60°C using 20 mM GG, in 20 mM sodium phosphate pH 6.0, 100 mM KCl, 5 mM $MgCl_2$ reaction buffer. The effect of pH on the activity of ga*Mha*GgH and *Mha*GgH-His_6 was determined at 55 and 42°C, respectively, using 20 mM sodium acetate (pH 4.0-5.5) or 20 mM sodium phosphate (pH 5.8-7.0). For ga*Mha*GgH, the effect of pH was also analysed at 37°C using 20 mM sodium acetate (pH 5.0-5.5), sodium phosphate (pH 6.0-8.0), Tris-HCl (7.0-9.0) or CAPSO (9.0-9.5) buffers. The pH values were adjusted using the conversion factor ($\Delta pKa/\Delta T$ [°C]) of 0.002 for sodium acetate, -0.0028 for sodium phosphate and -0.031 for Tris-HCl (Good, Winget et al. 1966). The effect of metallic cations on ga*Mha*GgH activity was examined by adding 5 mM chloride salts of Mg^{2+}, Mn^{2+}, Co^{2+}, Ca^{2+}, Fe^{2+}, Zn^{2+} or Cu^{2+} to the standard reaction. Reactions without metallic cations or with 5 mM EDTA were performed as controls. The effect of KCl concentration on ga*Mha*GgH activity was tested from 40 to 250 mM KCl. The half-life of the enzymes was evaluated by incubating *Mha*GgH aliquots (25 µL of a 2.8 µg μL^{-1} protein solution in storage buffer) at 42°C (*Mha*GgH-His_6) or 55°C (ga*Mha*GgH). Reaction aliquots were collected at different time points (up to 24h), cooled on ice, and the residual activity was quantified under the optimized reaction conditions.

Kinetic parameters of ga*Mha*GgH activity against 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 concentration of the enzyme (2.75 μM) was incubated with increasing concentrations of GG or MG in 20 mM sodium phosphate pH 6.0, 100 mM KCl, 5 mM $MgCl_2$ at 50°C. Maximum velocity (V_{max}) and half constant ($K_{0.5}$) were calculated with Prism (GraphPad Software), using the allosteric sigmoidal equation. All experiments were performed in triplicate by Susana Alarico, (Molecular Mycobacteriology Group, CNC, University of Coimbra, Portugal).

3.2.11. Crystallization of ga*Mha*GgH variants

Initial crystallization conditions were screened at 20°C with commercial sparse-matrix crystallization screens. Sitting-drop vapour-diffusion experiments were set up in 96-well CrystalQuick plates (Greiner Bio-One) using a Cartesian PixSys 4200 crystallization robot (Genomic Solutions) at the High Throughput Crystallization Laboratory (HTX Lab) of the EMBL Grenoble Outstation (Grenoble, France). The drops consisted of equal volumes (100 nL) of ga*Mha*GgH (at 9.5 mg mL^{-1} in storage buffer) and crystallization solution and were equilibrated against an 88 μL reservoir. Three-dimensional wild-type ga*Mha*GgH crystals appeared in condition 95 of the Morpheus sparse-matrix screen (Molecular Dimensions) after one day. The crystals were optimized in-house using 24-well Cryschem M plates (Hampton Research) from drops composed of equal volumes (1 μL) of protein and precipitant solutions equilibrated against 300 μL of 0.1 M Tris-Bicine pH 8.5, 0.1 M amino acids (glutamate, alanine, glycine, lysine and serine at equal molar concentration) and 24-34% GOL_P4K (2:1 of (v/v) glycerol to (w/v) PEG 4,000) as precipitant. Crystals were obtained after 1 to 2 d, and those growing from less than 30% GOL_P4K were first transferred to a solution containing at least 30% GOL_P4K before flash-cooling in liquid nitrogen. Both SeMet-ga*Mha*GgH and the ga*Mha*GgH point mutants were crystallized in the same conditions, although wild-type ga*Mha*GgH macro-seeds were employed to promote crystal growth of the point mutants. The complexes of D182A with GG, MG and GGlycolate were obtained by soaking the crystals in mother liquor supplemented with 100 mM of ligand for 2 h (GG or MG) or 25 min (GGlycolate) before flash-cooling in liquid nitrogen. The complexes of E419A with GG, MG, GGlycolate and GGlycerol were obtained by soaking the crystals in mother liquor supplemented with 100 mM of ligand for 5 (GG), 50 (MG), 40 (GGlycolate) or 10 min (GGlycerol). Crystals of wild-type ga*Mha*GgH were also obtained in the absence of serine, by omitting this compound from the amino acid mixture of the crystallization buffer. An additional crystallization condition was

identified in-house, in 24-well plates at 20°C, yielding crystals within a day from drops containing equal volumes (1 µL) of protein (at 9.5 mg mL^{-1} in storage buffer) and precipitant solution, equilibrated against 300 µL of solution 22 (0.1 M ADA pH 6.5, 1.0 M ammonium sulfate) of the MembFac sparse-matrix screen (Hampton Research). Crystals of Q434F mutant were also obtained in this last crystallization condition, containing 0.4-1.0 M ammonium sulfate. The ga*Mha*GgH and Q434F crystals were cryo-protected with Perfluoropolyether PFO-X175/08 (Hampton Research) or the crystallization condition supplemented with 30% (w/v) glucose, respectively, before being flash-cooled in liquid nitrogen.

3.2.12. Data collection and processing

Diffraction data were collected from cryo-cooled (100 K) single crystals at beamlines ID23-2 (Flot, Mairs et al. 2010), ID29 (de Sanctis, Beteva et al. 2012), ID30A-1 (Bowler, Nurizzo et al. 2015, Svensson, Malbet-Monaco et al. 2015), ID30A-3 (Theveneau, Baker et al. 2013) and ID30B (Mueller-Dieckmann, Bowler et al. 2015) of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) and PROXIMA-2A (Duran, Le Couster et al. 2013) of the French National Synchrotron Source (SOLEIL, Gif-sur-Yvette, France) (see Table 3.2 for details). All data sets were automatically processed by the Grenoble Automatic Data ProcEssing pipeline (Monaco, Gordon et al. 2013), except those collected at ID30B-3 and PROXIMA-2A beamlines, which were processed with XDS (Kabsch 2010) and reduced with utilities from the CCP4 program suite (Winn, Ballard et al. 2011). The data sets from the selenomethionine-substituted crystals were scaled with XSCALE (Kabsch 2010). X-ray diffraction data collection and processing statistics are summarized in Table 3.2.

3.2.13. Structure determination, model building and refinement

The structure of ga*Mha*GgH was solved by two-wavelength anomalous diffraction using the anomalous signal of a selenium-substituted ga*Mha*GgH crystal with the SHELXC/SHELXD/SHELXE pipeline (Sheldrick 2010) and the HKL2MAP GUI (Pape and Schneider 2004). Automated model building was subsequently performed with ARP/wARP (Langer, Cohen et al. 2008). The refined coordinates were used as search model to solve the structure of all other ga*Mha*GgH variants and complexes by molecular replacement using PHASER (McCoy, Grosse-Kunstleve et al. 2007). Alternating cycles of

model building with Coot (Emsley, Lohkamp et al. 2010) and refinement with PHENIX (Adams, Afonine et al. 2010) were performed until model completion. Refined coordinates and structure factors were deposited at the Protein Data Bank (Berman, Westbrook et al. 2000) under the PDB accession codes indicated in Table 3.3. Refinement statistics are summarized in Table 3.3.

3.2.14. Analysis of crystallographic structures

The crystallographic models were superposed with Coot (Emsley, Lohkamp et al. 2010) and the secondary structure elements were identified with DSSP (Kabsch and Sander 1983, Touw, Baakman et al. 2015). The interface area between the monomers was determined using PISA (Krissinel and Henrick 2007). The molecular surface electrostatic potential was calculated with APBS (Baker, Sept et al. 2001) using the AMBER force field (Cornell, Cieplak et al. 1995). Figures depicting molecular models were created using PyMOL (Schrödinger).

3.2.15. SAXS measurement and analysis

SAXS measurements were performed at beamline BM29 (Pernot, Round et al. 2013) of the ESRF (Grenoble, France) with radiation of 0.9919 Å wavelength on a Pilatus 1M detector (Dectris). Protein was loaded onto a Superdex 200 3.2/300 GL (GE Healthcare) column 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^{-1}. Data were processed and analyzed with the ATSAS package (Petoukhov, Franke et al. 2012). A Guinier plot was calculated using PRIMUS QT (Konarev, Volkov et al. 2003). The theoretical scattering curve from the crystallographic model was fitted to the experimental scattering curve with CRYSOL (Svergun, Barberato et al. 1995).

3.2.16. MST analysis

Interactions between catalytically inactive D182A or E419A ga*Mha*GgH mutants and GG or MG were assessed by microscale thermophoresis (MST) on a Monolith NT.115 instrument (NanoTemper Technologies). Purified proteins were labelled using the

Monolith NT™ Protein Labelling Kit RED-NHS (NanoTemper Technologies) according to the manufacturer's instructions after buffer exchange to 50 mM HEPES pH 7.8, 300 mM NaCl, 0.05% Tween®-20 (labelling buffer). Labelled proteins were separated from the free dye on a gravity flow column using 50 mM Tris-HCl pH 7.8, 150 mM NaCl, 10 mM $MgCl_2$, 0.05% Tween®-20 (MST buffer) as mobile phase. Protein concentration was estimated by measuring the absorbance at 280 nm. A constant concentration of labelled protein [160 nM, diluted 1:25 in 100 mM Tris-HCl pH 8.0, 400 mM NaCl (MST assay buffer)] was incubated with increasing concentrations of substrate (6.10 µM - 200 mM) in MST assay buffer in a 10 µL reaction volume. Measurements were performed at room temperature on Monolith™ NT.115 MST Premium Coated Capillaries (NanoTemper Technologies) with 50% LED power and 40% MST power. Three and two independent experiments were performed for D182A and E419A, respectively. Data were analysed and images were prepared using PALMIST (Scheuermann, Padrick et al. 2016).

3.3. Results and discussion

3.3.1. *Mha*GgH production and crystallization

*Mha*GgH variants were expressed in *E. coli* and purified to homogeneity by a combination of immobilized metal-affinity and size exclusion chromatography. Initially, the crystallization experiments were performed using the *Mha*GgH-His_6 construct, containing a C-terminal hexahistidine tag linked to the native *Mha*GgH sequence (Table 3.1), and for which biochemical characterization was previously performed (Alarico, Costa et al. 2014). The initial crystallization trials were performed at EMBL Grenoble Outstation HTX Lab (France). Spherulites were obtained in drops composed of 2.5 to 10% PEG (3,350, 5,000, 6,000, 8,000 or MME 5,000) and buffered from pH 8.0 to 9.5 with 100 mM Tris-HCl, bicine or CHES. The quality of the crystals was subsequently improved to thin individual crystal plates by adjusting the crystallization solution composition (precipitant type and concentration, buffer type and pH and additives) and vapour-diffusion rate (oil barrier, protein-to-precipitant drop ratio and drop-to-reservoir ratio) and by micro [cat whisker, Seed Bead™ Kit (Hampton research)] and macroseeding. Similar results were also obtained by microseed matrix screening (Ireton and Stoddard 2004) (Figure 3.1A), however none of the crystals tested diffracted X-rays to high resolution.

Since the C-terminal tag could be interfering with crystallization, a new construct was produced. The new construct contained a N-terminal hexahistidine tag separated from the native *Mha*GgH sequence by a TEV recognition site that allowed proteolytic tag removal (Table 3.1). The TEV protease cleavage resulted in an additional Gly-Ala dipeptide N-terminal to the native *Mha*GgH amino acid sequence and therefore this construct was named ga*Mha*GgH. Initial crystallization conditions were identified using extensive sampling of commercial sparse-matrix crystallization screens at HTX Lab (France). Three-dimensional crystals were obtained in condition 95 of the Morpheus sparse-matrix screen (Molecular Dimensions) within 1 day. These crystals were reproduced in-house (Figure 3.1B).

Table 3.1 – Macromolecule production information.

	*Mha*GgH-His$_6$	ga*Mha*GgH
Source organism	*M. hassiacum* DSM 44199	
Expression vector	pET30a	pETM11
Complete amino-acid sequence of the construct produced[a]	MPHDPSFTPTQLAARAAYLLRGNDL GTMTTAAPLLYPHMWSWDAAFVAIG LAPLSVERAVVELDTLLSAQWRNGM IPHIVFANGVDGYFPGPARWATATL ADNAPRNRLTSGITQPPVHAIAVQR ILEHARTRGRSTRAVAEAFLDRRWG DLMRWHRWLAECRDRNERGRITLYH GWESGMDNSPRWDSAYANVVPGKLP EYQRADNVIITDPSQRPSDGEYDRY LWLLEEMKAVRYDDERLPSVMSFQV EDVFFSAIFSVACQVLAEIGEDYKR PHADVKDLYLWAERFRAGVVETTDQ RTGAARDFDVLAEKWLVTETAAQFA PLLCGGLPHDRERALLKLLEGPRFC GHPDLKYGLIPSTSPVSRDFRPREY WRGPVWPVLTWLFSWCFARRGWAER ARLLRQEGLRQASDGSFAEYYEPFT GEPLGSMQQSWTAAAVLDWLGKLAA ALEHHHHHH	*MKHHHHHHPMSDYDIPTTENLYFQ*G AMPHDPSFTPTQLAARAAYLLRGND LGTMTTAAPLLYPHMWSWDAAFVAI GLAPLSVERAVVELDTLLSAQWRNG MIPHIVFANGVDGYFPGPARWATAT LADNAPRNRLTSGITQPPVHAIAVQ RILEHARTRGRSTRAVAEAFLDRRW GDLMRWHRWLAECRDRNERGRITLY HGWESGMDNSPRWDSAYANVVPGKL PEYQRADNVIITDPSQRPSDGEYDR YLWLLEEMKAVRYDDERLPSVMSFQ VEDVFFSAIFSVACQVLAEIGEDYK RPHADVKDLYLWAERFRAGVVETTD QRTGAARDFDVLAEKWLVTETAAQF APLLCGGLPHDRERALLKLLEGPRF CGHPDLKYGLIPSTSPVSRDFRPRE YWRGPVWPVLTWLFSWCFARRGWAE RARLLRQEGLRQASDGSFAEYYEPF TGEPLGSMQQSWTAAAVLDWLG
No of amino acids	459	472 (tagged), 448 (untagged)
Theoretical molecular weigth (Da)	52,288	53,895 (tagged), 50,897 (untagged)
Mutants produced	None	D43A, D182A, E419A, Q434F

[a] The residues removed by TEV protease cleavage are italicized. The sequence of native *Mha*GgH is underlined.

In order to solve the phase problem selenomethionine-containing ga*Mha*GgH was also produced. This variant was considerable less soluble than the unlabelled protein under the same experimental conditions. Therefore, SeMet-ga*Mha*GgH could only be concentrated to a fourfold lower concentration than the unlabelled material. Under crystallization conditions similar to the wild-type enzyme, two SeMet-gaMhaGgH crystals were obtained after four months (Figure 3.1C). The slower crystallogenesis of selenium-labelled ga*Mha*GgH is most likely due to the lower sample concentration.

Point mutants of active site residues potentially involved in catalysis and/or substrate binding were also produced (Table 3.1). These mutants (D43A, D182A, E419A, Q434F) were produced and crystallized in condition identical to the wild-type ga*Mha*GgH, albeit with the addition of wild-type seeds to promote crystal growth (Figure 3.1E to G). The crystals of mutant Q434F grew at a significantly slower rate and reached a considerable smaller size than the other mutants.

Three-dimensional crystals of ga*Mha*GgH were also obtained in-house within 1 day in another crystallization condition containing 0.1 M ADA pH 6.5, 1.0 M ammonium sulfate (Figure 3.1D). Due to the difficulty in crystallizing the Q434F mutant, the ability of this condition to promote Q434F crystal growth was also tested. Crystals of Q434F were easily obtained without requiring the addition of wild-type crystal seeds to promote crystal growth (Figure 3.1H).

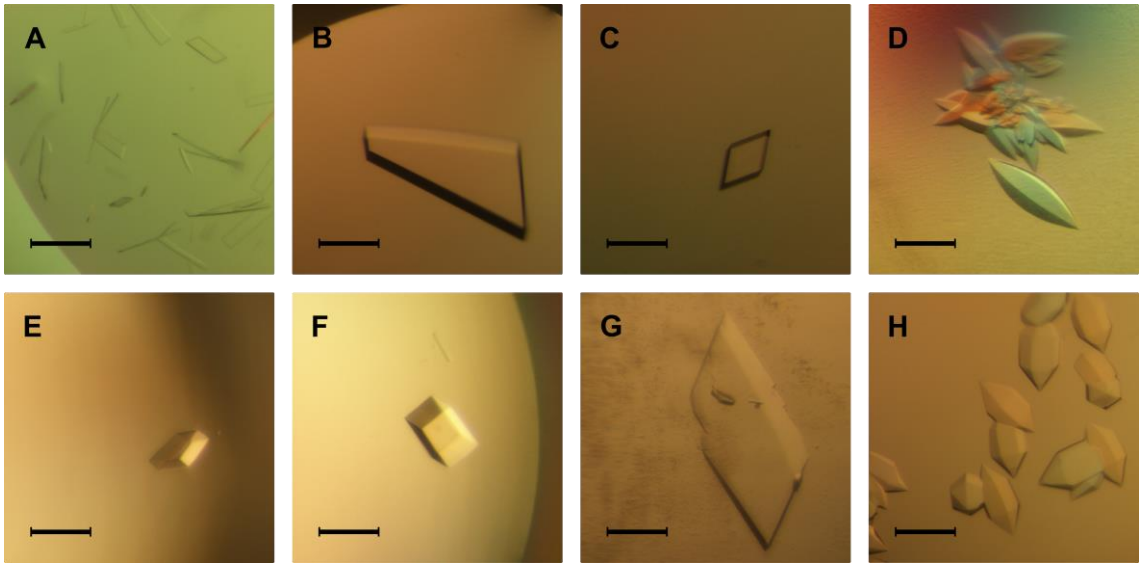


Figure 3.1 – Crystals from *Mha*GgH variants. A – Crystals of *Mha*GgH-His_6 obtained by microseed matrix screening. B – ga*Mha*GgH crystal grown from a mixture of Tris-Bicine, amino acids and GOL_P4K. C – SeMet-ga*Mha*GgH crystal obtained in the same crystallization condition as in B. D – ga*Mha*GgH crystal obtained from the ammonium sulfate containing crystallization solution. E to G – Crystals of the ga*Mha*GgH point mutants D43A (E), D182A (F) and E419A (G) obtained by macroseeding from the same crystallization condition as in B. H – Crystals from mutant Q434F obtained in the same crystallization condition as in D. The scale bars represent a length of 0.2 mm.

3.3.2. Overall structure of ga*Mha*GgH

The X-ray diffraction from cryo-protected crystals was measured at synchrotron radiation sources (Table 3.2; Supplemental figure S3.1). Two diffraction data sets were recorded from a single SeMet-ga*Mha*GgH crystal at wavelengths corresponding to the peak and the inflexion point of the *K* absorption edge of selenium (Supplemental figure S3.2). The crystal belonged to the monoclinic space group $P2_1$ and diffracted X-rays to ~2.0 Å (Supplemental figure S3.3A). The collected data sets were used to solved the phase

problem by two-wavelength anomalous diffraction phasing with the SHELXC/SHELXD/SHELXE pipeline (Sheldrick 2010) and the HKL2MAP GUI (Pape and Schneider 2004). SHELXD identified 27 of the 36 (nine per molecule) selenium atoms present in the asymmetric unit (CC_{all} = 23.6%; CC_{weak} = 17.8%), corresponding to SeMet residues 153, 232, 246 and 432 of all molecules, as well as SeMet residue 28 of molecules B and C, SeMet residue 39 of molecules A, B and C, and SeMet residues 75 and 181 of molecules A, B and D. Experimental phasing with density modification using SHELXE resulted in excellent electron-density maps, with a final FOM of 0.652 and a pseudo-free CC of 71.3%. The autotracing function of SHELXE produced a 1,411-residue poly-Ala model (CC for partial structure against native data = 36.1%). Subsequent automated model building with ARP/wARP (Langer, Cohen et al. 2008) using the SHELXE-derived phases resulted in an initial model with 1,677 (out of 1,792) residues built and sequenced. The crystallographic model of ga*Mha*GgH was further improved through alternating cycles of manual building with Coot (Emsley, Lohkamp et al. 2010) and refinement with PHENIX (Adams, Afonine et al. 2010). The asymmetric unit contains four ga*Mha*GgH monomers, corresponding to a Matthews coefficient of 3.07 $Å^3$ Da^{-1} and a solvent content of 59.9%.

Refined coordinates of the SeMet-ga*Mha*GgH model were used to solve the phase problem of a higher-resolution data set (1.68 Å) collected from a native ga*Mha*GgH crystal. Although crystallizing in identical conditions, the unlabelled ga*Mha*GgH crystals belong to space group $P2_12_12$ (Supplemental figure S3.3B). The crystal contains two molecules in the asymmetric unit, corresponding to a Matthews coefficient of 2.95 $Å^3$ Da^{-1} and a solvent content of 58.3%. The crystallographic structure was improved by alternating cycles of manual model building with Coot (Emsley, Lohkamp et al. 2010) and refinement with PHENIX (Adams, Afonine et al. 2010). The structure solution and refinement statistics are summarized in table 3.3.

Table 3.2 – Data collection and processing statistics[a] (continued on next page).

Crystal	SeMet ga*Mha*GgH		*ga*Mha GgH	*ga*Mha GgH	*ga*Mha GgH	D43A	D182A	E419A
	Peak	Inflexion point	SER-GOL	w/o Serine	Apo	SER-GOL	SER-GOL	SER-GOL
Synchrotron radiation facility	ESRF		ESRF	ESRF	ESRF	ESRF	ESRF	ESRF
Beamline	ID29		ID30B	ID30A-3	ID29	ID30A-1	ID30A-3	ID30A-1
Detector	Dectris PILATUS3 6M		Dectris PILATUS 6M	Dectris EIGER X 4M	Dectris PILATUS3 6M	Dectris PILATUS3 2M	Dectris EIGER X 4M	Dectris PILATUS3 2M
Wavelength (Å)	0.97909	0.97924	0.97265	0.96770	0.96863	0.96598	0.96770	0.96600
Frames	1,500	1,500	2,400	1,200	1,300	2,180	2,500	640
Rotation (°)	0.1	0.1	0.05	0.1	0.05	0.05	0.05	0.2
Space group	$P2_1$		$P2_12_12$	$P2_12_12$	$P6_222$	$P2_12_12$	$P2_12_12$	$P2_12_12$
Unit cell parameters	a=90.8 Å; b=86.1 Å; c=159.7 Å; β=93.0°	a=90.9 Å; b=86.3 Å; c=159.7 Å; β=93.0°	a=86.0 Å; b=158.8 Å; c=87.8 Å	a=85.9 Å; b=159.3 Å; c=91.2 Å	a=b=167.0 Å; c=243.3 Å	a=85.9 Å; b=159.1 Å; c=88.2 Å	a=86.3 Å; b=158.1 Å; c=87.7 Å	a=86.2 Å; b=159.4 Å; c=88.4 Å
Resolution range (Å)	57.4 – 2.04 (2.07 – 2.04)	58.6 – 2.06 (2.09 – 2.06)	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)
Reflections (measured/unique)	428,683/151,619	419,987/148,763	595,791/135,623	390,265/84,736	474,316/66,272	460,425/114,314	581,967/119,738	338,529/73,877
Completeness (%)	96.5 (82.1)	97.4 (87.6)	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)	2.8 (2.7)	4.4 (4.5)	4.6 (4.8)	7.2 (7.3)	4.0 (4.0)	4.9 (5.0)	4.6 (4.0)
Anomalous completeness	68.3 (53.3)	85.0 (71.9)						
Anomalous multiplicity	1.2 (1.7)	1.3 (1.5)						
R_{merge}[b]	0.105 (0.637)	0.105 (0.561)	0.047 (0.860)	0.084 (1.276)	0.109 (1.165)	0.057 (0.824)	0.070 (0.987)	0.094 (1.032)
R_{meas}[c]	0.128 (0.776)	0.129 (0.690)	0.053 (0.972)	0.094 (1.426)	0.117 (1.254)	0.066 (0.954)	0.079 (1.102)	0.106 (1.176)
$R_{p.i.m.}$[d]	0.072 (0.438)	0.074 (0.397)	0.025 (0.444)	0.042 (0.622)	0.043 (0.459)	0.032 (0.472)	0.035 (0.480)	0.048 (0.551)
Half-set correlation $CC_{1/2}$	0.990 (0.659)	0.990 (0.657)	0.999 (0.637)	0.998 (0.559)	0.998 (0.673)	0.999 (0.529)	0.998 (0.579)	0.997 (0.625)
Half-set anomalous correlation CC_{Anom}	0.355 (-0.028)	0.166 (0.018)						
Mean (I)/σ (I)	7.6 (1.6)	7.1 (1.8)	16.4 (1.7)	12.1 (1.2)	11.9 (1.8)	13.0 (1.7)	11.6 (1.6)	10.0 (1.3)
Protomers per asymmetric unit	4	4	2	2	2	2	2	2
Matthews coefficient ($Å^3Da^{-1}$)	3.07	3.08	2.95	3.07	4.82	2.97	2.94	2.99
Solvent content (%)	59.9	60.1	58.3	59.9	74.5	58.6	58.2	58.9
Average mosaicity	0.17	0.20	0.04	0.17	0.08	0.10	0.06	0.23

Table 3.2 – Data collection and processing statistics[a] (continued from previous page).

Crystal	D182A GG	D182A MG	D182A GGlycolate	E419A GG	E419A MG	E419A GGlycolate	E419A GGycerol	Q434F SER-GOL	Q434F Apo
Synchrotron radiation facility	ESRF	ESRF	ESRF	ESRF	ESRF	SOLEIL	ESRF	ESRF	ESRF
Beamline	ID30A-1	ID30A-1	ID23-2	ID30A-1	ID30B	PROXIMA 2	ID30A-3	ID30A-3	ID30A-3
Detector	Dectris PILATUS3 2M	Dectris PILATUS3 2M	Dectris PILATUS3 2M	Dectris PILATUS3 2M	Dectris PILATUS 6M	Dectris EIGER X 9M	Dectris EIGER X 4M	Dectris EIGER X 4M	Dectris EIGER X 4M
Wavelength (Å)	0.96600	0.96600	0.87290	0.96599	0.97625	0.98011	0.96770	0.96770	0.96770
Frames	3,200	900	3,200	2,500	3,600	1,800	2,200	3,600	3,600
Rotation (°)	0.05	0.2	0.05	0.05	0.05	0.1	0.05	0.1	0.05
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P6_222$
Unit cell parameters	a=85.3 Å; b=159.6 Å; c=91.0 Å	a=85.3 Å; b=159.6 Å; c=91.2 Å	a=86.2 Å; b=158.9 Å; c=88.0 Å	a=86.1 Å; b=159.1 Å; c=88.4 Å	a=87.8 Å; b=158.2 Å; c=87.6 Å	a=86.9 Å; b=157.7 Å; c=87.6 Å	a=86.9 Å; b=158.8 Å; c=87.7 Å	a=85.6 Å; b=158.8 Å; c=88.3 Å	a=b=169.2 Å; c=241.8 Å
Resolution range (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)	48.6 – 2.32 (2.40 – 2.32)	47.9 – 2.60 (2.67 – 2.60)
Reflections (measured/unique)	801,298/133,666	780,685/117,039	547,439/91,224	301,215/64,817	475,955/74,395	500,266/75,133	322,372/76,309	712,653/52,533	1,280,221/62,925
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)	99.3 (93.5)	99.5 (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)	13.6/10.4	20.3 (21.0)
Anomalous completeness									
Anomalous multiplicity									
R_{merge}[b]	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)	0.115 (1.492)	0.123 (1.923)
R_{meas}[c]	0.077 (1.183)	0.080 (1.141)	0.139 (1.340)	0.102 (0.897)	0.094 (1.069)	0.135 (1.526)	0.104 (1.027)	0.119 (1.568)	0.126 (1.969)
$R_{p.i.m.}$[d]	0.031 (0.474)	0.031 (0.436)	0.056 (0.533)	0.047 (0.417)	0.037 (0.423)	0.052 (0.579)	0.049 (0.488)	0.032 (0.471)	0.028 (0.425)
Half-set correlation $CC_{1/2}$	0.999 (0.659)	0.999 (0.660)	0.994 (0.560)	0.998 (0.642)	0.998 (0.715)	0.997 (0.575)	0.997 (0.552)	0.999 (0.536)	0.999 (0.805)
Half-set anomalous correlation CC_{Anom}									
Mean (I)/σ (I)	13.8 (1.8)	16.5 (1.7)	8.8 (1.5)	11.3 (1.8)	12.1 (1.5)	8.2 (1.3)	9.4 (1.6)	15.4 (1.6)	18.5 (2.0)
Protomers per asymmetric unit	2	2	2	2	2	2	2	2	2
Matthews coefficient ($Å^3Da^{-1}$)	3.05	3.06	2.97	2.98	3.00	2.96	2.98	2.95	4.91
Solvent content (%)	59.7	59.8	58.6	58.8	59.0	58.4	58.8	58.3	75.0
Average mosaicity	0.07	0.13	0.05	0.10	0.09	0.14	0.09	0.15	0.08

[a] Values in parenthesis correspond to the outermost resolution shell.

[b] $R_{merge} = \sum_{hkl} \sum_i \left| I_{i(hkl)} - \langle I_{hkl} \rangle \right| / \sum_{hkl} \sum_i I_{i(hkl)}$, where $I_{i(hkl)}$ is the observed intensity and $\langle I_{hkl} \rangle$ is the average intensity of multiple observations of symmetry-related reflections.

[c] $R_{meas} = \sum_{hkl} \sqrt{\frac{n}{n-1}} \sum_{i=1}^{n} \left| I_{i(hkl)} - \langle I_{hkl} \rangle \right| / \sum_{hkl} \sum_i I_{i(hkl)}$, where $I_{i(hkl)}$ is the observed intensity and $\langle I_{hkl} \rangle$ is the average intensity of multiple observations of symmetry-related reflections.

[d] $R_{p.i.m.} = \sum_{hkl} \sqrt{\frac{1}{n-1}} \sum_{i=1}^{n} \left| I_{i(hkl)} - \langle I_{hkl} \rangle \right| / \sum_{hkl} \sum_i I_{i(hkl)}$, where $I_{i(hkl)}$ is the observed intensity and $\langle I_{hkl} \rangle$ is the average intensity of multiple observations of symmetry-related reflections.

Table 3.3 – Structure solution and refinement statistics (continued on next page).

Crystal	SeMet ga*Mha*GgH	ga*Mha*GgH SER-GOL	ga*Mha*GgH w/o Serine	ga*Mha*GgH Apo	D43A SER-GOL	D182A SER-GOL	E419A SER-GOL
PDB code	5OHZ	5OI0	5OHC		5OIV	5OI1	5OIE
Resolution range (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_{factor}[a]/Free R_{factor}[b] (%)	19.9/24.5	14.6/17.2	16.7/21.0	16.8/20.8	14.9/17.9	14.6/17.2	16.7/20.8
Unique reflections (working/test set)	255,146/12,657	135,565/6,808	84,685/4,330	66,186/3,360	114,235/5,735	119,673/5,953	73,714/3,718
Completeness (working/test) (%)	82.5/4.96	98.8/5.02	99.4/5.11	99.6/5.08	98.8/5.02	98.7/4.97	98.6/5.04
Wilson B-factor (Å^2)	28.4	25.1	33.4	53.0	29.3	26.0	39.9
Total number of atoms	15,955	8,595	8,115	7,607	8,365	8,421	7,767
B-factor (Å^2)							
Overall	36.1	33.4	39.1	60.0	38.1	34.2	49.4
Protein	35.5	31.7	38.2	60.0	36.7	32.7	48.9
Ligands	46.2	48.4	54.6		56.5	48.5	71.8
Waters	41.3	47.3	47.2	60.8	50.4	47.8	54.3
RMSD							
Bond lengths (Å)	0.010	0.010	0.010	0.010	0.010	0.010	0.010
Bonds angles (°)	1.099	1.031	0.993	1.101	1.009	1.019	1.004
DPI (Å)	0.164	0.073	0.130	0.180	0.086	0.080	0.140
Polypeptide length	A: 443 (Asp4 – Gly446) B/D: 442 (Pro5 – Gly446) C: 444 (His3 – Gly446)	445 (Pro2 – Gly446)	445 (Pro2 – Gly446)	A: 443 (Asp4 – Gly446) B: 448 (Gly-1 – Gly446)	445 (Pro2 – Gly446)	445 (Pro2 – Gly446)	445 (Pro2 – Gly446)
Polypeptide gaps	0	0	0	0	0	0	0
Water molecules	1,289	838	705	365	761	776	420
Ligands	SER, GOL	SER, GOL	GOL		SER, GOL	SER, GOL	SER, GOL

Table 3.3 – Structure solution and refinement statistics (continued from previous page).

Crystal	D182A GG	D182A MG	D182A GGlycolate	E419A GG	E419A MG	E419A GGlycolate	E419A GGlycerol	Q434F SER-GOL	Q434F Apo
PDB code	5OIW	5OJ4	5ONZ	5OJU	5OJV	5OO2	5ONT		
Resolution range (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	41.3 – 2.32	46.6 – 2.60
R_{factor}[c]/Free R_{factor}[d] (%)	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	15.3/20.3	18.2/22.1
Unique reflections (working/test set)	133,615/6,717	116,974/5,861	91,157/4,566	64,760/3,281	74,327/3,733	75,046/3,625	76,246/3,909	52,473/2,595	62,782/3,089
Completeness (working/test) (%)	99.3/5.03	99.4/5.01	99.7/5.01	99.7/5.07	97.8/5.02	99.9/4.83	99.2/5.13	99.3/4.95	99.4/4.92
Wilson B-factor ($Å^2$)	24.9	25.6	28.3	35.7	39.5	39.7	33.7	50.8	60.4
Total number of atoms	8,376	8,421	8,283	7,817	7,840	7,868	7,944	7,503	7,473
B-factor ($Å^2$)									
Overall	33.8	33.5	36.3	44.0	44.3	52.7	44.0	60.0	65.4
Protein	32.1	31.8	34.9	43.6	43.6	52.3	43.3	59.8	65.3
Ligands	49.2	48.5	49.1	56.2	55.8	64.3	53.7	81.1	
Waters	45.9	46.5	48.6	47.8	53.0	57.0	51.3	62.2	68.5
RMSD									
Bond lengths (Å)	0.010	0.010	0.010	0.009	0.009	0.010	0.010	0.009	0.010
Bonds angles (°)	1.025	0.986	0.984	1.005	0.985	1.030	1.011	0.974	1.139
DPI (Å)	0.072	0.083	0.108	0.150	0.130	0.135	0.129	0.1787	0.1987
Polypeptide length	445 (Pro2 – Gly446)	445 (Pro2 – Gly446)	445 (Pro2 – Gly446)	445 (Pro2 – Gly446)	445 (Pro2 – Gly446)	445 (Pro2 – Gly446)	A: 440 (Pro2 – Gly446) B: 437 (Pro2 – Gly446) 2	445 (Pro2 – Gly446)	A: 440 (Asp4 – Gly446) B: 438 (Pro5 – Gly446 2
Polypeptide gaps	0	0	0	0	0	0	(A: Asp206 – Ile210; B: Gln203 – Ile210)	0	(A: Tyr202 – Asp206; B: Arg204 – Asn207)
Water molecules	847	846	677	419	463	426	566	255	280
Ligands	9WN	2M8	GOL, SER, 9YW	9WN	GOL, SER, 2M8	GOL, SER, 9YW	A0K	GOL, SER	

The two macromolecules present in the asymmetric unit (hereby termed molecule A and molecule B) were modelled from residue Pro2 to Gly446. The overall globular *Mha*GgH monomer (Figure 3.2) is composed of two domains: an $(\alpha/\alpha)_6$-barrel domain that in *Mha*GgH encompasses helices α2, α4, α6, α8, α10 and α12 in the inner layer and α1, α3, α5, α7, α9 and α11 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 A'-region (residues 163 to 252) and B'-region (residues 68 to 118). Five mobile loops are also noteworthy: loop-A (between α1 and β1, residues 23-38), loop-B (between B'β1 and B'α1a, residues 81-91), loop-C (between A' α2b and A'α3, residues 193-205), loop-D (between α9 and α10, residues 346-381) and loop-E (between β6 and α12, residues 430-434).

*Mha*GgH displays 68% secondary structure identity with the single structurally characterized mannosylglycerate hydrolase (*Thermus thermophilus* HB8 MgH, Tt8MGH; PDB accession codes 4WVA, 4WVB and 4WVC) (Miyazaki, Ichikawa et al. 2015), despite the much lower 36% sequence identity (calculated with PDBeFold (http://www.ebi.ac.uk/msd-srv/ssm) (Krissinel and Henrick 2004)). So far, only other three MgH enzymes have been biochemically characterized: the orthologues from *Selaginella moellendorffii* (Nobre, Empadinhas et al. 2013), from *T. thermophilus* HB27 (99% amino acid sequence identity to Tt8MgH), and from *Rubrobacter radiotolerans* (Alarico, Empadinhas et al. 2013). Despite relatively low overall amino acid sequence conservation, substrate interacting residues are highly conserved among all characterized MgH enzymes (Supplemental figure S3.4).

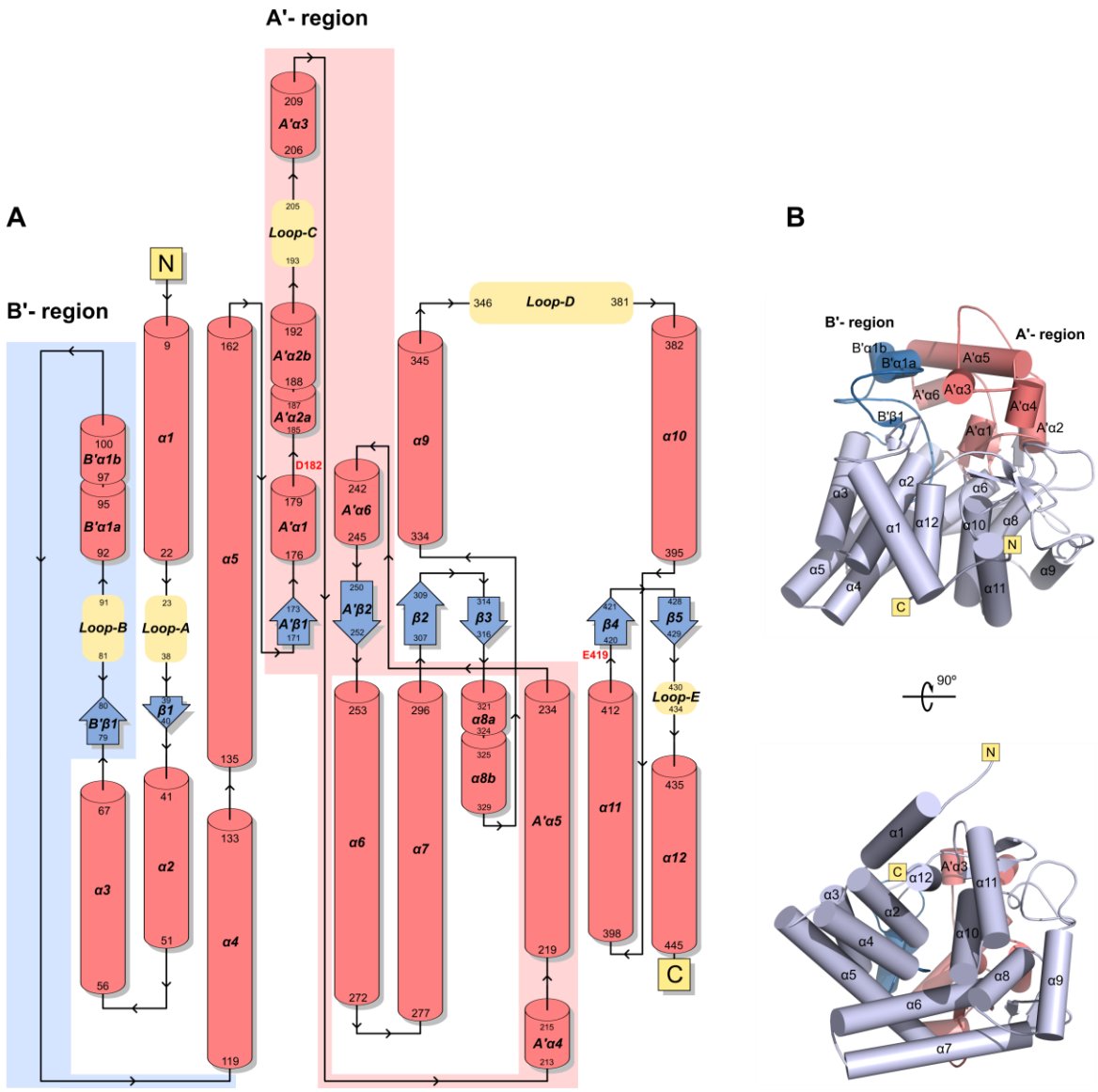


Figure 3.2 – Three-dimensional structure of *Mha*GgH. A – Topology diagram of ga*Mha*GgH with helices (α1 to α12, A'α1 to A'α6, and B'α1), strands (β1 to β5, A'β1 to A'β2, and B'β2) and mobile loops (Loop-A to Loop-E) represented as salmon cylinders, blue arrows and yellow ovals, respectively. Catalytic residues are labelled in red. The A' and B' regions are highlighted in pink and blue, respectively. B – Overall structure of the ga*Mha*GgH monomer (carton representation). Rotation of the top view 90° around x-axis originates the bottom view. The $(\alpha/\alpha)_6$ domain is coloured mauve, and A' and B'-regions are coloured salmon and blue, respectively. The N and C-termini are indicated in yellow.

3.3.3. Quaternary structure of ga*Mha*GgH

The molecular size of ga*Mha*GgH in solution was evaluated by size exclusion chromatography and DLS (Figure 3.3A and B). Despite its theoretical molecular weight of 50.9 kDa, the ga*Mha*GgH variant displayed an apparent molecular weight of 178.65 kDa, corresponding to approximately 3.5 times the expected mass of the monomer and compatible with a trimeric or a tetrameric organization. On the other hand, the *Mha*GgH-His_6 variant displayed an apparent molecular weight of 92.64 kDa, only 1.8-fold higher than the molecular weight of the monomer and compatible with a dimer (Figure 3.3A). The DLS analysis of the *Mha*GgH-His_6 and ga*Mha*GgH variants reveals an increase in the hydration radius from 5.75 to 7.34 nm, also supporting a difference in the oligomeric organization of the two variants (Figure 3.3B). Moreover, DLS analysis also revealed a lower polydispersity index for the ga*Mha*GgH solution, indicating higher homogeneity.

The thermal stability of *Mha*GgH-His_6 and ga*Mha*GgH was evaluated by differential scanning fluorimetry. Both variants produced a typical denaturation curve with a clear and sharp temperature transition, allowing the straightforward estimation of the T_m from the inflexion point of the melting curve. The ga*Mha*GgH variant displayed significantly higher thermal stability (T_m = 62ºC) than *Mha*GgH-His_6 (T_m = 56ºC) in the experimental conditions used (Figure 3.3D).

Altogether, these results suggest that the hexahistidine tag present at the C-terminus of the *Mha*GgH-His_6 variant affects protein stability probably by interfering with protein oligomerization.

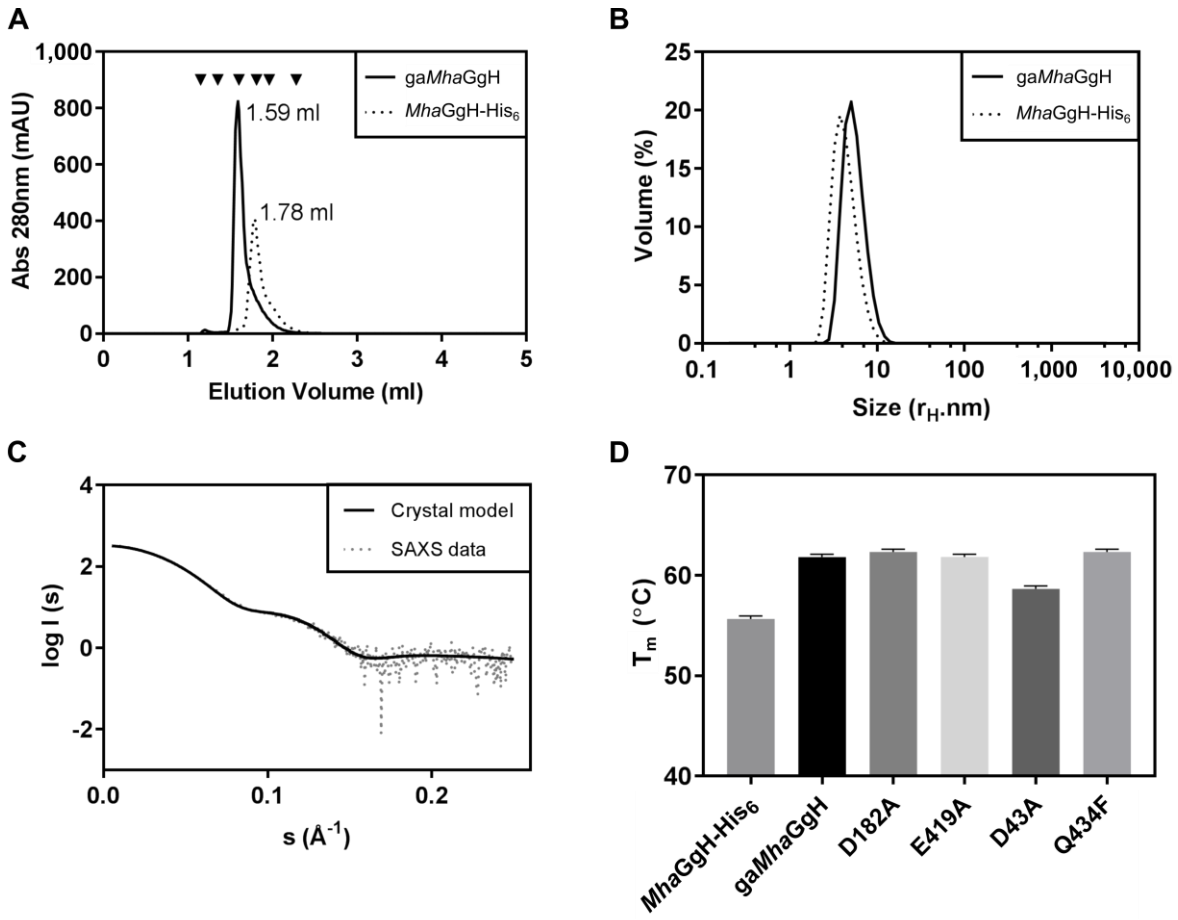


Figure 3.3 – Biophysical characterization of *Mha*GgH variants. A – Analytical size-exclusion chromatography of *Mha*GgH-His$_6$ (dotted line) and ga*Mha*GgH (solid line) variants. The standards used for column calibration (see Material and Methods) are indicated as downward black triangles. B – Analysis of *Mha*GgH-His$_6$ (dotted line) and ga*Mha*GgH (solid line) by DLS. The ga*Mha*GgH sample displayed a larger hydrodynamic radius (R_H = 7.34 nm) and lower polydispersity index (PdI = 0.092) than the *Mha*GgH-His$_6$ variant (R_H = 5.75 nm; PdI = 0.201). C – Superposition of the experimental small X-ray scattering (SAXS) data and the theoretical SAXS curve calculated from the tetrameric crystallographic model of *Mha*GgH. D – Melting temperature of *Mha*GgH variants determined by differential scanning fluorimetry, highlighting the lower stability of the *Mha*GgH-His$_6$ variant.

In the crystals, ga*Mha*GgH is arranged as a dimer of dimers with approximate dimensions 85 x 80 x 65 Å^3 (Figure 3.4). The total surface area of each monomer is ~17,298 Å^2. All monomers of the tetramer share an interface with each other. The largest interface area (~900 Å^2) occurs between molecules A and C (interface A:C), and molecules B and D, and is maintained by 14 (A:C) and 12 (B:D) hydrogen bonds. The interface area between the A:B and C:D dimers is ~478 Å^2, and includes 10 salt bridges. The interface between molecules A and D and molecules B and C is the smallest one with an area of ~261 Å^2, involving a single

hydrogen bond. The total interface area per monomer is approximately 1,636 Å^2 (9.5% of the total surface area). No covalent bonds are found at the intermonomer interfaces (Supplemental table S3.1) (*PISA* analysis) (Krissinel and Henrick 2007).

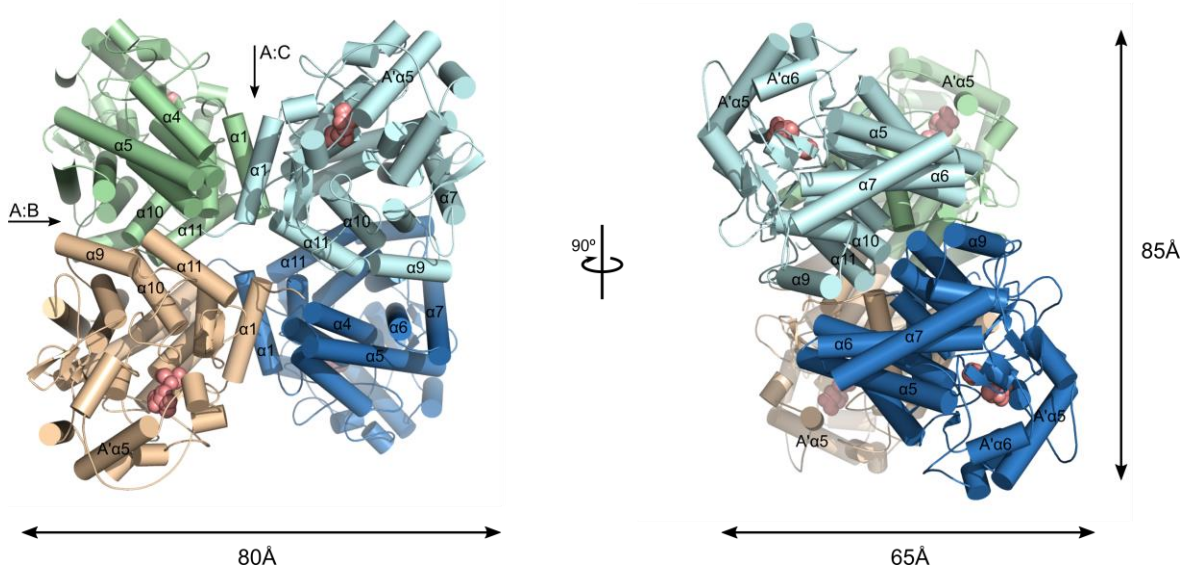


Figure 3.4 – Quaternary structure of ga*Mha*GgH. The monomers (cartoon representation) are coloured in green (molecule A), wheat (molecule B), cyan (molecule C) and blue (molecule D). The interfaces A:B or A:C are indicated. The active site is highlighlited in salmon. The α-helices are numbered. The approximate dimensions of the homotetramer are indicated.

The evaluation of ga*Mha*GgH dimensions in solution was also performed by SAXS. A Guinier analysis of the SAXS data was performed with PRIMUS QT (Konarev, Volkov et al. 2003), allowing the direct estimation of a radius of gyration of 41.9 Å, which is similar to the approximate dimensions of the crystallographic model (Figure 3.4). Also, a theoretical SAXS curve was generated from the atomic structure data and fitted to the experimental scattering curves with CRYSOL (Svergun, Barberato et al. 1995), revealing a good agreement between the SAXS data and the crystallographic model (Figure 3.3C).

The structural analysis of the ga*Mha*GgH tetramer revealed that the C-termini of molecules A and C, and of molecules B and D (Figure 3.5) are hydrogen bound. The C-terminus of each molecule adopts two different conformations (Figure 3.5B and C), the most frequent of which (52% occupancy in the highest resolution structure) (Figure 3.5B) is the only conformation observed in lower resolution models. In this conformation, the carboxyl group of Gly446 is

stabilised by two hydrogen bonds with the side chains of Thr8 and Thr10, located at helix α1 of the neighbour molecule. Interestingly, these two threonine residues are separated by a proline (Pro9), which confers rigidity to this segment. Water-mediated contacts are also present, involving His129, Leu442, Asp443, and Gln11. The other conformation of the C-terminus is illustrated in figure 3.5C, with Gly446 establishing polar contacts with the side chain of Gln11 and the main chain of Leu442. Additionally, the carbonyl group of Leu445 is hydrogen bonded to the side chain of Thr10. A glycerol molecule is likely to contribute to the stabilisation of this alternative conformation by interacting with Gly446 (Figure 3.5C). However, in glycerol-free crystals (Figure 3.1D) the C-terminus of GgH seems to adopt a single conformation similar to the conformation less frequently found in glycerol-containing crystals (Figure 3.5C and D). While the latter is stabilised by polar contacts with the glycerol molecule, in glycerol-free crystals the C-terminus is stabilised by a hydrogen bond with His129.

The biophysical analysis of the *Mha*GgH-His_6 variant pointed towards the formation of a dimer in solution, suggesting that the hexahistidine C-terminal tag interferes with tetramer formation by avoiding the interaction between dimers. Indeed, the space available at the centre of the tetramer is too small to accommodate four tags of 13 amino acids each (Table 3.1). Thus, the hexahistidine tags are likely to physically impair the contact between dimers. Also, the C-terminal tag may affect the hydrogen bond network between the C-terminus and helix α1 occurring between molecules A and C, and molecules B and D (Figure 3.5). However, the largest intermonomer contact area in the ga*Mha*GgH tetramer is found between these molecules, doubling the contact area observed between A-B or C-D dimers, and in which other interactions are also present (Supplemental table S3.1). Therefore, it is more likely that the *Mha*GgH-His_6 variant oligomerizes forming dimers equivalent to those corresponding to molecules A and C (or B and D) in the ga*Mha*GgH structure.

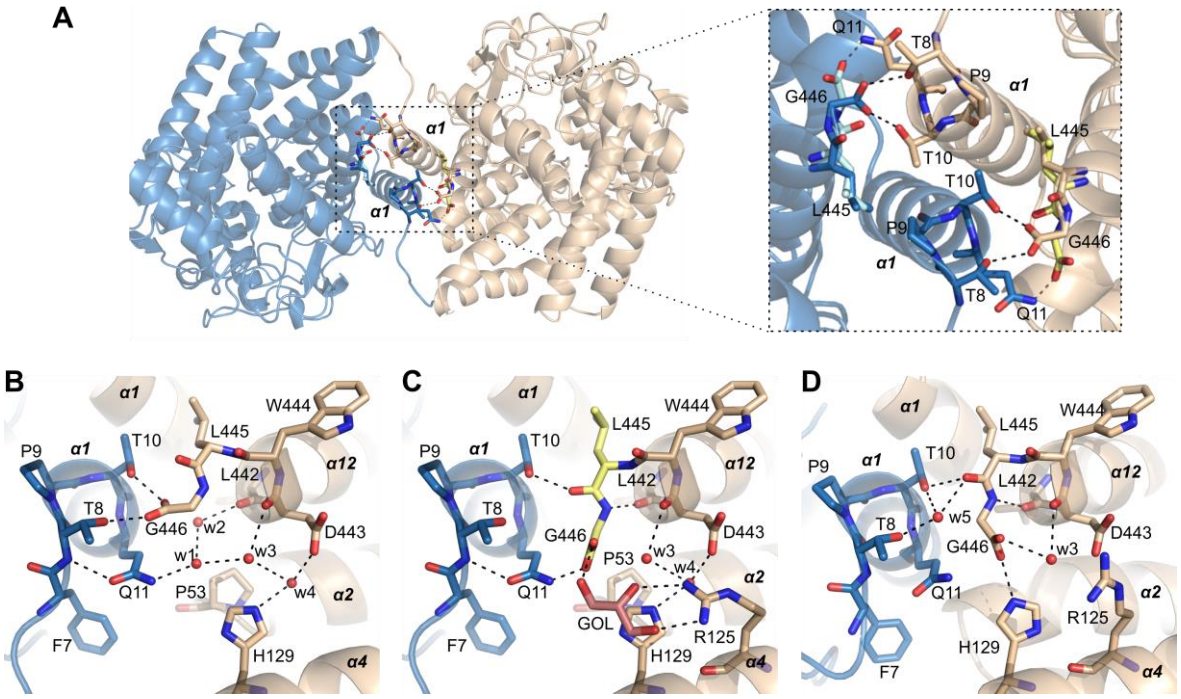


Figure 3.5 – Contribution of the C-terminus for oligomerization. A – Interface between monomers B (wheat) and D (blue), highlighting the polar contacts between the C-terminus of one monomer and helix α1 of the neighbouring monomer. The two alternative conformations of the C-terminus of molecules B and D are coloured in different shades of yellow and blue, respectively (inset). Hydrogen bonds are represented by dashed lines. B – Network of polar contacts at the C-terminus, with Gly446 stabilised by polar contacts with Thr8 and Thr10 of the neighbour monomer. Hydrogen bonds are represented by dashed lines. C – Alternative conformation of the C-terminus, stabilised by an intramolecular hydrogen bond between Leu442 and Gly446, and by intermolecular contacts between Leu445 and Thr10, and Gly446 and Glu11. A glycerol (GOL) molecule (salmon) is at hydrogen bonding distance of Gly446 and may contribute for the alternative conformation of the C-terminus. Hydrogen bonds are represented as in B. D – C-terminus of GgH in glycerol-free crystals, stabilised by intramolecular hydrogen bonds with His129, Leu442 and Asp443 and by intermolecular contacts between Leu445 and Thr8 and Thr10. Hydrogen bonds are represented as in B.

3.3.4. Biochemical characterization of *Mha*GgH variants

The biochemical characterization of the *Mha*GgH-His_6 variant was performed previously (Alarico, Costa et al. 2014). Unlike its bacterial homologues from *T. thermophilus* HB27 and *R. radiotolerans*, which display identical activity against GG and MG (Alarico, Empadinhas et al. 2013), *Mha*GgH-His_6 seems to be more efficient for GG than MG. Since the quaternary structure was affected by the presence of the hexahistidine tag, a new biochemical characterization was performed for the untagged variant.

The activity of ga*Mha*GgH against GG and MG was first evaluated by TLC (Figure 3.6), confirming the ability of this enzyme to hydrolyse both substrates. Despite lower efficiency for MG, the ga*Mha*GgH variant was able to produce more mannose compared to the *Mha*GgH-His_6 variant for a given reaction time. The increased production of mannose is probably associated to the higher stability of ga*Mha*GgH.

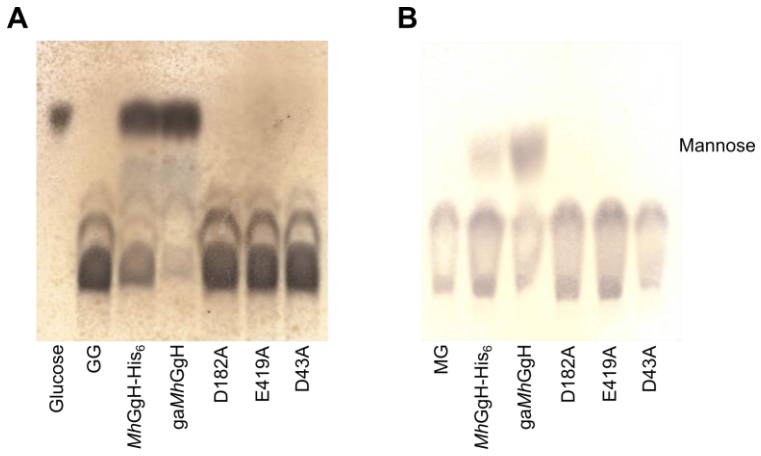


Figure 3.6 - Hydrolysis of GG and MG by *Mha*GgH variants. The different *Mha*GgH variants were incubated with GG (A) or MG (B) and their ability to hydrolyse the substrate was evaluated by TLC. Both *Mha*GgH-His_6 and ga*Mha*GgH variants were able to hydrolyse GG and MG, however a lower residual substrate amount was detected for the ga*Mha*GgH variant. None of the point mutants (D182A, E419A and D43A) produced a detectable amount of glucose or mannose.

Under the conditions tested, ga*Mha*GgH displays maximum activity at 55°C (Figure 3.7A), which is 13°C higher than the previously reported value for *Mha*GgH-His_6 variant (42°C) (Alarico, Costa et al. 2014), in agreement with the optimal temperature of growth of *M. hassiacum* (50°C) (Tiago, Maranha et al. 2012). The higher optimal temperature of activity of

ga*Mha*GgH variant reflects its higher thermal stability (Figure 3.3D), which in turns explains its longer half-life (Figure 3.7B).

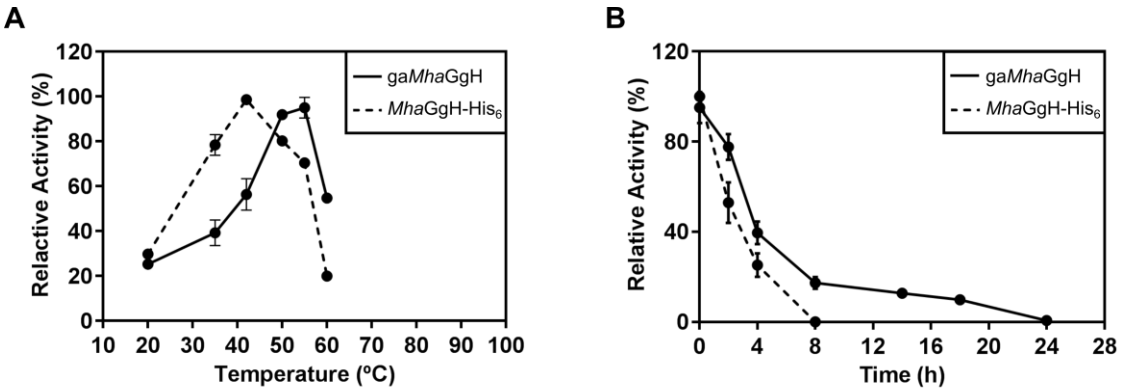


Figure 3.7 – Effect of temperature on the activity of ga*Mha*GgH and *Mha*GgH-His$_6$. A – The temperature profile was determined between 20 and 60°C. The optimal temperature of activity for ga*Mha*GgH (55°C) is significantly higher than that of *Mha*GgH-His$_6$ (42°C) (Alarico, Costa et al. 2014). B – Time-dependent activity profile of ga*Mha*GgH (55°C) and *Mha*GgH-His$_6$ (42°C), highlighting the higher resistance of the ga*Mha*GgH variant to thermal inactivation. (Experiments performed by Susana Alarico, Molecular Mycobacteriology Group, CNC, University of Coimbra, Portugal).

At the optimal temperature of activity, the ga*Mha*GgH variant displayed maximum activity at pH 6.0, which is very similar to the value reported for the *Mha*GgH-His$_6$ variant (pH 5.8) (Alarico, Costa et al. 2014). As also observed for the *Mha*GgH-His$_6$ variant, the activity of ga*Mha*GgH seems to be slightly increased by the presence of divalent cations, especially by Mg^{2+} (~25% increase). The activity of ga*Mha*GgH was maximal in the presence of 100 mM KCl, which allowed optimizing the buffer composition for the downstream biochemical analysis of ga*Mha*GgH activity.

The kinetics parameters for ga*Mha*GgH were determined at 50°C (Supplemental figure S3.5A and C). A more adequate fitting of the experimental data into an allosteric sigmoidal curve was observed for both *Mha*GgH-His$_6$ and ga*Mha*GgH variants, with a Hill coefficient higher than 1, which supports the hypothesis of positive cooperativity (Table 3.4). Moreover, the calculated Hill coefficient was higher for the ga*Mha*GgH variant than for *Mha*GgH-His$_6$, suggesting a higher cooperative effect in the tetramer compared to the dimer.

The kinetic values found for the hydrolysis of GG and MG by ga*Mha*GgH reflect an approximately 10-fold higher hydrolysis efficiency for GG. However, the kinetic curve obtained for MG is incomplete, resulting in a large standard-deviation for the value of K_{half},

and a larger range of MG concentrations would have to be tested. Nevertheless, the different efficiency for both substrates is clear. Additionally, an activity curve was also measured at 37°C and pH 8.0, where the reaction is slower (Supplemental figure S3.6B). The sigmoidal character of this curve is more evident, supporting the fitting of the data to an allosteric kinetic model.

Table 3.4 – Kinetic parameters of GG and MG hydrolysis at 50°C. Experimental data were analyzed using the allosteric kinetic model. The higher *h* value for the ga*Mha*GgH variant suggested an increased cooperative effect. A lower affinity for MG is expected due to the higher estimated K_{half} value.

Kinetic parameters	GG		MG
	*Mha*GgH-His_6	ga*Mha*GgH	ga*Mha*GgH
V_{max} (µmol/min mg protein)	9.34 ± 0.58	3.60 ± 0.18	3.09 ± 0.66
K_{half} (mM)	6.53 ± 0.76	9.36 ± 0.69	84.18 ± 30.27
h	1.44 ± 0.15	1.77 ± 0.20	1.29 ± 0.23
R^2	0.993	0.974	0.990

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

In the orthorhombic crystals, the ga*Mha*GgH molecules adopt a closed conformation, concomitant to the presence of two ligands – a molecule of glycerol and a molecule of L-serine, components of the crystallization buffer – at the active site (ga*Mha*GgH-SER-GOL). The glycerol molecule occupies subsite -1 and serine subsite +1 of the active site, inducing the closed state of ga*Mha*GgH and rendering them inaccessible to the solvent. Interestingly, the enzyme crystallized in the absence of serine also adopts a similar conformation, with subsite +1 occupied by water molecules (Supplemental figure S3.6).

In an alternative crystallization condition, ga*Mha*GgH crystallized in space group $P6_222$, with two protomers in the asymmetric unit (Supplemental figure S3.7). These two molecules correspond to dimer A-C and were modelled from Asp4 (molecule A) or Gly-1 (molecule C) to Gly446. The active site of gaMhaGgH contains only solvent (hence the name ga*Mha*GgH-Apo) and adopts an open conformation.

In the open conformation, the active site is accessible to the exterior through a negatively charged tunnel (Figure 3.8A). The tunnel is lined by the side chains of Gln115, Asp212, Gln215, Asp43, Ser214, Gln434, Tyr88 and the carbonyl groups of Trp177 and Gly180.

Interestingly, the putative catalytic residues, Asp182 and Glu419, are pointing away from the active site cavity. Asp182 is stabilised by polar contacts with the side chains of Tyr191 and Tyr225 and with the carbonyl group of Arg216 through a water molecule (Supplemental figure S3.8). The side chain of Glu419 is hydrogen bonded to the side chain of Ser435 and the amide nitrogens of Met432 and Thr437 (Supplemental figure S3.8).

When serine and glycerol molecules occupy the active site, the cavity completely closes to the solvent (Figure 3.8B). The ligands are stabilised and oriented mainly by hydrogen bonds with the binding site residues, and the putative catalytic residues are facing the lumen of the active site cavity.

The two structures of ga*Mha*GgH, corresponding to open and closed conformations, reveal the structural modifications occurring upon substrate binding (Figure 3.9). The active site is surrounded by mobile loops that allow active site exposure and guarantee a polar surface for substrate binding and accommodation. 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). The access to the active site is also 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 for the formation of an additional helix in this cap (residues 206-209) that is absent in the open conformation.

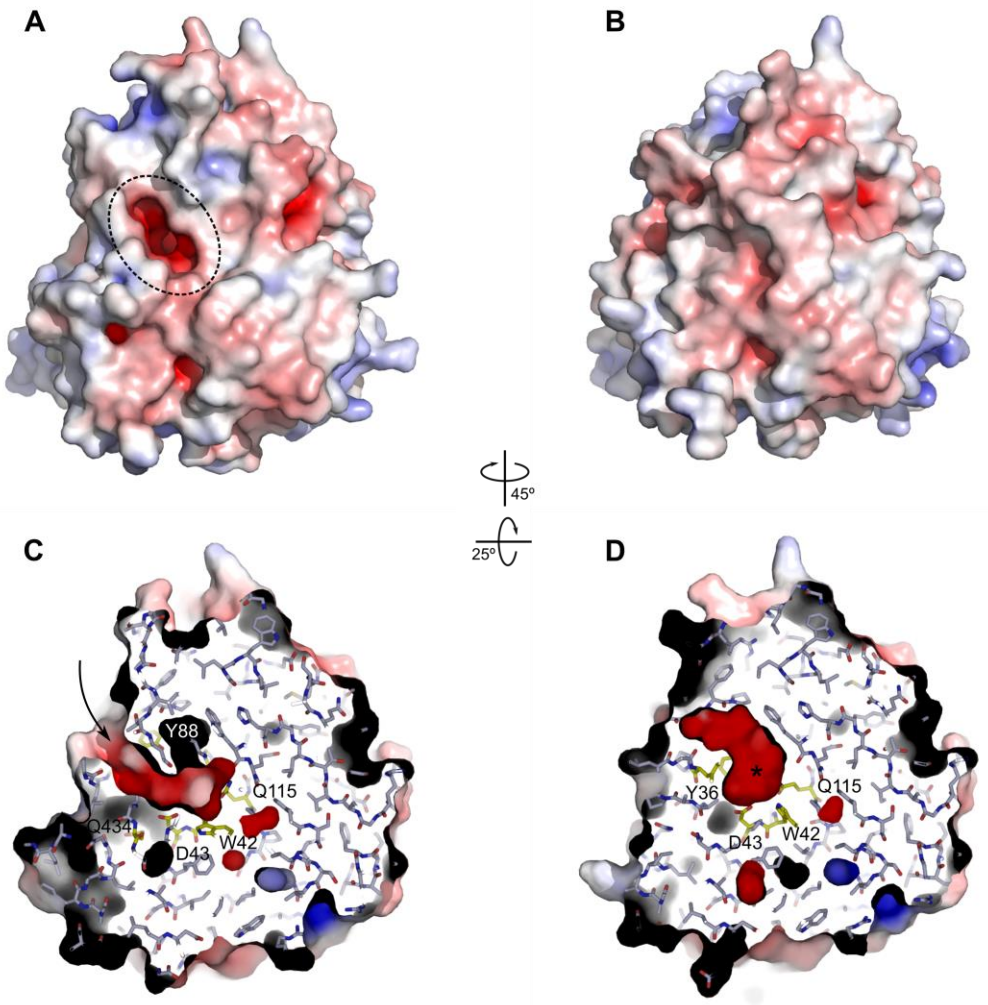


Figure 3.8 – Open and closed conformations of *Mha*GgH. A and B – Surface electrostatic potential (contoured from -8kT/e (red) to 8kT/e (blue)) of the *Mha*GgH monomer in open (A) and closed (B) conformations. In the open conformation, an opening leading to an acidic cavity is observed on the surface of the molecule (dashed ellipse), which is absent in the closed conformation state. C and D – Cross section of *Mha*GgH monomer in open (C) and closed (D) conformation. In the open state, a negatively charged tunnel (arrow) connects the active site cavity to the exterior of the molecule. In the closed state, the active site cavity (marked with an asterisk) becomes inaccessible to solvent. Substrate-binding residues are highlighted in yellow. The molecular surface was coloured by electrostatic potential as in A and B.

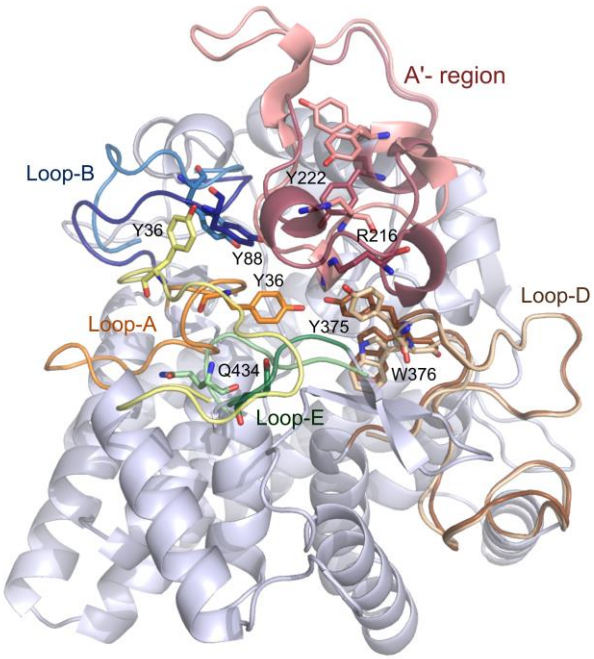


Figure 3.9 – Conformational changes in ga*Mha*GgH induced by substrate binding. Open (lighter hues) and closed (darker hues) states of monomeric ga*Mha*GgH are shown. The A'-region (flexible segment), and loops A, B, D and E are coloured salmon, blue, yellow, green and brown, respectively. Some of the substrate-interacting residues present in the highlighted regions [Tyr36 (loop A), Tyr88 (loop B), Arg216, Tyr222 (A'-region), Tyr375, Trp376 (loop D) and Gln434 (loop E)] are represented as sticks.

A particularly significant modification is observed in the segment Arg21-Ala31 (loop A) (Figure 3.9). In the open conformation, this loop displays higher mobility and covers the segment Ser431-Ser435 that contains the substrate-interacting residue Gln434. In the closed conformation, Gln434 faces the active site and binds to the substrate, while Gln433 moves outward (~6Å), forcing the segment Arg21-Ala31 to move away from its initial position. Moreover, upon substrate binding this segment is stabilised by polar contacts with the substrate *via* Tyr36 and with helices α3 and α12.

In the open conformation structure, the segment Arg21-Ala31 is localized close to a symmetry axis relating molecules A and C. The electron density map in this region suggests the presence of alternative conformations. Therefore, it is possible that this segment interacts with the equivalent region of the neighbour molecule, and thus impacts enzyme activity.

3.3.6. Molecular details of substrate and substrate-analogue binding

3.3.6.1. Inactive point mutants of ga*Mha*GgH

In order to understand the molecular determinants underlying substrate binding and specificity, crystals of wild-type ga*Mha*GgH were also obtained in the presence of its substrates (GG and MG), substrate analogues (GGlycerol and GGlycerate) and products of reaction (glucose, mannose and glycerate) or soaked into the crystallization condition containing these compounds. However, none of these ligands could be found in the active site of wild-type ga*Mha*GgH crystals. To avoid substrate hydrolysis during crystallization or soaking, three catalytically inactive point mutants were produced. The mutations were performed based on sequence alignment with other characterized MgHs and on the ga*Mha*GgH-SER-GOL ternary complex structure. Two putative catalytic residues (Asp182 and Glu419) and one substrate holder residue (Asp43) were identified (Suplemental figure S3.4) and replaced by alanine to produce three point mutants (D43A, D182A and E419A).

The ga*Mha*GgH point mutants were produced using the protocol developed for wild-type ga*Mha*GgH. The solubility of the mutants was very similar to that of the wild-type enzyme, except for the D182A mutant which was considerably less soluble. The thermal stability of these mutants was evaluated by differential scanning fluorimetry (Figure 3.3D) and no significant difference was observed, except for mutant D43A that displayed a 3°C lower T_m than wild-type ga*Mha*GgH.

The ability of the mutants to hydrolyse GG and MG was analysed by TLC (Figure 3.6). In the conditions tested, none of the mutants produced detectable amounts of glucose or mannose and thus were considered inactive.

3.3.6.2. Inactive point mutants in complex with substrates

The inactive point mutants produced were crystallized in the same condition as wild-type ga*Mha*GgH. Since the mutants were less crystallogenic, macroseeds of wild-type ga*Mha*GgH were used to accelerate crystallization. Crystals of the mutants were soaked into crystallization buffer containing GG or MG to yield binary complexes.

The crystals obtained belong to space group $P2_12_12$, diffract to 1.75-2.1 Å and contain 2 molecules per asymmetric unit, corresponding to dimer A-B of the tetramer (Table 3.2). Analysis of the active site of mutants D182A and E419A revealed the presence of residual

electron density, compatible with the substrates (Figure 3.10). However, for mutant D43A none of the substrates was found in the active site, independently on the soaking time, reflecting the important contribution of Asp43 for substrate binding. This residue is a substrate holder and establishes two hydrogen bonds with the glucose or mannose moiety of the substrate (Figure 3.11).

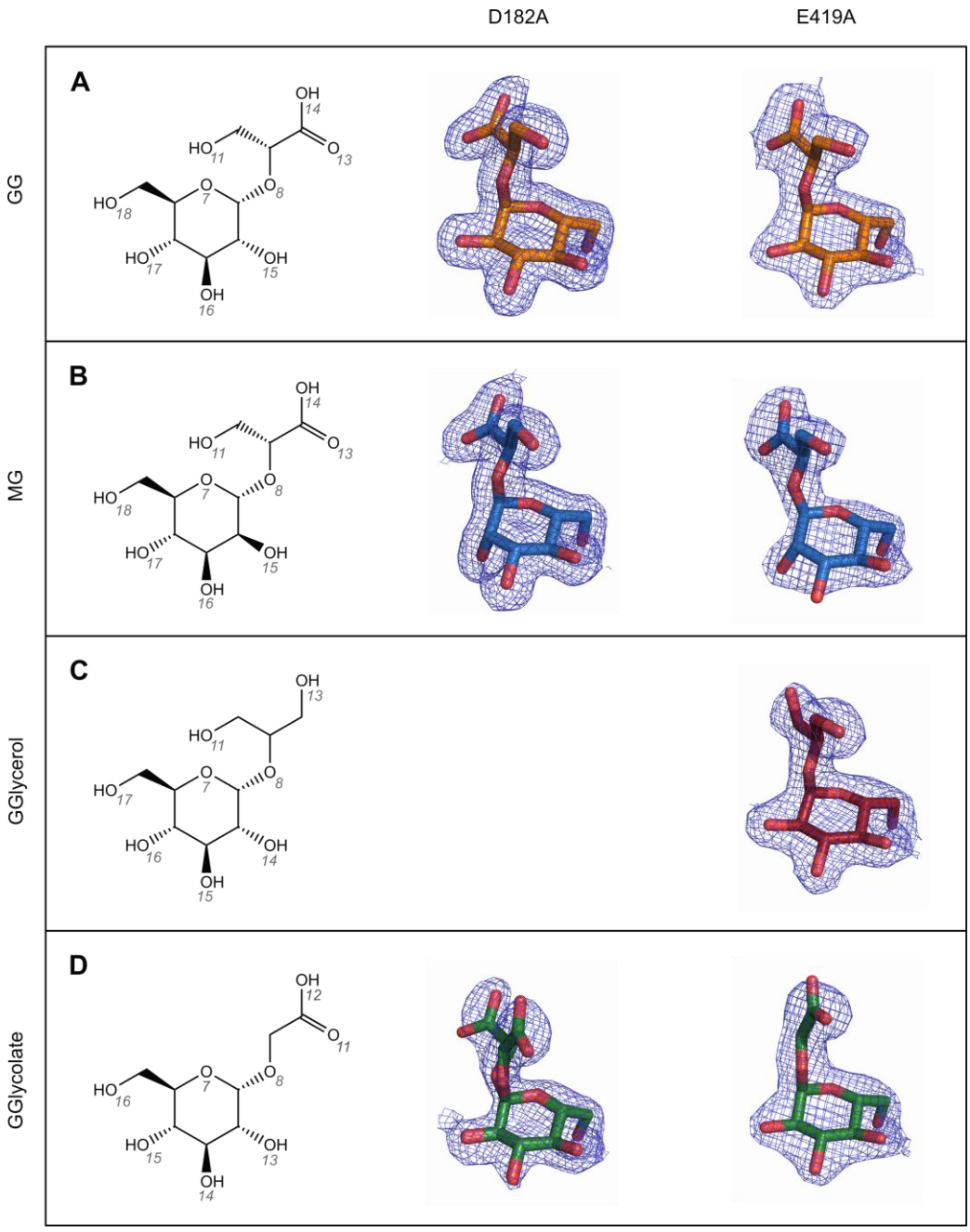


Figure 3.10 – Substrates and substrate analogues of *Mha*GgH. The substrate GG (A) and MG (B), and the substrate analogues GGlycerol (C) and GGlycolate (D) were soaked into the crystals of mutants D182A (middle) and E419A (right) to yield binary complexes. Chemical representation of the ligands is on the left. The oxygen atom numbering is indicated. The D182A-GGlycerol complex could not be obtained. The 2Fo-Fc electron density maps (blue) are contoured at 1 σ.

The ga*Mha*GgH D182A and E419A mutants in complex with each substrate were modelled from Pro2 to Gly446. No significant differences were observed in the active site of the D182A-GG and E419A-GG complexes (RMSD=0.177 Å, for 14 aligned atoms). In both

cases, the glucose and the glycerate moieties of GG occupy subsites -1 and +1 of the active site, respectively. The glucose moiety is stabilised mainly by hydrogen bonds with the side chains of residues Trp42, Asp43, Gln115, Asp182, Tyr375, Trp376 and Gln434 and with the carbonyl group of Gly180 (Figure 3.11A). The glucose moiety is also oriented by hydrophobic contacts with Tyr36, Trp40, Trp376, Trp436 and Trp381, and also interacts with Glu419 through a water-mediated contact. The glycerate moiety of the substrate is stabilised by hydrogen bonds with Trp40, Tyr88, Asp 182, Arg216 and Tyr375. In the E419A mutant, the Tyr36-to-glycerate distance is decreased (from 3.75 to 3.00 Å), suggesting that in this mutant Tyr36 may also establish a polar contact with the glycerate moiety. Water-mediated contacts with the side chains of Tyr222 and Trp177 also contribute for substrate binding, as do hydrophobic contacts with Trp177. In the case of mutant D182A, a water molecule is occupying the space of Asp182 side chain (w4 in Figure 3.11A), whereas in E419A the contacts established by Glu419 are mimicked by two water molecules. The substrate-interacting residues in the GG-mutants complexes are organized in a very similar way to that observed for the ga*Mha*GgH-SER-GOL (RMSD=0.191 Å, for 14 aligned atoms). The serine and glycerol molecules present in the active site of the ga*Mha*GgH-SER-GOL structure partially mimic a substrate, with serine superposing nicely to the glycerate moiety and the glycerol establishing the two hydrogen bonds with the substrate holder residue (Asp43) (Figure 3.11B). In the high resolution structure of wild-type ga*Mha*GgH, it is clear that bound serine can adopt more than one conformation. The conformation more frequently found (57% occupancy) is the one that better superposes to the glycerate moiety in the other complexes. Altogether, these results suggest that the GG-mutants and ga*Mha*GgH-SER-GOL complexes are likely to represent the substrate-binding mode of GG for active ga*Mha*GgH.

The active site of both mutants is also very similar when MG is present (RMSD=0.106 Å, for 10 aligned atoms) (Figure 3.11C). However, a better superposition is observed for the active sites containing GG, independently on the catalytic residue mutated. On the other hand, the worse superposition of the active sites containing MG, suggests that each active site undergoes a distinct rearrangement to accommodate this substrate. Due to the lower plasticity of the active site of wild-type ga*Mha*GgH, the binding of MG may be energetically more demanding than GG. Nevertheless, a similar hydrogen bonding network seems to be established for both substrates (Figure 3.11A and C), with the most significant difference observed for atom O15 of the substrate. In the MG-E419A mutant, O15 loses the hydrogen bond with Asp182 and Tyr375, and becomes at hydrogen bonding distance of the putative nucleophilic water (Figure 3.11C).

In the active site, the glycosidic oxygen of both substrates is within hydrogen bonding distance of the catalytic residue Asp182, and Glu419 establishes a water-mediated contact

with substrates. The distances between the catalytic residues and substrates, and the presence of a single water mediating the Glu419-substrate contact suggest that *Mha*GgH is likely to hydrolyse GG and MG using an inverting mechanism that requires Asp182, Glu419 and a water molecule, as acid, base and nucleophile, respectively (Figure 3.11D). During hydrolysis, the negatively charged Glu419 is likely to activate the water molecule that performs a nucleophilic attack on the anomeric carbon, while Asp182 donates a proton to the leaving glycerate. Inversion of the anomeric configuration of the glucose is expected. In a previous work, hydrolysis of MG by an inverting mechanism was confirmed for the homologous MgHs enzymes from *T. thermophilus* HB27 and *R. radiotolerans* (Alarico, Empadinhas et al. 2013). The ratio between α and β configurations of the mannose produced during MG hydrolysis was determined by NMR and found to be higher than the standard α:β-equilibrium ratio of the mannose anomers in solution. Since the α:β-equilibrium ratio of glucose ($\Delta G^{o}_{\alpha\rightarrow\beta}$ = -0.35 kcal/mol) is reached faster than for mannose ($\Delta G^{o}_{\alpha\rightarrow\beta}$ = 0.45 kcal/mol), the possible shift occurring upon hydrolysis may be difficult to detect (Miljković 2010).

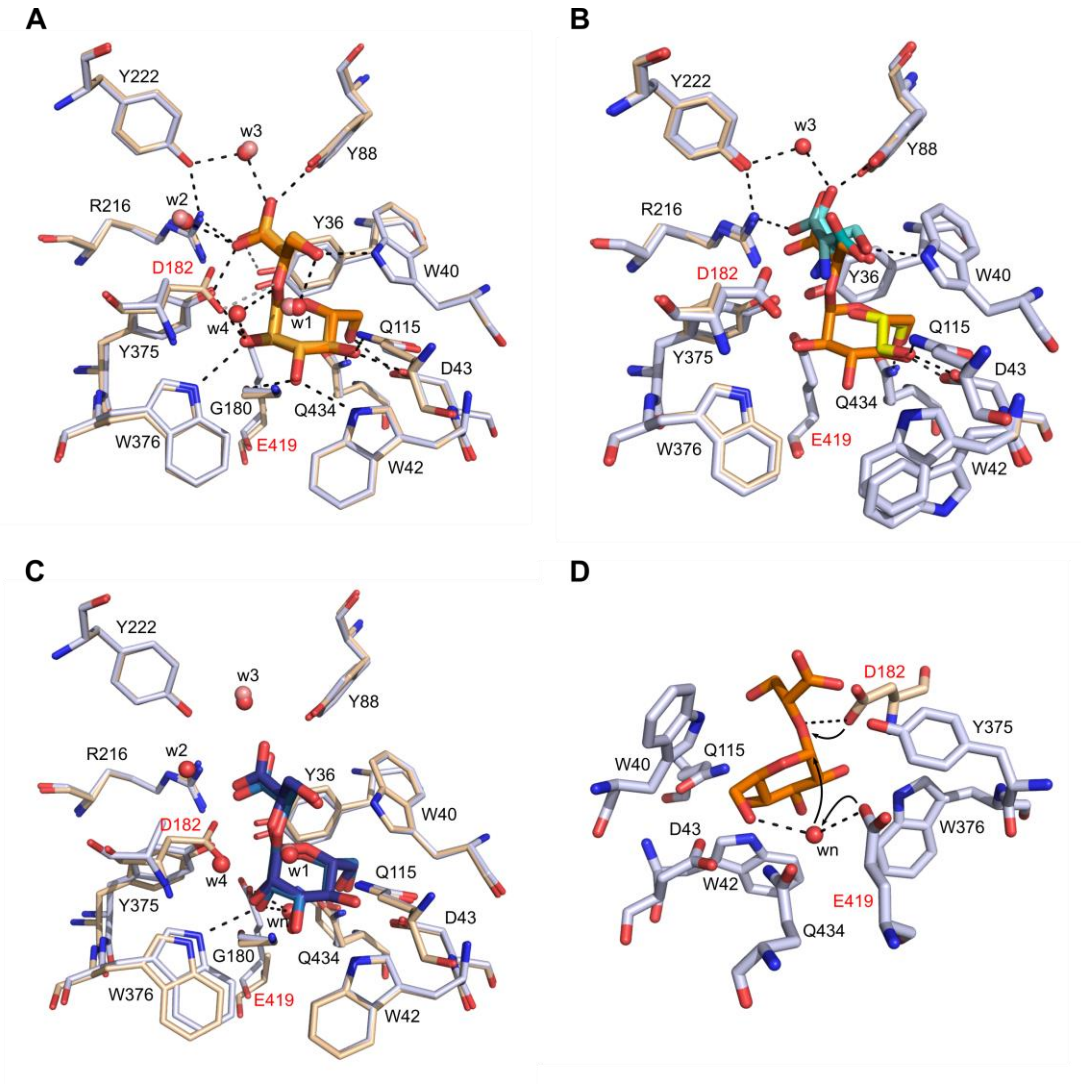


Figure 3.11 – The active site of ga*Mha*GgH variants in complex with substrates. A – Superposition of the active site region of ga*Mha*GgH mutants D182A and E419A in complex with GG. The position of GG (dark or light orange for D182A or E419A, respectively) in the active site of D182A (light blue) and E419A (wheat) is stabilised mainly by direct hydrogen bonds or by water-mediated contacts (dashed lines) with the side chains of the labelled residues. The polar contacts between GG and mutant D182A residues are represented by black dashed lines, while the alternative contacts observed in mutant E419A are coloured grey. Water molecules (labelled w1 to w4) of mutants D182A or E419A are represented by red or salmon spheres, respectively. Catalytic residues (D182 and E419) are highlighted in red. B – Superposition of the active site region of the ga*Mha*GgH D182A-GG complex (light blue with ligand in orange) to that of the ga*Mha*GgH-SER-GOL ternary complex (wheat). The ligands serine (cyan) and glycerol (yellow) mimic the substrate, allowing closure and stabilisation of the active site as observed for substrate-mutant complexes. The serine molecule may adopt two positions (coloured in different shades of cyan), with the darker shade of cyan indicating the more frequent position. Hydrogen bonds between glycerol or serine (dark cyan) and the residues of the active site are represented by dashed lines. C – Superposition of the active site region of ga*Mha*GgH

mutants D182A and E419A in complex with MG. The hydrogen bonding network involved in stabilising MG in the active site of the ga*Mha*GgH variants 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 are coloured as in A, highlighting the absence of w2 and nucleophilic water (wn) in the E419A-MG complex. D – Active site region of ga*Mha*GgH in complex with GG. The position of GG (orange) relative to the catalytic residues and the nucleophilic water are shown. The proposed mechanism of hydrolysis is illustrated by the arrows.

3.3.6.2.1. Binding affinity of the substrates GG and MG

The difference in hydrolysis efficiency of *Mha*GgH for GG and MG may be a consequence of different binding affinities. Thus, the binding affinities of the substrates were also analysed. Given the limited availability of the synthetic substrates, the binding affinities of the substrates were evaluated by MST at the Structural Biology & Genomics platform of the Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC) (Strasbourg, France), to which access was provided by Instruct (www.structuralbiology.eu).

Determining the binding affinity of a substrate to its processing enzyme by MST may be complicated by the formation of the reaction products. Measurements performed on an ongoing enzymatic reaction can yield hard to interpret results, arising from all possible complexation states: from free enzyme to enzyme-substrate and enzyme-product complexes. In order to circumvent this problem, the two mutants for the catalytic residues were used. As demonstrated by TLC, D182A and E419A are catalytically inactive against GG and MG, but still able to bind to the substrates, as indicated by the three-dimensional structures of mutant-substrate complexes. Also, the analysis of both mutants can allow the identification of the individual contribution of the residues for substrate binding.

The determination of binding affinities in a Monolith NT.115 equipment requires proteins to be labelled. The *Mha*GgH mutants were labelled with *N*-hydroxysuccinimide-ester dye that reacts efficiently with primary amines, forming highly stable dye-protein conjugates. Primary amines are found in lysines, which are usually solvent accessible and therefore suitable for labelling reactions, and at the protein N-terminus. Recombinant ga*Mha*GgH contains seven lysine residues per monomer and all are at the molecule surface and accessible to the solvent. Moreover, the N-terminus is disordered in the three-dimensional structures, suggesting that it is also prone to be labelled. Altogether, thirty two amines per ga*Mha*GgH molecule are likely to be labelled.

The binding affinity of the mutants for the substrates was evaluated based on changes in the thermophoresis (particle movement in a temperature gradient) of fluorescently labelled-proteins in presence of different substrate concentrations. Because low affinity was expected, the substrate concentration was increased to the highest achievable value with the existing stock solution.

Three independent measurements were performed for mutant D182A, whereas for E419A only two measurements were possible in the allocated time slot for the equipment. The D182A mutant displayed higher affinity for both substrates than variant E419A (Figure 3.12). This difference is likely related to mutated residue position and the corresponding interactions. Asp182 is localized deeply in the active site, establishing contacts almost exclusively with the substrate. Mutation of this residue does not affect the position of the other substrate-interacting residues. On the other hand, Glu419 is localized in a mobile segment and stabilised by polar contacts with the backbone amine of Ser431, the side chain of Tyr36 and through a solvent molecule with Arg216 and Tyr420. These interactions are important for active site closure and substrate-enzyme complex stabilisation. The substitution of Glu419 for an alanine is accompanied by the loss of this network of polar contacts and thus substrate binding is severely affected.

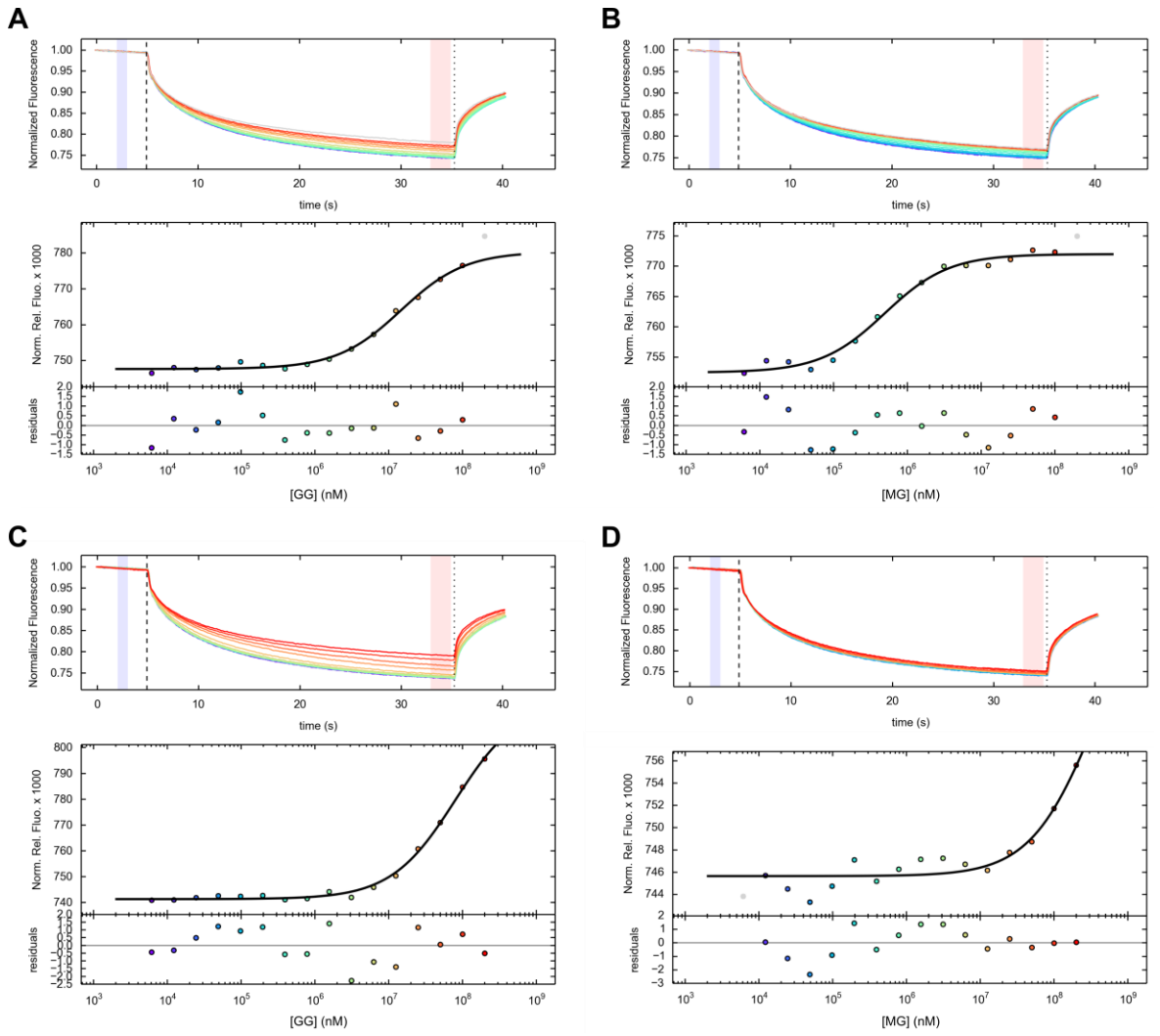


Figure 3.12 – Effect of GG and MG in the thermophoretic movement of catalytically inactive ga*Mha*GgH mutants. The binding affinities of D182A (A and B) and E419A (C and D) for GG (A and C) or MG (B and D) were analysed by microscale thermophoresis. The thermophoresis of the fluorescent protein in presence of different substrate concentrations (increasing concentrations are coloured blue to red) through a temperature gradient was detected and quantified (upper graphic). The microscopic temperature gradient is induced by turning on (dashed line) and off (dotted line) an infrared laser. The fluorescence average before (light blue area) and after heating (light pink area) was calculated and plotted against substrate to yield the dose-response curve (lower graphic). Sigmoid curves were obtained for the titration of D182A with GG (A) or MG (B), which allows the determination of the binding affinity of the interaction. The binding curve obtained for E419A with GG (C) or MG (D) lacks the plateau of the high ligand dose and the binding affinity cannot be determined.

Table 3.5 – Dissociation constant for GG and MG.

	K_D GG (mM)	K_D MG (mM)
D182A	15.33 ± 1.53	0.35 ± 0.13
E419A	≥ 80	≥ 300

The ga*Mha*GgH D182A mutant displayed an approximately 100-fold lower K_D for MG (~0.35 mM) than for GG (~15 mM) (Table 3.5). Structurally, the active site of mutant D182A containing MG or GG is similar (RMSD=0.065 Å, for 12 aligned atoms). The hydrogen bonds observed between MG and residues Tyr36, Trp40, Trp42, Asp43, Tyr88, Gln115, Arg216, Tyr222 and Gln434 are also present in the GG-mutant complex with identical distances (Figure 3.13A). The major differences in the active site of the complexes are induced by the position of the substrate O15 atom. Atom O15 of GG establishes a larger number of contacts compared to the same atom in MG. In GG, O15 is hydrogen bonded to Tyr375, Trp376 and water molecule w4, while in MG O15 loses all but the interaction with Trp376 and establishes a new polar contact with the nucleophilic water. Although conserved, the distance between O15 and Trp376 is ~0.2 Å longer in the MG-mutant complex ($DPI_{GG\text{-}D182A\ complex}$=0.072 Å; $DPI_{MG\text{-}D182A\ complex}$=0.083 Å). In consequence, other structural modifications in the active site are also observed. In the MG-D182A complex, Tyr375 is bonded to water w4, inducing a ~1.1 Å shift in Tyr375 position compared to the GG-D182A complex. Also, a ~0.3 Å inward movement of the segment Arg371-Arg377 is noticed. The loss of the hydrogen bonding with water w4 in the MG complex is also accompanied by a ~0.9 Å backward movement of the Ala182 α-carbon. While in GG-D182A complex the carbonyl group of Ala182 is bonded to Tyr191 by a water mediated contact, in the MG-D182A complex the backward movement of the Ala182 allows formation of a direct hydrogen bond with Tyr191.

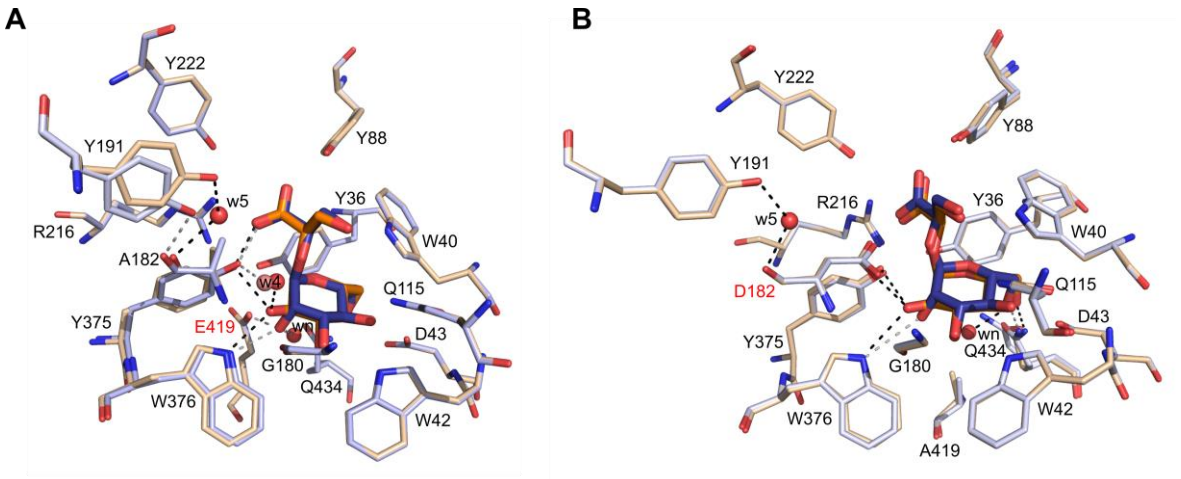


Figure 3.13 – Active site of ga*Mha*GgH D182A and E419A mutants in complex with GG or MG. A – Superposition of the active site of mutant D182A in complex with GG (orange with residues in wheat) or MG (blue with residues in light blue). The hydrogen bonds between the active site residues and GG or MG are represented as black or grey dashed lines, respectively. B – Superposition of the active site of mutant E419A in complex with GG (orange with residues in wheat) or MG (blue with residues in light blue). The hydrogen bonds are represented as in A.

The ga*Mha*GgH E419A mutant displayed a considerable lower affinity for both substrates compared to the D182A mutant. Despite using the same substrate concentration range as for D182A, the plateau corresponding to the substrate-mutant complex was not reached and higher substrate concentrations would be necessary. Since a typical sigmoidal binding curve was not obtained, the K_D values of GG or MG binding to E419A could not be determined. Nevertheless, the shape of the binding curve suggests higher affinity for GG than for MG (Figure 3.12C and D). The data obtained suggest that E419A-GG and E419A-MG interactions are likely to have K_D values higher than 80 and 300 mM, respectively (Table 3.5).

This difference in affinity is also reflected in the lower occupancy of MG in the active site of the E419A mutant. Despite MG being present in both molecules of the asymmetric unit, in one of them its occupancy refined to a value of ~42%, co-existing with glycerol and serine in the remaining molecules.

Structurally, the active site of E419A in complex with GG or MG is identical (RMSD=0.08 Å, for 12 aligned atoms), with the active site residues occupying similar positions (Figure 3.13B). However, the superposition of GG and MG is clearly worse than for mutant D182A-substrate complexes. The hydrogen bonds with Tyr36, Asp43, Trp40, Trp42, Tyr88, Arg216, Tyr222 and Gln115 are present for both substrates. The most significant changes are

induced by atom O15 of the substrate. In the case of the E419A-GG complex, O15 is able to establish polar contacts with Asp182, Tyr375 and Trp376, whereas in the E419A-MG complex the hydrogen bond to Trp376 is the only one preserved, with an increase of ~0.3 Å in length. In contrast to the E419A-GG complex, the nucleophilic water is absent when MG is in the active site of the E419A mutant. Thus, residues Asp182, Tyr375 and Trp376 are involved in the specific recognition of glucose moiety of GG. Since all these residues are preserved in the E419A mutant, it is more likely to reflect the different binding affinity of wild-type ga*Mha*GgH for GG and MG.

In mutant D182A, the substrates are identically oriented because Ala182 can shift ~0.9 Å allowing the movement of Tyr375. In mutant E418A, the position of Tyr375 is not allowed to change due to the presence of the side chain of Asp182. This residue is stabilised by water-mediated contacts with Gln115 and Trp177 in the E419A-GG complex, while in the E419A-MG complex the water mediating the contact to Trp177 is absent, and Asp182 is directly bound to the glycosidic oxygen of MG. In the E419A-MG complex, MG O18 is shifted ~0.4 Å, impairing hydrogen bonding to Gln434 and the nucleophilic water. Indeed, the nucleophilic water is absent in the MG-E419A complex and Gln434 adopts two conformations (~50% occupancy each). Since Glu419 contributes for Gln434 orientation through hydrophobic contacts, its absence allows Gln434 to establish new water-mediated contacts with the amide of Ser431.
A Cremer-Pople analysis (Cremer-Pople parameter calculator, http://enzyme13.bt.a.u-tokyo.ac.jp/CP/) was also performed for the cyclic moiety of GG and MG present in the active site of the mutants. Independently of the substrate or mutant, the moieties of glucose or mannose adopt a $^4C_1$ chair conformation with an α-anomeric configuration.

3.3.6.3. Inactive mutants in complex with substrate analogues

The activity of ga*Mha*GgH was also tested against the substrate analogues GGlycerol and GGlycolate. These compounds are composed of glucose and glycerol or glycolate moieties. Despite their similarity to GG, ga*Mha*GgH was unable to hydrolyse these compounds (Supplemental figure S3.9). In order to clarify the molecular determinants impairing the hydrolysis of these compounds, GGlycerol and GGlycolate were soaked into crystals of wild-type ga*Mha*GgH, and of mutants D182A and E419A.

Only the structure of the E419A mutant in complex with both substrate analogues and that of D182A in complex with GGlycolate could be obtained (Figure 3.10). Similar to the substrate-

mutant complexes, these structures were obtained from crystals belonging to space group $P2_12_12$ that diffract to between 1.9 and 2.1 Å and contain 2 molecules in the asymmetric unit Table 3.3). The structures of E419A-GGlycolate, E419A-GGlycerol and D182A-GGlycolate complexes were modelled from Pro2 to Gly446.

Comparing the active site of E419A in complex with GGlycolate or GG (RMSD=0.357 Å, for 13 aligned atoms), no significant differences are observed in the residues interacting with the glucose moiety (Figure 3.14A). In the E419A-GGlycolate complex, the polar contacts with Tyr36, Arg216, Tyr222 and Tyr375 are absent. Therefore, the Glu201-Ser218 segment has higher mobility, resulting in a poorly defined electron density map in this region. In molecule B, the segment Gln203-Ile210 was modelled from residual electron density and the occupancy of the side chains of these residues was set to zero. In the E419A-GGlycolate complex, the glycolate moiety is stabilised by hydrogen bonds to Trp40 and Tyr88, and by a water-mediated contact with Gln115 and Trp177. Interestingly, the catalytic water is absent in the E419A-GGlycolate complex. Although GGlycolate is present in the active site of both molecules of the asymmetric unit, in molecule A it displays an occupancy of ~36%, alternating with serine and glycerol. As described before, the presence of these compounds in the active site contributes for active site closure by mimicking the substrate. In this molecule, the residues known to interact with the substrate can be nicely superposed to the E419A-GG complex. Due to the high occupancy of the serine and glycerol over the GGlycolate, the active site arrangement may not represent a realistic model of GGlycolate binding. Indeed, important contacts for the stabilisation of the closed state, such as substrate binding to Arg216 and Tyr375, are absent when GGlycolate is present.

In the case of mutant D182A, GGlycolate is present in the active sites of both molecules in the asymmetric unit, although with partial occupancy. In molecule A, GGlycolate seems to adopt two different positions (~33 and 23% occupancy), whereas in molecule B a single conformation is present (~44% occupancy) (Figure 3.14B), in both cases alternating with serine and glycerol molecules. In mutant D182A, GGlycolate can adopt an identical position to that in mutant E419A, with the glycolate moiety establishing contacts with Trp40, Tyr88, Gln115 and Trp177. Nevertheless, in this mutant GGlycolate seems to preferentially adopt a different position, stabilised by hydrogen bonds to Tyr88, Arg216, Tyr222 and Tyr375. Also, the glycosidic oxygen is likely to establish a polar contact with water molecule w4, contributing for GGlycolate stabilisation. The presence of Glu419 contributes for orienting GGlycolate into the most frequent position through hydrogen bonding with Tyr36 and Arg216. The correct positioning of Arg216 also orients Tyr375 to face the active site. In mutant D182A, the most frequent conformation of GGlycolate is a close mimic of the ga*Mha*GgH substrate.

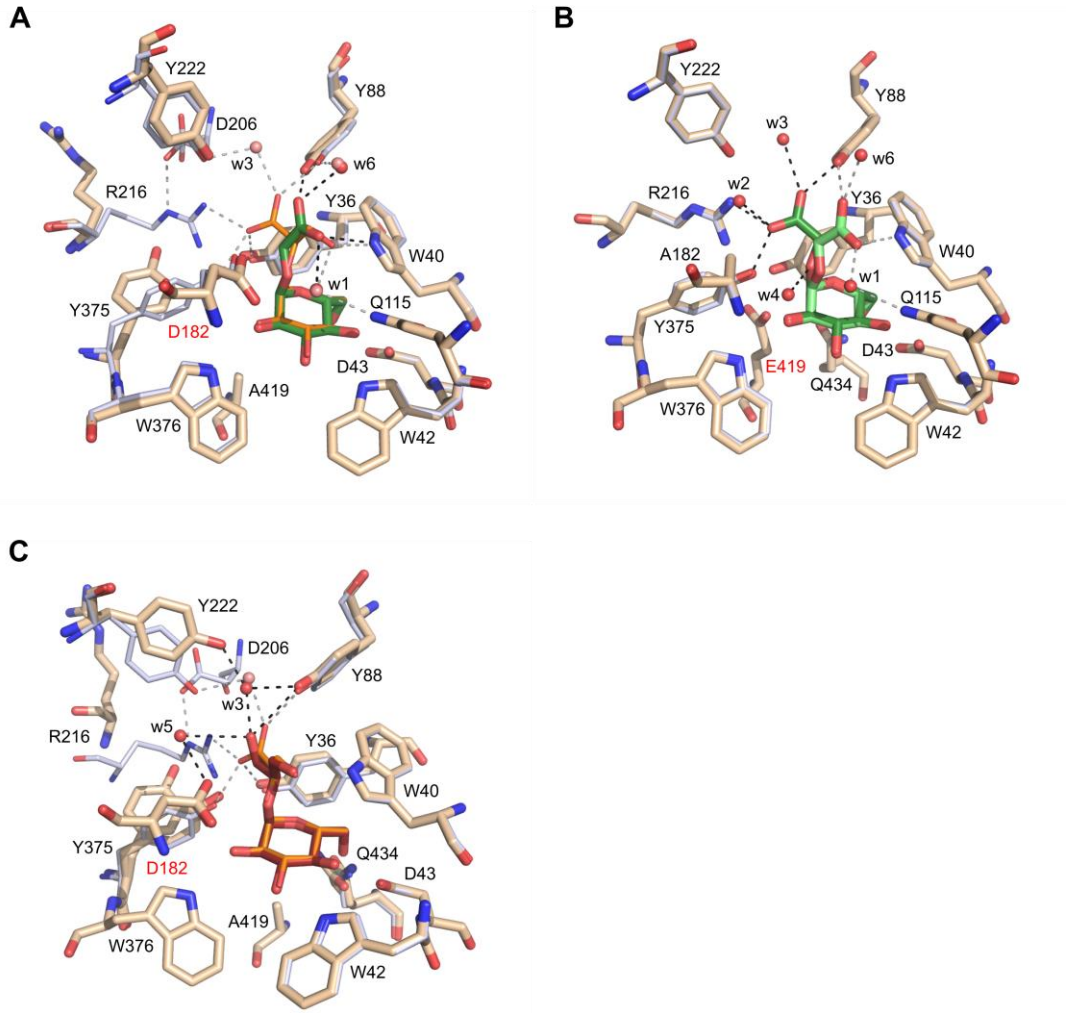


Figure 3.14 – Alternative ga*Mha*GgH ligands: GGlycolate and GGlycerol. A – Superposition of the active site region of ga*Mha*GgH mutant E419A in complex with GGlycolate (ligand and interacting residues are in green and wheat, respectively) and with GG (ligand and interacting residues are in orange and light blue, respectively). The grey dashed lines represent the contacts between mutant E419A residues and GG that are absent in the E419A-GGlycerol complex. Black dashed lines represent newly established contacts. B – Superposition of the active site region of ga*Mha*GgH mutant D182A in complex with GGlycolate (ligand and interacting residues are in green and wheat, respectively) and with GG (interacting residues are in light blue; GG is not shown). The bound GGlycolate molecule adopts two conformations (dark and light green) with the most frequent one coloured dark green. Polar contacts are represented as grey or black dashed lines depending on GGlycolate conformation. C – Superposition of the active site region of mutant E419A in complex with GGlycerol (ligand and interacting residues are in dark red and wheat, respectively) and with GG (ligand and interacting residues are in orange and light blue, respectively). The hydrogen bonds between mutant E419A and GG absent in the E419A-GGlycerol complex are represented by grey dashed lines. The new contacts are represented by black dashed lines.

In the structure of the E419-GGlycerol complex, segments Asp206-Ile210 of molecule A and Gln203-Ile210 of molecule B were not modelled, due to the absence of interpretable electron density in these regions. GGlycerol was found in both molecules of the asymmetric unit, seemingly with full occupancy. The residues of subsite -1 are in a similar position to that found in the E419A-GG complex, establishing identical contacts with the glucose moiety (Figure 3.14C). A single exception is observed for the position of Gln434 that can adopt two different conformations in the E419A-GGlycerol complex. The most significant differences are in subsite +1. The absence of one oxygen atom in the glycerol moiety compared to glycerate affects hydrogen bonding to Tyr36, Arg216, and Tyr375. Since Arg216 is not stabilised by contacts with the substrate, the hydrogen bond between Arg216 and Asp206 is also lost, resulting in higher mobility of the segment Glu201-Ser218. The glycerol moiety of GGlycerol is poorly stabilised by polar contacts with Tyr88, and with Asp182 and Tyr222 by water-mediated contacts. Nevertheless, the hydrogen bonds with Trp40, and Gln115 through a water molecule are still conserved in the reducing end of GGlycerol.

Several contacts are absent in the E419A-GGlycerol complex, especially those implicated in active site closure. The replacement of Asp182 by an alanine in the D182A mutant represents an additional lost bond that is likely to contribute for the reduction of GGlycerol binding. Moreover, the loss of a hydrogen bond with Tyr375 may destabilise several interactions necessary to properly orient the substrate and the catalytic residue Glu419 in the D182A mutant. Indeed, serine and glycerol were always found in the active site of the D182A mutant, regardless of the GGlycerol concentration used (up to 100mM) or the time of soaking. The ternary complex of D182A, serine and glycerol is likely more stable than the D182A-GGlycerol complex.

Despite all attempts to obtain the structure of the wild-type ga*Mha*GgH in complex with these substrate analogues, the active site of ga*Mha*GgH was always occupied by serine and glycerol. In the mutants, replacement of the catalytic residues by a small alanine allowed the rearrangement of the active site to accommodate the substrates, GG and MG, and likely the substrate analogues, GGlycerol and GGlycolate. Since a lower plasticity of the active site of wild-type ga*Mha*GgH is expected, the ligands that can be accepted are more limited. Substrate binding in *Mha*GgH is a well-coordinated event that involves flexible loops and highly conserved residues. This enzyme evolved to harbour specific substrates, especially GG. The presence of glycerol and serine in the active site of wild-type ga*Mha*GgH indicates a higher affinity for these molecules than for the substrate analogues.

3.3.7. Modulation of *Mha*GgH substrate specificity: an attempt

The structure of a *Mha*GgH homologue from *T. thermophilus* HB8 (Tt8MGH) was published recently (Miyazaki, Ichikawa et al. 2015). Although no biochemical study was performed for Tt8MGH, an enzyme with 99% amino acid sequence identity (Tt27MGH) was previously characterized (Alarico, Empadinhas et al. 2013). Tt27MGH is able to hydrolyse GG and MG with comparable efficiency. The high degree of amino acid sequence conservation between Tt8MGH and Tt27MGH suggests that the enzymatic activity is also conserved. Thus, the differences in *Mha*GgH and Tt27MGH active sites are likely to be responsible for the discrepancy in substrate preference.

A single substitution in the substrate-interacting residues is present in Tt8MGH and *Mha*GgH. While in *Mha*GgH position 434 (*Mha*GgH numbering) is occupied by a glutamine, in Tt8MGH this residue is a phenylalanine (Supplemental figure S3.4). As described above, residue Gln434 is localized in subsite -1 of the *Mha*GgH active site that hosts the glucose or mannose moiety of GG or MG, respectively. Therefore, it is likely that this substitution might contribute for substrate preference.

In order to evaluate the impact of the substitution of Gln434 for a phenylalanine in substrate hydrolysis by ga*Mha*GgH, a Q434F mutant was produced. The thermal stability of the Q434F mutant was evaluated and no significant differences were observed compared to the wild-type variant, suggesting that the mutation is unlikely to affect the protein folding and stability (Figure 3.3D).

The impact of the glutamine-to-phenylalanine substitution in ga*Mha*GgH activity was also analysed. Under the conditions tested, the mutant displayed considerable lower activity than wild-type ga*Mha*GgH (Supplemental figure S3.10). Also, no alteration in substrate preference was noticed, and apparently mutant Q434F is still more efficient hydrolysing GG than MG.

The Q434F mutant was crystallized in the same experimental conditions as ga*Mha*GgH and both open and closed conformations were obtained. While the crystals containing the enzyme in open conformation were easily obtained, the crystals of the closed conformation were only obtained by cross-seeding with wild-type crystal seeds. Moreover, the crystals of closed Q434F displayed a slower and poorer growth compared to the other mutants. X-ray diffraction data were collected from both types of crystals and the structures were solved by molecular replacement using the SeMet-ga*Mha*GgH coordinates as search model.

The active site residues of the ga*Mha*GgH Q434F mutant display an arrangement similar to that of wild-type ga*Mha*GgH or mutant D182A in complex with GG (Figure 3.15A). However,

a severe alteration is observed in the position of Trp436 (Figure 3.15B). This residue, together with Trp42, Trp381 and Trp376, contributes for substrate orientation through hydrophobic contacts. The movement of Trp436 alters the hydrophobic surface that accommodates the glucose ring, affecting substrate binding. Moreover, since the crystals corresponding to the closed conformation of the enzyme were difficult to grow, it is likely that replacement of Gln434 by a phenylalanine affects active site closure.

The position of the active site residues of Tt8MGH in complex with glucose (PDB accession code: 4WVB) is also comparable to that of ga*Mha*GgH Q434F (Figure 3.15C). However, when analysing the neighbourhood of Trp436 in wild-type ga*Mha*GgH, it becomes clear why there is not a phenylalanine in position 434 of ga*Mha*GgH (Figure 3.15D). In Tt8MGH, Glu393, Ile356 and Ser411 residues are all connected by a water-mediated hydrogen bonding network. Also, Trp410 is in an adequate position to establish an additional hydrogen bond with the same water molecule (labelled w in Figure 3.15D). In ga*Mha*GgH, a similar interaction is observed, with the catalytic Glu419 hydrogen bound via a water molecule to Thr437 and Val383. However, the substitution of isoleucine for a less flexible proline in ga*Mha*GgH forces the following residue (Val383) to interact with the water molecule. In consequence, Trp436 is repelled by the approximation of Val383 and moves towards Gln434. This movement is allowed due to the presence of glutamine, less bulky than phenylalanine. When residue 434 is replaced by a phenylalanine in ga*Mha*GgH, Trp436 is pushed away from both Phe434 and Val383. This movement also impacts the neighbouring residues, such as Phe46, slightly affecting their position.

In *Mha*GgH, the substitution of a glutamine for a phenylalanine represents a dramatic change. Besides their chemical differences, the residue size has also an impact in the restricted space of the closed active site of *Mha*GgH. When Gln434 is replaced by a phenylalanine, it interferes with the active site arrangement and/or substrate binding and orientation, and ultimately enzyme function. During evolution several amino acid substitutions are likely to occur. While some modifications are structural and biochemically well tolerated, others require additional substitutions to minimize their impact. In *Mha*GgH, Gln434 is well tolerated due to the presence of proline and valine at positions 382 and 383, respectively. These three residues, located on both sides of Trp436, contribute to positioning its side chain appropriately for substrate binding. Concomitant replacement of the residues at positions 382 and 383 by isoleucine and asparagine, respectively, would result in a triple mutant more likely to succeed as a mimic of the active site of the Tt8MGH enzyme.

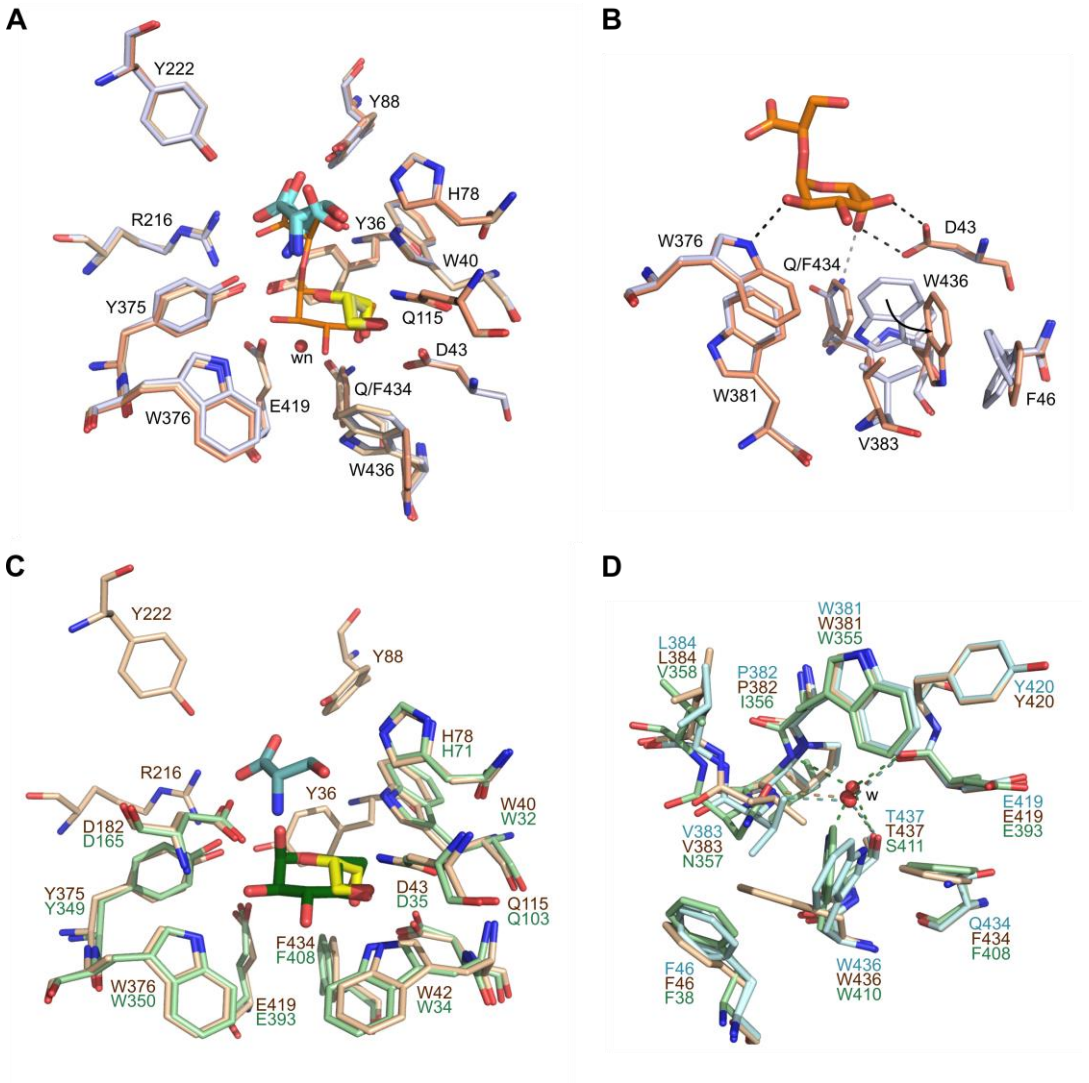
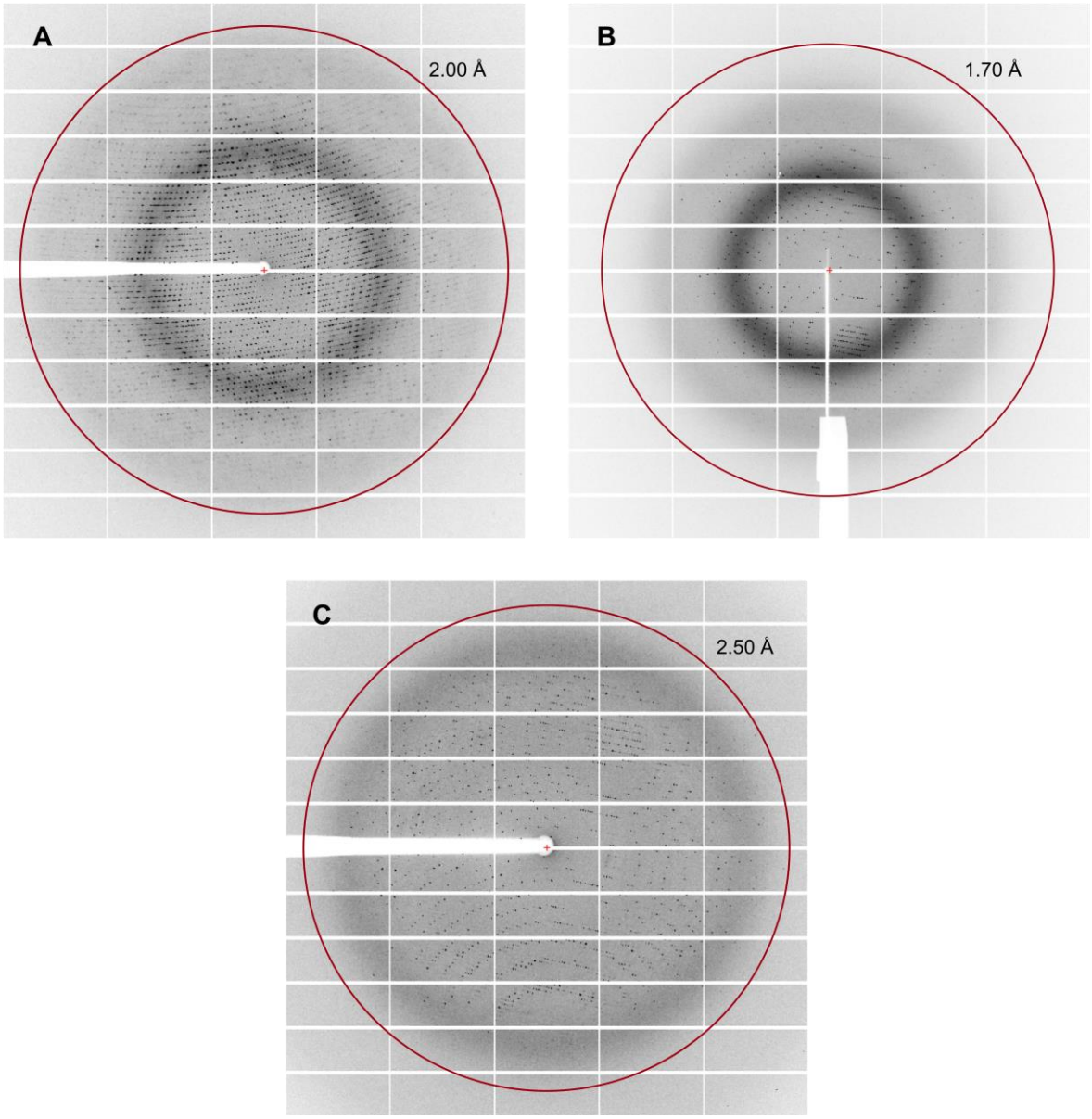
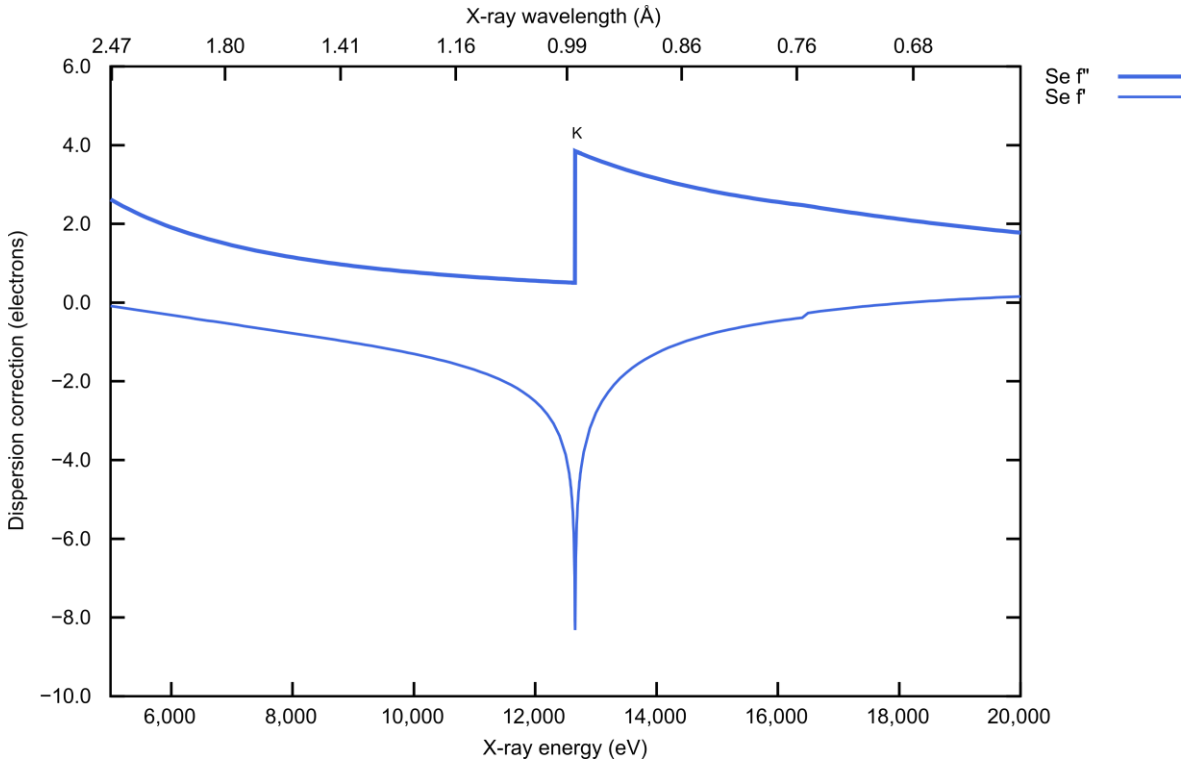


Figure 3.15 – Active site of ga*Mha*GgH Q434F mutant. A – Superposition of the active site region of wild-type ga*Mha*GgH and mutants Q434F and D182A. The active site of Q343F (salmon) and ga*Mha*GgH (wheat) is occupied by serine (cyan) and glycerol (yellow), while that of D182A (light blue) has GG (orange). B – Superposition of the active site region of the Q434F (salmon) and D182A (light blue) mutants in complex with GG (orange), highlighting the hydrophobic surface where the glucose ring of the substrate is accommodated. Some of the hydrogen bonds between GG and the D182A mutant are represented as black dashed lines, whereas the hydrogen bond that is absent upon replacement of Glu434 for phenylalanine is coloured grey. The position of Trp436 is severely affected by the amino acid substitution, as indicated by the arrow. Proteins are coloured as in A. C – Superposition of the active site region of ga*Mha*GgH Q434F and Tt8MGH in complex with glucose (dark green) (PDB accession code: 4WVB). The active site content of mutant Q434F is coloured as in A. The residues of ga*Mha*GgH Q434F and Tt8MGH are represented by wheat and pale green sticks, respectively. D – Superposition of the Trp436 neighbourhood in Tt8MGH (pale green), ga*Mha*GgH (pale cyan) and Q434F (wheat). The hydrogen bonds in Tt8MGH, ga*Mha*GgH and Q434F are coloured green, teal and wheat, respectively.

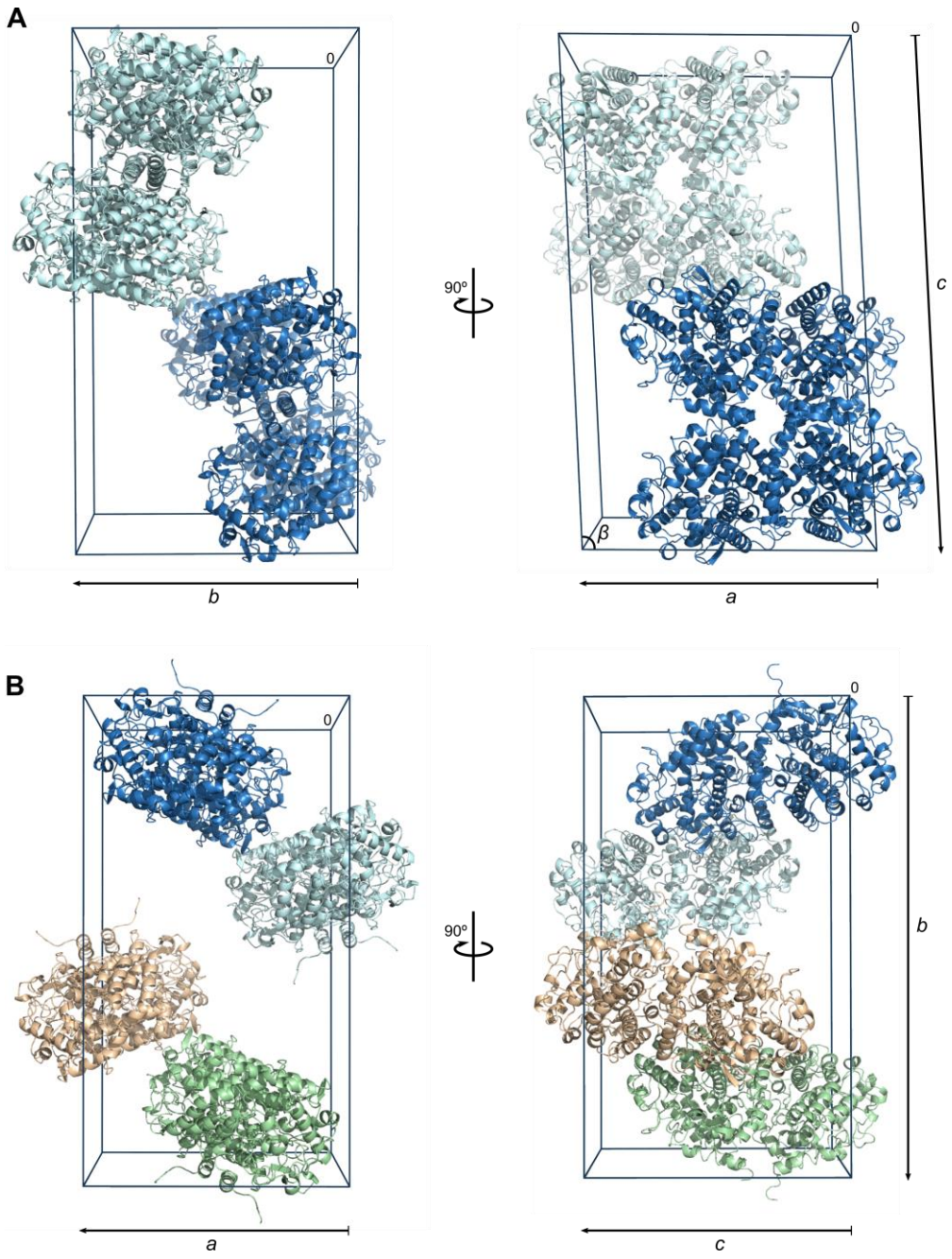
3.4. Supplemental information



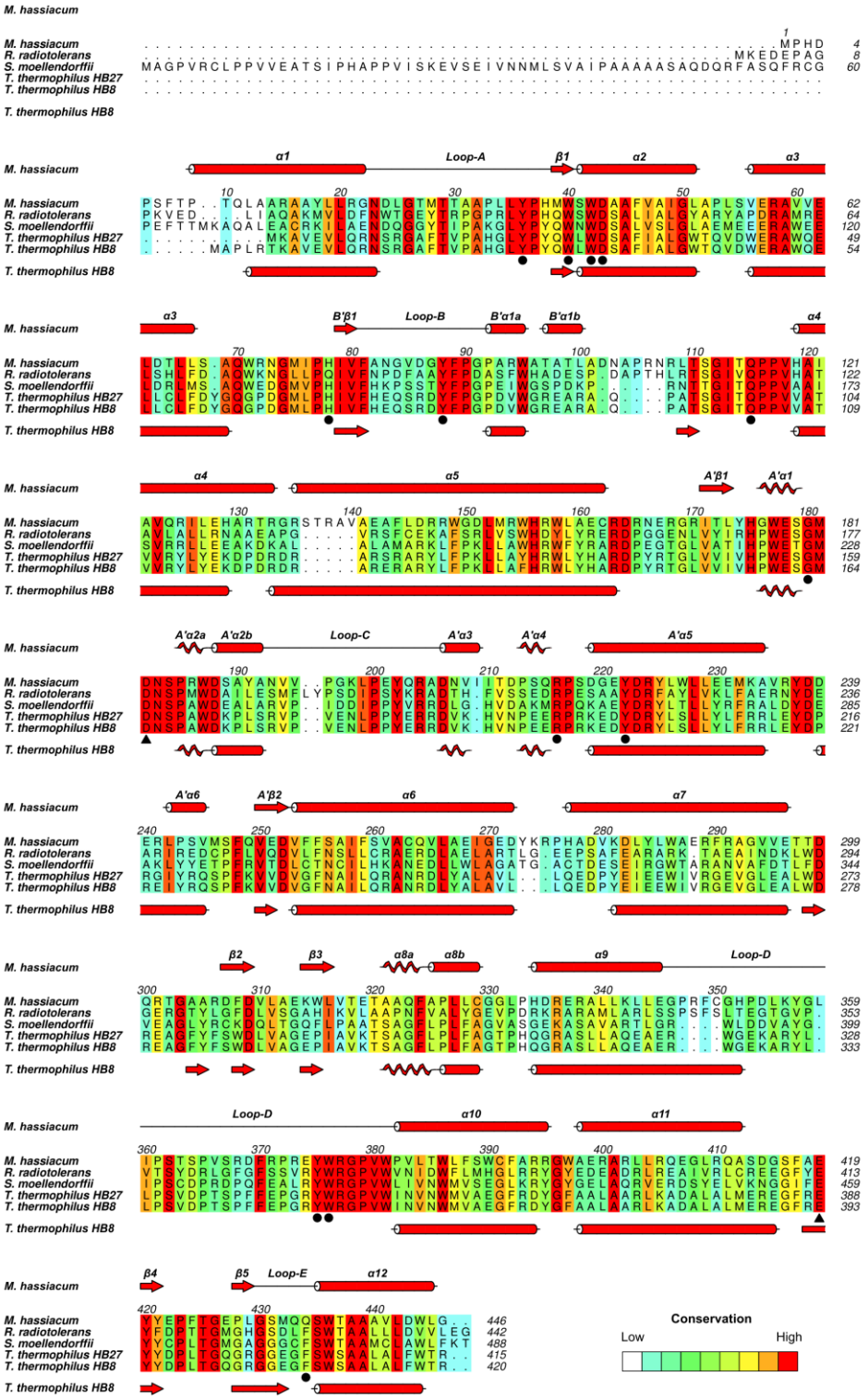
Supplemental figure S3.1 – Diffraction patterns from ga*Mha*GgH crystals. A and B – Diffraction patterns from SeMet-ga*Mha*GgH (A) and ga*Mha*GgH (B) crystals belonging to space groups $P2_1$ and $P2_12_12$, respectively. C – Diffraction pattern from a ga*Mha*GgH crystal belonging to the hexagonal space group $P6_222$. Each diffraction pattern was obtained by merging five consecutive 0.1° rotation images. The resolution limits corresponding to the circumscribed red circles are indicated.



Supplemental figure S3.2 – Theoretical scattering factors of selenium. The theoretical values of f″ (thick line) and f′ (thin line) for selenium were plotted over a range of energies, from 5,004 eV (2.47 Å) to 20,000 eV (0.62 Å). The plot was generated using the online tool X-ray Anomalous Scattering (http://skuld.bmsc.washington.edu/scatter/AS_index.html). The theoretical K absorption edge of selenium (~12,658 eV, 0.97Å) is labelled.



Supplemental figure S3.3 – Unit cell arrangement of monoclinic and orthorhombic ga*Mha*GgH crystals. The unit cells of ga*Mha*GgH crystals belonging to space group $P2_1$ (A) or $P2_12_12$ (B) are represented in two different orientations. The origin of the Cartesian coordinate system is indicated as 0. The unit cell *a*, *b* and *c*-axes are indicated and are ~90 Å (A) or ~86 Å (B), ~86 Å (A) or ~159 Å (B), and ~159 Å (A) or ~88 Å (B) in length, respectively. The angle β (~93º) of the monoclinic crystal unit cell is indicated. The molecules of the asymmetric unit are shown in blue, while the symmetry equivalent molecules are cyan, wheat or green.



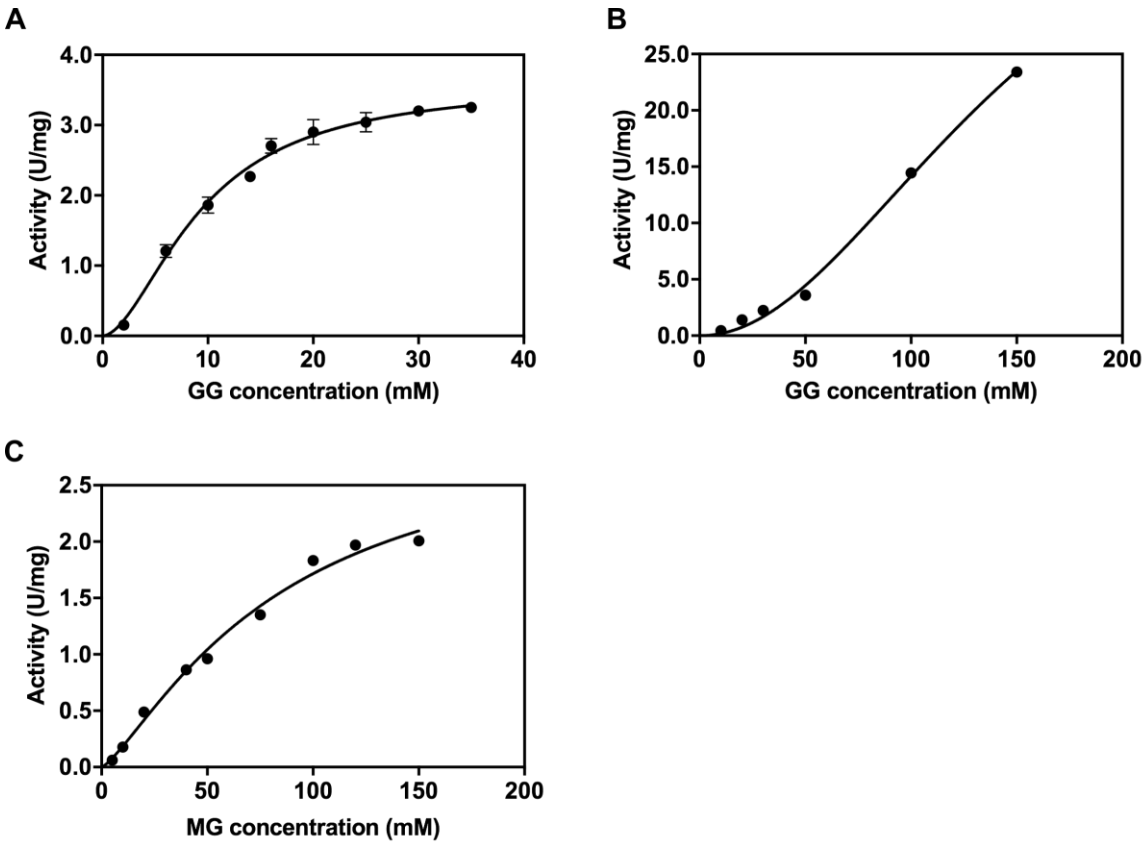
Supplemental figure S3.4 – Amino acid sequence alignment of characterized mannosylglycerate/glucosylglycerate hydrolases. Amino acid sequences of *Mycobacterium hassiacum* GgH (GenBank: EKF25940.1) and of MgH enzymes from *Rubrobacter radiotolerans* (GenBank: AFC76324.1), *Selaginella moellendorffii* (GenBank: XP_002961898.1), *Thermus thermophilus* HB27 (GenBank: WP_011173059.1) and *T. thermophilus* HB8 (GenBank: WP_011228344.1) were aligned with Clustal Omega (Sievers, Wilm et al. 2011). Secondary structure elements of *M. hassiacum* GgH and *T. thermophilus* HB8 MgH were identified with DSSP (Kabsch and Sander 1983) and are

indicated above and below the alignment, respectively. Residues are coloured from white to red according to increasing conservation. Catalytic residues are indicated by black triangles and other substrate-interacting residues by black circles. Figure prepared with ALINE (Bond and Schuttelkopf 2009).

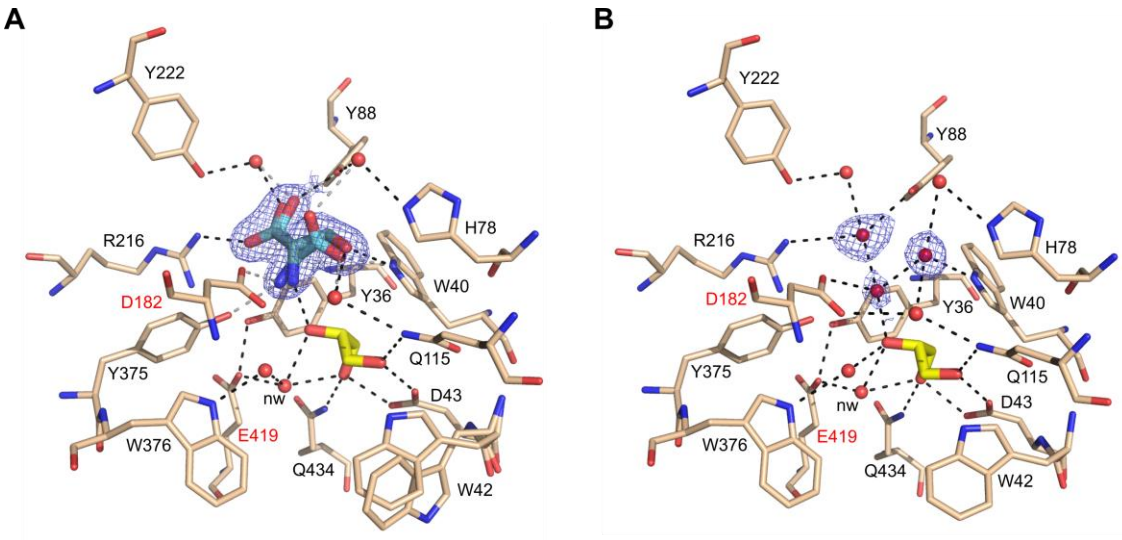
Supplemental table S3.1 – Analysis of the quaternary structure of ga*Mha*GgH. The total surface and total interface areas of each monomer (upper table) and the inter-monomer area and interactions (lower table) were determined using PISA (Krissinel and Henrick 2007).

Monomer	Total surface area (Å^2)	Total interface area (Å^2)
A	17,378	1,637
B	17,218	1,633
C	17,381	1,639
D	17,216	1,633
Average	*17,298*	*1,636*

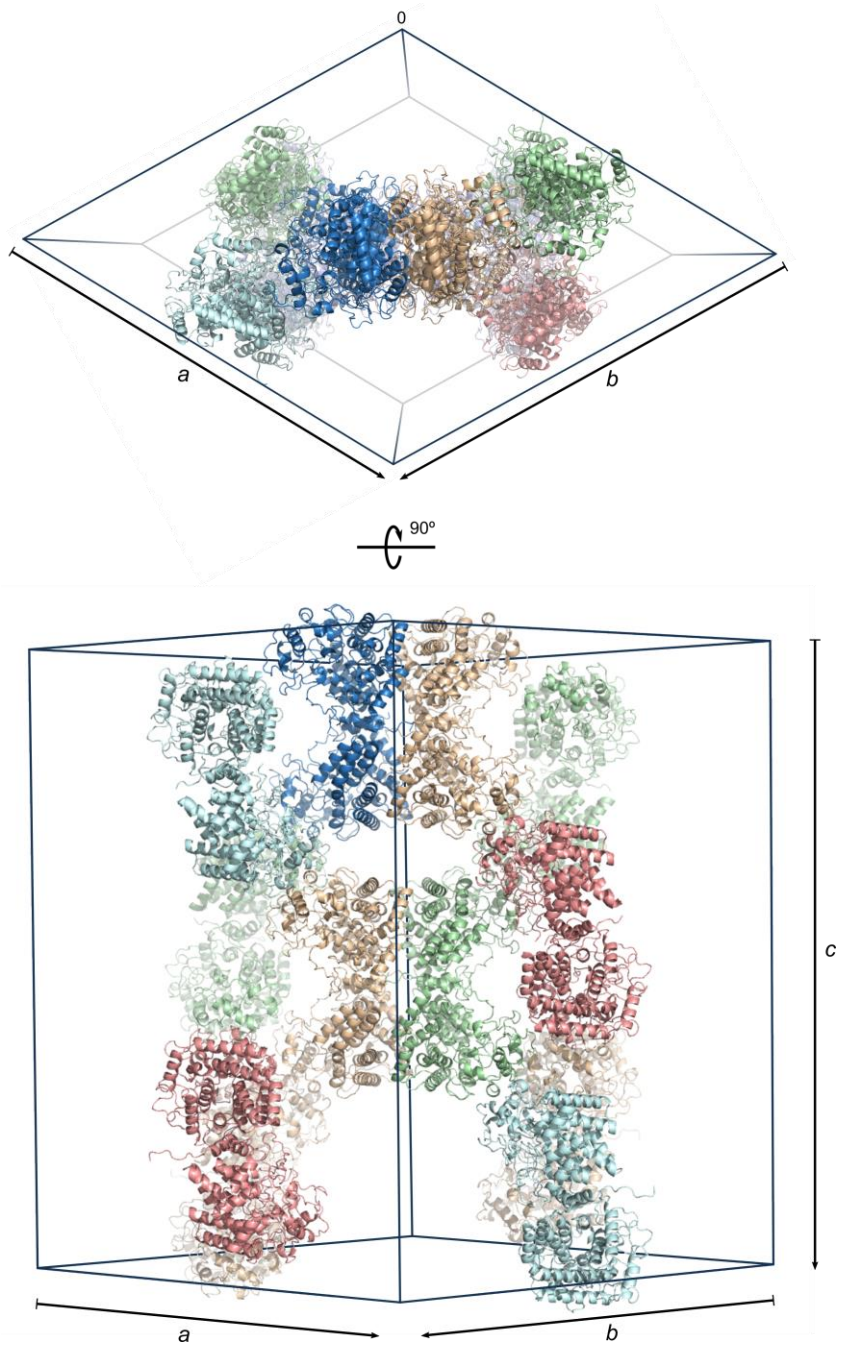
Interface	Interface area (Å^2)	Interface interaction	
		Hydrogen bonds	Salt bridges
A:B	477	0	10
C:D	478	0	10
A:C	900	14	0
B:D	895	12	0
A:D	260	1	0
B:C	261	1	0



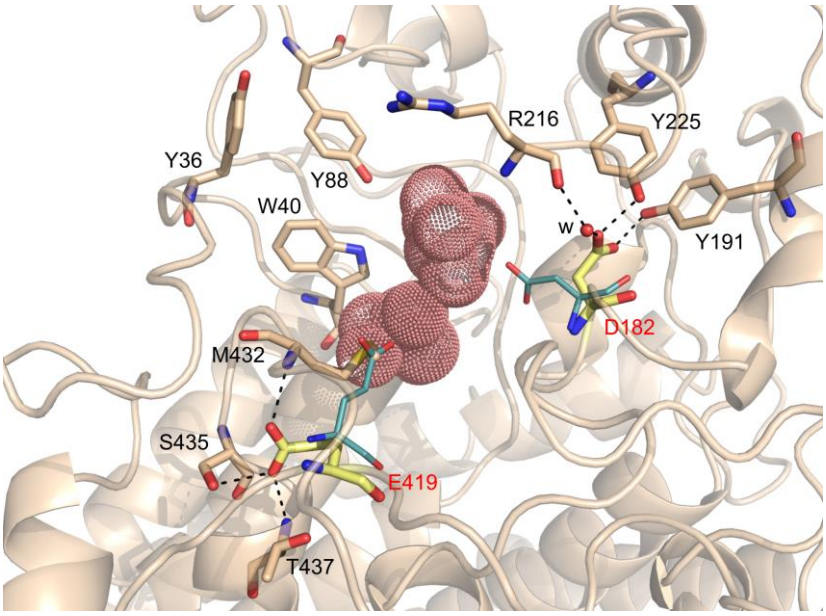
Supplemental figure S3.5 – Kinetic curves of GG and MG hydrolysis by ga*Mha*GgH. The sigmoidal shape of the experimental curve suggests that a cooperative effect between the tetramer subunits may exist. A – Kinetic curve obtained at 50°C and pH 6.0 using GG as substrate. B – Kinetic curve obtained at 37°C and pH 8.0 using GG as substrate. C – Kinetic curve of MG hydrolysis obtained at 50°C and pH 6.0. (Experiments performed by Susana Alarico, Molecular Mycobacteriology Group, CNC, University of Coimbra, Portugal)



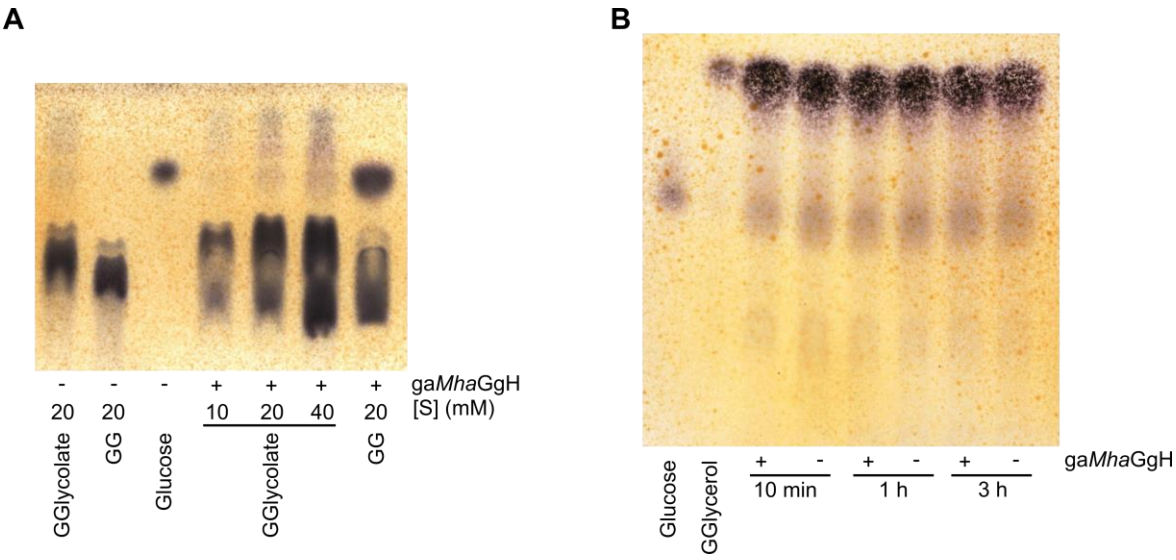
Supplemental figure S3.6 – Active site of ga*Mha*GgH. A – Active site of ga*Mha*GgH crystallized in a condition containing serine and glycerol. Subsites -1 and +1 are occupied by glycerol (yellow) and L-serine (cyan) molecules (ga*Mha*GgH-SER-GOL complex), respectively. L-serine adopts two different conformations (different shades of cyan) in the active site. Serine and glycerol are stabilised by direct and water-mediated polar contacts (dashed lines) with the active site residues. The catalytic residues are highlighted in red. The nucleophilic water (nw) is indicated. The 2Fo-Fc electron density map is contoured at 1 σ. B – Active site of ga*Mha*GgH crystallized in the absence of serine in the crystallization condition. Three water molecules (coloured crimson) replace the serine molecule present in A. The position of the substrate-binding residues is similar to that in the ga*Mha*GgH-SER-GOL complex. The catalytic residues and the nucleophilic water are indicated as in A. The 2Fo-Fc electron density map is contoured as in A.



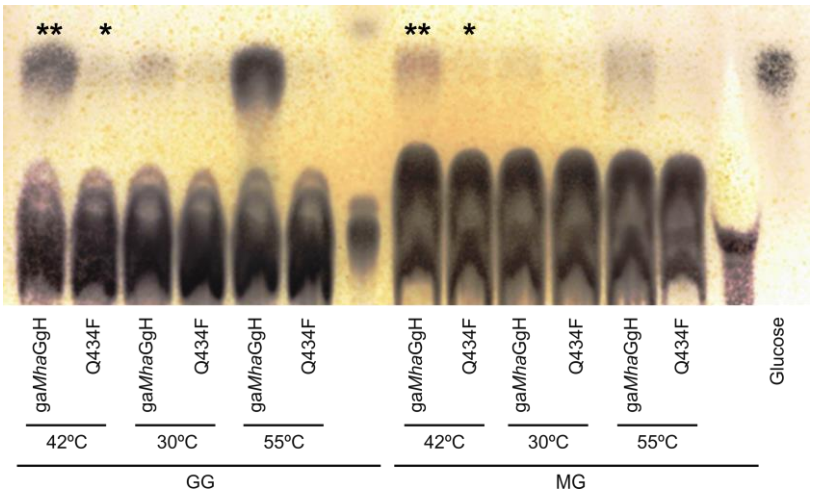
Supplemental figure S3.7 – Unit cell arrangement of hexagonal ga*Mha*GgH crystals. The unit cell of the ga*Mha*GgH crystals belonging to space group $P6_222$ is represented in two different orientations. The origin of the Cartesian coordinate system is indicated as 0. The unit cell *a*, *b* and *c*-axes are indicated and are ~167, ~167 and ~243 Å in length, respectively. The molecules of the asymmetric unit are shown in blue, while the symmetry equivalent molecules are cyan, wheat, green or salmon.



Supplemental figure S3.8 – Active site of ga*Mha*GgH in open state. The catalytic residues (yellow sticks) are facing away from the active site cavity (salmon spheres), stabilised by direct hydrogen bonds and by water (w)-mediated contacts with the neighbour residues (dashed lines). Upon substrate binding, the catalytic residues (teal sticks) turn towards the active site cavity, establishing new polar contacts (not represented).



Supplemental figure S3.9 – Analysis of GGlycolate and GGlycerol hydrolysis by ga*Mha*GgH. The ability of ga*Mha*GgH to hydrolyse GGlycolate (A) and GGlycerol (B) was evaluated by TLC. A – Different concentrations of GGlycolate were incubated with ga*Mha*GgH at 50°C for 2 h. Glucose production was not detected for any of the concentrations tested. B – The ga*Mha*GgH was incubated with GGlycerol for 10 min, 1h and 3h at 50°C. The production of glucose from GGlycerol was not observed. (Experiments performed by Susana Alarico, Molecular Mycobacteriology Group, CNC, University of Coimbra, Portugal)



Supplemental figure S3.10 – Analysis of GG and MG hydrolysis by ga*Mha*GgH mutant Q434F. The ability of the Q434F mutant to hydrolyse GG and MG was evaluated by TLC. Wild-type ga*Mha*GgH and Q434F mutant (~2.7 μM) were incubated with 20 mM GG or MG at different temperatures for 1 h. A considerable lower production of glucose or mannose was observed for mutant Q434F (marked with asterisk) compared to the wild-type enzyme (marked with a double asterisk). (Experiments performed by Susana Alarico, Molecular Mycobacteriology Group, CNC, University of Coimbra, Portugal)

Chapter 4

General discussion and conclusion

Mycobacteria constitute a large and diverse group of organisms. Although most of them are environmental, many of these ubiquitous bacteria are opportunistic pathogens capable of causing infections in humans, especially in immunocompromised or immunosuppressed individuals. Moreover, mycobacteria also comprehend highly aggressive species, including the causative agents of tuberculosis. Currently, tuberculosis is one of the top 10 causes of death worldwide and the leading cause of death by a single infectious agent. Mycobacteria in general display high resilience against stress conditions, including antibiotics and disinfectant action. Thus, the treatment of mycobacterial infections is usually challenging and demands long duration multidrug therapy. Worse, the emergence of multidrug and extensively drug-resistant strains limits the ability to control mycobacterial infections. Altogether, the increasing drug resistance of mycobacteria and the long duration required for the treatment urges for the development of new, shorter and more efficient anti-mycobacterial therapies.

Polymethylated polysaccharides are unique carbohydrates found exclusively in mycobacteria and related genera, including *Nocardia* and *Streptomyces.* Although these molecules were discovered more than 50 years ago, their physiological role and biosynthesis are still not fully understood. PMPSs are subdivided into 3-*O*-methylmannose polysaccharides and 6-*O*-methylglucose lipopolysaccharides. MGLPs are the only type of PMPSs common to all mycobacteria examined thus far and to a few other members of the *Actinobacteria* phylum. The genes proposed to participate in MGLPs biosynthesis are highly conserved among mycobacteria, and several of them were considered essential for mycobacterial *in vitro* growth, highlighting their contribution for survival and their potential as drug targets.

In this thesis, two mycobacterial enzymes are structurally characterized: glucosyl-3-phosphoglycerate phosphatase and glucosylglycerate hydrolase. GpgP and GgH are involved in the metabolism of glucosylglycerate, one of the earlier intermediates of MGLPs. While GpgP catalyses the dephosphorylation of glucosyl-3-phosphoglycerate producing GG (Chapter 2), GgH converts the later to glucose and glycerate (Chapter 3). Although the structure of a recombinant *M. tuberculosis* GpgP was recently reported (Zheng, Jiang et al. 2014), the structure described in this thesis was determined at a considerable higher resolution (~1.5 Å), allowing a better interpretation of the active site content.

The mycobacterial GpgP was first annotated as a phosphoglycerate mutase due to its sequence similarities to other enzymes of this type. However, biochemical analysis of a recombinant *M. tuberculosis* GpgP revealed that it displays a phosphatase activity –

preferentially on GPG – and only residual phosphoglycerate mutase activity, leading to its reclassification as glucosyl-3-phosphoglycerate phosphatase (Mendes, Maranha et al. 2011).

*Mtu*GpgP is a homodimeric protein, composed of central β-strands flanked on both sides by α-helices. The electrostatic surface of *Mtu*GpgP reveals two positively charged clefts, clearly shaped by both monomers, where the active sites are located. The contribution of both monomers for the active site clefts suggests that the dimer is likely the minimal functional unit, which was underscored by the inability of individual monomers to dephosphorylate a synthetic substrate (Zheng, Jiang et al. 2014).

The substrate binding mode of *Mtu*GpgP was elucidated by the structure of its complex with vanadate (Figure 2.5). The large electron density blob found in the active site of *Mtu*GpgP co-crystallized with vanadate can be explained by the presence of three chemical species: a vanadate ion and two trivanadate glycerol esters. Vanadate was found covalently linked to the catalytic histidine and stabilised through a network of polar contacts with neighbour residues, including three highly conserved amino acids: two arginines (Arg10 and Arg60) and one histidine (His159). The vanadate-histidine complex is a mimic of the phosphohistidine transiently formed during the course of the reaction, which allowed the identification of the residues implicated in substrate binding and catalysis.

The trivanadate glycerol esters described here are novel. These molecules highlighted the presence of a hydrophobic surface, lined by residues Met22, Leu87, and Trp109, that is likely to contribute for accommodating the glucose ring of the substrate. This surface is surrounded by polar residues (His95, Arg110 and Arg123) that may participate in glucose stabilisation through polar contacts.

In the vanadate-*Mtu*GpgP structure, the proton donor Glu84 is located within hydrogen bonding distance of the oxygen that is protonated during the phosphorylation of the catalytic histidine, suggesting that it is likely to donate a proton directly to the leaving group. In the trivanadate-glycerol-*Mtu*GpgP complex, the distance between the vanadium atom V3 of trivanadate and Glu84 is compatible with a covalent bond, which suggests that Glu84 may have performed a nucleophilic attack directly to the vanadate. The structures of *Mtu*GpgP in complex with vanadate and trivanadate glycerol suggest that Glu84 is able to shift between protonated and ionized forms as expected.

The structure of *Mtu*GpgP in complex with vanadate and GG was also obtained (Figure 2.7). The ternary complex suggests that GG may establish transient interactions with the

residues lining the active site cleft, which are likely to facilitate the inward and outward movement of substrate and of product, respectively. The substrate binding mode of *Mtu*GpgP could be better clarified by the structure of wild-type and/or catalytic inactive *Mtu*GpgP in complex with the substrate GPG. Catalytic inactive *Mtu*GpgP could likely be obtained by replacing Glu84 and/or His11 by alanine. Indeed, a catalytically inactive mutant of *Mtu*GpgP was successfully generated by replacing His11 by an alanine (data not shown). Crystals of wild-type and catalytically inactive *Mtu*GpgP were soaked into the crystallization condition containing substrate GPG or synthetic substrate PNPP. However, none of the ligands were found in the active site of the macromolecules. Different times of soaking and higher concentrations of ligands should be used in the future.

*Mtu*GpgP contains a C-terminal hexahistidine tag linked to the native *M. tuberculosis* GpgP sequence by a seven amino acid residue linker. In the higher-resolution structure, the linker and the hexahistidine tag are organized into an α-helix. Although the molecules in the asymmetric unit have different crystallographic contacts, the C-terminal helices are in equivalent positions. In the iodide-containing crystal, the different crystal packing results in an increased space between macromolecules at this region, allowing the engineered C-terminal tail to have higher mobility, as suggested by the absence of electron density map that limits the C-terminal modelling from the end of native GpgP sequence.

Although scarce, there are reports pointing to contribution of the C-terminus for the mutase activity. Removal of the C-terminal region led to a markedly decreased mutase activity for the *Saccharomyces cerevisiae* phosphoglycerate mutase (*S. cerevisiae* PGAM) (Walter, Nairn et al. 1999). Also, the C-terminal tail of *S. cerevisiae* PGAM showed higher resistance to proteolysis in presence of the cofactor 2,3-bisphosphoglycerate, suggesting the participation of the C-terminal in substrate binding. The crystallographic study of the cofactor-dependent phosphoglycerate mutase from *Escherichia coli* showed that the C-terminal region displayed higher mobility in the apo form of the enzyme, while in phosphohistidine-activated and vanadate-inactivated forms the C-terminal was found to interact with the residues implicated in substrate binding (Bond, White et al. 2001, Bond, White et al. 2002). A similar observation was made for the cofactor-dependent phosphoglycerate mutase from *Bacillus stearothermophilus* (PhoE): its C-terminus could only be modelled for the phosphate and vanadate complexes (Rigden, Littlejohn et al. 2003). The interaction of the C-terminal with the substrate-binding residues is thus likely to offer additional stabilisation of the active conformation of the catalytic core.

Since the C-terminal tail of mutases can have variable length and no evident sequence pattern, it is not possible to predict protein function based on the C-terminal sequence alone. The overall structure of *B. stearothermophilus* PhoE, *E.coli* PGM and *Mtu*GpgP is similar, with the C-terminal tails displaying different length but occupying similar positions. The amino acid segment of *Mtu*GpgP preceding the engineered tag seems to be unstructured and mobile, as supported by the iodide-containing structure, and thus suitable to interact with the active site residues. Therefore, it would be interesting to evaluate the impact of the engineered C-terminal tag in mycobacterial GpgP activity. Although more challenging to purify, it would be more reliable to evaluate the activity of a tag-free GpgP protein. On the other hand, the production of a recombinant GpgP containing an additional Gly-Ala dipeptide at the N-terminus using a similar strategy as for GgH, would facilitate the purification step. Also, due to the localization of the N-terminus in dimeric GpgP, the engineered N-terminal dipeptide would be unlikely to affect enzyme activity.

The biochemical and structural characterization of a recombinant GgH from *M. hassiacum* was reported in Chapter 3. GgH seems to be restricted to rapid-growing mycobacteria. The subdivision of mycobacteria into slow and rapid-growing mycobacteria is based on their growth rate, which reflects their phylogenetic divergency (Tortoli, Fedrizzi et al. 2017). Interestingly, when comparing the growth rate with pathogenicity, the highly pathogenic bacteria (such as *M. tuberculosis* and *M. leprae*) are in the slow-growing group, while non-pathogenic species (such as *M. phei* and *M. thermoresistible*) are among the rapid-growing mycobacteria. However, the link between virulence and growth rate is narrow and a number of RGM have also been associated to human infections, especially in immunosusceptible individuals, with *M. abscessus*, *M. chelonae* and *M. fortuitum* (formerly known as *M. ranae*) as the most common infectious agents (Brown-Elliott and Philley 2017). Alarmingly, RGM are not susceptible to the first-line anti-TB drugs and drug therapeutic regimens are selected based on their unique *in vitro* susceptibility patterns (Nunes-Costa, Alarico et al. 2016, Akram and Bhimji 2017, Brown-Elliott and Philley 2017). The treatment of rapid-growing mycobacterial infections is becoming a considerable clinical challenge for which the therapeutic solutions are scarce (Kasperbauer and De Groote 2015).

GgH from *M. hassiacum* was proposed to be involved in the decrease of intracellular GG, accumulated during mycobacterial growth under nitrogen-limiting conditions, by catalysing its hydrolysis to glucose and glycerate (Alarico, Costa et al. 2014). Although considered a compatible solute accumulated in diverse organisms to avoid the loss of water in high

salinity environments, the accumulation of GG in mycobacteria was shown to be independent of osmotic stress (Behrends, Williams et al. 2012). The accumulation of GG was exclusively linked to nitrogen starvation and, so far, no data is available regarding GG accumulation in other growth-limiting conditions.

Under stress, bacteria prepare themselves to enter a hibernation-like state known as dormancy that allows cells to keep viable until improvement of the environmental conditions (Rittershaus, Baek et al. 2013). In growth-limiting conditions, cells accumulate carbon stores to use as a rapidly mobilisable source of energy to promote cell growth and division and outcompete the neighbouring microorganisms. In agreement, the growth curve of *M. hassiacum* (Alarico, Costa et al. 2014) shows a small jump of growth after the addition of nitrogen to the medium, concomitant with the decrease of intracellular GG content. In *M. tuberculosis*, the dormancy state seems to the accompanied by the accumulation of triacylglycerol (Daniel, Deb et al. 2004) and wax esters (Sirakova, Deb et al. 2012). Since slow-growing mycobacteria, such as *M. tuberculosis*, seem to lack a GgH ortholog-encoding gene, it is likely that SGM and RGM have developed different strategies to survive to growth-limiting conditions. In order to clarify if RGM accumulate GG as a general response to stress, it would be fundamental to evaluate the accumulation of GG by other RGM and in different growth-limiting conditions, including in the absence of other nutrients, at lower oxygen levels, and at lower or higher pH and temperature.

The dormancy state is associated to a high relapse rate of a number of infections, including tuberculosis (Lewis 2007, Lipworth, Hammond et al. 2016). The drugs currently in use are mostly directed to cell growth and division. Since these functions are almost supressed in dormant cells, they are more likely to survive the treatment. Thus, targeting functions linked to dormancy and resuscitation would likely result in a more efficient treatment (Cano-Muniz, Anthony et al. 2018). GgH is likely involved in cell resuscitation of RGM by promoting the rapid mobilisation of energy upon relief of growing-limiting stress and thus a promising target for the development of new anti-RGM drugs. However, physiological evidence on GgH role in rapid-growing mycobacterial survival, dormancy and pathogenesis is still needed.

GgH from *M. hassiacum* was produced recombinantly with a C-terminal hexahistidine tag (Alarico, Costa et al. 2014). Although active, this recombinant protein did not crystallize into three-dimensional crystals. Under the hypothesis that the C-terminal tag could be affecting crystallization, a new construct was produced containing a N-terminal hexahistidine tag separated from the native *M. hassiacum* GgH sequence by a TEV cleavage peptide sequence. Removal of the N-terminal tag by proteolytic cleavage

resulted in a protein with a Gly-Ala dipeptide N-terminal to the native GgH sequence at the end of the purification procedure. Size-exclusion chromatography and dynamic light scattering analysis of both constructs suggested a change in the oligomeric state from the C-terminal tagged GgH (*Mha*GgH-His_6) to the untagged protein (ga*Mha*GgH), accompanied by the increase in both protein solution homogeneity and melting temperature. Unlike *Mha*GgH-His_6, the ga*Mha*GgH variant yielded crystals able to diffract X-rays to high resolution that allowed its structure to be determined.

M. hassiacum GgH is a homotetrametic protein, composed of a dimer of dimers (Figure 3.4). Each monomer contains the $(\alpha/\alpha)_6$-barrel domain typical of the family 63 of glycoside hydrolases (GH63), to which GgH belongs. The $(\alpha/\alpha)_6$-barrel domain is covered by a cap domain that can be subdivided into A' and B'-regions. The active site is localized between the $(\alpha/\alpha)_6$ and cap domains and opens to the solvent through a negatively charged tunnel. The active site of GgH displays a typical topology of monosaccharidases (Davies and Henrissat 1995). Upon substrate binding, the active site becomes inaccessible to the solvent, in a closed like state. The open and closed state of GgH is determined by the well-coordinated movement of several mobile loops that contain some of the substrate-interacting residues.

The biochemical data showed that GgH was able to hydrolyse *in vitro* GG and MG, with higher efficiency for the former. The difference in hydrolysis efficiency of GgH for GG and MG may result from distinct binding affinities. Since active GgH would lead to the release of reaction products and difficult result interpretation, catalytic inactive point mutants, ga*Mha*GgH D182A and E419A, were used to evaluate the binding affinity of GgH for the substrates. The ga*Mha*GgH E419A mutant loses Glu419-mediated polar contacts necessary to stabilise the closed state of the enzyme. Although displaying lower affinity for both substrates, this mutant is likely the most adequate to evaluate the molecular determinants of GG and MG binding specificities. The α-D-glucose and α-D-mannose differ in the stereochemistry at the C2 position; while the C2 hydroxyl group in α-D-glucose is equatorial, in α-D-mannose it is axial. The glucose and mannose moieties of the substrates are recognized by subsite -1. The C2 hydroxyl group of the glucose moiety of GG establishes polar contacts with Asp182, Tyr375 and Trp376, while that of the mannose moiety of MG is only hydrogen bonded to Trp376. Thus, these residues are involved in specific recognition of the non-reducing end of GG and likely in the higher efficiency of GgH to hydrolyse this substrate.

The contribution of subsite +1 to substrate recognition was also evaluated by using the substrate analogues glucosylglycerol and glucosylglycolate that differ from the substrate

GG in the reducing end moiety. *In vitro*, GgH seems to be unable to hydrolyse any of these compounds, and the structure of active GgH in complex with GGlycerol or GGlycolate could not be obtained in the experimental conditions used. However, the structures of the catalytically inactive mutants in complex with these substrate analogues were determined, revealing the important contribution of the reducing end of the substrate for active site closure. Also, since these compounds were found in the active site of catalytically inactive mutants concomitantly with glycerol and serine from the crystallization solution, the affinity of ga*Mha*GgH D182A and E419A for GGlycerol and GGlycolate is likely lower than for those molecules.

In *M. hassiacum* GgH, substrate binding is a well-coordinated event, involving highly conserved residues and a complex hydrogen bond network. During evolution, the active site of GgH was optimized to harbour specific substrates, such as GG. The presence of MG in *M. hassiacum* cells was not reported so far. Assuming that *M. hassiacum* is unable to produce MG, the ability of GgH to hydrolase this compound can be just a vestigial function from an ancestor enzyme. On the other hand, since cells are able to interchange molecules with the exterior using different strategies – from passive diffusion to active transport of specific molecules – it is also possible that *M. hassiacum* possesses adequate machinery for capturing MG from the exterior as a possible source of carbon and energy, and uses GgH for its hydrolysis.

Although the MGLP biosynthetic pathway is becoming clearer, there are still steps in MGLP maturation that need to be explored. Several genes were proposed to participate in MGLP biosynthesis based on their genomic neighbourhood and encoded protein sequence similarity to other characterized enzymes. However, not only physiological evidence on their contribution for MGLP production is still missing, as their specific function needs to be confirmed. Nevertheless, the results presented in this thesis contribute for a better comprehension of the biochemistry of the mycobacterial enzymes, GpgP and GgH, involved in the metabolism of GG, one of the earlier intermediates in MGLP biosynthesis and an important carbon store of rapid-growing mycobacteria.

References

Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 213-221.

Akram, S. M. and S. S. Bhimji. (2017, updated 2017 May 16). "Mycobacterium Chelonae." StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing, 2018, from https://www.ncbi.nlm.nih.gov/books/NBK430806/.

Alarico, S., M. Costa, M. S. Sousa, A. Maranha, E. C. Lourenco, T. Q. Faria, M. R. Ventura and N. Empadinhas (2014). "Mycobacterium hassiacum recovers from nitrogen starvation with up-regulation of a novel glucosylglycerate hydrolase and depletion of the accumulated glucosylglycerate." Sci Rep **4**: 6766.

Alarico, S., N. Empadinhas and M. S. da Costa (2013). "A new bacterial hydrolase specific for the compatible solutes alpha-D-mannopyranosyl-(1-->2)-D-glycerate and alpha-D-glucopyranosyl-(1-->2)-D-glycerate." Enzyme Microb Technol **52**(2): 77-83.

Albesa-Jove, D., F. Mendoza, A. Rodrigo-Unzueta, F. Gomollon-Bel, J. O. Cifuente, S. Urresti, N. Comino, H. Gomez, J. Romero-Garcia, J. M. Lluch, E. Sancho-Vaello, X. Biarnes, A. Planas, P. Merino, L. Masgrau and M. E. Guerin (2015). "A Native Ternary Complex Trapped in a Crystal Reveals the Catalytic Mechanism of a Retaining Glycosyltransferase." Angew Chem Int Ed Engl **54**(34): 9898-9902.

Albesa-Jove, D., J. Romero-Garcia, E. Sancho-Vaello, F. X. Contreras, A. Rodrigo-Unzueta, N. Comino, A. Carreras-Gonzalez, P. Arrasate, S. Urresti, X. Biarnes, A. Planas and M. E. Guerin (2017). "Structural Snapshots and Loop Dynamics along the Catalytic Cycle of Glycosyltransferase GpgS." Structure **25**(7): 1034-1044 e1033.

Alcaide, F., M. J. Pena, D. Perez-Risco, D. Camprubi, L. Gonzalez-Luquero, M. D. Grijota-Camino, J. Dorca and M. Santin (2017). "Increasing isolation of rapidly growing mycobacteria in a low-incidence setting of environmental mycobacteria, 1994-2015." Eur J Clin Microbiol Infect Dis **36**(8): 1425-1432.

Amaya, M. F., A. G. Watts, I. Damager, A. Wehenkel, T. Nguyen, A. Buschiazzo, G. Paris, A. C. Frasch, S. G. Withers and P. M. Alzari (2004). "Structural insights into the catalytic mechanism of Trypanosoma cruzi trans-sialidase." Structure **12**(5): 775-784.

Andries, K., P. Verhasselt, J. Guillemont, H. W. Gohlmann, J. M. Neefs, H. Winkler, J. Van Gestel, P. Timmerman, M. Zhu, E. Lee, P. Williams, D. de Chaffoy, E. Huitric, S. Hoffner, E. Cambau, C. Truffot-Pernot, N. Lounis and V. Jarlier (2005). "A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis." Science **307**(5707): 223-227.

Antonsson, B. (1997). "Phosphatidylinositol synthase from mammalian tissues." Biochim Biophys Acta **1348**(1-2): 179-186.

Anuchin, A. M., A. L. Mulyukin, N. E. Suzina, V. I. Duda, G. I. El-Registan and A. S. Kaprelyants (2009). "Dormant forms of Mycobacterium smegmatis with distinct morphology." Microbiology **155**(Pt 4): 1071-1079.

Baba, T., K. Kaneda, E. Kusunose, M. Kusunose and I. Yano (1989). "Thermally adaptive changes of mycolic acids in Mycobacterium smegmatis." J Biochem **106**(1): 81-86.

Baker, N. A., D. Sept, S. Joseph, M. J. Holst and J. A. McCammon (2001). "Electrostatics of nanosystems: application to microtubules and the ribosome." Proc Natl Acad Sci U S A **98**(18): 10037-10041.

Balazsi, G., A. van Oudenaarden and J. J. Collins (2011). "Cellular decision making and biological noise: from microbes to mammals." Cell **144**(6): 910-925.

Banis, R. J., D. O. Peterson and K. Bloch (1977). "Mycobacterium smegmatis fatty acid synthetase. Polysaccharide stimulation of the rate-limiting step." J Biol Chem **252**(16): 5740-5744.

Barker, M. K. and D. R. Rose (2013). "Specificity of Processing alpha-glucosidase I is guided by the substrate conformation: crystallographic and in silico studies." J Biol Chem **288**(19): 13563-13574.

Beblo-Vranesevic, K., E. A. Galinski, R. Rachel, H. Huber and P. Rettberg (2017). "Influence of osmotic stress on desiccation and irradiation tolerance of (hyper)-thermophilic microorganisms." Arch Microbiol **199**(1): 17-28.

Behrends, V., K. J. Williams, V. A. Jenkins, B. D. Robertson and J. G. Bundy (2012). "Free glucosylglycerate is a novel marker of nitrogen stress in Mycobacterium smegmatis." J Proteome Res **11**(7): 3888-3896.

Bergeron, R., Y. Machida and K. Bloch (1975). "Complex formation between mycobacterial polysaccharides or cyclodextrins and palmitoyl coenzyme A." J Biol Chem **250**(4): 1223-1230.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne (2000). "The Protein Data Bank." Nucleic Acids Res **28**(1): 235-242.

Bernstein, J., W. A. Lott, B. A. Steinberg and H. L. Yale (1952). "Chemotherapy of experimental tuberculosis. V. Isonicotinic acid hydrazide (nydrazid) and related compounds." Am Rev Tuberc **65**(4): 357-364.

Bhamidi, S., M. S. Scherman, C. D. Rithner, J. E. Prenni, D. Chatterjee, K. H. Khoo and M. R. McNeil (2008). "The identification and location of succinyl residues and the characterization of the interior arabinan region allow for a model of the complete primary structure of Mycobacterium tuberculosis mycolyl arabinogalactan." J Biol Chem **283**(19): 12992-13000.

Bond, C. S. and A. W. Schuttelkopf (2009). "ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments." Acta Crystallogr D Biol Crystallogr **65**(Pt 5): 510-512.

Bond, C. S., M. F. White and W. N. Hunter (2001). "High resolution structure of the phosphohistidine-activated form of Escherichia coli cofactor-dependent phosphoglycerate mutase." J Biol Chem **276**(5): 3247-3253.

Bond, C. S., M. F. White and W. N. Hunter (2002). "Mechanistic implications for Escherichia coli cofactor-dependent phosphoglycerate mutase based on the high-resolution crystal structure of a vanadate complex." J Mol Biol **316**(5): 1071-1081.

Bones, A. M. and J. T. Rossiter (1996). "The myrosinase-glucosinolate system, its organisation and biochemistry." Physiologia Plantarum **97**(1): 194-208.

Bourassa, L. and A. Camilli (2009). "Glycogen contributes to the environmental persistence and transmission of Vibrio cholerae." Mol Microbiol **72**(1): 124-138.

Bowler, M. W., D. Nurizzo, R. Barrett, A. Beteva, M. Bodin, H. Caserotto, S. Delageniere, F. Dobias, D. Flot, T. Giraud, N. Guichard, M. Guijarro, M. Lentini, G. A. Leonard, S. McSweeney, M. Oskarsson, W. Schmidt, A. Snigirev, D. von Stetten, J. Surr, O. Svensson, P. Theveneau and C. Mueller-Dieckmann (2015). "MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules." J Synchrotron Radiat **22**(Pt 6): 1540-1547.

Brennan, P. and C. E. Ballou (1968). "Phosphatidylmyoinositol monomannoside in Propionibacterium shermanii." Biochem Biophys Res Commun **30**(1): 69-75.

Brennan, P. J. and D. P. Lehane (1971). "The phospholipids of corynebacteria." Lipids **6**(6): 401-409.

Brennan, P. J. and H. Nikaido (1995). "The envelope of mycobacteria." Annu Rev Biochem **64**: 29-63.

Brown-Elliott, B. A. and J. V. Philley (2017). "Rapidly Growing Mycobacteria." Microbiol Spectr **5**(1).

Burmeister, W. P., S. Cottaz, H. Driguez, R. Iori, S. Palmieri and B. Henrissat (1997). "The crystal structures of Sinapis alba myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase." Structure **5**(5): 663-675.

Burmeister, W. P., S. Cottaz, P. Rollin, A. Vasella and B. Henrissat (2000). "High resolution X-ray crystallography shows that ascorbate is a cofactor for myrosinase and substitutes for the function of the catalytic base." J Biol Chem **275**(50): 39385-39393.

Candy, D. J. and J. Baddiley (1966). "3-O-methyl-D-mannose from Streptomyces griseus." Biochem J **98**(1): 15-18.

Caner, S., N. Nguyen, A. Aguda, R. Zhang, Y. T. Pan, S. G. Withers and G. D. Brayer (2013). "The structure of the Mycobacterium smegmatis trehalose synthase reveals an unusual active site configuration and acarbose-binding mode." Glycobiology **23**(9): 1075-1083.

Cano-Muniz, S., R. Anthony, S. Niemann and J. C. Alffenaar (2018). "New Approaches and Therapeutic Options for Mycobacterium tuberculosis in a Dormant State." Clin Microbiol Rev **31**(1).

Chapman, G. B., J. H. Hanks and J. H. Wallace (1959). "An electron microscope study of the disposition and fine structure of Mycobacterium lepraemurium in mouse spleen." J Bacteriol **77**(2): 205-211.

Chatterjee, D., S. W. Hunter, M. McNeil and P. J. Brennan (1992). "Lipoarabinomannan. Multiglycosylated form of the mycobacterial mannosylphosphatidylinositols." J Biol Chem **267**(9): 6228-6233.

Chong, A. K., M. S. Pegg, N. R. Taylor and M. von Itzstein (1992). "Evidence for a sialosyl cation transition-state complex in the reaction of sialidase from influenza virus." Eur J Biochem **207**(1): 335-343.

Cole, S. T. (2016). "Inhibiting Mycobacterium tuberculosis within and without." Philos Trans R Soc Lond B Biol Sci **371**(1707).

Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead and B. G. Barrell (1998). "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence." Nature **393**(6685): 537-544.

Connolly, L. E., P. H. Edelstein and L. Ramakrishnan (2007). "Why is long-term therapy required to cure tuberculosis?" PLoS Med **4**(3): e120.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman (1995). "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules." Journal of the American Chemical Society **117**(19): 5179-5197.

Corper, H. J. and M. L. Cohn (1933). "The viability and virulence of old cultures of tubercle bacilli. Studies in twelve-year broth cultures maintained at incubator temperature." Am Rev Tuberc **28**: 856-874.

Costa, J., N. Empadinhas and M. S. da Costa (2007). "Glucosylglycerate biosynthesis in the deepest lineage of the Bacteria: characterization of the thermophilic proteins GpgS and GpgP from Persephonella marina." J Bacteriol **189**(5): 1648-1654.

Costa, J., N. Empadinhas, L. Goncalves, P. Lamosa, H. Santos and M. S. da Costa (2006). "Characterization of the biosynthetic pathway of glucosylglycerate in the archaeon Methanococcoides burtonii." J Bacteriol **188**(3): 1022-1030.

Cottaz, S., B. Henrissat and H. Driguez (1996). "Mechanism-based inhibition and stereochemistry of glucosinolate hydrolysis by myrosinase." Biochemistry **35**(48): 15256-15259.

Cowtan, K. (1994). "DM: an automated procedure for phase improvement by density modification." Joint CCP4 and ESF-EACBM newsletter on protein crystallography **31**: 34-38.

Cowtan, K. (2000). "General quadratic functions in real and reciprocal space and their application to likelihood phasing." Acta Crystallographica Section D **56**(12): 1612-1621.

Cowtan, K. (2006). "The Buccaneer software for automated model building. 1. Tracing protein chains." Acta Crystallogr D Biol Crystallogr **62**(Pt 9): 1002-1011.

Crans, D. C., J. J. Smee, E. Gaidamauskas and L. Yang (2004). "The chemistry and biochemistry of vanadium and the biological activities exerted by vanadium compounds." Chem Rev **104**(2): 849-902.

Crans, D. C., M. L. Tarlton and C. C. McLauchlan (2014). "Trigonal Bipyramidal or Square Pyramidal Coordination Geometry? Investigating the Most Potent Geometry for Vanadium Phosphatase Inhibitors." European Journal of Inorganic Chemistry **2014**(27): 4450-4468.

d'Avo, A. F., S. Cunha, A. Mingote, P. Lamosa, M. S. da Costa and J. Costa (2013). "A Unique Pool of Compatible Solutes on Rhodopirellula baltica, Member of the Deep-Branching Phylum Planctomycetes." PLoS One **8**(6): e68289.

Daffe, M. and P. Draper (1998). "The envelope layers of mycobacteria with reference to their pathogenicity." Adv Microb Physiol **39**: 131-203.

Daniel, J., C. Deb, V. S. Dubey, T. D. Sirakova, B. Abomoelak, H. R. Morbidoni and P. E. Kolattukudy (2004). "Induction of a novel class of diacylglycerol acyltransferases and triacylglycerol accumulation in Mycobacterium tuberculosis as it goes into a dormancy-like state in culture." J Bacteriol **186**(15): 5017-5030.

Davies, D. R. and W. G. Hol (2004). "The power of vanadate in crystallographic investigations of phosphoryl transfer enzymes." FEBS Lett **577**(3): 315-321.

Davies, D. R., H. Interthal, J. J. Champoux and W. G. Hol (2002). "Insights into substrate binding and catalytic mechanism of human tyrosyl-DNA phosphodiesterase (Tdp1) from vanadate and tungstate-inhibited structures." J Mol Biol **324**(5): 917-932.

Davies, D. R., B. L. Staker, J. A. Abendroth, T. E. Edwards, R. Hartley, J. Leonard, H. Kim, A. L. Rychel, S. N. Hewitt, P. J. Myler and L. J. Stewart (2011). "An ensemble of structures of Burkholderia pseudomallei 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase." Acta Crystallogr Sect F Struct Biol Cryst Commun **67**(Pt 9): 1044-1050.

Davies, G. and B. Henrissat (1995). "Structures and mechanisms of glycosyl hydrolases." Structure **3**(9): 853-859.

Dawson, D. J. (2000). "Mycobacterial terminology." J Clin Microbiol **38**(10): 3913.

de Jonge, M. R., L. H. Koymans, J. E. Guillemont, A. Koul and K. Andries (2007). "A computational model of the inhibition of Mycobacterium tuberculosis ATPase by a new drug candidate R207910." Proteins **67**(4): 971-980.

de Sanctis, D., A. Beteva, H. Caserotto, F. Dobias, J. Gabadinho, T. Giraud, A. Gobbo, M. Guijarro, M. Lentini, B. Lavault, T. Mairs, S. McSweeney, S. Petitdemange, V. Rey-Bakaikoa, J. Surr, P. Theveneau, G. A. Leonard and C. Mueller-Dieckmann (2012). "ID29: a high-intensity highly automated ESRF beamline for macromolecular crystallography experiments exploiting anomalous scattering." J Synchrotron Radiat **19**(Pt 3): 455-461.

DeJesus, M. A., E. R. Gerrick, W. Xu, S. W. Park, J. E. Long, C. C. Boutte, E. J. Rubin, D. Schnappinger, S. Ehrt, S. M. Fortune, C. M. Sassetti and T. R. Ioerger (2017). "Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis." MBio **8**(1).

Dennis Chasteen, N. (1983). The biochemistry of vanadium. Copper, Molybdenum, and Vanadium in Biological Systems. B. A. Averill, L. R. Briggs, N. D. Chasteen et al. Berlin, Heidelberg, Springer Berlin Heidelberg**:** 105-138.

Dhar, N. and J. D. McKinney (2007). "Microbial phenotypic heterogeneity and antibiotic tolerance." Curr Opin Microbiol **10**(1): 30-38.

Downing, K. J., V. V. Mischenko, M. O. Shleeva, D. I. Young, M. Young, A. S. Kaprelyants, A. S. Apt and V. Mizrahi (2005). "Mutants of Mycobacterium tuberculosis lacking three of the five rpf-like genes are defective for growth in vivo and for resuscitation in vitro." Infect Immun **73**(5): 3038-3043.

Draper, P., K. H. Khoo, D. Chatterjee, A. Dell and H. R. Morris (1997). "Galactosamine in walls of slow-growing mycobacteria." Biochem J **327 (Pt 2)**: 519-525.

Drouillard, S., S. Armand, G. J. Davies, C. E. Vorgias and B. Henrissat (1997). "Serratia marcescens chitobiase is a retaining glycosidase utilizing substrate acetamido group participation." Biochem J **328 (Pt 3)**: 945-949.

Dupeux, F., M. Rower, G. Seroul, D. Blot and J. A. Marquez (2011). "A thermal stability assay can help to estimate the crystallization likelihood of biological samples." Acta Crystallogr D Biol Crystallogr **67**(Pt 11): 915-919.

Duran, D., S. Le Couster, K. Desjardins, A. Delmotte, G. Fox, R. Meijers, T. Moreno, M. Savko and W. Shepard (2013). "PROXIMA 2A – A New Fully Tunable Micro-focus Beamline for Macromolecular Crystallography." Journal of Physics Conferences Series **425**: 012005.

Eldar, A. and M. B. Elowitz (2010). "Functional roles for noise in genetic circuits." Nature **467**(7312): 167-173.

Empadinhas, N., L. Albuquerque, V. Mendes, S. Macedo-Ribeiro and M. S. da Costa (2008). "Identification of the mycobacterial glucosyl-3-phosphoglycerate synthase." FEMS Microbiol Lett **280**(2): 195-202.

Empadinhas, N. and M. S. da Costa (2011). "Diversity, biological roles and biosynthetic pathways for sugar-glycerate containing compatible solutes in bacteria and archaea." Environ Microbiol **13**(8): 2056-2077.

Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.

Engelbrecht, F., K. Marin and M. Hagemann (1999). "Expression of the ggpS gene, involved in osmolyte synthesis in the marine cyanobacterium *Synechococcus* sp. strain PCC 7002, revealed regulatory differences between this strain and the freshwater strain *Synechocystis* sp. strain PCC 6803." Appl Environ Microbiol **65**(11): 4822-4829.

Esteban, J. and M. C. Muñoz-Egea (2016). "Mycobacterium bovis and Other Uncommon Members of the Mycobacterium tuberculosis Complex." Microbiol Spectr **4**(6).

Falkinham, J. O. (2010). "Impact of human activities on the ecology of nontuberculous mycobacteria." Future Microbiol **5**(6): 951-960.

Faria, T. Q., A. Mingote, F. Siopa, R. Ventura, C. Maycock and H. Santos (2008). "Design of new enzyme stabilizers inspired by glycosides of hyperthermophilic microorganisms." Carbohydr Res **343**(18): 3025-3033.

Ferguson, J. A. and C. E. Ballou (1970). "Biosynthesis of a mycobacterial lipopolysaccharide. Properties of the polysaccharide methyltransferase." J Biol Chem **245**(16): 4213-4223.

Fernandes, C., V. Mendes, J. Costa, N. Empadinhas, C. Jorge, P. Lamosa, H. Santos and M. S. da Costa (2010). "Two alternative pathways for the synthesis of the rare compatible solute mannosylglucosylglycerate in Petrotoga mobilis." J Bacteriol **192**(6): 1624-1633.

Ferreira, C., A. R. Soares, P. Lamosa, M. A. Santos and M. S. da Costa (2016). "Comparison of the compatible solute pool of two slightly halophilic planctomycetes species, Gimesia maris and Rubinisphaera brasiliensis." Extremophiles **20**(6): 811-820.

Finkel, S. E. (2006). "Long-term survival during stationary phase: evolution and the GASP phenotype." Nat Rev Microbiol **4**(2): 113-120.

Finn, R. D., P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. A. Salazar, J. Tate and A. Bateman (2016). "The Pfam protein families database: towards a more sustainable future." Nucleic Acids Res **44**(D1): D279-285.

Flick, P. K. and K. Bloch (1974). "*In vitro* alterations of the product distribution of the fatty synthetase from Mycobacterium phlei." J Biol Chem **249**(4): 1031-1036.

Flick, P. K. and K. Bloch (1975). "Reversible inhibition of the fatty acid synthetase complex from Mycobacterium smegmatis by palmitoyl-coenzyme A." J Biol Chem **250**(9): 3348-3351.

Flot, D., T. Mairs, T. Giraud, M. Guijarro, M. Lesourd, V. Rey, D. van Brussel, C. Morawe, C. Borel, O. Hignette, J. Chavanne, D. Nurizzo, S. McSweeney and E. Mitchell (2010). "The ID23-2 structural biology microfocus beamline at the ESRF." J Synchrotron Radiat **17**(1): 107-118.

Foster, J. D., S. E. Young, T. D. Brandt and R. C. Nordlie (1998). "Tungstate: a potent inhibitor of multifunctional glucose-6-phosphatase." Arch Biochem Biophys **354**(1): 125-132.

Fothergill-Gilmore, L. A. and H. C. Watson (1989). "The phosphoglycerate mutases." Adv Enzymol Relat Areas Mol Biol **62**: 227-313.

Fraga, J., A. Maranha, V. Mendes, P. J. Pereira, N. Empadinhas and S. Macedo-Ribeiro (2015). "Structure of mycobacterial maltokinase, the missing link in the essential GlgE-pathway." Sci Rep **5**: 8026.

Friedberg, I. (2006). "Automated protein function prediction--the genomic challenge." Brief Bioinform **7**(3): 225-242.

Gomez, J. E. and J. D. McKinney (2004). "M. tuberculosis persistence, latency, and drug tolerance." Tuberculosis (Edinb) **84**(1-2): 29-44.

Good, N. E., G. D. Winget, W. Winter, T. N. Connolly, S. Izawa and R. M. M. Singh (1966). "Hydrogen Ion Buffers for Biological Research*." Biochemistry **5**(2): 467-477.

Gordon, J. A. (1991). Use of vanadate as protein-phosphotyrosine phosphatase inhibitor. Methods in Enzymology, Academic Press. **201:** 477-482.

Gray, G. R. and C. E. Ballou (1971). "Isolation and characterization of a polysaccharide containing 3-O-methyl-D-mannose from Mycobacterium phlei." J Biol Chem **246**(22): 6835-6842.

Grellert, E. and C. E. Ballou (1972). "Biosynthesis of a mycobacterial lipopolysaccharide. Evidence for an acylpolysaccharide methyltransferase." J Biol Chem **247**(10): 3236-3241.

Griffin, J. E., J. D. Gawronski, M. A. Dejesus, T. R. Ioerger, B. J. Akerley and C. M. Sassetti (2011). "High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism." PLoS Pathog **7**(9): e1002251.

Griffith, D. E., T. Aksamit, B. A. Brown-Elliott, A. Catanzaro, C. Daley, F. Gordin, S. M. Holland, R. Horsburgh, G. Huitt, M. F. Iademarco, M. Iseman, K. Olivier, S. Ruoss, C. F. von Reyn, R. J. Wallace, Jr., K. Winthrop, A. T. S. M. D. Subcommittee, S. American Thoracic and A. Infectious Disease Society of (2007). "An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases." Am J Respir Crit Care Med **175**(4): 367-416.

Griffith, D. E., B. A. Brown-Elliott, J. L. Benwill and R. J. Wallace, Jr. (2015). "Mycobacterium abscessus. "Pleased to meet you, hope you guess my name..."." Ann Am Thorac Soc **12**(3): 436-439.

Haagsma, A. C., I. Podasca, A. Koul, K. Andries, J. Guillemont, H. Lill and D. Bald (2011). "Probing the interaction of the diarylquinoline TMC207 with its target mycobacterial ATP synthase." PLoS One **6**(8): e23575.

Hamada, K., M. Kato, T. Shimizu, K. Ihara, T. Mizuno and T. Hakoshima (2005). "Crystal structure of the protein histidine phosphatase SixA in the multistep His-Asp phosphorelay." Genes Cells **10**(1): 1-11.

Han, X. Y., Y. H. Seo, K. C. Sizer, T. Schoberle, G. S. May, J. S. Spencer, W. Li and R. G. Nair (2008). "A new Mycobacterium species causing diffuse lepromatous leprosy." Am J Clin Pathol **130**(6): 856-864.

Hao, Q. (2004). "ABS: a program to determine absolute configuration and evaluate anomalous scatterer substructure." Journal of Applied Crystallography **37**(3): 498-499.

Harris, L. S. and G. R. Gray (1977). "Acetylated methylmannose polysaccharide of Streptomyces." J Biol Chem **252**(8): 2470-2477.

Henkle, E. and K. L. Winthrop (2015). "Nontuberculous mycobacteria infections in immunosuppressed hosts." Clin Chest Med **36**(1): 91-99.

Hoffmann, C., A. Leis, M. Niederweis, J. M. Plitzko and H. Engelhardt (2008). "Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure." Proc Natl Acad Sci U S A **105**(10): 3963-3967.

Hosaka, K. and S. Yamashita (1984). "Regulatory role of phosphatidate phosphatase in triacylglycerol synthesis of Saccharomyces cerevisiae." Biochim Biophys Acta **796**(1): 110-117.

Hunter, S. W., H. Gaylord and P. J. Brennan (1986). "Structure and antigenicity of the phosphorylated lipopolysaccharide antigens from the leprosy and tubercle bacilli." J Biol Chem **261**(26): 12345-12351.

Ilton, M., A. W. Jevans, E. D. McCarthy, D. Vance, H. B. White, 3rd and K. Bloch (1971). "Fatty acid synthetase activity in Mycobacterium phlei: regulation by polysaccharides." Proc Natl Acad Sci U S A **68**(1): 87-91.

Ireton, G. C. and B. L. Stoddard (2004). "Microseed matrix screening to improve crystals of yeast cytosine deaminase." Acta Crystallogr D Biol Crystallogr **60**(Pt 3): 601-605.

Jackson, M. (2014). "The mycobacterial cell envelope-lipids." Cold Spring Harb Perspect Med **4**(10).

Jackson, M. and P. J. Brennan (2009). "Polymethylated polysaccharides from Mycobacterium species revisited." J Biol Chem **284**(4): 1949-1953.

Jackson, M., G. Stadthagen and B. Gicquel (2007). "Long-chain multiple methyl-branched fatty acid-containing lipids of Mycobacterium tuberculosis: biosynthesis, transport, regulation and biological activities." Tuberculosis (Edinb) **87**(2): 78-86.

Jankute, M., J. A. Cox, J. Harrison and G. S. Besra (2015). "Assembly of the Mycobacterial Cell Wall." Annu Rev Microbiol **69**: 405-423.

Jankute, M., S. Grover, H. L. Birch and G. S. Besra (2014). "Genetics of Mycobacterial Arabinogalactan and Lipoarabinomannan Assembly." Microbiol Spectr **2**(4): MGM2-0013-2013.

Jones, S. E. and J. T. Lennon (2010). "Dormancy contributes to the maintenance of microbial diversity." Proc Natl Acad Sci U S A **107**(13): 5881-5886.

Jorge, C. D., P. Lamosa and H. Santos (2007). "Alpha-D-mannopyranosyl-(1-->2)-alpha-D-glucopyranosyl-(1-->2)-glycerate in the thermophilic bacterium Petrotoga miotherma--structure, cellular content and function." FEBS J **274**(12): 3120-3127.

Kabsch, W. (2010). "Xds." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 125-132.

Kabsch, W. and C. Sander (1983). "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features." Biopolymers **22**(12): 2577-2637.

Kalscheuer, R., K. Syson, U. Veeraraghavan, B. Weinrick, K. E. Biermann, Z. Liu, J. C. Sacchettini, G. Besra, S. Bornemann and W. R. Jacobs, Jr. (2010). "Self-poisoning of Mycobacterium tuberculosis by targeting GlgE in an alpha-glucan pathway." Nat Chem Biol **6**(5): 376-384.

Kamisango, K., A. Dell and C. E. Ballou (1987). "Biosynthesis of the mycobacterial O-methylglucose lipopolysaccharide. Characterization of putative intermediates in the initiation, elongation, and termination reactions." J Biol Chem **262**(10): 4580-4586.

Kana, B. D., B. G. Gordhan, K. J. Downing, N. Sung, G. Vostroktunova, E. E. Machowski, L. Tsenova, M. Young, A. Kaprelyants, G. Kaplan and V. Mizrahi (2008). "The resuscitation-promoting factors of Mycobacterium tuberculosis are required for virulence and resuscitation from dormancy but are collectively dispensable for growth in vitro." Mol Microbiol **67**(3): 672-684.

Kao, Y. H., L. Lerner and T. G. Warner (1997). "Stereoselectivity of the Chinese hamster ovary cell sialidase: sialoside hydrolysis with overall retention of configuration." Glycobiology **7**(4): 559-563.

Kapopoulou, A., J. M. Lew and S. T. Cole (2011). "The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes." Tuberculosis (Edinb) **91**(1): 8-13.

Kasperbauer, S. H. and M. A. De Groote (2015). "The treatment of rapidly growing mycobacterial infections." Clin Chest Med **36**(1): 67-78.

Kataoka, T. and S. Nojima (1967). "The phospholipid compositions of some Actinomycetes." Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism **144**(3): 681-683.

Kaur, D., M. E. Guerin, H. Skovierova, P. J. Brennan and M. Jackson (2009). "Chapter 2: Biogenesis of the cell wall and other glycoconjugates of Mycobacterium tuberculosis." Adv Appl Microbiol **69**: 23-78.

Kaur, D., H. Pham, G. Larrouy-Maumus, M. Riviere, V. Vissa, M. E. Guerin, G. Puzo, P. J. Brennan and M. Jackson (2009). "Initiation of methylglucose lipopolysaccharide biosynthesis in mycobacteria." PLoS One **4**(5): e5447.

Kawaguchi, A. and K. Bloch (1976). "Inhibition of glutamate dehydrogenase and malate dehydrogenases by palmitoyl coenzyme A." J Biol Chem **251**(5): 1406-1412.

Keller, J. and C. E. Ballou (1968). "The 6-O-methylglucose-containing lipopolysaccharide of Mycobacterium phlei. Identification of the lipid components." J Biol Chem **243**(11): 2905-2910.

Khoo, K. H., A. Dell, H. R. Morris, P. J. Brennan and D. Chatterjee (1995). "Structural definition of acylated phosphatidylinositol mannosides from Mycobacterium tuberculosis: definition of a common anchor for lipomannan and lipoarabinomannan." Glycobiology **5**(1): 117-127.

Klahn, S., C. Steglich, W. R. Hess and M. Hagemann (2010). "Glucosylglycerate: a secondary compatible solute common to marine cyanobacteria from nitrogen-poor environments." Environ Microbiol **12**(1): 83-94.

Kleist, S., M. Ulbrich, N. Bill, K. Schmidt-Hohagen, R. Geffers and D. Schomburg (2017). "Dealing with salinity extremes and nitrogen limitation - an unexpected strategy of the marine bacterium Dinoroseobacter shibae." Environ Microbiol **19**(3): 894-908.

Kohanski, M. A., D. J. Dwyer and J. J. Collins (2010). "How antibiotics kill bacteria: from targets to networks." Nat Rev Microbiol **8**(6): 423-435.

Koliwer-Brandl, H., K. Syson, R. van de Weerd, G. Chandra, B. Appelmelk, M. Alber, T. R. Ioerger, W. R. Jacobs, Jr., J. Geurtsen, S. Bornemann and R. Kalscheuer (2016).

"Metabolic Network for the Biosynthesis of Intra- and Extracellular alpha-Glucans Required for Virulence of Mycobacterium tuberculosis." PLoS Pathog **12**(8): e1005768.

Kollman, V. H., J. L. Hanners, R. E. London, E. G. Adame and T. E. Walker (1979). "Photosynthetic preparation and characterization of ^{13}C-labelled carbohydrates in *Agmenellum quadruplicatum*." Carbohydr Res **73**: 193-202.

Konarev, P. V., V. V. Volkov, A. V. Sokolova, M. H. J. Koch and D. I. Svergun (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." Journal of Applied Crystallography **36**(5): 1277-1282.

Koshland, D. E. (1953). "STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS." Biological Reviews **28**(4): 416-436.

Krissinel, E. and K. Henrick (2004). "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2256-2268.

Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." J Mol Biol **372**(3): 774-797.

Kumar, G., S. Guan and P. A. Frantom (2014). "Biochemical characterization of the retaining glycosyltransferase glucosyl-3-phosphoglycerate synthase from Mycobacterium tuberculosis." Arch Biochem Biophys **564**: 120-127.

Kurakata, Y., A. Uechi, H. Yoshida, S. Kamitori, Y. Sakano, A. Nishikawa and T. Tonozuka (2008). "Structural insights into the substrate specificity and function of Escherichia coli K12 YgjK, a glucosidase belonging to the glycoside hydrolase family 63." J Mol Biol **381**(1): 116-128.

Lamosa, P., M. V. Rodrigues, L. G. Goncalves, J. Carr, R. Ventura, C. Maycock, N. D. Raven and H. Santos (2013). "Organic solutes in the deepest phylogenetic branches of the Bacteria: identification of alpha(1-6)glucosyl-alpha(1-2)glucosylglycerate in Persephonella marina." Extremophiles **17**(1): 137-146.

Langer, G., S. X. Cohen, V. S. Lamzin and A. Perrakis (2008). "Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7." Nat Protoc **3**(7): 1171-1179.

Layre, E., D. C. Paepe, G. Larrouy-Maumus, J. Vaubourgeix, S. Mundayoor, B. Lindner, G. Puzo and M. Gilleron (2011). "Deciphering sulfoglycolipids of Mycobacterium tuberculosis." J Lipid Res **52**(6): 1098-1110.

Lee, D. C., M. A. Cottrill, C. W. Forsberg and Z. Jia (2003). "Functional insights revealed by the crystal structures of Escherichia coli glucose-1-phosphatase." J Biol Chem **278**(33): 31412-31418.

Lee, Y. C. (1966). "Isolation and characterization of lipopolysaccharides containing 6-O-methyl-D-glucose from Mycobacterium species." J Biol Chem **241**(8): 1899-1908.

Lee, Y. C. and C. E. Ballou (1964). "6-O-Methyl-D-Glucose from Mycobacteria." J Biol Chem **239**: PC3602-3603.

Lemassu, A. and M. Daffé (1994). "Structural features of the exocellular polysaccharides of Mycobacterium tuberculosis." Biochem J **297 (Pt 2)**: 351-357.

Levin, B. R. and D. E. Rozen (2006). "Non-inherited antibiotic resistance." Nat Rev Microbiol **4**(7): 556-562.

Lewis, K. (2007). "Persister cells, dormancy and infectious disease." Nat Rev Microbiol **5**(1): 48-56.

Lewis, K. (2010). "Persister cells." Annu Rev Microbiol **64**: 357-372.

Li, J., X. Guan, N. Shaw, W. Chen, Y. Dong, X. Xu, X. Li and Z. Rao (2014). "Homotypic dimerization of a maltose kinase for molecular scaffolding." Sci Rep **4**: 6418.

Lin, K., L. Li, J. J. Correia and S. J. Pilkis (1992). "Glu327 is part of a catalytic triad in rat liver fructose-2,6-bisphosphatase." J Biol Chem **267**(10): 6556-6562.

Lindenberger, J. J., S. K. Veleti, B. N. Wilson, S. J. Sucheck and D. R. Ronning (2015). "Crystal structures of Mycobacterium tuberculosis GlgE and complexes with non-covalent inhibitors." Sci Rep **5**: 12830.

Lipworth, S., R. J. Hammond, V. O. Baron, Y. Hu, A. Coates and S. H. Gillespie (2016). "Defining dormancy in mycobacterial disease." Tuberculosis (Edinb) **99**: 131-142.

Liu, L., I. Siuda, M. R. Richards, J. Renaud, E. N. Kitova, P. M. Mayer, D. P. Tieleman, T. L. Lowary and J. S. Klassen (2016). "Structure and Stability of Carbohydrate-Lipid Interactions. Methylmannose Polysaccharide-Fatty Acid Complexes." Chembiochem **17**(16): 1571-1578.

Liu, Q. P., G. Sulzenbacher, H. Yuan, E. P. Bennett, G. Pietz, K. Saunders, J. Spence, E. Nudelman, S. B. Levery, T. White, J. M. Neveu, W. S. Lane, Y. Bourne, M. L. Olsson, B. Henrissat and H. Clausen (2007). "Bacterial glycosidases for the production of universal red blood cells." Nat Biotechnol **25**(4): 454-464.

Lombard, V., H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat (2014). "The carbohydrate-active enzymes database (CAZy) in 2013." Nucleic Acids Res **42**(Database issue): D490-495.

Lopez-Varela, E., A. L. Garcia-Basteiro, B. Santiago, D. Wagner, J. van Ingen and B. Kampmann (2015). "Non-tuberculous mycobacteria in children: muddying the waters of tuberculosis diagnosis." Lancet Respir Med **3**(3): 244-256.

Lourenco, E. C., C. D. Maycock and M. Rita Ventura (2009). "Synthesis of potassium (2R)-2-O-alpha-d-glucopyranosyl-(1-->6)-alpha-d-glucopyranosyl-2,3-dihydroxypropa noate a natural compatible solute." Carbohydr Res **344**(15): 2073-2078.

Lourenco, E. C. and M. R. Ventura (2011). "The synthesis of compatible solute analogues-solvent effects on selective glycosylation." Carbohydr Res **346**(2): 163-168.

Machida, Y. and K. Bloch (1973). "Complex formation between mycobacterial polysaccharides and fatty acyl-CoA derivatives." Proc Natl Acad Sci U S A **70**(4): 1146-1148.

Maggi, N., C. R. Pasqualucci, R. Ballotta and P. Sensi (1966). "Rifampicin: a new orally active rifamycin." Chemotherapy **11**(5): 285-292.

Maggio, J. E. (1980). "Structure of a mycobacterial polysaccharide-fatty acyl-CoA complex: nuclear magnetic resonance studies." Proc Natl Acad Sci U S A **77**(5): 2582-2586.

Maitra, S. K. and C. E. Ballou (1977). "Heterogeneity and refined structures of 3-O-methyl-D-mannose polysaccharides from Mycobacterium smegmatis." J Biol Chem **252**(8): 2459-2469.

Maloney, D. H. and C. E. Ballou (1980). "Polymethylpolysaccharide synthesis in an ethionine-resistant mutant of Mycobacterium smegmatis." J Bacteriol **141**(3): 1217-1221.

Maranha, A., P. J. Moynihan, V. Miranda, E. Correia Lourenco, D. Nunes-Costa, J. S. Fraga, P. Jose Barbosa Pereira, S. Macedo-Ribeiro, M. R. Ventura, A. J. Clarke and N. Empadinhas (2015). "Octanoylation of early intermediates of mycobacterial methylglucose lipopolysaccharides." Sci Rep **5**: 13610.

Mark, B. L., D. J. Vocadlo, S. Knapp, B. L. Triggs-Raine, S. G. Withers and M. N. James (2001). "Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase." J Biol Chem **276**(13): 10330-10337.

Marrakchi, H., M. A. Laneelle and M. Daffé (2014). "Mycolic acids: structures, biosynthesis, and beyond." Chem Biol **21**(1): 67-85.

McCarter, J. D. and S. G. Withers (1994). "Mechanisms of enzymatic glycoside hydrolysis." Curr Opin Struct Biol **4**(6): 885-892.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.

McNeil, M., M. Daffe and P. J. Brennan (1990). "Evidence for the nature of the link between the arabinogalactan and peptidoglycan of mycobacterial cell walls." J Biol Chem **265**(30): 18200-18206.

McNeil, M., M. Daffe and P. J. Brennan (1991). "Location of the mycolyl ester substituents in the cell walls of mycobacteria." J Biol Chem **266**(20): 13217-13223.

Mendes, V., A. Maranha, S. Alarico, M. S. da Costa and N. Empadinhas (2011). "Mycobacterium tuberculosis Rv2419c, the missing glucosyl-3-phosphoglycerate phosphatase for the second step in methylglucose lipopolysaccharide biosynthesis." Sci Rep **1**: 177.

Mendes, V., A. Maranha, S. Alarico and N. Empadinhas (2012). "Biosynthesis of mycobacterial methylglucose lipopolysaccharides." Nat Prod Rep **29**(8): 834-844.

Mendes, V., A. Maranha, P. Lamosa, M. S. da Costa and N. Empadinhas (2010). "Biochemical characterization of the maltokinase from Mycobacterium bovis BCG." BMC Biochem **11**: 21.

Miah, F., M. J. Bibb, J. E. Barclay, K. C. Findlay and S. Bornemann (2016). "Developmental delay in a Streptomyces venezuelae glgE null mutant is associated with the accumulation of alpha-maltose 1-phosphate." Microbiology **162**(7): 1208-1219.

Miah, F., H. Koliwer-Brandl, M. Rejzek, R. A. Field, R. Kalscheuer and S. Bornemann (2013). "Flux through trehalose synthase flows from trehalose to the alpha anomer of maltose in mycobacteria." Chem Biol **20**(4): 487-493.

Miljković, M. (2010). Conformational Analysis of Monosaccharides. Carbohydrates, Springer, New York, NY.

Minnikin, D. E. (1982). "Complex lipids: their chemistry, biosynthesis and roles." The biology of mycobacteria: 94-184.

Mita, M. and I. Yasumasu (1981). "Inhibition of dihydrofolate reductase by palmitoyl coenzyme A." Int J Biochem **13**(2): 229-232.

Miyazaki, T., M. Ichikawa, H. Iino, A. Nishikawa and T. Tonozuka (2015). "Crystal structure and substrate-binding mode of GH63 mannosylglycerate hydrolase from Thermus thermophilus HB8." J Struct Biol **190**(1): 21-30.

Moise, G., N. M. Gallup, A. N. Alexandrova, A. C. Hengge and S. J. Johnson (2015). "Conservative tryptophan mutants of the protein tyrosine phosphatase YopH exhibit impaired WPD-loop function and crystallize with divanadate esters in their active sites." Biochemistry **54**(42): 6490-6500.

Monaco, S., E. Gordon, M. W. Bowler, S. Delageniere, M. Guijarro, D. Spruce, O. Svensson, S. M. McSweeney, A. A. McCarthy, G. Leonard and M. H. Nanao (2013). "Automatic processing of macromolecular crystallography X-ray diffraction data at the ESRF." J Appl Crystallogr **46**(Pt 3): 804-810.

Mueller-Dieckmann, C., M. W. Bowler, P. Carpentier, D. Flot, A. A. McCarthy, M. H. Nanao, D. Nurizzo, P. Pernot, A. Popov, A. Round, A. Royant, D. de Sanctis, D. von Stetten and G. A. Leonard (2015). "The status of the macromolecular crystallography beamlines at the European Synchrotron Radiation Facility." The European Physical Journal Plus **130**(4): 70.

Murakami, T., T. Kanai, H. Takata, T. Kuriki and T. Imanaka (2006). "A novel branching enzyme of the GH-57 family in the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1." J Bacteriol **188**(16): 5915-5924.

Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**(Pt 3): 240-255.

Nessar, R., E. Cambau, J. M. Reyrat, A. Murray and B. Gicquel (2012). "Mycobacterium abscessus: a new antibiotic nightmare." J Antimicrob Chemother **67**(4): 810-818.

Nikaido, H., S. H. Kim and E. Y. Rosenberg (1993). "Physical organization of lipids in the cell wall of Mycobacterium chelonae." Mol Microbiol **8**(6): 1025-1030.

Nikawa, J. and S. Yamashita (1997). "Phosphatidylinositol synthase from yeast." Biochim Biophys Acta **1348**(1-2): 173-178.

Nobre, A., S. Alarico, A. Maranha, V. Mendes and N. Empadinhas (2014). "The molecular biology of mycobacterial trehalose in the quest for advanced tuberculosis therapies." Microbiology **160**(Pt 8): 1547-1570.

Nobre, A., N. Empadinhas, M. F. Nobre, E. C. Lourenco, C. Maycock, M. R. Ventura, A. Mingote and M. S. da Costa (2013). "The plant Selaginella moellendorffii possesses enzymes for synthesis and hydrolysis of the compatible solutes mannosylglycerate and glucosylglycerate." Planta **237**(3): 891-901.

Nouioui, I., V. Sangal, L. Carro, K. Teramoto, M. Jando, M. D. C. Montero-Calasanz, J. M. Igual, I. Sutcliffe, M. Goodfellow and H. P. Klenk (2017). "Two novel species of rapidly growing mycobacteria: Mycobacterium lehmannii sp. nov. and Mycobacterium neumannii sp. nov." Int J Syst Evol Microbiol.

Nunes-Costa, D., S. Alarico, M. P. Dalcolmo, M. Correia-Neves and N. Empadinhas (2016). "The looming tide of nontuberculous mycobacterial infections in Portugal and Brazil." Tuberculosis (Edinb) **96**: 107-119.

Nunes-Costa, D., A. Maranha, M. Costa, S. Alarico and N. Empadinhas (2017). "Glucosylglycerate metabolism, bioversatility and mycobacterial survival." Glycobiology **27**(3): 213-227.

Nurizzo, D., T. Mairs, M. Guijarro, V. Rey, J. Meyer, P. Fajardo, J. Chavanne, J. C. Biasci, S. McSweeney and E. Mitchell (2006). "The ID23-1 structural biology beamline at the ESRF." J Synchrotron Radiat **13**(Pt 3): 227-238.

Ortalo-Magne, A., M. A. Dupont, A. Lemassu, A. B. Andersen, P. Gounon and M. Daffé (1995). "Molecular composition of the outermost capsular material of the tubercle bacillus." Microbiology **141 (Pt 7)**: 1609-1620.

Ortalo-Magné, A., A. Lemassu, M. A. Laneelle, F. Bardou, G. Silve, P. Gounon, G. Marchal and M. Daffé (1996). "Identification of the surface-exposed lipids on the cell envelopes of Mycobacterium tuberculosis and other mycobacterial species." J Bacteriol **178**(2): 456-461.

Ostanin, K. and R. L. Van Etten (1993). "Asp304 of Escherichia coli acid phosphatase is involved in leaving group protonation." J Biol Chem **268**(28): 20778-20784.

Otwinowski, Z. (1991). "Maximum likelihood refinement of heavy atom parameters." Isomorphous replacement and anomalous scattering: 80-86.

Palomo, M., T. Pijning, T. Booiman, J. M. Dobruchowska, J. van der Vlist, S. Kralj, A. Planas, K. Loos, J. P. Kamerling, B. W. Dijkstra, M. J. van der Maarel, L. Dijkhuizen and H. Leemhuis (2011). "Thermus thermophilus glycoside hydrolase family 57 branching enzyme: crystal structure, mechanism of action, and products formed." J Biol Chem **286**(5): 3520-3530.

Panjikar, S., V. Parthasarathy, V. S. Lamzin, M. S. Weiss and P. A. Tucker (2005). "Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment." Acta Crystallogr D Biol Crystallogr **61**(Pt 4): 449-457.

Panjikar, S., V. Parthasarathy, V. S. Lamzin, M. S. Weiss and P. A. Tucker (2009). "On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination." Acta Crystallogr D Biol Crystallogr **65**(Pt 10): 1089-1097.

Pape, T. and T. R. Schneider (2004). "HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs." Journal of Applied Crystallography **37**(5): 843-844.

Peck, A., F. Sunden, L. D. Andrews, V. S. Pande and D. Herschlag (2016). "Tungstate as a Transition State Analog for Catalysis by Alkaline Phosphatase." J Mol Biol **428**(13): 2758-2768.

Pereira, P. J., N. Empadinhas, L. Albuquerque, B. Sa-Moura, M. S. da Costa and S. Macedo-Ribeiro (2008). "Mycobacterium tuberculosis glucosyl-3-phosphoglycerate synthase: structure of a key enzyme in methylglucose lipopolysaccharide biosynthesis." PLoS One **3**(11): e3748.

Pernot, P., A. Round, R. Barrett, A. De Maria Antolinos, A. Gobbo, E. Gordon, J. Huet, J. Kieffer, M. Lentini, M. Mattenet, C. Morawe, C. Mueller-Dieckmann, S. Ohlsson, W. Schmid, J. Surr, P. Theveneau, L. Zerrad and S. McSweeney (2013). "Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution." J Synchrotron Radiat **20**(Pt 4): 660-664.

Petoukhov, M. V., D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. Mertens, P. V. Konarev and D. I. Svergun (2012). "New developments in the ATSAS program package for small-angle scattering data analysis." J Appl Crystallogr **45**(Pt 2): 342-350.

Piersimoni, C. and C. Scarparo (2008). "Pulmonary infections associated with non-tuberculous mycobacteria in immunocompetent patients." Lancet Infect Dis **8**(5): 323-334.

Pitarque, S., G. Larrouy-Maumus, B. Payre, M. Jackson, G. Puzo and J. Nigou (2008). "The immunomodulatory lipoglycans, lipoarabinomannan and lipomannan, are exposed at the mycobacterial cell surface." Tuberculosis (Edinb) **88**(6): 560-565.

Pommier, M. T. and G. Michel (1981). "Structure of 2',3'-di-O-acyl-alpha-D-glucopyranosyl-(1 leads to 2)-D-glyceric acid, a new glycolipid from Nocardia caviae." Eur J Biochem **118**(2): 329-333.

Pommier, M. T. and G. Michel (1986). "Isolation and characterization of an O-methylglucose-containing lipopolysaccharide produced by Nocardia otitidis-caviarum." J Gen Microbiol **132**(9): 2433-2441.

Primm, T. P., C. A. Lucero and J. O. Falkinham, 3rd (2004). "Health impacts of environmental mycobacteria." Clin Microbiol Rev **17**(1): 98-106.

Rajan, S. S., X. Yang, F. Collart, V. L. Yip, S. G. Withers, A. Varrot, J. Thompson, G. J. Davies and W. F. Anderson (2004). "Novel catalytic mechanism of glycoside hydrolysis based on the structure of an NAD+/Mn2+ -dependent phospho-alpha-glucosidase from Bacillus subtilis." Structure **12**(9): 1619-1629.

Rashid, A. M., S. F. Batey, K. Syson, H. Koliwer-Brandl, F. Miah, J. E. Barclay, K. C. Findlay, K. P. Nartowski, Y. Z. Khimyak, R. Kalscheuer and S. Bornemann (2016). "Assembly of alpha-Glucan by GlgE and GlgB in Mycobacteria and Streptomycetes." Biochemistry **55**(23): 3270-3284.

Rastogi, N., R. Hellio and H. L. David (1991). "A new insight into the mycobacterial cell envelope architecture by the localization of antigens in ultrathin sections." Zentralbl Bakteriol **275**(3): 287-302.

Rath, S. P., K. K. Rajak, S. Mondal and A. Chakravorty (1998). "Synthesis and structure of vanadate esters of glycerol and propane-1,3-diol." Journal of the Chemical Society, Dalton Transactions(12): 2097-2102.

Rehder, D. (2013). "Vanadium. Its role for humans." Met Ions Life Sci **13**: 139-169.

Rehder, D. (2015). "The role of vanadium in biology." Metallomics **7**(5): 730-742.

Reiter, N. J., D. J. White and F. Rusnak (2002). "Inhibition of bacteriophage lambda protein phosphatase by organic and oxoanion inhibitors." Biochemistry **41**(3): 1051-1059.

Rengarajan, J., B. R. Bloom and E. J. Rubin (2005). "Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages." Proc Natl Acad Sci U S A **102**(23): 8327-8332.

Rigden, D. J. (2008). "The histidine phosphatase superfamily: structure and function." Biochem J **409**(2): 333-348.

Rigden, D. J., J. E. Littlejohn, K. Henderson and M. J. Jedrzejas (2003). "Structures of phosphate and trivanadate complexes of Bacillus stearothermophilus phosphatase PhoE: structural and functional analysis in the cofactor-dependent phosphoglycerate mutase superfamily." J Mol Biol **325**(3): 411-420.

Rittershaus, E. S., S. H. Baek and C. M. Sassetti (2013). "The normalcy of dormancy: common themes in microbial quiescence." Cell Host Microbe **13**(6): 643-651.

Roy, R., V. Usha, A. Kermani, D. J. Scott, E. I. Hyde, G. S. Besra, L. J. Alderwick and K. Futterer (2013). "Synthesis of alpha-glucan in mycobacteria involves a hetero-octameric complex of trehalose synthase TreS and Maltokinase Pep2." ACS Chem Biol **8**(10): 2245-2255.

Saier, M. H., Jr. and C. E. Ballou (1968). "The 6-O-methylglucose-containig lipopolysaccharide of Mycobacterium phlei. Complet structure of the polysaccharide." J Biol Chem **243**(16): 4332-4341.

Sambou, T., P. Dinadayala, G. Stadthagen, N. Barilone, Y. Bordat, P. Constant, F. Levillain, O. Neyrolles, B. Gicquel, A. Lemassu, M. Daffe and M. Jackson (2008). "Capsular glucan and intracellular glycogen of Mycobacterium tuberculosis: biosynthesis and impact on the persistence in mice." Mol Microbiol **70**(3): 762-774.

Sani, M., E. N. Houben, J. Geurtsen, J. Pierson, K. de Punder, M. van Zon, B. Wever, S. R. Piersma, C. R. Jimenez, M. Daffe, B. J. Appelmelk, W. Bitter, N. van der Wel and P. J. Peters (2010). "Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins." PLoS Pathog **6**(3): e1000794.

Santos, C. R., C. C. Tonoli, D. M. Trindade, C. Betzel, H. Takata, T. Kuriki, T. Kanai, T. Imanaka, R. K. Arni and M. T. Murakami (2011). "Structural basis for branching-enzyme activity of glycoside hydrolase family 57: structure and stability studies of a novel branching enzyme from the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1." Proteins **79**(2): 547-557.

Sassetti, C. M., D. H. Boyd and E. J. Rubin (2003). "Genes required for mycobacterial growth defined by high density mutagenesis." Mol Microbiol **48**(1): 77-84.

Sato, S., D. Kitamoto and H. Habe (2014). "In vitro evaluation of glyceric acid and its glucosyl derivative, alpha-glucosylglyceric acid, as cell proliferation inducers and protective solutes." Biosci Biotechnol Biochem **78**(7): 1183-1186.

Sawangwan, T., C. Goedl and B. Nidetzky (2010). "Glucosylglycerol and glucosylglycerate as enzyme stabilizers." Biotechnol J **5**(2): 187-191.

Scheuermann, T. H., S. B. Padrick, K. H. Gardner and C. A. Brautigam (2016). "On the acquisition and analysis of microscale thermophoresis data." Analytical Biochemistry **496**: 79-93.

Seibert, F. B. (1949). "The isolation of three different proteins and two polysaccharides from tuberculin by alcohol fractionation; their chemical and biological properties." Am Rev Tuberc **59**(1): 86-101.

Shahraki, A. H., P. Heidarieh, S. Z. Bostanabad, A. D. Khosravi, M. Hashemzadeh, S. Khandan, M. Biranvand, D. E. Schraufnagel and M. Mirsaeidi (2015). ""Multidrug-resistant tuberculosis" may be nontuberculous mycobacteria." Eur J Intern Med **26**(4): 279-284.

Sheldrick, G. M. (2010). "Experimental phasing with SHELXC/D/E: combining chain tracing with density modification." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 479-485.

Shi, L., B. M. Sutter, X. Ye and B. P. Tu (2010). "Trehalose is a key determinant of the quiescent metabolic state that fuels cell cycle progression upon return to growth." Mol Biol Cell **21**(12): 1982-1990.

Shleeva, M., G. V. Mukamolova, M. Young, H. D. Williams and A. S. Kaprelyants (2004). "Formation of 'non-culturable' cells of Mycobacterium smegmatis in stationary phase in response to growth under suboptimal conditions and their Rpf-mediated resuscitation." Microbiology **150**(Pt 6): 1687-1697.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson and D. G. Higgins (2011). "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." Mol Syst Biol **7**: 539.

Sillje, H. H., J. W. Paalman, E. G. ter Schure, S. Q. Olsthoorn, A. J. Verkleij, J. Boonstra and C. T. Verrips (1999). "Function of trehalose and glycogen in cell cycle progression and cell viability in Saccharomyces cerevisiae." J Bacteriol **181**(2): 396-400.

Singh, P., A. Benjak, V. J. Schuenemann, A. Herbig, C. Avanzi, P. Busso, K. Nieselt, J. Krause, L. Vera-Cabrera and S. T. Cole (2015). "Insight into the evolution and origin of leprosy bacilli from the genome sequence of Mycobacterium lepromatosis." Proc Natl Acad Sci U S A **112**(14): 4459-4464.

Sirakova, T. D., C. Deb, J. Daniel, H. D. Singh, H. Maamar, V. S. Dubey and P. E. Kolattukudy (2012). "Wax ester synthesis is required for Mycobacterium tuberculosis to enter in vitro dormancy." PLoS One **7**(12): e51641.

Smith, C. S., A. Aerts, P. Saunderson, J. Kawuma, E. Kita and M. Virmond (2017). "Multidrug therapy for leprosy: a game changer on the path to elimination." Lancet Infect Dis **17**(9): e293-e297.

Smith, W. L. and C. E. Ballou (1973). "The 6-O-methylglucose-containing lipopolysaccharides of Mycobacterium phlei. Locations of the neutral and acidic acyl groups." J Biol Chem **248**(20): 7118-7125.

Stadthagen, G., T. Sambou, M. Guerin, N. Barilone, F. Boudou, J. Kordulakova, P. Charles, P. M. Alzari, A. Lemassu, M. Daffe, G. Puzo, B. Gicquel, M. Riviere and M. Jackson (2007). "Genetic basis for the biosynthesis of methylglucose lipopolysaccharides in Mycobacterium tuberculosis." J Biol Chem **282**(37): 27270-27276.

Steinmann, P., S. G. Reed, F. Mirza, T. D. Hollingsworth and J. H. Richardus (2017). "Innovative tools and approaches to end the transmission of Mycobacterium leprae." Lancet Infect Dis **17**(9): e298-e305.

Svensson, O., S. Malbet-Monaco, A. Popov, D. Nurizzo and M. W. Bowler (2015). "Fully automatic characterization and data collection from crystals of biological macromolecules." Acta Crystallogr D Biol Crystallogr **71**(Pt 8): 1757-1767.

Svergun, D., C. Barberato and M. H. J. Koch (1995). "CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates." Journal of Applied Crystallography **28**(6): 768-773.

Syson, K., C. E. Stevenson, M. Rejzek, S. A. Fairhurst, A. Nair, C. J. Bruton, R. A. Field, K. F. Chater, D. M. Lawson and S. Bornemann (2011). "Structure of Streptomyces maltosyltransferase GlgE, a homologue of a genetically validated anti-tuberculosis target." J Biol Chem **286**(44): 38298-38310.

Terwilliger, T. C. (2000). "Maximum-likelihood density modification." Acta Crystallogr D Biol Crystallogr **56**(Pt 8): 965-972.

Terwilliger, T. C. (2003). "Automated main-chain model building by template matching and iterative fragment extension." Acta Crystallographica Section D-Biological Crystallography **59**: 38-44.

Terwisscha van Scheltinga, A. C., S. Armand, K. H. Kalk, A. Isogai, B. Henrissat and B. W. Dijkstra (1995). "Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and X-ray structure of a complex with allosamidin: evidence for substrate assisted catalysis." Biochemistry **34**(48): 15619-15623.

Theveneau, P., R. Baker, R. Barrett, A. Beteva, M. W. Bowler, P. Carpentier, H. Caserotto, D. d. Sanctis, F. Dobias, D. Flot, M. Guijarro, T. Giraud, M. Lentini, G. A. Leonard, M. Mattenet, A. A. McCarthy, S. M. McSweeney, C. Morawe, M. Nanao, D. Nurizzo, S. Ohlsson, P. Pernot, A. N. Popov, A. Round, A. Royant, W. Schmid, A. Snigirev, J. Surr and C. Mueller-Dieckmann (2013). "The Upgrade Programme for the Structural Biology beamlines at the European Synchrotron Radiation Facility – High throughput sample evaluation and automation." Journal of Physics: Conference Series **425**(1): 012001.

Thomas, J. P., C. O. Baughn, R. G. Wilkinson and R. G. Shepherd (1961). "A new synthetic compound with antituberculous activity in mice: ethambutol (dextro-2,2'-(ethylenediimino)-di-l-butanol)." Am Rev Respir Dis **83**: 891-893.

Tiago, I., A. Maranha, V. Mendes, S. Alarico, P. J. Moynihan, A. J. Clarke, S. Macedo-Ribeiro, P. J. Pereira and N. Empadinhas (2012). "Genome sequence of Mycobacterium hassiacum DSM 44199, a rare source of heat-stable mycobacterial proteins." J Bacteriol **194**(24): 7010-7011.

Tian, X. X., A. Li, I. V. Farrugia, X. Mo, D. Crich and M. J. Groves (2000). "Isolation and identification of poly-alpha-(1-->4)-linked 3-O-methyl-D-mannopyranose from a hot-water extract of Mycobacterium vaccae." Carbohydr Res **324**(1): 38-44.

Todeschini, A. R., L. Mendonca-Previato, J. O. Previato, A. Varki and H. van Halbeek (2000). "Trans-sialidase from Trypanosoma cruzi catalyzes sialoside hydrolysis with retention of configuration." Glycobiology **10**(2): 213-221.

Toriyama, S., I. Yano, M. Masui, E. Kusunose, M. Kusunose and N. Akimori (1980). "Regulation of cell wall mycolic acid biosynthesis in acid-fast bacteria. I. Temperature-induced changes in mycolic acid molecular species and related compounds in Mycobacterium phlei." J Biochem **88**(1): 211-221.

Tortoli, E., T. Fedrizzi, C. J. Meehan, A. Trovato, A. Grottola, E. Giacobazzi, G. F. Serpini, S. Tagliazucchi, A. Fabio, C. Bettua, R. Bertorelli, F. Frascaro, V. De Sanctis, M. Pecorari, O. Jousson, N. Segata and D. M. Cirillo (2017). "The new phylogeny of the genus Mycobacterium: The old and the news." Infect Genet Evol **56**: 19-25.

Touw, W. G., C. Baakman, J. Black, T. A. te Beek, E. Krieger, R. P. Joosten and G. Vriend (2015). "A series of PDB-related databanks for everyday needs." Nucleic Acids Res **43**(Database issue): D364-368.

Tuffal, G., R. Albigot, M. Riviere and G. Puzo (1998). "Newly found 2-N-acetyl-2,6-dideoxy-beta-glucopyranose containing methyl glucose polysaccharides in M.bovis BCG: revised structure of the mycobacterial methyl glucose lipopolysaccharides." Glycobiology **8**(7): 675-684.

Tuffal, G., C. Ponthus, C. Picard, M. Riviere and G. Puzo (1995). "Structural elucidation of novel methylglucose-containing polysaccharides from Mycobacterium xenopi." Eur J Biochem **233**(1): 377-383.

Tung, K. K. and C. E. Ballou (1973). "Biosynthesis of a mycobacterial lipopolysaccharide. Properties of the polysaccharide: acyl coenzyme A acyltransferase reaction." J Biol Chem **248**(20): 7126-7133.

Vance, D. E., O. Mitsuhashi and K. Bloch (1973). "Purification and properties of the fatty acid synthetase from Mycobacterium phlei." J Biol Chem **248**(7): 2303-2309.

Vocadlo, D. J. and S. G. Withers (2005). "Detailed comparative analysis of the catalytic mechanisms of beta-N-acetylglucosaminidases from families 3 and 20 of glycoside hydrolases." Biochemistry **44**(38): 12809-12818.

Walter, R. A., J. Nairn, D. Duncan, N. C. Price, S. M. Kelly, D. J. Rigden and L. A. Fothergill-Gilmore (1999). "The role of the C-terminal region in phosphoglycerate mutase." Biochem J **337 (Pt 1)**: 89-95.

Wang, Y., L. Liu, Z. Wei, Z. Cheng, Y. Lin and W. Gong (2006). "Seeing the process of histidine phosphorylation in human bisphosphoglycerate mutase." J Biol Chem **281**(51): 39642-39648.

Watanabe, M., Y. Aoyagi, H. Mitome, T. Fujita, H. Naoki, M. Ridell and D. E. Minnikin (2002). "Location of functional groups in mycobacterial meromycolate chains; the

recognition of new structural principles in mycolic acids." Microbiology **148**(Pt 6): 1881-1902.

Watanabe, M., Y. Aoyagi, M. Ridell and D. E. Minnikin (2001). "Separation and characterization of individual mycolic acids in representative mycobacteria." Microbiology **147**(Pt 7): 1825-1837.

Watson, J. N., V. Dookhun, T. J. Borgford and A. J. Bennet (2003). "Mutagenesis of the conserved active-site tyrosine changes a retaining sialidase into an inverting sialidase." Biochemistry **42**(43): 12682-12690.

Watts, A. G., I. Damager, M. L. Amaya, A. Buschiazzo, P. Alzari, A. C. Frasch and S. G. Withers (2003). "Trypanosoma cruzi trans-sialidase operates through a covalent sialyl-enzyme intermediate: tyrosine is the catalytic nucleophile." J Am Chem Soc **125**(25): 7532-7533.

Weisman, L. S. and C. E. Ballou (1984). "Biosynthesis of the mycobacterial methylmannose polysaccharide. Identification of a 3-O-methyltransferase." J Biol Chem **259**(6): 3464-3469.

Weisman, L. S. and C. E. Ballou (1984). "Biosynthesis of the mycobacterial methylmannose polysaccharide. Identification of an α1→4-mannosyltransferase." J Biol Chem **259**(6): 3457-3463.

WHO (2013). The use of bedaquiline in the treatment of multidrug-resistant tuberculosis. Interim policy guidance. Geneva, Switzerland, World Health Organization.

WHO (2017). A 2016 review of available evidence on the use of bedaquiline in the treatment of multidrug-resistant tuberculosis. Geneva, Switzerland, World Health Organization.

WHO (2017). Global tuberculosis report 2017. Geneva, Switzerland, World Health Organization.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson (2011). "Overview of the CCP4 suite and current developments." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 235-242.

Wishart, D. S., Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox and M. Wilson (2017). "DrugBank 5.0: a major update to the DrugBank database for 2018." Nucleic Acids Res.

Withers, S. and S. Williams. (2017, 27 April 2017). "Glycoside hydrolases." CAZypedia Retrieved August 2017, 2017, from URL http://www.cazypedia.org.

Wood, W. I., D. O. Peterson and K. Bloch (1977). "Mycobacterium smegmatis fatty acid synthetase. A mechanism based on steady state rates and product distributions." J Biol Chem **252**(16): 5745-5749.

Xia, L. (2013). Studies of a mycobacterial α-(1→4)-mannosyltransferase involved in 3-O-methyl-mannose polysaccharide biosynthesis. Doctor of Philosophy, University of Alberta.

Xia, L., R. B. Zheng and T. L. Lowary (2012). "Revisiting the specificity of an alpha-(1-->4)-mannosyltransferase involved in mycobacterial methylmannose polysaccharide biosynthesis." Chembiochem **13**(8): 1139-1151.

Yabusaki, K. K. and C. E. Ballou (1978). "Interaction of mycobacterial polymethylpolysaccharides with paranaric acid and palmitoyl-coenzyme A: structural specificity and monomeric dissociation constants." Proc Natl Acad Sci U S A **75**(2): 691-695.

Yabusaki, K. K. and C. E. Ballou (1979). "Effect of polymethylpolysaccharides on the hydrolysis of palmitoyl coenzyme A by a thioesterase from Mycobacterium smegmatis." J Biol Chem **254**(24): 12314-12317.

Yabusaki, K. K., R. E. Cohen and C. E. Ballou (1979). "Conformational changes associated with complex formation between a mycobacterial polymethylpolysaccharide and palmitic acid." J Biol Chem **254**(15): 7282-7286.

Yamada, H., R. E. Cohen and C. E. Ballou (1979). "Characterization of 3-O-methyl-D-mannose polysaccharide precursors in Mycobacterium smegmatis." J Biol Chem **254**(6): 1972-1979.

Yano, I., Y. Furukawa and M. Kusunose (1969). "Phospholipids of Nocardia coeliaca." J Bacteriol **98**(1): 124-130.

Yeager, R. L., W. G. Munroe and F. I. Dessau (1952). "Pyrazinamide (aldinamide) in the treatment of pulmonary tuberculosis." Am Rev Tuberc **65**(5): 523-546.

Yip, V. L., A. Varrot, G. J. Davies, S. S. Rajan, X. Yang, J. Thompson, W. F. Anderson and S. G. Withers (2004). "An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 beta-glycosidase from Thermotoga maritima." J Am Chem Soc **126**(27): 8354-8355.

Zhang, M., M. Zhou, R. L. Van Etten and C. V. Stauffacher (1997). "Crystal structure of bovine low molecular weight phosphotyrosyl phosphatase complexed with the transition state analog vanadate." Biochemistry **36**(1): 15-23.

Zheng, Q., D. Jiang, W. Zhang, Q. Zhang, Q. Zhao, J. Jin, X. Li, H. Yang, M. Bartlam, N. Shaw, W. Zhou and Z. Rao (2014). "Mechanism of dephosphorylation of glucosyl-3-phosphoglycerate by a histidine phosphatase." J Biol Chem **289**(31): 21242-21251.

Zuber, B., M. Chami, C. Houssin, J. Dubochet, G. Griffiths and M. Daffe (2008). "Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state." J Bacteriol **190**(16): 5672-5680.

Zumla, A., P. Nahid and S. T. Cole (2013). "Advances in the development of new tuberculosis drugs and treatment regimens." Nat Rev Drug Discov **12**(5): 388-404.