UNIVERSIDADE D
COIMBRA

Daniela Cristina Nunes Costa

BIOCHEMICAL AND STRUCTURAL STUDIES OF A NOVEL GLUCOSAMINE KINASE AND A UNIQUE GLYCOSYLTRANSFERASE:

NEW ANTIMICROBIAL OPPORTUNITIES IN ACTINOBACTERIAL GLYCOBIOLOGY

Tese no âmbito do Doutoramento em Biologia Experimental e Biomedicina, ramo de Biologia Molecular, Celular e do Desenvolvimento, orientada pelo Doutor Nuno Miguel da Silva Empadinhas e coorientada pela Doutora Sandra de Macedo Ribeiro e pela Doutora Sandra Morais Cardoso e apresentada ao Instituto de Investigação Interdisciplinar da Universidade de Coimbra.

Dezembro de 2022

Daniela Cristina Nunes Costa

Biochemical and structural studies of a novel glucosamine kinase and a unique glycosyltransferase: new antimicrobial opportunities in actinobacterial glycobiology

Tese de Doutoramento em Biologia Experimental e Biomedicina, ramo de Biologia Molecular, Celular e do Desenvolvimento, orientada pelo Doutor Nuno Miguel da Silva Empadinhas e coorientada pela Doutora Sandra de Macedo Ribeiro e pela Doutora Sandra Morais Cardoso e apresentada ao Instituto de Investigação Interdisciplinar da Universidade de Coimbra

Dezembro de 2022

UNIVERSIDADE D
COIMBRA

Acknowledgments

Quero agradecer em primeiro lugar ao Doutor Nuno Empadinhas que, mais do que orientador científico deste trabalho, foi uma fonte constante de incentivo e apoio a todos os níveis. Sempre disponível para partilhar a sua experiência e conhecimentos, debater novas hipóteses e contrariar o meu frequente pessimismo. Sempre compreensivo mesmo quando a vida me foi trocando as voltas e obrigando a estabelecer prioridades. Agradeço toda a confiança que depositou em mim ao longo destes anos e a liberdade que me deu para explorar e testar novas ideias, tentando sempre encorajar a minha criatividade científica.

À Doutora Sandra de Macedo Ribeiro, que co-orientou este trabalho, por me ter recebido no seu laboratório, por todo o conhecimento partilhado, pelas muitas horas dedicadas a ensinar-me, especialmente quando o confinamento obrigou a que esse ensino ocorresse à distância. Agradeço toda a disponibilidade, paciência e empatia, sem o seu apoio não teria sido possível continuar a trabalhar a partir de casa quando assim teve que ser.

À Doutora Sandra Morais Cardoso por assumir a responsabilidade interna da tese.

A todos os meus companheiros de laboratório ao longo destes anos, que tanto me ensinaram e que me acompanharam no dia-a-dia. Em particular, um obrigada muito especial à Susana e à Ana, minhas "orientadoras de bancada" que tanto tempo dedicaram a ensinar-me, sempre pacientes e prontas a ajudar, foram-me transmitindo a sua experiência, conhecimentos e paixão pela ciência diariamente desde que entrei no grupo em 2013. Agradeço à Mafalda, à Inês, à Mariana e à Sara, colegas de bancada e de escritório que tornaram este percurso mais divertido e enriquecedor.

Ao Doutor Pedro Pereira, uma especial nota de agradecimento pela disponibilidade e preciosas dicas informáticas quando foi necessário preparar o meu computador para analisar dados de cristalografia à distância, sempre disponível para resolver os problemas que iam surgindo.

Aos membros dos grupos de trabalho da Doutora Sandra Macedo-Ribeiro e do Doutor Pedro Pereira no i3S, que tão bem me acolheram e integraram no laboratório durante as minhas viagens ao Porto. Um obrigada especial à Alexandra, ao Jose Antonio e ao Jorge pela preciosa colaboração, partilha de conhecimento e conversas esclarecedoras.

À Prof. Paula Veríssimo pela constante disponibilidade e todo o conhecimento partilhado sobre enzimologia, não só no decurso deste trabalho, mas desde a licenciatura e ao longo de todo o meu percurso científico, tendo em muito contribuído para a minha formação até aqui.

Às pessoas do 1° piso do CNC, em especial à equipa técnica do CNC pela ajuda constante e preciosa, e ao pessoal do almoço pelos momentos de descontração.

Aos meus amigos, com um agradecimento especial ao Marcelo, ao Pedro e ao João, que mesmo durante a pandemia se mantiveram presentes e graças aos nossos "copos virtuais" me foram distraindo e animando com a sua boa disposição, brincadeiras e palavras de incentivo. À malta do grupo de Teatro Amador de Brasfemes, em especial à Adriana, por me terem desafiado a juntar-me ao projecto enquanto fotógrafa, "obrigando-me" a sair de casa para desanuviar e me divertir mesmo em noites de Inverno em que não apetecia nada mas depois me sabia e fazia tão bem. À malta da anilhagem científica de aves, em especial ao Sr. Paulo, que nestes últimos meses mais difíceis me "abriu a porta" das suas sessões que para mim não foram só de anilhagem mas principalmente de terapia.

À minha família, especialmente aos meus pais e aos meus avós pelo apoio incondicional mesmo nos momentos mais difíceis, por todo o carinho e compreensão, pelos valores que me transmitiram ao longo da vida e pelo espírito de entreajuda constante, especialmente nestes últimos anos em que não tivemos alternativa senão unir-nos mais do que nunca. Aos que me viram começar este percurso mas não me viram acabar, a minha mãe, a avó Isolina e o tio Gil, sei o orgulho que tinham em mim e estou grata por tudo o que me deram em vida e pelas memórias que me deixaram.

To my boyfriend, Efraín, for the constant love and companionship, for being an unconditional source of emotional support, always there to remind me to keep taking care of myself and to never give up. Thank you for remaining close even while the pandemic forced us to be apart for so long, and for crossing the ocean as soon as it became possible just to recharge my batteries and give me the strength I needed to finish this work.

A todos os que contribuíram de alguma forma para a elaboração deste trabalho, o meu sincero obrigada.

Daniela

Funding

This work was primarily supported by Fundação para a Ciência e a Tecnologia (FCT) through grants PTDC/BTM-TEC/29221/2017 and PTDC/BIA-MIC/0122/2021 and in the framework of projects UIDB/04539/2020, UIDP/04539/2020 and LA/P/0058/2020 of the Research Unit and Associate Laboratory, Center for Innovative Biomedicine and Biotechnology (CIBB). DN-C acknowledges the support of FCT through PhD scholarship SFRH/BD/117777/2016.

Table of Contents

Abstract

The emergence and spread of antimicrobial resistance mechanisms, including against last-line antibiotics, coupled with decades of disinvestment in the discovery and development of new antimicrobials to replenish our arsenal, is creating a perfect storm that threatens many of the achievements on which modern medicine relies. Despite being recognised as one of the most serious global health threats by the World Health Organisation, there is still a lot to be done to tackle this crisis, from intensifying basic research on the subject to adopting effective policies. From a basic research point of view, one group of bacteria, commonly called actinomycetes, stands out as particularly interesting to study since it comprises both prolific antibiotic producers, streptomycetes, and some of the most threatening antibiotic-resistant pathogens, mycobacteria. Since actinomycetes are a source of problems but can also offer solutions, understanding their physiology and metabolism is critical. Their glycobiology is of special interest, given that both the beneficial and pathogenic sides of actinomycetes are tightly linked to the unique sugar-containing molecules they produce.

While pursuing the characterisation of a duplicate gene of the mycobacterial maltokinase, an enzyme involved in α-1,4-glucan biosynthesis through a non-canonical pathway which has attracted significant interest as a potential drug target in mycobacteria, we identified a novel family of actinobacterial glucosamine kinases (GlcNK). The phylogenetic distribution of the gene encoding this GlcNK was surprising and suggested links to secondary metabolism, leading us to focus on the homologue from *Streptacidiphilus jiangxiensis*, member of a prolific family of antibiotic producers. Its structural characterisation unveiled details of the catalytic mechanism of this novel family of glucosamine kinases, while bioinformatic analyses of the oddly structured biosynthetic gene cluster it belongs to revealed hints of its participation in the biosynthesis of a unique secondary metabolite with potential antimicrobial activity. Within the topic of α-1,4-glucan biosynthesis in mycobacteria, another enzyme was also investigated, glucosyl-3-phosphoglycerate synthase (GpgS), which catalyses the first committed step in the biosynthesis of the unique intracellular methylglucose lipopolysaccharides (MGLP). GpgS has been extensively studied and has garnered interest as a potential drug target, but high-resolution 3D structures, essential to guide rational drug design efforts, were still missing. After establishing *Mycobacterium hassiacum* as a genuine thermophilic species of mycobacteria and predicting it to be a source of stable mycobacterial proteins, high-resolution structures of its GpgS homologue were successfully obtained. These structures not only provided new insights into the active site of this important enzyme, but also suggested intriguing and yet unsuspected links with other crucial metabolic pathways, including the folate pathway, the target of many important antimicrobials.

The results presented in this work demonstrate how studying the glycobiology of actinomycetes can, sometimes unexpectedly, open new avenues of research with the potential to spur the discovery of novel antimicrobial strategies.

Resumo

O surgimento e disseminação de mecanismos de resistência antimicrobiana, incluindo contra antibióticos de última linha, juntamente com décadas de desinvestimento na descoberta e desenvolvimento de novos compostos antimicrobianos para reforçar o nosso arsenal terapêutico, estão a criar uma tempestade perfeita que ameaça muitas das conquistas em que a medicina moderna assenta. Apesar de reconhecida como uma das mais sérias ameaças à saúde global pela Organização Mundial de Saúde, ainda há muito por fazer para responder a esta crise, desde intensificar a investigação fundamental na área a adoptar políticas eficazes. No que respeita à investigação fundamental, um grupo de bactérias, os actinomicetos, destaca-se como sendo de interesse especial por conter tanto produtores prolíficos de antibióticos, os streptomicetos, como alguns dos mais ameaçadores e resistentes agentes patogénicos, as micobactérias. Sendo os actinomicetos uma fonte de problemas, mas também de soluções, é fundamental estudar a sua fisiologia e metabolismo e, em particular, a sua glicobiologia já que tanto os aspectos benéficos como os ligados à sua patogenicidade dependem em grande parte dos carboidratos únicos que produzem.

Ao estudar um gene duplicado da maltocinase, uma enzima envolvida na biossíntese de α-1,4-glucanos através de uma via não-canónica que tem atraído interesse como potencial alvo terapêutico nas micobactérias, identificámos uma família nova de glucosamina cinases (GlcNK). A distribuição filogenética surpreendente do gene que codifica esta GlcNK sugeriu ligações ao metabolismo secundário, levando-nos a focar no homólogo de *Streptacidiphilus jiangxiensis*, membro de uma família prolífica de produtores de antibióticos. A sua caracterização estrutural mostrou detalhes do mecanismo catalítico desta nova família de GlcNK, e análises bioinformáticas do cluster biossintético a que pertence revelaram pistas da sua participação na produção de um metabolito secundário único e com potencial actividade microbiana. Também em relação à biossíntese de α-1,4-glucanos em micobactérias, investigámos a enzima glicosil-3-phosphoglicerato sintase (GpgS), que catalisa o primeiro passo da via de biossíntese dos lipopolissacáridos de metilglucose (MGLP). Esta enzima tem sido amplamente estudada e tem despertado interesse enquanto potencial alvo terapêutico, mas até à data faltavam estruturas 3D de alta resolução para guiar esforços de desenho racional de fármacos. Após demonstrar que a micobactéria *Mycobacterium hassiacum* é um termófilo genuíno e prever que seria uma fonte de proteínas estáveis, foi possível obter estruturas de alta resolução da sua GpgS, que permitem conhecer melhor detalhes do local activo desta importante enzima e sugerem ligações intrigantes a outras vias metabólicas cruciais, incluindo a via de biossíntese do ácido fólico, conhecida por ser o alvo de importantes compostos antimicrobianos.

Assim, os resultados apresentados neste trabalho exemplificam como o estudo da glicobiologia dos actinomicetos pode, por vezes de forma inesperada, abrir novos caminhos de investigação e estimular a descoberta de novas estratégias antimicrobianas.

List of Abbreviations

3-PGA	3-phosphoglycerate
ABC	ATP-binding cassette
AG	arabinogalactan
APH	aminoglycoside phosphotransferase
Ara*f*	D-arabinofuranosyl
BGC	biosynthetic gene cluster
CAZYmes	carbohydrate-active enzymes
CBD	chitin-binding domain
DGG	glucosyl-α-1,6-glucosylglycerate
DKP	diketopiperazine
ELK	eukaryotic-like kinase
ePK	eukaryotic protein kinase
F6P	fructose-6-phosphate
FAS	fatty acid synthase
G1P	glucose-1-phosphate
G6P	glucose-6-phosphate
Gal*f*	D-galactofuranosyl
GalN	galactosamine
GG	glucosylglycerate
Glc	glucose
GlcN	glucosamine
GlcN1P	glucosamine-1-phosphate
GlcN6P	glucosamine-6-phosphate
GlcNAc	*N*-acetylglucosamine
GlcNAc1P	*N*-acetylglucosamine-1-phosphate
GlcNAc6P	*N*-acetylglucosamine-6-phosphate
GLPs	glycopeptidolipids
GND	genome neighbourhood diagrams
GNN	genome neighbourhood network
GPG	glucosyl-3-phosphoglycerate
GT	glycosyltransferase
IM	inner membrane
Ins	*myo*-inositol
Ins-P	*myo*-inositol-1-phosphate
LAMs	lipoarabinomannans
LMs	lipomannans
LOS	lipooligosaccharides

MIP	maltose-1-phosphate
MA	mycolic acid
mAGP	mycoloylarabinogalactan-peptidoglycan complex
MDR	multidrug-resistant
MG	mannosylglycerate
MGLP	6-*O*-methylglucose lipopolysaccharides
MMP	3-*O*-methylmannose polysaccharides
MSH	mycothiol
MTR	methylthioribose
MurGly	*N*-glycolylmuramic acid
MurNAc	*N*-acetylmuramic acid
NTM	nontuberculous mycobacteria
OM	outer membrane
*p*ABA	*p*-aminobenzoic acid
PDIMs	phthiocerol dimycocerosates
PEP	phosphoenolpyruvate
PG	peptidoglycan
PGLs	phenolic glycolipids
*p*HBA	*p*-hydroxybenzaldehyde
Pi	inorganic phosphate
PIMs	phosphatidylinositol mannosides
PKA	eukaryotic cAMP-dependent protein kinase
PKL	protein kinase-like
PMPS	polymethylated polysaccharides
PTS	phosphotransferase system
RGM	rapidly growing mycobacteria
RMSD	root mean square deviation
SGM	slowly growing mycobacteria
SIS	sugar isomerase
SM	secondary metabolites
SSN	sequence similarity networks
TB	tuberculosis
TDM	trehalose dimycolates
TLC	thin-layer chromatography
TMM	trehalose monomycolates
WHO	World Health Organization
WSH	water of standardised hardness
XDR	extensively drug-resistant

List of Enzymes

APH	aminoglycoside phosphotransferases, kinases acting on aminoglycoside antibiotics, involved in antibiotic resistance
CDPS	cyclodipeptide synthases, catalyse the formation of cyclodipeptides using aminoacylated-tRNA as substrates, involved in secondary metabolism
FolP1	dihydropteroate synthase, condenses *p*-aminobenzoic acid with 6-hydroxymethyl-7,8-dihydropterin diphosphate to form 7,8-dihydropteroate, involved in folate biosynthesis
FolP2	paralogue of FolP1, unknown function
GgH	glucosylglycerate hydrolase, cleaves glucosylglycerate to release free glucose and glycerate, involved in recovery from nitrogen stress in some NTM
GlcNK	glucosamine kinase, phosphorylates glucosamine to glucosamine-6-phosphate
GlgA	glycogen synthase, glycosyltransferase which elongates α-1,4-linked glucose chains using ADP-glucose, involved in glycogen biosynthesis (canonical pathway)
GlgB	α-1,4-glucan branching enzyme, transfers oligoglucans from the non-reducing end of a linear α-1,4-chain to the C-6 hydroxyl group of a glucose residue within the chain, involved in glycogen biosynthesis (canonical pathway)
GlgC	ADP-glucose pyrophosphorylase, produces ADP-glucose from glucose-1-phosphate and ATP, the first committed step in glycogen biosynthesis (canonical pathway)
GlgE	maltosyltransferase, uses maltose-1-phosphate to elongate α-1,4-glucan chains two glucoses at a time, involved in α-1,4-glucan biosynthesis
GlgM	maltose-1-phosphate synthase, condenses ADP-glucose and glucose-1-phosphate to produce maltose-1-phosphate, involved in α-1,4-glucan biosynthesis
GlmM	phosphoglucosamine mutase, converts glucosamine-6-phosphate into glucosamine-1-phosphate

GlmS	glucosamine-6-phosphate synthase, transfers an amino group from glutamine to fructose-6-phosphate producing glutamate and glucosamine-6-phosphate, the first committed step in amino sugar biosynthesis
GlmU	*N*-Acetylglucosamine-1-phosphate uridyltransferase, bifunctional enzyme whose acetyltransferase domain produces *N*-acetylglucosamine-1-phosphate and whose uridyltransferase domain subsequently forms UDP-GlcNAc, a cell wall building block
GpgP	glucosyl-3-phosphoglycerate phosphatase, dephosphorylates glucosyl-3-phosphoglycerate, involved in glucosylglycerate biosynthesis
GpgS	glucosyl-3-phosphoglycerate synthase, condenses NDP-glucose and 3-phosphoglycerate to form glucosyl-3-phosphoglycerate, involved in glucosylglycerate biosynthesis
LOG	"lonely guy", cytokinin riboside 5′-monophosphate phosphoribohydrolase, converts inactive cytokinin nucleotides to the biologically active free-base form, involved in cytokinin biosynthesis
Mak	maltokinase, phosphorylates maltose to maltose-1-phosphate, involved in α-1,4-glucan biosynthesis
MpgP	mannosyl-3-phosphoglycerate phosphatase, involved in mannosylglycerate biosynthesis
MpgS	mannosyl-3-phosphoglycerate synthase, involved in mannosylglycerate biosynthesis
MshA-D	enzymes involved in mycothiol biosynthesis
MurA	UDP-*N*-acetylglucosamine enolpyruvyl transferase, transfers the enolpyruvyl moiety of phosphoenolpyruvate to UDP-GlcNAc, involved in peptidoglycan biosynthesis
MurB	UDP-*N*-acetylenolpyruvoylglucosamine reductase, catalyses the NADPH dependent reduction of the enolpyruvate moiety of the MurA product to D-lactate producing UDP-MurNAc, involved in peptidoglycan biosynthesis
MurC-F	transferases which assemble the peptide bridge of PG directly onto UDP-MurNAc
MurG	*N*-acetylglucosamine transferase, uses UDP-GlcNAc to incorporate GlcNAc into peptidoglycan

MurQ	*N*-acetylmuramic acid 6-phosphate hydrolase, cleaves the lactate group of MurNAc6P generating GlcNAc6P, involved in peptidoglycan recycling
NagA	*N*-acetylglucosamine-6-phosphate deacetylase, converts GlcNAc6P to GlcN6P, involved in peptidoglycan recycling
NagB	glucosamine-6-phosphate deaminase, converts glucosamine-6-phosphate to fructose-6-phosphate, involved in peptidoglycan recycling
NamH	UDP-*N*-acetylmuramic acid hydroxylase, catalyses the NADPH-dependent hydroxylation of UDP-MurNAc to UDP-MurGly, precursor of the unique MurGly residues in mycobacterial peptidoglycan
NRPS	nonribosomal peptide synthases, modular enzymes which produce nonribosomal peptides, usually involved in secondary metabolism
OctT	octanoyltransferase proposed to be involved in MGLP biosynthesis
OtsA	trehalose-6-phospate synthase, condenses NDP-glucose and glucose-6-phosphate to form trehalose-6-phosphate, involved in trehalose biosynthesis
OtsB	trehalose-6-phosphate phosphatase, dephosphorylates trehalose-6-phosphate, involved in trehalose biosynthesis
PGM	phosphoglucomutase, interconverts glucose 1-phosphate and glucose 6-phosphate
PKS	polyketide synthases, modular enzymes which produce polyketides, often involved in secondary metabolism
Rv3032	putative α-1,4-glucosyltransferase, involved in MGLP biosynthesis
TreS	trehalose synthase, isomerises between maltose and trehalose, involved in maltose biosynthesis from trehalose
TreY	maltooligosyltrehalose synthase, isomerises a terminal maltose unit of a glycogen chain to trehalose, involved in trehalose biosynthesis
TreZ	maltooligosyltrehalose trehalohydrolase, cleaves the product of TreY to release free trehalose, involved in trehalose biosynthesis
UDPGP	UTP–glucose-1-phosphate uridylyltransferase, catalyses the formation of UDP-glucose from glucose-1-phosphate and UTP

List of Figures

List of Tables

Chapter 1 – General Introduction

Chapter 1 – Table of Contents

1. The threat of a post-antibiotic world

A world without antibiotics was a reality less than a century ago. Before the discovery of antimicrobial compounds, physicians were faced with infected wounds or common infections such as pneumonia for which they could do little but wait and let the infection run its natural course. In a world without antibiotics, common infectious diseases could not be treated, any infected wound could become life-threatening, and even the simplest surgery could result in a deadly infection. Thus, it is no surprise that the discovery and introduction into clinical practice of these compounds resulted in a true medical revolution, contributing to a decrease in mortality and an unprecedented increase in average life expectancy (Hutchings et al., 2019). Two key scientific landmarks are acknowledged as starting points of this revolution: Ehrlich's seminal work leading to the synthesis of the first antimicrobial agent, salvarsan, marketed in 1910 as the first effective treatment against syphilis (Bosch & Rosich, 2008); and the discovery and development of the first natural antibiotic, penicillin, between 1928 and 1945 by the combined work of Fleming, Florey, Heatley and Chain (Gaynes, 2017).

Although Ehrlich's work established the principles of chemotherapy and showed the potential of synthetic chemistry and compound screening as valuable tools in pharmacology, the discovery of penicillin was particularly important as it was also the discovery of a seemingly endless source of bioactive natural products, directly leading to what is called the "golden age" of antibiotic discovery that peaked in the mid-1950s (Hutchings et al., 2019). Antibiotics are secondary metabolites (SM) produced by bacteria and fungi to kill or inhibit the growth of other competing microorganisms within their natural environment. During this golden age of discovery, many novel antibiotics with varied spectrums of activity were isolated at a fast rate, allowing humanity to quickly harness a diverse and substantial arsenal of weapons to fight infectious diseases (Hutchings et al., 2019).

Right at the beginning of the 20th century, Ehrlich discovered that microorganisms were able to acquire resistance to antimicrobial compounds through continued exposure (Bosch & Rosich, 2008). The same observation was made by Fleming with penicillin, leading him to end his Nobel acceptance speech with a warning that misuse of penicillin could render it useless due to the ease with which microbes were able to acquire resistance when exposed to low doses of the antibiotic (Fleming, 1945). However, and despite these early warnings, widespread antibiotic (mis)use and a decline in the rate of antibiotic

discovery and development has led, in recent decades, to an explosion of antimicrobial resistance, which threatens to render the antibiotics at our disposal obsolete and leave humanity once again at the mercy of once easily-handled infections. Indeed, the emergence of extensively drug-resistant and pan-drug-resistant pathogens indicates that we may be heading towards a post-antibiotic era (Theuretzbacher et al., 2020). The World Health Organization (WHO) has recognised this problem as a global health emergency that "threatens the very core of modern medicine" and which requires an "all-out effort" (WHO, 2015). To promote a more focused approach, in 2017 the WHO established a priority list of pathogens towards which action needs to be urgently directed, attributing critical priority to *Mycobacterium tuberculosis* and drug-resistant gram-negative bacteria such as *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacteriaceae* (Tacconelli et al., 2018).

Currently, the antimicrobial development pipeline is weak and lacking in innovation (Theuretzbacher et al., 2020). Although it is crucial to invest into prevention and alternatives to antimicrobials, such as phage therapy or predatory bacteria (Gordillo Altamirano & Barr, 2019; Tyson & Sockett, 2017), leaving the pipeline to dry completely is an unacceptable risk. To find new replacement molecules that can enter the pipeline and replenish our antimicrobial arsenal, both Ehrlich's and Fleming's approaches remain valid avenues of research, which can now be re-invented in light of modern scientific advances, namely the unprecedented availability of genomic and structural data. Following Ehrlich's footsteps, identification of novel drug targets and the detailed study of their structure now allows for synthetic chemistry to be guided by rational drug design and molecular docking strategies in order to discover and develop novel chemical compounds with therapeutic potential (Chernov et al., 2019). Following Fleming's footsteps, genome sequencing and innovative cultivation strategies have revealed that fungi and bacteria still harbour many undiscovered SM which were missed during the "golden age" of antibiotic discovery, due to not being easily produced under the laboratory growth conditions used at the time (Belknap et al., 2020; Katz & Baltz, 2016; Ling et al., 2015). To answer WHO's call for an all-out effort, all avenues of research should be pursued and combined to maximise our chances of finding urgently needed antimicrobials and avoid the worst scenarios of a post-antibiotic world.

2. The class *Actinomycetia*

Actinomycetia (formerly known as *Actinobacteria*) is the largest class within the phylum *Actinobacteria*, one of the largest lineages in the domain Bacteria (Salam et al., 2020). The common name "actinomycetes", meaning "ray fungi", is often used to refer to members of this group, alluding to their typical mycelial growth (forming rays/threads) which led them to be initially classified as fungi (Wiliams, 1990). According to the most recent taxonomic update, this class now includes 34 orders and important families such as the *Actinomycetaceae*, *Bifidobacteriaceae*, *Corynebacteriaceae*, *Gordoniaceae*, *Mycobacteriaceae*, *Nocardiaceae* and *Streptomycetaceae* to name a few (Salam et al., 2020). Highly diverse in terms of morphology, physiology and metabolic activities, members of this medically important class are generally characterised as gram-positive bacteria with a high G+C content (Gao & Gupta, 2012).

Most species within *Actinomycetia* are environmental saprophytes and this group is well represented among microbial soil and water communities in both terrestrial and marine environments (Bhatti et al., 2017; Subramani & Aalbersberg, 2012). It also includes notorious pathogens responsible for well-known diseases such as tuberculosis (*Mycobacterium tuberculosis*) and diphtheria (*Corynebacterium diphtheriae*), important and ubiquitous members of the human microbiome such as *Bifidobacterium* species, and prolific antibiotic producers namely within the genus *Streptomyces* (Doroghazi & Metcalf, 2013; O'Callaghan & van Sinderen, 2016). Standing as essential members of microbial ecosystems, as important players in plant and animal health, as a key source of antibiotics and other important SM, and as relevant agents of disease, actinomycetes are undoubtedly of great importance at many levels. Thus, there is an obvious interest in strengthening our understanding of both pathogenic and environmental species, discovering new enzymes, deciphering unique metabolic pathways, and gathering knowledge that may prove helpful in humanity's ongoing struggle with infectious diseases.

2.1. Mycobacteria: the family *Mycobacteriaceae*

One of the most clinically important families within *Actinomycetia* is the *Mycobacteriaceae*, better known for harbouring the etiological agents of tuberculosis (TB) and leprosy. Members of this family are described as aerobic to microaerophilic, rod-shaped, nonmotile and non-sporulating bacteria with abnormal grain-stain and acid–alcohol-fast properties attributed to a waxy, mycolic acid-rich cell wall (Gupta et al., 2018).

Despite these shared characteristics, mycobacteria are a diverse group of microorganisms with nearly 200 validly published species to date (https://lpsn.dsmz.de/genus/mycobacterium) and including not only obligate human pathogens, but also opportunistic bacteria with a broad range of clinical manifestations and environmental saprophytes inhabiting an impressive variety of environments (see sections 2.1.1, 2.1.2 and 2.1.3). Considering this diversity, mycobacteria have been traditionally divided into groups, as shown in Figure 1.1, and have also been the subject of a recent major taxonomic reclassification as discussed below.

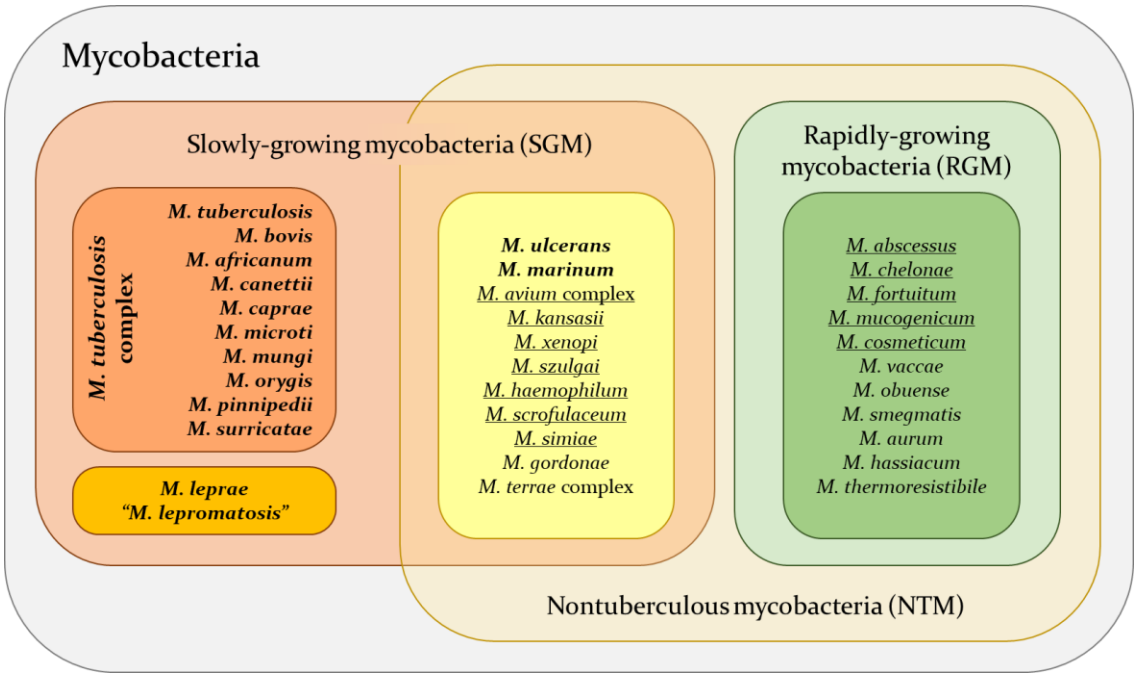


Figure 1.1. Members of the family *Mycobacteriaceae*. Representative species of mycobacteria are grouped according to growth rate and their association with disease. Bolded names represent true pathogens, underlined names represent opportunistic pathogens, and the remaining species are only rarely associated with disease. Adapted from Johansen et al., 2020.

Other traditional divisions of mycobacteria have been based on the diseases they are associated with and on their growth rate and pigment production. Based on the disease caused, they can be divided into three groups: species that cause TB, comprising the *Mycobacterium tuberculosis* complex (Gagneux, 2018); species that cause leprosy, namely *Mycobacterium leprae* and "*Mycobacterium lepromatosis*" (Han & Silva, 2014); and nontuberculous mycobacteria (NTM, also called atypical or environmental mycobacteria), some of which can be opportunistic and cause pulmonary and other infections (Primm et al., 2004). Historically, NTM have been further divided into four groups according to the Runyon classification, introduced in 1959 by Ernest Runyon to facilitate their study upon recognition of their diversity and clinical significance (Runyon, 1959). Group I, II and III

are slowly growing mycobacteria (meaning they take longer than seven days to form visible colonies on solid media) which produce a yellow-orange pigment only when exposed to light (photochromogens, group I), regardless of exposure to light (scotochromogens, group II) or never produce pigment (nonchromogens, group III). Group IV are nonpigmented rapidly growing mycobacteria. While this classification is outdated, the division between slowly growing (SGM) and rapidly growing mycobacteria (RGM) is still in use and is generally supported by phylogenetic studies, which have revealed that rapid growers are more ancestral than slow growers, and that some species with intermediate growth belong to a clade that is interposed between the SGM and RGM species (Gupta et al., 2018; Tortoli, Fedrizzi, et al., 2017). The most pathogenic and well-known species, *M. tuberculosis* and *M. leprae*, belong to the SGM group.

Until recently, the family *Mycobacteriaceae* contained only one genus, *Mycobacterium*, accommodating all mycobacterial species described to date. A recent proposal by Gupta and co-workers proposed four novel genera within *Mycobacteriaceae* and emended the genus *Mycobacterium* to encompass only one of the five distinct mycobacterial clades, as supported by phylogenomics and comparative genomic studies (Gupta et al., 2018). Under this classification, SGM are divided into the genera *Mycobacterium*, *Mycolicibacter* and *Mycolicibacillus*, and RGM into the genera *Mycolicibacterium* and *Mycobacteroides*. Thorough analysis of the fundamental properties of cell morphology supports this separation into five distinct, monophyletic lineages (Yamada et al., 2020). However, the novel nomenclature has not been fully embraced and has been criticised by medical researchers for bringing unnecessary confusion and disruption into the healthcare setting, thus doing a disservice to clinicians and patients (Tortoli et al., 2019). It should be noted that both nomenclatures remain in use and are valid according to the principles and rules outlined in the Bacteriological Code. As such, the new genus and species names are heterotypic synonyms of the old names, in this case the basonym *Mycobacterium* (Tindall, 1999).

2.1.1. Tuberculosis and leprosy, millennia-old diseases

Two of the oldest diseases to have plagued humankind, tuberculosis and leprosy, are caused by mycobacteria. Both can leave characteristic palaeopathological marks that make them recognisable in human and animal remains, which together with detection of DNA sequences and lipid biomarkers can be used to diagnose ancient infections (Donoghue et

al., 2017). Definitive evidence of TB has been found in human remains from the early phases of the Neolithic (8.000-10.000 years ago) and in a 17.000-year-old Pleistocene bovid (Baker et al., 2015; Hershkovitz et al., 2008; Lee et al., 2015; Rothschild et al., 2001). The oldest skeletons with evidence of leprosy are also pre-historical, dating from the Late Copper Age, and molecular evidence has been found in a 2200-year-old Egyptian mummy (Köhler et al., 2017; Neukamm et al., 2020). In agreement with archaeological evidence, genomic studies also suggest an ancient relationship between humans and these pathogenic mycobacteria, supporting a model of co-evolution with the human host (Gagneux, 2018; Han & Silva, 2014).

In addition to being among the most ancient diseases known, they were also among the first to be recognised as being caused by bacteria. The etiological agent of leprosy, *M. leprae* or Hansen's bacillus, was discovered in 1873 by Gerhard Hansen and it was the first bacterium to be identified as a cause of disease in humans (Hansen, 1874). However, its study was hindered by the inability to culture the bacterium *in vitro*, which remains a challenge to this day, and difficulty in finding a reliable animal model, which only happened almost a century later with the discovery that *M. leprae* could be successfully replicated in mice footpads and could cause disease in armadillos (Shepard, 1960; Storrs et al., 1974). *M. leprae* is an intracellular parasite with a downsized genome, known to infect tissue macrophages of the dermis and Schwann cells of the peripheral nerves (Cole et al., 2001; Scollard et al., 2015). This leads to extensive skin and nerve damage, which may be permanent and result in severe disability, disfigurement and the associated social stigma and discrimination (Suzuki et al., 2012).

Less than a decade after Hansen's discovery, in 1882, Robert Koch identified *M. tuberculosis*, or Koch's bacillus, as the etiological agent of TB (Koch, 1882). Since then, other closely related species able to cause TB in humans and/or in animals have been found and are collectively known as the *M. tuberculosis* complex (Gagneux, 2018). The members of this complex diverge in terms of host species preference and geographical distribution, but all share a remarkable ability to survive and multiply within lung macrophages upon inhalation by ingeniously manipulating the cell's phagocytic and signalling pathways (Cambier et al., 2014; Quigley et al., 2017). This strategy allows *M. tuberculosis* to persist in the infected host for decades in a latent asymptomatic stage. When *M. tuberculosis* is able to overcome this dormant state, either for reasons related with the virulence of the strain or with the immune status of the host, active pulmonary TB develops, with consequent host lung damage and dissemination to new uninfected

hosts (Ernst, 2012; Gengenbacher & Kaufmann, 2012; Lee et al., 2011). Although pulmonary TB is the most common presentation of the disease, *M. tuberculosis* has tropism for other organs besides the lung. TB can develop anywhere in body, originating cases of extrapulmonary TB and, when dissemination through the bloodstream occurs, of a serious condition called miliary TB (Peto et al., 2009; Sharma & Mohan, 2004; van Crevel et al., 2002).

Unfortunately, these millenary diseases still persist nowadays as important health problems worldwide. TB is still the leading cause of death by a single infectious agent and it is estimated that approximately 10 million people fell ill and 1.4 million died in 2019 with TB (WHO, 2020b). Leprosy has been eliminated as a public health problem since 2000 but its incidence remains high in some countries, with around 200.000 new cases reported every year, and it represents a significant cause of preventable lifelong disability (Smith et al., 2017; WHO, 2020a).

Many factors contribute to the difficulty in controlling these diseases, particularly among low socioeconomic status populations (Furin et al., 2019; Smith et al., 2017). Both have long incubation periods (from months to years) due to mycobacteria's slow growth and therapeutic regimens are also long and involve the combination of multiple drugs, often with significant side-effects, which lower quality of life and can affect patient's compliance with the treatment (Behr et al., 2018; de Andrade et al., 2019; Koo et al., 2020; Smith et al., 2017). Only one 100 year old vaccine is available to prevent mycobacterial infection and its effectiveness is suboptimal (Setia et al., 2006; Zwerling et al., 2011). Furthermore, mycobacteria are inherently resistant to most antimicrobials, due in part to their unique and impermeable cell envelope, and emergence of resistance against the few effective antimycobacterial drugs is a cause for worry (Gygli et al., 2017; Machado et al., 2018).

In leprosy's case, first-line treatment consists of rifampicin, clofazimine, and dapsone for 6-12 months, and second-line treatment may include minocycline, clarithromycin and quinolone and have a duration of 18 months (WHO, 2017). Resistance to rifampicin, dapsone and quinolones has been emerging and the lack of investment in new therapies, due to its status as a neglected tropical disease, could hamper treatment and worsen its epidemiological status (Cambau et al., 2018).

TB's case is even more alarming, which is why this pathogen is considered by the WHO a priority for the development of new antimicrobials (Tacconelli et al., 2018). First-

line TB treatment combines isoniazid, rifampicin, pyrazinamide and ethambutol, and several second-line drugs are available, but the emergence of multi-drug (MDR) and extensively drug-resistant (XDR) strains has led to almost impossible to treat cases of TB (Abubakar et al., 2013). After 40 years without discovering any new effective drug against TB, some progress has been made recently with the introduction of three new drugs, bedaquiline, delamanid and pretomanid, and promising clinical trials aiming to simplify the treatment of XDR-TB (Furin et al., 2019). However, toxicity and long treatment regimens remain a problem, and resistance to bedaquiline and delamanid has already been reported (Bloemberg et al., 2015; Polsfuss et al., 2019), so the quest for new and better therapies must continue.

2.1.2. Nontuberculous mycobacteria, an emerging threat

The obligate pathogens *M. leprae* and *M. tuberculosis* are the most infamous mycobacteria due to their pathogenicity. Nonetheless, they are far from being representative members of the *Mycobacteriaceae* family. The vast majority of mycobacteria are free-living saprophytes, ubiquitous in the environment, although not necessarily innocuous (Falkinham, 2009a). Thus, their lifestyle is closer to what is associated with typical soil and water-dwelling actinomycetes. Even rapidly growing mycobacteria grow much slower than most members of water and soil microbial communities, which could suggest they are poor competitors in some environments, but their remarkable ability to adapt to extreme stress conditions gives them an advantage when it comes to colonising and thriving in harsher environments (Falkinham, 2009b).

Generally, NTM have a high tolerance to pH and temperature extremes, desiccation, low nutrient and low oxygen levels, and to the presence of organic pollutants, heavy metals, disinfectants and antibiotics (Burgess et al., 2017; de Carvalho et al., 2020; Harland et al., 2008; Kirschner et al., 1999; Le Dantec et al., 2002; Meissner & Falkinham, 1984; Nasiri et al., 2017; Norton et al., 2004; Schulze-Robbecke & Buchholtz, 1992; Smeulders et al., 1999; Wang et al., 2006). Consequently, some natural habitats where NTM seem to thrive are acidic and nutrient-poor environments such as peat-rich boreal forest soils and waters, acidic brown-water swamps, estuaries, coastal swamps and acidic hot springs (Falkinham, 2009b; Niva et al., 2006; Santos et al., 2007). Another ecological niche NTM are remarkably adapted to is tap water, where nutrient scarcity, high temperatures and chlorination provide a hostile environment for many microbes but not for oligotrophic,

thermotolerant and chlorine-resistant NTM (Le Dantec et al., 2002; Schulze-Robbecke & Buchholtz, 1992). Furthermore, the ability of NTM to form biofilms, due to their high surface hydrophobicity, allows them to attach strongly to surfaces and avoid being diluted and washed out by high water flow rates (Falkinham, 2021; Williams et al., 2009). As a result, NTM can stably colonize point-of-use water sources such as showerheads, forming biofilms that render them even more resistant to antimicrobials and disinfectants (Feazel et al., 2009; Gebert et al., 2018). Unfortunately, this means NTM can become enriched in water distribution systems, including in the healthcare setting, significantly increasing human exposure to these opportunistic pathogens and causing nosocomial outbreaks (Decker & Palmore, 2014).

In developed countries with sophisticated water distribution systems, NTM infections are reported to be on a sustained rising trend, often surpassing the incidence of new TB infections (Johansen et al., 2020; Schildkraut et al., 2020). Their clinical manifestations can take many forms, including pulmonary disease, lymphadenitis, skin, soft tissue and osteoarticular infections and, in immunocompromised patients, disseminated infection (Swenson et al., 2018). As opportunistic bacteria, NTM disease is often associated with an underlying pathology affecting the immune system or the structure of respiratory airways, or with the colonisation of wounds and injuries (Swenson et al., 2018). Several hypotheses have been proposed to explain the rising trend of NTM infections in recent decades including: the increased overlap between human and mycobacterial habitats; NTM enrichment in water distributions systems due to disinfection regimens that favour their selection; greater prevalence of predisposing factors in the human population (e.g., aging, lowered immunocompetence, underlying lung disorders); and increased awareness by physicians, leading to more cases being diagnosed and reported (Nunes-Costa et al., 2016). These infections are notably difficult to treat due to NTM's intrinsic antimicrobial resistance and poor correlation between *in vitro* susceptibility and *in vivo* response to antibiotics (Griffith, 2010).

Although the large majority of NTM have been found in clinical samples, meaning that they are able to colonise or infect humans (Nunes-Costa et al., 2016), not all species are equally capable of causing disease. At the most pathogenic end of the spectrum is *Mycobacterium ulcerans*, responsible for the third most common mycobacterial disease after TB and leprosy, the tropical disease Buruli ulcer, which causes skin and bone lesions and can lead to permanent disability (Walsh et al., 2011). Tissue damage is mediated by mycolactone, a toxin produced only by *M. ulcerans* and closely related SGM, such as the

fish pathogen *Mycobacterium marinum*, also a common cause of cutaneous infections in humans (Yip et al., 2007). Pulmonary NTM disease, caused by the inhalation of aerosolized mycobacteria, is also common and frequently associated with members of the *Mycobacterium avium* complex, *Mycobacterium kansasii*, *Mycobacterium szulgai*, *Mycobacterium xenopi* and *Mycobacterium malmoense* (Glassroth, 2008; McGrath et al., 2010; Prevots & Marras, 2015). NTM lymphadenitis is observed almost exclusively in children and HIV-infected patients and is mostly associated with *M. avium* complex, *M. malmoense* and *Mycobacterium scrofulaceum* (Griffith et al., 2007). Finally, the disseminated disease commonly observed in severely immunocompromised patients frequently involves *M. avium* complex and *Mycobacterium genavense* (Tortoli, 2009). All the species mentioned before are SGM belonging to the same clade as *M. tuberculosis* and *M. leprae* (Gupta et al., 2018). However, some RGM species are also commonly associated with both pulmonary and cutaneous disease, notably *Mycobacterium abscessus* (*Mycobacteroides abscessus*) and *Mycobacterium fortuitum* (*Mycolicibacterium fortuitum*) which have been a frequent cause of nosocomial outbreaks (Tortoli, 2009).

Infections caused by *M. abscessus* are the most difficult to treat, with an average rate of treatment success below 50% (Johansen et al., 2020). This can be attributed to its unusually high intrinsic and acquired antibiotic resistance, conjugated with the frequent lack of correlation between *in vitro* susceptibility profiles and *in vivo* response to treatment, which has led it to be described as an antibiotic and clinical "nightmare" (Lopeman et al., 2019; Nessar et al., 2012). Due to the poor response to chemotherapy-based treatments in some patients, resection surgery is often recommended to improve clinical outcomes (Griffith et al., 2007). Phage therapy has recently emerged as a promising therapeutic tool, after some successful outcomes in untreatable cases of *M. abscessus* infection in patients who received this treatment under compassionate use (Dedrick et al., 2019, 2022). However, no clinical trials have been conducted yet on the use of bacteriophages against mycobacteria (Senhaji-Kacha et al., 2021). Furthermore, while there has been a recent effort to invest into strengthening the TB drug pipeline, the NTM drug pipeline is nearly empty and mostly comprised of repurposed drugs and anti-TB compounds (Johansen et al., 2020; Wu et al., 2018). However, it is common for NTM to not respond to classic TB drug regimens, and some of the new TB drugs show decreased activity against NTM (Wu et al., 2018). Bedaquiline, for example, is bactericidal against *M. tuberculosis* but only bacteriostatic against several NTM species, including *M. abscessus* and *M. avium* (Aguilar-Ayala et al., 2017; Dupont et al., 2017; Lounis et al., 2009). Thus, a

dedicated drug pipeline is essential to adequately manage the emerging threat of NTM infections. This is especially important since *M. abscessus* has a rapidly evolving genome with a low threshold for foreign DNA acquisition, suggesting it might easily acquire novel virulence factors and could potentially further evolve into a human-adapted pathogen (Choo et al., 2014). This evolution might already be in course, and it is worrying that studies suggesting human-to-human transmission of *M. abscessus* in cystic fibrosis centres are accumulating, challenging the long-held idea that NTM infections are exclusively acquired from the environment (Bryant et al., 2013, 2016; Ruis et al., 2021; Tortoli, Kohl, et al., 2017; Yan et al., 2020).

Interestingly, the clinical relevance of NTM goes beyond their role as opportunistic pathogens. Indeed, some species, notably *Mycobacterium vaccae* (*Mycolicibacterium vaccae*) and *Mycobacterium obuense* (*Mycolicibacterium obuense*), have found use as therapeutic agents with a diverse range of applications (Stanford & Stanford, 2012). Attenuated and heat-killed preparations of these NTM show immunomodulatory properties which have been investigated in cancer therapy (Cananzi et al., 2013; Dalgleish et al., 2016; Fowler et al., 2014), in the prevention and treatment of leprosy and TB (Huang & Hsieh, 2017; Stanford et al., 1990) and in the management of conditions such as psoriasis and asthma (Balagon et al., 2000; Dalbeth et al., 2004; Jiang et al., 2020). Furthermore, and as discussed in the next section, environmental mycobacteria can serve as valuable surrogates to study their pathogenic counterparts. Thus, the in-depth study of NTM and their diversity is essential to provide insights that can guide interventions into multiple areas of human health.

2.1.3. Thermophilic mycobacteria

As mentioned in the previous section, NTM are remarkably resilient bacteria able to tolerate, among other stress conditions, temperature extremes. A psychrotolerant species of mycobacteria capable of growth at 4°C has been described (Trujillo et al., 2004). Many species, including those frequently found in tap water and associated with community and healthcare-acquired infections, are heat-resistant (Schulze-Robbecke & Buchholtz, 1992). The thermotolerant *Mycobacterium avium* subsp. *paratuberculosis*, a common cattle pathogen that can be present in high numbers in raw milk and is suspected of being an etiological agent in human Crohn's disease, has been detected in pasteurised dairy products, suggesting a disturbing ability to resist pasteurisation (Eltholth et al., 2009;

Peterz et al., 2016). This is interesting given that mycobacteria do not form spores, so their ability to resist heat is an intrinsic property of mycobacterial cells accomplished through mechanisms differing from those used by sporulating bacteria. Optimal growth at around 45°C has been observed for some species, including *M. xenopi*, a common agent of NTM pulmonary disease (Griffith et al., 2007). A strain of *Mycobacterium parascrofulaceum* isolated from a hot spring was demonstrated to grow at 56°C (Santos et al., 2007). Finally, some species of mycobacteria are not only thermotolerant but also fit the definition of thermophilic since they grow optimally at or above 50°C (Vieille & Zeikus, 2001). This is the case for *Mycobacterium thermoresistibile* (*Mycolicibacterium thermoresistibile*) growing optimally at 52°C (Wolfe & Moore, 1992) and *Mycobacterium hassiacum* (*Mycolicibacterium hassiacum*) growing optimally at 50°C and up to 65°C (Schröder et al., 1997).

What is remarkable about these thermophilic species of mycobacteria is that they are capable of colonising and infecting humans, albeit rarely, making them rare but genuine thermophilic opportunistic pathogens. *M. hassiacum* has been associated with cases of peritonitis, cystitis and pulmonary disease (Jiang et al., 2013; Salzer et al., 2020; Tortoli et al., 1998), while *M. thermoresistibile* has been implicated in cases of osteomyelitis and skin and lung infections (Suy et al., 2013). True thermophiles acting as pathogens is an extremely rare occurrence since they are in theory not well adapted to human body temperature. Although clinically relevant bacterial isolates able to grow at 50°C were reported for the first time in 1985, they were not identified to the species level and research on this subject is scarce (Rabkin et al., 1985).

An interesting feature of thermophilic mycobacteria is that many of their essential pathways and enzymes are remarkably similar to those of *M. tuberculosis* and other pathogenic species. At the same time, their thermophilic lifestyle means that their proteins are essentially thermostable versions of mycobacterial proteins, which comes with several practical advantages, namely the well-known propensity of thermotolerant proteins to be more amenable to biochemical and structural studies, since they are in general more soluble and stable *in vitro* (Edwards et al., 2012). Because there is a high degree of active site similarity between mycobacterial enzymes sharing over 55% sequence identity, it is possible to use the orthologues of thermophilic species as surrogates for the *M. tuberculosis* enzymes which are more difficult to obtain and study (Baugh et al., 2015). Therefore, thermophilic mycobacteria should be seen as more than just an oddity within

the field of mycobacterial research, and also as an invaluable tool in the study of essential mycobacterial pathways and novel drug targets.

2.2. The family *Streptomycetaceae*

Streptomycetaceae are another clinically important family within *Actinomycetia*, although for different reasons than *Mycobacteriaceae*. Unlike the latter, it doesn't harbour important obligate pathogens, instead being known for the remarkable antibiotic-producing abilities of some of its members, commonly called streptomycetes. Indeed, its most famous member is arguably *Streptomyces griseus*, the source of many antibiotics including the first antibiotic against TB to be discovered, streptomycin (Woodruff, 2014).

Most species within this family belong to the genus *Streptomyces*, which includes nearly 700 validly published species to date (https://lpsn.dsmz.de/genus/streptomyces). It also comprises the genera *Embleya*, *Kitasatospora*, *Streptacidiphilus* and *Yinghuangia* (Salam et al., 2020). They are aerobic, non-motile, sporulating (unlike mycobacteria), gram-positive bacteria with a high G+C content and which form a highly branched substrate mycelium with aerial hyphae (Nouioui et al., 2018). Typically seen as soil-dwelling microorganisms, streptomycetes are ubiquitous across terrestrial and aquatic ecosystems, inhabiting environments as diverse as sea water, marine sediments, arid desert soils, artic glaciers and volcanic sediments (Sivalingam et al., 2019). They have been found in close association with plants, algae, corals, insects and bats, highlighting their ecological importance (Braña et al., 2015; Hamm et al., 2020; Kaltenpoth et al., 2005; Schrey & Tarkka, 2008). Although undoubtedly a mostly beneficial group of microorganisms, some species are plant pathogens and some can be infrequent causes of human infection (Dunne et al., 1998; Li et al., 2019).

The proficiency of streptomycetes as producers of bioactive compounds is arguably unmatched. They are not only the source of the majority of antibiotics, but also produce other clinically important compounds such as antifungal, anticancer, antiparasitic and immunosuppressive agents (Nett et al., 2009). Although by the 1990s frequent re-discovery of known antibiotics suggested this prolific source had been exhausted, the availability of whole genome sequences has since revealed a yet unexplored hidden diversity even in extensively screened species, as well as validating the notion that discovery of new species increases the pool of novel compounds available for screening ("new bugs = new drugs") (Ward & Allenby, 2018).

Additionally, some *Streptomyces* species are also important laboratory tools in the field of molecular biology, namely as heterologous hosts for recombinant protein production and for the cloning and overexpression of large biosynthetic gene clusters (BGCs) and consequent overproduction of their respective SM (Berini et al., 2020; Nah et al., 2017). This means that not only are streptomycetes a large reservoir of SM, but they also provide laboratory tools to obtain them in large enough quantities for screening and drug development studies.

2.2.1. The genus *Streptacidiphilus*

The genus *Streptomyces* has been extensively screened for novel bioactive compounds, but other members of the family remain largely underexplored. This is the case for the closely related genus *Streptacidiphilus*, a genus of acidophilic streptomycetes described for the first time in 2003 (Kim et al., 2003) but whose biosynthetic potential has only very recently begun to be explored (Elsayed et al., 2020; Hwang et al., 2019; Malik et al., 2020). Currently, the genus comprises 15 known species, most of which have been isolated from acidic forest soils, although there is one single report of isolation of a new species from a human clinical specimen (Nouioui et al., 2019).

A genome mining effort analysing 11 *Streptacidiphilus* genomes indicates that members of this genus possess unique BGCs that aren't present in other streptomycetes and which could encode novel SM with potentially interesting bioactivities, ranging from antibacterial, antifungal, antimalarial and antitumor effects (Malik et al., 2020). The first *Streptacidiphilus* SM to be discovered were the acidiphilamides, five new modified tripeptides with autophagy inhibition activity (Hwang et al., 2019). This is a potentially relevant finding since there is a current interest in developing autophagy inhibitors for their use in cancer treatment (Dolgin, 2019). The only other report to date of *Streptacidiphilus* SM is the discovery of an unusual family of spirotetronate polyketides, the streptaspironates (Elsayed et al., 2020). With acidiphilamides and streptaspironates, this relatively new and underexplored genus has already contributed to enrich the known chemical space of natural products, only in the last three years. However, it is predicted that many more novel compounds remain to be discovered by studying these acidophilic streptomycetes (Malik et al., 2020).

3. The glycobiology of mycobacteria and related actinomycetes

The glycobiology of *Actinomycetia* is a vast and diverse field of study with many ramifications. Indeed, sugar-containing molecules are involved in both the pathogenic and beneficial facets of these microorganisms, as well as in their ability to adapt to different ecological niches. Prominent examples are the glycoconjugates of the mycobacterial cell envelope which play a role in infection and in evading the host immune system, or the aminoglycoside antibiotics produced by *Streptomyces* species (ex., streptomycin) and which help fight infectious diseases such as TB. The importance of characterising the glycome, referring to the set of all glycans in an organism, is increasingly recognised particularly in the case of pathogens due to its ample involvement in bacteria-host interactions and the nearly infinite structural diversity it encompasses (Imperiali, 2019). Glycomic approaches have provided new insights into the mycobacterial glycome, for example with the recent discovery of a novel mycobacterial polysaccharide through the development of a new purification methodology (Larrouy-Maumus et al., 2015), or the systematic investigation of interactions between mycobacterial glycans and human cell receptors through the development of a synthetic glycan array (Zheng et al., 2017).

Behind the immense chemical diversity found in the carbohydrates and glycoconjugates of actinomycetes is an equally impressive variety of biocatalysts that assemble, modify, recycle and degrade these molecules. Carbohydrate-active enzymes (CAZYmes), including but not limited to glycoside hydrolases and glycosyltransferases, are extremely important not only because of their crucial physiological roles with ample implications in human health, but also for their countless biotechnological applications (Chao & Jongkees, 2019). Consequently, the in-depth study of these enzymes and their reaction mechanisms yields opportunities for drug discovery (Yakovlieva & Walvoort, 2020), expands our toolkit for chemical synthesis approaches that can answer the demand from pharmaceutical and chemical industries (André et al., 2014), and provides solutions for sustainable biomass conversion and biofuel production (Chettri et al., 2020), among others.

The following sections do not offer a comprehensive overview of this vast area of knowledge, but they highlight its diversity while focusing on the microorganisms, pathways, enzymes and sugars that were the subject of the work carried out in the scope of this thesis. Specifically, the focus will be placed on mycobacteria and the carbohydrates

that make them unique, as well as in the resulting opportunities for biomedical applications in the field of infectious diseases.

3.1. Mycobacterial cell envelope glycoconjugates

Actinobacteria, and thus actinomycetes, are often referred to as a gram-positive group of bacteria. While this is true for many members of this class, a remarkable exception evolved in the order *Corynebacteriales*, which includes mycobacteria. Instead of the typical monoderm cell envelope of gram-positive bacteria, these microorganisms have a diderm envelope that differs from that of gram-negative bacteria (Figure 1.2) (Rahlwes et al., 2019). The most distinguishing feature of this unique cell envelope is the presence of a layer of mycolic acids (MAs), among the longest fatty acids found in nature with up to 100 carbon atoms per molecule, and hence these bacteria are sometimes called the "mycolata" (Rahlwes et al., 2019). Notably, the thick, impermeable and complex nature of this envelope plays a role in the extraordinary resilience of mycobacteria to environmental stresses and their intrinsic resistance to many antibiotics and disinfectants (Bansal-Mutalik & Nikaido, 2014).

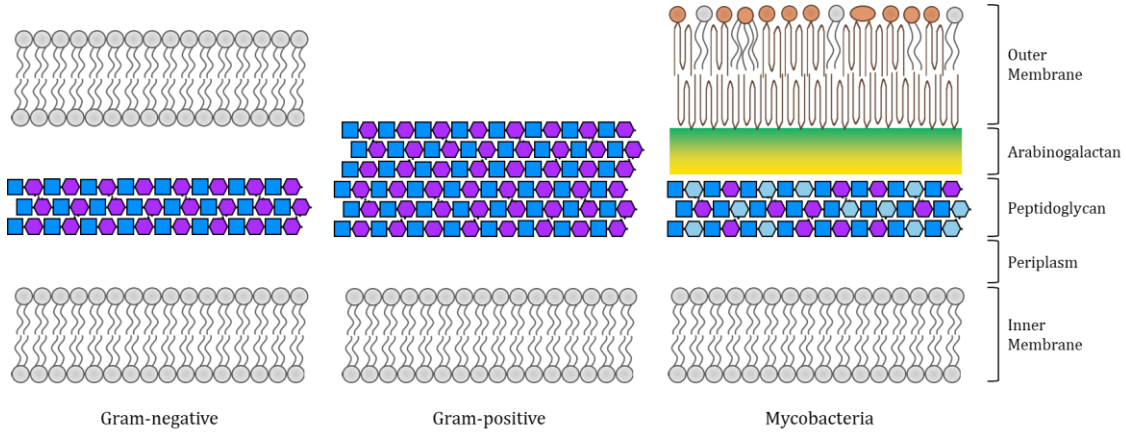


Figure 1.2. Schematic representation highlighting the differences in cell envelope architecture between gram-negative, gram-positive and "mycolata" bacteria. Both gram-negative bacteria and mycobacteria possess two lipid bilayers, distinguished in the latter by the presence of mycolic acids (represented by "w" shaped lines) and unique glycolipids (represented by brown polar heads). The thickness of each layer is variable between species and not drawn to scale.

The bacterial cell wall is a traditional target of many classes of antibiotics (Hutchings et al., 2019). In the case of mycobacteria, two of the four first-line drugs target the wall: isoniazid inhibits MA production and ethambutol blocks arabinogalactan (AG) synthesis (Dulberger et al., 2020). Furthermore, the outer layer of the cell envelope is the first point

of interaction between mycobacteria and the immune system of the host, and in pathogenic species its composition seems to be fine-tuned to ensure persistence and successful infection (Dulberger et al., 2020).

The mycobacterial cell envelope can be divided into four layers: i) inner membrane (IM), ii) cell wall core, iii) outer membrane (OM) and iv) capsule, all rich in sugar-containing molecules (Figure 1.3). The IM is composed of standard cytoplasmic membrane phospholipids, but it also contains phosphatidylinositol mannosides (PIMs) unique to mycobacteria and related actinomycetes (Rahlwes et al., 2019). These are glycosylated phosphatidylinositols with up to 4 fatty acids and up to 6 mannose units that seem to be confined to the IM, being its major component and likely contributing to the low permeability of the cell envelope (Bansal-Mutalik & Nikaido, 2014). They are also precursors of lipomannans (LMs), in which the mannan chain is extended reaching up to 34 mannose units, and lipoarabinomannans (LAMs), containing an additional polymer of ~70 arabinofuranose residues attached to the mannan chain of LMs (Rahlwes et al., 2019). The precise location of LMs and LAMs is still debated. Although they are synthesised from PIMs in the IM, share the same lipid anchors, and are present in actinomycetes that lack an OM, their presence in the OM of mycobacteria has been unequivocally demonstrated (Bansal-Mutalik & Nikaido, 2014; Minnikin et al., 2015). In any case, PIMs, LMs and LAMs are critical for cell wall integrity and strongly immunomodulatory, playing key roles in innate immunity evasion and granuloma formation (Mishra et al., 2011; Vergne et al., 2004).

The cell wall core consists of a covalently-bound complex of peptidoglycan (PG), AG and MAs, the mycoloylarabinogalactan-peptidoglycan (mAGP) complex. PG, composed of a linear glycan chain crosslinked by peptides, is conserved in all bacteria with a cell wall and is responsible for maintaining cell shape and protecting against osmotic pressure (Alderwick et al., 2015). The amino acids of the crosslinking peptide bridge vary across taxa, while the glycan chain is mostly conserved and consists of alternating β-1,4-linked *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) residues. However, a distinctive feature is found in the PG of mycobacteria and related actinomycetes, in which some muramic acid residues are glycolylated rather than acetylated, forming *N*-glycolylmuramic acid (MurGly) (Raymond et al., 2005). It is not known precisely why this modification evolved, but it has been hypothesised that *N*-glycolyl groups strengthen the PG's mesh-like structure through additional hydrogen bonding and that it may protect against lysozyme degradation (Alderwick et al., 2015). Some of the muramic acid residues

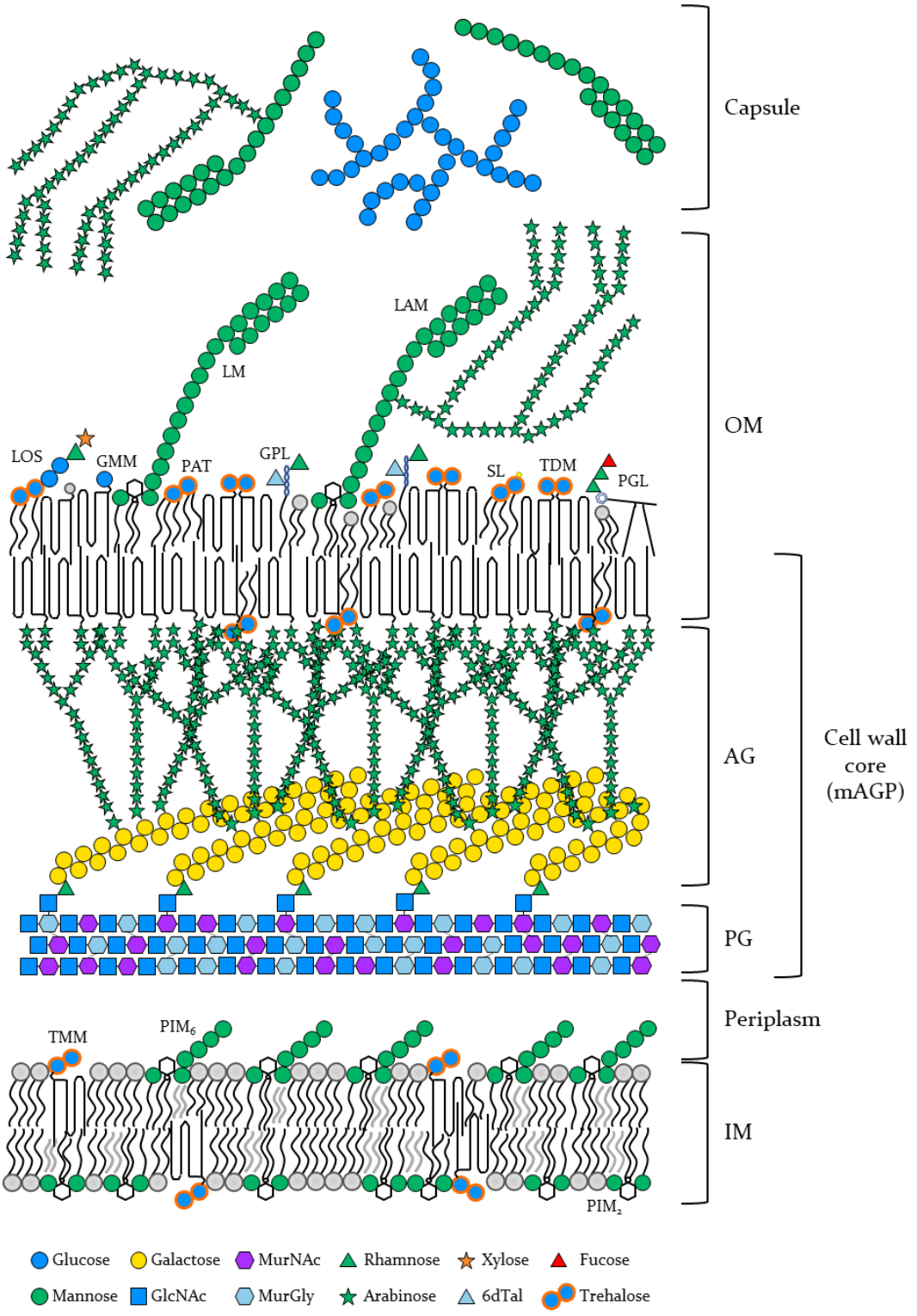


Figure 1.3. Schematic representation of the mycobacterial cell envelope highlighting its major glycan components. Not drawn to scale. The glycolipids represented in the inner membrane (IM) are: TMM, trehalose monomycolates; PIM_2 and PIM_6, phosphatidylinositol mannosides containing 2 and 6 mannose residues, respectively. The glycolipids represented in the outer membrane (OM) are: LOS, lipooligosaccharides; GMM, glucose monomycolates; LM, lipomannans; PAT, pentaacyl trehaloses; GPL, glycopeptidolipids; LAM, lipoarabinomannans; SL, sulfoglycolipids; TDM, trehalose dimycolates; and PGL, phenolic glycolipids. Arabinomannan, α-D-glucan and mannan are represented in the capsule layer. GlcNAc, *N*-acetylglucosamine; MurNAc, *N*-acetylmuramic acid; MurGly, *N*-glycolylmuramic acid; 6dTal, 6-deoxy-talose. Adapted from Bansal-Mutalik & Nikaido, 2014; Daffé et al., 2014; Minnikin et al., 2015; Rahlwes et al., 2019.

of the mycobacterial PG are covalently attached to a disaccharide linker unit formed by GlcNAc and rhamnose, the latter residue being bound to the galactan domain of AG (Alderwick et al., 2015). AG is composed of a galactan domain, a linear chain of ~30 D-galactofuranosyl (Gal*f*) residues with alternating β-1,5 and β-1,6 links, and an arabinan domain, consisting of 3 highly branched chains of ~23 D-arabinofuranosyl (Ara*f*) residues per each galactan chain. Together, the PG and AG layers form a dense yet flexible macro polymer whose three-dimensional structure is not fully understood, but which must provide a platform to anchor the MAs that form both the final layer of the cell wall core and the inner leaflet of the OM (Minnikin et al., 2015).

The mycobacterial OM, also called mycomembrane, is a lipid bilayer mainly consisting of high molecular weight MAs, which in mycobacteria contain 60-90 carbon atoms per molecule. Despite the length of their chains, MAs seem to adopt a folded conformation, explaining why electron microscopy studies indicate only a small difference in thickness between the inner and outer membranes (Groenewald et al., 2014; Hoffmann et al., 2008; Zuber et al., 2008). As mentioned above, some MAs are bound to AG, while others are free or esterified with polar molecules, including with trehalose with which they form trehalose monomycolates (TMM) and dimycolates (TDM). TMM is the precursor of TDM and also the source of the MA chains transferred to the AG. Inhibitors targeting the specific mycolyltransferases responsible for transferring MAs from TMM to both TDM and AG (Ag85 complex) and the membrane transporter that translocates TMM to the OM (MmpL3) are under investigation as promising new TB drug targets (Nobre et al., 2014). TDM (cord factor) is highly antigenic and plays a myriad of roles in the immunopathogenesis of TB, notably in intracellular survival (Welsh et al., 2013). Other trehalose-containing lipids found in the OM are diacyl, triacyl and pentaacyl trehaloses, sulfoglycolipids (consisting of sulfonated acyltrehalose), and lipooligosaccharides (LOS), all involved in virulence modulation (Nobre et al., 2014). LOS, absent from *M. tuberculosis* but present in many NTM species, consist of acylated trehalose linked to an oligosaccharide whose composition is species-dependent, and are involved in biofilm formation and immune modulation (Rahlwes et al., 2019). In addition to glucose, 3-*O*-methyl-rhamnose, xylose and fucose, some unusual and unique sugars can be found in the oligosaccharides modifying the acylated trehalose core of LOS, such as *N*-acyl kansosamine in *M. kansasii* or caryophyllose and a 4,6-dideoxy-galactose substituted with a 3-hydroxy-3-methylated-pyrrolidone group in *M. marinum* (Hunter et al., 1984; Sarkar et al., 2011).

Glycopeptidolipids (GLPs) are another type of glycoconjugate found in the cell envelope of some NTM, but not in *M. tuberculosis*, and which are important virulence factors in pathogenic species such as *M. avium* and *M. abscessus* (Mukherjee & Chatterji, 2012). The core structure of GPLs consists of a fatty-acyl tetrapeptide glycosylated with 6-deoxy-talose and methyl rhamnose at two different residues of the peptide. 6-deoxy-talose can be further glycosylated and the monosaccharides can be methylated and acetylated giving rise to great structural diversity and serovar-specific GPLs. These can be more or less polar and thus modulate the hydrophobicity of the cell envelope, with consequences in biofilm formation, sliding motility and host-pathogen interactions (Gutiérrez et al., 2018; Miyamoto et al., 2010; Mukherjee & Chatterji, 2012). The mycobacterial OM also harbours phenolic glycolipids (PGLs), present in *M. tuberculosis*, *M. leprae* and a few slowly growing pathogenic NTM, which play important roles in host-pathogen interactions (Arbues et al., 2014). They are derived from the highly apolar lipids phthiocerol dimycocerosates (PDIMs), but in which this lipid moiety is linked to a phenolic residue, further modified with a saccharide of variable length and composition, with methylated forms of rhamnose, fucose and glucose being encountered (Pérez et al., 2004; Tabouret et al., 2010). Interestingly, *M. tuberculosis* secretes glycosylated *p*-hydroxybenzoic acid methyl esters containing the same carbohydrate moiety found in its PGLs (Pérez et al., 2004).

Finally, at the outermost layer of the mycobacterial cell envelope there is a loose capsule-like structure formed mainly by polysaccharides, but also by lipids and proteins with ratios, composition and thickness varying across species. In general, lipids are the minor component, although glycolipids (mostly species-specific GPLs and PGLs) are found in abundant quantities in the capsules of some species such as *M. leprae* and *M. avium* (Daffé et al., 2014). In other species, such as *M. smegmatis*, proteins are the major constituent and the capsule is thin, while in *M. tuberculosi*s this structure is much thicker and mainly composed of polysaccharides (Daffé et al., 2014; Rahlwes et al., 2019). Three types of capsular polysaccharides are found in the mycobacterial capsule: α-glucan, arabinomannan, and mannan. The latter two are structurally identical to the carbohydrate moieties of LAM and LM and are presumed to derive from these glycolipids. α-glucan is the most abundant of the three and is structurally related to cytoplasmic glycogen but smaller in size, consisting of an α-1,4-glucose polymer with extensive α-1,6 branching (Rahlwes et al., 2019). As with most of the other unique elements of the mycobacterial cell

envelope, the capsule components have been shown to play important roles in pathogenesis and immunomodulation (Daffé et al., 2014).

3.2. Free solutes: trehalose and glucosylglycerate

Microbial resistance to environmental stress can be accomplished through several mechanisms. In mycobacteria, as mentioned, an impervious cell envelope is a key source of resilience, but other strategies are also present, including the accumulation of small solutes in response to specific environmental challenges. Accumulation of compatible solutes is a widespread microbial strategy to cope with abiotic stress agents like osmotic stress, heat, cold, oxidative stress, dehydration, desiccation and nutrient starvation (Elbein et al., 2003; Empadinhas & da Costa, 2011; Nunes-Costa et al., 2017). Two small carbohydrates are especially important in mycobacterial physiology and adaptation to stress: the disaccharide trehalose (α-D-glucopyranosyl-1,1-α-D-glucopyranoside) and the glycoside glucosylglycerate (α-D-glucopyranosyl-1,2-D-glycerate) (Figure 1.4). While both are precursors of vital macromolecules (see section 3.1 for trehalose and section 3.3 for glucosylglycerate), they are also accumulated in the cytoplasm in their free form under certain conditions.

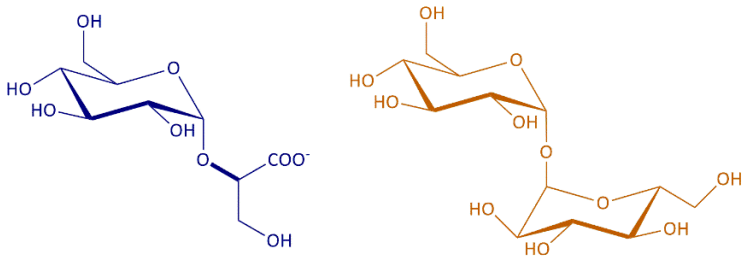


Figure 1.4. Chemical structures of glucosylglycerate (left, blue) and trehalose (right, dark orange).

Free trehalose is ubiquitous in nature and an important molecule in plants, insects, fungi and bacteria, serving as an energy and carbon reserve, as a sensing compound and growth regulator, and as a stabiliser and protectant of proteins and membranes against temperature extremes, dehydration and oxidative damage (Elbein et al., 2003). In actinomycetes, its function is not strictly structural as part of cell envelope glycolipids. Indeed, several *Streptomyces* strains accumulate trehalose in their spores (for example, trehalose levels in spores of *S. griseus* can reach up to 25% of their dry weight) which is thought to stabilise macromolecules as cells lose their water content, and to provide energy during the first stages of spore germination (Bobek et al., 2017). Although

mycobacteria are non-sporulating, they can also resist desiccation and stressed mycobacterial cells are able to enter a dormant stage with reduced metabolic activity. It has recently been shown that *M. smegmatis* cells accumulate trehalose as they enter dormancy and its levels are associated with maintaining viability during this critical stage, and also that the breakdown of accumulated trehalose is involved in the early events of resuscitation (Shleeva et al., 2017). Thus, the function of free trehalose in mycobacteria seems to mimic what is observed for trehalose-accumulating sporulating microorganisms such as yeasts and streptomycetes (Bobek et al., 2017; Shleeva et al., 2017).

Out of the five natural pathways that can contribute to trehalose biosynthesis, mycobacteria possess genes for at least three of them (Nobre et al., 2014). The OtsA/OtsB pathway produces trehalose through the sequential action of a trehalose-6-phospate synthase that condenses NDP-glucose and glucose-6-phosphate (OtsA) and a phosphatase that dephosphorylates trehalose-6-phosphate (OtsB). Alternatively, the TreY/TreZ pathway can synthetise trehalose from glycogen through the action of a maltooligosyltrehalose synthase (TreY) that isomerises a terminal maltose unit of a glycogen chain to trehalose, which is then cleaved by a maltooligosyltrehalose trehalohydrolase (TreZ) to release free trehalose. Finally, the enzyme trehalose synthase (TreS) is able to isomerise between maltose and trehalose (Figure 1.7) (Nobre et al., 2014). The three pathways were initially thought to be redundant, and indeed early evidence suggested that an *M. smegmatis* triple mutant with all pathways disabled could only grow when supplemented with trehalose, while each of the three pathways was able to recover the mutant (Woodruff et al., 2004). However, in *M. tuberculosis* the OtsA/OtsB pathway is essential for growth while the TreY/TreZ pathway is not, and TreS seems to be important during the late stages of pathogenesis in mice but not essential (Minato et al., 2019; Murphy et al., 2005). Furthermore, subsequent genetic studies complemented NMR analyses in *M. smegmatis* refuted its involvement in trehalose synthesis, suggesting that the metabolic flux through this enzyme *in vivo* is towards the production of maltose from trehalose (Miah et al., 2013). These results suggest that all three pathways play different roles *in vivo* and underline the vital importance of trehalose in mycobacterial physiology.

Glucosylglycerate (GG) was originally identified over 50 years ago as a structural component of a unique mycobacterial polysaccharide, and later as a minor free solute in a few unrelated prokaryotes. Once thought to be a rare microbial metabolite, it is now known that accumulation of GG is a widespread microbial strategy to cope with salt-stress in nitrogen poor environments. Like trehalose, GG plays a role in actinomycetes as a

precursor of macromolecules, namely as the polar head of a glycolipid found in *Nocardia otitidiscaviarum* whose function, taxonomic distribution and biosynthetic pathway remain completely unknown, and attached to the reducing end of 6-*O*-methylglucose lipopolysaccharides (MGLP) (discussed in detail in section 3.3) (Nunes-Costa et al., 2017). Also similarly to what has been documented for trehalose, free GG has been shown to stabilise and protect proteins and nucleic acids, which could be of benefit to organisms under environmental stress conditions (Faria et al., 2008; Sato et al., 2014; Sawangwan et al., 2010). Its crucial role in mycobacteria as a free solute has only recently been discovered. Unlike what is commonly observed in other prokaryotes that accumulate GG as a response to osmotic stress when nitrogen levels are low, mycobacteria accumulate this glycoside in the absence of salt stress and in response to nitrogen starvation alone (Alarico et al., 2014; Behrends et al., 2012). In *M. smegmatis*, the key gene for GG biosynthesis was found to be up-regulated about 20-fold upon nitrogen restriction, leading to GG accumulation, and the corresponding mutant had defective growth under nitrogen-limited conditions (Behrends et al., 2012). Likewise, *M. hassiacum* was found to accumulate GG under nitrogen deprivation and to rapidly decrease its levels when nitrogen became available by up-regulating a GG-specific hydrolase (Alarico et al., 2014). These results suggest that tight control of free GG levels in rapidly growing mycobacteria is necessary for optimal cell fitness, with immediate mobilisation of accumulated GG presumably required to resume biosynthetic processes upon nitrogen replenishment (Nunes-Costa et al., 2017). It has not been investigated if free GG is also accumulated by slowly growing species such as *M. tuberculosis*, but the genomic absence of the specific GG-hydrolase identified in *M. hassiacum* indicates that the adaptation strategies of RGM and SGM to nitrogen starvation might be different (Nunes-Costa et al., 2017)

3.2.1. GG biosynthesis in mycobacteria

Three biosynthetic pathways yielding GG are presumed to exist in nature, but only one is known to be present in mycobacteria (Nunes-Costa et al., 2017). This is a two-step pathway involving a specific glycosyltransferase, glucosyl-3-phosphoglycerate synthase (GpgS), that condenses NDP-glucose and 3-phosphoglycerate (3-PGA) to form the phosphorylated intermediate glucosyl-3-phosphoglycerate (GPG) which is then dephosphorylated by a specific phosphatase, glucosyl-3-phosphoglycerate phosphatase (GpgP), producing GG (Figure 1.5). This route was first identified in the psychrotolerant

archaeon *Methanococoides burtonii* (Costa et al., 2006), but although by then it had been known for decades that mycobacteria produce GG, homologues to the archaeal enzymes could not be found in mycobacterial genomes (Nunes-Costa et al., 2017). Later, the identification of an ambiguous glycosyltransferase able to produce both GG and the related glycoside mannosylglycerate (MG) in the actinobacterium *Rubrobacter xylanophilus* led to the discovery of the divergent mycobacterial GpgS and other "actinobacterial-type" GpgS homologues (Empadinhas et al., 2008). In *M. tuberculosis*, GpgS is considered essential for growth (Minato et al., 2019) which quickly prompted the determination of its three-dimensional structure so as to provide an experimental scaffold for inhibitor design (Pereira et al., 2008). Since then, investigation of the structure and catalytic mechanism and dynamics of the mycobacterial GpgS has attracted significant interest, although high-resolution structures are still missing, hampering rational drug design efforts (Albesa-Jové et al., 2015; Albesa-Jové, Romero-García, et al., 2017; Urresti et al., 2012).

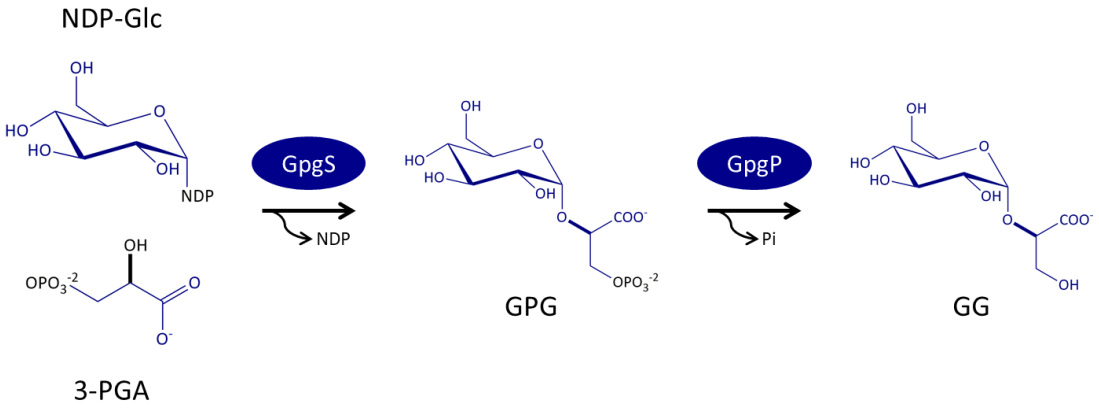


Figure 1.5. Pathway for the biosynthesis of glucosylglycerate (GG) in mycobacteria. NDP-Glc, nucleoside diphosphate glucose; 3-PGA, 3-phosphoglycerate; GPG, glucosyl-3-phosphoglycerate; GpgS, glucosyl-3-phosphoglycerate synthase; GpgP, glucosyl-3-phosphoglycerate phosphatase.

Like with GpgS, the discovery of the mycobacterial GpgP was also not straightforward due to lack of sequence homology with typical GpgPs found in other GG-accumulating bacteria and archaea. Indeed, while typical GpgPs belong to the dehalogenase-like hydrolase superfamily, the mycobacterial enzyme is part of the histidine phosphatase superfamily instead and was misannotated as a phosphoglycerate mutase (Mendes et al., 2011; Zheng et al., 2014). The phylogenetic distribution of this second family of GpgPs is rather limited, being restricted almost exclusively to the *Corynebacterineae* (Nunes-Costa et al., 2017). Unlike GpgS, this gene is not considered essential for *M. tuberculosis* growth, suggesting that other phosphatase(s) can

dephosphorylate GPG in the cell (Minato et al., 2019). Although in prokaryotes genes encoding enzymes involved in the same metabolic pathways are often grouped into operons to facilitate co-transcription, and this is often the case with *gpgS* and *gpgP* genes in some organisms, it is not always the case (Empadinhas & da Costa, 2011; Ma & Xu, 2013). In mycobacteria, the two genes are not associated in the genome and it has been proposed that this genomic arrangement may confer metabolic flexibility, with each enzyme participating in the same metabolic pathway but also possibly being independently regulated and additionally involved in different pathways (Nunes-Costa et al., 2017).

In RGM, free GG levels are controlled not only by the two enzymes participating in the two-step pathway outlined above (Figure 1.5), but also by a specific GG hydrolase, GgH, which mobilises accumulated GG when nitrogen levels are re-established (see section 3.2) and whose three-dimensional structure has been recently unveiled (Alarico et al., 2014; Cereija et al., 2019). In SGM, the absence of a GgH homologue could indicate that they only use GG as a MGLP precursor or that an alternative non-homologous hydrolase is responsible for this activity, but evidence either way is still lacking.

3.3. Intracellular polysaccharides

Three polysaccharides can be found in the cytoplasm of mycobacteria and related actinomycetes: glycogen, MGLP and 3-*O*-methylmannose polysaccharides (MMP). The last two macromolecules are collectively known as polymethylated polysaccharides (PMPS) due to being extensively methylated. Unlike the universal carbon and energy storage molecule glycogen, PMPS have a very limited phylogenetic distribution and their physiological function is still not fully understood (Mendes et al., 2012).

Hints on the function of PMPS were provided by early studies showing their ability to bind long-chain fatty acids and form stable 1:1 complexes, as well as by their activating effect on fatty acid synthase I (FAS-I) (Machida & Bloch, 1973). These observations led to the proposal that PMPS had a regulatory role in fatty acid synthesis by simultaneously sheltering newly synthetised acyl chains from lipolytic degradation, preventing their accumulation in the cytoplasm from disrupting metabolism, and protecting FAS systems from end product inhibition by accelerating the rate of diffusion of newly synthetised acyl chains and promoting enzyme turnover (Bloch & Vance, 1977). Consequently, and since these fatty acids are also the precursors of MAs, an essential role in cell envelope assembly was anticipated for PMPS, but it has yet to be experimentally demonstrated. The first

attempt at studying their function *in vivo* was carried out in a *M. smegmatis* spontaneous mutant with reduced PMPS levels (50% of MGLP and 7% of MMP when compared to wildtype). No severe anomalies were observed, except for a slower growth rate and a shift in fatty acid composition towards higher levels of short-chain and lower levels of long-chain fatty acids (Maloney & Ballou, 1980). The elucidation of some of the genes involved in MGLP biosynthesis allowed a more targeted approach, with the inactivation of specific genes in both *M. tuberculosis* and *M. smegmatis*, although total MGLP abolition was not achieved and fatty acid metabolism was not affected (Kaur et al., 2009; Stadthagen et al., 2007). Nonetheless, these mutants' impaired growth at supraoptimal temperatures gave further clues about the physiological role of MGLP, implying it in mycobacterial adaptation to thermal stress.

The least investigated of these polysaccharides is by far MMP, possibly due to its absence from most pathogenic mycobacteria, including from species of the *M. tuberculosis* complex. Indeed, MMP has a scattered distribution across mycobacteria being typically associated with RGM (Mendes et al., 2012). It consists of a linear chain of 11–14 units of α-1,4–linked 3-*O*-methyl-D-mannose, with an unmethylated mannose at the non-reducing end and the reducing end blocked by a methyl group (Figure 1.6). The methyltransferase responsible for blocking the 1-OH position of MMP's reducing end was recently identified in a gene cluster likely encoding all the activities required for MMP synthesis (Ripoll-Rozada et al., 2019). MMP is unique to mycobacteria, although an acetylated version has been found in *Streptomyces griseus* (Kari & Gray, 1979). In the genome of this species, as well as in some *Nocardia* genomes, the recently identified genes for MMP biosynthesis are present in a differently organised cluster (Ripoll-Rozada et al., 2019).

Unlike MMP, MGLP (or highly conserved genes for its biosynthesis, see section 3.3.1) has been found in all mycobacteria investigated thus far as well as in some *Nocardia* strains (Mendes et al., 2012). Remarkably, even *M. leprae* produces MGLP despite its extremely reduced genome and massive gene decay, suggesting that this polysaccharide plays an indispensable role in mycobacterial physiology (Cole et al., 2001; Hunter et al., 1986). MGLP is composed of 15–20 glucose and methylglucose residues with a variable degree of acylation. Its backbone consists of α-1,4–linked glucoses and 6-*O*-methylglucoses, with a 3-*O*-methylglucose at the non-reducing end and GG at the reducing end. GG is α-1,6-linked to the backbone, and two branching glucoses are β-1,3-linked to glucoses 1 and 3. Acetate, propionate or isobutyrate are esterified to the methylglucose units close to the nonreducing end, succinate is esterified to both branching glucoses and an octanoate

moiety is esterified to glucose 1 of the backbone (Figure 1.6) (Maranha et al., 2015). The α-1,4–linkages of the main chain result in stereochemical constraints that are thought to induce an helical conformation in solution, presumably with the methyl groups facing the inner cylindrical hydrophobic cavity to accommodate acyl chains (Jackson & Brennan, 2009). It has also been suggested that the octanoyl group may have a specific effect in stabilising the polysaccharide's helical conformation, as well as in anchoring MGLP intermediates to the cytoplasmic membrane (Hindsgaul & Ballou, 1984; Smith & Ballou, 1973). However, experimental evidence to support all these hypotheses is still lacking.

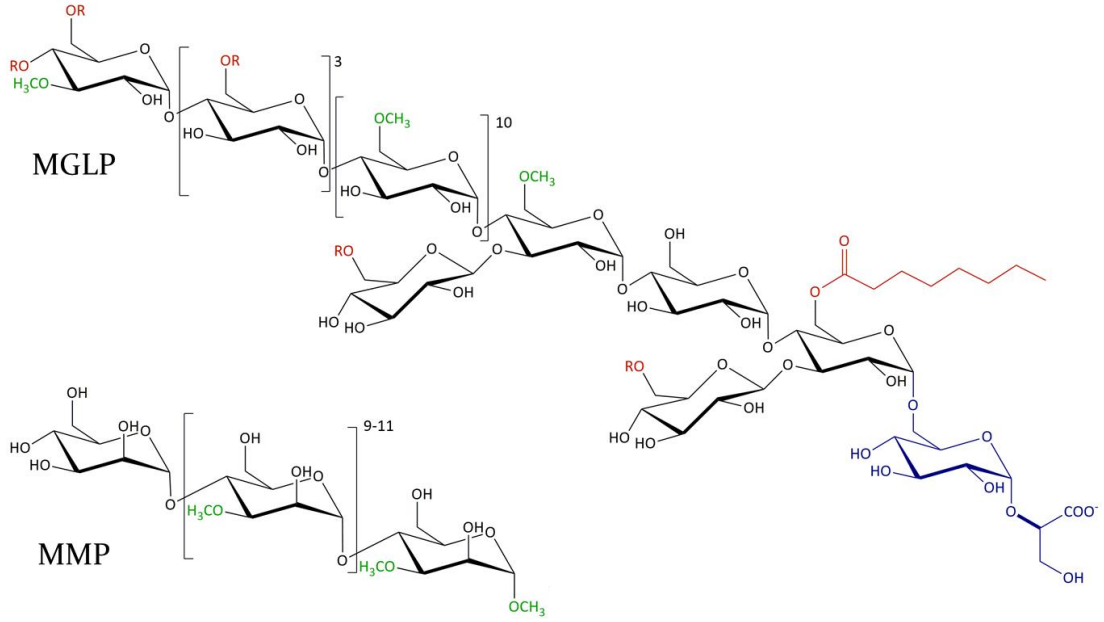


Figure 1.6. Scheme of the general chemical structures of MGLP from *M. bovis* BCG and MMP from *M. smegmatis*. The glucosylglycerate moiety of MGLP is coloured blue, methyl groups are coloured green and acyl groups (R) are coloured red. Adapted from Maranha et al., 2015; Mendes et al., 2012.

3.3.1. Biosynthesis of MGLP and other α-1,4-glucans

There are three α-(1,4)-glucans found in mycobacteria: glycogen, capsular α-glucan, and MGLP. As discussed in sections 3.1 and 3.3, all three have important roles in mycobacteria and share a basic backbone. Additionally, and as expected from their similar structures, they also probably share some biosynthetic steps.

The canonical pathway for glycogen synthesis in bacteria proceeds through the synthesis of ADP-glucose by the ADP-glucose pyrophosphorylase GlgC, the elongation of α-1,4-linked glucose chains by the glycogen synthase GlgA, which uses ADP-glucose as a substrate, and the generation of α-1,6-linked branches by the branching enzyme GlgB, which transfers oligoglucans from the non-reducing end of a linear α-1,4-chain to the C-6

hydroxyl group of a glucose residue within the chain. This classical pathway was thought to be present in mycobacteria, together with a second α-1,4-glucosyltransferase (Rv3032) partially redundant with GlgA and an alternative pathway, using maltose-1-phosphate (M1P) as the donor of glucose units (Bornemann, 2016; Sambou et al., 2008). The gene encoding Rv3032 was one of the first genes in *M. tuberculosis* H37Rv to be implicated in MGLP biosynthesis, but in one study disruption of this gene was found to decrease not only MGLP but also glycogen production, while capsular glucan levels were unaffected (Stadthagen et al., 2007). On the other hand, disruption of the *M. tuberculosis* GlgA homologue led to decreased levels of capsular glucan and wildtype levels of both glycogen and MGLP (Sambou et al., 2008). Complementation of the *glgA* mutant with a wildtype copy of *rv3032* restored glucan production, and simultaneous deletion of both *glgA* and *rv3032* was unsuccessful, leading the authors to suggest that the two enzymes could partially replace each other (Sambou et al., 2008).

An alternative pathway was subsequently discovered to contribute to α-1,4-glucan biosynthesis in mycobacteria and other bacteria (Elbein et al., 2010; Kalscheuer et al., 2010). The first step of this pathway is catalysed by TreS (initially thought to participate in trehalose biosynthesis, see section 3.2), which converts trehalose into maltose. Maltose is subsequently phosphorylated by maltokinase (Mak), producing M1P (Mendes et al., 2010) which is then used by the maltosyltransferase GlgE to elongate α-1,4-glucan chains two glucoses at a time (Figure 1.7). The branching enzyme GlgB is common to both the classical and the GlgE pathway (Bornemann, 2016). Interestingly, recent evidence showed that there is an alternative route to M1P production and that the canonical pathway may after all not be present in mycobacteria, since the mycobacterial GlgA homologue was shown to have a different activity than what was initially assumed. Thus, this enzyme was renamed as GlgM and identified as a maltose-1-phosphate synthase which uses ADP-glucose as a donor and glucose-1-phosphate (G1P), rather than a glycogen chain, as its preferential acceptor (Figure 1.7) (Koliwer-Brandl et al., 2016; Syson et al., 2020). Unlike previously reported, these authors found that Rv3032 does not contribute substantially to glycogen production, does not rescue glucan deficient mutants, and that disrupting Rv3032 and the GlgA homologue simultaneously is not lethal to *M. tuberculosis* (Koliwer-Brandl et al., 2016). According to this revision of the metabolic network behind α-1,4-glucan production in mycobacteria (Figure 1.7), Rv3032 would be involved in MGLP biosynthesis, and the GlgE pathway in capsular glucan and glycogen biosynthesis. However, the contribution of the GlgE pathway to MGLP production has not been directly

investigated. Given that *rv3032* deletion in *M. tuberculosis* does not lead to complete abolition of MGLP production, it is possible that the GlgE pathway can also contribute to elongating the MGLP backbone (Stadthagen et al., 2007). Unlike Rv3032, GlgE is essential for *M. tuberculosis* and has been validated as a drug target, since its disruption leads to accumulation of M1P, a toxic metabolite (Kalscheuer et al., 2010; Veleti et al., 2014). It should also be noted that while Rv3032 has been undeniably implicated in MGLP biosynthesis and its sequence indicates it is a glucosyltransferase, its identification as an α-1,4-glucosyltransferase that elongates linear α-1,4-glucan chains was based on the observation that its inactivation affected glycogen levels, an observation that has since been disputed (Koliwer-Brandl et al., 2016; Stadthagen et al., 2007). Since this enzyme has not been biochemically characterised, its true activity and preferred substrates remain unknown.

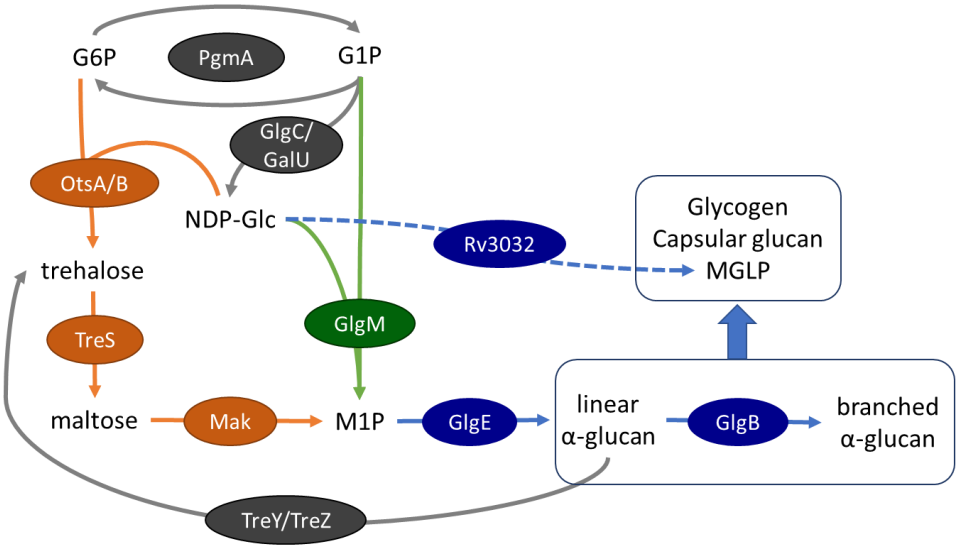


Figure 1.7. Pathways for the biosynthesis of α-1,4-glucans in mycobacteria and related actinomycetes. Two alternative pathways yield maltose-1-phosphate (M1P), one through TreS/Mak (depicted in orange) and the other through GlgM (depicted in green). Depicted in blue are the enzymes involved in α-1,4-glucan elongation and branching. The dashed line indicates a hypothetical enzyme activity. G6P, glucose-6-phosphate; G1P, glucose-1-phosphate; NDP-Glc, ADP-glucose or UDP-glucose; PgmA, phosphoglucomutase; GlgC, ADP-glucose pyrophosphorylase; GalU, UDP-glucose pyrophosphorylase; OtsA/B, trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase; TreS, trehalose synthase; Mak, maltokinase; GlgM, maltose-1-phosphate synthase; GlgE, maltosyltransferase; GlgB, α-1,4-glucan branching enzyme. Adapted from Koliwer-Brandl et al., 2016.

The chemical structure of MGLP is much more complex than that of glycogen and capsular glucan, therefore the shared biosynthetic steps are restricted to the elongation of its backbone. Some of the other enzymes involved in the MGLP biosynthetic pathway have been identified, but there are still several missing and hypothetical-only steps. The primer for MGLP biosynthesis is GG (Kamisango et al., 1987), and thus the enzymes GpgS and

GpgP (see section 3.2.1) participate in the first two steps of the currently accepted pathway (Figure 1.8). The third step seems to be the transfer of one glucose residue to GG to form glucosyl-α-1,6-glucosylglycerate (DGG), isolated from *M. smegmatis* cell extracts in 1987 and identified as a MGLP precursor (Kamisango et al., 1987). Based on genomic context and sequence analysis, a gene coding for a putative glycoside hydrolase of the GH-57 family of branching enzymes (Rv3031) found in the vicinity of Rv3032 was proposed to be involved in this step, but experimental evidence is still lacking (Stadthagen et al., 2007). Recently, a gene encoding an octanoyltransferase (OctT) able to transfer an octanoyl group to the C6 OH of DGG's second glucose was characterised and proposed to mediate the fourth step of the pathway (Maranha et al., 2015). This discovery led to a revision of the accepted structure of MGLP, which previously displayed the octanoyl group esterified to glycerate (Jackson & Brennan, 2009). Surprisingly, OctT also octanoylated GG albeit to a lesser extent, and also in position 6 of glucose, which if occurring *in vivo* would block the transfer of the second glucose to GG and halt the MGLP pathway. Although a regulatory mechanism was proposed, the physiological importance of this activity has yet to be elucidated (Maranha et al., 2015).

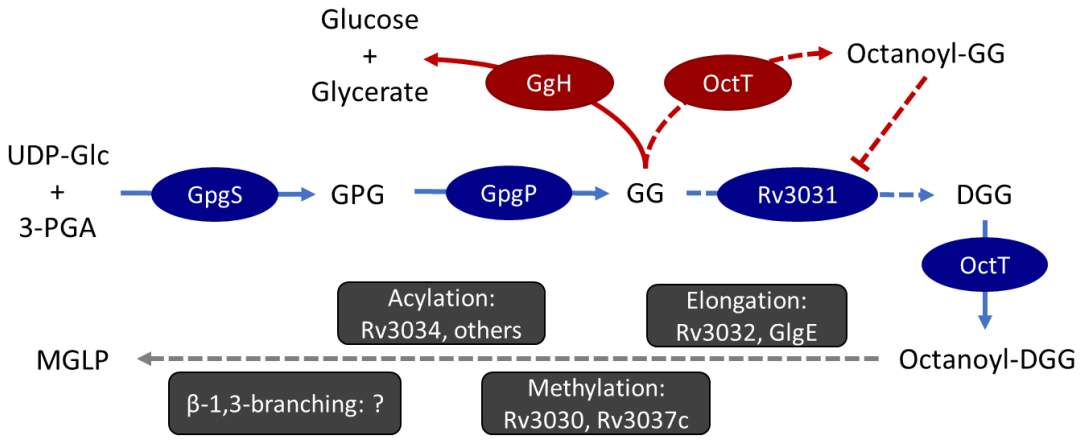


Figure 1.8. Pathway for MGLP biosynthesis. The initial steps are depicted in blue, late steps in grey and hypothetical regulatory steps in red. Dashed lines represent hypothetical or experimentally unconfirmed functions. UDP-Glc, UDP-glucose; 3-PGA, 3-phosphoglycerate; GPG, glucosyl-3-phosphoglycerate; GG, glucosylglycerate; DGG, diglucosylglycerate; GpgS, glucosyl-3-phosphoglycerate synthase; GpgP, glucosyl-3-phosphoglycerate phosphatase; GgH, glucosylglycerate hydrolase; OctT, octanoyltransferase; GlgE, maltosyltransferase.

How the pathway proceeds after the formation of octanoyl-DGG is unclear. Early studies with *M. smegmatis* and *M. phlei* extracts pointed towards a model of alternate glycosylation, acylation and methylation which is still accepted today (Grellert & Ballou, 1972; Kamisango et al., 1987). The putative methyltransferase Rv3030, identified in the

operon of Rv3032, was confirmed to be involved in MGLP production since its inactivation in *M. smegmatis* led to decreased MGLP levels (Stadthagen et al., 2007). Also in the same genomic region, a gene encoding an additional putative methyltransferase (Rv3037c) and a putative acetyltransferase (Rv3034) were proposed to also participate in MGLP biosynthesis, but experimental evidence is lacking (Stadthagen et al., 2007). The timing and enzymes involved in the transfer of the two β-1,3-branching glucoses and their succinylation are unknown.

Both MGLP and GG have been implicated in mycobacterial adaptation, albeit to different abiotic stresses: GG in the response to nitrogen starvation and MGLP in thermal adaptation (Nunes-Costa et al., 2017). Thus, mycobacteria, especially those known to accumulate GG, must dynamically tune GG and MGLP levels. Since the biosynthesis of both compounds is interconnected, regulatory mechanisms at the early stages of MGLP synthesis would theoretically be necessary to, for example, allow mycobacteria to rapidly and substantially increase GG levels without undesirable downstream effects on MGLP biosynthetic enzymes. How this is accomplished is also not known, although the activity of OctT on GG, blocking the C6 OH where the next glucose should be added, and the hydrolysis of GG by the specific hydrolase GgH (see section 3.2) have both been proposed as possible regulatory nodes (Figure 1.8) (Alarico et al., 2014; Maranha et al., 2015).

3.4. Amino sugar metabolism

Amino sugars play many important roles in the biology of actinomycetes. They have a structural role as the building blocks of peptidoglycan, thus preserving cell integrity (see section 3.1). Furthermore, they are present in mycothiol which replaces glutathione as the major protectant against oxidative stress in *Actinobacteria* (Newton et al., 2008), as well as being part of a wide range of antibiotics produced by actinomycetes and other bacteria (Kudo & Eguchi, 2009; Nedal & Zotchev, 2004). Additionally, amino sugars can represent an important source of carbon, nitrogen and energy in actinomycetes inhabiting chitin-rich environments and function as signalling molecules and regulators of both primary and secondary metabolism (Świątek-Połatyńska et al., 2015; Świątek et al., 2012).

3.4.1. Amino sugars in primary metabolism and in the biosynthesis of cell envelope components

The pathway for amino sugar biosynthesis in bacteria starts with the conversion of the glycolysis intermediate fructose-6-phosphate (F6P) to glucosamine-6-phosphate (GlcN6P) by the enzyme glucosamine-6-phosphate synthase (GlmS). GlmS is composed of two domains, a sugar isomerase domain (SIS domain) and a glutaminase domain. The latter cleaves the amino group from glutamine, which is transferred to F6P producing glutamate and GlcN6P. This is an important metabolic node intersecting carbon and nitrogen metabolism and the first committed step in amino sugar biosynthesis, essential for PG production (Moraes et al., 2015). GlcN6P is converted by phosphoglucosamine mutase (GlmM) to glucosamine-1-phosphate (GlcN1P), the substrate of the bifunctional enzyme GlmU whose acetyltransferase domain catalyses the formation of *N*-acetylglucosamine-1-phosphate (GlcNAc1P), and whose uridylyltransferase domain subsequently forms UDP-GlcNAc, a cell wall building block (Maitra et al., 2019). GlmU has been structurally characterised in *M. tuberculosis* (Zhang et al., 2009) and has garnered attention as a promising drug target due to: i) the absence of homologues in eukaryotes, which employ a slightly different pathway by acetylating GlcN6P and not GlcN1P; ii) the demonstration of its essential role *in vitro* as well as in *ex vivo* and *in vivo* models, where GlmU depletion was found to lead to pathogen clearance in infected murine lungs and bacillary load reduction in infected mice; and iii) the observation that both domains of GlmU are independently essential for *M. tuberculosis* growth, thus showing that inhibiting either domain would be effective and disrupting both domains would block two critical reactions at once (Soni et al., 2015). Interestingly, a recent report identified a possible alternative, GlmU-independent fate for GlcN1P in the actinomycete *Rhodococcus jostii*, with the discovery that a duplicated copy of the gene encoding UDP-glucose pyrophosphorylase uses GlcN1P more efficiently than G1P, forming UDP-GlcN (Cereijo et al., 2021). The physiological relevance of this finding is so far unknown.

As mentioned in section 3.1, the carbohydrate portion of PG consists of alternating GlcNAc and MurNAc, with MurGly replacing some MurNAc residues. The pathway described above accounts for the formation of UDP-GlcNAc, which is directly incorporated into PG by the *N*-acetylglucosamine transferase MurG (Maitra et al., 2019). Concomitantly, UDP-GlcNAc is also converted to UDP-MurNAc by the sequential action of MurA, which transfers the enolpyruvyl moiety of phosphoenolpyruvate (PEP) to UDP-GlcNAc, and MurB, a NADPH dependent oxidoreductase that reduces the enolpyruvate

moiety to D-lactate (Maitra et al., 2019). MurA is the target of the broad-spectrum antibiotic fosfomycin, produced by some *Streptomyces* strains, however it has been found to be ineffective against *M. tuberculosis* due to a specific mutation in the active site of this enzyme, showing that the innate antibiotic resistance of mycobacteria is not exclusively traceable to the impermeably of the cell envelope (De Smet et al., 1999). A series of transferases, MurC-F, assemble the peptide bridge of PG directly onto UDP-MurNAc and a translocase finally transfers this assembly to the membrane where the remaining modifications happen, including the transfer of the GlcNAc residue (Maitra et al., 2019). The origin of the unique MurGly residues in the PG of all mycobacteria except *M. leprae* has been traced to the NADPH-dependent hydroxylation of UDP-MurNAc to UDP-MurGly by the hydroxylase NamH (Raymond et al., 2005).

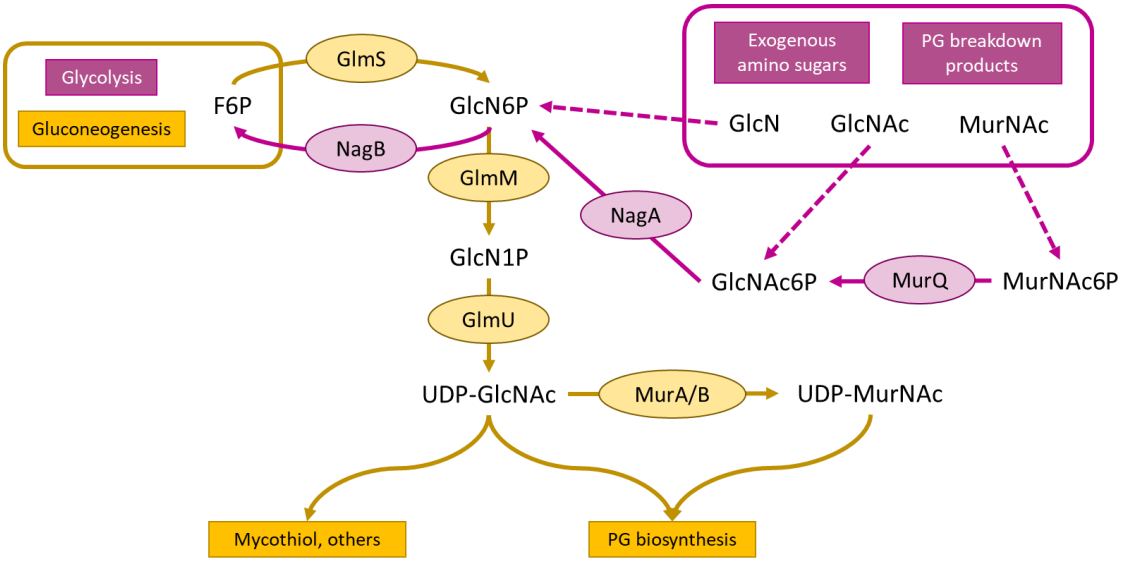


Figure 1.9. Pathways for amino sugar biosynthesis and degradation in the context of central metabolism and cell wall assembly. Anabolic and catabolic steps and pathways are coloured yellow and pink, respectively. Dashed lines represent phosphorylation by the phosphotransferase system or intracellular kinases. F6P, fructose-6-phosphate; GlcN6P, glucosamine-6-phosphate; GlcN1P, glucosamine-1-phosphate; UDP-GlcNAc, UDP-*N*-acetylglucosamine; UDP-MurNAc, UDP-*N*-acetyl-muramic acid; GlcN, glucosamine; GlcNAc, *N*-acetylglucosamine; MurNAc, *N*-acetyl-muramic acid; GlcNAc6P, *N*-acetylglucosamine-6-phosphate; MurNAc6P, *N*-acetyl-muramic acid-6-phosphate; PG, peptidoglycan.

Another amino sugar found in the cell wall of some mycobacteria is galactosamine (GalN). Non-acetylated hexosamine residues are rather uncommon in prokaryotes, but the existence of GalN in the cell envelope of SGM (and apparent absence in RGM) such as *M. tuberculosis* is well documented (Draper et al., 1997). This GalN residue is α-linked to the branched arabinan domain of AG with a stoichiometry of one GalN residue per AG molecule (Peng et al., 2012). The complete pathway for its biosynthesis has not been established, but is likely to have as a precursor UDP-GlcNAc which can be converted to

UDP-GalNAc by a UDP-GalNAc 4-epimerase, which interconverts the two substrates. An enzyme with such activity has recently been characterised in *M. tuberculosis* (Pardeshi et al., 2017). Interestingly, this enzyme is located in close proximity to a polyprenyl-phospho-*N*-acetylgalactosaminyl synthase (PpgS) that was implicated in the origin of the GalN residue through genetic studies and shown to utilise UDP-GalNAc (Škovierová et al., 2010). The same study identified a membrane-associated glycosyltransferase likely responsible for the transfer to AG, given that its disruption led to the abolition of the production of galactosaminylated AG. A deacetylase is also predicted to be involved in this pathway, although it has not been identified.

Amino sugars resulting from degradation and recycling of cellular components and from uptake from the external environment can be used as energy, carbon and nitrogen sources through catabolic pathways. PG turnover is an obvious and major source of amino sugars and most bacteria are able to fully recycle the GlcNAc and MurNAc residues released during this process (Borisova et al., 2016). GlcNAc, and GlcN to a lesser extent, can also be taken up from the environment by actinomycetes following the degradation of chitin, the abundant polysaccharide found in the cell walls of fungi and in the exoskeletons of arthropods, such as crustaceans and insects, and which represents a major nutrient for soil-dwelling streptomycetes (Świątek et al., 2012). Catabolism of these monosaccharides begins with their phosphorylation, either during transport through specific phosphotransferase systems (PTS) or by intracellular kinases. Key enzymes directing these compounds back to central metabolism are MurQ, which cleaves the lactate group of MurNAc; NagA, a deacetylase that converts GlcNAc6P to GlcN6P; and NagB, a deaminase converting GlcN6P to F6P which can then enter the glycolytic pathway. As expected, environmental saprophytes such as *Streptomyces* spp. and *M. smegmatis* have a more complete set of metabolic genes compared to obligate pathogens such as *M. tuberculosis* (Moynihan et al., 2019; Świątek et al., 2012; Titgemeyer et al., 2007). Indeed, *M. tuberculosis* lacks both MurQ and NagB homologues, unlike *M. smegmatis*, and is apparently unable to catabolise GlcNAc, although it possesses a hydrolase that cleaves the GlcNAc-MurNAc bond, and can utilise the lactate moiety of MurNAc for growth through the action of an as-yet unidentified enzyme (Moynihan et al., 2019). NagA, conserved in mycobacteria, has also been functionally and structurally characterised in *M. tuberculosis* (Ahangar et al., 2018).

3.4.2. Mycothiol

Mycothiol (MSH) is the major low molecular weight thiol found in actinomycetes, consisting of D-glucosamine (GlcN) α-1,1-linked to *myo*-inositol (Ins) and amide linked to the carboxyl group of an acetylated cysteine reside (Figure 1.10). It is unique to *Actinobacteria*, and its primary role is to maintain intracellular redox homeostasis, paralleling the function of glutathione (Jothivasan & Hamilton, 2008). It serves as a cofactor in detoxification reactions and may also be involved in the biodegradation of recalcitrant xenobiotics (Rawat & Av-Gay, 2007).

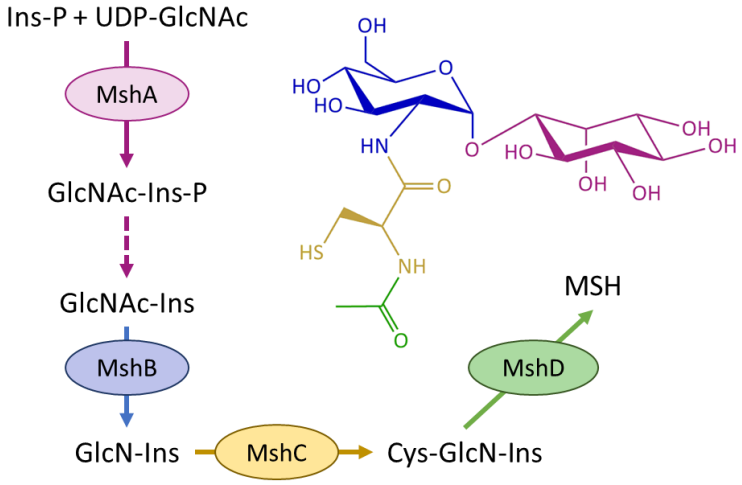


Figure 1.10. Structure and biosynthetic pathway of mycothiol (MSH). Structurally, mycothiol is composed of *myo*-inositol (purple), glucosamine (blue), cysteine (yellow) and an acetyl group (green). The dashed line represents a step catalysed by a yet unidentified enzyme. Adapted from Jothivasan & Hamilton, 2008.

The MHS biosynthetic pathway was first investigated in mycobacteria, which gave the thiol its designation (Newton et al., 2008). The first step of this pathways is the conversion of glucose-6-phosphate (G6P) into *myo*-inositol-1-phosphate (Ins-P) by the inositol phosphate synthase. Subsequently, the glycosyltransferase MshA transfers the amino sugar moiety from UDP-GlcNAc to Ins-P forming the intermediate GlcNAc-Ins-P, which is then dephosphorylated by a yet unidentified phosphatase and deacetylated by the metalloprotein MshB to yield GlcN-Ins. Finally, the Cys-tRNA synthetase homolog MshC catalyses the ATP-dependent ligation of cysteine and the acetyltransferase MshD acetylates the amino group of cysteine using acetyl-CoA, yielding MSH (Figure 1.10). In general, the genes involved in this biosynthetic pathway are not arranged into operons, being scattered throughout the genome in the species investigated (Newton et al., 2008). Since *M. tuberculosis* likely depends on the protective role of MSH to survive the oxidative

environment encountered in macrophages, several steps of this pathway have attracted interest as potential drug targets (Newton et al., 2008).

Interestingly, *M. smegmatis* mutants unable to produce MSH have been reported to actively transport MSH from the external environment to generate cellular levels compatible with those of the wild-type strain. Once transferred to a MSH-free medium, the degradation of MSH could be followed in these mutants revealing MSH's catabolic fate, with all of its components eventually entering the glycolytic and Krebs cycle pathways (Bzymek et al., 2007). The only step of MSH degradation that could be attributed to a known enzyme was the cleavage of the amide bong linking the cysteine and GlcN residues by the amidase Mca, yielding acetyl-Cys and GlcN-Ins (Bzymek et al., 2007). The deacetylation of cysteine, cleavage of GlcN-Ins and likely phosphorylation of GlcN and oxidation of Ins are predicted but yet unassigned activities.

3.4.3. Secondary metabolites and antibiotic production

Secondary metabolism is one of the most remarkable aspects of the biology of actinomycetes and the incredible structural diversity associated with it is in no small part due to amino sugars (Kudo & Eguchi, 2009; Nedal & Zotchev, 2004). Amino sugars can be found modifying several classes of antibiotics and other bioactive compounds, the carbohydrate moiety often being essential for the bioactivity of the molecule (Křen & Řezanka, 2008; Weymouth-Wilson, 1997). This is the case for the important antibiotic streptomycin, the first antituberculous agent to be discovered, which includes a *N*-methyl-L-glucosamine moiety (Figure 1.11) that has been observed by X-ray structural analysis to establish key interactions with streptomycin's target, the 30S ribosomal subunit (Carter et al., 2000).

Although the enzymes involved in the synthesis of the *N*-methyl-L-glucosamine moiety have not been characterised, a tentative pathway has been proposed based on sequence analysis of the genes present in the streptomycin BGC in the producers *Streptomyces griseus* and *S. glaucescens*. The proposed pathway begins with the activation of G1P by a cytidylyltransferase, yielding CDP-glucose, and subsequent modification of the nucleotide sugar, including amino and methyl transfer reactions, before transfer to the scaffold by a glycosyltransferase (Flatt & Mahmud, 2007). Streptomycin is an aminoglycoside antibiotic, a class known for containing unusual and unique amino sugars. Other aminoglycosides have better characterised pathways, as is the case for neomycin,

produced by *S. fradiae* and containing the unique amino sugar neosamine (2,6-diamino-2,6-dideoxy-glucose) (Figure 1.11). In this biosynthetic pathway, a GlcNAc moiety is transferred from UDP-GlcNAc to the aminocyclitol scaffold (2-deoxystreptamine) by a specific glycosyltransferase, and is subsequently deacetylated, oxidised by a FAD-dependent dehydrogenase, and aminated by an aminotransferase, all biochemically characterised enzymes (Kudo & Eguchi, 2009). Neomycin possesses a second neosamine residue which also originates from UDP-GlcNAc through the action of a second glycosyltransferase and is modified after transfer by the same deacetylase, dehydrogenase and aminotransferase acting on the first residue (Kudo & Eguchi, 2009).

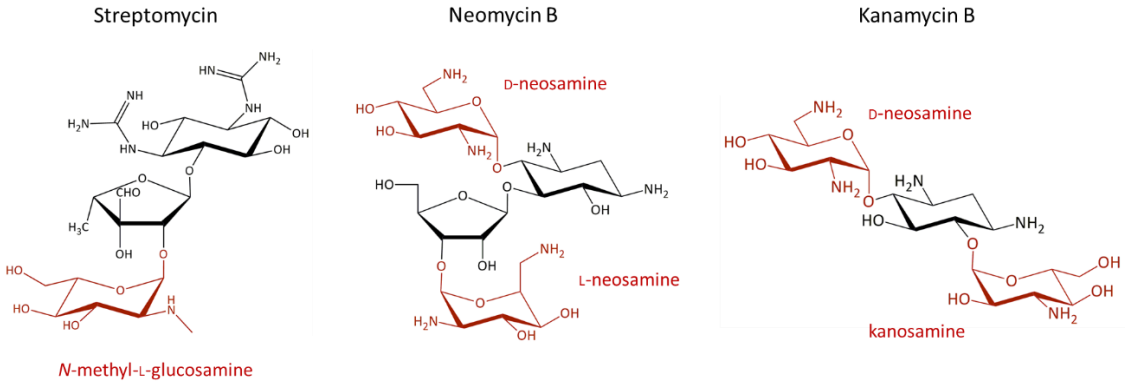


Figure 1.11. Structures of the aminoglycoside antibiotics streptomycin, neomycin B and kanamycin B with the amino sugar moieties highlighted in red.

Neosamine is also found in the related aminoglycoside butirosin. Interestingly, a nucleotidylyltransferase capable of utilising GlcN1P to produce NDP-GlcN was identified in the butirosin BGC, suggesting a pathway using GlcN, rather than GlcNAc, as a precursor and dispensing with the deacetylation step described for neomycin (Kudo et al., 2005). Although butirosin is produced by the firmicute *Bacillus circulans*, homologues of this enzyme were also detected in the BGC of the aminoglycoside kanamycin in *Streptomyces kanamyceticus* and in the BGCs of the glycopeptides teicoplanin, balhimycin and chloroeremomycin produced by other actinomycetes (Kudo et al., 2005). Kanamycin represents an interesting example of how chemical diversity can be generated within a biosynthetic pathway. Containing two sugar moieties modifying the aminocyclitol ring, kanamycin is, like many other antibiotics, a mixture of compounds. The substrate promiscuity of the two glycosyltransferases involved in the pathway is in part responsible for generating the various forms of kanamycin, since the first glycosyltransferase can utilise both UDP-GlcNAc and UDP-Glc, while the second glycosyltransferase can use UDP-kanosamine (3-amino-3-deoxy-D-glucose, Figure 1.11) or UDP-Glc (Park et al., 2011).

The ability of each enzyme to use UDP-GlcN has not been reported, although the presence of the GlcN1P nucleotidylyltransferase in its operon suggests it may also be a possible substrate (Kudo et al., 2005).

In addition to aminoglycosides, unusual amino sugars are also common structural components of many glycopeptide and macrolide antibiotics, as well as of the important anticancer drugs anthracyclines (Chen et al., 2000; Metsä-Ketelä et al., 2008; Nedal & Zotchev, 2004). The biosynthetic pathways for these carbohydrate substituents are compound-specific and too numerous to go into detail, however they generally follow the basic scheme outlined in the examples given above for some aminoglycosides. First, an activated sugar is generated from a phosphorylated hexose/hexosamine by a nucleotidylyltransferase. This intermediate NDP-sugar may (or may not) be modified by dehydrogenases, aminotransferases, methyltransferases, (de)acyltransferases, epimerases, isomerases among other tailoring enzymes before transfer by a pathway-specific glycosyltransferase. Once transferred to the SM scaffold, the sugar moiety can be (further) modified by the tailoring enzyme activities mentioned above, giving rise to great structural diversity. As a rule, biosynthetic genes involved in secondary metabolism tend to be in close proximity in the genome of the producing strain (associated into BGCs), and have a restricted and often scattered phylogenetic distribution.

In addition to being the building blocks of many SM, amino sugars also act as key signalling molecules controlling antibiotic production and other important cellular processes. In streptomycetes, antibiotic production is known to correlate temporarily with important developmental events that coordinate the transition between vegetative growth and sporogenesis, according to nutrient availability (Bibb, 2005; van Bergeijk et al., 2020). Sporogenesis is a stress response and a key event in the lifecycle of sporulating bacteria that requires precise timing and perception of environmental cues (Dworkin & Losick, 2005; González-Pastor et al., 2003). Committing to sporogenesis too late may mean that not enough nutrients are available to complete the developmental programme, but suspending vegetative growth too early means a lost opportunity for maximising biomass accumulation and thus increasing the number of spores released. GlcNAc is one of the main signalling molecules involved in this tightly controlled process in streptomycetes and has been observed to exert a dual effect on *Streptomyces* cultures: under nutrient-rich conditions, the presence of GlcNAc blocks antibiotic production and sporogenesis initiation, whereas under nutrient-scarcity GlcNAc elicits the onset of both (Rigali et al., 2008). GlcNAc uptake and metabolism is under the control of the transcriptional

repressor DasR, also involved in the regulation of antibiotic production, development onset and primary metabolism pathways (Świątek-Połatyńska et al., 2015). Notably, GlcN6P, resulting from GlcNAc metabolism and representing a metabolic node between primary and secondary metabolic pathways, acts as an effector molecule and inhibits DasR from binding to DNA, relieving the repression of BGCs involved in secondary metabolism (Rigali et al., 2006).

In streptomycetes, uptake of exogenous GlcN monomers and oligosaccharides resulting from the extracellular degradation of chitin or chitosan has been suggested to contribute to the intracellular GlcN6P pool allowing its assimilation and incorporation into SM biosynthesis, but while the specific transporters have been identified, experimental evidence of actinobacterial GlcN kinases is still lacking (Viens et al., 2015).

4. Objectives

Actinomycetes are a diverse group of bacteria which includes pathogens and antibiotic-providers, a source of problems and a source of solutions. Although often studied in separate, these two facets are such a fundamental part of the biology of these bacteria that research into one area may easily lead to the other when their metabolism is closely investigated. In the present work, we sought to investigate some unanswered questions relating to previously characterised mycobacterial enzymes considered essential for the growth of *M. tuberculosis* and for the environmental adaptation of opportunistic NTM. The two enzymes selected were Mak, involved in the biosynthesis of the backbone of mycobacterial α-1,4-glucans, and GpgS, responsible for the first committed step in GG and MGLP biosynthesis. Both have been proposed as potential new drug targets, since blocking their activity could have a deleterious effect on mycobacteria. Mak has been biochemically and structurally characterised, but our genomic surveys revealed that some mycobacteria carry a second copy of this gene, whose activity we sought to investigate through biochemical and structural studies. Strikingly, our preliminary results indicated that the two Mak paralogues did not have redundant activities and that the second copy might be involved in secondary and amino sugar metabolism in only a few species of actinomycetes. We deemed important to pursue this avenue of research given the multifold importance of amino sugars, especially in the context of secondary metabolism. In addition, and although GpgS has also been biochemically and structurally characterised and considered an attractive target for drug design efforts, high-resolution structures to guide rational drug design and provide deeper structural insights are still missing, reason why we explored the potential of thermophilic mycobacteria to fill this gap. Both avenues of research were successfully pursued through an approach intersecting the fields of microbiology, bioinformatics, enzymology and structural biology, and both gave rise to new and unexpected questions laying the foundations for future research avenues.

Chapter 2 – A novel enzyme family with glucosamine kinase activity in *Actinomycetia*: characterisation and structural insights

Part of this chapter is published in:

Manso JA*, Nunes-Costa D*, Macedo-Ribeiro S, Empadinhas N, Pereira PJB. 2019. Molecular fingerprints for a novel enzyme family in Actinobacteria with glucosamine kinase activity. *mBio* 10:e00239-19. https://doi .org/10.1128/mBio.00239-19.

*These authors contributed equally to this work.

Chapter 2 – Table of Contents

Abstract

Secondary metabolites (SM) produced by members of the class *Actinomycetia* have long been the main source of antibiotics and other small molecules with therapeutic and industrial applications. Secondary metabolism in these microorganisms is tightly controlled by environmental and physiological factors and deeply intertwined with central metabolic pathways. Phosphorylation of exogenous glucosamine has been suggested as a mechanism for incorporation of this extracellular material into amino sugar-containing SM, as well as a source of the important signalling molecule glucosamine-6-phosphate, involved in SM biosynthesis regulation. However, experimental evidence of specific glucosamine kinases in *Actinomycetia* is lacking. Here, we present the discovery of a unique family of actinobacterial glucosamine kinases related to maltokinases and describe the structural and biochemical characterisation of one of its members, from the soil actinomycete *Streptacidiphilus jiangxiensis*, including structural evidence of phosphoryl transfer to glucosamine. Conservation of glucosamine-contacting residues across a large number of uncharacterised actinobacterial proteins unveiled a specific glucosamine binding sequence motif and global genetic context analysis revealed a link to biosynthetic gene clusters (BGCs) containing secondary metabolism-associated genes. The present work reveals a previously "hidden" family of sugar kinases and heralds the discovery of novel pathways and SM within the medically important class *Actinomycetia*.

Introduction

Actinomycetes, and particularly members of the genus *Streptomyces*, are still the largest natural source of antibiotics and other bioactive compounds used in medicine and industry today (Genilloud, 2017; Hutchings et al., 2019). In recent decades, and following a golden era of antibiotic discovery, interest and investment into natural product discovery fell abruptly as antimicrobial resistance escalated worldwide, limiting present and future treatment options for numerous infectious diseases (Brown & Wright, 2016; Hutchings et al., 2019). However, genomic-era advances revealed that many novel compounds remain to be unearthed, even among extensively screened antibiotic-producing strains, leading to genome mining efforts aimed at cryptic BGCs whose chemical products are unknown but predicted to be significantly novel (Belknap et al., 2020; Challis, 2008; Du & van Wezel, 2018; Ward & Allenby, 2018). The chemical diversity associated with secondary metabolism is partly driven by the ability of enzymes to evolve and diversify their

functions, allowing the emergence of novel pathways and, consequently, novel bioactive compounds. One important source of metabolic innovation is gene duplication followed by neofunctionalisation, leading to the evolution of new enzyme families with divergent functions from a common ancestor (Díaz-Mejía et al., 2007; Scossa & Fernie, 2020). In addition to metabolic innovation within biosynthetic pathways, the complexity of secondary metabolism also requires the evolution of intricate signalling and regulatory mechanisms that can govern this important aspect of the biology of microorganisms (Liu et al., 2013; van der Heul et al., 2018). Indeed, SM play diverse roles in the organisms that produce them, being mainly associated with environmental interactions and adaptation responses such as nutrient acquisition (siderophores), protection from abiotic stresses (pigments, compatible solutes), interference competition (biocides) or signalling and communication (quorum sensing and interspecies cross-talk) (Cornforth & Foster, 2013; Hibbing et al., 2010; van Bergeijk et al., 2020). Consequently, SM production must be precisely tuned according to environmental stimuli, but also according to the metabolic status of the organism.

In streptomycetes, two of the most important molecules known to be associated with SM biosynthesis are *N*-acetylglucosamine (GlcNAc) and glucosamine-6-phosphate (GlcN6P). In addition to serving as building blocks for many amino sugar-containing SM (Kudo et al., 2005; Kudo & Eguchi, 2009; Park et al., 2011), they regulate SM biosynthesis through a signalling pathway controlled by the transcriptional repressor DasR (Rigali et al., 2006, 2008; Świątek-Połatyńska et al., 2015; Świątek et al., 2012). GlcNAc, one of the major components of the bacterial cell wall and a primary source of carbon and nitrogen for actinomycetes inhabiting chitin-rich environments, has been found to have a dual regulatory effect on streptomycetes, blocking or eliciting SM biosynthesis depending on the nutritional status of the culture (Rigali et al., 2008). GlcNAc uptake and metabolism is under direct control of DasR, which is in turn a target for the effector molecule GlcN6P, able to bind this transcriptional repressor and prevent its binding to DNA, relieving the repression of BGCs involved in SM production (Rigali et al., 2006; Świątek-Połatyńska et al., 2015). The crystal structures of DasR and of its ortholog NagR from *Bacillus subtilis* in complex with GlcN6P have been reported (Fillenberg et al., 2015, 2016). However, to date only two metabolic enzymes that lead to production of intracellular GlcN6P have been described in actinobacteria: NagA (EC 3.5.1.25), which deacetylates GlcNAc6P formed by phosphorylation of acquired extracellular GlcNAc by a phosphotransferase system (Ahangar et al., 2018), and GlmS (EC 2.6.1.16), which produces GlcN6P by transamination

of the glycolytic intermediate fructose-6P (Milewski, 2002). Homologues of these two enzyme classes are present in the genomes of several *Streptomycetaceae*. On the other hand, it was recently demonstrated that the *csnR-K* operon, conserved in *Actinomycetia*, controls the uptake of chitosan-derived D-glucosamine (GlcN) oligosaccharides and, at a lower scale, of monomeric GlcN (Viens et al., 2015). Phosphorylation of exogenous GlcN could thus constitute an alternative source of intracellular GlcN6P. Although GlcN kinases (GlcNK) (EC 2.7.1.8) have been identified in the gammaproteobacterium *Vibrio cholerae* (Park et al., 2002) and in the hyperthermophilic archaeon *Thermococcus kodakarensis* (Aslam et al., 2018), and their existence in actinobacteria has been anticipated (Viens et al., 2015), experimental evidence for this activity in this medically important phylum is still missing.

Here, we identify and characterise a unique GlcNK from the soil bacterium *Streptacidiphilus jiangxiensis*, member of the family *Streptomycetaceae* and belonging to a genus whose secondary metabolism is an underexplored but promising source of novel bioactive compounds (Huang et al., 2004; Malik et al., 2020). This GlcNK is not a sequence homologue of the isofunctional enzymes from *Vibrio* or from archaea (Aslam et al., 2018; Park et al., 2002). Instead, its closest homologues are maltokinases (Mak), ubiquitous enzymes in the bacterial domain which phosphorylate maltose and have been implicated, particularly in mycobacteria, in glycogen metabolism and in the biosynthesis of essential α-1,4-glucans (Fraga et al., 2015). Like Mak, the GlcNK reported in this work is a member of the eukaryotic-like kinase (ELK) superfamily, which is distantly related to eukaryotic protein kinases (ePKs) and also includes other prokaryotic kinases acting on small molecules, such as aminoglycoside phosphotransferases (APH) and methylthioribose (MTR) kinase (Kannan et al., 2007). The crystal structure of the quaternary complex between this kinase and GlcN, ADP, inorganic phosphate, and Mg^{2+} provides unparalleled structural evidence of a transition state of the phosphoryl-transfer mechanism in this family. Additionally, the molecular determinants for GlcN phosphorylation, with high evolutionary conservation in several representative families of the class *Actinomycetia*, have been determined, unveiling a sequence motif for the classification of a large number of uncharacterised or misannotated phosphotransferases into this unique family of actinobacterial GlcNKs.

Methods

Identification, sequence analyses and genomic context

glcNK genes were initially detected as *mak* paralogues in the genomes of some mycobacteria and other actinomycetes and hypothesised to encode functionally redundant copies of the canonical Mak. Sequences of *mak* and *mak* paralogues were identified by BLAST searches using the sequences of characterised mycobacterial maltokinases (EC 2.7.1.175) as queries in the KEGG SSDB (https://www.kegg.jp/kegg/ssdb) and NCBI databases. These databases were also used for inspection of the genomic context of each *mak* homologue in genomes which returned two BLAST hits. Sequence Similarity Networks (SSN) were generated with the Enzyme Function Initiative's Enzyme Similarity Tool (EFI-EST), which allows the visual segregation of homologous sequences into clusters based on sequence similarity and reveals potential functional divergence within an enzyme family (Zallot et al., 2019). To build the SSN, the amino acid sequence of the *S. jiangxiensis* GlcNK (*Sj*GlcNK) was used to retrieve homologous sequences from the UniProt database using a negative log of e-value threshold of 5. A maximum sequence length restriction of 600 residues was applied to discard multidomain fusion proteins (in some organisms, *mak* and *treS* are fused in a single open reading frame) and an alignment score threshold of 35 was selected. Generated SSNs were visualised and edited with Cytoscape 3.8.2. Sequences belonging to unclassified taxa were manually removed.

For genomic context analyses, the Enzyme Function Initiative's Genome Neighbourhood Tool (EFI-GNT) was used to generate a Genome Neighbourhood Network (GNN) based on the edited SSN. This tool uses each sequence considered in an SSN to collect information about its neighbouring genes from the European Nucleotide Archive database, assigns each identified gene to a Pfam family, and calculates co-occurrence frequencies (Zallot et al., 2019). A neighbourhood size of 20 (number of neighbouring genes upstream and downstream of the query sequence) and a minimal co-occurrence percentage lower limit of 30% were set. The resulting GNN and the associated Genome Neighbourhood Diagrams (GND) were examined and some species were selected for further analysis with the *anti*biotics and *s*econdary *m*etabolite *a*nalysis *sh*ell, antiSMASH version 5.0 (Blin et al., 2019), using default parameters and enabling the ClusterBlast and MIBiG (Kautsar et al., 2020) comparison algorithm.

Multiple sequence alignments for selected sequences were generated with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and edited with Aline version

1.0.025. Molecular weight predictions were estimated with the ProtParam tool (http://web.expasy.org/protparam/). The *Ms*GlcNK model was built with Swiss-Model (https://swissmodel.expasy.org/) using the experimentally solved *Sj*GlcNK structure as a template (PDB entry: 6HWL).

Strains and culture conditions

Streptacidiphilus jiangxiensis 33214 (DSM 45096), obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (Germany), was cultivated at 28°C for 6 days in ISP2 agar medium at pH 5.5 (4 g/L yeast extract, 10 g/L malt extract, 4 g/L glucose, 20 g/L agar). *Mycobacterium smegmatis* mc^2155 (ATCC 700084), obtained from LGC Standards S.L.U. (Spain), was cultivated for 5 days at 30°C in a glycerol-based agar medium at pH 7.0 (20 g/L glycerol, 5 g/L casamino acids, 1 g/L fumaric acid, 1 g/L K_2HPO_4, 0.3 g/L $MgSO_4$, 0.02 g/L $FeSO_4$, 2 g/L Tween 80) (Brennan & Ballou, 1967).

Cloning and site-directed mutagenesis

Chromosomal DNA from *S. jiangxiensis* and from *M. smegmatis* was isolated from cultures with the Microbial gDNA Isolation kit (NZYTech) and the quality and concentration of extracted DNA were evaluated in a Nanodrop 2000 (Thermo Scientific). The nucleotide sequences of the *glcNK* genes from *S. jiangxiensis* and *M. smegmatis* were retrieved from the PATRIC database (http://www.patricbrc.org/, genome ID 235985.3) and from the KEGG database (https://www.genome.jp/kegg/genome.html, genome ID T02191), respectively, and used for primer design.

The *S. jiangxiensis* gene was amplified from chromosomal DNA by PCR with the forward primer 5'-ACATTTCTGCATATGACCCCGAACTGGT and the reverse primer 5'-GATCAAGCTTGTCCTTCAGTCTTTCCG, designed to remove the stop codon and introduce *Nde*I and *Hind*III restriction sites (underlined), allowing in frame insertion into the pET30a(+) expression vector with transcription of its C-terminal hexahistidine tag for purification purposes. Amplification was carried out in 50 µL mixtures with KOD hot start DNA polymerase (Novagen) using 130 ng of template DNA and the annealing temperature was set as a gradient between 53-59°C for 7 mixtures prepared, to overcome initial difficulties finding the ideal annealing temperature. PCR mixtures were analysed by agarose gel electrophoresis, the bands corresponding to the PCR product of the expected size (~1300 bp, present at annealing temperatures 58.5°C and 58.9°C of the gradient) were excised from the gel, and the DNA was purified from agarose with the JETquick Gel Extraction Spin Kit (Genomed). The amplified gene and pET30a(+) (Novagen) were both

incubated with *Nde*I and *Hind*III restriction enzymes (New England BioLabs) and the digested products were purified with the JETquick PCR Product Purification Spin Kit (Genomed). The linearised vector was further dephosphorylated with Alkaline Phosphatase Calf Intestinal (New England BioLabs), to prevent self-ligation, and purified. The digested gene and the linearised and dephosphorylated vector were ligated with T4 DNA Express Link™ ligase (Invitrogen) and transformed into chemically competent *E. coli* DH5α cells using standard procedures. Recombinant plasmids were isolated from kanamycin-resistant colonies with the ZR Plasmid Miniprep™-Classic Kit (Zymo Research) and sequenced (GATC Biotech) to confirm the identity of the inserts. Constructs confirmed to carry a copy of the *S. jiangxiensis glcNK* gene were subsequently transformed into chemically competent *E. coli* BL21 Star™(DE3) for protein expression.

The *M. smegmatis* gene had been previously cloned and the procedure is described in detail in (Nunes-Costa, 2015). Briefly, the sequence was amplified from chromosomal DNA using the forward primer 5'-TAATATCATATGATCGAGCTCGACC and the reverse primer 5'- TATAAGCTTCGTGCTTGTCCC (restriction sites underlined). Amplification products were cloned into the *Nde*I and *Hind*III sites of pET-30a(+), in frame with the vector-encoded C-terminal hexahistidine tag.

Single point mutations were introduced by site-directed mutagenesis using the QuikChange method (Stratagene).

Protein expression and purification

*Sj*GlcNK used for enzyme assays was recombinantly expressed in *E. coli* BL21 Star™(DE3) cells grown in LB medium containing kanamycin (30 μg/mL) at 37°C in an orbital shaker at 130 rpm. Upon reaching mid-exponential phase (OD_{610}≈0.8), growth temperature was gradually decreased to 20°C and gene expression was induced by adding IPTG to a final concentration of 0.5 mM. Cells were harvested 17 h later by centrifugation (9000 × *g*, 10 min, 4°C), resuspended in Buffer A (20 mM sodium phosphate buffer pH 7.4 with 0.5 M NaCl and 20 mM imidazole) and frozen at -20°C. Before purification, cells were thawed in the presence of DNase I (10 μg/mL) and disrupted by sonication on ice with five 60 Hz pulses of 45 s (15 s interval between pulses) per each 5 mL of cell suspension, with care to not let the suspension overheat. The lysate was centrifuged (15000 x *g*, 20 min, 4°C) to remove cell debris and protein production was confirmed in 12% SDS-PAGE gels. The His-tagged recombinant *Sj*GlcNK protein was isolated from filtered lysates (0.2 μm syringe filter) with a prepacked Ni-Sepharose™ High-Performance column (HisTrap™ HP

5 mL, GE Healthcare) equilibrated with Buffer A. After loading the extract onto the column, a washing step was performed by increasing the imidazole concentration to 44 mM. Elution was subsequently carried out in the same buffer by increasing the imidazole concentration in steps to 92, 212, 356 and 500 mM and collecting 2 mL fractions. The purity of each fraction was evaluated by SDS-PAGE and the purest fractions were pooled, diluted 10-fold in 20 mM BTP buffer pH 7.5, concentrated by ultrafiltration in 30 kDa cutoff centricons (Amicon) and equilibrated with the same buffer containing 50 mM NaCl. Purity was evaluated by SDS-PAGE and enzyme concentration was determined by the Bradford assay (BioRad). The recombinant enzyme was dispensed in single-use aliquots, flash frozen in liquid nitrogen and stored at -80°C.

*Ms*GlcNK for enzyme assays was obtained with a similar protocol and is described in detail in (Nunes-Costa, 2015).

Analysis of enzyme activity by thin-layer chromatography

Kinase activity was assessed in 50 µl mixtures typically containing 50 mM BTP pH 8.0, 10 mM $MgCl_2$, 10 mM phosphate acceptor, 5 mM ATP, and 2 µg of pure recombinant enzyme incubated at 37°C, with equivalent mixtures without enzyme used as negative controls. Potential phosphate donors other than ATP were tested, namely UTP, GTP, ADP and inorganic phosphate. A wide range of phosphate acceptors were tested and are listed in Table 2.1.

Table 2.1. Complete list of phosphate acceptors tested by TLC with *Sj*GlcNK and *Ms*GlcNK in the presence of ATP and $MgCl_2$.

Monosaccharides	**Hexoses:** D-glucose, L-glucose, D-galactose, D-mannose, fructose, D-tagatose, D-allose **Modified hexoses:** D-glucosamine, D-galactosamine, D-mannosamine, *N*-acetyl-D-glucosamine, *N*-acetyl-D-mannosamine, *N*-acetyl-D-galactosamine, *N*-acetyl-D-muramic acid, glucuronic acid, 2-deoxyglucose, 1-*O*-methylglucose, 3-*O*-methylglucose, 1-*O*-methylmannose **Pentoses:** arabinose, xylose
Disaccharides	maltose, isomaltose, trehalose, turanose, leucrose, saccharose, chitobiose
Trisaccharides	maltotriose, raffinose
Glycosylated antibiotics	**Aminoglycosides**: kanamycin, streptomycin, gentamicin, hygromycin B **Others**: clarithromycin, streptozotocin
Other glycosides	glucosylglycerate, glucosylglycerol
Other molecules	glycerate, glycolic acid, glutamine, mannitol

ATPase activity was assessed in the absence of a phosphate acceptor in 50 µl mixtures containing 50 mM BTP pH 8.0, 10 mM $MgCl_2$ or 5 mM $ZnSO_4$, 5 mM ATP, and 2 µg of pure recombinant enzyme, incubated overnight at 37°C.

Optimal pH and temperature, concentration of sugars and ATP, divalent cations and ideal incubation times were chosen based on the previous characterisation of *Ms*GlcNK (Nunes-Costa, 2015) and on confirmation of each variable by thin-layer chromatography (TLC) analysis for *Sj*GlcNK. Kinase activity was tested in the 5.5-9.5 pH range using 50 mM BTP buffers and 50 mM MES buffer for the lowest pH value, with $MgCl_2$ concentrations of 2.5, 5, 10, 15 and 20 mM and at 30, 37 and 45°C.

To visualise and qualitatively evaluate product formation, reaction and control mixtures were spotted on Silica 60 gel plates (Merck), 25 µL per spot, and developed with one of two solvent systems: acetic acid-ethyl acetate-water-ammonia 25% (6:6:2:1, vol/vol) or ethanol-water (7:3, vol/vol). Sugars and glycosides were visualized by spraying the dried plate with an α-naphthol–based staining solution (prepared by mixing 40.5 mL ethanol, 10.5 mL α-naphthol 15% (w/v) ethanol solution, 4 mL distilled water and 6.5 mL concentrated H_2SO_4) and charring at 120°C (Jacin & Mishkin, 1965). For plates in which it was necessary to stain amino sugars without staining ATP and ADP, a ninhydrin-based staining solution (prepared by mixing 100 mL ninhydrin 0.2% (w/v) *n*-butanol solution, 0.5 mL glacial acetic acid and 4.5 mL water) was used instead.

Purification of GlcN6P and NMR analysis

To confirm the identity of the *Sj*GlcNK product when incubated with GlcN and ATP, a 2 mL reaction mixture (50 mM BTP pH 8.0, 20 mM $MgCl_2$, 20 mM GlcN, 10 mM ATP, and 120 µg *Sj*GlcNK) was incubated overnight at 37°C and separated by TLC. Reaction conditions were optimised to ensure all ATP was depleted after incubation. The mixture was spotted on Silica 60 gel plates (Merck), 50 µL per spot, and developed with a methanol-chloroform-acetic acid-water (4:1:1:1, vol/vol) solvent system which allows the separation of the phosphorylated product from the other components in the reaction mixture (Aslam et al., 2018). After drying, the marginal lanes of the TLC plates were stained to identify the spot corresponding to the *Sj*GlcNK product, and the remaining unstained lanes were scrapped in the same region. The product was extracted from the silica with ultrapure water (2 extraction steps consisting of washing the silica with 6 mL ultrapure water for 20 min with agitation, followed by centrifugation to remove the silica, all steps at room temperature), lyophilised, and further purified on a Sephadex G10

column (GE Healthcare) with a water flow of 1 mL/min, followed by a second lyophilisation step before NMR analysis. All NMR spectra were acquired on an Avance III 400 spectrometer from Bruker operating at a central proton frequency of 400.13 MHz equipped with a BBI(F)-z H-X-D (5-mm) probe at 25°C with presaturation of the water signal.

Enzyme assays and K_m estimation

Quantification of G6P and ADP released by *Ms*GlcNK was determined as previously described (Nunes-Costa, 2015). 50 µL reactions containing 50 mM BTP pH 8.0, 20 mM Glc, 5 mM ATP, 10 mM $MgCl_2$ and 2 µg *Ms*GlcNK, and controls without enzyme, were incubated at 37°C and stopped after 0, 5, 10 or 15 min by cooling on ethanol-ice, followed by enzyme inactivation with 5 µL of 5 N HCl and neutralisation with 5 µL of 5 N NaOH. For ADP quantification, the terminated reactions were mixed with 1 mL of a detection reagent containing 50 mM Tris-HCl pH 7.4, 2.5 mM $MgCl_2$, 2.5 mM KCl, 3 U of pyruvate kinase/lactate dehydrogenase (Sigma-Aldrich), 0.3 mM NADH and 2.5 mM PEP and further incubated at 30°C for 10 min. NADH decrease, proportional to ADP release, was measured in 1 mL cuvettes at 340 nm using a UV-Vis Spectrophotometer (PerkinElmer). For G6P quantification, terminated reactions were mixed with 100 µL of a detection reagent containing 50 mM BTP pH 8.0, 5 mM $MgCl_2$, 0.3 U of glucose 6-phosphate dehydrogenase from *Leuconostoc mesenteroides* (Sigma-Aldrich) and 1.5 mM NAD and further incubated at 30°C for 10 min. NADH formation, proportional to G6P release, was measured in 96-well microtiter plates at 340 nm using a SpectraMax® Plus 384 Microplate Reader (Molecular Devices). ADP standard curves were prepared in the 0-250 nmol range and G6P standard curves were prepared in the 0-100 nmol range.

K_m values for each sugar were estimated with the ADP-Glo™ kinase Assay Kit (Promega) (Zegzouti et al., 2009).

Crystallisation and structure determination

Crystallisation and structure determination procedures are described in detail in:

Manso JA*, Nunes-Costa D*, Macedo- Ribeiro S, Empadinhas N, Pereira PJB. 2019. Molecular fingerprints for a novel enzyme family in Actinobacteria with glucosamine kinase activity. mBio 10:e00239-19. https://doi .org/10.1128/mBio.00239-19. *equal contributions.

Results

A rare Mak paralogue with a scattered phylogenetic distribution within *Actinomycetia* is associated with different BGCs

In some mycobacterial genomes, BLAST searches using the amino acid sequences of characterised mycobacterial maltokinases returned two hits, while in other mycobacterial genomes such searches only returned one hit corresponding to the canonical Mak, often located adjacent to the *treS* gene. The co-occurrence of two copies of *mak* homologues was also observed for other genomes within the class *Actinomycetia*. In addition to genomic context, paralogues could be distinguished through analysis of SSNs (Figure 2.1). In each species, the sequence identity between the canonical Mak and its paralogue was approximately 30%. Common automatic annotations assigned to these genes across different databases were "maltokinase" and "aminoglycoside phosphotransferase".

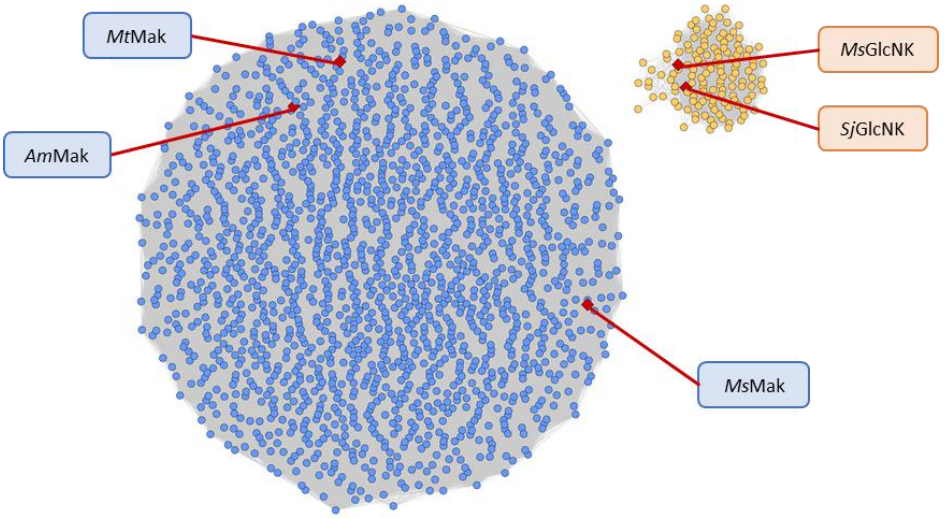


Figure 2.1. SSN for the "maltokinase" family showing the existence of two clearly defined clusters within the family. Each circle (node) represents a sequence or group of sequences sharing ≥95% identity. Grey lines interconnect pairs of sequences with an alignment score ≥35. Nodes corresponding to functionally characterised proteins are represented by red diamonds and labelled, namely the maltokinases from *M. smegmatis* (*Ms*Mak) (Roy et al., 2013), *M. tuberculosis* (*Mt*Mak) (Mendes et al., 2010) and *Actinoplanes missouriensis* (*Am*Mak) (Jarling et al., 2004), and the glucosamine kinases from *S. jiangxiensis* (*Sj*GlcNK) and *M. smegmatis* (*Ms*GlcNK) (present work).

A BLAST search using the amino acid sequence of *Sj*GlcNK as query against the UniProt database retrieved 1906 homologous sequences with less than 95% sequence identity between them and belonging to genomes with clear taxonomic assignment (i.e., sequences from metagenomic data and genomes with unclear or missing taxonomic classification to at least the genus level were removed). Of these, 1775 clustered together in the SSN with the functionally characterised Mak from *M. tuberculosis* (Mendes et al.,

2010), *M. smegmatis* (Roy et al., 2013) and *Actinoplanes missouriensis* (Jarling et al., 2004). The other 131 sequences formed a separate cluster which included *Sj*GlcNK and all the other orthologues identified in genomes which returned two BLAST hits when queried with a Mak sequence. This clear separation in the SSN suggested that these paralogues form a possible functionally divergent group within the Mak family (Figure 2.1). Furthermore, the SSN shows that GlcNK homologues have a much narrower distribution than Mak homologues and are present in only a few members of the class *Actinomycetia*, unlike the widespread Mak. The phylogenetic distribution of GlcNK homologues is also scattered, since for each genus orthologues are present in a few species and absent in most (Table 2.2).

Table 2.2. Taxonomic distribution of detected GlcNK sequences. Some sequences pertain to genomes from isolates not yet assigned to species with a validly published name, so the universe of genomes searched is larger than the number of described species in each genus. However, the number of BLAST hits is consistently and significantly lower than the number of described species, showing that most species in each genus do not contain a GlcNK orthologue.

Family	Genus	Number of detected GlcNK sequences[a]	Number of described species in the genus[b]
Bogoriellaceae	*Georgenia*	1	13
Cryptosporangiaceae	*Cryptosporangium*	1	8
Frankiaceae	*Frankia*	3	12
Microbacteriaceae	*Agromyces*	12	36
	Microcella	1	2
	Leifsonia	1	12
Micrococcaceae	*Arthrobacter*	3	54
Micromonosporaceae	*Micromonospora*	5	106
	Actinoplanes	1	47
Mycobacteriaceae	*Mycobacterium*	60	193
Nocardioidaceae	*Mumia*	3	3
	Nocardioides	3	124
Streptomycetaceae	*Streptacidiphilus*	1	15
	Streptomyces	1	680
	Kitasatospora	1	33
Streptosporangiaceae	*Nonomuraea*	16	64
	Planomonospora	1	6
	Sinosporangium	1	3
	Sphaerisporangium	1	12
	Streptosporangium	4	29
	Herbidospora	2	7
	Microbispora	6	12
	Acrocarpospora	3	4

[a]Only the sequences used to calculate the SSN were counted (less than 95% sequence identity between them).
[b]Number of species with a validly published and correct name according to the List of Prokaryotic names with Standing in Nomenclature, as of April 2021.

Genomic context analysis revealed a tight association with two other genes encoding putative enzymes, a putative sugar isomerase (SIS) of the AgaS superfamily (COG2222/PF01380) and a putative glycosyltransferase (GT) with homology to members of the GT1 family (PF00201), both with a co-occurrence ≥ 90% which indicates they likely participate in the same metabolic pathway (Figure 2.2A). A putative aldehyde dehydrogenase (PF00171) and a putative FAD-dependent oxidoreductase (PF01266) were also detected as frequent neighbours, but with a lower co-occurrence percentage. Interestingly, components of ATP-binding cassette (ABC) transporters were commonly found in the vicinity of the *glcNK* gene. None of the identified neighbours have been characterised to date, however, sequenced-based annotations in some genomes provided valuable clues to discover the function of the enzyme encoded by this *mak* paralogue. The presence of a GT and the homology with Mak suggested it to be a sugar kinase. Additionally, the putative sugar isomerase, SIS, has homology with the sugar isomerising domain of GlmS involved in GlcN6P biosynthesis. The dehydrogenase is often found annotated as "gamma-aminobutyraldehyde dehydrogenase", the FAD-dependent oxidoreductase as "amino acid dehydrogenase (deaminating)", and the ABC transporters as "spermidine/putrescine ABC transporter". Although these automatic annotations are far from being reliable, they hint at a common theme involving amine functional groups.

In several species, these genes were predicted to be included in larger BGCs by genome mining tools specialised in detecting SM gene clusters such as antiSMASH (Figure 2.2B). This was the case for *S. jiangxiensis*, in which two other genes with possible sugar-modifying activities are present in the cluster, annotated as a putative phosphoglucomutase (PGM) and a putative UTP–glucose-1-phosphate uridylyltransferase (UDPGP), thus completing a potential pathway for sugar activation and transfer. The scaffold to which this sugar is hypothetically transferred to might be produced by the cyclodipeptide synthase (CDPS) found in the same cluster. A curious genomic organisation was also detected in some mycobacterial species, where the proximity of the *glcNK* gene to nonribosomal peptide synthetases (NRPS) and/or polyketide synthases (PKS) suggests it may be involved in the assembly of glycosylated peptides/polyketides. Since the *S. jiangxiensis* homologue was part of an apparently more complete set of sugar-modifying genes and given the phylogenetic relationship of this organism with antibiotic-producing *Streptomyces*, this gene was selected for biochemical and structural studies. A mycobacterial homologue, previously identified as a kinase with very low affinity for glucose, was also further investigated.

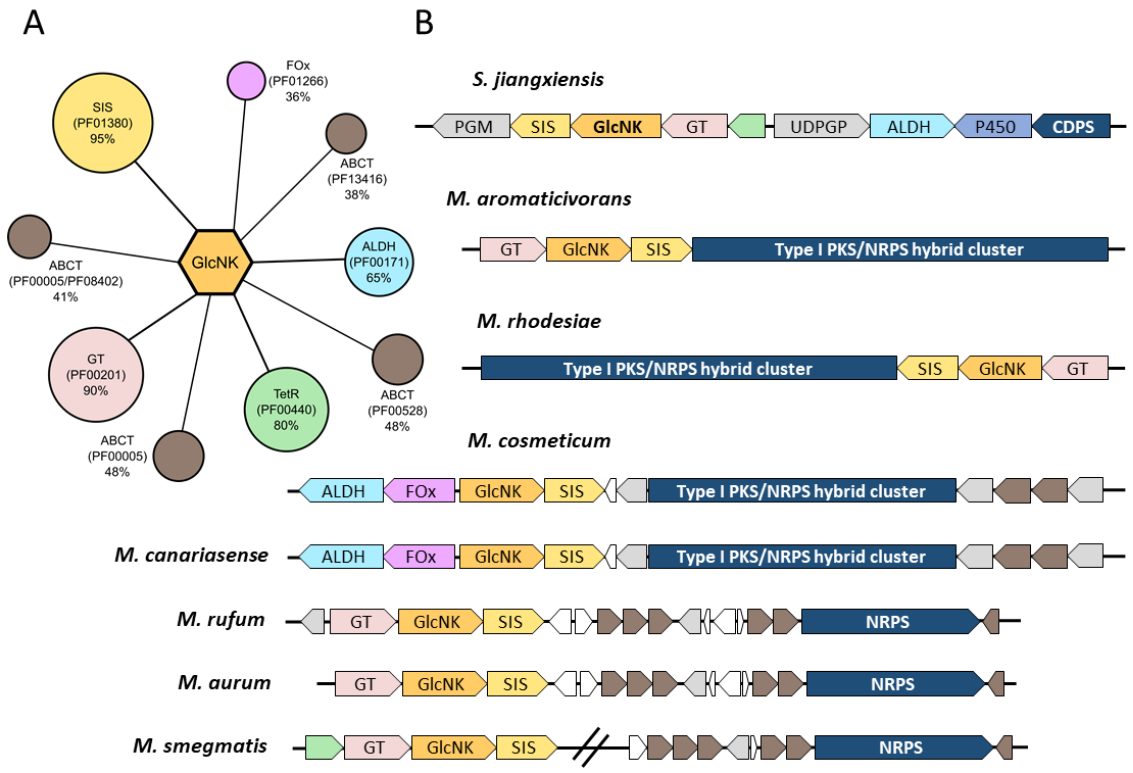


Figure 2.2. Genomic context of GlcNK homologues. **(A)** GNN for the SSN cluster corresponding to *Sj*GlcNK orthologues. Genomic neighbours with over 30% co-occurrence within a window of 20 ORFs upstream and downstream of the *glcNK* gene are shown. Pfam families are shown in brackets and the size of each circle is proportional to the co-occurrence, shown in percentage below the Pfam family. **(B)** Genomic context diagrams for *Sj*GlcNK and orthologues in other selected species, drawn according to antiSMASH results (not drawn to scale). *glcNK* is predicted to be within the borders of BGCs in all represented species except for *M. smegmatis*, which has a cluster homologous to *M. rufum* and *M. aurum* but with the sugar-modifying genes, including *glcNK*, located elsewhere on the chromosome. ORFs of unknown function are coloured white and putative enzymes whose annotation was omitted for simplicity are coloured light grey. ORFs responsible for scaffold biosynthesis are coloured dark blue, the remaining colours match those in panel A. SIS, sugar isomerase; GT, glycosyltransferase; TetR, TetR-family of transcriptional regulators; ALDH, aldehyde dehydrogenase; ABCT, ATP-binding cassette (ABC) transporter components; FOx, FAD-dependent oxidoreductase; PGM, phosphoglucomutase; UDPGP, UTP–glucose-1-phosphate uridylyltransferase; CDPS, cyclodipeptide synthase; NRPS, nonribosomal peptide synthetase; PKS, polyketide synthase (all putative functions).

The *M. smegmatis* and *S. jiangxiensis* Mak paralogues encode sugar kinases active with GlcN and with different catalytic properties

High-level expression in soluble form was achieved for the his-tagged recombinant enzymes from *M. smegmatis* and *S. jiangxiensis*, allowing the purification of both proteins as confirmed by SDS-PAGE (Figure 2.3). Gel migration suggests that the orthologue from *S. jiangxiensis* has a higher molecular weight, in accordance with ProtParam predictions (43.9 kDa and 48.2 kDa for recombinant *Ms*GlcNK and *Sj*GlcNK, respectively) The yields obtained were approx. 11 mg and 38 mg of recombinant protein per liter of *E. coli* culture for *Ms*GlcNK and *Sj*GlcNK, respectively.

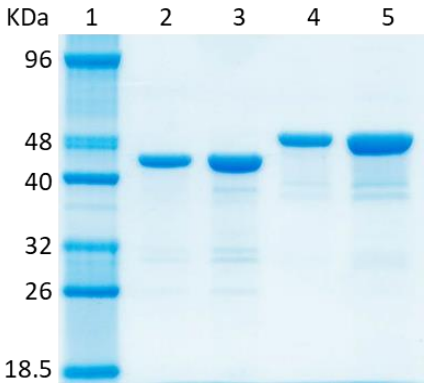


Figure 2.3. SDS-PAGE analysis of purified recombinant GlcNKs. Lane 1: molecular weight marker; lanes 2 and 3: 2 µg and 6 µg of purified recombinant GlcNK from *M. smegmatis*; lanes 4 and 5: 2 µg and 6 µg of purified recombinant GlcNK from *S. jiangxiensis*.

In order to decipher the function of these enzymes, their sequences were aligned and compared with the closest characterised homologues, mycobacterial maltokinases (Figure 2.4). Despite a slightly shorter sequence when compared to Mak, the important catalytic, nucleotide-binding and magnesium-binding motifs are highly conserved in GlcNKs. These motifs include the nucleotide-positioning P-loop ([117]VD**Q**T**N**E**SV**[124]), which contains the residues interacting with the terminal γ-phosphate of ATP, the region containing the conserved phosphate-binding lysine residue ([130]**A**VV**K**W[134]), the catalytic ([296]DV**HGD**F**H**V**GQI**[306]) and magnesium-binding ([317]**DF**D[319]) loops, as well as the [338]**DV**V/A[340] segment which together with the conserved histidine residue [230]**H** forms an hydrogen network stabilising and interlinking the nucleotide and sugar binding pockets (Fraga et al., 2015) (sequences provided in brackets correspond to *Sj*GlcNK, bolded residues are strictly conserved in the multiple sequence alignment shown in Figure 2.4).

As expected from analysing motif conservation, both *Sj*GlcNK and *Ms*GlcNK were confirmed to display kinase activity, although they were unable to phosphorylate maltose *in vitro* (Figure 2.5A), indicating an incorrect annotation of their function and confirming the functional divergence suggested by SSN analyses (Figure 2.1). Phosphorylation of glucose by *Ms*GlcNK had been previously demonstrated, however, determination of kinetic parameters revealed a high Michaelis constant (K_m > 70 mM) indicating a very low affinity for this substrate (Nunes-Costa, 2015). Since glucose phosphorylation by *Sj*GlcNK was found to be equally inefficient (K_m > 100 mM) (Figure 2.5C), substrate specificity was evaluated using a wide range of possible phosphate acceptors (Table 2.1), including 30 sugars. Due to clues obtained from genomic context analyses and database automatic annotations, several amino sugars were included in the screening as well as amino sugar-containing antibiotics. Of all substrates tested, phosphotransfer activity was only detected

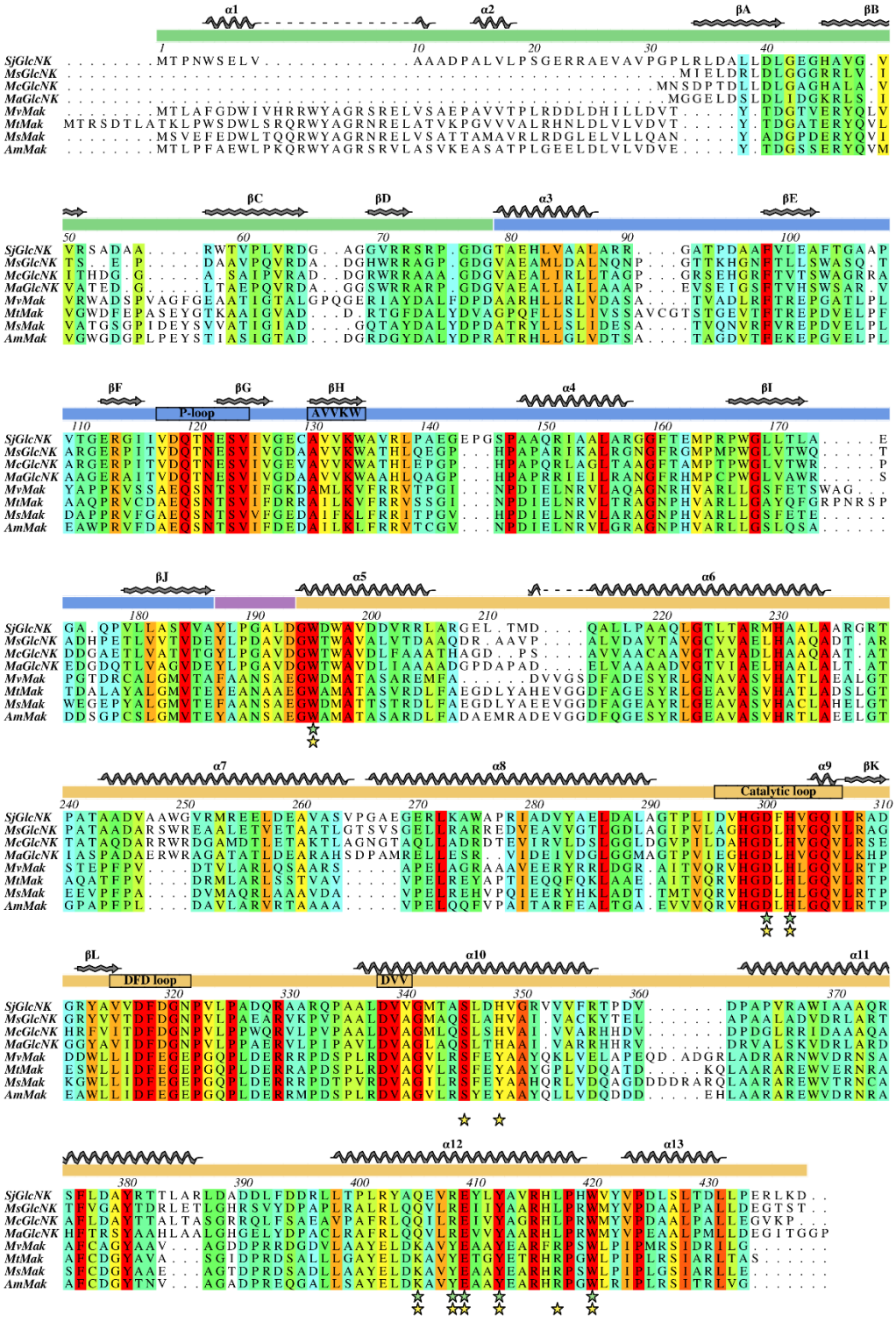


Figure 2.4. Multiple amino acid sequence alignment of *Sj*GlcNK with representative orthologues and characterised maltokinases. Residues are coloured according to conservation (red: identical residues; orange to blue: decreasing conservation; white: dissimilar). Amino acid numbers, structural subdomains, and secondary structure elements are based on the *Sj*GlcNK structure and are represented above the alignment. Subdomains are represented by coloured lines (green: N-terminal cap; blue: intermediate subdomain; magenta: linker; orange: C-lobe) and the main conserved motifs associated with phosphotransfer activity are labelled. Residues participating in GlcN and maltose binding are indicated below the alignment by green and yellow stars, respectively. UniProt entries for sequences used in the alignment: GlcNK – *S. jiangxiensis* (A0A1H7TQR5), *M. smegmatis* mc^{2}155 (I7FJX8), *M. cosmeticum* (W9AJ24), *M. aromaticivorans* JS19b1 (A0A064CHT8); Mak – *M. tuberculosis* H37Rv (O07177), *M. vanbaalenii* PYR-1 (A1TH50), *M. smegmatis* mc^{2}155 (A0R6D9), *Actinoplanes missouriensis* (Q7WUM3).

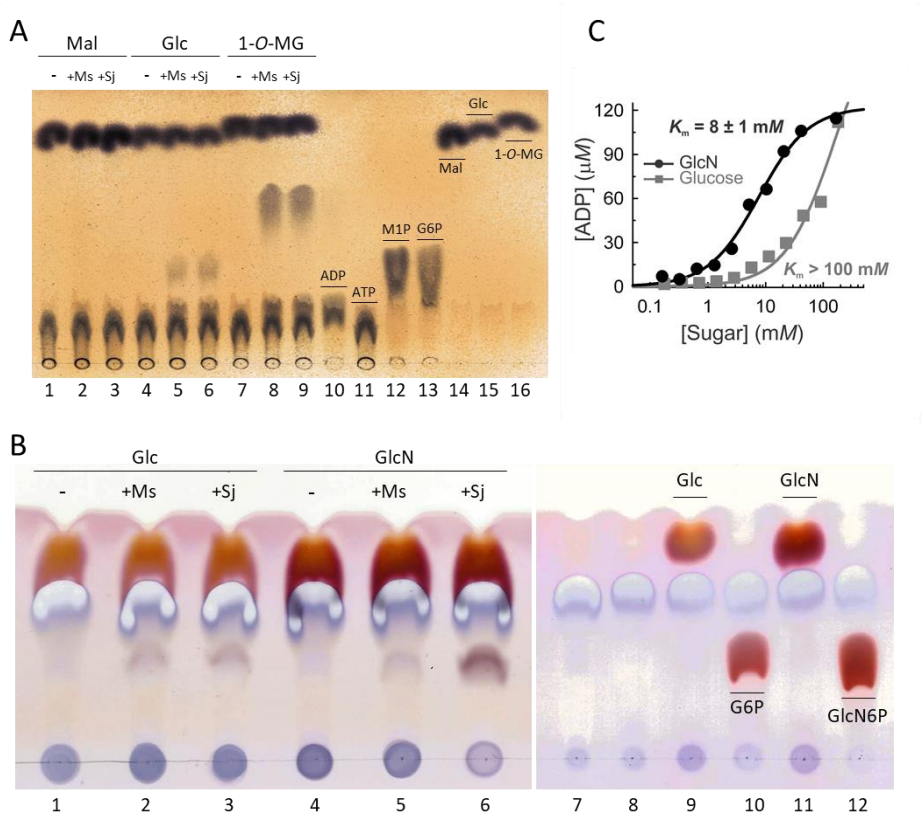


Figure 2.5. Substrate specificity and kinase activities of the *M. smegmatis* and *S. jiangxiensis* recombinant enzymes. **(A)** TLC analysis of maltose, Glc and 1-*O*-methylglucose phosphorylation by GlcNKs. Lanes 1, 4 and 7: Mal, Glc and 1-*O*-MG control mixtures without enzyme; lanes 2, 5 and 8: Mal, Glc and 1-*O*-MG reaction mixtures with *Ms*GlcNK; lanes 3, 6 and 9: Mal, Glc and 1-*O*-MG reaction mixtures with *Sj*GlcNK; lanes 10-16: standards of ADP, ATP, M1P, G6P, Mal, Glc and 1-*O*-MG. Plates were developed in an ethanol-water (7:3, vol/vol) solvent system and stained with an α-naphthol-based solution. Control and reaction mixtures were incubated at 37°C for 2 h. Maltose is not phosphorylated by either enzyme. **(B)** TLC analysis of Glc and GlcN phosphorylation by GlcNKs. Lanes 1 and 4: Glc and GlcN control mixtures without enzyme; lanes 2 and 5: Glc and GlcN reaction mixtures with *Ms*GlcNK; lanes 3 and 6: Glc and GlcN reaction mixtures with *Sj*GlcNK; lanes 7-12: standards of ATP, ADP, Glc, G6P, GlcN and GlcN6P. Plates were developed in an acetic acid-ethyl acetate-water-ammonia 25% (6:6:2:1, vol/vol) solvent system and stained with a ninhydrin-based solution, which does not stain ADP and ATP as shown in lanes 7 and 8. Control and reaction mixtures were incubated at 37°C for 1 h. The most efficient reaction was observed for *Sj*GlcNK with GlcN. **(C)** Kinetics of *Sj*GlcNK activity towards Glc and GlcN. ADP release was quantified with the ADP-Glo kinase assay and experimental data were fitted to the Michaelis-Menten equation. The measured K_m for GlcN (black) is more than 10-fold lower than that for glucose (grey). Panel A and B show representative plates of at least 3 separate experiments. Panel C was prepared by José A. Manso and adapted with permission from Manso et al., 2019.

for D-glucose, 1-*O*-methyl-D-glucose and D-glucosamine, using ATP as the phosphate donor (Figure 2.5A,B). Despite the ability of the Mak from *M. tuberculosis* to utilise UTP and GTP efficiently, phosphoryl transfer was not observed in the presence of these NDPs (data not shown). TLC comparative analysis indicated that for the *S. jiangxiensis* enzyme, the reaction with GlcN was more efficient than with Glc (Figure 2.5B). This was confirmed by estimating the Michaelis constant for GlcN (K_m = 8±1 mM) which was found to be much lower than for Glc, suggesting that *Sj*GlcNK is better described as a GlcN kinase (EC 2.7.1.8) (Figure 2.5C). The preference for GlcN was confirmed for the *M. smegmatis* enzyme (K_m =

14 ± 2 mM), albeit this orthologue was found to be less efficient than the one from *S. jiangxiensis* under the *in vitro* conditions tested (Figure 2.5B).

The identity of the *Ms*GlcNK product when incubated with ATP and glucose had previously been identified as G6P (Nunes-Costa, 2015). To confirm if GlcN was also being phosphorylated at position 6, preparative scale purification of the *Sj*GlcNK product was carried out, followed by ^{1}H-NMR analysis (Figure 2.6). Separation of the phosphorylated product from the other reaction components was successfully achieved by TLC using an adequate mobile phase (Figure 2.6A). Comparison between the signals of the ^{1}H-NMR spectra for the enzymatic product and for a GlcN6P commercial standard shows a perfect alignment, and spiking the sample with the standard revealed no new signals and a slight increase of those already present, thus confirming the identity of the enzymatic product as GlcN6P (Figure 2.6B).

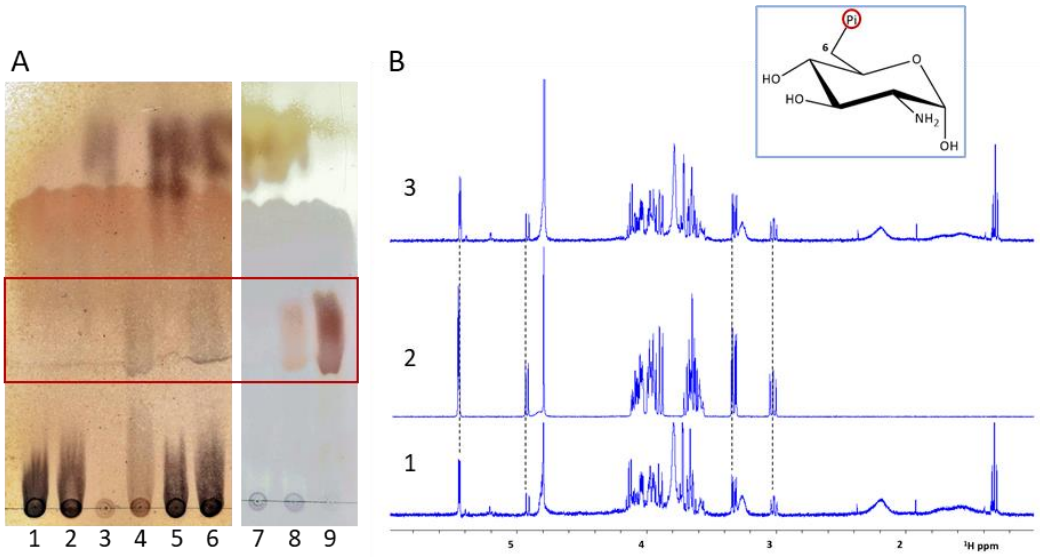


Figure 2.6. Separation and NMR analysis of the *Sj*GlcNK product. **(A)** TLC plates showing the migration pattern of the reaction mixture used for product purification, when developed in a methanol-chloroform-acetic acid-water (4:1:1:1, vol/vol) solvent system. The plate on the left was stained with an α-naphthol-based solution and the plate on the right was stained with a ninhydrin-based solution. The red box corresponds to the area that was scrapped in unstained plates used for product purification. Lanes 1-4: standards of ATP, ADP, GlcN and GlcN6P; lanes 5 and 7: control mixture without enzyme; lanes 6 and 8: reaction mixture (see methods); lane 9: GlcN6P standard. **(B)** ^{1}H-NMR spectra of *Sj*GlcNK product. 1: sample purified from TLC; 2: GlcN6P standard (Sigma-Aldrich) in D_2O; 3: sample 1 spiked with an aliquot of standard 2. Inset: chemical structure of GlcN6P. NMR spectra figures were prepared by Pedro Lamosa at CERMAX, ITQB-NOVA (Oeiras, Portugal).

We had previously observed the ability of *Ms*GlcNK to hydrolyse ATP non-productively, even in the absence of a phosphate acceptor (Nunes-Costa, 2015). Employing ADP and G6P detection methods in parallel, it was possible to confirm that in the presence of 20 mM glucose and 5 mM ATP there is approx. 30% more ADP being released than G6P

(Figure 2.7A). The same behaviour was observed by TLC analysis in the presence of GlcN, with GlcN6P production not matching ADP release (Figure 2.7C). Interestingly, the uncoupling of the hydrolysis and phosphotransfer activities was markedly exacerbated by Zn^{2+}, and in the absence of a phosphate acceptor, Zn^{2+} was found to stimulate enzymatic ATP hydrolysis (Figure 2.7B). The *S. jiangxiensis* homologue, on the other hand, did not

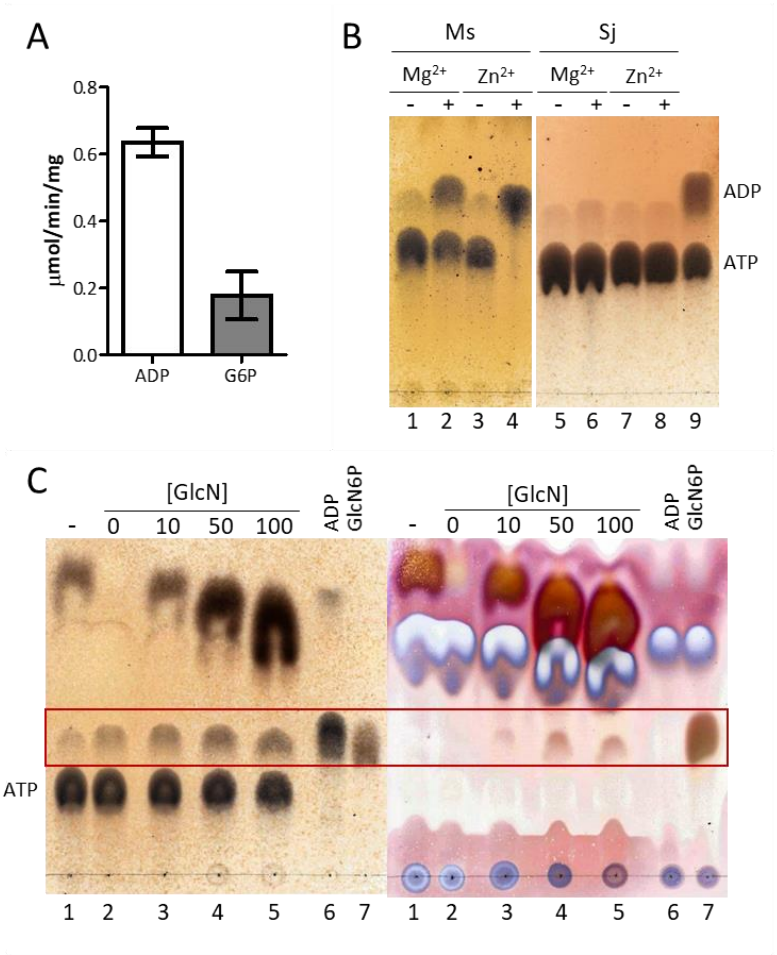


Figure 2.7. ATPase activity of *Ms*GlcNK. **(A)** Quantification of ADP and G6P released by *Ms*GlcNK in the presence of Glc. Reactions (50 mM BTP pH 8.0, 20 mM Glc, 5 mM ATP, 10 mM $MgCl_2$, 2 µg *Ms*GlcNK) were incubated for 15 min at 37°C. Plotted data are the mean of four separate experiments. **(B)** TLC analysis of ATP hydrolysis by *Ms*GlcNK and *Sj*GlcNK in the absence of a phosphate acceptor and in the presence of 10 mM $MgCl_2$ or 5 mM $ZnSO_4$ ($ZnCl_2$ was also tested, with equivalent results – not shown). Reactions (50 mM BTP pH 8.0, 5 mM ATP, 2 µg *Ms*GlcNK) and controls without enzyme were incubated overnight at 37°C. Lanes 1 and 5: control mixtures containing Mg^{2+}; lanes 2 and 6: reaction mixtures containing Mg^{2+}; lanes 3 and 7: control mixtures containing Zn^{2+}; lanes 4 and 8: reaction mixtures containing Zn^{2+}; lane 9: standards of ATP and ADP. **(C)** TLC analysis of ADP and GlcN6P released by *Ms*GlcNK in the presence of GlcN. Reactions (50 mM BTP pH 8.0, 0-100 mM GlcN, 5 mM ATP, 10 mM $MgCl_2$, 2 µg *Ms*GlcNK) and controls without enzyme were incubated for 1 h at 37°C. The mobile phase used does not separate ADP and GlcN6P, which migrate within the boxed area, but the two staining solutions differentiate between them. The plate on the left was stained with α-naphthol which stains ADP and GlcN6P standards (last two lanes on each plate), while the equivalent plate on the right was stained with ninhydrin which stains GlcN6P but not ADP. Lane 1: control mixture without enzyme; lanes 2-5: reaction mixtures with 0, 10, 50 and 100 mM GlcN; lanes 6 and 7: standards of ADP and GlcN6P. The faint spots on lanes 1 and 6 of the α-naphthol-stained plate are due to non-enzymatic degradation of ATP to ADP and ADP to AMP, respectively, as confirmed in separate TLC runs (not shown). Plates shown in panels B and C are representative of at least 3 separate experiments.

hydrolyse ATP even when incubated overnight in the presence of Zn^{2+} (Figure 2.7B). Whether the *M. smegmatis* homologue can function as an efficient GlcNK *in vivo* or under yet-unknown *in vitro* conditions remains to be elucidated, as well as the molecular basis for the different properties observed for the two homologues. While these results are puzzling and difficult to understand from a physiological point of view, the existence of kinases with significant uncoupled ATPase activity *in vitro* has been reported before (Kim et al., 2006), which warrants future research on this enigmatic result.

*Sj*GlcNK displays a typical ELK fold similar to maltokinases and high conformational flexibility

Crystallisation and structure determination of *Ms*GlcNK was not successful, but the crystal structure of *Sj*GlcNK could be solved at a resolution of 1.98 Å, revealing the same typical ELK fold observed in maltokinases (Fraga et al., 2015; Scheeff & Bourne, 2005). The overall structure consists of an N-lobe, formed by the N-terminal cap and intermediate subdomain, and a C-lobe, both separated by a linker, with a relative arrangement that results in an overall bean shape (Figure 2.8A). An acidic groove is formed between the two lobes, where the catalytic loop and magnesium-binding DFD loop can be found, close to the P-loop in the intermediate subdomain (Figure 2.8B). The N-terminal cap is composed of three long antiparallel β-strands (βA, βB and βC) which form a curved β-sheet enclosing the two α-helices, α1 and α2, and the shorter strand βD (Figure 2.4 and Figure 2.8C). The intermediate subdomain consists of a β-sheet composed of six strands (βF-βG-βH-βJ-βI-βE) flanked by the helixes α3 and α4 (Figure 2.4 and Figure 2.8D). The C-lobe, connected to the intermediate subdomain by a seven-residue linker, comprises a two-stranded β-sheet (βK-βL), two 4-helical bundles (α5-α6-α10-α11 and α7-α8-α12-α13) and an additional helix α9 downstream of the catalytic loop (Figure 2.4 and Figure 2.8E). While the intermediate subdomain and the C-lobe are almost identical to the corresponding subdomains in Mak (Figure 2.8D,E), the N-terminal cap is significantly different, being shorter and missing some secondary structure elements found in Mak (Figure 2.8C). In Mak, the function of the N-terminal cap subdomain is unknown, but it seems to be involved in the interaction with TreS, with which it forms a hetero-octameric complex, stimulating Mak's kinase activity (Kermani et al., 2019; Roy et al., 2013). Indeed, Mak orthologues are found in some organisms fused to TreS, and the unique N-terminal cap subdomain of Mak is conserved in these bifunctional TreS-Mak units (Chandra et al., 2011). Thus, the different N-terminal caps of Mak and GlcNK can be easily explained by their participation in distinct pathways.

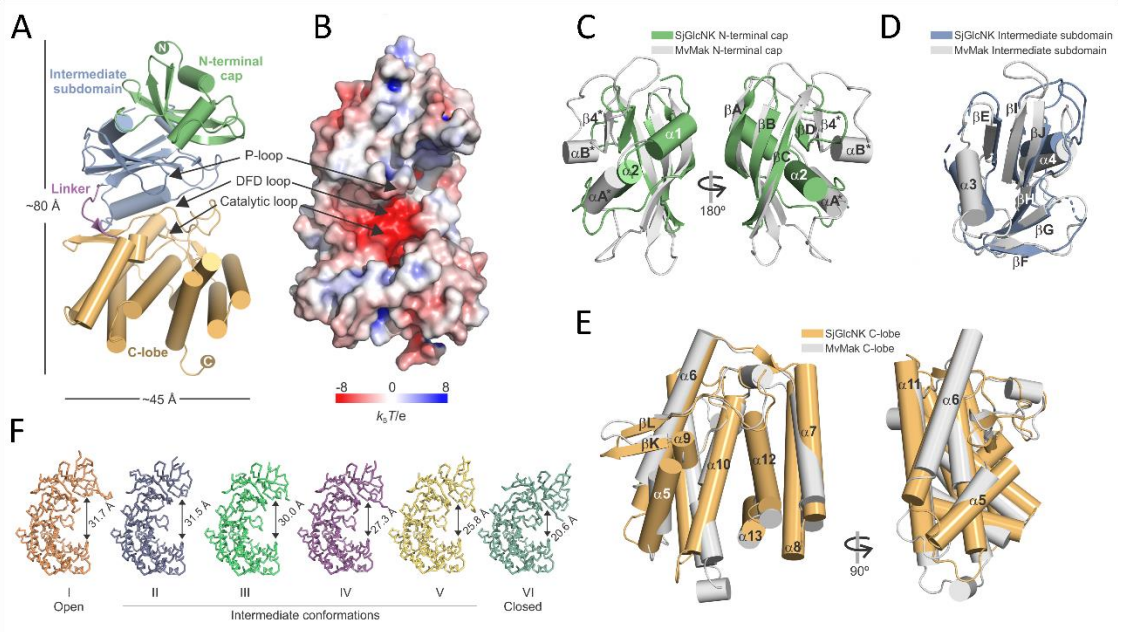


Figure 2.8. Overall structure of *Sj*GlcNK and comparison with *Mv*Mak. **(A)** Cartoon representation of the 3D structure of *Sj*GlcNK with the N-terminal cap, intermediate subdomain, linker, and C-lobe labelled and coloured as in Figure 2.4. Arrows indicate the positions of the P-loop, DFD loop, and catalytic loop. **(B)** Solid surface representation of *Sj*GlcNK coloured according to the electrostatic potential contoured from -8 (red) to 8 (blue) k_BTe^{-1} (k_B, Boltzmann's constant; T, temperature [K]; e, charge of an electron) highlighting the acidic nature of the active site located in the groove between the intermediate subdomain and the C-lobe. **(C)** Structural superposition of the N-terminal caps (two views rotated by 180° around y), **(D)** intermediate subdomains, and **(E)** C-lobes (two views rotated by 90° around y) of *Sj*GlcNK (coloured as in panel A) and *Mv*Mak (light grey; PDB entry 4U94). Secondary structure elements are labelled as in Figure 2.4. **(F)** Ribbon representation of six crystallographic models of *Sj*GlcNK, highlighting their different conformational states. Differences in the relative position of the two lobes (measured distance between the Cα atoms of A15 and H416 is shown for each state) indicate conformational flexibility. The structures are labelled with Roman numerals and ordered from left to right, according to decreasing interlobe distance. All figures were prepared by José A. Manso and adapted with permission from Manso et al., 2019.

*Sj*GlcNK was crystallised in two different monoclinic ($P2_1$) forms, one containing two molecules in the asymmetric unit and the other containing four, making up a total of six individual molecules for which a 3D structure could be determined. Interestingly, the six monomers were found to correspond to six different conformations of *Sj*GlcNK, in which the individual structures of each subdomain are highly conserved but the relative positions of the N- and C-lobes are strikingly different between conformers, suggesting an opening/closing movement with the seven-residue linker functioning as a hinge (Figure 2.8F). Across the six conformers, the distance between equivalent atoms of each lobe varies from 31.7 Å (open conformation) to 20.6 Å (closed conformation), with four intermediate conformers providing snapshots of this transition (Figure 2.8F). The closed state likely corresponds to the active conformation of the enzyme, since the catalytic lysine K113, which in the open state sits far apart from the active site, approaches and establishes contacts with the catalytic aspartate D317 as the two lobes move towards each other.

Crystal structure of a productive complex of *Sj*GlcNK reveals the mechanism of phosphoryl transfer

Co-crystallisation of *Sj*GlcNK with a molar excess of its substrates, ATP and GlcN, resulted in a crystal structure of a productive complex solved at a resolution of 2.15 Å (Figure 2.9A), providing a snapshot of the phosphoryl transfer reaction. Indeed, GlcN and the two hydrolysis products of ATP, ADP and inorganic phosphate (Pi), could be clearly identified in the electron density map at the active site of the enzyme, as well as two magnesium ion co-factors (Figure 2.9B). All components were caught midreaction, providing detailed mechanistic insights into how phosphoryl transfer is accomplished (Figure 2.9C). As expected, *Sj*GlcNK displayed a closed conformation when complexed with its substrates, confirming that the closed state conformer corresponds to the active conformation (Figure 2.9A). In this structure, ADP is located in the cleft between the two lobes with the adenine group accommodated into a hydrophobic pocket within the hinge region, binding in a similar way to what was described for Mak (Fraga et al., 2015). In particular, adenine establishes hydrophobic contacts with Y187 and L188 at the hinge, with V131, P164, and V185 at the intermediate subdomain, and with L307 and V316 at the C-lobe. Both the adenine and ribose moieties establish polar contacts with the main chain of A186 and L188 and with the side chain of D193, residues located at the hinge. This extensive network of contacts between lobes is likely involved in favouring the closed conformation of the enzyme upon nucleotide binding. In this complex, the catalytic lysine K133 is seen hydrogen bonding with the α and β-phosphate groups of ADP (Figure 2.9B,C), and the positive charge of this conserved residue increases the electrophilicity of the γ-phosphate group of ATP, facilitating its hydrolysis as has been described for other kinases (Matte et al., 1998). The leaving phosphate, seen in this crystal perfectly oriented for transfer to position 6 of GlcN, is stabilised by D300 and by the two magnesium ions (Figure 2.9B,C) whose octahedral coordination drives ATP breakdown similar to what is observed for the eukaryotic cAMP-dependent protein kinase (PKA) (Bastidas et al., 2013; Gerlits et al., 2015; Kovalevsky et al., 2012). The magnesium ions are coordinated by the catalytic residues D317 (in the DFD loop) and Q305 (in the catalytic loop), stabilising the observed transition state (Figure 2.9B,C).

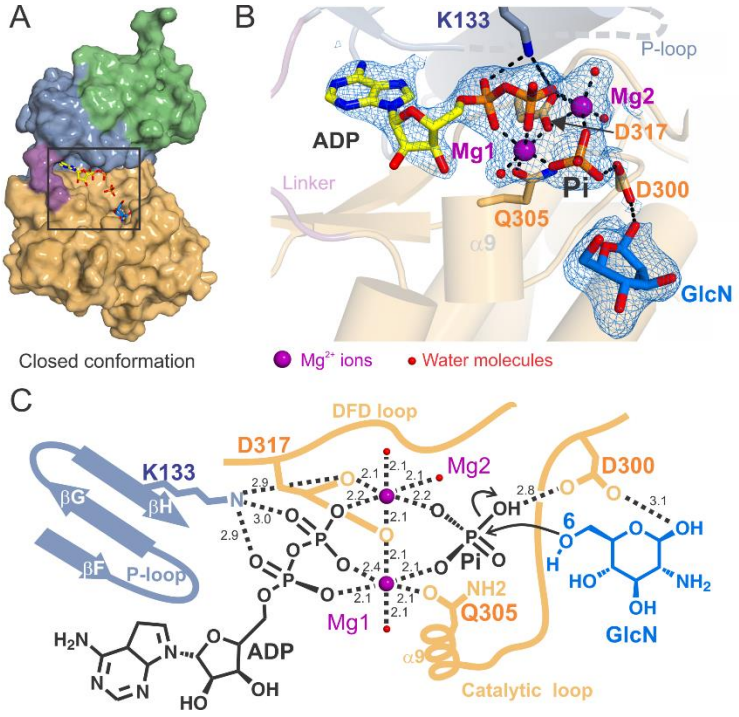


Figure 2.9. The crystal structure of a productive complex of *Sj*GlcNK unveils the mechanism of phosphate transfer. **(A)** Surface representation of the *Sj*GlcNK-GlcN-ADP-Pi quaternary complex with subdomains coloured as in Figure 2.8A. The substrates in the active site are boxed. **(B)** Closeup of the enzyme active site in the quaternary complex with the electron density map (2*mFo-DFc* contoured at 1.0 σ) around ADP, Pi, GlcN, Mg^{2+} ions and Mg^{2+}-coordinating water molecules shown as a blue mesh. Substrates are shown as sticks with oxygen atoms red, nitrogen dark blue, phosphorus orange, and carbon yellow (ADP) or blue (GlcN). Metal coordination and hydrogen bonds are indicated by black dashed lines. The protein is depicted in cartoon representation and colour coded as in Figure 2.8A. Catalytic residues K133, D300, Q305, and D317 are represented as sticks (carbon atoms coloured according to subdomain colour scheme). **(C)** Schematic representation of the transition state of GlcN phosphorylation by *Sj*GlcNK. Magnesium ion coordination and polar interactions are shown as black dashed lines, labelled with the respective distances in Å. Magnesium ions and water molecules are shown as magenta and red spheres, respectively. Figure prepared by José A. Manso and adapted with permission from Manso et al., 2019.

Since the 3D structure of *Ms*GlcNK could not be experimentally determined, a structural model was obtained using the active conformation of *Sj*GlcNK as a template (Figure 2.10). All the important binding and catalytic residues mentioned above are conserved in *Ms*GlcNK except for V185 and V316, replaced by an aspartate and a threonine, respectively, and A186, replaced by a glutamate. These substitutions in residues lining the nucleotide binding pocket might contribute to an overall less hydrophobic and more acidic environment, as suggested by analysis of the surface electrostatic potential of the *Ms*GlcNK model compared with that of the experimental structure of *Sj*GlcNK (Figure 2.10). However, it is not clear if this difference can affect the binding and stability of the nucleotide and contribute to the observed ATPase activity of *Ms*GlcNK (Figure 2.7). The distinct catalytic properties of both enzymes may also be related to a difference in the mobility of the two lobes, and indeed the substitution of Gly190 in *Sj*GlcNK for an aspartate in *Ms*GlcNK at the hinge section suggests this region could be more flexible in

the *S. jiangxiensis* orthologue, but its true impact on subdomain movement is difficult to anticipate without molecular dynamic studies. The experimental determination of the structure of the *M. smegmatis* orthologue, particularly in complex with zinc ions, could also be valuable to elucidate this question. Notably, zinc cannot replace magnesium in driving phosphoryl transfer with either enzyme and has no effect on *Sj*GlcNK kinase activity, but stimulates *Ms*GlcNK-catalysed ATP breakdown. Furthermore, the magnesium binding residues, D317 and Q305, are conserved in both proteins (Figure 2.4 andFigure 2.9) and do not offer clues regarding the different behaviour of each protein in the presence of different divalent cations. These observations suggest that zinc binds *Ms*GlcNK and participates in catalysis in a manner that cannot be predicted from the *Sj*GlcNK structure and must be experimentally elucidated.

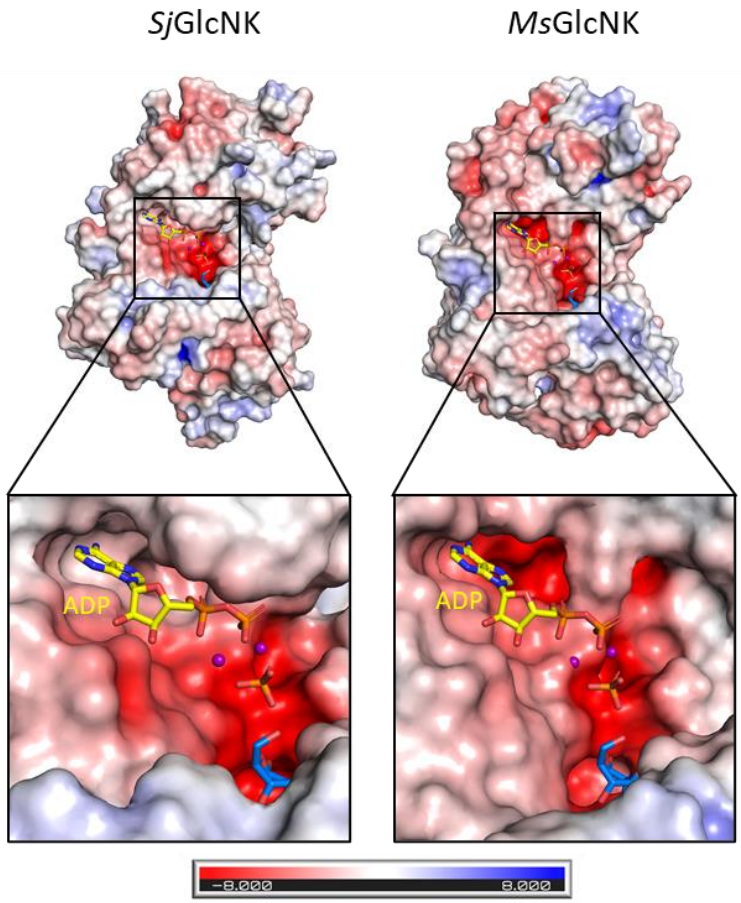


Figure 2.10. Solid surface representation of *Sj*GlcNK and of a *Ms*GlcNK model, highlighting the differences in electrostatic potential found in substrate-binding pockets. Surface is coloured according to the electrostatic potential contoured as in Figure 2.8B and shows a more acidic environment in the predicted ADP binding pocket of the *M. smegmatis* orthologue, particularly in the region predicted to interact with adenine. The *Ms*GlcNK model was obtained with Swiss-Model using the *Sj*GlcNK productive complex (PDB entry: 6HWL) as a template (target-template sequence identity = 43.32%; GMQE = 0.68; QMEAN = -3.03). The target model was superposed with the template and the ligands shown (yellow: ADP; purple: magnesium ions; orange: Pi; blue: GlcN) are positioned as in the template. GlcN and Pi could be modelled and their positions in the model match those in the template perfectly, but ADP could not (a ligand is considered for modelling if it has at least three coordinating residues in the protein and those residues are conserved in the target–template alignment). Figures were prepared with PyMOL version 2.4.1 using the APBS Electrostatics Plugin.

The molecular determinants for substrate specificity are highly conserved

The overall structures of Mak and GlcNK are very similar and superposition of the C-lobes of *Mt*Mak and *Sj*GlcNK in complex with their preferred substrates reveals that the sugar-binding pocket is located in the same region for both, framed by the catalytic loop and helices α5, α10 and α12, and lined by residues which are mostly conserved between the two enzymes. However, careful analysis of residue conservation in the regions relevant for sugar binding reveals the molecular determinants responsible for substrate specificity (Figure 2.11). Although six strictly conserved residues are involved in both maltose and GlcN binding (W195, D300, H302, E409, Y412, and W420), key differences could be identified and correlated with the structural and chemical differences between the two sugars. Namely, two residues interacting with the second glucose moiety of maltose in *Mt*Mak, Y370 and R438, are replaced by H348 and L417 in *Sj*GlcNK, unable to establish equivalent contacts, which could explain the inability of this enzyme to phosphorylate maltose (Figure 2.5A). Additionally, *Mt*Mak maltose-contacting residues K426 and Y429 are replaced by Q405 and R408 in all sequences of GlcNK homologues analysed (Figure 2.11B). In *Sj*GlcNK, R408 doesn't establish direct contacts with GlcN but is involved in stabilising the conformation of Q405, positioning it to hydrogen bond with the amino group of GlcN (Figure 2.11A). This interaction between Q405 and GlcN also explains the preference for GlcN over Glc, which lacks an amino group. Indeed, when this glutamine residue was replaced by an adenine in an *Sj*GlcNK mutant, the activity with GlcN was largely reduced but the activity with Glc was not affected (Figure 2.11C). These results suggest that Q405 does not interact with Glc, explaining the reduced affinity of the enzyme for this substrate (Figure 2.5C).

Although conserved between Mak and GlcNK sequences, the glutamate residue in the sugar-binding pocket seems to also play a key role in GlcN binding. While in *Mt*Mak E430 establishes a single hydrogen bond with a hydroxyl group of maltose, the equivalent E409 in *Sj*GlcNK can establish two hydrogen bonds, with the amino group and with one hydroxyl group of GlcN (Figure 2.11A). Furthermore, its negative charge contributes to the overall acidic environment of the sugar-binding pocket (Figure 2.10), which might also play a role in tethering GlcN given the mildly basic nature of its amino group (pKa 7.58 (Bichsel & Von Gunten, 2000)). Thus, it is conceivable that GlcN is protonated in the enzyme's active site resulting in a stronger interaction with the negatively charged E409 and adding an additional electrostatic component to GlcN's specific recognition by

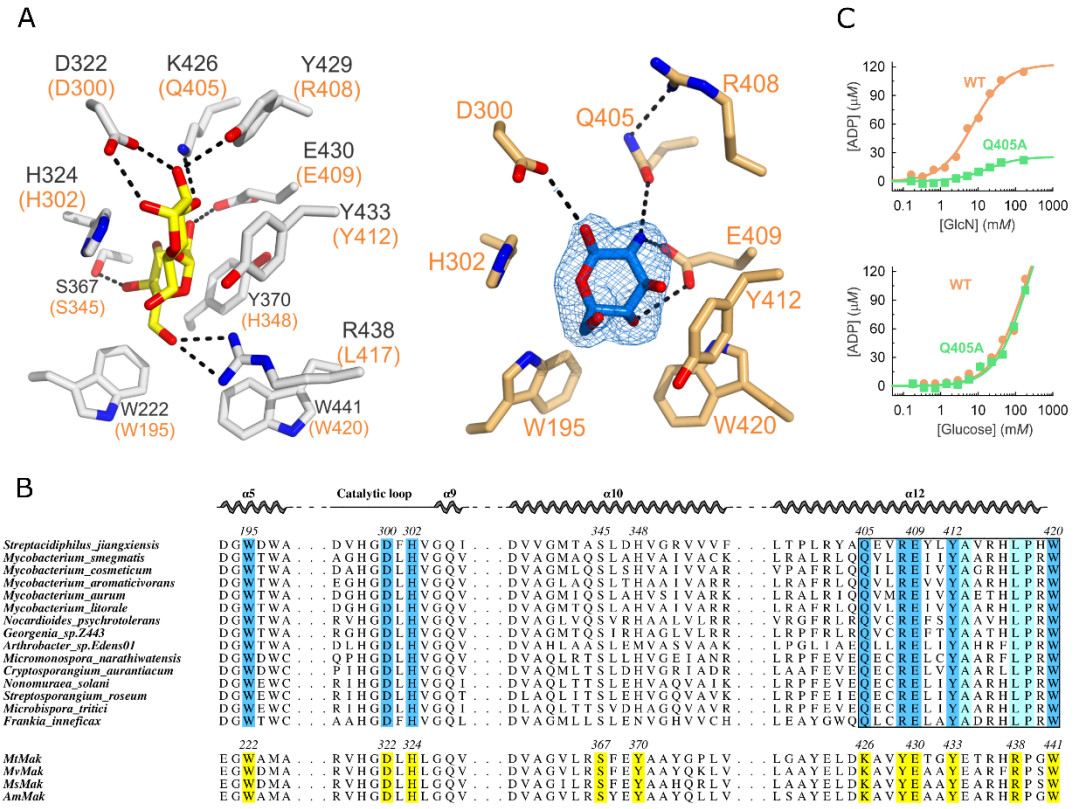


Figure 2.11. Molecular determinants for substrate specificity in *Sj*GlcNK. **(A)** Closeup view of the sugar binding sites of *Mt*Mak (PDB entry 4O7P) (left) and *Sj*GlcNK (right). Amino acids involved in sugar binding and sugars are shown as sticks with oxygen atoms red, nitrogen dark blue, and carbon light grey (*Mt*Mak) or light orange (*Sj*GlcNK) or yellow (maltose) or blue (GlcN). Residues interacting with maltose are labelled in black and the equivalent residues in *Sj*GlcNK are indicated in orange parentheses. Residues interacting with GlcN are labelled in orange. The electron density map (2*mFo-DFc* contoured at 1.0 σ) around the bound GlcN molecule is shown as a blue mesh. Dashed black lines represent polar contacts. **(B)** Multiple amino acid sequence alignment of the regions involved in sugar binding. Mak sequences are labelled with the codes used in Figure 2.4 and GlcNK sequences are labelled with the species name. Maltose and GlcN-contacting residues are coloured yellow and blue, respectively, and the other strictly conserved residues completing the proposed GlcNK consensus signature sequence (boxed) are coloured turquoise. Secondary structure elements for *Sj*GlcNK are represented above the alignment. Residue numbers based on *Sj*GlcNK and *Mt*Mak are given above the respective sequences. UniProt entries for additional GlcNK sequences (not used in Figure 2.4): *Mycobacterium litorale* (A0A1U9PBG4), *Mycobacterium aurum* (A0A448ITV9) *Frankia inneficax* (E3JA12), *Streptosporangium roseum* (D2ATB4), *Nonomuraea solani* (A0A1H6B4L3), *Micromonospora narathiwatensis* (A0A1A8ZPS6), *Nocardioides psychrotolerans* (A0A1I3PBM5), *Cryptosporangium aurantiacum* (A0A1M7H263), *Georgenia* sp. Z443 (A0A5B8CB00), *Arthrobacter* sp. Edens01 (A0A0P7GHY6), *Microbispora tritici* (A0A5D0NSI9). **(C)** Effect of the structure-based amino acid replacement Q405A on the activity of *Sj*GlcNK with GlcN (top) and glucose (bottom). Data were fitted to the Michaelis-Menten equation. Figures in Panels A and C were prepared by José A. Manso and adapted with permission from Manso et al., 2019. Panel B was prepared with Aline.

*Sj*GlcNK. In light of these observations, the region containing Q405, R408 and E409 is predicted to harbour an important motif of GlcN recognition. Close inspection of a multiple sequence alignment including 15 sequences of GlcNK homologues from several different families within the class *Actinomycetia* shows strict conservation of these residues and of Y412 and W420 (also conserved in Mak) as well as the non-contacting residues A413, L417 and P418 (Figure 2.11B). Notably, a BLAST search with the consensus

sequence "Q-x(2)-RE-x(2)-YA-x(3)-LP-x-W" in the nonredundant protein sequence database identified 132 sequences, annotated as putative maltokinases, hypothetical proteins, or aminoglycoside phosphotransferases of unknown function, most of which corresponded to sequences that clustered with *Sj*GlcNK in the SSN (Figure 2.1). This consensus sequence is therefore proposed as a signature motif to distinguish GlcNK proteins from their maltokinase homologues in the genomes of several actinomycetes.

Discussion

The impressive chemical diversity of SM has a genomic basis, with unique sets of enzymes being involved in their often-complex and unique biosynthetic pathways. One known path to metabolic novelty is gene duplication followed by functional divergence, which allows the evolution of different functions within an enzyme superfamily (Díaz-Mejía et al., 2007; Scossa & Fernie, 2020). A notable example of protein evolution can be found in the protein kinase-like (PKL) superfamily, present in all three domains of life and which includes ePKs involved in cell signalling and regulatory pathways in eukaryotes, as well as ELKs which phosphorylate a vast array of substrates in prokaryotes, including proteins, antibiotics, sugars, lipids and amino acids (Kannan et al., 2007; Scheeff & Bourne, 2005). Kinases within this superfamily share the same structural fold and the biochemical function of ATP-dependent phosphoryl transfer, although they differ greatly in preferred substrates and display very low sequence identity between them (7-17%), revealing extensive modifications over long evolutionary timescales and an impressive functional plasticity (Kannan et al., 2007; Scheeff & Bourne, 2005). Early on, paralogue detection and sequence comparison methods were used to discover novel ELK families in bacteria and archaea (Leonard et al., 1998) and structural studies assigned to this superfamily then unsuspected members such as the APHs (Hon et al., 1997). In this work, we report for the first time glucosamine kinase activity in a member of this superfamily (Figure 2.5B,C), further expanding the array of substrates known to be phosphorylated by ELKs. This novel GlcNK was discovered as a paralogue of Mak, a bacterial enzyme which phosphorylate maltose in the context of a metabolic pathway involved in α-1,4-glucan biosynthesis (Elbein et al., 2010; Fraga et al., 2015; Kalscheuer et al., 2010).

The PKL core fold consists of a bilobal architecture, with the N-lobe and C-lobe connected by a short hinge section, and with an exposed catalytic cleft between the two lobes where phosphoryl transfer takes place. The N-lobe is mainly composed of β-strands

and is responsible for positioning ATP, while the C-lobe is mainly helical in nature and recognises the phosphate acceptor substrate (Scheeff & Bourne, 2005). In addition to this basic kinase core, which we confirmed to be present in *Sj*GlcNK (Figure 2.8A), other domains can be found in the different members of the superfamily, often involved in regulatory and signalling mechanisms or in protein-protein interactions (Pérez et al., 2008; Walker et al., 1999). The N-terminal cap found in Mak is structurally similar to the cystatin family of protease inhibitors and has been proposed to act as an anchoring point tethering Mak and its metabolic partner TreS to the site of α-(1,4)-glucan biosynthesis (Fraga et al., 2015). Mak and TreS are fused in many organisms and form a hetero-octameric complex that stimulates Mak activity in mycobacteria (Kermani et al., 2019; Roy et al., 2013). Interestingly, truncation of *M. smegmatis* Mak N-terminal cap substantially reduces maltokinase activity and affinity to TreS, showing its involvement in the TreS-Mak interaction (Kermani et al., 2019). *Sj*GlcNK possesses an N-terminal cap topologically similar to the one found in Mak, but 10-14 residues shorter (Figure 2.8C). A possible functional role of this subdomain in GlcNKs and its involvement in putative interactions with other proteins participating in the same pathway requires further investigation.

The high conformational flexibility of *Sj*GlcNK (Figure 2.8F) is a noteworthy observation. This opening/closing movement has been described as essential for ePKs such as PKA to carry out their catalytic cycle, with ATP binding being the trigger that induces lobe closure (Cheng et al., 1998; Hayward, 2004). Likewise, small-angle X-ray scattering (SAXS) experiments demonstrated that adding ATP to *Sj*GlcNK markedly increases the proportion of closed conformations of the enzyme in solution (Manso et al., 2019), suggesting a similar ATP-induced closing movement. However, no evidence of such lobe movements has been found for ELKs closely related to GlcNK, including Mak (Fraga et al., 2015), APHs (Burk et al., 2001; Young et al., 2009) and MTR kinases (Ku et al., 2007). For these ELKs, crystal structures of the apo and nucleotide-bound enzymes consistently show minor differences in the relative positions of both lobes, suggesting conformational rigidity. This has often been ascribed to few direct contacts between the adenosine moiety and residues of the C-lobe (Fraga et al., 2015; Ku et al., 2007; Young et al., 2009). The single polar contact between the ribosyl moiety and *Sj*GlcNK is through D193, located at the interface between the C-lobe and the linker, and is analogous to the equivalent glutamate residue in Mak (Figure 2.4, PDB entry 4U98) and to the glutamate considered to contribute to lobe closure in PKA (PDB entry 4XW4, Hayward, 2004). These residues are positioned in such a way as to possibly interact with both ribosyl hydroxyl groups. In

APH(3′)-IIIa, a serine residue from the C-lobe hydrogen bonds with one ribosyl hydroxyl group (PDB entry 1J7U) and in MTR kinase (PDB entry 2PUP) and APH(2″)-IIa (PDB entry 3HAV) no polar contacts are observed. The actual determinants of conformational flexibility remain to be confirmed, but our results go against the prevailing notion that ELKs display negligible lobe movement.

The K_m of 8 ± 1 mM towards GlcN estimated for *Sj*GlcN (Figure 2.5C) is relatively high, which might raise questions about whether GlcN really is the physiological substrate of this kinase. However, it should be noted that it is similar to the K_m of *Mv*Mak for maltose, 7.3 ± 0.9 mM (Fraga et al., 2015), although higher than in *Mt*Mak, 2.5 ± 0.4 mM (Mendes et al., 2010). Furthermore, it has been noted that enzymes involved in secondary metabolism, which we propose to be the case for *Sj*GlcN (Figure 2.2), are often observed to display kinetic parameters indicative of a lower catalytic efficiency compared to enzymes involved in primary metabolism (Bar-Even et al., 2011). In any case, conditions *in vivo* can be very different from *in vitro* experiments with recombinant enzymes, especially when multienzyme complexes and other regulatory mechanisms are involved. While it cannot be ruled out at this point that a rare sugar that was not part of our screening (Table 2.1) may be a better substrate for this kinase, the evidence from structural and mutational studies supports its assignment as a GlcNK. Indeed, the identification of the conserved residue Q405, which interacts with the amino group of GlcN, unveiled the main molecular determinant conferring specificity to GlcNK (Figure 2.11). Additionally, the consensus sequence ^{405}Q-x(2)-RE-x(2)-YA-x(3)-LP-x-W^{420} found in helix α12 was identified as a signature of this family of glucosamine kinases, based on structural evidence and analyses of sequence conservation across a large number of sequences from actinomycetes containing a *mak* paralogue in their genomes.

The orthologue from *M. smegmatis*, previously shown to phosphorylate glucose with a K_m over 70 mM (Nunes-Costa, 2015), contains this consensus sequence and indeed, we confirmed that it also phosphorylates GlcN, albeit not as efficiently as *Sj*GlcNK (Figure 2.5B). Intriguingly, *Ms*GlcNK displays uncoupled ATPase activity while *Sj*GlcNK does not (Figure 2.7). Transfer of phosphate from ATP to water by kinases of the PKL family has been described before (Ahuja et al., 2017; Kim et al., 2006; Kovalevsky et al., 2012). For some APHs, this uncoupled ATPase activity is not only non-negligible *in vitro*, but it has been demonstrated to represent a burden *in vivo* for bacteria expressing these enzymes constitutively (Kim et al., 2006). In the presence of aminoglycoside antibiotics, bacteria benefit from the resistance conferred by these antibiotic-inactivating kinases and absorb

the cost of increased ATP turnover, while in their absence there is a stronger selective pressure to lose the plasmid containing the gene compared to a similar situation in which the resistance enzyme does not consume ATP superfluously. This example illustrates how enzymes with less than optimal and even seemingly deleterious properties can be kept by organisms if they confer specific selective advantages under certain conditions. In *M. smegmatis*, *glcNK* is found in the same operon as a gene encoding a putative Tet repressor protein (Figure 2.2B) and based on the observed genomic organisation, it is predicted to be co-transcribed with and to regulate its adjacent genes (Cuthbertson & Nodwell, 2013). This suggests that *Ms*GlcNK may be under transcriptional repression, potentially a strategy that would mitigate the effect of its intrinsic ATPase activity on the organism's fitness. Could this ATPase activity and its enhancement by zinc ions have a non-obvious physiological purpose or is it merely a trivial property of this specific enzyme? Further studies would be required to begin tackling this question, including structural analyses of *Ms*GlcNK, particularly in complex with zinc, the characterisation of other mycobacterial homologues, and the identification of putative GlcNK interactors. These future studies could also help answering the question of why *Sj*GlcNK is unable to efficiently transfer phosphate from ATP to water like *Ms*GlcNK (Figure 2.7C). Careful analysis of the residues involved in catalysis and lining the nucleotide binding pocket revealed few amino acid substitutions whose main effect appears to be on the electrostatic potential of the ATP binding pocket, less hydrophobic in *Ms*GlcNK (Figure 2.10). However, small differences in non-obvious residues distant from the active site can have a major influence in catalysis, as demonstrated by a PKA mutant in which a single amino acid substitution far from the catalytic site and causing no conformational changes to the protein's structure turned out to have a profound effect on catalysis, disrupting the transfer of phosphate from ATP to the substrate (kinase activity) but not to water (ATPase activity) (Ahuja et al., 2017).

Glucosamine kinase activity (EC 2.7.1.8) seems to have evolved independently at least three times, from unrelated fold superfamilies. Based on their sequences, the proteobacterial GlcN-specific kinase identified in *Vibrio cholerae* belongs to the ASKHA (acetate and sugar kinase/hsp70/actin) superfamily and is thus predicted to adopt the ribonuclease H-like fold (Cheek et al., 2002; Park et al., 2002), while the one found in the archaeon *Thermococcus kodakarensis* is an ADP-dependent kinase belonging to the ribokinase superfamily and predicted to adopt a Rossman-like fold (Aslam et al., 2018; Cheek et al., 2002). Neither is related to *Sj*GlcNK, shown here to be a member of the PKL superfamily. In both *V. cholerae* and *T. kodakarensis*, these GlcN kinases were implicated

in the chitin degradation pathway of each organism. Chitin is an ubiquitous polymer of GlcNAc units which can be partly deacetylated by the action of extracellular chitin deacetylases produced by some organisms, yielding segments rich in GlcN residues (chitosan) (Zhao et al., 2010). GlcNAc and GlcN-rich segments are substrates for extracellular chitinases and chitosanases, respectively, and the resulting oligo, di and monosaccharides can be taken up by microorganisms and further hydrolysed and phosphorylated for use as sources of energy, carbon and nitrogen (Aslam et al., 2018; Park et al., 2002; Rigali et al., 2006; Viens et al., 2015). In actinomycetes, generally known as good chitin and chitosan decomposers, a highly conserved transcriptional regulator responsive to the presence of chitosan oligosaccharides and D-glucosamine (CsnR) has been identified, as well as an ABC transporter (CsnEFG) responsible for the uptake of these compounds (Dubeau et al., 2011; Lacombe-Harvey et al., 2018; Viens et al., 2015). A possible fate for the resulting intracellular GlcN monomers in some actinomycetes might be phosphorylation by the unique GlcNK described in this work, the first actinobacterial enzyme found to display this activity. Since this GlcNK has a limited and scattered distribution among actinomycetes, while CsnR and CsnEFG are widespread, an alternative route for GlcN phosphorylation is expected to exist especially in species lacking a GlcNK homologue. A likely candidate is present in the operon containing *csnR*, annotated as a putative sugar kinase (*csnK*) of the ASKHA superfamily but only very distantly related to the *Vibrio* enzyme (15% amino acid sequence identity).

Experimental evidence to elucidate the physiological role of this novel family of kinases is still lacking, but some hypothesis can be postulated based on bioinformatic analyses. Although GlcNK activity, previously unreported in actinobacteria, potentially fills a gap in the chitin/chitosan degradation pathway of these organisms, its genomic context (Figure 2.2), rareness and sparse phylogenetic distribution (Table 2.2) suggest a different role, linked to secondary metabolism or the biosynthesis of species-specific molecules. Several lines of evidence support this hypothesis. First, the close proximity of a glycosyltransferase gene, co-occurring as a neighbour in 90% of the genomes analysed and found in the same operon, hints at a biosynthetic rather than catabolic gene cluster (Figure 2.2). This GT belongs to the GT1 family, a group which in bacteria is involved in the glycosylation of diverse chemical scaffolds such as polyketides, macrolides and glycolipopeptides, playing important roles in the biosynthesis of numerous natural products including antibiotics (Zhang et al., 2020). Interestingly, the gene cluster of *S. jiangxiensis* also harbours genes encoding a putative PGM, predicted to transfer a

phosphate group from carbon 6 to carbon 1 of a phosphorylated hexose, and a putative UDPGP, predicted to synthesise a sugar nucleotide from the PGM product that could be a substrate for a GT (Figure 2.12A). It is therefore tempting to speculate that the GT transfers an activated amino sugar to a specific scaffold. This sugar moiety might not be GlcN, since the putative sugar isomerase present in the operon could convert it into a different sugar. The phosphosugar recognition domain of this putative isomerase shows homology with the one found in GlmS, suggesting that GlcN6P, the product of GlcNK, could be its substrate. The second line of evidence linking GlcNK to secondary metabolism is the presence of genes putatively involved in the biosynthesis of natural product scaffolds that often serve as substrates for GTs of the GT1 family, which causes the cluster containing GlcNK to be detected by antiSMASH as a BGC in many genomes. In *S. jiangxiensis*, this scaffold would be a cyclodipeptide formed by the putative CDPS and P450 enzymes (Skinnider et al., 2018), while in mycobacterial genomes it would be a lipidic or peptidolipidic scaffold synthesised by the putative PKS and NRPS modules found in these clusters (Figure 2.2B). Unlike streptomycetes, mycobacteria are not known for their rich secondary metabolism and antibiotic production capabilities. Nevertheless, mycobacteria synthesise many species-specific unique molecules, particularly cell envelope components such as glycopeptidolipids and lipooligosaccharides, whose scaffolds are assembled by PKS/NRPS modules to which specific GTs transfer sugar residues (Doroghazi & Metcalf, 2013; Etienne et al., 2009; Tatham et al., 2012). Finally, the phylogenetic distribution pattern of GlcNK is itself indicative of a role in secondary metabolism, since while genes involved in central metabolic pathways are often deeply conserved within a phylogenetic group, secondary metabolic pathways show a more divergent distribution, often being species or strain-specific (Choudoir et al., 2018; Doroghazi & Metcalf, 2013). Additionally, it should be noted that the product of *Sj*GlcNK, GlcN6P, is an elicitor of SM biosynthesis through its direct interaction with the transcriptional repressor DasR (Rigali et al., 2006), potentially implicating the activity of this novel kinase in controlling the secondary metabolism of *S. jiangxiensis* at a higher-order level (Figure 2.12B).

The presence of kinases in BGCs is often associated with self-resistance mechanisms, since phosphorylation is a common strategy to produce inactive forms of antibiotics and prevent autotoxicity, but they can also be involved in catalysing intermediate steps within the biosynthetic pathways of some SM (Draelos et al., 2021). The GlcNK identified in this work may allow the incorporation of GlcN from extracellular material (chitin/chitosan) into the secondary metabolism of some actinomycetes. Particularly, in *S. jiangxiensis* we

propose that GlcNK is part of a biosynthetic pathway for a glycosylated diketopiperazine (DKP) (Figure 2.12B). DKPs are structurally diverse molecules with a wide range of biological activities of pharmacological and industrial interest, including antibacterial, antifungal, antiviral, immunosuppressant, herbicidal, anticancer and neuroprotective effects (Borgman et al., 2019; Cornacchia et al., 2012). The presence of glycosyltransferases in BGCs containing a CDPS gene is a rare occurrence (Skinnider et al., 2018), of which the *Sj*GlcNK cluster is an example, and no pathways associated with this type of BGC organisation have been characterised. In fact, until very recently no natural glycosylated DKPs had been reported, even though they had been synthetised in the scope of the development of novel antiparasitic drugs (Martins-Teixeira et al., 2013). To the best of our knowledge there has been a single report of a natural glycosylated DKP to date, a compound with broad antimicrobial activity isolated from a marine *Streptomyces* strain whose genome has not been sequenced (Chen et al., 2018).

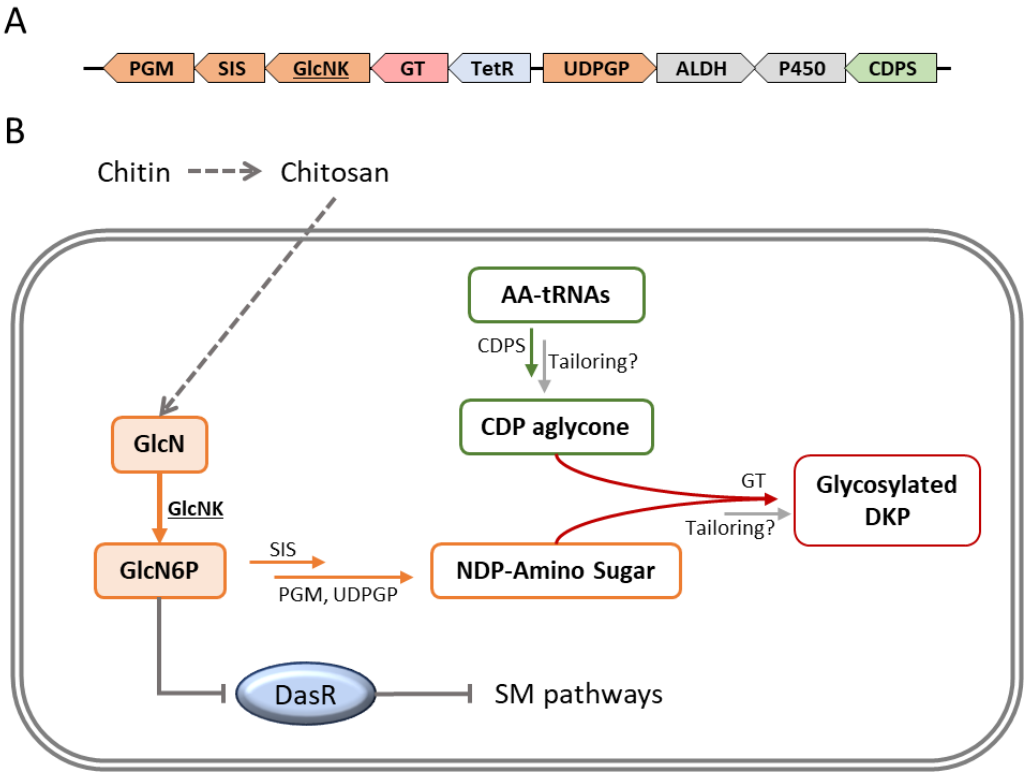


Figure 2.12. Putative role of GlcNK in *Streptacidiphilus jiangxiensis*. **(A)** *Sj*GlcNK biosynthetic gene cluster. PGM, phosphosugar mutase; SIS, phosphosugar isomerase; GT, glycosyltransferase; TetR, TetR-family of transcriptional regulators; UDPGP, sugar-phosphate nucleotydyltransferase; ALDH, aldehyde dehydrogenase; P450, cytochrome P450; CDPS, cyclodipeptide synthase (all putative functions based on manual sequence homology analyses). **(B)** Predicted pathway for the biosynthesis of a glycosylated diketopiperazine (DKP) based on putative gene annotations. GlcN originating from the catabolism and uptake of chitin/chitosan is phosphorylated by *Sj*GlcNK yielding GlcN6P, which binds the DasR receptor relieving the repression of BGCs involved in secondary metabolism pathways. Concomitantly, GlcN6P might be used as a precursor for the biosynthesis of a novel glycosylated DKP, whose scaffold is likely a cyclodipeptide aglycone synthetised by a CDPS enzyme from aminoacyl-tRNAs (AA-tRNAs). ALDH and P450 might be involved in tailoring reactions at undefined steps of the pathway.

In conclusion, the results presented here i) expand the array of substrates known to be phosphorylated by ELKs and reveal a level of conformational flexibility and interdomain movement previously unreported in members of this superfamily; ii) represent the first report of a GlcN-specific kinase in actinobacteria and identify a signature sequence motif that can be used for the reannotation of a large number of sequences into this novel ELK family; and iii) shed light into a BGC predicted to be responsible for the biosynthesis of a novel and rare natural product in an underexplored streptomycete, with potentially interesting biological activities.

Acknowledgements and work contributions

All crystallography procedures (with corresponding data collection and processing, structure determination and figure preparation) were performed by José A. Manso, Sandra de Macedo-Ribeiro and Pedro J. B. Pereira at Instituto de Investigação e Inovação em Saúde (i3S, Porto), with the exception of Figure 2.10. DNA manipulations with the QuikChange method (Stratagene) and determination of kinetic parameters with the ADP-Glo kinase assay (Promega) were performed by José A. Manso. NMR data was acquired and interpreted by Pedro Lamosa from CERMAX, ITQB-NOVA (Oeiras, Portugal).

The work in this chapter was supported by Fundo Europeu de Desenvolvimento Regional (FEDER) through the COMPETE 2020-Operational Programme for Competitiveness and Internationalisation (POCI) and by Portuguese funds through Fundação para a Ciência e a Tecnologia (FCT) in the framework of projects PTDC/BTM-TEC/29221/2017 (POCI-01-0145-FEDER-029221), project Institute for Research and Innovation in Health Sciences (POCI-01-0145-FEDER-007274), and the Structured Program on Bioengineered Therapies for Infectious Diseases and Tissue Regeneration (Norte-01-0145-FEDER-000012), funded by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement.

Chapter 3 – Structural insights into an essential enzyme afforded by the highly thermostable GpgS from the genuine mycobacterial thermophile *Mycobacterium hassiacum*

Part of this chapter is published in: Alarico S*, Nunes-Costa D*, Silva A, Costa M, Macedo-Ribeiro S, Empadinhas N. 2020. A genuine mycobacterial thermophile: *Mycobacterium hassiacum* growth, survival and Gpgs stability at near-pasteurization temperatures. Microbiology 166(5):474-483. https://doi.org/10.1099/mic.0.000898.

A paper describing the structural studies of GpgS from *M. hassiacum*, authored by Nunes-Costa D*, Silva A*, Manso JA, Alarico S, Empadinhas N, Pereira PJB, Macedo-Ribeiro S, is in final preparation to be submitted for publication soon.

*These authors contributed equally to this work.

Chapter 3 – Table of Contents

Abstract

Mycobacterium hassiacum is so far the most thermophilic among mycobacteria as it grows optimally at 50°C and up to 65°C. Since this and other NTM thrive in diverse natural and artificial environments, from where they may access and infect humans, we deemed essential to probe *M. hassiacum* resistance to heat, a strategy routinely used to control microbial growth in water-supply systems, as well as in the food and drink industries. In addition to possibly being a threat in its own right in rare occasions, *M. hassiacum* is also a good surrogate for studying other mycobacterial species more often associated with infection. First, this thermophilic species is an ideal surrogate to validate thermal disinfection strategies, which should always take into account the unusual resistance of NTM and the threat they pose to human health. Secondly, *M. hassiacum* is predicted to be a source of stable proteins that may provide more detailed structures of potential drug targets. Here, we investigate *M. hassiacum* growth at near-pasteurization temperatures and at different pHs, demonstrating that it is indeed a genuine thermophile growing optimally under slightly alkaline conditions and able to withstand exposure to 70°C for several minutes. We also characterise its thermostable glucosyl-3-phosphoglycerate synthase (GpgS), an enzyme considered essential for *M. tuberculosis* growth and associated with both nitrogen starvation and thermal stress in different NTM species. We confirm the potential of *M. hassiacum* as a source of enzymes amenable to structural studies, since its GpgS orthologue yielded not only the first mycobacterial GpgS X-ray three-dimensional structure at near atomic resolution (1.11 Å), but also several high-resolution structures under different pH conditions and in complex with different ligands. The results presented here provide new insights into the properties of this essential enzyme, hint at possible intersections between distinct metabolic pathways in mycobacteria, and explore new possibilities for the development of GpgS inhibitors.

Introduction

Members of the family *Mycobacteriaceae* share characteristics which grant them an unusual resilience to stress, even though each has evolved to adapt to strikingly different environments and lifestyles (Cole et al., 2001; Falkinham, 2009a; Galagan, 2014). This group includes obligate human pathogens such as *Mycobacterium tuberculosis* and *Mycobacterium leprae*, poikilotherm pathogens such as *Mycobacterium marinum* and environmental saprophytes ranging from psychrotolerant species able to grow at 4°C

(Santos et al., 2015; Trujillo et al., 2004) to genuine thermophiles growing optimally at 50°C and up to 65°C (Santos et al., 2007, 2015; Schröder et al., 1997). Some environmental mycobacteria, commonly called nontuberculous mycobacteria (NTM), are thermotolerant species growing optimally around 45°C such as *Mycobacterium xenopi*, *Mycobacterium thermoresistibile* and *Mycobacterium phlei*, many of which can opportunistically infect humans, with *M. xenopi* being a particularly important etiological agent of NTM lung disease (Griffith et al., 2007; Karnam et al., 2011). Routine strategies to inactivate pathogens from foodstuffs and from water distribution systems often resort to high-temperature treatments, namely pasteurization or superheat-and-flush methods that involve raising hot water temperature to 71–77 °C, so that temperature at the outlets reaches approx. 65 °C, and flushing at this temperature for 10–30 min (Whiley et al., 2017). Thus, the ability of some NTM to survive high temperatures is a probable factor underlying their ubiquity in artificial water systems (Gebert et al., 2018), which in turn are a source of both nosocomial and community-acquired NTM infections and outbreaks (Brown-Elliott et al., 2011; Falkinham, 2011; Schulze-Robbecke & Buchholtz, 1992; Wagner et al., 2019).

Mycobacterium hassiacum is, to date, the most thermophilic among the known mycobacteria and appears to be capable of colonising the human body and, although infrequently, of causing human disease (Deinhardt-Emmer et al., 2018; Jiang et al., 2013; Salzer et al., 2020; Schröder et al., 1997; Tortoli et al., 1998). The regular detection of *M. hassiacum* in water distribution systems provides ample opportunity for human contact (Gebert et al., 2018). *M. hassiacum* is able to grow at temperatures between 30 and 65 °C, anticipating a higher thermotolerant profile than most known NTM and raising questions about the adequacy of standard thermal inactivation strategies. At a time when NTM infections are rising (Johansen et al., 2020; Wagner et al., 2019), efforts to understand their lifestyle and physiology as well as the pathways underlying their superior resistance to stress are urgent. In this context, the deeper our knowledge of the key enzymes of essential pathways crucial for their unusual resilience, the closer we will be to developing better strategies to fight mycobacterial pathogens.

Biosynthesis of the glycoside glucosylglycerate (GG) has been identified as a key aspect of mycobacterial physiology contributing to their survival and ability to adapt to abiotic stress (Nunes-Costa et al., 2017). In the NTM species *Mycobacterium smegmatis* and *M. hassiacum*, GG accumulation has been shown to be involved in the adaptation to nitrogen scarcity, a stress often encountered in natural and artificial oligotrophic environments, such as water distribution systems, where NTM are known to thrive

(Alarico et al., 2014; Behrends et al., 2012). GG is also the precursor of the unique methylglucose lipopolysaccharides (MGLP), proposed to modulate fatty acid metabolism and the assembly of the mycobacterial cell envelope (Mendes et al., 2012), as well as implicated in adaptation to thermal stress given the inability of mutants deficient in MGLP production to grow at supraoptimal temperatures (Kaur et al., 2009; Stadthagen et al., 2007). The first step of GG biosynthesis in mycobacteria is catalysed by GpgS, whose proposed essentiality for *M. tuberculosis* growth (Minato et al., 2019) has earned it a lot of attention from structural biologists (Albesa-Jové et al., 2015; Albesa-Jové, Romero-García, et al., 2017; Fulton et al., 2008; Kumar et al., 2014; Pereira et al., 2008; Urresti et al., 2012) since it was first identified and functionally characterised (Empadinhas et al., 2008).

GpgS is a glycosyltransferase (GT) catalysing the transfer of glucose from UDP-glucose (UDP-Glc), its preferred donor, to 3-phosphoglycerate (3-PGA), forming the GG precursor glucosyl-3-phosphoglycerate (GPG). GTs fulfil myriad functions across all domains of life, being involved in crucial biosynthetic, signalling and detoxification pathways. They can be classified according to sequence, structural fold and catalytic mechanism. As of November 2021, the carbohydrate-active enzyme (CAZy) database (http://www.cazy.org) listed 114 different families (GT1 to GT114) to which characterised GTs have been assigned based on sequence similarity and biochemical activity, reflecting the vast diversity of substrates that can be utilised and types of glycosidic linkages that can be created by this large class of enzymes. In contrast, this sequence diversity is supported by a relatively small number of structural folds, with the majority of characterised GTs being classified into two main folds, GT-A and GT-B, containing one or two Rossmann fold domains, respectively (Lairson et al., 2008). Only a few other GT folds have been identified, mainly in membrane-associated GTs using donors other than nucleotide-activated sugars (Albuquerque-Wendt et al., 2019; Kattke et al., 2019; Yuan et al., 2007; Zhang et al., 2014). According to catalytic mechanism, GTs can be classified into "inverting" or "retaining" depending on whether the stereochemistry of the anomeric carbon is inverted or retained in the glycosidic product relative to the sugar donor, a property that is generally conserved among GTs belonging to the same family but is not correlated with the structural fold (Lairson et al., 2008). Mycobacterial GpgSs are retaining enzymes displaying a GT-A fold and have been classified into the GT81 family, present only in prokaryotes and initially created to accommodate the distantly related GpgS from the archaeon *Methanococcoides burtonii* and its homologues, and which also includes the ambiguous mannosyl-3-phosphoglycerate synthase (MpgS) from *Rubrobacter*

xylanophilus capable of synthesising both GG and the related glycoside mannosylglycerate (MG) (Costa et al., 2006; Empadinhas et al., 2008, 2011; Taujale et al., 2020). To date, the structures of three members of this family have been reported, namely the *R. xylanophilus* MpgS/GpgS (Empadinhas et al., 2011) and two mycobacterial GpgSs, from *M. avium* subsp. *paratuberculosis* (Fulton et al., 2008) and from *M. tuberculosis* (Albesa-Jové et al., 2015; Albesa-Jové, Romero-García, et al., 2017; Pereira et al., 2008; Urresti et al., 2012).

These structural studies have confirmed that GpgS possesses the typical GT-A architecture, with an N-terminal nucleotide-binding subdomain displaying the Rossman-like fold, a C-terminal acceptor-binding subdomain, and the signature catalytic DXD motif, characteristic of metal-dependent GTs, found at the juncture of both subdomains and involved in divalent metal ion coordination (Figure 3.1) (Fulton et al., 2008; Pereira et al., 2008; Taujale et al., 2020). Although some GTs display high intrinsic flexibility and undergo substantial conformational changes and inter-domain movements upon substrate binding (Albesa-Jové & Guerin, 2016; Rodrigo-Unzueta et al., 2020), GpgS seems to display a mostly rigid structure with preformed substrate-binding sites, possibly existing in a state very close to the active conformation even in its unliganded form (Fulton et al., 2008; Pereira et al., 2008; Urresti et al., 2012). The exceptions are two flexible loops (Figure 3.1) directly involved in substrate binding and catalysis which seem to adopt different conformations in the presence or absence of ligands. One of these loops, named the *L* loop, is located in the vicinity of the nucleotide-binding site and contains an important and widely conserved histidine residue involved in metal coordination, as well as residues which interact with the donor substrate UDP-Glc (Albesa-Jové, Romero-García, et al., 2017; Fulton et al., 2008; Taujale et al., 2020). This loop has been observed to adopt "open/inactive" and "closed/active" conformations in crystals and has been proposed to play a critical role in the assembly of the active form of GpgS and its catalytic cycle (Albesa-Jové, Romero-García, et al., 2017). A second flexible loop (loop 165-184) located in the vicinity of the acceptor substrate binding site shows evidence of undergoing conformational changes during the catalytic cycle and has been proposed to play a key role in facilitating product release (Albesa-Jové, Romero-García, et al., 2017). Indeed, both loops are also found in many other GT-A enzymes where they have been associated with similar functional roles (Qasba et al., 2005; Romero-García et al., 2013). The importance of short dynamic loops for GT activity has been long recognised along with the observation that most GTs display a high loop to secondary structure ratio, an attribute that also holds true for GpgS (Albesa-Jové & Guerin, 2016; Breton et al., 2006; Pereira et al., 2008).

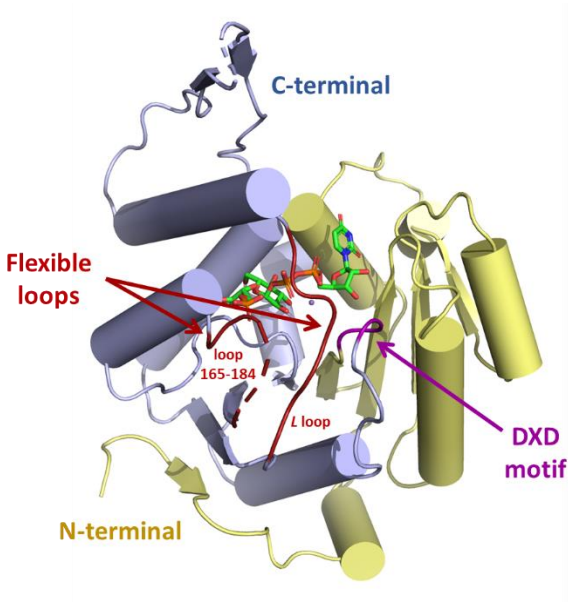


Figure 3.1. Cartoon representation of the overall three-dimensional structure of *M. tuberculosis* GpgS in complex with UDP-glucose and 3-phosphoglycerate, solved to 2.35 Å resolution (PDB accession code 4Y6N) (Albesa-Jové et al., 2015). The N-terminal and C-terminal domains are coloured yellow and blue, respectively, and the DXD motif bridging both domains is coloured purple. The two flexible loops involved in substrate binding and catalysis are coloured red. The substrates are shown as sticks and coloured with oxygen atoms red, nitrogen blue, phosphorous orange and carbon green. The divalent metal ion, manganese, is shown as a violet sphere.

The reaction mechanism for retaining GTs was initially assumed to be analogous to the one employed by retaining glycosidases, which utilise a double-displacement mechanism with transient formation of a covalent glycosyl-enzyme intermediate and in which net retention of the anomeric configuration is achieved through two consecutive inverting steps (Cerqueira et al., 2012). However, the lack of a suitable enzyme nucleophile in the active site of most retaining GTs that could act as an intermediate acceptor suggested that these enzymes must follow a different mechanism (Moremen & Haltiwanger, 2019). Indeed, several recent X-ray structures of native retaining GTs in complex with their substrates as well as QM/MM (quantum-mechanics/molecular-mechanics) studies support a S_Ni-like mechanism, with the β-phosphate of the donor substrate acting as the catalytic base which deprotonates the acceptor hydroxyl group and facilitates a same-sided nucleophilic attack of the anomeric carbon, resulting in retention of the anomeric configuration (Albesa-Jové, Sainz-Polo, et al., 2017; Lira-Navarrete et al., 2014; Yan & Liu, 2020; Yu et al., 2015). This mechanism has also been proposed for *M. tuberculosis* GpgS (Albesa-Jové et al., 2015). However, all structures of GpgS-substrate complexes reported to date, and which have been used to infer details of its catalytic mechanism and loop dynamics, were solved at resolutions between 2-3 Å, with the associated limitations in determining atom positions with confidence from the

experimental data. Detailed structures at full-atomic resolution could provide additional mechanistic information and serve as better scaffolds for rational drug design efforts.

In this work, we confirm that *M. hassiacum* grows optimally at 50°C, up to 65°C and retains viability at near-pasteurization temperatures, making it the most thermophilic mycobacterial species described to date and a unique case of a thermophile that is able to colonise and, although very rarely, infect humans. We have also characterised its highly thermostable GpgS, sharing 73% amino acid identity with the *M. tuberculosis* homologue, and therefore being a potential surrogate to study the drugability of this essential enzyme in the urgent quest for novel antimycobacterial compounds (Baugh et al., 2015). Based on its properties, we predicted the *M. hassiacum* homologue would be more amenable to crystallisation studies and, indeed, we were able to solve the three-dimensional structure of a mycobacterial GpgS at near atomic resolution for the first time. Due to its high expression levels, easy purification and ability to readily yield well-diffracting crystals, it was possible to carry out a comprehensive comparative study of the effect of pH, magnesium concentration and ligand binding on the active site and flexible loops of GpgS. Furthermore, the detection of an intriguing molecule at the enzyme's active site and a thorough analysis of its genomic context hinted at a possible crosstalk between MGLP biosynthesis and other crucial metabolic pathways in mycobacteria, opening new and unexpected avenues of research.

Methods

M. hassiacum growth conditions and thermotolerance

Mycobacterium hassiacum DSM 44199 was grown at 50°C, pH 7.2 on GPHF medium plates (DSM 553 medium (Schröder et al., 1997)). For the determination of growth rates at different temperatures and pH, cells were cultivated in metal-capped flasks, containing 200 mL of GPHF medium supplemented with 0.2% (w/v) Tween 80, with or without 0.8% (w/v) glycerol, and stirred at 150 rpm with continuous aeration. Growth was monitored at 610 nm for several days at 50°C in media at pH 5.0, 6.0, 7.2 or 8.0 and at temperatures of 37, 50, 60 or 65°C, at pH 8.0.

Thermotolerance was examined by exposing to different temperatures an *M. hassiacum* cell suspension prepared in either rich medium (GPHF) or water of standardised hardness (WSH, prepared as described previously (Schulze-Robbecke & Buchholtz, 1992)). For the rich medium experiment, the suspension was obtained by

cultivating cells in glycerol-supplemented GPHF medium (pH 8.0) at 50°C to an OD_{610} of 0.5-0.7 and diluting to 0.2 in the same medium. For the water experiment, cells were incubated in GPHF medium plates for 48 hours after which colony material was transferred to a tube containing 5 ml of WSH and three glass beads 4 mm in diameter and vortexed for 3 min. The suspension was allowed to settle for 30 min to remove the remaining cell aggregates and the supernatant was carefully transferred to another tube, diluted to an OD_{610} of 0.2 and starved for 24 hours at 22°C with gentle shaking (protocol adapted from Schulze-Robbecke & Buchholtz, 1992). GPHF and WSH cell suspensions were distributed through 150 µl aliquots, which were further incubated at temperatures of 50, 60, 65 and 70°C in a water bath, for 0, 0.5, 1, 2, 5, 10, 15, 30 and 60 min. Duplicate samples were placed on ice at each time point. Culturability was evaluated by spotting 10 µl of each heat-treated sample on GPHF agar followed by incubation of the plates at 50°C for 5 days. Metabolic viability was measured by incubating 100 µl of heat-treated sample with 10 µl of AlamarBlue reagent kit (GeneCopoeia) for 3 hours at 37°C and measuring absorbance at 570 nm and 600 nm.

Cloning, heterologous expression and purification of the *M. leprae* GpgS (*Ml*GpgS)

The *M. leprae gpgS* nucleotide sequence was retrieved from the Mycobrowser database (https://mycobrowser.epfl.ch/genes/ML1064), optimised for expression in *E. coli* and synthetised (GenScript) with the insertion of a *Nde*I restriction site at the 5' terminal and replacement of the native stop codon by a Met codon followed by a *Sap*I restriction site at the 3' terminal. The restriction sites were included to allow cloning into the pTXB1 expression vector and fusion with a C-terminal self-cleavable intein tag containing a chitin-binding domain (intein-CBD tag), for purification with the IMPACT system (New England BioLabs). The Met codon was added to encode a terminal methionine residue between *Ml*GpgS and the intein tag to improve controllable intein self-cleavage during purification. The plasmid containing this synthetic gene (pUC57-*gpgS*) and the pTXB1 vector were both incubated with *Nde*I and *Sap*I restriction enzymes (New England BioLabs). The pUC57-*gpgS* digestion mixture was analysed by agarose gel electrophoresis, the band corresponding to the *gpgS* gene (~1000 bp) was excised from the gel, and the DNA was purified from agarose with the GeneJET Gel Extraction Kit (Thermo Scientific). The linearised pTXB1 vector was purified with the JETquick PCR Product Purification Spin Kit (Genomed) and further dephosphorylated with Alkaline Phosphatase Calf Intestinal (New England BioLabs), to prevent self-ligation, and purified again with the same kit. The digested gene and the linearised and dephosphorylated vector were ligated with Anza™

T4 DNA Ligase Master Mix (Invitrogen) and transformed into chemically competent *E. coli* DH5α cells using standard procedures. Recombinant plasmids were isolated from kanamycin-resistant colonies with the ZR Plasmid Miniprep™-Classic Kit (Zymo Research) and sequenced (GATC Biotech) to confirm the identity of the inserts. Constructs confirmed to carry a copy of the synthetic *M. leprae gpgS* gene were subsequently transformed into chemically competent *E. coli* BL21(DE3) for protein expression.

An *E. coli* BL21(DE3) clone carrying the recombinant pTXB1-*gpgS* plasmid was grown in LB medium containing ampicillin (100 µg/mL) at 37°C in an orbital shaker at 120 rpm. Upon reaching mid-exponential phase (OD_{610}≈0.8), growth temperature was gradually decreased to 20°C and gene expression was induced by adding IPTG to a final concentration of 0.5 mM. Cells were harvested 17 h later by centrifugation (9000 × *g*, 10 min, 4°C), resuspended in 20 mM Tris-HCl pH 8.5 with 0.5 M NaCl (Column Buffer) and frozen at -20°C. Before purification, cells were thawed in the presence of DNase I (10 µg/mL) and disrupted by sonication on ice with five 60 Hz pulses of 45 s (15 s interval between pulses) per each 5 mL of cell suspension, with care to not let the suspension overheat. The lysate was centrifuged (15000 x *g*, 30 min, 4°C) to remove cell debris, filtered (0.45 µm syringe filter) and loaded onto a chitin column equilibrated with Column Buffer. After washing the column with Column Buffer to remove unbound proteins, on-column cleavage of the intein-CBD tag was induced by quickly flushing the column with 3 bed volumes of Cleavage Buffer (20 mM Tris-HCl pH 8.5 with 0.5 M NaCl and 50 mM DTT). Flow was stopped and the column was incubated at 6°C overnight to allow thiol-induced intein cleavage. Following its release from the column-bound intein-CBD tag, *Ml*GpgS was eluted by restarting the flow with Column Buffer and collecting 4 mL fractions. The purity of each fraction was evaluated by SDS-PAGE and the purest fractions were pooled, diluted 10-fold in 20 mM BTP buffer pH 7.5, concentrated by ultrafiltration in 30 kDa cutoff centricons (Amicon) and equilibrated with the same buffer. Purity was evaluated by SDS-PAGE and enzyme concentration was determined by the Bradford assay (BioRad). The recombinant enzyme was dispensed in single-use aliquots, flash frozen in liquid nitrogen and stored at -80°C.

Cloning, heterologous expression, and purification of the *M. hassiacum* GpgS (*Mh*GpgS)

Genomic DNA was purified (Nzytech Microbial gDNA Isolation kit) from *M. hassiacum* cultivated in glycerol-supplemented GPHF medium as described above. The full *gpgS* gene sequence retrieved from the *M. hassiacum* genome (Tiago et al., 2012) was amplified by PCR using AccuPrime GC rich DNA polymerase (Invitrogen), with primers 5′-GCGCATATGACACTGGTACCG and 5′-ATAAGCTTGCCGCGCAGGGTG designed to introduce 5′ *Nde*I and 3′ *Hind*III restriction sites (underlined), removing the native stop codon to allow cloning into the corresponding sites of pET30a vector and translation of the vector-encoded C-terminal hexahistidine tag. The construct was sequenced (LGC genomics, Germany) and transformed into *E. coli* BL21, which was grown at 37°C in LB medium containing kanamycin (30 µg/ml) to mid-exponential phase of growth (OD_{610nm}=0.9–1). Expression was induced with 0.5 mM IPTG and growth continued for further 4 h. Cells were harvested by centrifugation (9000 x *g*, 10 min, 4 °C) and resuspended in 20 mM sodium phosphate buffer pH 7.4 with 0.5 M NaCl and 20 mM imidazole (Buffer A). Before protein purification, 10 µg/ml DNaseI was added to the suspension and cells were disrupted by sonication, followed by centrifugation (15000 x *g*, 30 min, 4 °C) to remove debris.

His-tagged recombinant *Mh*GpgS was purified from filtered lysates (0.45 µm syringe filter) in a prepacked Ni-Sepharose™ High-Performance column (HisTrap™ HP 5 mL, GE Healthcare) equilibrated with Buffer A (see above). After loading the extract onto the column, a washing step was performed by increasing the imidazole concentration to 44 mM, and elution was subsequently carried out in the same buffer with 212 mM imidazole. The purity of each fraction was evaluated by SDS-PAGE and the purest fractions were pooled, diluted 10-fold in 50 mM BTP buffer pH 7.5, concentrated by ultrafiltration in 30 kDa cutoff centricons (Amicon) and equilibrated with the same buffer. Purity was evaluated by SDS-PAGE and enzyme concentration was determined by the Bradford assay (BioRad). The recombinant enzyme was dispensed in single-use aliquots, flash frozen in liquid nitrogen and stored at -80°C. Oligomerization in solution of *Mh*GpgS was estimated by gel filtration on a Superdex 200 column previously calibrated with the molecular mass standards of albumin (67 kDa), aldolase (158 kDa), catalase (232 kDa) and ferritin (440 kDa). Blue Dextran 2000 (GE Healthcare) was used to determine the void volume.

Analysis of enzyme activity by thin-layer chromatography

GpgS activity was detected in 50 µL mixtures containing 25 mM BTP pH 7.5, 5 mM $MgCl_2$, 5 mM 3-PGA, 5 mM UDP-Glc, and pure recombinant enzyme (3 µg *Ml*GpgS or 2 µg *Mh*GpgS) incubated at 30 or 50°C for 1 h, with equivalent mixtures without enzyme used as negative controls. Potential donors other than UDP-Glc were probed, namely ADP-glucose, GDP-glucose and GDP-mannose. To visualise product formation, reaction and control mixtures were spotted on Silica 60 gel plates (Merck), 25 µL per spot, and developed with the solvent system acetic acid-ethyl acetate-water-ammonia 25% (6:6:2:1, vol/vol). The plate was dried, sprayed with an α-naphthol–based staining solution (prepared by mixing 40.5 mL ethanol, 10.5 mL α-naphthol 15% (w/v) ethanol solution, 4 mL distilled water and 6.5 mL concentrated H_2SO_4) and charred at 120°C (Jacin & Mishkin, 1965). GG and GPG were used in the TLC as standards for comparative purposes.

Enzyme assays

To quantify GPG production, a previously described (Empadinhas et al., 2011) coupled-enzyme assay was used, relying on the specific dephosphorylation of the GpgS product by the mannosyl-3-phosphoglycerate phosphatase from *Thermus thermophilus* (*Tt*MpgP) (Empadinhas et al., 2003). Reaction mixtures (50 µl) containing 25 mM BTP buffer pH 7.5, 10 mM $MgCl_2$, 2.5 mM of each substrate, 1.0 µg of *Mh*GpgS and 5.0 µg of *Tt*MpgP were incubated at 50°C and stopped after 0.5, 1, 2 or 3 min by cooling on ethanol-ice. Control mixtures prepared without the addition of *Mh*GpgS were also incubated at 50°C for 3 min. GPG formed under each condition was measured through the quantification of free phosphate released by the coupled action of *Mh*GpgS/*Tt*MpgP using the Ames method (Ames, 1966). Briefly, cooled reactions were diluted to a total volume of 300 µL with 25 mM BTP buffer pH 7.5, mixed with 700 µL of freshly prepared detection reagent (prepared by mixing 1 part ascorbic acid 10% and 6 parts 0.42% ammonium molybdate·$4H_2O$ in 1 *N* H_2SO_4) and incubated at 45°C for 20 min. Absorbance was measured in 1 mL cuvettes at 820 nm using a UV-Vis Spectrophotometer (PerkinElmer).

Biochemical characterisation of the recombinant *Mh*GpgS

The biochemical characterisation of *Mh*GpgS, including the determination of its substrate specificity, temperature, pH and cation dependence profiles and kinetic parameters was performed by Susana Alarico and is described in detail in: Alarico S*, Nunes-Costa D*, Silva A, Costa M, Macedo- Ribeiro S, Empadinhas N. 2020. A genuine

mycobacterial thermophile: *Mycobacterium hassiacum* growth, survival and Gpgs stability at near-pasteurization temperatures. *Microbiology* 166(5):474-483. *equal contributions.

*Mh*GpgS thermal stability characterisation

Thermal stability was determined by first incubating recombinant *Mh*GpgS aliquots (30 µl of a solution of 1 mg/ml) in 25 mM BTP pH 7.5 at 37 and 50°C. For thermal stability at 60°C, incubation of *Mh*GpgS aliquots was done in the presence of 5 mM of each substrate. Aliquots were withdrawn at defined times (up to 5 days) and examined for residual activity in enzyme assays as described above. Thermal stability of recombinant GpgSs from *M. smegmatis* (*Ms*GpgS) and *M. bovis* (*Mb*GpgS) was determined as previously described (Empadinhas et al., 2008).

To further characterise *Mh*GpgS thermostability, its melting temperature (Tm) was determined by a thermofluor assay (Ericsson et al., 2006) using SYPRO Orange, at different pH and in the presence or absence of its natural ligands. Briefly, *Mh*GpgS was diluted to a concentration of 0.8 mg/ml using buffers containing 50 mM NaCl and 2 mM $MgCl_2$ and whose pH was varied between 5.5 and 9.0 (0.1 mM Bis-Tris pH 5.5–6.0; 0.1 mM BTP pH 6.5–9.0), in the absence or presence of 20 mM 3-PGA, 20 mM UDP or 20 mM UDP-Glc. This thermofluor assay protocol was performed by Alexandra Silva (IBMC/i3S) and is described in detail in: Alarico S*, Nunes-Costa D*, Silva A, Costa M, Macedo- Ribeiro S, Empadinhas N. 2020. A genuine mycobacterial thermophile: *Mycobacterium hassiacum* growth, survival and Gpgs stability at near-pasteurization temperatures. *Microbiology* 166(5):474-483. *equal contributions.

Crystallisation of the recombinant *Mh*GpgS

Initial crystallisation trials were performed at 293K using the sitting-drop vapour diffusion method with the Index HT sparse matrix crystallisation kit (Hampton Research). Crystals were observed in four different conditions: condition 9 (0.1 M Bis-Tris pH 5.5, 3 M sodium chloride); condition 85 (0.2 M magnesium chloride hexahydrate, 0.1 M Tris pH 8.5, 25% (w/v) polyethylene glycol 3350); condition 86 (0.2 M potassium sodium tartrate tetrahydrate pH 7.2, 20% (w/v) polyethylene glycol 3350); and condition 91 (0.15 M DL-malic acid pH 7.1, 20% (w/v) polyethylene glycol 3350). All four conditions were used to obtain *apoMh*GpgS crystals, which grew from drops consisting of 1 µL protein solution (11.5 mg/mL in 50 mM BTP pH 8.8, 50 mM NaCl) and 1 µL crystallisation buffer, equilibrated against a 300 µL reservoir. Crystals of *Mh*GpgS in complex with UDP and UDP-Glc were obtained by co-crystallisation or soaking. For co-crystallisation, GpgS (11.5 mg/mL in 50

mM BTP pH 8.8, 50 mM NaCl) was incubated with UDP (0.2 mM) or UDP-Glc (UPG) (0.2 mM) for 1 hour on ice and then crystallised under the same conditions as the apo enzyme. For the soaking experiments, the GpgS crystals were soaked in a drop of crystallisation buffer supplemented with 0.2 mM of UDP or UPG. Prior to data collection, cryoprotection for crystals grown in conditions 85, 86 and 91 was performed through serial transfer of crystals to mother liquor containing increasing concentrations of glycerol (up to 10%) or by crystal dehydration soaking the crystals in mother liquor with slightly higher concentrations of precipitant supplemented with 0.85 M sodium malonate (condition 9). These experiments were done by Alexandra Silva (IBMC/i3S).

To obtain crystals of *Mh*GpgS in complex with benzoic-like compounds, both co-crystallisation and soaking experiments were performed. For co-crystallisation, crystals were grown from drops consisting of 1 µL protein solution (14 mg/mL *Mh*GpgS in 50 mM BTP buffer pH 7.5, pre-incubated for 1 h at 4°C with 10 mM $MgCl_2$ and either 20 mM *p*-aminobenzoic acid (*p*ABA) or 20 mM *p*-hydroxybenzaldehyde (*p*HBA)) and 1 µL crystallisation buffer, under the conditions described above. Prior to data collection, crystals were cryoprotected in the corresponding crystallisation buffer supplemented with 10% (vol/vol) glycerol and flash frozen in liquid nitrogen.

Data collection and processing

Diffraction data were collected from single crystals at 100 K on beamlines ID29 (de Sanctis et al., 2012) or ID23-2 (Flot et al., 2010) of the European Synchrotron Radiation Facility (ESRF) (Grenoble, France), or on beamline BL13-XALOC (Juanhuix et al., 2014) of the ALBA-CELLS synchrotron (Cerdanyola del Vallès, Spain). Diffraction data were processed with XDS (Kabsch, 2010) and reduced with utilities from the CCP4 software suite (Winn et al., 2011). A summary of data collection and processing statistics is presented in Table 3.1 (*Mh*GpgS apo structures), Table 3.2 (*Mh*GpgS·UDP and *Mh*GpgS·*p*ABA complexes) and Table 3.3 (*Mh*GpgS·UPG and *Mh*GpgS·*p*HBA complexes).

Table 3.1. Data collection and refinement statistics for *apoMh*GpgS structures at different pH.

Crystal	***apoMh*GpgS pH 5.5**	***apoMh*GpgS pH 7.1**	***apoMh*GpgS pH 7.2**	***apoMh*GpgS pH 8.5**
PDB code	**7P8G**	**7P5L**	**7QCP**	**7PVL**
Data collection				
X-ray beamline	ID29 (ESRF)	ID29 (ESRF)	ID29 (ESRF)	ID29 (ESRF)
Space group	I 4_1	P 2_1 2_1 2_1	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Cell constants				
a, b, c (Å)	101.4, 101.4, 122.6	71.1, 90.6, 95.3	71.3, 90.6, 95.3	71.0, 89.9, 95.6
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	0.99987	0.96112	0.96112	0.96112
Resolution range (Å)	46.6-2.13 (2.20-2.13)	50.0 - 1.22 (1.26 - 1.22)	50.0-1.11 (1.15-1.11)	50.0-1.43 (1.48-1.43)
Total/unique reflections	159768/34136 (16164/3388)	782392/182244 (63391/17739)	2977865/242579 (232191/23553)	699283/113499 (48791/10985)
Multiplicity	4.7 (4.8)	4.3 (3.6)	12.3 (9.9)	6.2 (4.4)
Completeness (%)	98.6 (98.9)	99.8 (99.7)	99.9 (99.6)	99.9 (99.7)
R_{merge}	0.050 (0.568)	0.052 (0.625)	0.066 (1.116)	0.066 (0.767)
$R_{measure}$	0.056 (0.635)	0.060 (0.734)	0.069 (1.178)	0.072 (0.871)
R_{pim}	0.024 (0.276)	0.028 (0.379)	0.019 (0.371)	0.028 (0.407)
Mean $I/\sigma I$	15.0 (1.9)	12.7 (1.9)	16.7 (2.0)	17.0 (1.8)
Refinement				
Resolution range (Å)	42.5-2.13	47.6-1.22	47.8-1.11	44.95-1.43
R_{work}/R_{free} (%)	20.5/22.9	14.3/16.4	14.2/15.6	14.8/18.1
Unique reflections, work/free	34056/1715	182144/9129	242455/12152	113413/5679
No. of:				
Non-hydrogen atoms	2404	5238	5169	5330
Amino acid residues	303	581	577	587
Water molecules	87	614	655	788
Ligands at active site	1 BGC, 1 MLI, 4 Cl^-	1 MLT, 1 LMR	2 TLA	2 Cl^-, 3 Mg^{2+}
Wilson *B* factor (Å^2)	46.0	12.2	12.2	14.0
RMSD bond length (Å)	0.007	0.009	0.006	0.005
RMSD bond angle (°)	0.90	1.01	0.91	0.74
Ramachandran plot				
Favoured (%)	98.0	98.1	98.6	98.8
Allowed (%)	2.0	1.9	1.4	1.2
Outliers (%)	0.0	0.0	0.0	0.0
Average B-factor	58.4	19.3	19.2	20.6
Macromolecules	58.5	17.9	17.3	18.0
Ligands	60.2	47.9	52.9	35.0
Solvent	54.0	29.4	31.8	35.9
Molecules in asymmetric unit	1	2	2	2

Statistics for the outermost resolution shell are shown in parentheses.

BGC – β-D-glucose, MLI – malonate, MLT – D-Malate, LMR – L-Malate, TLA – L-Tartrate.

Table 3.2. Data collection and refinement statistics for *Mh*GpgS·UDP complexes at different pH and for the *Mh*GpgS·*p*ABA complex.

Crystal	***Mh*GpgS·UDP pH 5.5**	***Mh*GpgS·UDP pH 7.2**	***Mh*GpgS·UDP pH 8.5**	***Mh*GpgS·*p*ABA pH 5.5**
PDB code	**7PDO**	**7QI9**	**7QIB**	**7PD5**
Method	Co-crystallisation	Co-crystallisation	Soaking	Soaking
Data collection				
X-ray beamline	ID29 (ESRF)	ID29 (ESRF)	ID23-2 (ESRF)	ID29 (ESRF)
Space group	I 4_1	P 2_1 2_1 2_1	P 2_1 2_1 2_1	I 4_1
Cell constants				
a, b, c (Å)	101.3, 101.3, 122.5	71.6, 90.4, 95.5	70.0, 90.0, 95.2	100.3, 100.3, 123.9
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	0.96112	0.96112	0.97934	0.97915
Resolution range (Å)	50.0-1.97 (2.04-1.97)	50.0-1.23 (1.27-1.23)	47.8-1.60 (1.68-1.60)	46.7-1.95 (2.00-1.95)
Total/unique reflections	301079/43448 (27756/4232)	786310/171306 (49576/13931)	433325/80077 (59085/11135)	201388/44461 (14143/3161)
Multiplicity	6.9 (6.6)	4.6 (3.6)	5.4 (5.3)	4.5 (4.5)
Completeness (%)	99.6 (99.0)	95.6 (80.5)	99.3 (95.7)	99.9 (99.9)
R_{merge}	0.059 (0.881)	0.040 (0.554)	0.064 (1.24)	0.035 (2.157)
$R_{measure}$	0.064 (0.957)	0.052 (0.716)	0.071 (1.38)	0.045 (2.879)
R_{pim}	0.024 (0.371)	0.024 (0.357)	0.030 (0.589)	0.021 (1.353)
$CC_{1/2}$ (%)	-	-	0.999 (0.594)	1.00 (0.419)
Mean $I/\sigma I$	17.1 (1.7)	15.1 (1.8)	13.6 (1.4)	15.5 (0.6)
Refinement				
Resolution range (Å)	46.6-1.97	48.4-1.23	47.8-1.60	46.7-1.95
R_{work}/R_{free} (%)	18.3/19.6	14.2/16.2	17.4/19.6	19.1-21.3
Unique reflections, work/free	43384/2184	171142/8578	79824/3878	44294/2177
No. of:				
Non-hydrogen atoms	2499	5159	4831	2341
Amino acid residues	301	576	577	302
Water molecules	184	618	445	82
Ligands at active site	1 UDP, 4 Cl^-	2 UDP	2 UDP, 6 Mg^{2+}, 2 Cl^-	1 *p*ABA, 1 BTB, 5 Cl^-
Wilson *B* factor (Å^2)	38.2	13.8	23.5	52.0
RMSD bond length (Å)	0.006	0.007	0.006	0.007
RMSD bond angle (°)	0.85	0.94	0.78	0.92
Ramachandran plot				
Favoured (%)	97.6	98.8	98.8	99.0
Allowed (%)	2.4	1.2	1.2	1.0
Outliers (%)	0.0	0.0	0.0	0.0
Average B-factor	51.8	21.2	31.8	68.2
Macromolecules	51.8	19.4	30.9	68.5
Ligands	51.0	27.9	34.2	70.1
Solvent	52.2	33.4	39.7	61.4
Molecules in asymmetric unit	1	2	2	1

Statistics for the outermost resolution shell are shown in parentheses.
*p*ABA – *p*-aminobenzoic acid, BTB – Bis-Tris buffer.

Table 3.3. Data collection and refinement statistics for *Mh*GpgS·UDP-Glc complexes at different pH and for the *Mh*GpgS·pHBA complex.

Crystal	***Mh*GpgS·UPG pH 5.5**	***Mh*GpgS·UPG pH 7.2**	***Mh*GpgS·UPG pH 8.5**	***Mh*GpgS·*p*HBA pH 7.1**
PDB code	**7PE4**	*	**7QOQ**	**7PHO**
Method	Co-crystallisation	Co-crystallisation	Soaking	Co-crystallisation
Data collection				
X-ray beamline	ID29 (ESRF)	ID29 (ESRF)	ID23-2 (ESRF)	BL13-XALOC (ALBA)
Space group	$I\,4_1$	$P\,2_1\,2_1\,2_1$	$P\,2_1\,2_1\,2_1$	$P\,2_1\,2_1\,2_1$
Cell constants				
a, b, c (Å)	101.9, 101.9, 122.1	71.4, 90.5, 95.3	71.1, 90.4, 96.0	70.5, 90.5, 95.3
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	0.96112	0.96112	0.87260	0.97926
Resolution range (Å)	50.0-2.05 (2.12-2.05)	50.0-1.25 (1.29-1.25)	48.3-1.93 (2.03-1.93)	65.7-1.27 (1.29-1.27)
Total/unique reflections	269589/38917 (26249/3768)	762940/168628 (54661/15671)	238444/46655 (25738/5966)	873161/161347 (40060/8041)
Multiplicity	6.9 (7.0)	4.5 (3.5)	5.1 (4.3)	5.4 (5.0)
Completeness (%)	99.8 (99.6)	98.9 (94.4)	98.1 (87.3)	99.9 (100)
R_{merge}	0.085 (0.973)	0.047 (0.590)	0.195 (1.106)	0.036 (1.115)
$R_{measure}$	0.092 (1.052)	0.053 (0.691)	0.217 (1.251)	0.040 (1.248)
R_{pim}	0.035 (0.398)	0.024 (0.352)	0.094 (0.573)	0.017 (0.552)
$CC_{1/2}$ (%)	-	-	0.989 (0.441)	1.00 (0.579)
Mean $I/\sigma I$	14.7 (1.8)	14.6 (1.8)	7.9 (1.4)	18.2 (1.1)
Refinement				
Resolution range (Å)	46.7-2.05	48.3-1.25	45.2-1.93	56.7-1.27
R_{work}/R_{free} (%)	16.0/19.2	15.5/17.3	17.6/22.7	14.5/16.4
Unique reflections, work/free	38879/1950	168531/8442	46470/2348	160726/7909
No. of:				
Non-hydrogen atoms	2740	5088	5070	5228
Amino acid residues	310	576	593	604
Water molecules	307	551	570	458
Ligands at active site	1 UPG, 1 Mg^{2+}, 3 Cl^-, 3 BCT	2 UDP, 2 TLA	2 UMP, 4 Mg^{2+}, 1 Cl^-	7 *p*HBA, 1 MLT, 2 GOL, 1 BCT
Wilson *B* factor (Å²)	35.7	13.1	19.9	16.6
RMSD bond length (Å)	0.007	0.011	0.007	0.008
RMSD bond angle (°)	0.90	1.13	0.80	0.94
Ramachandran plot				
Favoured (%)	97.7	98.6	98.1	98.2
Allowed (%)	2.3	1.4	1.7	1.8
Outliers (%)	0.0	0.0	0.2	0.0
Average B-factor	46.1	22.0	23.0	23.9
Macromolecules	45.6	20.2	22.2	22.6
Ligands	46.2	49.2	20.4	34.8
Solvent	50.5	33.2	29.5	34.7
Molecules in asymmetric unit	1	2	2	2

Statistics for the outermost resolution shell are shown in parentheses.
UPG – UDP-Glc, BCT – bicarbonate ion, TLA – L-Tartrate, *p*HBA – *p*-hydroxybenzaldehyde, MLT – D-Malate, GOL – glycerol.
* submission in preparation.

Structure determination, model building and refinement

The structure of *apoMh*GpgS at pH 7.1 was solved by molecular replacement with PHASER (McCoy et al., 2007), using the coordinates of *Mt*GpgS (PDB accession code 3E25) as a search model. Alternating cycles of model building with Coot (Emsley et al., 2010) and refinement with PHENIX (Adams et al., 2010) were performed until model completion. The refined coordinates were used as a search model to solve all the other *Mh*GpgS structures and complexes. Refined coordinates and structure factors were deposited in the Protein Data Bank (Berman et al., 2000). A summary of refinement statistics is presented in Tables 3.1-3.3.

Analysis of crystallographic structures

The models of the crystallographic structures were superposed and root mean square deviation (RMSD) values were calculated with SUPERPOSE (Krissinel & Henrick, 2004). Secondary-structure elements were identified with DSSP (Kabsch & Sander, 1983). Buried residues, the dimer interface and interactions between residues at the dimer-dimer interface were determined using the PDBePISA server (Krissinel & Henrick, 2007). Cavities and pockets were detected with Fpocket (Le Guilloux et al., 2009). The surface electrostatic potential was calculated with the APBS Electrostatics plugin for PyMOL (Jurrus et al., 2018). Hydrogen bonds, salt bridges and protein-ligand interactions were studied with Coot (Emsley et al., 2010). All figures depicting molecular models and electron density maps were created with PyMOL version 2.4.1 (Schrödinger).

Sequence and genomic context analysis

Multiple sequence alignments for selected GpgS sequences were generated with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and edited with Aline version 1.0.025. Molecular weight predictions and amino acid content were calculated with the ProtParam tool (http://web.expasy.org/protparam/).

Genomic context was analysed with the Enzyme Function Initiative's Genome Neighbourhood Tool (EFI-GNT) (Zallot et al., 2019). First, the amino acid sequence of the *MhGpgS* was used to retrieve homologous sequences from the UniProt database using a negative log of e-value threshold of 5. The output was filtered by applying a maximum sequence length restriction of 360 residues to eliminate possible multidomain proteins and an alignment score threshold of 35 was initially selected. The generated Sequence Similarity Network (SSN) was visualised and edited with Cytoscape 3.8.2. The alignment score threshold was raised iteratively to achieve sequence segregation on a phylogenetic

basis, until GpgS orthologues from members of the order *Corynebacteriales* clustered together and separately from sequences belonging to different taxa. The *Corynebacteriales* cluster, obtained after raising the alignment score threshold to 112, was used as input to generate a Genome Neighbourhood Network (GNN) with a neighbourhood size of 10 (number of neighbouring genes upstream and downstream of the query sequence) and a minimal co-occurrence percentage lower limit of 70%. The resulting GNN was visualised and edited with Cytoscape and the genomic environment for selected species was drawn based on the generated Genome Neighbourhood Diagrams (GND).

Thermal shift and inhibition assays in the presence of *p*ABA and *p*HBA

The effect of *p*ABA and *p*HBA in the melting temperature of *Mh*GpgS was evaluated by a thermofluor assay (Ericsson et al., 2006) using SYPRO Orange, an environmentally sensitive dye. SYPRO Orange (Invitrogen) 5000× concentrated in DMSO was diluted in 50 mM BTP pH 7.5 buffer to a 10x concentration. 25 µL solutions containing *Mh*GpgS (0.8 mg/ml) and *p*ABA and *p*HBA (enzyme:ligand ratios of 1:10, 1:30 and 1:60) were prepared in 50 mM BTP pH 7.5 buffer, mixed with an equal volume of SYPRO Orange 10x, and loaded in a white 96-well PCR plate (Bio-Rad). Controls with only enzyme or only ligand were also prepared. The thermofluor assay was performed in an iCycler iQ5 Multi-color Real Time PCR detection system (Bio-Rad) running the following protocol: heating from 25 to 85 °C with a 30 s hold time every 0.5 °C, followed by a fluorescence reading using Cy3 dye filter (excitation/emission, 545/585). The melting curves were analysed using the CFX Manager software (Bio-Rad) to calculate the melting temperature (Tm) from the minimum value of the negative first derivative of the melting curve.

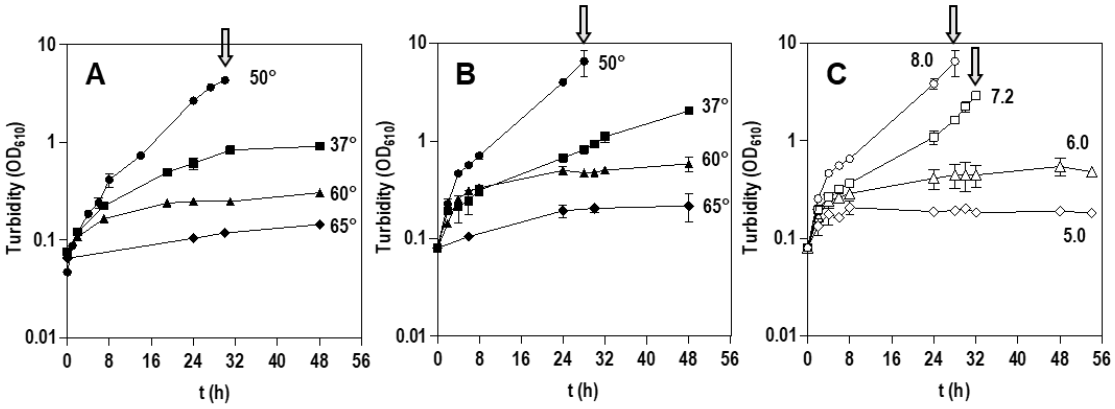
Inhibition assays were carried out with the UDP-Glo™ Glycosyltransferase Assay Kit (Promega). *Mh*GpgS was incubated with increasing concentrations of *p*ABA or *p*HBA in 25 mM BTP pH 7.5 at RT for 20 min. The reactions (containing 25 mM BTP buffer pH 7.5, 20 mM $MgCl_2$, 1 mM 3-PGA, 100 µM UDP-Glc, 14 nM *Mh*GpgS, and variable concentrations of inhibitor) were initiated by mixing 40 µL of *Mh*GpgS pre-incubated with inhibitor and 10 µL of substrate mix containing $MgCl_2$, 3-PGA and UDP-Glc and incubated at RT for 10 min. UDP production was quantified by mixing 10 µL of each reaction with 10 µL of UDP Detection Reagent in an opaque white 384-well plate, incubating at RT for 1 h, and recording the luminescence signal using a SpectraMax iD3 microplate reader. UDP standard curves were prepared in the 0.5 to 25 µM range and reaction conditions were optimised by conducting GpgS titration curves in the 1.8 to 114 nM range and time-course experiments in which the reaction was stopped after 1, 3, 7, 10, 15 or 20 min incubation.

Interference of the inhibitors with the assay was corrected by conducting parallel mock reaction mixtures containing all assay components, except the enzyme, and 1 µM of UDP in the presence of variable concentrations of inhibitor, and applying a correction formula as previously described (Shapiro et al., 2009): the luminescent signal of a reaction in the presence of X mM inhibitor was multiplied by the signal of a mock reaction containing 1 µM UDP in the absence of inhibitor, and divided by the signal of a mock reaction containing UDP and X mM inhibitor. Curve fitting was performed using GraphPad Prism® version 9, using the [inhibitor] vs. normalized response equation to obtain the IC_{50} values.

Results

M. hassiacum is a genuine thermophile growing optimally at 50°C and withstanding near-pasteurisation temperatures

M. hassiacum is a thermophilic (Figure 3.2A-B) and slightly alkaliphilic (Figure 3.2C) rapidly growing mycobacterium that grows optimally at 50°C and pH 8.0 in GPHF medium, describing a monophasic curve (μ=0.242±0.032 h^{-1}) (Figure 3.2A). Glycerol supplementation (0.8%) of GPHF medium enhanced exponential cell growth rate almost 2-fold (μ=0.440±0.016 h^{-1}) at the same temperature and pH (Figure 3.2B,C). Moreover, growth in glycerol-supplemented GPHF medium exhibited a growth rate shift detected at around 4 h of growth in all conditions except at a growth temperature of 65°C or at a growth pH of 5.0 (Figure 3.2B,C). The maximal specific growth rate (μ) determined under optimal growth conditions in the glycerol-supplemented medium decreased to about 24% of maximum rate after the shift, and in all conditions the specific rates were lower during the post-shift phase (Figure 3.2B,C). These shifts are reminiscent of diauxic growth, and the fact that they were not observed in the absence of glycerol could suggest sequential utilization of carbon sources. Diauxic growth has been reported for the pyrene-degrading isolate *Mycobacterium* sp. KMS (Miller et al., 2004; Zhang & Anderson, 2013) but this nutritional strategy is not used by the model mycobacteria *M. smegmatis* or *M. tuberculosis* (Aung et al., 2015; De Carvalho et al., 2010), and further studies would be required to clarify the cause of the shift observed for *M. hassiacum*.



Temp. °C	μ
37 (■)	0.148±0.006 [μ (0-7h)]
50 (●)	0.242±0.032 [μ (0-8h)]
60 (▲)	0.055±0.001 [μ (0-19h)]
65 (♦)	0.028±0.010 [μ (0-24h)]

Temp. °C	μ (0-4h)	μ (4-24h)
37 (■)	0.302±0.056	0.046±0.008
50 (●)	0.440±0.016	0.107±0.005
60 (▲)	0.289±0.016	0.022±0.004
65 (♦)	0.041±0.006 [μ (0-24h)]	

pH	μ (0-4h)	μ (4-24h)
5.0 (◊)	0.191±0.081	n.d.
6.0 (Δ)	0.274±0.058	0.024±0.007
7.2 (□)	0.302±0.032	0.069±0.005
8.0 (○)	0.440±0.016	0.107±0.005

Figure 3.2. Growth of *M. hassiacum* under different temperatures **(A, B)** or pH **(C)** in GPHF **(A)** and glycerol-supplemented GPHF **(B, C)** media. Growth rates (μ) reflect early exponential phases up to 4 h of growth (before growth shifts) and during exponential phases between 4 and 24 h growth. At 65°C (pH 8.0) *M. hassiacum* exhibits monophasic growth and at pH 5.0 (50°C) cells entered stationary phase after 4–6 h. Arrows indicate the turbidity above which cells aggregated and accurate measurements were no longer possible. Data are representative of three independent experiments. Error bars are the standard error of the mean. All graphs were prepared by Susana Alarico and adapted with permission from Alarico et al., 2020.

In order to further characterise the thermal resistance of *M. hassiacum* beyond its optimal growing temperature of 50°C, cell culturability and viability assays at higher temperatures (60, 65 and 70°C) were performed in rich medium (GPHF) and also in water, reproducing the oligotrophic conditions NTM face in the environment. As expected, no loss of viability was observed at 60°C by either colony formation (Figure 3.3A) or AlamarBlue (Figure 3.3B) assays after 1 h exposure. *M. hassiacum* was found to be able to withstand 65°C for 10 min and 70°C for 2 min without considerable loss of colony forming units (Figure 3.3A). After 10 min at 65°C, *M. hassiacum* retained 89% and 84% of reducing potential in rich medium and in water, respectively, and after 2 min at 70°C it retained 79% and 76%. After 60 min at 65°C, metabolic viability decreased to 20% and 27% in GPHF and water, respectively (Figure 3.3B). Inactivation was observed after 10 min at 70°C, with no visible colony formation although still retaining reducing potential values of 17% in GPHF medium and 8% in water. These results indicate that nutrient-depletion has no appreciable effect on the resistance of *M. hassiacum* cultures to high temperatures under the conditions tested.

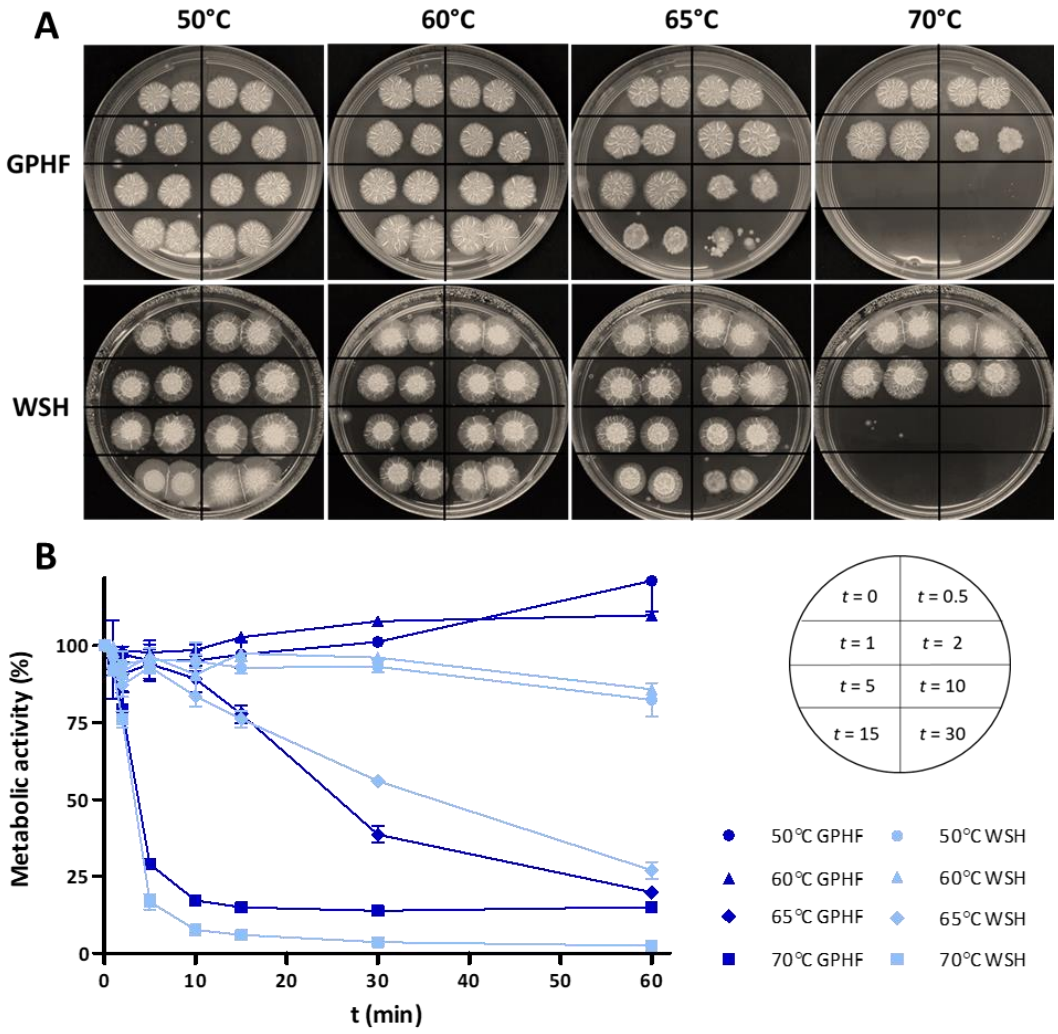


Figure 3.3. Thermotolerance profile of *M. hassiacum* exponentially growing cultures (GPHF) and of starved cells suspended in water (WSH). Duplicate samples were exposed to 50, 60, 65 and 70 °C at different time points. **(A)** Culturability after temperature exposure was assayed by spotting samples in duplicate on GPHF agar as displayed in the plate scheme below panel A (time is given in min). One representative of at least two independent experiments is shown. **(B)** Metabolic viability after temperature exposure was measured by the AlamarBlue assay. Viability of exponentially growing cultures in rich medium and of starved cell suspensions are represented in dark blue and light blue, respectively. Data comes from two independent experiments and error bars show the standard deviation.

GpgS is functional in *M. leprae* and is likely a core mycobacterial protein

The *M. tuberculosis gpgS* orthologue gene has been considered essential for growth by transposon mutagenesis studies (Minato et al., 2019). *M. leprae* has an extremely downsized genome compared with most mycobacterial species, and less than half of it is predicted to encode functional genes (Cole et al., 2001). Therefore, genes which have remained functional in *M. leprae* despite an extreme reductive evolution process may be considered core mycobacterial genes, part of the genus' "essential toolkit". In order to further investigate the importance of this enzyme and if it can be considered a core

mycobacterial protein, we recombinantly expressed the *M. leprae gpgS* orthologue to confirm if it encodes a functional GpgS. The codon optimised gene showed low levels of expression but the recombinant tagged protein could be produced in soluble form. SDS-PAGE analysis after affinity purification revealed a band corresponding to the predicted molecular weight of the native *Ml*GpgS protein (35 KDa), as well as a contaminant of higher molecular weight (Figure 3.4A). Although we did not confirm the identity of this contaminant, we do not expect it to affect activity results since *E. coli* extracts are consistently unable to carry out 3-PGA glycosylation (data not shown). The yield obtained was approx. 1.3 mg of protein per liter of *E. coli* culture. Incubation of the partially purified *Ml*GpgS with 3-PGA and UDP-Glc led to production of GPG, demonstrating that the GpgS *M. leprae* orthologue is fully functional (Figure 3.4B). *Ml*GpgS was also able to use ADP-glucose and GDP-glucose as glycosyl donors, but not GDP-mannose (results not shown).

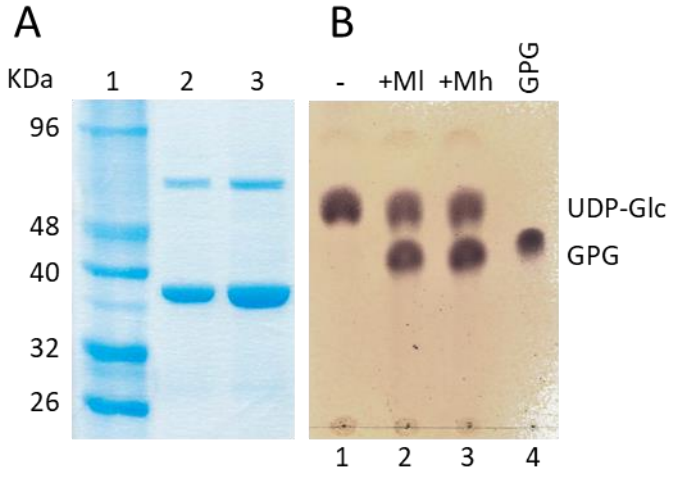


Figure 3.4. The *M. leprae gpgS* gene encodes a functional enzyme. **(A)** SDS-PAGE analysis of *M. leprae* GpgS after intein-mediated affinity purification with a chitin-binding tag. Lane 1: molecular weight marker; lanes 2 and 3: 3 µg and 6 µg of *Ml*GpgS. **(B)** TLC analysis of *Ml*GpgS activity. Reactions (25 mM BTP pH 7.5, 5 mM 3-PGA, 5 mM UDP-Glc, 5 mM $MgCl_2$, 3 µg GpgS) and control without enzyme were incubated for 1 h at 30°C. Lane 1: control mixture without enzyme; lane 2: reaction with *Ml*GpgS; lane 3: reaction with *Mh*GpgS; lane 4: GPG standard.

*Mh*GpgS is a highly thermostable glucosyl-3-phosphoglycerate synthase

The thermophilic nature of *M. hassiacum* (Figures 3.2A,B and 3.3) suggests it might be a source of thermostable enzymes, more amenable to structural studies and thus good surrogates to study essential mycobacterial enzymes. As a proof of concept, and due to its essentiality in *M. tuberculosis* and retained functionality in *M. leprae* (Figure 3.4), the enzyme GpgS was selected for characterisation. The 945 bp *gpgS* gene from *M. hassiacum* codes for a polypeptide with 314 residues and a calculated molecular weight of 33.6 kDa.

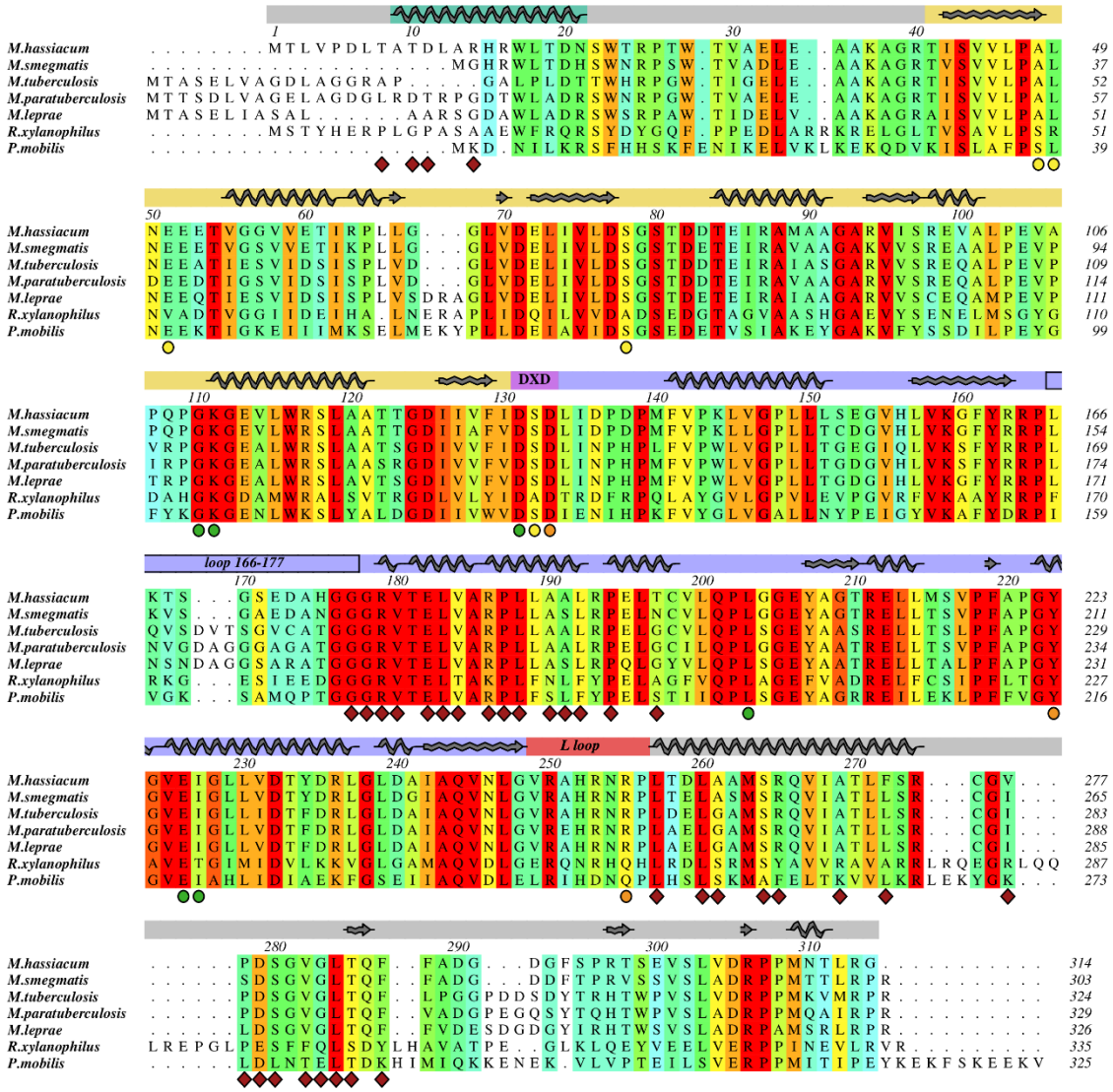


Figure 3.5. Multiple amino acid sequence alignment of GpgS homologues whose function has been confirmed. Residues are coloured according to conservation (red: identical residues, orange to blue: decreasing conservation; white: dissimilar). Amino acid numbers, structural subdomains, and secondary structure elements are based on the *Mh*GpgS 3D structure and represented above the alignment. Subdomains and motifs are represented by coloured boxes (N-terminal helix: teal; nucleotide-binding domain: yellow; acceptor binding domain: slate blue; DXD catalytic motif: purple; *L* loop: red). Residues that participate in the dimer interface and in UDP-Glc binding are indicated below the alignment by diamonds and circles, respectively (uridine moiety interacting residues: yellow circles; metal and phosphate coordination: orange circles; glucose moiety interacting residues: green circles). The dimer interface was calculated with PDBePISA server using the *apoMh*GpgS structure at pH 7.2.

BLAST analyses with its amino acid sequence showed 73-90% amino acid identity with other mycobacterial GpgSs, namely the characterised homologues from M. *bovis*/*M. tuberculosis* (73%), *M. avium* subsp. *paratuberculosis* (77%) and *M. smegmatis* (88%) (Empadinhas et al., 2008; Fulton et al., 2008). Homologues were also found in other thermotolerant mycobacteria such as *M. thermoresistibile* and *M. xenopi* (both 81%), in highly relevant pathogenic NTM most frequently colonising water systems including *M. abscessus* (81%) and *M. avium* (77%), and even in mycobacteria with extensively

downsized genomes, namely *M. leprae* (confirmed to be a *bona fide* GpgS, Figure 3.4), "*M. lepromatosis*" and "*M uberis*" (all 74%). Furthermore, *Mh*GpgS shows 43% and 37% amino acid identity with the characterised homologues from *Petrotoga mobilis* and from *Rubrobacter xylanophilus*, respectively (Empadinhas et al., 2011; Fernandes et al., 2010), but is only distantly related to isofunctional enzymes identified in the archaeon *Methanococcoides burtonii* (Costa et al., 2006), in the aquificale *Persephonella* marina (Costa et al., 2007) and in two cyanobacteria (Klähn et al., 2010). A multiple sequence alignment with all GpgS homologues whose function has been confirmed to date is shown in Figure 3.5, highlighting key conserved residues and motifs which will be discussed in following sections.

Expression of *M. hassiacum gpgS* in *E. coli* resulted in the production of homogeneous recombinant protein in soluble form at very high levels, which could be purified to homogeneity in a single affinity purification step, with yields typically close to 60 mg of protein per liter of *E. coli* culture (Figure 3.6A). The tagged enzyme displayed an apparent molecular weight of 76.9 kDa in solution, corresponding to 2.2 times the expected mass of the monomer (33.6 KDa plus 1.5 kDa hexahistidine tag), compatible with a dimeric arrangement. Activity assays carried out at 50 or 60°C led to GPG synthesis from UDP-Glc, ADP-glucose and GDP-glucose as donor substrates (with decreasing efficiency), and with 3-PGA as acceptor. GDP-mannose was also probed as a possible sugar donor but mannosyl-3-phosphoglycerate product was not detected (results not shown). Enzyme activity was maximal near 60°C and at pH 7.4–8.0, retaining 40% of maximal activity at 70°C and about 20% of maximal activity at 80°C (Figure 3.6B,C). *Mh*GpgS was dependent on divalent cations in the following order of efficiency: Mg^{2+} (20 mM) > Mn^{2+} (10 mM) (Figure 3.6D) while other cations tested did not stimulate enzyme activity at the concentrations tested.

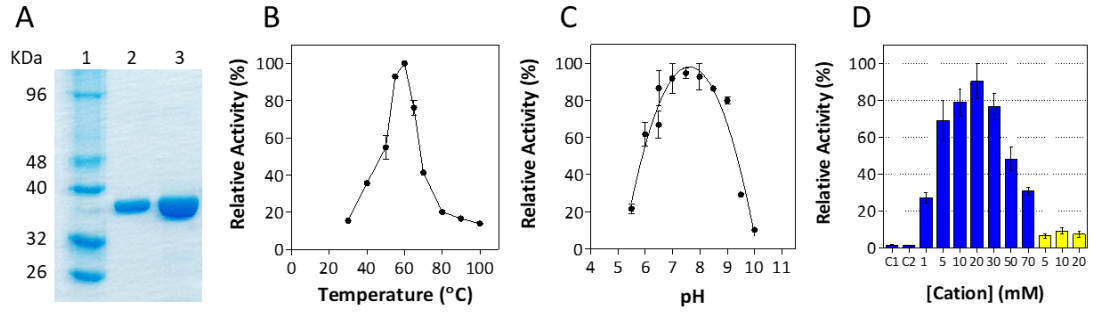


Figure 3.6. Biochemical characterisation of *Mh*GpgS. **(A)** SDS-PAGE analysis of *Mh*GpgS after Ni-affinity purification. Lane 1: molecular weight marker; lanes 2 and 3: 2 µg and 6 µg of *Mh*GpgS. **(B)** Temperature profile, **(C)** pH dependence, and **(D)** effect of divalent cations Mg^{2+} (blue) and Mn^{2+} (yellow) at different concentrations (C1 and C2, enzyme activity without cations and in presence of EDTA, respectively). Data are the mean values of three independent experiments. Panels B, C and D were prepared by Susana Alarico.

The enzyme was highly stable at 37 and 50°C (Figure 3.7A) with estimated half-lives of 7.8±1.4 and 6.4±2.4 days, respectively. After 1 week at room temperature or after 1 month at 0°C, it retained about 100% activity, and it could withstand several cycles of freeze-thaw without detectable loss of activity (results not shown). At 60°C, the pure recombinant enzyme denaturated quickly, after 5 min of incubation. Incubation in the presence of its substrates stabilized the enzyme, although its half-life at 60°C remained comparatively low at 12.1±2.1 min (Figure 3.7A). To further test the thermostability of *Mh*GpgS, the melting temperature (Tm) of the protein was determined at different pH values and in the presence of the nucleotide UDP, the preferred sugar donor UDP-Glc and the acceptor 3-PGA. The maximal values of Tm for *Mh*GpgS were observed at pH ranging from 7.0 to 9.0 (Figure 3.7B) suggesting that this protein is more stable in neutral to alkaline conditions. Interestingly, *Mh*GpgS thermoresistance was more pronounced in the presence of the protein's preferential substrates, the sugar donor UDP-Glc or the acceptor 3-PGA (Figure 3.7B). The highest Tm of 63.5°C was observed at pH 7.5 when both ligands were present demonstrating that *Mh*GpgS stability increases when both substrate-binding sites are occupied (Figure 3.7B).

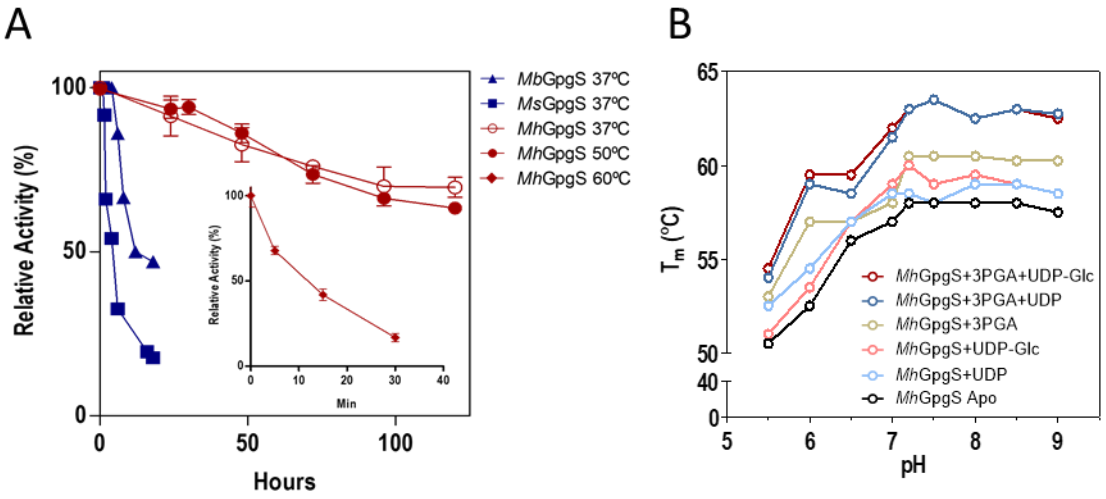


Figure 3.7. Thermostability of recombinant *Mh*GpgS. **(A)** Thermal inactivation of *Mh*GpgS activity monitored during 5 days at 37°C, 50°C and during 30 min at 60°C (after this time the protein precipitates) and of the recombinant GpgSs from *M. bovis* (*Mb*GpgS) and *M. smegmatis* (*Ms*GpgS) monitored during 18 h at 37°C. Data are the mean values of three independent experiments. **(B)** Melting temperature (Tm) of recombinant *Mh*GpgS in the presence of donor and/or acceptor substrates at different pH values. Data are the mean values of four independent experiments. Panel B was prepared by Alexandra Silva and adapted with permission from Alarico et al., 2020.

Taken together, these results confirm the higher thermostability of *Mh*GpgS when compared with the other characterised mycobacterial homologues, which when combined with the high expression levels in *E. coli* and ease of purification, make it an ideal candidate for thorough structural studies.

High-resolution structures of *Mh*GpgS reveal a typical mycobacterial GpgS architecture

A preliminary crystallisation screening yielded *Mh*GpgS crystals under a variety of non-overlapping conditions, making it possible to obtain high resolution structures at acidic, neutral and basic pH and confirming its predicted amenability to structural studies. At pH 5.5, tetragonal crystals belonging to the space group I 4_1 were obtained, which diffracted to 2.13 Å resolution and contained one molecule of *Mh*GpgS in the asymmetric unit. At pH 7.1, 7.2 and 8.5, orthorhombic crystals belonging to space group P 2_1 2_1 2_1 were produced instead and diffracted to 1.22, 1.11 and 1.43 Å respectively, containing two molecules in the asymmetric unit (designated as chains A and B).

Like its closest homologues, *Mh*GpgS displays a typical GT-A fold with a core catalytical domain, subdivided into a nucleotide-binding subdomain at the N-terminal (residues 41-130), which adopts the Rossman-fold, and an acceptor-binding subdomain at the C-terminal (residues 134-248), bridged by the signature catalytic motif DXD (^{131}DSD^{133}) (Figure 3.8A). The acceptor binding sub-domain is followed by a loop (*L-loop*, ^{249}VRAHRNRP^{256}), positioned at the divide between both subdomains, whose flexibility has been implicated in the GpgS catalytic cycle (Albesa-Jové, Romero-García, et al., 2017) and which contains the histidine residue (His252) suggested to be indispensable for pH-dependent metal coordination and donor binding (Fulton et al., 2008). The three-dimensional structure of *Mh*GpgS closely matches the previously reported structures of *Mt*GpgS and *Ma*GpgS (Albesa-Jové, Romero-García, et al., 2017; Fulton et al., 2008; Pereira et al., 2008). Indeed, *apoMh*GpgS at pH 5.5 shows overall RMSD values of 0.95 Å over 271 structurally equivalent Cα atoms when superposed with *apoMt*GpgS at pH 6.5, and of 1.1 Å over 266 structurally equivalent Cα atoms when superposed with *apoMa*GpgS at pH 5.5 (Figure 3.8B). Interestingly, the most relevant conformational difference between *apoMh*GpgS at pH 5.5 and the apo structures of its mycobacterial homologues at similar pH values can be found precisely in this loop. While two alternate conformations have been observed in *apoMt*GpgS at pH 6.5, "open" and "closed", and only the closed conformation has been observed in *apoMa*GpgS at pH 5.5, we could only model the open conformation in *apoMh*GpgS at pH 5.5 (Figure 3.8C).

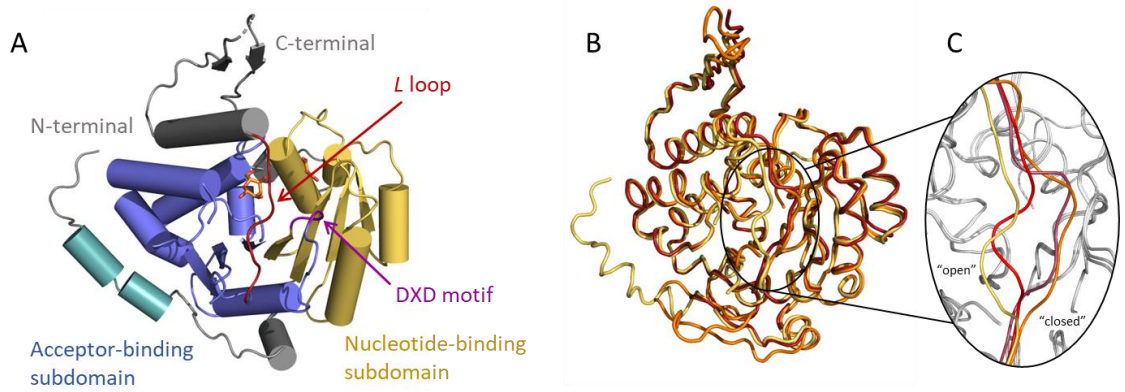


Figure 3.8. Overall structure of *Mh*GpgS. **(A)** Cartoon representation of the 3D structure of *Mh*GpgS with the catalytic core labelled and coloured as in Figure 3.5 (N-terminal helix: teal; nucleotide-binding domain: yellow; acceptor binding domain: slate blue; DXD catalytic motif: purple; *L* loop: red). Glucose, a contaminant that co-purified with *Mh*GpgS, and malonate, used as cryo-protectant, were modelled in the electron density at *Mh*GpgS active site and represented as sticks with carbon atoms orange and oxygen atoms red. **(B)** Ribbon representation of the superposed *apoMh*GpgS at pH 5.5 (yellow), *apoMt*GpgS at pH 6.5 (red; PDB entry 5JQQ) and *apoMa*GpgS at pH 5.5 (orange; PDB entry 3CKJ) 3D structures. **(C)** Close-up of the superposed structures from panel A showing the *L* loop conformation. Only the loop residues are coloured: *apoMh*GpgS: yellow; *apoMt*GpgS: closed conformation in rose red, open conformation in red; *apoMa*GpgS: orange.

Sharing 73% amino acid identity with *Mt*GpgS, the determinants of *Mh*GpgS's higher thermostability are not immediately evident. Global sequence and structural analyses do not show significant differences in amino acid composition that could account for differences in intrinsic flexibility (such as Gly and Pro content), hydrophobicity and content of hydrophobic buried residues, electrostatic surface potential, volume of cavities, and number of electrostatic interactions between residues (Ang et al., 2021; Petsko, 2001). Thus, a closer inspection of each residue substitution was carried out. A sequence alignment between these two homologues reveals a non-conserved N-terminal region (Figure 3.9A) which comprises an α-helix in all *Mh*GpgS structures and complexes obtained, but is unstructured in all *Mt*GpgS structures reported to date and is not predicted to assume a helical conformation by AlphaFold (Varadi et al., 2022). Indeed, the *M. tuberculosis* homologue has a higher Gly content in this region, making it more flexible, and two Pro substitutions, possibly disrupting the helix (Figure 3.9A). Interestingly, in *Mh*GpgS several residues in this region are located at the dimerization interface, interacting with a β-hairpin at the C-terminal. This hairpin had been previously shown to be implicated in *MtGpgS* dimerization (Pereira et al., 2008), but its interaction with the N-terminal segment observed in *M. hassiacum* is novel and might strengthen the dimer-dimer interaction, contributing to thermal stability (Figure 3.9B). Likewise, the substitution of Gly203 and Leu278 in *Mt*GpgS for Thr197 and Phe272 in *Mh*GpgS also appear to contribute to a tighter packing at the dimer interface. Furthermore, the N-

terminal region in *Mh*GpgS might also contribute to stabilise the monomer, by packing against the enzyme's core through interactions with the acceptor binding subdomain, namely at the Leu192-Val199 and Leu239-Ile242 regions. Val4, Leu7 and Leu12 form a well-defined hydrophobic patch that packs against the conserved residue Pro194, hindering its exposure to the solvent. Trp17 inserts into a hydrophobic pocket composed of the aliphatic portion of the side chain of Asn21, the conserved Leu239 and Cys198. The side chain of Asn21 further stabilises this region through hydrogen bonds with the carbonyl groups of Asp240 and Ile242 (Figure 3.9C). Taken together, these interactions may account for a more compact structure, making *Mh*GpgS more stable, despite an otherwise nearly identical architecture (Figure 3.8B).

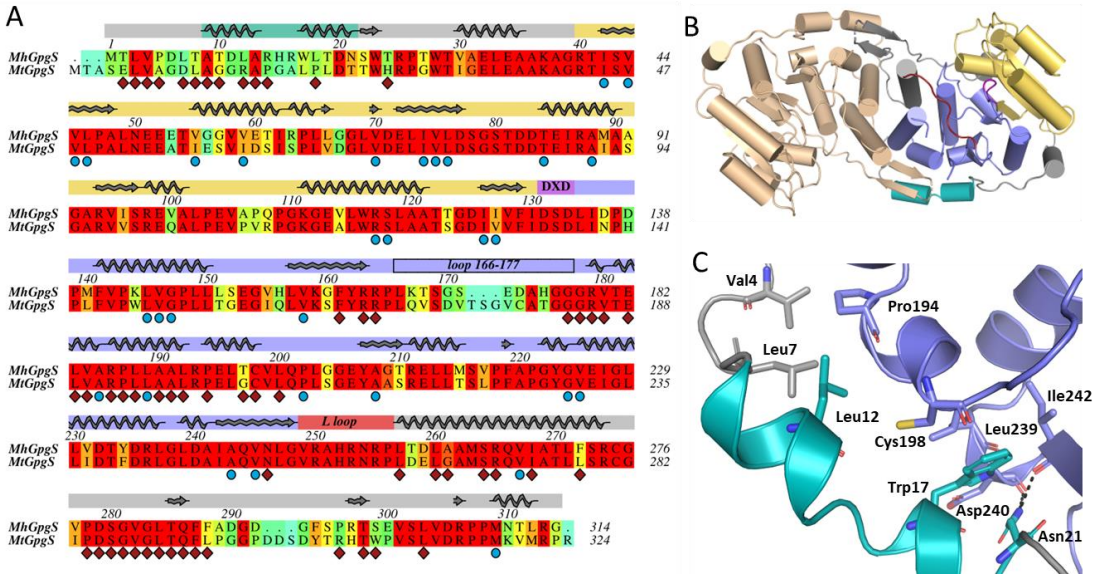


Figure 3.9. *Mh*GpgS dimer and possible determinants of thermostability. **(A)** Multiple amino acid sequence alignment of *Mh*GpgS and *Mt*GpgS. Residues are coloured according to conservation (red: identical residues, orange to blue: decreasing conservation; white: dissimilar). Amino acid numbers, structural subdomains, and secondary structure elements are based on the *apoMh*GpgS structure at pH 5.5 and represented above the alignment. Subdomains and motifs are represented by coloured boxes (N-terminal helix: teal; nucleotide-binding domain: yellow; acceptor binding domain: slate blue; DXD catalytic motif: purple; *L* loop: red). Residues that participate in the dimer interface and buried residues are indicated below the alignment by diamonds and circles, respectively. Buried residues and the dimer interface were calculated with PDBePISA server using the *apoMh*GpgS structure at pH 5.5. **(B)** Cartoon representation of the *Mh*GpgS dimer based on the apo structure at pH 5.5. The monomer on the left is colour coded as in panel A, the monomer on the right was obtained by the symmetry operation -x,-y-1,z and is coloured beige. **(C)** Close-up of *Mh*GpgS showing some of the stabilising interactions between the N-terminal helix and the core of the enzyme. Structural subdomains are coloured as in panel A, selected residue side chains are shown as sticks with carbon atoms colour coded, nitrogen atoms blue, oxygen atoms red and sulfur atoms yellow. Hydrogen bonds are indicated by black dashed lines.

On the other hand, some residue substitutions identified by the sequence alignment seem to indicate a greater flexibility for *Mh*GpgS in certain regions, particularly in the vicinity of the active site (Figure 3.9A). Examples of this are Gly56 and Gly57 (Glu and Ser

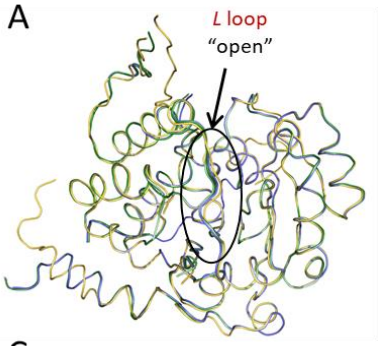
in *Mt*GpgS), as well as Val59 and Val55 (both Ile in *Mt*GpgS), which are in the neighbourhood of the nucleotide binding pocket and the catalytic DXD motif, and which suggest a less compact structure. This is also the case for Gly204 (Ser in *Mt*GpgS), which is located in a conserved loop at the base of the sugar binding pocket. This increased structural flexibility at the active site may allow *Mh*GpgS to retain a relatively high activity at suboptimal temperatures, compensating the effect of other substitutions which confer stability but also rigidity.

*Mh*GpgS displays a stable fold not influenced by variations in pH

Given that *Mh*GpgS easily crystallises at different pH values, we analysed the effect of pH on the conformation of the enzyme, both in the absence and presence of ligands, and focusing on the *L* loop histidine whose ionizable imidazole group was previously suggested to be required for metal coordination and sugar-donor binding (at pH 5.5, below the pK_a of histidine, the imidazole group is protonated and thus cannot coordinate the divalent metal ion required for catalysis) (Fulton et al., 2008). To perform this analysis, we started by superposing all molecules of the apo enzyme obtained at different pH values (Figure 3.10A). As mentioned, at neutral and basic pH, *apoMh*GpgS crystallises as a dimer in the asymmetric unit with the two molecules (A and B) establishing extensive contacts at the dimer interface, as described previously for *Mt*GpgS (Pereira et al., 2008). The RMSD values between the Cα backbones of molecules A and B at the same pH ranged between 0.45 and 0.52 Å, and the single chain in the structure at pH 5.5 was found to more closely match the A (average RMSD 0.72 Å) rather than the B (average RMSD 0.83 Å) chains and to be the most distinct (Figure 3.10B). Indeed, structures at neutral and basic pH were nearly identical to each other over the length of the polypeptide chain, while at acidic pH small shifts in the backbone conformation could be observed in some regions. In part, these minor structural changes may be caused by crystal packing differences due to the acidic structure belonging to a different space group (Figure 3.10C). Of note, regions 160-180 (flanking residues) and 248-258 show backbone shifts between all structures (Figure 3.10C) and comprise the two flexible loops which have been implicated in catalysis (Albesa-Jové, Romero-García, et al., 2017). A closer look at these regions shows a significant conformational change of the His252 side chain, which at neutral and alkaline pH establishes a stacking interaction with the π-plane formed by the protonated side chain of Arg164 (Figure 3.10D). At pH 5.5, the side chain of His252 is also protonated, and the repulsive electrostatic interaction between His^+ and Arg^+ could explain the different position of the His252 side chain. Interestingly, this interaction was not observed in

previously reported *apoMt*GpgS and *apoMa*GpgS structures, in which the Arg164 side chain could not be modelled, while the His252 side chain position was highly variable between structures (Figure 3.10E).

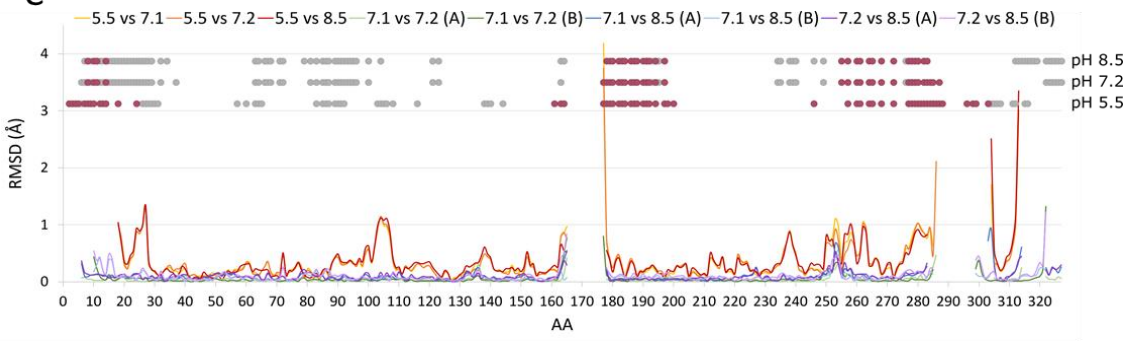
A



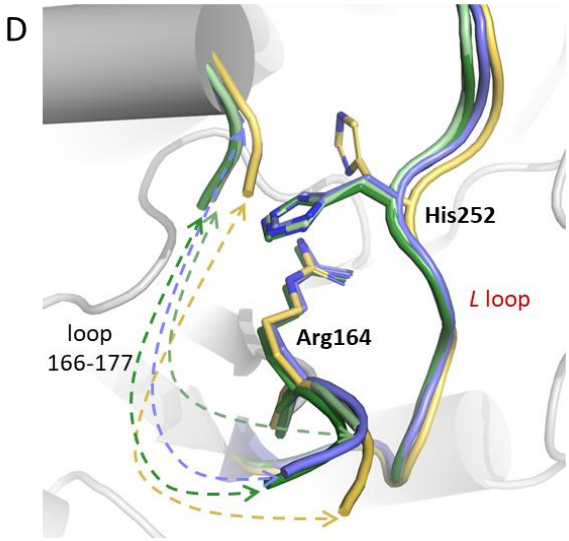
B

pH (chain)	5.5	7.1 (A)	7.1 (B)	7.2 (A)	7.2 (B)	8.5 (A)
7.1 (A)	0.7039	-	-	-	-	-
7.1 (B)	0.8246	0.5155	-	-	-	-
7.2 (A)	0.7384	0.0711	0.4866	-	-	-
7.2 (B)	0.8766	0.4766	0.1276	0.4459	-	-
8.5 (A)	0.7038	0.1456	0.4497	0.1273	0.4557	-
8.5 (B)	0.7794	0.4975	0.1108	0.4721	0.1652	0.4790

C



D



E

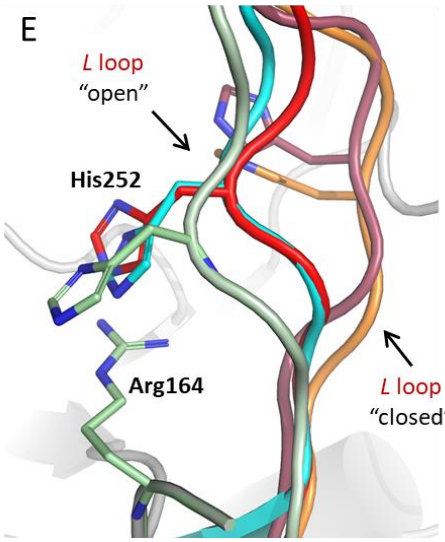


Figure 3.10. Comparison of *Mh*GpgS structures obtained in acidic, neutral and alkaline conditions. **(A)** Ribbon representation of the superposed *apoMh*GpgS 3D structures at pH 5.5 (yellow), pH 7.1 (light green), pH 7.2 (green) and pH 8.5 (blue). The *L* loop is circled. **(B)** Table of RMSD values between the backbone Cα atoms of all chains. **(C)** Variation of RMSD values over the length of the polypeptide chain for each structure superposition. For three representative chains (single chain at pH 5.5 and chain A at pH 7.2 and 8.5), residues involved in dimerization and crystal packing contacts are indicated by red and grey dots, respectively. **(D)** Close-up of the superposed structures from panel A showing two flexible loops where conformational shifts were observed. Only the loops are coloured as in panel A, with the structure at pH 5.5 shown as cartoon in grey and residue side chains shown as sticks with carbon atoms colour coded and nitrogen atoms blue. **(E)** Close-up of the structural superposition of the *apoMh*GpgS structure at pH 7.1 (light green) and previously published *apo*GpgS structures (*apoMt*GpgS at pH 6.5: red for the open conformation, rose red for the closed conformation, PDB entry 5JQQ; *apoMt*GpgS at pH 8: cyan, PDB entry 3E26; *apoMa*GpgS at pH 5.5: orange, PDB entry 3CKJ), showing the *L* loop and His side chain conformations.

The UDP binding pocket is rigid and not sensitive to pH

Crystals of *Mh*GpgS in complex with UDP were obtained in acidic (pH 5.5), neutral (pH 7.2) and basic (pH 8.5) conditions, yielding crystals belonging to the same space groups as the respective apo structures and which diffracted to 1.97, 1.23 and 1.60 Å resolution, respectively. UDP could be easily located in the electron density map (Figure 3.11B-D) and its binding had almost no effect on the overall conformation of the enzyme, with the following RMSD values between the backbones of *apoMh*GpgS and *Mh*GpgS·UDP complex: 0.21 Å at pH 5.5, 0.16 and 0.13 Å for the A and B chains at pH 7.2, and 0.29 and 0.19 Å for the A and B chains at pH 8.5. This suggests that the structure of the UDP binding pocket is rigid and not influenced by variations in pH. The regions with more pronounced conformational shifts, consistently observed across all crystallisation conditions, are the *L* loop and the loop 166-177 (Figure 3.11A). However, these conformational changes appear to result from their inherent flexibility and not from ligand binding. Indeed, when comparing chains A and B of the same crystallographic structure, these regions also show slight backbone shifts (data not shown). Interestingly, the stacking interaction between Arg164 and His252 described above is also observed in the *Mh*GpgS·UDP complexes at neutral and basic but not at acidic pH, consistent with what was observed for the apo structures (Figure 3.11B-D).

The binding mode of the uridine moiety to *Mh*GpgS is equivalent in all three structures. Briefly, uracil O2' hydrogen bonds with the main chain amide of Ala48, the uracil N3' with the side chain of Ser78, and the uracil O4' interacts with the main chain amide of Gly110 through a conserved water molecule. The uracil moiety also establishes hydrophobic contacts with the aliphatic chains of Leu49 and Lys111. The ribose hydroxyl O2 hydrogen bonds with the main chain amide of Leu49 and with the side chain of Glu51, while the O3 interacts with both the main and side chains of Ser132. All the aforementioned residues are highly conserved in mycobacterial GpgS sequences (Figure 3.5) and this binding mode does not differ from what has been described for *Mt*GpgS and *Ma*GpgS (Albesa-Jové, Romero-García, et al., 2017; Fulton et al., 2008; Pereira et al., 2008). With the exception of Lys111, these uridine interacting residues are not conserved in the *R. xylanophilus* GpgS (Figure 3.5) which exclusively utilises GDP-glucose (Empadinhas et al., 2011).

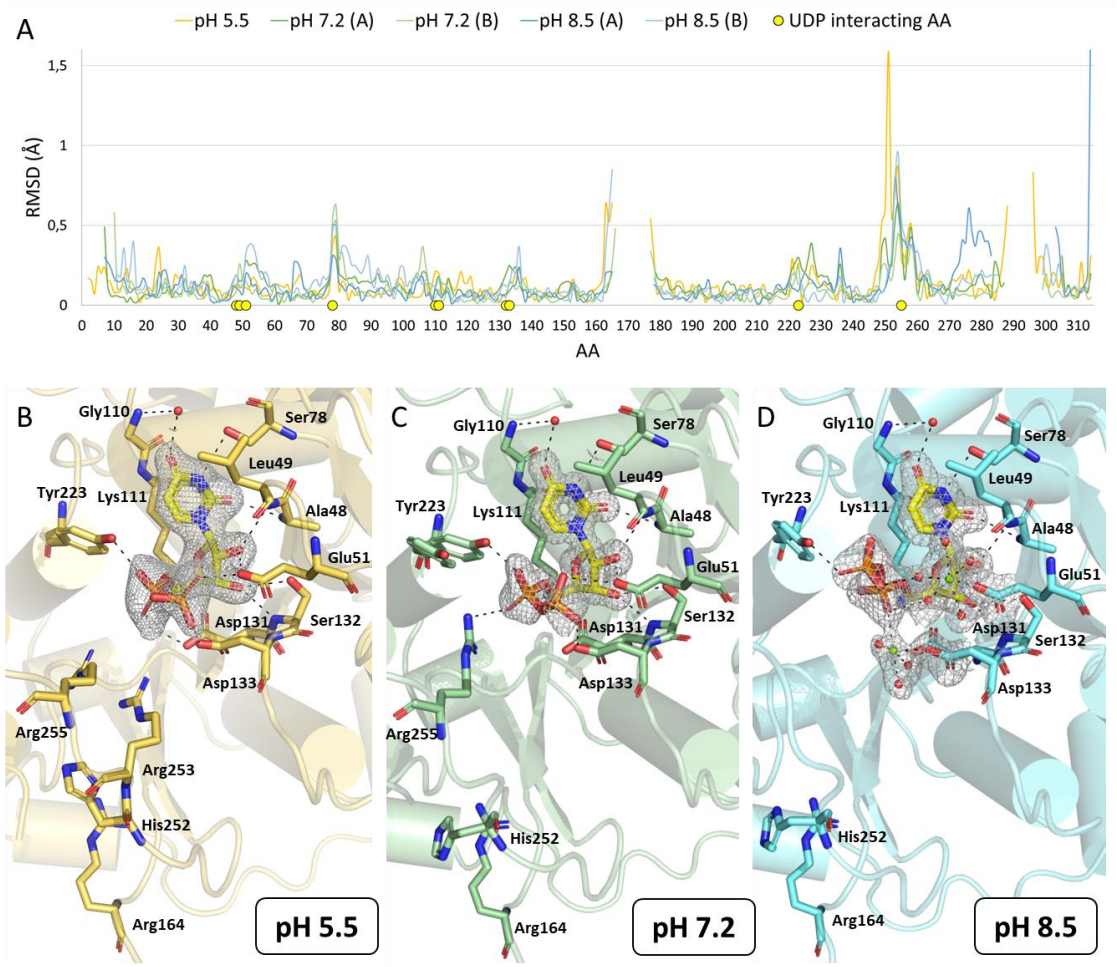


Figure 3.11. *Mh*GpgS·UDP complexes obtained in acidic, neutral and alkaline conditions. **(A)** Variation of backbone RMSD values over the length of the polypeptide chain between each *Mh*GpgS·UDP complex chain and the corresponding apo structure chain at the same pH. Residues involved in UDP binding are indicated by yellow circles. **(B)**, **(C)** and **(D)** Close-up of the enzyme active site in different crystallisation conditions. The electron density maps (2*mFo-DFc* contoured at 1.0 σ) are shown as a grey mesh around UDP, Mg^{2+} ions and Mg^{2+}-coordinating side chain groups and water molecules. UDP is shown as sticks with oxygen atoms red, nitrogen dark blue, phosphorus orange, and carbon bright yellow. Magnesium ions and water molecules are shown as green and red spheres, respectively. Metal coordination and hydrogen bonds are indicated by black dashed lines. The protein is depicted in cartoon representation and colour coded according to pH (pH 5.5: yellow; pH 7.2: light green; pH 8.5: cyan). Selected residues are represented as sticks with oxygen atoms red, nitrogen dark blue and carbon coloured according to pH.

On the other hand, differences in how the phosphate moiety is positioned and interacts with the active site could be observed between the three *Mh*GpgS·UDP complexes (Figure 3.11B-D). In the structures at pH 5.5 and 7.2, no metal ion is seen coordinating the phosphate moiety, which is stabilised by interactions between the α-phosphate and the hydroxyl group of Tyr223, as well as between the β-phosphate and the carboxyl groups of Glu51 and Asp 133, all three residues being highly conserved. Additionally, the side chain of Arg255 interacts with the α-phosphate in the structure at pH 7.2 (Figure 3.11C), but not at pH 5.5, in which this residue adopts a different conformation due to a stacking interaction with Arg253 (Figure 3.11B), whose side chain is

disordered in the two other complexes. Likely due to the high magnesium concentration of the alkaline crystallisation buffer used, at pH 8.5 two magnesium ions are seen coordinating the phosphate moiety, which assumes a different orientation (Figure 3.11D). In this structure, the first magnesium is coordinated by Asp133 and the α-phosphate, and, although it likely adopts a distorted octahedral coordination with four waters fulfilling its coordination sphere, only three of these waters are visible in the electron density. The side chain of Arg255 is disordered, and Tyr223 adopts a different conformation to hydrogen bond with the β-phosphate instead of interacting with the α-phosphate. The second magnesium ion is coordinated by the side chain of Glu51 and five water molecules, two of which are involved in hydrogen bonds with both the α- and β-phosphate groups. This second cation coordination site had not been observed before in GpgS. Interestingly, His252 is distant from both metal coordination sites and the *L* loop remains in its open conformation, just as observed for the *apoMh*GpgS structures, but in stark contrast to previously reported complexes of metal-bound GpgS in which this conserved histidine always participates in cation coordination along with the also conserved second aspartate of the DXD motif (Asp133 in *Mh*GpgS) (Albesa-Jové, Romero-García, et al., 2017; Fulton et al., 2008).

The *L* loop histidine does not play a role in metal coordination and UDP-Glc binding to *Mh*GpgS

UDP-Glc binding to *Mh*GpgS was initially investigated through co-crystallisation experiments at the different pH values, but UDP-Glc could only be located in the electron density map at pH 5.5, while at pH 7.2 and 8.5 only UDP or the uridine monophosphate moiety could be located. To account for the possibility of enzymatic cleavage during the long incubation times of the co-crystallisation experiment (2-3 days), soaking experiments were subsequently carried out, but UDP-Glc still could not be located at the active site in complexes obtained under neutral and basic conditions. Of all the acquired datasets, the best diffracting crystals were obtained by co-crystallisation at pH 5.5 and pH 7.2 and by soaking at pH 8.5, diffracting up to 2.05, 1.25 and 1.93 Å resolution, respectively. The complexes belonged to the same space group as the respective apo structures.

At pH 5.5, structural comparisons between *apoMh*GpgS, *Mh*GpgS·UDP and *Mh*GpgS·UDP-Glc further confirm the negligible effect of sugar donor binding on the overall conformation of the enzyme, as reflected by backbone RMSD values of 0.21 and 0.24 Å between the apo enzyme and the complexes with UDP and UDP-Glc, respectively, and of 0.20 Å between the two complexes. In addition to the flexible regions comprising

the *L* loop and loop 166-177, only slight conformational shifts are observed over the polypeptide chain (Figure 3.12A). The uridine moiety in the *Mh*GpgS·UDP-Glc complex binds exactly in the same way as described for the complex with UDP (Figure 3.11C), but the phosphate moiety adopts a different orientation. Tyr223 is still directly involved in stabilising the α-phosphate, but unlike in the complex with UDP, a magnesium ion coordinates both phosphate groups. Asp133, which directly interacted with the β-phosphate in the complex with UDP, is now part of the coordination sphere of the magnesium ion, and the β-phosphate is within hydrogen bonding distance of Arg255 (Figure 3.12C). The glucose moiety of UDP-Glc occupies a pocket which, interestingly, is not empty in the apo structure at pH 5.5, where a glucose molecule which possibly co-purified with GpgS could be modelled in the electron density, hydrogen bonding with the side chains of Asp131 and Glu226 and the main chain of Leu203 (Figure 3.12B). These highly conserved residues are also involved in the binding of the sugar donor's glucose moiety to *Mh*GpgS, as can be observed in the *Mh*GpgS·UDP-Glc complex: the side chains of Lys111 and Asp131 both establish hydrogen bonds with the O3' hydroxyl group of glucose, while the side chain of Glu226 interacts with both the O4' and O6' hydroxyl groups. Additional glucose-contacting residues are Leu203, whose carbonyl group is within hydrogen bonding distance of the O6' hydroxyl group, and Ile227 and Tyr223 which interact with O6' through a water molecule. The O2' hydroxyl group of glucose interacts with two water molecules, one of which belonging to the coordination sphere of the magnesium ion (Figure 3.12C). Overall, the active site remains nearly identical between the apo structure and the complex with UDP-Glc, with the only differences being observed for the side chains of Ser78, Lys111 and Arg255 (Figure 3.12B,C).

This binding mode is similar to what has been described for the other mycobacterial GpgSs, except we do not observe the involvement of Arg250 in glucose binding as described for *Mt*GpgS·UDP-Glc (PDB entry 5JSX, pH 7.5) nor the participation of His252 in metal coordination as reported for that complex and also for *Ma*GpgS·UDP-Glc (PDB entry 3CKQ, pH 7.0). This suggests that the *L* loop may not be as essential for metal and sugar donor binding as previous evidence suggested. Interestingly, in a previous attempt at obtaining a GpgS·UDP-Glc complex at pH 5.5, Fulton *et al.* (2008) could only locate UMP in the electron density, but not the β-phosphate and glucose moiety. The authors suggested that at this pH, the protonated histidine side chain would be unable to coordinate the metal ion, without which the phosphate and sugar moieties would not bind

strongly enough to be visible in the electron density, but our results are not consistent with this hypothesis.

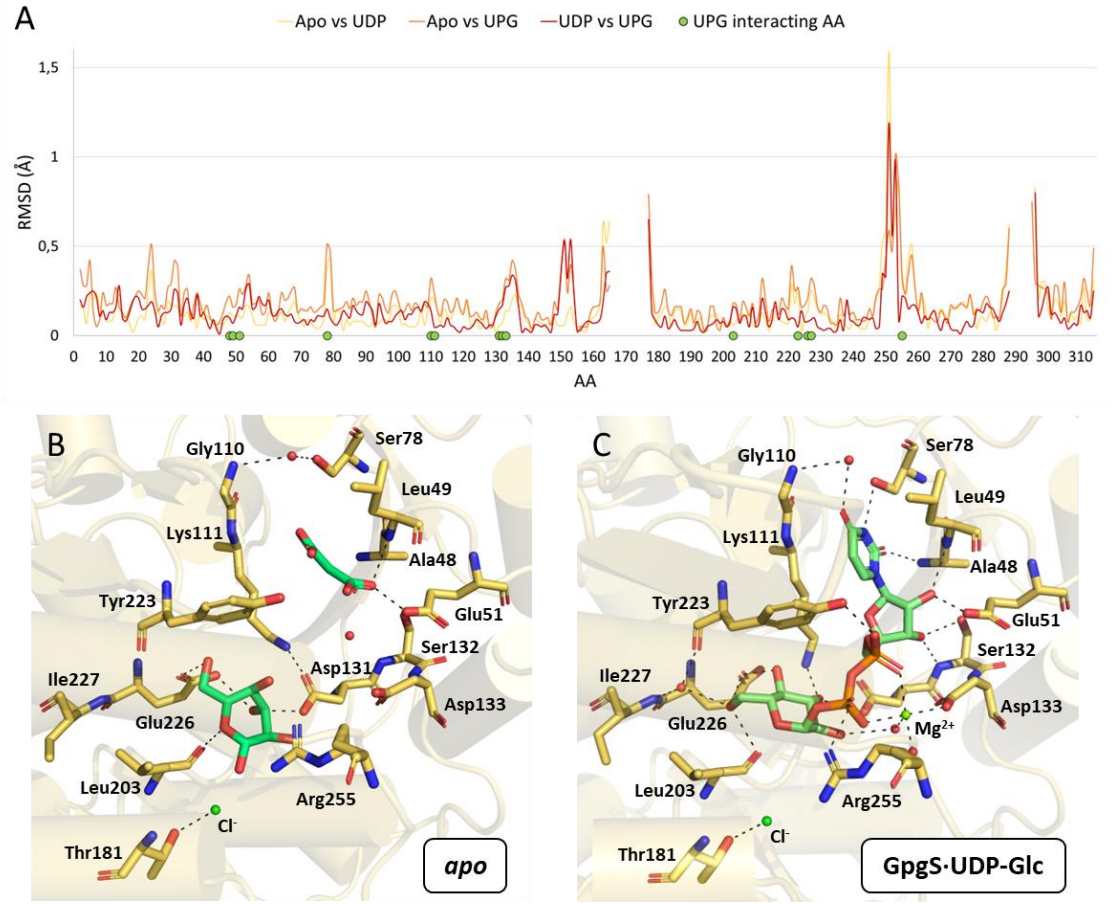


Figure 3.12. *Mh*GpgS·UDP-Glc complex at pH 5.5. **(A)** Variation of backbone RMSD values over the length of the polypeptide chain between each structure obtained at pH 5.5, showing the effect of UDP and UDP-Glc (UPG) binding on the conformation of the *Mh*GpgS backbone. Residues involved in UDP-Glc binding are indicated by green circles. **(B)** and **(C)** Close-up of the enzyme active site showing a side-by-side comparison between the apo structure (B) and *Mh*GpgS·UDP-Glc complex (C) at pH 5.5. UDP-Glc, glucose and malonate are shown as sticks with oxygen atoms red, nitrogen dark blue, phosphorus orange, and carbon bright green. Magnesium/chloride ions and water molecules are shown as green and red spheres, respectively. Metal coordination and hydrogen bonds are indicated by black dashed lines. The protein is depicted in cartoon representation and selected residues are represented as sticks with oxygen atoms red, nitrogen dark blue and carbon yellow.

In fact, of all the conditions tried in this work, we could only locate UDP-Glc in the electron density at pH 5.5. At pH 8.5, only UMP was found, likely due to the chemical instability of UDP-Glc under alkaline conditions in the presence of high magnesium concentrations (Hill et al., 2017). In this structure, uridine binds in the same way as described previously and two magnesium ions are present at the active site (Figure 3.13B), similarly to what is observed for the complex with UDP under the same crystallisation conditions (Figure 3.11D). At pH 7.2, the β-phosphate could also be located in the electron density, hydrogen bonding with Tyr223, although it is not buried within the glucose

binding pocket as in the equivalent complex with UDP-Glc, being instead exposed to solvent, which suggests that cleavage might have occurred (Figure 3.13A). Furthermore, the occupancy of UDP in this structure is below unity, and an unidentifiable region of positive electron density is observed in the glucose binding pocket (not shown), indicating that a reaction might be taking place in some of the active sites. These crystals were prepared in the absence of the enzyme's natural acceptor substrate, but its pocket is occupied by tartrate (also with an occupancy below unity), a component of the crystallisation buffer, which could facilitate UDP-Glc cleavage. Tartrate hydrogen bonds with the main and side chains of Thr181, with the main chain of Val180 and with the side chain of Arg255, as well as interacting with the main chain amide groups of Gly178 and Arg179 through two water molecules (Figure 3.13A). Residues 178-181 are highly conserved (Figure 3.5) and Thr181 (Thr187 in *Mt*GpgS) is a key residue involved in 3-PGA binding (Albesa-Jové, Romero-García, et al., 2017)

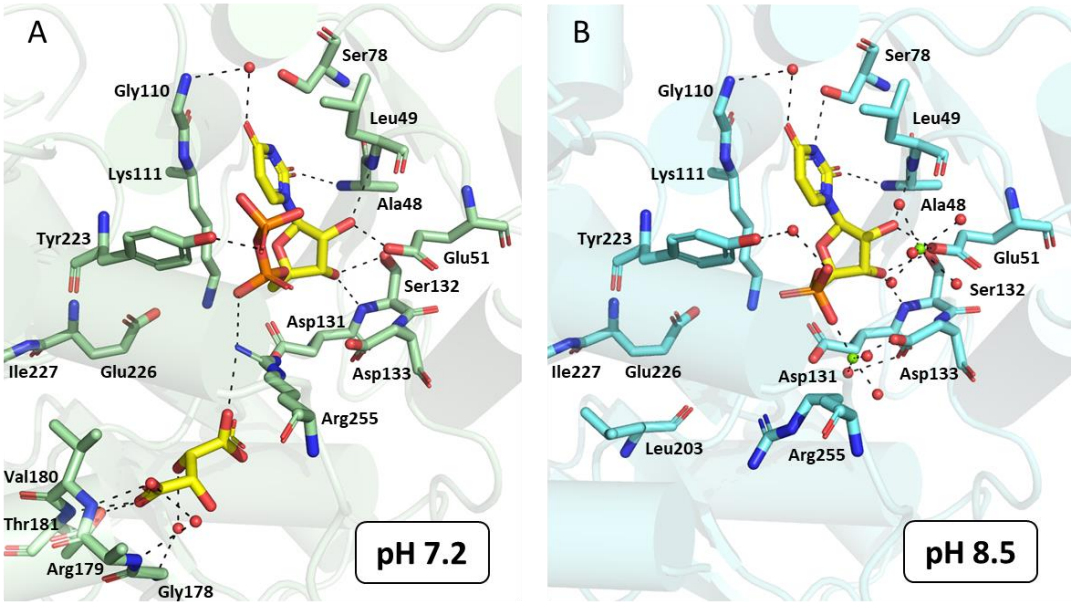


Figure 3.13. Close-up of the active site for the crystal structures obtained by co-crystallisation of *Mh*GpgS with UDP-Glc at pH 7.2 **(A)** and by soaking *Mh*GpgS crystals with UDP-Glc at pH 8.5 **(B)**. UDP, UMP and tartrate (crystallisation buffer component) are shown as sticks with oxygen atoms red, nitrogen dark blue, phosphorus orange, and carbon bright yellow. Magnesium ions and water molecules are shown as green and red spheres, respectively. Metal coordination and hydrogen bonds are indicated by black dashed lines. The protein is depicted in cartoon representation and selected residues are represented as sticks with oxygen atoms red, nitrogen dark blue and carbon green (A) or blue (B).

Substrate binding pockets are occupied even in the apo structures of *Mh*GpgS

As discussed in previous sections, the substrate binding pockets of *Mh*GpgS are rigid, unaltered by pH changes, and the catalytic site seems to be pre-formed, with all relevant residues set in place even in the absence of the natural ligands or magnesium.

This may partly be explained by the fact that all crystallisation experiments were conducted at 20°C, well below the enzyme's optimal activity temperature range (Figure 3.6B), and thus in conditions where the flexibility of a thermostable enzyme is expected to be low (Radestock & Gohlke, 2011). It is interesting to note, however, that none of the apo crystal structures obtained truly represent an unliganded form of the enzyme, making comparisons between apo and ligand-bound structures inevitably biased. In the apo structure at pH 5.5, for example, regions of positive electron density at the active site could be assigned as follows (Figure 3.12B): malonate, used as a cryo-protectant, was located at the uridine binding pocket; β-glucose, which seems to have co-purified with the enzyme, was located at the glucose binding pocket; and a chloride ion, present in high concentrations in the crystallisation buffer, was observed in close proximity to Thr181 (within 3.1 Å of its side chain hydroxyl group), occupying the position where the carboxylic group of 3-PGA was located in previously published complexes of *Mt*GpgS with 3-PGA (e.g., PDB entries 4Y6N and 5JQX). A chloride ion also occupies this position in the apo structure at pH 8.5, located at 3.0 Å distance of Thr181, as well as in all complexes with UDP and UDP-Glc obtained at pH 5.5 and 8.5. In the apo structures at pH 7.1 and 7.2, this binding site is occupied by the carboxylic acid buffers used in crystallisation, malate and tartrate, respectively, with an equivalent binding mode to that shown in (Figure 3.13A).

Although regions of positive electron density at the ligand binding pockets are observed in all the apo structures, not all could be assigned to compounds present during protein production, purification or crystallisation in sufficient amounts to be expected to co-crystallise with *Mh*GpgS. The most intriguing example is found in the uridine binding pocket of the structure with the highest resolution obtained in this study, *apoMh*GpgS at pH 7.2. Occupying the position of the uracil ring is a positive electron density that seems to suggest the presence of 6-membered ring molecule, with at least one substituent group in close proximity to the side chain of Glu51, and with the atom in the *para* position possibly within hydrogen binding distance of a water molecule coordinated by the amide main chain group of Gly110 and the side chain hydroxyl of Ser78 (Figure 3.14B). In *apoMh*GpgS at pH 7.1, a similar situation is observed although the ring is not closed, possible due to a lower occupancy of the unidentified ligand (Figure 3.14A). Since no matching compound was among the known components of the growth medium or purification and crystallisation buffers, we hypothesised that an *E. coli* metabolite with high affinity towards GpgS might have co-purified with the enzyme and searched the *Escherichia coli* Metabolome Database (ECMDB) (Sajed et al., 2016) for candidates.

Attempts to fit and refine benzyl alcohol (PDB code 0I0), pyrimidine (P1R), pyrazinoic acid (VGL), nicotinamide (NCA), nicotinic acid (NIO) and L-pipecolic acid (YCP) yielded unsatisfactory results.

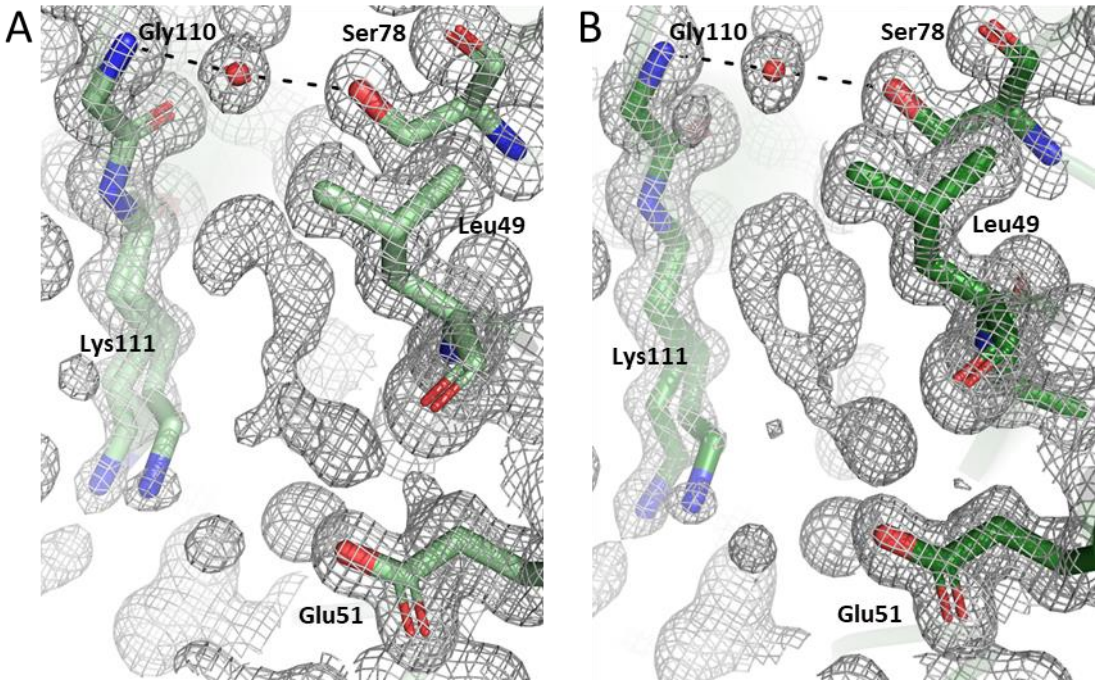


Figure 3.14. Close-up of the uridine binding pocket for the apo crystal structures of *Mh*GpgS obtained at pH 7.1 **(A)** and 7.2 **(B)**. Selected residues are represented as sticks with oxygen atoms red, nitrogen dark blue and carbon light green (A) or dark green (B). Hydrogen bonds are indicated by black dashed lines and the electron density maps (2*mFo-DFc* contoured at 1.0 σ) are shown as a grey mesh. The unassigned electron density is at the centre of the figures.

Genomic analyses in the *Corynebacteriales* hint at an intriguing GpgS partner

Even though we could not decipher the identity of the electron density shown in Figure 3.14, we sought to investigate if there could be a physiologically relevant interaction between GpgS and a molecule containing a 6-membered ring other than its natural substrate. Intriguingly, genomic context analyses in 167 *Corynebacteriales* GpgS orthologues with less than 95% sequence identity between them reveal a tight association between GpgS and two enzymes whose predicted or confirmed functions involve small 6-membered ring molecules (Figure 3.15). The closest genomic neighbour of *gpgS* found in this analysis is *folP2*, co-occurring in 92% of the genomes and immediately adjacent to *gpgS* (average distance of 1). FolP2, is a paralogue of the enzyme dihydropteroate synthase (FolP1) which participates in folate biosynthesis, an essential pathway in mycobacteria. One of the natural substrates of FolP1 is *p*ABA (Figure 3.15B) and although previous studies have shown that FolP2 is not a *bona fide* dihydropteroate synthase, the *p*ABA binding

pocket is conserved and evidence suggests that FolP2 retains the ability to bind the *p*ABA analogue dapsone (Gengenbacher et al., 2008).

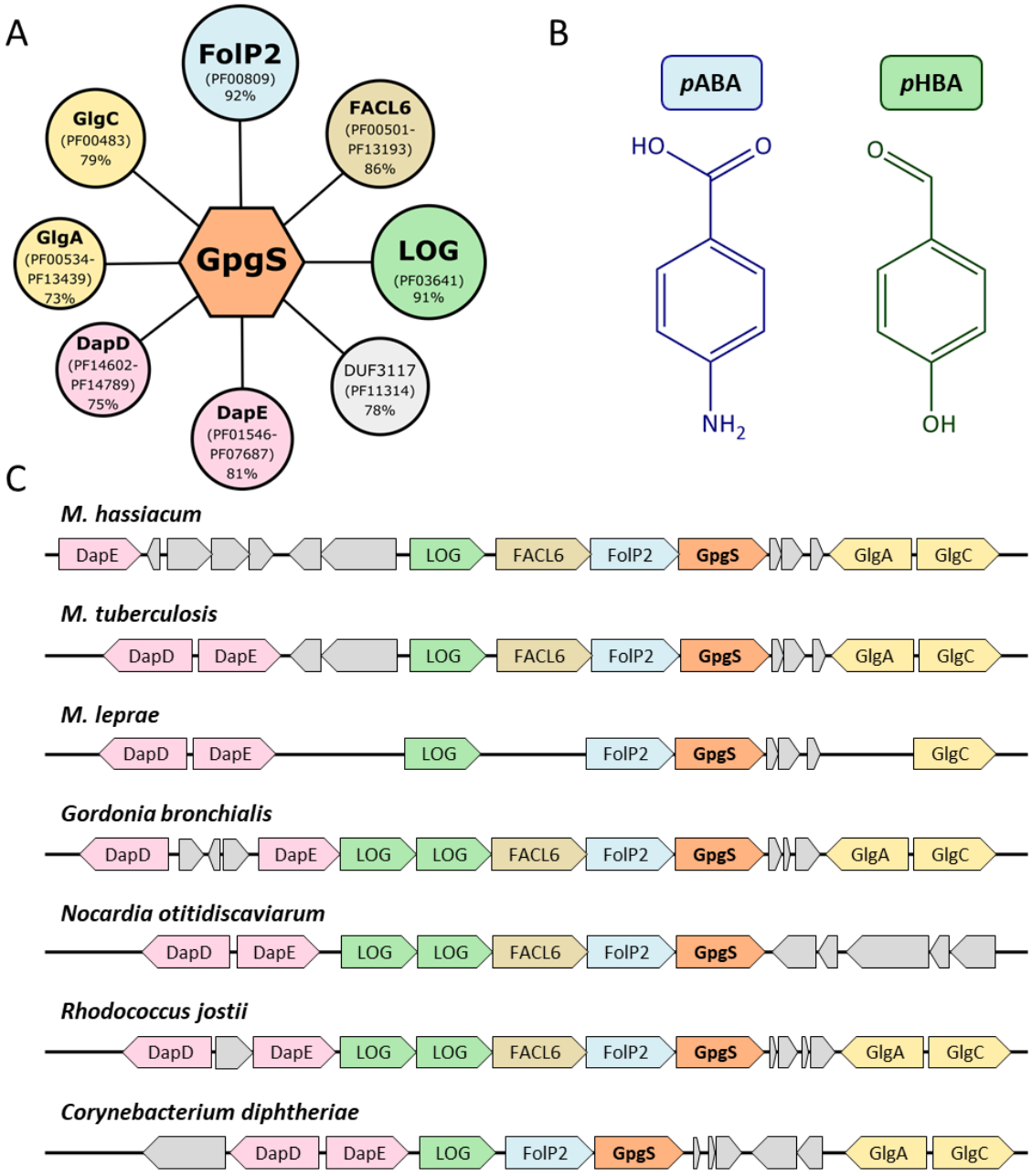


Figure 3.15. Genomic context of GpgS in the order *Corynebacteriales*. **(A)** GNN for a cluster of 167 GpgS orthologues in the order *Corynebacteriales*. Genomic neighbours with over 70% co-occurrence within a window of 10 ORFs upstream and downstream of the *gpgS* gene are shown. Pfam families are shown in brackets and the size of each circle is proportional to the co-occurrence, shown in percentage below the Pfam family. **(B)** Genomic context diagrams for GpgS orthologues in selected species. ORFs are coloured coded as in panel A and not drawn to scale, ORFs not shown in the GNN and of unknown function are coloured grey. Pseudogenes in *M. leprae* are omitted. FolP2, inactive paralogue of dihydropteroate synthase; FACL6, acyl-CoA synthetase; LOG, cytokinin riboside 5′-monophosphate phosphoribohydrolase; DapD, tetrahydrodipicolinate succinylase; DapE, *N*-succinyl-L,L-diaminopimelic acid desuccinylase; GlgA, glycogen synthase; GlgC, glucose-1-phosphate adenylyltransferase; *p*ABA, *p*-aminobenzoic acid; *p*HBA, *p*-hydroxybenzaldehyde.

The next neighbouring gene with higher co-occurrence (91%) is *log*, at an average distance of 3 ORFs from GpgS. LOG has been shown to participate in cytokinin biosynthesis in mycobacteria and increased LOG activity has been linked to the accumulation of toxic cytokinin breakdown products, such as *p*HBA (Figure 3.15B) (Samanovic et al., 2015). Also with a high co-occurrence and in close proximity to *gpgS* are the genes encoding FACL6, an acyl-CoA synthetase involved in fatty-acid metabolism that has been shown to be up-regulated when *M. tuberculosis* cells enter dormancy (Daniel et al., 2014), DapD and DapE, involved in lysine biosynthesis (Schuldt et al., 2009; Usha et al., 2016), GlgA and GlgC, which participate in glycogen biosynthesis (Koliwer-Brandl et al., 2016), and a domain of unknown function (Figure 3.15A).

These results, together with those presented in Figure 3.14, prompted us to investigate if *p*ABA and/or *p*HBA could bind GpgS at the uridine pocket and modulate its activity.

*p*ABA and *p*HBA bind *Mh*GpgS at the uracil binding pocket, but do not affect thermal stability and are poor inhibitors

Complexes between *Mh*GpgS and *p*ABA or *p*HBA were successfully obtained through soaking or co-crystallization. *Mh*GpgS·*p*ABA tetragonal crystals, belonging to the space group I 4_1, were obtained at pH 5.5 and diffracted up to 1.95 Å resolution. *Mh*GpgS·*p*HBA orthorhombic crystals, belonging to space group P 2_1 2_1 2_1, were obtained at pH 7.1 and diffracted to 1.27 Å resolution. In both cases, the compounds could be readily located in the electron density map at the uridine binding site, as expected (Figure 3.16C,D).

*p*ABA binding had no effect on the conformation of GpgS, as shown by a backbone RMSD value of 0.35 Å between the *Mh*GpgS·*p*ABA complex and the apo structure obtained under the same crystallisation conditions. Shifts along the polypeptide chain were only observed at the regions already identified as inherently flexible in previous comparisons (Figure 3.16A). *p*ABA hydrogen bonds to the side chain of Ser78 through its amino group and establishes hydrophobic contacts with the aliphatic chains of Leu49 and Lys111. Its amino group also interacts with the main chain amide of Gly110 through a conserved water molecule, similarly to what is observed for the uracil O4' carbonyl group in the complexes with UDP and UDP-Glc. The carboxylic group of *p*ABA is coordinated by four water molecules which mediate interactions with Glu51, Ser132, Lys111 and Tyr223, as well as with

a Bis-Tris molecule (from the crystallisation buffer) which was modelled in the electron density map at the glucose binding site (Figure 3.16C).

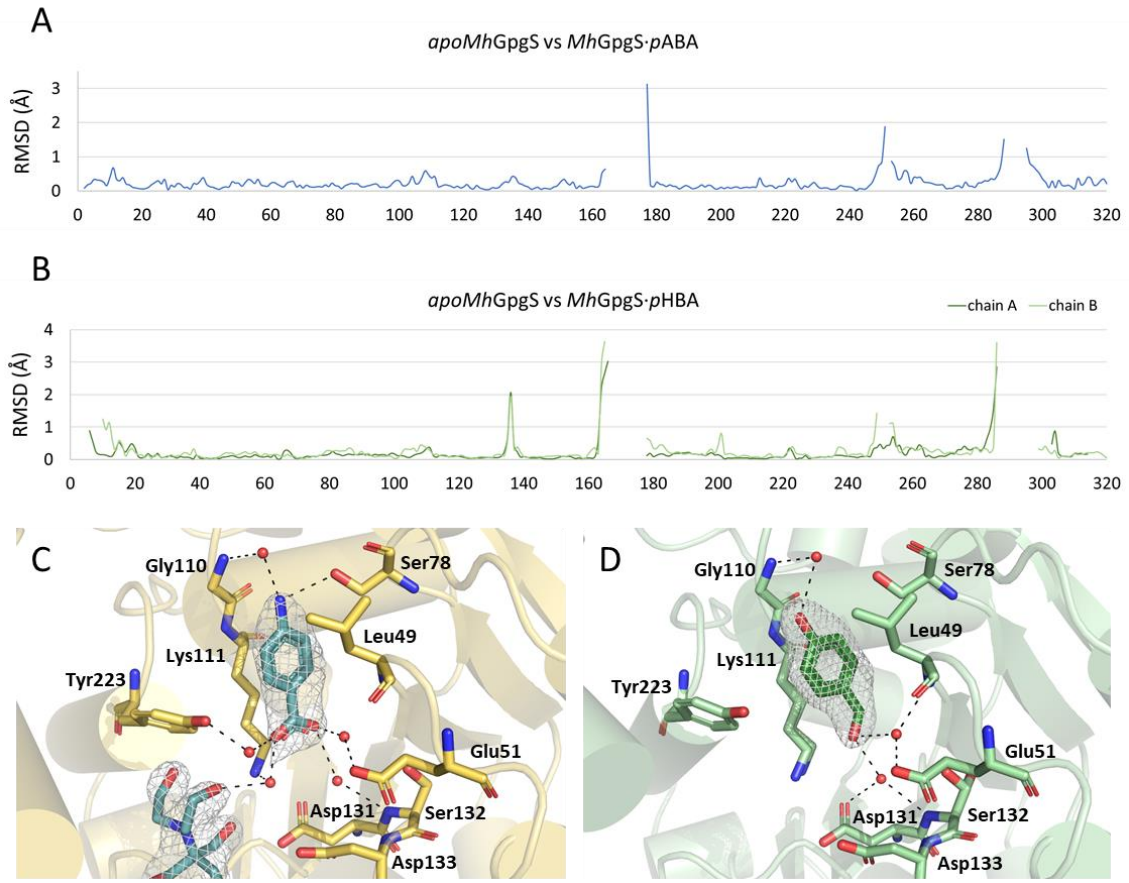


Figure 3.16. *Mh*GpgS complexes with *p*ABA and *p*HBA. **(A)** Variation of backbone RMSD values over the length of the polypeptide chain between the *Mh*GpgS·*p*ABA complex and the apo structure at pH 5.5. **(B)** Variation of backbone RMSD values over the length of the polypeptide chain between the *Mh*GpgS·*p*HBA complex and the *apo* structure at pH 7.1. **(C)** and **(D)** Close-up of the uridine binding pocket in the *Mh*GpgS·*p*ABA and *Mh*GpgS·*p*HBA complexes. The electron density maps (2*mFo-DFc* contoured at 1.0 σ) are shown as a grey mesh around (C) *p*ABA and Bis-Tris or (D) *p*HBA. Ligands are shown as sticks with oxygen atoms red, nitrogen dark blue, and carbon teal (for *p*ABA and Bis-Tris) or dark green (for *p*HBA). Hydrogen bonds are indicated by black dashed lines. The protein is depicted in cartoon representation and colour coded according to pH (pH 5.5: yellow; pH 7.1: light green). Selected residues are represented as sticks with oxygen atoms red, nitrogen dark blue and carbon coloured according to pH.

Interestingly, *p*HBA could be located not only at the uridine binding pocket, but also at other sites within the crystal contact interface. However, the effect of *p*HBA binding on the conformation of the enzyme was still negligible, with backbone RMSD values of 0.42 and 0.46 for chains A and B, respectively, when compared with the apo structure obtained under equivalent crystallisation conditions. In addition to the flexible regions seen undergoing slight conformational shifts in all previous structure comparisons, part of the loop containing the DSD catalytic motif, namely residues 135-139, adopt a slightly different conformation in the *Mh*GpgS·*p*HBA complex when compared with the apo structure at

pH 7.1, but the reason for this is not clear (Figure 3.16B). The binding mode of *p*HBA is similar to that observed for the uracil ring, with the aromatic ring stacked between the aliphatic chains of Leu49 and Lys111 and the hydroxyl group interacting with the main chain amide of Gly110 through a water molecule. The carbonyl group of *p*HBA hydrogen bonds with two water molecules which mediate interactions with Leu49, Glu51 and Asp131 (Figure 3.16D).

These binding modes do not reveal a particularly strong interaction between the ligands and GpgS, and this is in accordance with thermal shift experiments which showed no impact of *p*ABA or *p*HBA binding on the Tm of *Mh*GpgS (Figure 3.17A). It should be noted that UDP binding alone was found to have only a very slight effect on the Tm of *Mh*GpgS (Figure 3.7B), so it is not surprising that molecules which only mimic the uracil moiety, and cannot establish stabilising interactions equivalent to those of the ribosyl moiety, have an even smaller effect. The effect of these small ligands on the activity of *Mh*GpgS suggests that they are poor competitive inhibitors, with IC_{50} values in the range of 6 mM for both compounds, as revealed by preliminary inhibition assays (Figure 3.17B,C).

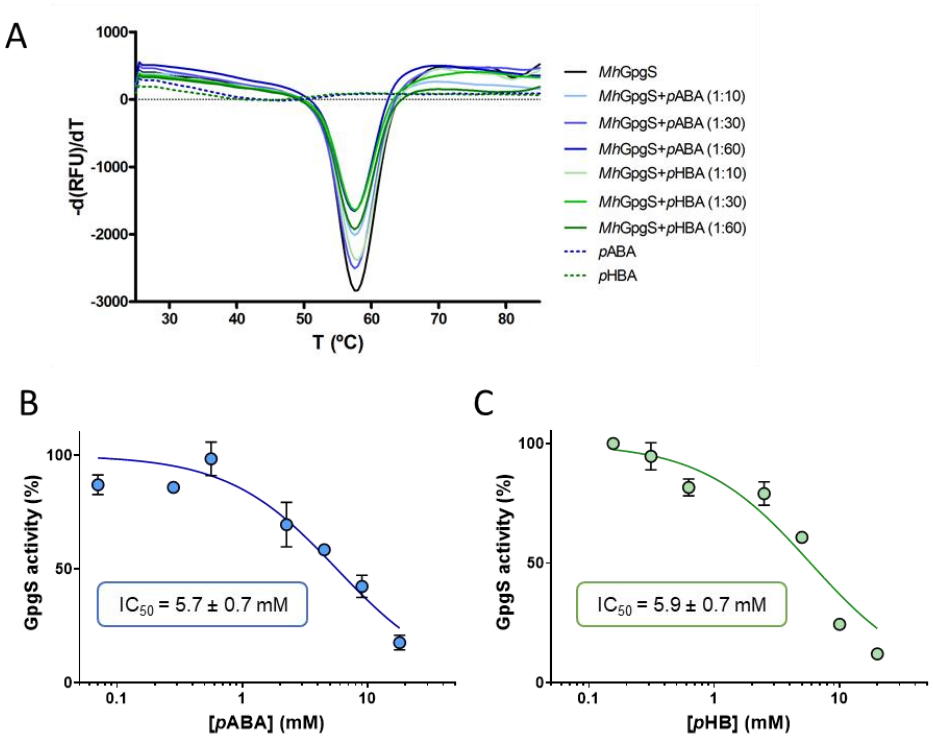


Figure 3.17. Effects of *p*ABA and *p*HBA on *Mh*GpgS stability and activity. **(A)** Thermal shift assay of the recombinant *Mh*GpgS in the presence of each ligand. The negative first derivative plots of fluorescence versus temperature show that the melting temperature (Tm) at pH 7.5 is not altered by the presence of *p*ABA or *p*HBA. Data are the mean values of four experiments. **(B)** and **(C)** Determination of IC_{50} for *p*ABA and *p*HBA. Data are the mean values of at least two experiments. Curve fitting was performed using GraphPad Prism® version 9 and the [inhibitor] vs. normalized response equation. IC_{50}±SEM is reported in an insert on the graph. The asymmetrical 95% confidence intervals are 4.3 to 7.5 mM for *p*ABA and 4.7 to 7.4 mM for *p*HB.

Discussion

Human pathogens are, as a rule, mesophilic organisms that struggle to grow and survive when exposed to temperatures higher than those found in the human host and for that reason, thermal treatment strategies are routinely used to decontaminate foodstuffs, medical equipment (Rutala & Weber, 2013) and hospital-water distribution systems (Whiley et al., 2017). Since heat exposure is one of the abiotic stresses NTM are highly resistant to (Schulze-Robbecke & Buchholtz, 1992), successful elimination of these opportunistic pathogens is challenging but essential, and indeed the United States Food and Drug Administration recommends the use of a thermophilic mycobacterium species to validate intermediate- and high-level thermal disinfection of medical devices, although not specifying which species should be used (FDA, 2002). *M. hassiacum*, growing optimally at 50 °C and up to 65 °C (Figure 3.2), is to the best of our knowledge the most thermophilic NTM identified to date and, although its clinical significance is still uncertain, it has been implicated in a few cases of human colonization or infection (Deinhardt-Emmer et al., 2018; Jiang et al., 2013; Salzer et al., 2020; Schröder et al., 1997; Tortoli et al., 1998). This species has been frequently detected in drinking-water distribution systems of high-income countries, characterised as oligotrophic environments to which mycobacteria successfully adapt (Gebert et al., 2018). As a rarely isolated but still potentially infective thermophilic NTM, and as a likely ideal surrogate for evaluating the effectiveness of thermal disinfection procedures, it is of utmost importance to investigate the ability of *M. hassiacum* to grow under stress, namely its heat tolerance.

Our results exploring the culturability and metabolic viability of *M. hassiacum* suspensions exposed to high temperatures indicate that, in principle, heat treatment at or above 70°C for at least 5 min should be enough to inactivate planktonic *M. hassiacum*, while temperatures around or below 65°C are clearly insufficient for this purpose (Figure 3.3). Shortly after publishing these results (Alarico et al., 2020), another study was published which found that 1-min treatments at 70°C or 75° were both insufficient to completely kill *M. hassiacum* in a washer-disinfector and complete killing after 1 min was only achieved at 80°C (Haas et al., 2020), complementing our findings. Although our thermotolerance results suggest that *M. hassiacum* progressively loses viability when exposed to 65°C over long periods of time (Figure 3.3), we observed slow but sustained growth at this temperature in liquid medium over several days (Figure 3.2A,B), which could suggest that some cells in the population survive and undergo the necessary adaptations to grow at this temperature. Results with *M. avium*, a common cause of

pulmonary NTM disease often reported to colonise household plumbing, show that this species undergoes an adaptative, reversible response to high-temperatures, with cells grown at 42°C being significantly more resistant to subsequent exposure to 60°C and 65°C when compared with cells grown at 25°C (Guenette et al., 2020). While the mechanisms behind this adaptative response are still not fully understood, temperature-induced changes to the structure and fatty acid composition of the mycobacterial cell wall have been observed in several species and could play a key role in heat adaptation (Alibaud et al., 2010; Baba et al., 1989; Kremer et al., 2002; Toriyama et al., 1980). Moreover, the unique polysaccharide MGLP, postulated to modulate fatty acid biosynthesis, has also been implicated in thermal adaptation in mycobacteria as suggested by the observation that mutants deficient in MGLP synthesis become temperature-sensitive (Kaur et al., 2009; Stadthagen et al., 2007). Regardless of the mechanisms involved, NTM heat adaptation should be taken into account when validating and implementing thermal disinfection protocols.

As a genuine thermophile, *M. hassiacum* is also of interest as a source of thermostable and tractable mycobacterial proteins. Rational drug design, one of the many strategies for the development of new antimycobacterial therapies, requires the identification of promising targets and the availability of their high-resolution three-dimensional structures to serve as templates against which inhibitors can be optimised (Baugh et al., 2015). Often, the poor solubility and crystallisation properties of enzymes of interest hinders the acquisition of these high-resolution structures and, in these cases, the use of homologues from other organisms with high sequence identity and conserved active site shape and chemistry can be useful (Baugh et al., 2015). Since thermostability is often correlated with higher solubility and increased amenability to crystallisation, and thermophiles are a prime source of thermostable enzymes, the existence of thermophilic mycobacteria is an obvious opportunity (Edwards et al., 2012). In this work we chose GpgS, proposed to be essential for *M. tuberculosis* and critical for *M. smegmatis* growth at supraoptimal temperatures (Kaur et al., 2009; Minato et al., 2019), as a proof of concept. We further established that possessing an active GpgS is crucial for mycobacteria by demonstrating that even *M. leprae*, which has undergone extensive genome reduction and pseudogenisation (Cole et al., 2001), maintains a functional GpgS (Figure 3.4). However, the *M. leprae* GpgS expressed poorly in *E. coli*, likely due to its low solubility and stability.

On the other hand, the *M. hassiacum* GpgS displayed very high expression levels and could be concentrated up to 60 mg/ml without precipitating, indicating high

solubility. It was optimally active at 60°C and retained about 40% of maximal activity at 70°C (Figure 3.6B), displayed a half-life of days even at 50°C (Figure 3.7A) and high melting temperatures (Figure 3.7B), highlighting its thermostability. These results are in contrast with what has been reported for *Mt*GpgS, which is optimally active at 45°C and displays an half-life of only a few hours at 37°C (Empadinhas et al., 2008). As expected, this increased stability translated into favourable crystallisation properties for *Mh*GpgS, which produced well-diffracting crystals in four different crystallisation buffers spanning acidic, neutral and basic conditions. The highest resolution structure obtained was solved to 1.11 Å, and the lowest to 2.13 Å (Table 3.1). Among all *Mh*GpgS structures analysed in this work, an average resolution of 1.95 Å (median of 1.51 Å) was seen, significantly higher than the values obtained for all structures of mycobacterial GpgSs reported to date (resolutions between 1.80-3.23 Å, with an average of 2.56 Å and median of 2.59 Å). Sharing 73% sequence identity with *Mt*GpgS, a conserved active site (Figure 3.5) and a nearly identical overall structure (Figure 3.8B), *Mh*GpgS could indeed be a good surrogate template for drug design efforts targeting this essential enzyme. For this purpose, however, it is also essential to understand how the two homologues differ from each other.

The biochemical characterisation of *Mh*GpgS (Figure 3.6B-D) revealed that the main difference between this enzyme and the previously characterised *M. tuberculosis* homologue (Empadinhas et al., 2008) is its optimal temperature and thermostability, since no relevant differences in substrate specificity, optimal pH and activation by cations were observed. Numerous studies have compared thermophilic and mesophilic homologues to understand the structural basis of protein thermostability and, while many contributing factors have been identified, no universal rule exists and each protein seems to have evolved its own set of stabilising modifications leading to thermostability (Petsko, 2001). In the case of *Mh*GpgS, it was possible to identify only a few differences that could justify its different properties when compared with *Mt*GpgS, mostly related to oligomerisation (Figure 3.9). Interactions at the dimer-dimer interface seem to be stronger in *Mh*GpgS, not only through specific amino acid modifications that contribute to better packing, but also through an additional α-helix at the N-terminal, absent in its mesophilic counterparts from *M. tuberculosis* and *M. avium paratuberculosis*. This α-helix stabilises the N-terminal, contributes to oligomerisation through its interactions with a C-terminal β-hairpin of the opposite monomer, and packs against the catalytic core of the enzyme stabilising it as well. Increased secondary structure content has been previously identified as one of the contributing factors for increased protein thermostability (Ding et al., 2012),

and for example the lactate dehydrogenase from the hyperthermophile *Thermotoga maritima* is also stabilised by an additional α-helix which is not present in homologues from mesophilic organisms (Auerbach et al., 1998). Interestingly, an N-terminal α-helix equivalent to the one observed in *Mh*GpgS is also present in all published three-dimensional structures of the *R. xylanophilus* homologue (PDB entries: 3FIY, 3KIA, 3O3P), equally packing against the core of the monomer and participating in the dimer-dimer interface, namely through contacts with the C-terminal β-hairpin (Empadinhas et al., 2011). *R. xylanophilus* is a thermophile with optimal growth temperature of 60°C and its MpgS/GpgS is optimally active at 70-75°C. It is thus tempting to speculate that stabilisation of this N-terminal segment is a contributing factor to the thermostability of both the *M. hassiacum* and *R. xylanophilus* enzymes, although further studies are needed to test this hypothesis.

An increase in stabilising interactions, necessary for enzymes to maintain an active conformation at high temperatures, tends to be associated with an increase in overall rigidity and a lack of flexibility at lower temperatures which can compromise the enzyme's ability to carry out its catalytic cycle below its optimal temperature (Radestock & Gohlke, 2011). Although previous studies with *Mt*GpgS already hinted at an overall rigid enzyme, undergoing very few conformational changes upon substrate binding (Pereira et al., 2008; Urresti et al., 2012), our results with *Mh*GpgS not only corroborate these previous reports but point to an even more rigid enzyme which maintains its conformation nearly unaltered regardless of pH or ligand binding (Figures 3.10, 3.11 and 3.12). In the case of *Mh*GpgS, this rigidity seems to extend to the flexible *L* loop, observed to adopt alternate conformations in *Mt*GpgS (Albesa-Jové, Romero-García, et al., 2017), which we could not confirm in the crystal structures of *Mh*GpgS (Figure 3.10E). It is not obvious which could be the sequence/structural basis for such a difference in *L* loop flexibility between both enzymes. It should be noted that all crystal structures of *Mh*GpgS were obtained at 20°C and reflect its properties at this temperature. Structural studies carried out at temperatures closer to the enzyme's optimum might show more evidence of flexibility. Taken together, the evidence suggests that the mycobacterial GpgS possesses an inherently stable and rigid fold which is made more rigid by stabilising modifications in *M. hassiacum*. A strategy to compensate increased rigidity at low temperatures in enzymes from thermophilic organisms is through adaptive mutations conferring local flexibility in regions of the protein important for catalysis, maintaining a balance between overall structural stability and local conformational flexibility (Radestock & Gohlke, 2011).

Accordingly, comparing the amino acid sequence of *Mt*GpgS and *Mh*GpgS revealed several flexibility-conferring substitutions, all situated in the vicinity of the enzyme's active site, and which could have the mentioned balancing role, ensuring the enzyme retains enough flexibility at suboptimal temperatures. Nevertheless, *Mh*GpgS retains only 36% of maximal activity at 40°C (Figure 3.6B). The fact that the main effect of *gpgS* disruption in an *M. smegmatis* mutant is hindering growth at supraoptimal temperatures (Kaur et al., 2009) suggests that maintaining high levels of GpgS activity may be more important to mycobacteria at higher temperatures, possibly justifying trading off activity at low temperatures for thermostability.

An interesting observation stemming from this study is that, unlike previously reported for other mycobacterial GpgSs (Albesa-Jové, Romero-García, et al., 2017; Fulton et al., 2008; Pereira et al., 2008), *Mh*GpgS can coordinate the metal cation without the participation of the *L* loop's conserved histidine (His252 in *Mh*GpgS). In fact, the *L* loop in *Mh*Gpgs is consistently seen in what is considered to be the open/inactive conformation, with His252 too distant from the active site to coordinate the cation. Nonetheless, several *Mh*GpgS structures show a coordinated magnesium ion at the active site (Figures 3.11D, 3.12C and 3.13B). In *Mt*GpgS, mutation of this histidine by alanine (His258Ala) had a negligible effect on the turnover number but substantially increased K_m for both substrates, resulting in a dramatic drop in efficiency (Kumar et al., 2014). This observation, together with crystallographic evidence invariably showing this histidine as part of the cation's coordination sphere, suggests it has a crucial role in catalysis. Mutagenesis studies with *Mh*GpgS would be required to elucidate if the importance of this residue for catalysis differs from what has been reported for the other studied mycobacterial GpgSs. Additionally, obtaining high-resolution structures of *Mh*GpgS in complex with both UDP-Glc and 3-PGA would also be of interest to gain further insights into the catalytic mechanism of this enzyme and to determine if binding of both substrates could bring the *L* loop into the closed/active conformation observed in *Mt*GpgS. So far, our attempts at obtaining such a complex with *Mh*GpgS have failed as we could never detect 3-PGA in crystals obtained through co-crystallisation or soaking experiments.

Another unexpected finding was the puzzling electronic density found in the apo structures of *Mh*GpgS, occupying the uridine binding pocket, which despite many attempts could not be assigned to any known molecule present during protein production and purification or in buffers used for storage, crystallisation and cryoprotection, leaving the hypothesis of an *E. coli* metabolite which co-purified with GpgS as the only other

explanation to explore (Figure 3.14). However, even the identification of candidate compounds in an *E. coli* metabolome database did not yield satisfactory results, and the identity of the electronic density remains elusive. Since we also conducted, in parallel, a close inspection of the genomic context of GpgS, it was impossible not to notice the similarity between the overall shape of the unidentified electronic density and the chemical structure of two compounds, *p*ABA and *p*HBA, associated with two GpgS genomic neighbours, FolP2 and LOG (Figure 3.15).

The function of FolP2 is unknown, since despite its homology with the dihydropteroate synthase FolP1 involved in folate biosynthesis, FolP2 lacks this activity (Gengenbacher et al., 2008). Nonetheless, a role in folate metabolism is anticipated given that FolP2 has been shown to bind the antifolate drug dapsone, a *p*ABA analogue and FolP1 inhibitor, and that knocking out the *folp2* gene in *M. smegmatis* and *M. tuberculosis* increases susceptibility to sulfamethoxazole, another antifolate targeting FolP1 (Gengenbacher et al., 2008; Liu et al., 2015; Wang et al., 2021). Moreover, a study found that pathogenic *Nocardia* strains can evolve resistance to the combined action of two antifolates, sulfamethoxazole and trimethoprim, through adaptative mutations in *folp2* and, unexpectedly, in *gpgS* as well (Mehta et al., 2018). How mutations in GpgS could lead to increased resistance to antifolates is perplexing and our results indicating that *p*ABA can bind the active site of *Mh*GpgS (Figure 3.16C) could be a further hint of a yet undiscovered role for GpgS in folate metabolism. Also unexpected is the function of LOG, which was shown to synthetise cytokinins in *M. tuberculosis*, demonstrating for the first time that these phytohormones are not restricted to plants and plant-associated microbes (Samanovic et al., 2015). The function of cytokinins in mycobacteria is still unclear, but exacerbation of cytokinin production seems to increase the susceptibility of *M. smegmatis* to antifolates (Guzzo et al., 2021), suggesting cytokinin and folate metabolism may be linked. Furthermore, exacerbation of cytokinin biosynthesis also leads to the accumulation of a toxic break-down product, *p*HBA (Samanovic et al., 2015), whose structural similarity to *p*ABA is interesting to note and which according to our results also binds *Mh*GpgS (Figure 3.16D). Although it is tempting to speculate that GpgS may function as a node connecting GG and MGLP biosynthesis with cytokinin signalling pathways and/or folate metabolism, the GpgS-*p*ABA and GpgS-*p*HBA interactions found in the present work are too weak to support any such conclusion. Indeed, neither compound affects the conformation (Figure 3.16A,B) or thermal stability of the enzyme (Figure 3.17A), and they display a very low inhibitory activity (Figure 3.17B,C). However, and given the

abovementioned studies which also hint at interactions between these pathways, it is an avenue of research worth pursuing.

Besides giving clues for possible links to other mycobacterial pathways, the ability of GpgS to accommodate small-molecule aromatic compounds different from its substrates in its active site may be a positive indicator of its druggability (Brown et al., 2018; Schneider, 2004) and the results presented here could be a first step in the identification and development of GpgS inhibitors. Although the inhibitory activity of *p*HBA and *p*ABA is very poor, that is to be expected given their small size, which results in a limited molecular interaction surface between the ligand and the target (Jhoti et al., 2013; Negatu et al., 2020). However, such small compounds, even though poor inhibitors, are often attractive starting points for lead optimisation (Jhoti et al., 2013) and, in the specific case of drugs with antimycobacterial activity, there is an historically important role for unusually small compounds with useful *in vivo* activities (Gopal & Dick, 2014; Negatu et al., 2018).

In conclusion, the findings reported here confirm the ability of *M. hassiacum* to survive near pasteurization temperatures, making it the most thermotolerant species of mycobacteria known. *M. hassiacum* is rarely isolated but may be more ubiquitous in the environment than it is assumed, since a recent study found it to be one of the most abundant species in showerhead biofilms by cultivation-independent methods, while cultivation-dependent methods failed to detect it (Gebert et al., 2018). Although a rare cause of human infection and less pathogenic than other opportunistic mycobacteria, it may represent a good indicator of the effectiveness of disinfection strategies. Moreover, *M. hassiacum* is a promising source of stable mycobacterial proteins, facilitating the study of essential enzymes with the goal of developing new therapeutic strategies to control mycobacterial infections. Using GpgS as an example, the results presented here offer both high-resolution structures of this essential enzyme which can be used as templates for future rational drug design efforts, and also potential starting compounds for lead optimisation. Finally, an intriguing relationship between GpgS, the obscure FolP2 and the still poorly understood cytokinin signalling pathway in mycobacteria is tentatively suggested, but future work is required to confirm it and elucidate its nature.

Acknowledgements and work contributions

Growth of *M. hassiacum* in different media and under different temperature and pH conditions were partly performed by Susana Alarico. Cloning of the *M. hassiacum gpgS* gene, determination of *Mh*GpgS substrate specificity, temperature, pH and cation dependence profiles and kinetic parameters were performed by Susana Alarico. Crystallisation of *Mh*GpgS apo structures and complexes with UDP and UDP-Glc, and thermal shift experiments in the presence of UDP, UDP-Glc and 3-PGA, were performed by Alexandra Silva. Data collection, structure determination and refinement were performed in collaboration with Alexandra Silva, J. Antonio Manso and Pedro Pereira.

The work in this chapter was supported by Fundo Europeu de Desenvolvimento Regional (FEDER) through the COMPETE 2020-Operational Programme for Competitiveness and Internationalisation (POCI), PORTUGAL 2020 and by Portuguese funds through Fundação para a Ciência e a Tecnologia (FCT) in the framework of projects PTDC/BTM-TEC/29221/2017 (POCI-01–0145-FEDER-029221), PTDC/BIA-MIC/0122/2021 and UID/BIM/4293 (POCI-01–0145-FEDER-007274).

Chapter 4 – Concluding Remarks and Future Directions

Historically, serendipity has been credited with many important scientific discoveries, and its role in basic research is undeniable (Yaqub, 2018). In words widely credited to Isaac Asimov, 'The most exciting phrase to hear in science, the one that heralds new discoveries, is not "Eureka!" but "That's funny …" or "That's odd…"'. In the course of the research presented in this thesis, the study of two potential therapeutic targets in mycobacteria led to unforeseen discoveries or to the formulation of new, unexpected hypothesis. In a way, both studies included "odd" but hypothesis-generating moments, namely identifying a maltokinase (Mak) paralogue within novel biosynthetic gene clusters (BGCs) and finding evidence of an unknown and unexpected molecule in the active site of glucosyl-3-phosphoglycerate synthase (GpgS). Despite these surprising results, the initial questions were nonetheless answered in the course of this work: the intriguing Mak paralogue found in the genomes of some mycobacteria is not functionally redundant with the canonical Mak (Chapter 2), and the thermophilic *M. hassiacum* can indeed be a valuable source of stable mycobacterial enzymes amenable to structural studies (Chapter 3). In this chapter, the new hypothesis and the future avenues of research that resulted from this work will be presented and discussed in the context of the overall objective behind it, finding new ways to tackle the global health threat of a post-antibiotic world.

Mak, which phosphorylates maltose and is involved in a non-canonical pathway of 1,4-glucan biosynthesis, and GpgS, responsible for synthesising the important solute GG and initiating MGLP biosynthesis, have attracted significant attention due to their participation in important pathways of the particularly threatening and antibiotic-resistant *Mycobacterium* genus (Fraga et al., 2015; Nunes-Costa et al., 2017). The long treatment regimens required to treat diseases caused by mycobacteria, and the high treatment failure rate associated, underscore the urgent need to develop alternatives (Johansen et al., 2020; Kumar et al., 2017). In order to find shorter, more effective treatments with less side effects, it is crucial to i) discover novel chemical space to replenish the drug development pipelines with new molecules displaying innovative chemistry; ii) identify and describe novel targets through the in-depth study of essential metabolic pathways; and iii) find novel modes of action by studying the interactions between new candidate compounds and their targets (Kumar et al., 2017; Miethke et al., 2021). It should be noted that this strategy is not only valid against the intrinsically resistant mycobacteria, but has also been proposed as a general roadmap to counterbalance the explosion of antibiotic resistance observed in many critical pathogens (Miethke et al., 2021). The global health threat posed by the spread of antibiotic resistance

is of such magnitude that it has led to calls for an "all-out", "leaving no stone unturned" approach (Gammon, 2014; WHO, 2015).

When studying essential pathways in the quest for novel potential therapeutic targets, it is critical to understand if there are compensatory mechanisms which will play a role if that pathway is blocked. Indeed, functional redundancy is one of the strategies organisms use to confer robustness to metabolic networks, reducing the likelihood that perturbations in the system will lead to the disruption of vital metabolic pathways and cause lethality (Sambamoorthy & Raman, 2018). Duplicate genes are often a source of functional redundancy (Kafri et al., 2009), and the discovery of a duplicate Mak gene in some NTM raised the question of whether these species had a second enzyme capable of phosphorylating maltose. This is not the case, and as shown, this duplicate gene seems to represent a case of neofunctionalization and not of redundancy. Unexpectedly, this enzyme phosphorylates glucosamine and its genomic context and phylogenetic distribution suggest a role in secondary metabolism in many actinomycetes.

Structurally, this glucosamine kinase (GlcNK) is very similar to Mak, adopting a eukaryotic-like kinase fold, but key substitutions in the sugar binding pocket determine its different substrate specificity. Additionally, the N-terminal cap subdomain, which in Mak has been proposed to mediate the interaction between Mak and its metabolic partner TreS (Kermani et al., 2019; Roy et al., 2013), is present but substantially different in GlcNK. This could indicate that GlcNK also interacts with a metabolic partner through its N-terminal cap subdomain. Although genomic analyses show that GlcNK is part of different gene clusters in different species, indicating that it could participate in different pathways, there is one gene annotated as a phosphosugar isomerase of unknown function that is always adjacent to GlcNK and could hypothetically be its closest metabolic partner. Whether GlcNK and this isomerase interact in a way similar to what has been observed for Mak and the isomerase TreS requires further investigation.

In the actinomycete *Streptacidiphilus jiangxiensis*, the gene coding for GlcNK is present in a rare and oddly structured BGC which also includes a putative cyclodipeptide synthase gene, and whose architecture suggests that its product is a glycosylated diketopiperazine. Rare and oddly structured BGCs are of particular interest since their products are likely to be significantly novel, expanding known chemical space and perhaps leading to the discovery of novel modes of action (Du & van Wezel, 2018). The only glycosylated diketopiperazine of natural origin reported to date was discovered recently and found to possess antimicrobial activity, but its mode of action and biosynthetic

pathway is unknown (Chen et al., 2018). Therefore, future efforts aimed at deciphering the biosynthetic pathway in which this novel GlcNK is involved could lead to the discovery of a novel and rare secondary metabolite with potentially useful applications.

Natural product discovery is a fundamental part of drug development, since it reveals chemical space that has already been explored and "optimised" by nature to interact with biological systems and to carry out specific biological functions. Another strategy is structure-based drug design, in which the three-dimensional structure of a target is used as a template to rationally design specific inhibitors (Batool et al., 2019). The two strategies are not necessarily separate from each other: natural products often display suboptimal efficacy and can be improved by chemical modifications to the original structure (Atanasov et al., 2021), so if the molecular target of the natural product is known, rational optimisation of its chemical structure is possible. For this purpose, however, it is crucial to have three-dimensional structures of the target available at the highest resolution possible (Baugh et al., 2015). Despite very recent advances showing that it is possible to reach near atomic resolution with the cryo-EM technique (Strack, 2020), X-ray crystallography is still the predominant and most accessible technique to obtain high-resolution structures of proteins, and its success is still dependent on obtaining high-quality and well-diffracting crystals which in many cases is challenging. Due to their increased stability, proteins of thermophilic organisms are often more amenable to *in vitro* manipulation and to crystallisation compared to proteins from mesophilic human pathogens. In the case of mycobacteria, thermophilic species such as *M. hassiacum* can be useful as a source of stable proteins, which are more amenable to structural studies but still share high sequence identity and an identical active site with their homologues from pathogenic mycobacteria such as *M. tuberculosis*.

Indeed, the GpgS homologue from *M. hassiacum* yielded X-ray crystallographic structures of much higher resolution than those which had been obtained for the pathogenic species *M. tuberculosis* and *M. avium paratuberculosis*. This, together with the ease of obtaining high-quality crystals in different crystallisation conditions, allowed for an in-depth study of this enzyme responsible for the first step of MGLP biosynthesis. The mycobacterial GpgS had previously been shown to be a conformationally stable and rigid protein since binding of its substrates led to almost no changes in conformation. This was also confirmed with the *M. hassiacum* homologue and it was further shown that pH variation also does not lead to significant conformational changes. Furthermore, it was shown that the UDP-glucose binding mode is conserved in *M. hassiacum*, suggesting its

suitability as a surrogate for future drug design studies. However, the inability to obtain complexes with the sugar acceptor, 3-phosphoglycerate, hindered further conclusions.

A surprising and intriguing result obtained from the high-resolution structures of *Mh*GpgS at neutral pH was the occupation of the uridine binding pocket by an unknown and apparently benzoic-like molecule. The identity of this molecule, likely an *E. coli* metabolite that co-purified with the recombinant enzyme, could not be elucidated, but the similarity between its predicted structure and the chemical structure of two metabolites (*p*ABA and *p*HBA) associated with GpgS's closest genomic neighbours (FolP2 and LOG) prompted co-crystallisation and soaking experiments to determine if these two metabolites could interact with the active site of GpgS. Indeed, both molecules could be detected at the active site of GpgS and showed weak inhibitory activity, to be expected given their small size and consequent limited molecular interaction surface between ligand and protein. Nonetheless, these results are intriguing. On the one hand, *p*ABA is an intermediate in the folate biosynthesis pathway and several structural analogues are used clinically as antimicrobials due to their ability to block folate biosynthesis (commonly called antifolates). A study in *Nocardia* recently showed that specific mutations in the *gpgS* gene conferred resistance to antifolates, but so far no explanation for this intriguing link has been found (Mehta et al., 2018). On the other hand, *p*HBA is a toxic cytokinin break-down product (Samanovic et al., 2015) and a recent study in *M. smegmatis* showed that exacerbation of cytokinin production increases susceptibility to antifolates through an also unknown mechanism (Guzzo et al., 2021). The interaction between these compounds and GpgS could represent another piece of the puzzle linking folate and cytokinin metabolism, curiously through an enzyme implicated in stress adaptation (Nunes-Costa et al., 2017), but the interaction described here seems to be too weak to justify any *in vivo* effect. It is possible, though, that other metabolites involved in the folate and/or cytokinin pathways bind GpgS with greater affinity than *p*ABA and *p*HBA, and this is an avenue of research that should be pursued in future studies.

Currently, many questions remain to be answered, not just regarding the possible relationship between GpgS and its genomic neighbours LOG and FolP2, but also about the neighbours themselves. Although LOG has been demonstrated to be involved in cytokinin biosynthesis (Samanovic et al., 2015), it is still unclear what role these "phyto"-hormones serve in mycobacteria. It is also unclear which biochemical reaction is catalysed by FolP2, since several lines of evidence suggest it is important for antifolate resistance (Liu et al., 2015; Mehta et al., 2018; Wang et al., 2021) but it seems to be unable to catalyse the same reaction as its paralogue and antifolate target FolP1 (Gengenbacher et al., 2008), despite

strict conservation of the key catalytic residues and only a few differences in the residues lining the substrate binding pockets. Addressing these questions, and further investigating possible links with GG and MGLP biosynthesis through GpgS, is crucial to better understand mycobacterial physiology and hopefully finding innovative therapeutic approaches or new ways to overcome resistance to antifolates, a clinically important class of antimycobacterial drugs.

In addition to being a source of stable mycobacterial proteins for in-depth structural studies which may help fight pathogenic mycobacteria, as shown with GpgS, *M. hassiacum* is also of interest as a threat in its own right, since it is a rare cause of opportunistic infections in humans (Deinhardt-Emmer et al., 2018; Jiang et al., 2013; Salzer et al., 2020; Schröder et al., 1997; Tortoli et al., 1998), and as a possible indicator to validate thermal disinfection procedures and help ensure that mycobacteria do not survive disinfection due to their unusual resistance to abiotic stresses.

The genus *Mycobacterium* includes some of the most ancient pathogens known to humankind and yet, both the prevention and treatment of infections caused by these bacteria remain a challenge. Prevention is hindered by the lack of adequate disinfection protocols especially in water distribution systems, which leads to the selective enrichment of NTM in tap water as other groups of bacteria, more susceptible to chlorination and high temperatures, are eliminated. This results in increased exposure to NTM, which in populations with increasingly debilitated immune systems, due to the rise in life expectancy and the high prevalence of chronic illnesses, probably explains the rising trend of NTM infections. Treatment is hindered by the intrinsic resistance of mycobacteria to most antibiotics, the emergence of resistance against those which used to work, and the long duration of current treatment regimens which include cocktails of antibiotics with serious side effects. Therefore, increased awareness is needed to adequately prevent these infections, and an intensification of research that can bring about new therapeutic options is also crucial.

Several aspects of mycobacterial physiology and metabolism are still not understood, and they may hold the key to some of the breakthroughs that are desperately needed in this field. In some cases, as shown by the findings reported here, the interconnection between different pathways and the phylogenetic proximity between human pathogens and antibiotic-producing actinomycetes can lead to surprising, often puzzling discoveries. Nevertheless, keeping the “leaving no stone unturned” approach in mind, unexpected findings can still point to new and promising avenues of research.

References

Abubakar, I., Zignol, M., Falzon, D., Raviglione, M., Ditiu, L., Masham, S., Adetifa, I., Ford, N., Cox, H., Lawn, S. D., Marais, B. J., McHugh, T. D., Mwaba, P., Bates, M., Lipman, M., Zijenah, L., Logan, S., McNerney, R., Zumla, A., … Zumla, A. (2013). Drug-resistant tuberculosis: Time for visionary political leadership. *The Lancet Infectious Diseases*, *13*(6), 529–539. https://doi.org/10.1016/S1473-3099(13)70030-6

Adams, P. D., Afonine, P. V, Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 2), 213–221. https://doi.org/10.1107/S0907444909052925

Aguilar-Ayala, D. A., Cnockaert, M., André, E., Andries, K., Gonzalez-Y-Merchand, J. A., Vandamme, P., Palomino, J. C., & Martin, A. (2017). *In vitro* activity of bedaquiline against rapidly growing nontuberculous mycobacteria. *Journal of Medical Microbiology*, *66*(8), 1140–1143. https://doi.org/10.1099/jmm.0.000537

Ahangar, M. S., Furze, C. M., Guy, C. S., Cooper, C., Maskew, K. S., Graham, B., Cameron, A. D., & Fullam, X. E. (2018). Structural and functional determination of homologs of the *Mycobacterium tuberculosis* N-acetylglucosamine-6-phosphate deacetylase (NagA). *Journal of Biological Chemistry*, *293*(25), 9770–9783. https://doi.org/10.1074/jbc.RA118.002597

Ahuja, L. G., Kornev, A. P., Mcclendon, C. L., Veglia, G., & Taylor, S. S. (2017). Mutation of a kinase allosteric node uncouples dynamics linked to phosphotransfer. *PNAS*, *114*(6), E931–E940. https://doi.org/10.1073/pnas.1620667114

Alarico, S., Costa, M., Sousa, M. S., Maranha, A., Lourenço, E. C., Faria, T. Q., Ventura, M. R., & Empadinhas, N. (2014). *Mycobacterium hassiacum* recovers from nitrogen starvation with up-regulation of a novel glucosylglycerate hydrolase and depletion of the accumulated glucosylglycerate. *Scientific Reports*, *4*, 6766. https://doi.org/10.1038/srep06766

Alarico, S., Nunes-Costa, D., Silva, A., Costa, M., Macedo-Ribeiro, S., & Empadinhas, N. (2020). A genuine mycobacterial thermophile: *Mycobacterium hassiacum* growth, survival and GpgS stability at near-pasteurization temperatures. *Microbiology (United Kingdom)*, *166*(5), 474–483. https://doi.org/10.1099/mic.0.000898

Albesa-Jové, D., & Guerin, M. E. (2016). The conformational plasticity of glycosyltransferases. *Current Opinion in Structural Biology*, *40*, 23–32. https://doi.org/10.1016/j.sbi.2016.07.007

Albesa-Jové, D., Mendoza, F., Rodrigo-Unzueta, A., Gomollón-Bel, F., Cifuente, J. O., Urresti, S., Comino, N., Gómez, H., Romero-García, J., Lluch, J. M., Sancho-Vaello, E., Biarnés, X., Planas, A., Merino, P., Masgrau, L., & Guerin, M. E. (2015). A Native Ternary Complex Trapped in a Crystal Reveals the Catalytic Mechanism of a Retaining Glycosyltransferase. *Angewandte Chemie - International Edition*, *54*(34), 9898–9902. https://doi.org/10.1002/anie.201504617

Albesa-Jové, D., Romero-García, J., Sancho-Vaello, E., Contreras, F. X., Rodrigo-Unzueta, A., Comino, N., Carreras-González, A., Arrasate, P., Urresti, S., Biarnés, X., Planas, A., & Guerin, M. E. (2017). Structural Snapshots and Loop Dynamics along the Catalytic Cycle of Glycosyltransferase GpgS. *Structure*, *25*(7), 1034–1044. https://doi.org/10.1016/j.str.2017.05.009

Albesa-Jové, D., Sainz-Polo, M. A., Marina, A., & Guerin, M. E. (2017). Structural Snapshots of α-1,3-Galactosyltransferase with Native Substrates: Insight into the Catalytic Mechanism of Retaining Glycosyltransferases. *Angewandte Chemie - International Edition*, *56*(47), 14853–14857. https://doi.org/10.1002/anie.201707922

Albuquerque-Wendt, A., Hütte, H. J., Buettner, F. F. R., Routier, F. H., & Bakker, H. (2019). Membrane topological model of glycosyltransferases of the GT-C superfamily. *International Journal of Molecular Sciences*, *20*(19), 4842. https://doi.org/10.3390/ijms20194842

Alderwick, L. J., Harrison, J., Lloyd, G. S., & Birch, H. L. (2015). The Mycobacterial Cell Wall - Peptidoglycan and Arabinogalactan. *Cold Spring Harbor Perspectives in Medicine*, *5*(8), a021113. https://doi.org/10.1101/cshperspect.a021113

Alibaud, L., Alahari, A., Trivelli, X., Ojha, A. K., Hatfull, G. F., Guerardel, Y., & Kremer, L. (2010). Temperature-dependent regulation of mycolic acid cyclopropanation in saprophytic mycobacteria: Role of the *Mycobacterium smegmatis* 1351 gene (MSMEG-1351) in cis-cyclopropanation of α-mycolates. *Journal of Biological Chemistry*, *285*(28), 21698–21707. https://doi.org/10.1074/jbc.M110.125724

Ames, B. (1966). Assay of inorganic phosphate, total phosphate and phosphatases. *Methods Enzymol*, *8*, 115–118.

André, I., Potocki-Véronèse, G., Barbe, S., Moulis, C., & Remaud-Siméon, M. (2014). CAZyme discovery and

design for sweet dreams. *Current Opinion in Chemical Biology*, *19*, 17–24. https://doi.org/10.1016/j.cbpa.2013.11.014

Ang, D. L., Hoque, M. Z., Hossain, M. A., Guerriero, G., Berni, R., Hausman, J. F., Bokhari, S. A., Bridge, W. J., & Siddiqui, K. S. (2021). Computational analysis of thermal adaptation in extremophilic chitinases: The achilles' heel in protein structure and industrial utilization. *Molecules*, *26*(3), 707. https://doi.org/10.3390/molecules26030707

Arbues, A., Lugo-Villarino, G., Neyrolles, O., Guilhot, C., & Astarie-Dequeker, C. (2014). Playing hide-and-seek with host macrophages through the use of mycobacterial cell envelope phthiocerol dimycocerosates and phenolic glycolipids. *Frontiers in Cellular and Infection Microbiology*, *4*, 173. https://doi.org/10.3389/fcimb.2014.00173

Aslam, M., Takahashi, N., Matsubara, K., Imanaka, T., Kanai, T., & Atomi, H. (2018). Identification of the glucosamine kinase in the chitinolytic pathway of *Thermococcus kodakarensis*. *Journal of Bioscience and Bioengineering*, *125*(3), 320–326. https://doi.org/10.1016/j.jbiosc.2017.10.005

Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., Orhan, I. E., Banach, M., Rollinger, J. M., Barreca, D., Weckwerth, W., Bauer, R., Bayer, E. A., Majeed, M., Bishayee, A., Bochkov, V., Bonn, G. K., Braidy, N., Bucar, F., Cifuentes, A., D'Onofrio, G., Bodkin, M., … Supuran, C. T. (2021). Natural products in drug discovery: advances and opportunities. *Nature Reviews Drug Discovery*, *20*(3), 200–216. https://doi.org/10.1038/s41573-020-00114-z

Auerbach, G., Ostendorp, R., Prade, L., Korndörfer, I., Dams, T., Huber, R., & Jaenicke, R. (1998). Lactate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima*: The crystal structure at 2.1 Å resolution reveals strategies for intrinsic protein stabilization. *Structure*, *6*(6), 769–781. https://doi.org/10.1016/S0969-2126(98)00078-1

Aung, H. L., Dixon, L. L., Smith, L. J., Sweeney, N. P., Robson, J. R., Berney, M., Buxton, R. S., Green, J., & Cook, G. M. (2015). Novel regulatory roles of cAMP receptor proteins in fast-growing environmental mycobacteria. *Microbiology (United Kingdom)*, *161*(3), 648–661. https://doi.org/10.1099/mic.0.000015

Baba, T., Kaneda, K., Kusunose, E., Kusunose, M., & Yano, I. (1989). Thermally adaptive changes of mycolic acids in *Mycobacterium smegmatis*. *Journal of Biochemistry*, *106*(1), 81–86. http://www.ncbi.nlm.nih.gov/pubmed/2777756

Baker, O., Lee, O. Y. C., Wu, H. H. T., Besra, G. S., Minnikin, D. E., Llewellyn, G., Williams, C. M., Maixner, F., O'Sullivan, N., Zink, A., Chamel, B., Khawam, R., Coqueugniot, E., Helmer, D., Le Mort, F., Perrin, P., Gourichon, L., Dutailly, B., Pálfi, G., … Dutour, O. (2015). Human tuberculosis predates domestication in ancient Syria. *Tuberculosis*, *95 Suppl 1*, S4–S12. https://doi.org/10.1016/j.tube.2015.02.001

Balagon, M. V, Walsh, D. S., Tan, P. L., Cellona, R. V, Abalos, R. M., Tan, E. V, Fajardo, T. T., Watson, J. D., & Walsh, G. P. (2000). Improvement in psoriasis after intradermal administration of heat-killed *Mycobacterium vaccae*. *International Journal of Dermatology*, *39*(1), 51–58. https://doi.org/10.1046/j.1365-4362.2000.00862.x

Bansal-Mutalik, R., & Nikaido, H. (2014). Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides. *PNAS*, *111*(13), 4958–4963. https://doi.org/10.1073/pnas.1403078111

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., & Milo, R. (2011). The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, *50*(21), 4402–4410. https://doi.org/10.1021/bi2002289

Bastidas, A. C., Deal, M. S., Steichen, J. M., Guo, Y., Wu, J., & Taylor, S. S. (2013). Phosphoryl transfer by protein kinase A is captured in a crystal lattice. *J Am Chem Soc*, *135*(12), 4788–4798. https://doi.org/10.1021/ja312237q

Batool, M., Ahmad, B., & Choi, S. (2019). A structure-based drug discovery paradigm. *International Journal of Molecular Sciences*, *20*(11), 2783. https://doi.org/10.3390/ijms20112783

Baugh, L., Phan, I., Begley, D. W., Clifton, M. C., Armour, B., Dranow, D. M., Taylor, B. M., Muruthi, M. M., Abendroth, J., Fairman, J. W., Fox III, D., Dieterich, S. H., Staker, B. L., Gardberg, A. S., Choi, R., Hewitt, S. N., Napuli, A. J., Myers, J., Barrett, L. K., … Myler, P. J. (2015). Increasing the Structural Coverage of Tuberculosis Drug Targets. *Tuberculosis (Edinb)*, *95*(2), 142–148. https://doi.org/10.1016/j.tube.2014.12.003.Increasing

Behr, M. A., Edelstein, P. H., & Ramakrishnan, L. (2018). Revisiting the timetable of tuberculosis. *BMJ*, *362*, k2738. https://doi.org/10.1136/bmj.k2738

Behrends, V., Williams, K. J., Jenkins, V. A., Robertson, B. D., & Bundy, J. G. (2012). Free glucosylglycerate is a novel marker of nitrogen stress in *Mycobacterium smegmatis*. *Journal of Proteome Research*, *11*(7), 3888–3896. https://doi.org/10.1021/pr300371b

Belknap, K. C., Park, C. J., Barth, B. M., & Andam, C. P. (2020). Genome mining of biosynthetic and

chemotherapeutic gene clusters in *Streptomyces* bacteria. *Scientific Reports*, *10*, 2003. https://doi.org/10.1038/s41598-020-58904-9

Berini, F., Marinelli, F., & Binda, E. (2020). *Streptomycetes*: Attractive Hosts for Recombinant Protein Production. *Frontiers in Microbiology*, *11*, 1958. https://doi.org/10.3389/fmicb.2020.01958

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. https://doi.org/10.1093/nar/28.1.235

Bhatti, A. A., Haq, S., & Bhat, R. A. (2017). Actinomycetes benefaction role in soil and plant health. *Microbial Pathogenesis*, *111*, 458–467. https://doi.org/10.1016/j.micpath.2017.09.036

Bibb, M. J. (2005). Regulation of secondary metabolism in streptomycetes. *Current Opinion in Microbiology*, *8*(2), 208–215. https://doi.org/10.1016/j.mib.2005.02.016

Bichsel, Y., & Von Gunten, U. (2000). Formation of iodo-trihalomethanes during disinfection and oxidation of iodide-containing waters. *Environmental Science and Technology*, *34*(13), 2784–2791. https://doi.org/10.1021/es9914590

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., & Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*, *47*(W1), W81–W87. https://doi.org/10.1093/nar/gkz310

Bloch, K., & Vance, D. (1977). Control mechanisms in the synthesis of saturated fatty acids. *Annual Review of Biochemistry*, *46*, 263–298. https://doi.org/10.1146/annurev.bi.46.070177.001403

Bloemberg, G. V, Keller, P. M., Stucki, D., Trauner, A., Borrell, S., Latshang, T., Coscolla, M., Rothe, T., Hömke, R., Ritter, C., Feldmann, J., Schulthess, B., Gagneux, S., & Böttger, E. C. (2015). Acquired Resistance to Bedaquiline and Delamanid in Therapy for Tuberculosis. In *The New England journal of medicine* (Vol. 373, Issue 20, pp. 1986–1988). https://doi.org/10.1056/NEJMc1505196

Bobek, J., Šmídová, K., & Čihák, M. (2017). A waking review: Old and novel insights into the spore germination in *Streptomyces*. *Frontiers in Microbiology*, *8*, 2205. https://doi.org/10.3389/fmicb.2017.02205

Borgman, P., Lopez, R. D., & Lane, A. L. (2019). The expanding spectrum of diketopiperazine natural product biosynthetic pathways containing cyclodipeptide synthases. *Organic and Biomolecular Chemistry*, *17*(9), 2305–2314. https://doi.org/10.1039/c8ob03063d

Borisova, M., Gaupp, R., Duckworth, A., Schneider, A., Dalügge, D., Mühleck, M., Deubel, D., Unsleber, S., Yu, W., Muth, G., Bischoff, M., Götz, F., & Mayer, C. (2016). Peptidoglycan recycling in gram-positive bacteria is crucial for survival in stationary phase. *MBio*, *7*(5), e00923-16. https://doi.org/10.1128/mBio.00923-16

Bornemann, S. (2016). α-Glucan biosynthesis and the GlgE pathway in *Mycobacterium tuberculosis*. *Biochemical Society Transactions*, *44*(1), 68–73. https://doi.org/10.1042/BST20150181

Bosch, F., & Rosich, L. (2008). The contributions of Paul Ehrlich to pharmacology: A tribute on the occasion of the centenary of his nobel prize. *Pharmacology*, *82*(3), 171–179. https://doi.org/10.1159/000149583

Braña, A. F., Fiedler, H. P., Nava, H., González, V., Sarmiento-Vizcaíno, A., Molina, A., Acuña, J. L., García, L. A., & Blanco, G. (2015). Two *Streptomyces* Species Producing Antibiotic, Antitumor, and Anti-Inflammatory Compounds Are Widespread Among Intertidal Macroalgae and Deep-Sea Coral Reef Invertebrates from the Central Cantabrian Sea. *Microbial Ecology*, *69*(3), 512–524. https://doi.org/10.1007/s00248-014-0508-0

Brennan, P., & Ballou, C. E. (1967). Biosynthesis of mannophosphoinositides by *Mycobacterium phlei*. The family of dimannophosphoinositides. *The Journal of Biochemical Chemistry*, *242*(13), 3046–3056.

Breton, C., Šnajdrová, L., Jeanneau, C., Koča, J., & Imberty, A. (2006). Structures and mechanisms of glycosyltransferases. *Glycobiology*, *16*(2), 29R-37R. https://doi.org/10.1093/glycob/cwj016

Brown-Elliott, B. A., Wallace, R. J., Tichindelean, C., Sarria, J. C., McNulty, S., Vasireddy, R., Bridge, L., Mayhall, C. G., Turenne, C., & Loeffelholz, M. (2011). Five-year outbreak of community- and hospital-acquired *Mycobacterium porcinum* infections related to public water supplies. *Journal of Clinical Microbiology*, *49*(12), 4231–4238. https://doi.org/10.1128/JCM.05122-11

Brown, E. D., & Wright, G. D. (2016). Antibacterial drug discovery in the resistance era. *Nature*, *529*(7586), 336–343. https://doi.org/10.1038/nature17042

Brown, K. K., Hann, M. M., Lakdawala, A. S., Santos, R., Thomas, P. J., & Todd, K. (2018). Approaches to target tractability assessment - a practical perspective. *MedChemComm*, *9*(4), 606–613. https://doi.org/10.1039/c7md00633k

Bryant, J. M., Grogono, D. M., Greaves, D., Foweraker, J., Roddick, I., Inns, T., Reacher, M., Haworth, C. S., Curran, M. D., Harris, S. R., Peacock, S. J., Parkhill, J., & Floto, R. A. (2013). Whole-genome sequencing to identify transmission of *Mycobacterium abscessus* between patients with cystic fibrosis: A retrospective cohort study. *The Lancet*, *381*(9877), 1551–1560. https://doi.org/10.1016/S0140-6736(13)60632-7

Bryant, J. M., Grogono, D. M., Rodriguez-Rincon, D., Everall, I., Brown, K. P., Moreno, P., Verma, D., Hill, E., Drijkoningen, J., Gilligan, P., Esther, C. R., Noone, P. G., Giddings, O., Bell, S. C., Thomson, R., Wainwright, C. E., Coulter, C., Pandey, S., Wood, M. E., … Floto, R. A. (2016). Population-level genomics identifies the emergence and global spread of a human transmissible multidrug-resistant nontuberculous mycobacterium. *Science*, *354*(6313), 751–757. https://doi.org/10.1126/science.aaf8156.Population-level

Burgess, W., Margolis, A., Gibbs, S., Duarte, R. S., & Jackson, M. (2017). Disinfectant susceptibility profiling of glutaraldehyde-resistant nontuberculous mycobacteria. *Infection Control and Hospital Epidemiology*, *38*(7), 784–791. https://doi.org/10.1017/ice.2017.75

Burk, D. L., Hon, W. C., Leung, A. K. W., & Berghuis, A. M. (2001). Structural analyses of nucleotide binding to an aminoglycoside phosphotransferase. *Biochemistry*, *40*(30), 8756–8764. https://doi.org/10.1021/bi010504p

Bzymek, K. P., Newton, G. L., Ta, P., & Fahey, R. C. (2007). Mycothiol import by *Mycobactenum smegmatis* and function as a resource for metabolic precursors and energy production. *Journal of Bacteriology*, *189*(19), 6796–6805. https://doi.org/10.1128/JB.00644-07

Cambau, E., Saunderson, P., Matsuoka, M., Cole, S. T., Kai, M., Suffys, P., Rosa, P. S., Williams, D., Gupta, U. D., Lavania, M., Cardona-Castro, N., Miyamoto, Y., Hagge, D., Srikantam, A., Hongseng, W., Indropo, A., Vissa, V., Johnson, R. C., Cauchoix, B., … Gillini, L. (2018). Antimicrobial resistance in leprosy: results of the first prospective open survey conducted by a WHO surveillance network for the period 2009-15. *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *24*(12), 1305–1310. https://doi.org/10.1016/j.cmi.2018.02.022

Cambier, C. J., Takaki, K. K., Larson, R. P., Hernandez, R. E., Tobin, D. M., Urdahl, K. B., Cosma, C. L., & Ramakrishnan, L. (2014). Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. *Nature*, *505*(7482), 218–222. https://doi.org/10.1038/nature12799.Mycobacteria

Cananzi, F. C. M., Mudan, S., Dunne, M., Belonwu, N., & Dalgleish, A. G. (2013). Long-term survival and outcome of patients originally given *Mycobacterium vaccae* for metastatic malignant melanoma. *Human Vaccines and Immunotherapeutics*, *9*(11), 2427–2433. https://doi.org/10.4161/hv.25618

Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., & Ramakrishnan, V. (2000). Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, *407*(6802), 340–348. https://doi.org/10.1038/35030019

Cereija, T. B., Alarico, S., Lourenco, E. C., Manso, J. A., Ventura, M. R., Empadinhas, N., Macedo-Ribeiro, S., & Pereira, P. J. B. (2019). The structural characterization of a glucosylglycerate hydrolase provides insights into the molecular mechanism of mycobacterial recovery from nitrogen starvation. *IUCrJ*, *6*(Pt 4), 572–585. https://doi.org/10.1107/S2052252519005372

Cereijo, A. E., Kuhn, M. L., Hernández, M. A., Ballicora, M. A., Iglesias, A. A., Alvarez, H. M., & Asencion Diez, M. D. (2021). Study of duplicated galU genes in *Rhodococcus jostii* and a putative new metabolic node for glucosamine-1P in rhodococci. *Biochimica et Biophysica Acta - General Subjects*, *1865*(1), 129727. https://doi.org/10.1016/j.bbagen.2020.129727

Cerqueira, N., Brás, N., Ramos, M. J., & Fernandes, P. A. (2012). Glycosidases – A Mechanistic Overview. In *Carbohydrates - Comprehensive Studies on Glycobiology and Glycotechnology*. IntechOpen. https://doi.org/10.5772/52019

Challis, G. L. (2008). Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology*, *154*(6), 1555–1569. https://doi.org/10.1099/mic.0.2008/018523-0

Chandra, G., Chater, K. F., & Bornemann, S. (2011). Unexpected and widespread connections between bacterial glycogen and trehalose metabolism. *Microbiology*, *157*(6), 1565–1572. https://doi.org/10.1099/mic.0.044263-0

Chao, L., & Jongkees, S. (2019). High-Throughput Approaches in Carbohydrate-Active Enzymology: Glycosidase and Glycosyl Transferase Inhibitors, Evolution, and Discovery. *Angewandte Chemie - International Edition*, *58*(37), 12750–12760. https://doi.org/10.1002/anie.201900055

Cheek, S., Zhang, H., & Grishin, N. V. (2002). Sequence and Structure Classification of Kinases. *Journal of Molecular Biology*, *320*(4), 855–881. https://doi.org/10.1016/S0022-2836(02)00538-7

Chen, H., Thomas, M. G., Hubbard, B. K., Losey, H. C., Walsh, C. T., & Burkart, M. D. (2000). Deoxysugars in

glycopeptide antibiotics: Enzymatic synthesis of TDP-L-epivancosamine in chloroeremomycin biosynthesis. *PNAS*, *97*(22), 11942–11947. https://doi.org/10.1073/pnas.210395097

Chen, S., Zhang, D., Chen, M., Zhang, Z., & Lian, X.-Y. (2018). A rare diketopiperazine glycoside from marine-sourced *Streptomyces* sp. ZZ446. *Natural Product Research*, *34*(7), 1046–1050. https://doi.org/10.1080/14786419.2018.1544978

Cheng, X., Shaltiel, S., & Taylor, S. S. (1998). Mapping substrate-induced conformational changes in cAMP-dependent protein kinase by protein footprinting. *Biochemistry*, *37*(40), 14005–14013. https://doi.org/10.1021/bi981057p

Chernov, V. M., Chernova, O. A., Mouzykantov, A. A., Lopukhov, L. L., & Aminov, R. I. (2019). Omics of antimicrobials and antimicrobial resistance. *Expert Opinion on Drug Discovery*, *14*(5), 455–468. https://doi.org/10.1080/17460441.2019.1588880

Chettri, D., Verma, A. K., & Verma, A. K. (2020). Innovations in CAZyme gene diversity and its modification for biorefinery applications. *Biotechnology Reports*, *28*, e00525. https://doi.org/10.1016/j.btre.2020.e00525

Choo, S. W., Wee, W. Y., Ngeow, Y. F., Mitchell, W., Tan, J. L., Wong, G. J., Zhao, Y., & Xiao, J. (2014). Genomic reconnaissance of clinical isolates of emerging human pathogen *Mycobacterium abscessus* reveals high evolutionary potential. *Scientific Reports*, *4*, 4061. https://doi.org/10.1038/srep04061

Choudoir, M. J., Pepe-Ranney, C., & Buckley, D. H. (2018). Diversification of secondary metabolite biosynthetic gene clusters coincides with lineage divergence in *Streptomyces*. *Antibiotics*, *7*(1), 12. https://doi.org/10.3390/antibiotics7010012

Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honoré, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R. M., Devlin, K., Duthoy, S., Feltwell, T., … Barrell, B. G. (2001). Massive gene decay in the leprosy bacillus. *Nature*, *409*(6823), 1007–1011. https://doi.org/10.1038/35059006

Cornacchia, C., Cacciatore, I., Baldassarre, L., Mollica, A., Feliciani, F., & Pinnen, F. (2012). 2,5-diketopiperazines as neuroprotective agents. *Mini Reviews in Medicinal Chemistry*, *12*(1), 2–12. https://doi.org/10.2174/138955712798868959

Cornforth, D. M., & Foster, K. R. (2013). Competition sensing: the social side of bacterial stress responses. *Nat Rev Microbiol*, *11*(4), 285–293. https://doi.org/10.1038/nrmicro2977

Costa, J., Empadinhas, N., & da Costa, M. S. (2007). Glucosylglycerate biosynthesis in the deepest lineage of the Bacteria: Characterization of the thermophilic proteins GpgS and GpgP from *Persephonella marina*. *Journal of Bacteriology*, *189*(5), 1648–1654. https://doi.org/10.1128/JB.00841-06

Costa, J., Empadinhas, N., Gonçalves, L., Lamosa, P., Santos, H., & da Costa, M. S. (2006). Characterization of the biosynthetic pathway of glucosylglycerate in the archaeon *Methanococcoides burtonii*. *Journal of Bacteriology*, *188*(3), 1022–1030. https://doi.org/10.1128/JB.188.3.1022-1030.2006

Cuthbertson, L., & Nodwell, J. R. (2013). The TetR family of regulators. *Microbiology and Molecular Biology Reviews*, *77*(3), 440–475. https://doi.org/10.1128/MMBR.00018-13

Daffé, M., Crick, D. C., & Jackson, M. (2014). Genetics of Capsular Polysaccharides and Cell Envelope (Glyco)lipids. *Microbiology Spectrum*, *2*(4), MGM2-0021–2013. https://doi.org/10.1128/microbiolspec.MGM2-0021-2013

Dalbeth, N., Yeoman, S., Dockerty, J. L., Highton, J., Robinson, E., Tan, P. L., Herman, D., & McQueen, F. M. (2004). A randomised placebo controlled trial of delipidated, deglycolipidated *Mycobacterium vaccae* as immunotherapy for psoriatic arthritis. *Annals of the Rheumatic Diseases*, *63*(6), 718–722. https://doi.org/10.1136/ard.2003.007104

Dalgleish, A. G., Stebbing, J., JA Adamson, D., Arif, S. S., Bidoli, P., Chang, D., Cheeseman, S., Diaz-Beveridge, R., Fernandez-Martos, C., Glynne-Jones, R., Granetto, C., Massuti, B., McAdam, K., McDermott, R., Munoz Martin, A. J., Papamichael, D., Pazo-Cid, R., Vieitez, J. M., Zaniboni, A., … Mudan, S. S. (2016). Randomised, open-label, phase II study of gemcitabine with and without IMM-101 for advanced pancreatic cancer. *British Journal of Cancer*, *115*(7), 789–796. https://doi.org/10.1038/bjc.2016.271

Daniel, J., Sirakova, T., & Kolattukudy, P. (2014). An Acyl-CoA synthetase in *Mycobacterium tuberculosis* involved in triacylglycerol accumulation during dormancy. *PLoS ONE*, *9*(12), e114877. https://doi.org/10.1371/journal.pone.0114877

de Andrade, K. V. F., Silva Nery, J., Moreira Pescarini, J., Ramond, A., de Souza Teles Santos, C. A., Ichihara, M. Y., Fernandes Penna, M. L., B Brickley, E., C Rodrigues, L., Smeeth, L., L Barreto, M., Martins Pereira, S., & Oliveira Penna, G. (2019). Geographic and socioeconomic factors associated with leprosy treatment default:

An analysis from the 100 Million Brazilian Cohort. *PLoS Neglected Tropical Diseases*, *13*(9), e0007714. https://doi.org/10.1371/journal.pntd.0007714

de Carvalho, C. C. C. R., Teixeira, R., & Fernandes, P. (2020). *Mycobacterium vaccae* adaptation to disinfectants and hand sanitisers, and evaluation of cross-tolerance with antimicrobials. *Antibiotics*, *9*(9), 544. https://doi.org/10.3390/antibiotics9090544

De Carvalho, L. P. S., Fischer, S. M., Marrero, J., Nathan, C., Ehrt, S., & Rhee, K. Y. (2010). Metabolomics of *Mycobacterium tuberculosis* reveals compartmentalized co-catabolism of carbon substrates. *Chemistry and Biology*, *17*(10), 1122–1131. https://doi.org/10.1016/j.chembiol.2010.08.009

de Sanctis, D., Beteva, A., Caserotto, H., Dobias, F., Gabadinho, J., Giraud, T., Gobbo, A., Guijarro, M., Lentini, M., Lavault, B., Mairs, T., McSweeney, S., Petitdemange, S., Rey-Bakaikoa, V., Surr, J., Theveneau, P., Leonard, G. A., & Mueller-Dieckmann, C. (2012). ID29: a high-intensity highly automated ESRF beamline for macromolecular crystallography experiments exploiting anomalous scattering. *Journal of Synchrotron Radiation*, *19*(Pt 3), 455–461. https://doi.org/10.1107/S0909049512009715

De Smet, K. A. L., Kempsell, K. E., Gallagher, A., Duncan, K., & Young, D. B. (1999). Alteration of a single amino acid residue reverses fosfomycin resistance of recombinant MurA from *Mycobacterium tuberculosis*. *Microbiology*, *145*(11), 3177–3184. https://doi.org/10.1099/00221287-145-11-3177

Decker, B. K., & Palmore, T. N. (2014). Hospital water and opportunities for infection prevention. *Current Infectious Disease Reports*, *16*(10), 432. https://doi.org/10.1007/s11908-014-0432-y

Dedrick, R. M., Guerrero-Bustamante, C. A., Garlena, R. A., Russell, D. A., Ford, K., Harris, K., Gilmour, K. C., Soothill, J., Jacobs-Sera, D., Schooley, R. T., Hatfull, G. F., & Spencer, H. (2019). Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant *Mycobacterium abscessus*. *Nature Medicine*, *25*(5), 730–733. https://doi.org/10.1038/s41591-019-0437-z

Dedrick, R. M., Smith, B. E., Cristinziano, M., Freeman, K. G., Jacobs-Sera, D., Belessis, Y., Whitney Brown, A., Cohen, K. A., Davidson, R. M., van Duin, D., Gainey, A., Garcia, C. B., Robert George, C. R., Haidar, G., Ip, W., Iredell, J., Khatami, A., Little, J. S., Malmivaara, K., … Hatfull, G. F. (2022). Phage Therapy of *Mycobacterium* Infections: Compassionate-use of Phages in Twenty Patients with Drug-Resistant Mycobacterial Disease. *Clinical Infectious Diseases*, ciac453. https://doi.org/10.1093/cid/ciac453

Deinhardt-Emmer, S., Höring, S., Mura, C., Hillemann, D., Hermann, B., Sachse, S., Bohnert, J., & Löffler, B. (2018). First Time Isolation of *Mycobacterium hassiacum* From a Respiratory Sample. *Clinical Medicine Insights: Circulatory, Respiratory and Pulmonary Medicine*, *12*, 1–3. https://doi.org/10.1177/1179548417747529

Díaz-Mejía, J. J., Pérez-Rueda, E., & Segovia, L. (2007). A network perspective on the evolution of metabolism by gene duplication. *Genome Biology*, *8*(2), R26. https://doi.org/10.1186/gb-2007-8-2-r26

Ding, Y., Cai, Y., Han, Y., & Zhao, B. (2012). Comparison of the structural basis for thermal stability between archaeal and bacterial proteins. *Extremophiles*, *16*(1), 67–78. https://doi.org/10.1007/s00792-011-0406-z

Dolgin, E. (2019). Anticancer autophagy inhibitors attract "resurgent" interest. *Nature Reviews. Drug Discovery*, *18*(6), 408–410. https://doi.org/10.1038/d41573-019-00072-1

Donoghue, H. D., Taylor, G. M., Stewart, G. R., Lee, O. Y. C., Wu, H. H. T., Besra, G. S., & Minnikin, D. E. (2017). Positive diagnosis of ancient leprosy and tuberculosis using ancient DNA and lipid biomarkers. *Diversity*, *9*(4), 46. https://doi.org/10.3390/d9040046

Doroghazi, J. R., & Metcalf, W. W. (2013). Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics*, *14*, 611. https://doi.org/10.1186/1471-2164-14-611

Draelos, M. M., Thanapipatsiri, A., Sucipto, H., & Yokoyama, K. (2021). Cryptic phosphorylation in nucleoside natural product biosynthesis. *Nature Chemical Biology*, *17*(2), 213–221. https://doi.org/10.1038/s41589-020-00656-8

Draper, P., Khoo, K. H., Chatterjee, D., Dell, A., & Morris, H. R. (1997). Galactosamine in walls of slow-growing mycobacteria. *Biochemical Journal*, *327*(2), 519–525. https://doi.org/10.1042/bj3270519

Du, C., & van Wezel, G. P. (2018). Mining For Microbial Gems: Integrating Proteomics In The Postgenomic Natural Product Discovery Pipeline. *Proteomics*, *18*(18), e1700332. https://doi.org/10.1002/pmic.201700332

Dubeau, M. P., Poulin-Laprade, D., Ghinet, M. G., & Brzezinski, R. (2011). Properties of CsnR, the transcriptional repressor of the chitosanase gene, csnA, of *Streptomyces lividans*. *Journal of Bacteriology*, *193*(10), 2441–2450. https://doi.org/10.1128/JB.01476-10

Dulberger, C. L., Rubin, E. J., & Boutte, C. C. (2020). The mycobacterial cell envelope — a moving target. *Nature Reviews Microbiology*, *18*(1), 47–59. https://doi.org/10.1038/s41579-019-0273-7

Dunne, E. F., Burman, W. J., & Wilson, M. L. (1998). *Streptomyces pneumonia* in a Patient with Human Immunodeficiency Virus Infection: Case Report and Review of the Literature on Invasive *Streptomyces* Infections. *Clinical Infectious Diseases*, *27*(1), 93–96. https://doi.org/10.1086/322645

Dupont, C., Viljoen, A., Thomas, S., Roquet-Banères, F., Herrmann, J. L., Pethe, K., & Kremer, L. (2017). Bedaquiline inhibits the ATP synthase in *Mycobacterium abscessus* and is effective in infected zebrafish. *Antimicrobial Agents and Chemotherapy*, *61*(11), 1–15. https://doi.org/10.1128/AAC.01225-17

Dworkin, J., & Losick, R. (2005). Developmental commitment in a bacterium. *Cell*, *121*(3), 401–409. https://doi.org/10.1016/j.cell.2005.02.032

Edwards, T. E., Liao, R., Phan, I., Myler, P. J., & Grundner, C. (2012). *Mycobacterium thermoresistibile* as a source of thermostable orthologs of *Mycobacterium tuberculosis* proteins. *Protein Science*, *21*(7), 1093–1096. https://doi.org/10.1002/pro.2084

Elbein, A. D., Pan, Y. T., Pastuszak, I., & Carroll, D. (2003). New insights on trehalose: A multifunctional molecule. *Glycobiology*, *13*(4), 17R-27R. https://doi.org/10.1093/glycob/cwg047

Elbein, A. D., Pastuszak, I., Tackett, A. J., Wilson, T., & Pan, Y. T. (2010). Last step in the conversion of trehalose to glycogen: a mycobacterial enzyme that transfers maltose from maltose 1-phosphate to glycogen. *The Journal of Biological Chemistry*, *285*(13), 9803–9812. https://doi.org/10.1074/jbc.M109.033944

Elsayed, S., Genta-Jouve, G., Carrión, V., Nibbering, P., Siegler, M. A., de Boer, W., Hankemeier, T., & van Wezel, G. P. (2020). Atypical Spirotetronate Polyketides Identified in the Under-explored Genus *Streptacidiphilus*. *The Journal of Organic Chemistry*, *85*(16), 10648–10657. https://doi.org/10.1021/acs.joc.0c01210

Eltholth, M. M., Marsh, V. R., Van Winden, S., & Guitian, F. J. (2009). Contamination of food products with *Mycobacterium avium paratuberculosis*: A systematic review. *Journal of Applied Microbiology*, *107*(4), 1061–1071. https://doi.org/10.1111/j.1365-2672.2009.04286.x

Empadinhas, N., Albuquerque, L., Henne, A., Santos, H., & Da Costa, M. S. (2003). The bacterium *Thermus thermophilus*, like hyperthermophilic archaea, uses a two-step pathway for the synthesis of mannosylglycerate. *Applied and Environmental Microbiology*, *69*(6), 3272–3279. https://doi.org/10.1128/AEM.69.6.3272-3279.2003

Empadinhas, N., Albuquerque, L., Mendes, V., Macedo-Ribeiro, S., & da Costa, M. S. (2008). Identification of the mycobacterial glucosyl-3-phosphoglycerate synthase. *FEMS Microbiology Letters*, *280*(2), 195–202. https://doi.org/10.1111/j.1574-6968.2007.01064.x

Empadinhas, N., & da Costa, M. S. (2011). Diversity, biological roles and biosynthetic pathways for sugar-glycerate containing compatible solutes in bacteria and archaea. *Environmental Microbiology*, *13*(8), 2056–2077. https://doi.org/10.1111/j.1462-2920.2010.02390.x

Empadinhas, N., Pereira, P. J. B., Albuquerque, L., Costa, J., Sá-Moura, B., Marques, A. T., Macedo-Ribeiro, S., & da Costa, M. S. (2011). Functional and structural characterization of a novel mannosyl-3-phosphoglycerate synthase from *Rubrobacter xylanophilus* reveals its dual substrate specificity. *Molecular Microbiology*, *79*(1), 76–93. https://doi.org/10.1111/j.1365-2958.2010.07432.x

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D*, *66*(4), 486–501. https://doi.org/10.1107/S0907444910007493

Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N., & Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry*, *357*(2), 289–298. https://doi.org/10.1016/j.ab.2006.07.027

Ernst, J. D. (2012). The immunological life cycle of tuberculosis. *Nature Reviews Immunology*, *12*(8), 581–591. https://doi.org/10.1038/nri3259

Etienne, G., Malaga, W., Laval, F., Lemassu, A., Guilhot, C., & Daffé, M. (2009). Identification of the polyketide synthase involved in the biosynthesis of the surface-exposed lipooligosaccharides in mycobacteria. *Journal of Bacteriology*, *191*(8), 2613–2621. https://doi.org/10.1128/JB.01235-08

Falkinham, J. O. (2009a). Surrounded by mycobacteria: nontuberculous mycobacteria in the human environment. *Journal of Applied Microbiology*, *107*(2), 356–367. https://doi.org/10.1111/j.1365-2672.2009.04161.x

Falkinham, J. O. (2009b). The biology of environmental mycobacteria. *Environmental Microbiology Reports*, *1*(6), 477–487. https://doi.org/10.1111/j.1758-2229.2009.00054.x

Falkinham, J. O. (2011). Nontuberculous mycobacteria from household plumbing of patients with

nontuberculous mycobacteria disease. *Emerging Infectious Diseases*, *17*(3), 419–424. https://doi.org/10.3201/eid1703.101510

Falkinham, J. O. (2021). Ecology of nontuberculous mycobacteria. *Microorganisms*, *9*(11), 1–10. https://doi.org/10.3390/microorganisms9112262

Faria, T. Q., Mingote, A., Siopa, F., Ventura, R., Maycock, C., & Santos, H. (2008). Design of new enzyme stabilizers inspired by glycosides of hyperthermophilic microorganisms. *Carbohydrate Research*, *343*(18), 3025–3033. https://doi.org/10.1016/j.carres.2008.08.030

FDA. (2002). *Medical Washers and Medical Washer-Disinfectors - Class II Special Controls Guidance Document for the Medical Device Industry and FDA Review Staff*. Available at: https://www.fda.gov/media/71730/download

Feazel, L. M., Baumgartner, L. K., Peterson, K. L., Frank, D. N., Harris, J. K., & Pace, N. R. (2009). Opportunistic pathogens enriched in showerhead biofilms. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(38), 16393–16399. https://doi.org/10.1073/pnas.0908446106

Fernandes, C., Mendes, V., Costa, J., Empadinhas, N., Jorge, C., Lamosa, P., Santos, H., & da Costa, M. S. (2010). Two alternative pathways for the synthesis of the rare compatible solute mannosylglucosylglycerate in *Petrotoga mobilis*. *Journal of Bacteriology*, *192*(6), 1624–1633. https://doi.org/10.1128/JB.01424-09

Fillenberg, S. B., Friess, M. D., Korner, S., Böckmann, R. A., & Muller, Y. A. (2016). Crystal structures of the global regulator dasr from *Streptomyces coelicolor*: Implications for the allosteric regulation of GntR/HutC Repressors. *PLoS ONE*, *11*(6), 1–23. https://doi.org/10.1371/journal.pone.0157691

Fillenberg, S. B., Grau, F. C., Seidel, G., & Muller, Y. A. (2015). Structural insight into operator dre-sites recognition and effector binding in the GntR/HutC transcription regulator NagR. *Nucleic Acids Research*, *43*(2), 1283–1296. https://doi.org/10.1093/nar/gku1374

Flatt, P. M., & Mahmud, T. (2007). Biosynthesis of aminocyclitol-aminoglycoside antibiotics and related compounds. *Natural Product Reports*, *24*(2), 358–392. https://doi.org/10.1039/b603816f

Fleming, A. (1945). Sir Alexander Fleming - Nobel Lecture: Penicillin. In *Nobel Lecture*, December 11. Available at: https://www.nobelprize.org/uploads/2018/06/fleming-lecture.pdf

Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., van Brussel, D., Morawe, C., Borel, C., Hignette, O., Chavanne, J., Nurizzo, D., McSweeney, S., & Mitchell, E. (2010). The ID23-2 structural biology microfocus beamline at the ESRF. *Journal of Synchrotron Radiation*, *17*(1), 107–118. https://doi.org/10.1107/S0909049509041168

Fowler, D., Dalgleish, A., & Liu, W. (2014). A heat-killed preparation of *Mycobacterium obuense* can reduce metastatic burden in vivo. *Journal for ImmunoTherapy of Cancer*, *2*(Suppl 3), P54. https://doi.org/10.1186/2051-1426-2-s3-p54

Fraga, J., Maranha, A., Mendes, V., Pereira, P. J. B., Empadinhas, N., & Macedo-Ribeiro, S. (2015). Structure of mycobacterial maltokinase, the missing link in the essential GlgE-pathway. *Scientific Reports*, *5*, 8026. https://doi.org/10.1038/srep08026

Fulton, Z., McAlister, A., Wilce, M. C. J., Brammananth, R., Zaker-Tabrizi, L., Perugini, M. A., Bottomley, S. P., Coppel, R. L., Crellin, P. K., Rossjohn, J., & Beddoe, T. (2008). Crystal structure of a UDP-glucose-specific glycosyltransferase from a *Mycobacterium* species. *Journal of Biological Chemistry*, *283*(41), 27881–27890. https://doi.org/10.1074/jbc.M801853200

Furin, J., Cox, H., & Pai, M. (2019). Tuberculosis. *The Lancet*, *393*(10181), 1642–1656. https://doi.org/10.1016/S0140-6736(19)30308-3

Gagneux, S. (2018). Ecology and evolution of *Mycobacterium tuberculosis*. *Nature Reviews Microbiology*, *16*(4), 202–213. https://doi.org/10.1038/nrmicro.2018.8

Galagan, J. E. (2014). Genomic insights into tuberculosis. *Nature Reviews. Genetics*, *15*(5), 307–320. https://doi.org/10.1038/nrg3664

Gammon, K. (2014). Drug discovery: Leaving no stone unturned. *Nature*, *509*(7498), S10-2. https://doi.org/10.1038/509S10a

Gao, B., & Gupta, R. S. (2012). Phylogenetic Framework and Molecular Signatures for the Main Clades of the Phylum Actinobacteria. *Microbiology and Molecular Biology Reviews*, *76*(1), 66–112. https://doi.org/10.1128/MMBR.05011-11

Gaynes, R. (2017). The Discovery of Penicillin—New Insights After More Than 75 Years of Clinical Use. *Emerging Infectious Diseases*, *23*(5), 849–853. https://doi.org/http://dx.doi.org/10.3201/eid2305.161556

Gebert, M. J., Delgado-Baquerizo, M., Oliverio, A. M., Webster, T. M., Nichols, L. M., Honda, J. R., Chan, E. D., Adjemian, J., Dunn, R. R., & Fierer, N. (2018). Ecological analyses of mycobacteria in showerhead biofilms and their relevance to human health. *MBio*, *9*(5), e01614-18. https://doi.org/10.1101/366088

Gengenbacher, M., & Kaufmann, S. H. E. (2012). *Mycobacterium tuberculosis*: Success through dormancy. *FEMS Microbiology Reviews*, *36*(3), 514–532. https://doi.org/10.1111/j.1574-6976.2012.00331.x

Gengenbacher, M., Xu, T., Niyomrattanakit, P., Spraggon, G., & Dick, T. (2008). Biochemical and structural characterization of the putative dihydropteroate synthase ortholog Rv1207 of *Mycobacterium tuberculosis*. *FEMS Microbiology Letters*, *287*(1), 128–135. https://doi.org/10.1111/j.1574-6968.2008.01302.x

Genilloud, O. (2017). Actinomycetes: Still a source of novel antibiotics. *Natural Product Reports*, *34*(10), 1203–1232. https://doi.org/10.1039/c7np00026j

Gerlits, O., Tian, J., Das, A., Langan, P., Heller, W. T., & Kovalevsky, A. (2015). Phosphoryl transfer reaction snapshots in crystals: Insights into the mechanism of protein kinase a catalytic subunit. *Journal of Biological Chemistry*, *290*(25), 15538–15548. https://doi.org/10.1074/jbc.M115.643213

Glassroth, J. (2008). Pulmonary disease due to nontuberculous mycobacteria. *Chest*, *133*(1), 243–251. https://doi.org/10.1378/chest.07-0358

González-Pastor, J. E., Hobbs, E. C., & Losick, R. (2003). Cannibalism by sporulating bacteria. *Science*, *301*(5632), 510–513. https://doi.org/10.1126/science.1086462

Gopal, P., & Dick, T. (2014). Reactive dirty fragments: Implications for tuberculosis drug discovery. *Current Opinion in Microbiology*, *21*, 7–12. https://doi.org/10.1016/j.mib.2014.06.015

Gordillo Altamirano, F. L., & Barr, J. J. (2019). Phage therapy in the postantibiotic era. *Clinical Microbiology Reviews*, *32*(2), e00066-18. https://doi.org/10.1128/CMR.00066-18

Grellert, E., & Ballou, C. E. (1972). Biosynthesis of a Mycobacterial Lipopolysaccharide. *Journal of Biological Chemistry*, *247*(10), 3236–3241.

Griffith, D. E. (2010). Nontuberculous mycobacterial lung disease. *Current Opinion in Infectious Diseases*, *23*(2), 185–190. https://doi.org/10.1097/QCO.0b013e328336ead6

Griffith, D. E., Aksamit, T., Brown-Elliott, B. A., Catanzaro, A., Daley, C., Gordin, F., Holland, S. M., Horsburgh, R., Huitt, G., Iademarco, M. F., Iseman, M., Olivier, K., Ruoss, S., Von Reyn, C. F., Wallace, R. J., & Winthrop, K. (2007). An official ATS/IDSA statement: Diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *American Journal of Respiratory and Critical Care Medicine*, *175*(4), 367–416. https://doi.org/10.1164/rccm.200604-571ST

Groenewald, W., Baird, M. S., Verschoor, J. A., Minnikin, D. E., & Croft, A. K. (2014). Differential spontaneous folding of mycolic acids from *Mycobacterium tuberculosis*. *Chemistry and Physics of Lipids*, *180*, 15–22. https://doi.org/10.1016/j.chemphyslip.2013.12.004

Guenette, S., Williams, M. D., & Falkinham, J. O. (2020). Growth temperature, trehalose, and susceptibility to heat in *Mycobacterium avium*. *Pathogens*, *9*(8), 657. https://doi.org/10.3390/pathogens9080657

Gupta, R. S., Lo, B., & Son, J. (2018). Phylogenomics and comparative genomic studies robustly support division of the genus *Mycobacterium* into an emended genus *Mycobacterium* and four novel genera. *Frontiers in Microbiology*, *9*, 67. https://doi.org/10.3389/fmicb.2018.00067

Gutiérrez, A. V., Viljoen, A., Ghigo, E., Herrmann, J. L., & Kremer, L. (2018). Glycopeptidolipids, a double-edged sword of the *Mycobacterium abscessus* complex. *Frontiers in Microbiology*, *9*, 1145. https://doi.org/10.3389/fmicb.2018.01145

Guzzo, M. B., Li, Q., Nguyen, H. V., Boom, W. H., & Nguyen, L. (2021). The Pup-proteasome system protects mycobacteria from antimicrobial antifolates. *Antimicrobial Agents and Chemotherapy*, *65*(4), e01967-20. https://doi.org/10.1128/AAC.01967-20

Gygli, S. M., Borrell, S., Trauner, A., & Gagneux, S. (2017). Antimicrobial resistance in *Mycobacterium tuberculosis*: Mechanistic and evolutionary perspectives. *FEMS Microbiology Reviews*, *41*(3), 354–373. https://doi.org/10.1093/femsre/fux011

Haas, B., Soto, K. J., Day, D. S., Roy, A. C., Gagnon, M. C., Alt, J. R., & Labrie, P. (2020). Mycobacterium hassiacum: A thermophilic *Mycobacterium* species to demonstrate thermal disinfection of medical devices. *BMC Research Notes*, *13*(1), 140. https://doi.org/10.1186/s13104-020-04978-7

Hamm, P. S., Dunlap, C. A., Mullowney, M. W., Caimi, N. A., Kelleher, N. L., Thomson, R. J., Porras-Alfaro, A., & Northup, D. E. (2020). *Streptomyces buecherae* sp. nov., an actinomycete isolated from multiple bat species. *Antonie van Leeuwenhoek*, *113*(12), 2213–2221. https://doi.org/10.1007/s10482-020-01493-4

Han, X. Y., & Silva, F. J. (2014). On the Age of Leprosy. *PLoS Neglected Tropical Diseases*, *8*(2), e2544. https://doi.org/10.1371/journal.pntd.0002544

Hansen, G. H. A. (1874). Undersogelser angaende spedalskhedens aasager. *Norsk Magazin for Laegervidenskaben*, *4*(Suppl.), 1–88.

Harland, C. W., Rabuka, D., Bertozzi, C. R., & Parthasarathy, R. (2008). The *Mycobacterium tuberculosis* virulence factor trehalose dimycolate imparts desiccation resistance to model mycobacterial membranes. *Biophysical Journal*, *94*(12), 4718–4724. https://doi.org/10.1529/biophysj.107.125542

Hayward, S. (2004). Identification of specific interactions that drive ligand-induced closure in five enzymes with classic domain movements. *Journal of Molecular Biology*, *339*(4), 1001–1021. https://doi.org/10.1016/j.jmb.2004.04.004

Hershkovitz, I., Donoghue, H. D., Minnikin, D. E., Besra, G. S., Lee, O. Y. C., Gernaey, A. M., Galili, E., Eshed, V., Greenblatt, C. L., Lemma, E., Bar-Gal, G. K., & Spigelman, M. (2008). Detection and molecular characterization of 9000-year-old *Mycobacterium tuberculosis* from a neolithic settlement in the Eastern mediterranean. *PLoS ONE*, *3*(10), e3426. https://doi.org/10.1371/journal.pone.0003426

Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microbiol*, *8*(1), 15–25. https://doi.org/10.1038/nrmicro2259

Hill, B. L., Figueroa, C. M., Diez, M. D. A., Lunn, J. E., Iglesias, A. A., & Ballicora, M. A. (2017). On the stability of nucleoside diphosphate glucose metabolites: Implications for studies of plant carbohydrate metabolism. *Journal of Experimental Botany*, *68*(13), 3331–3337. https://doi.org/10.1093/jxb/erx190

Hindsgaul, O., & Ballou, C. E. (1984). Affinity purification of mycobacterial polymethyl polysaccharides and a study of polysaccharide-lipid interactions by 1H NMR. *Biochemistry*, *23*(3), 577–584. http://www.ncbi.nlm.nih.gov/pubmed/6704384

Hoffmann, C., Leis, A., Niederweis, M., Plitzko, J. M., & Engelhardt, H. (2008). Disclosure of the mycobacterial outer membrane: Cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *PNAS*, *105*(10), 3963–3967. https://doi.org/10.1073/pnas.0709530105

Hon, W. C., McKay, G. a, Thompson, P. R., Sweet, R. M., Yang, D. S., Wright, G. D., & Berghuis, a M. (1997). Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases. *Cell*, *89*(6), 887–895. https://doi.org/10.1016/S0092-8674(00)80274-3

Huang, C. Y., & Hsieh, W. Y. (2017). Efficacy of *Mycobacterium vaccae* immunotherapy for patients with tuberculosis: A systematic review and meta-analysis. *Human Vaccines and Immunotherapeutics*, *13*(9), 1960–1971. https://doi.org/10.1080/21645515.2017.1335374

Huang, Y., Cui, Q., Wang, L., Rodriguez, C., Quintana, E., Goodfellow, M., & Liu, Z. (2004). *Streptacidiphilus jiangxiensis* sp. nov., a novel actinomycete isolated from acidic rhizosphere soil in China. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *86*(2), 159–165. https://doi.org/10.1023/B:ANTO.0000036124.18820.ae

Hunter, S. W., Fujiwara, T., Murphy, R. C., & Brennan, P. J. (1984). N-acylkansosamine. A novel N-acylamino sugar from the trehalose-containing lipooligosaccharide antigens of *Mycobacterium kansasii*. *The Journal of Biological Chemistry*, *259*(15), 9729–9734.

Hunter, S. W., Gaylor, H., & Brennan, P. J. (1986). Structure and antigenicity of the phosphorylated lipopolysaccharide antigens from the leprosy and tubercle bacilli. *Journal of Biological Chemistry*, *261*(26), 12345–12351. https://doi.org/10.1016/s0021-9258(18)67246-1

Hutchings, M., Truman, A., & Wilkinson, B. (2019). Antibiotics: past, present and future. *Current Opinion in Microbiology*, *51*, 72–80. https://doi.org/10.1016/j.mib.2019.10.008

Hwang, S., Yun, Y., Choi, W. H., Kim, S. B., Shin, J., Lee, M. J., & Oh, D.-C. (2019). Acidiphilamides A-E, Modified Peptides as Autophagy Inhibitors from an Acidophilic Actinobacterium, *Streptacidiphilus rugosus*. *Journal of Natural Products*, *82*(2), 341–348. https://doi.org/10.1021/acs.jnatprod.8b00828

Imperiali, B. (2019). Bacterial carbohydrate diversity - a Brave New World. *Current Opinion in Chemical Biology*, *53*, 1–8. https://doi.org/10.1016/j.cbpa.2019.04.026

Jacin, H., & Mishkin, A. (1965). Separation of carbohydrates on borate-impregnated silica gel G plates. *J Chromatogr*, *18*, 170–173.

Jackson, M., & Brennan, P. (2009). Polymethylated polysaccharides from *Mycobacterium* species revisited. *Journal of Biological Chemistry*, *248*(4), 1949–1953. https://doi.org/10.1074/jbc.R800047200

Jarling, M., Cauvet, T., Grundmeier, M., Kuhnert, K., & Pape, H. (2004). Isolation of makI from *Actinoplanes*

missouriensis and evidence that Pep2 from *Streptomyces coelicolor* is a maltokinase. *Journal of Basic Microbiology*, *44*(5), 360–373. https://doi.org/10.1002/jobm.200410403

Jhoti, H., Williams, G., Rees, D. C., & Murray, C. W. (2013). The "rule of three" for fragment-based drug discovery: Where are we now? *Nature Reviews Drug Discovery*, *12*(8), 644. https://doi.org/10.1038/nrd3926-c1

Jiang, S. H., Roberts, D. M., Clayton, P. A., & Jardine, M. (2013). Non-tuberculous mycobacterial PD peritonitis in Australia. *International Urology and Nephrology*, *45*(5), 1423–1428. https://doi.org/10.1007/s11255-012-0328-4

Jiang, X. H., Li, C. Q., Feng, G. Y., Luo, M. J., & Sun, Q. X. (2020). Inhalation of nebulized *Mycobacterium vaccae* can protect against allergic bronchial asthma in mice by regulating the TGF-β/Smad signal transduction pathway. *Allergy, Asthma and Clinical Immunology*, *16*, 59. https://doi.org/10.1186/s13223-020-00456-8

Johansen, M. D., Herrmann, J. L., & Kremer, L. (2020). Non-tuberculous mycobacteria and the rise of *Mycobacterium abscessus*. *Nature Reviews Microbiology*, *18*(7), 392–407. https://doi.org/10.1038/s41579-020-0331-1

Jothivasan, V. K., & Hamilton, C. J. (2008). Mycothiol: Synthesis, biosynthesis and biological functions of the major low molecular weight thiol in actinomycetes. *Natural Product Reports*, *25*(6), 1091–1117. https://doi.org/10.1039/b616489g

Juanhuix, J., Gil-Ortiz, F., Cuní, G., Colldelram, C., Nicolás, J., Lidón, J., Boter, E., Ruget, C., Ferrer, S., & Benach, J. (2014). Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the ALBA synchrotron. *Journal of Synchrotron Radiation*, *21*(Pt 4), 679–689. https://doi.org/10.1107/S160057751400825X

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., … Baker, N. A. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Science: A Publication of the Protein Society*, *27*(1), 112–128. https://doi.org/10.1002/pro.3280

Kabsch, W, & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, *22*(12), 2577–2637. https://doi.org/10.1002/bip.360221211

Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D*, *66*(2), 133–144. https://doi.org/10.1107/S0907444909047374

Kafri, R., Springer, M., & Pilpel, Y. (2009). Genetic Redundancy: New Tricks for Old Genes. *Cell*, *136*(3), 389–392. https://doi.org/10.1016/j.cell.2009.01.027

Kalscheuer, R., Syson, K., Veeraraghavan, U., Weinrick, B., Biermann, K. E., Liu, Z., Sacchettini, J. C., Besra, G., Bornemann, S., & Jacobs, W. R. (2010). Self-poisoning of *Mycobacterium tuberculosis* by targeting GlgE in an alpha-glucan pathway. *Nature Chemical Biology*, *6*(5), 376–384. https://doi.org/10.1038/nchembio.340

Kaltenpoth, M., Göttler, W., Herzner, G., & Strohm, E. (2005). Symbiotic Bacteria Protect Wasp Larvae from Fungal Infestation. *Current Biology*, *15*(5), 475–479. https://doi.org/10.1016/j.cub.2004.12.084

Kamisango, K., Dell, A., & Ballou, C. E. (1987). Biosynthesis of the Mycobacterial O-Methylglucose Lipopolysaccharide: Characterization of putative intermediates in the initiation, elongation, and termination reactions. *The Journal of Biological Chemistry*, *262*(10), 4580–4586.

Kannan, N., Taylor, S. S., Zhai, Y., Venter, J. C., & Manning, G. (2007). Structural and functional diversity of the microbial kinome. *PLoS Biology*, *5*(3), 0467–0478. https://doi.org/10.1371/journal.pbio.0050017

Kari, B. E., & Gray, G. R. (1979). Acetylated methylmannose polysaccharide of *Streptomyces griseus*. Locations of the acetyl groups. *The Journal of Biological Chemistry*, *254*(9), 3354–3357.

Karnam, S., Alla, V. M., Kwon, J., Harbert, T., Sharma, A., Airey, K., & Mooss, A. (2011). *Mycobacterium phlei*, a previously unreported cause of pacemaker infection: Thinking outside the box in cardiac device infections. *Cardiology Journal*, *18*(6), 687–690. https://doi.org/10.5603/CJ.2011.0034

Kattke, M. D., Gosschalk, J. E., Martinez, O. E., Kumar, G., Gale, R. T., Cascio, D., Sawaya, M. R., Philips, M., Brown, E. D., & Clubb, R. T. (2019). Structure and mechanism of TagA, a novel membrane-associated glycosyltransferase that produces wall teichoic acids in pathogenic bacteria. *PLoS Pathogens*, *15*(4), e1007723. https://doi.org/10.1371/journal.ppat.1007723

Katz, L., & Baltz, R. H. (2016). Natural product discovery: past, present, and future. *Journal of Industrial Microbiology & Biotechnology*, *43*(2–3), 155–176. https://doi.org/10.1007/s10295-015-1723-5

Kaur, D., Pham, H., Larrouy-Maumus, G., Rivière, M., Vissa, V., Guerin, M. E., Puzo, G., Brennan, P. J., & Jackson, M. (2009). Initiation of methylglucose lipopolysaccharide biosynthesis in mycobacteria. *PLoS ONE*,

4(5), e5447. https://doi.org/10.1371/journal.pone.0005447

Kautsar, S. A., Blin, K., Shaw, S., Navarro-Muñoz, J. C., Terlouw, B. R., van der Hooft, J. J. J., van Santen, J. A., Tracanna, V., Suarez Duran, H. G., Pascal Andreu, V., Selem-Mojica, N., Alanjary, M., Robinson, S. L., Lund, G., Epstein, S. C., Sisto, A. C., Charkoudian, L. K., Collemare, J., Linington, R. G., … Medema, M. H. (2020). MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Research*, *48*(D1), D454–D458. https://doi.org/10.1093/nar/gkz882

Kermani, A. A., Roy, R., Gopalasingam, C., Kocurek, K. I., Patel, T. R., Alderwick, L. J., Besra, G. S., & Fütterer, K. (2019). Crystal structure of the TreS:Pep2 complex, initiating α-glucan synthesis in the GlgE pathway of mycobacteria. *Journal of Biological Chemistry*, *294*(18), 7348–7359. https://doi.org/10.1074/jbc.RA118.004297

Kim, C., Joo, Y. C., Yan, H., Vakulenko, S. B., & Mobashery, S. (2006). Hydrolysis of ATP by aminoglycoside 3'-phosphotransferases: An unexpected cost to bacteria for harboring an antibiotic resistance enzyme. *Journal of Biological Chemistry*, *281*(11), 6964–6969. https://doi.org/10.1074/jbc.M513257200

Kim, S. B., Lonsdale, J., Seong, C. N., & Goodfellow, M. (2003). *Streptacidiphilus* gen. nov., acidophilic actinomycetes with wall chemotype I and emendation of the family *Streptomycetaceae* (Waksman and Henrici (1943)AL) emend. Rainey et al. 1997. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *83*(2), 107–116. https://doi.org/10.1023/A:1023397724023

Kirschner, R. a., Parker, B. C., & Falkinham, J. O. (1999). Humic and fulvic acids stimulate the growth of *Mycobacterium avium*. *FEMS Microbiology Ecology*, *30*(4), 327–332. https://doi.org/10.1016/S0168-6496(99)00069-0

Klähn, S., Steglich, C., Hess, W. R., & Hagemann, M. (2010). Glucosylglycerate: A secondary compatible solute common to marine cyanobacteria from nitrogen-poor environments. *Environmental Microbiology*, *12*(1), 83–94. https://doi.org/10.1111/j.1462-2920.2009.02045.x

Koch, R. (1882). Die Äetiologie der Tuberkulose. *Berliner Klinische Wochenschrift*, *15*, 221–230.

Köhler, K., Marcsik, A., Zádori, P., Biro, G., Szeniczey, T., Fábián, S., Serlegi, G., Marton, T., Donoghue, H. D., & Hajdu, T. (2017). Possible cases of leprosy from the Late Copper Age (3780-3650 cal BC) in Hungary. *PLoS ONE*, *12*(10), e0185966. https://doi.org/10.1371/journal.pone.0185966

Koliwer-Brandl, H., Syson, K., van de Weerd, R., Chandra, G., Appelmelk, B., Alber, M., Ioerger, T. R., Jacobs, W. R., Geurtsen, J., Bornemann, S., & Kalscheuer, R. (2016). Metabolic Network for the Biosynthesis of Intra- and Extracellular α-Glucans Required for Virulence of *Mycobacterium tuberculosis*. *PLoS Pathogens*, *12*(8), e1005768. https://doi.org/10.1371/journal.ppat.1005768

Koo, H.-K., Min, J., Kim, H. W., Lee, J., Kim, J. S., Park, J. S., & Lee, S.-S. (2020). Prediction of treatment failure and compliance in patients with tuberculosis. *BMC Infectious Diseases*, *20*(1), 622. https://doi.org/10.1186/s12879-020-05350-7

Kovalevsky, A. Y., Johnson, H., Hanson, B. L., Waltman, M. J., Fisher, S. Z., Taylor, S., & Langan, P. (2012). Low- and room-temperature X-ray structures of protein kinase A ternary complexes shed new light on its activity. *Acta Crystallographica Section D: Biological Crystallography*, *68*(Pt 7), 854–860. https://doi.org/10.1107/S0907444912014886

Kremer, L., Guérardel, Y., Gurcha, S. S., Locht, C., & Besra, G. S. (2002). Temperature-induced changes in the cell-wall components of *Mycobacterium thermoresistibile*. *Microbiology*, *148*(10), 3145–3154. https://doi.org/0002-5713

Křen, V., & Řezanka, T. (2008). Sweet antibiotics - The role of glycosidic residues in antibiotic and antitumor activity and their randomization. *FEMS Microbiology Reviews*, *32*(5), 858–889. https://doi.org/10.1111/j.1574-6976.2008.00124.x

Krissinel, E., & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica. Section D, Biological Crystallography*, *60*(Pt 12 Pt 1), 2256–2268. https://doi.org/10.1107/S0907444904026460

Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, *372*(3), 774–797. https://doi.org/10.1016/j.jmb.2007.05.022

Ku, S. Y., Yip, P., Cornell, K. a., Riscoe, M. K., Behr, J. B., Guillerm, G., & Howell, P. L. (2007). Structures of 5-methylthioribose kinase reveal substrate specificity and unusual mode of nucleotide binding. *Journal of Biological Chemistry*, *282*(30), 22195–22206. https://doi.org/10.1074/jbc.M611045200

Kudo, F., & Eguchi, T. (2009). Biosynthetic genes for aminoglycoside antibiotics. *Journal of Antibiotics*, *62*(9), 471–481. https://doi.org/10.1038/ja.2009.76

Kudo, F., Kawabe, K., Kuriki, H., Eguchi, T., & Kakinuma, K. (2005). A new family of glucose-1-phosphate/glucosamine-1-phosphate nucleotidylyltransferase in the biosynthetic pathways for antibiotics. *Journal of the American Chemical Society*, *127*(6), 1711–1718. https://doi.org/10.1021/ja044921b

Kumar, A., Chettiar, S., & Parish, T. (2017). Current challenges in drug discovery for tuberculosis. *Expert Opinion on Drug Discovery*, *12*(1), 1–4. https://doi.org/10.1080/17460441.2017.1255604

Kumar, G., Guan, S., & Frantom, P. A. (2014). Biochemical characterization of the retaining glycosyltransferase glucosyl-3-phosphoglycerate synthase from *Mycobacterium tuberculosis*. *Archives of Biochemistry and Biophysics*, *564*, 120–127. https://doi.org/10.1016/j.abb.2014.10.002

Lacombe-Harvey, M. È., Brzezinski, R., & Beaulieu, C. (2018). Chitinolytic functions in actinobacteria: ecology, enzymes, and evolution. *Applied Microbiology and Biotechnology*, *102*(17), 7219–7230. https://doi.org/10.1007/s00253-018-9149-4

Lairson, L. L., Henrissat, B., Davies, G. J., & Withers, S. G. (2008). Glycosyl transferases: Structures, functions, and mechanisms. *Annual Review of Biochemistry*, *77*, 521–555. https://doi.org/10.1146/annurev.biochem.76.061005.092322

Larrouy-Maumus, G., Gilleron, M., Skovierová, H., Zuberogoitia, S., Brennan, P. J., Puzo, G., Jackson, M., & Nigou, J. (2015). A glycomic approach reveals a new mycobacterial polysaccharide. *Glycobiology*, *25*(11), 1163–1171. https://doi.org/10.1093/glycob/cwv061

Le Dantec, C., Duguet, J., Montiel, A., Dumoutier, N., Dubrou, S., & Vincent, V. (2002). Chlorine Disinfection of Atypical Mycobacteria Isolated from a Water Distribution System. *Applied and Environmental Microbiology*, *68*(3), 1025–1032. https://doi.org/10.1128/AEM.68.3.1025

Le Guilloux, V., Schmidtke, P., & Tuffery, P. (2009). Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics*, *10*, 168. https://doi.org/10.1186/1471-2105-10-168

Lee, J., Repasy, T., Papavinasasundaram, K., Sassetti, C., & Kornfeld, H. (2011). *Mycobacterium tuberculosis* induces an atypical cell death mode to escape from infected macrophages. *PLoS ONE*, *6*(3), e18367. https://doi.org/10.1371/journal.pone.0018367

Lee, O. Y. C., Wu, H. H. T., Besra, G. S., Rothschild, B. M., Spigelman, M., Hershkovitz, I., Bar-Gal, G. K., Donoghue, H. D., & Minnikin, D. E. (2015). Lipid biomarkers provide evolutionary signposts for the oldest known cases of tuberculosis. *Tuberculosis*, *95*(S1), S127–S132. https://doi.org/10.1016/j.tube.2015.02.013

Leonard, C. J., Aravind, L., & Koonin, E. V. (1998). Novel families of putative protein kinases in bacteria and archaea: Evolution of the "eukaryotic" protein kinase superfamily. *Genome Research*, *8*(10), 1038–1047. https://doi.org/10.1101/gr.8.10.1038

Li, Y., Liu, J., Díaz-Cruz, G., Cheng, Z., & Bignell, D. R. D. (2019). Virulence mechanisms of plant-pathogenic *Streptomyces* species: An updated review. *Microbiology (United Kingdom)*, *165*(10), 1025–1040. https://doi.org/10.1099/mic.0.000818

Ling, L. L., Schneider, T., Peoples, A. J., Spoering, A. L., Engels, I., Conlon, B. P., Mueller, A., Hughes, D. E., Epstein, S., Jones, M., Lazarides, L., Steadman, V. a, Cohen, D. R., Felix, C. R., Fetterman, K. A., Millett, W. P., Nitti, A. G., Zullo, A. M., Chen, C., & Lewis, K. (2015). A new antibiotic kills pathogens without detectable resistance. *Nature*, *517*(7535), 455–459. https://doi.org/10.1038/nature14098

Lira-Navarrete, E., Iglesias-Fernández, J., Zandberg, W. F., Compañón, I., Kong, Y., Corzana, F., Pinto, B. M., Clausen, H., Peregrina, J. M., Vocadlo, D. J., Rovira, C., & Hurtado-Guerrero, R. (2014). Substrate-guided front-face reaction revealed by combined structural snapshots and metadynamics for the polypeptide N-acetylgalactosaminyltransferase 2. *Angewandte Chemie - International Edition*, *53*(31), 8206–8210. https://doi.org/10.1002/anie.201402781

Liu, G., Chater, K. F., Chandra, G., Niu, G., & Tan, H. (2013). Molecular regulation of antibiotic biosynthesis in streptomyces. *Microbiology and Molecular Biology Reviews*, *77*(1), 112–143. https://doi.org/10.1128/MMBR.00054-12

Liu, T., Wang, B., Guo, J., Zhou, Y., Julius, M., Njire, M., Cao, Y., Wu, T., Liu, Z., Wang, C., Xu, Y., & Zhang, T. (2015). Role of folP1 and folP2 genes in the action of sulfamethoxazole and trimethoprim against mycobacteria. *Journal of Microbiology and Biotechnology*, *25*(9), 1559–1567. https://doi.org/10.4014/jmb.1503.03053

Lopeman, R. C., Harrison, J., Desai, M., & Cox, J. A. G. (2019). *Mycobacterium abscessus*: Environmental bacterium turned clinical nightmare. *Microorganisms*, *7*(3), 90. https://doi.org/10.3390/microorganisms7030090

Lounis, N., Gevers, T., Van Den Berg, J., Vranckx, L., & Andries, K. (2009). ATP synthase inhibition of *Mycobacterium avium* is not bactericidal. *Antimicrobial Agents and Chemotherapy*, *53*(11), 4927–4929.

https://doi.org/10.1128/AAC.00689-09

Ma, Q., & Xu, Y. (2013). Global Genomic Arrangement of Bacterial Genes Is Closely Tied with the Total Transcriptional Efficiency. *Genomics, Proteomics and Bioinformatics*, *11*(1), 66–71. https://doi.org/10.1016/j.gpb.2013.01.004

Machado, D., Lecorche, E., Mougari, F., Cambau, E., & Viveiros, M. (2018). Insights on *Mycobacterium leprae* efflux pumps and their implications in drug resistance and virulence. *Frontiers in Microbiology*, *9*, 3072. https://doi.org/10.3389/fmicb.2018.03072

Machida, Y., & Bloch, K. (1973). Complex formation between mycobacterial polysaccharides and fatty acyl-CoA derivatives. *Proceedings of the National Academy of Sciences of the United States of America*, *70*(4), 1146–1148. https://doi.org/10.1073/pnas.70.4.1146

Maitra, A., Munshi, T., Healy, J., Martin, L. T., Vollmer, W., Keep, N. H., & Bhakta, S. (2019). Cell wall peptidoglycan in *Mycobacterium tuberculosis*: An Achilles' heel for the TB-causing pathogen. *FEMS Microbiology Reviews*, *43*(5), 548–575. https://doi.org/10.1093/femsre/fuz016

Malik, A., Kim, Y. R., & Kim, S. B. (2020). Genome mining of the genus *Streptacidiphilus* for biosynthetic and biodegradation potential. *Genes*, *11*(10), 1166. https://doi.org/10.3390/genes11101166

Maloney, D. H., & Ballou, C. E. (1980). Polymethylpolysaccharide synthesis in an ethionine-resistant mutant of *Mycobacterium smegmatis*. *Journal of Bacteriology*, *141*(3), 1217–1221. https://doi.org/10.1128/jb.141.3.1217-1221.1980

Manso, J. A., Nunes-Costa, D., Macedo-Ribeiro, S., Empadinhas, N., & Pereira, P. J. B. (2019). Molecular Fingerprints for a Novel Enzyme Family in Actinobacteria with Glucosamine Kinase Activity. *MBio*, *10*(3), e00239-19. https://doi.org/https://doi .org/10.1128/mBio.00239-19

Maranha, A., Moynihan, P. J., Miranda, V., Correia Lourenço, E., Nunes-Costa, D., Fraga, J. S., José Barbosa Pereira, P., Macedo-Ribeiro, S., Ventura, M. R., Clarke, A. J., & Empadinhas, N. (2015). Octanoylation of early intermediates of mycobacterial methylglucose lipopolysaccharides. *Scientific Reports*, *5*, 13610. https://doi.org/10.1038/srep13610

Martins-Teixeira, M. B., Campo, V. L., Biondo, M., Sesti-Costa, R., Carneiro, Z. A., Silva, J. S., & Carvalho, I. (2013). α-Selective glycosylation affords mucin-related GalNAc amino acids and diketopiperazines active on *Trypanosoma cruzi*. *Bioorganic and Medicinal Chemistry*, *21*(7), 1978–1987. https://doi.org/10.1016/j.bmc.2013.01.027

Matte, A., Tari, L. W., & Delbaere, L. T. (1998). How do kinases transfer phosphoryl groups? *Structure*, *6*(4), 413–419. https://doi.org/10.1016/S0969-2126(98)00043-4

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, *40*(4), 658–674. https://doi.org/10.1107/S0021889807021206

McGrath, E. E., Blades, Z., McCabe, J., Jarry, H., & Anderson, P. B. (2010). Nontuberculous mycobacteria and the lung: from suspicion to treatment. *Lung*, *188*(4), 269–282. https://doi.org/10.1007/s00408-010-9240-9

Mehta, H., Weng, J., Prater, A., Elworth, R. A. L., Han, X., & Shamoo, Y. (2018). Pathogenic *Nocardia cyriacigeorgica* and *Nocardia nova* Evolve To Resist Trimethoprim-Sulfamethoxazole by both Expected and Unexpected Pathways. *Antimicrobial Agents and Chemotherapy*, *62*(7), e00364-18. https://doi.org/10.1128/AAC.00364-18

Meissner, P. S., & Falkinham, J. O. (1984). Plasmid-encoded mercuric reductase in *Mycobacterium scrofulaceum*. *Journal of Bacteriology*, *157*(2), 669–672.

Mendes, V., Maranha, A., Alarico, S., da Costa, M. S., & Empadinhas, N. (2011). *Mycobacterium tuberculosis* Rv2419c, the missing glucosyl-3-phosphoglycerate phosphatase for the second step in methylglucose lipopolysaccharide biosynthesis. *Scientific Reports*, *1*, 177. https://doi.org/10.1038/srep00177

Mendes, V., Maranha, A., Alarico, S., & Empadinhas, N. (2012). Biosynthesis of mycobacterial methylglucose lipopolysaccharides. *Natural Product Reports*, *29*(8), 834–844. https://doi.org/10.1039/c2np20014g

Mendes, V., Maranha, A., Lamosa, P., da Costa, M. S., & Empadinhas, N. (2010). Biochemical characterization of the maltokinase from *Mycobacterium bovis* BCG. *BMC Biochemistry*, *11*, 21. https://doi.org/10.1186/1471-2091-11-21

Metsä-Ketelä, M., Niemi, J., Mäntsälä, P., & Schneider, G. (2008). Anthracycline biosynthesis: Genes, enzymes and mechanisms. In K. Krohn (Ed.), *Anthracycline Chemistry and Biology I* (Vol. 282, pp. 101–140). Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/128-2007-14

Miah, F., Koliwer-Brandl, H., Rejzek, M., Field, R. A., Kalscheuer, R., & Bornemann, S. (2013). Flux through trehalose synthase flows from trehalose to the alpha anomer of maltose in mycobacteria. *Chemistry & Biology*, *20*(4), 487–493. https://doi.org/10.1016/j.chembiol.2013.02.014

Miethke, M., Pieroni, M., Weber, T., Brönstrup, M., Hammann, P., Halby, L., Arimondo, P. B., Glaser, P., Aigle, B., Bode, H. B., Moreira, R., Li, Y., Luzhetskyy, A., Medema, M. H., Pernodet, J. L., Stadler, M., Tormo, J. R., Genilloud, O., Truman, A. W., … Müller, R. (2021). Towards the sustainable discovery and development of new antibiotics. *Nature Reviews Chemistry*, *5*(10), 726–749. https://doi.org/10.1038/s41570-021-00313-1

Milewski, S. (2002). Glucosamine-6-phosphate synthase - the multi-facets enzyme. *Biochimica et Biophysica Acta*, *1597*(2), 173–192. https://doi.org/10.1016/s0167-4838(02)00318-7

Miller, C. D., Hall, K., Liang, Y. N., Nieman, K., Sorensen, D., Issa, B., Anderson, A. J., & Sims, R. C. (2004). Isolation and characterization of polycyclic aromatic hydrocarbon-degrading mycobacterium isolates from soil. *Microbial Ecology*, *48*(2), 230–238. https://doi.org/10.1007/s00248-003-1044-5

Minato, Y., Gohl, D. M., Thiede, J. M., Maruyama, F., Baughn, A. D., & Harcombe, W. R. (2019). Genomewide Assessment of *Mycobacterium tuberculosis* Conditionally Essential Metabolic Pathways. *Molecular Biology and Physiology*, *4*, e00070-19.

Minnikin, D. E., Lee, O. Y.-C., Wu, H. H. T., Nataraj, V., Donoghue, H. D., Ridell, M., Watanabe, M., Alderwick, L., Bhatt, A., & Besra, G. S. (2015). Pathophysiological Implications of Cell Envelope Structure in *Mycobacterium tuberculosis* and Related Taxa. In *Tuberculosis - Expanding Knowledge*. IntechOpen. https://doi.org/10.5772/59585

Mishra, A. K., Driessen, N. N., Appelmelk, B. J., & Besra, G. S. (2011). Lipoarabinomannan and related glycoconjugates: structure, biogenesis and role in *Mycobacterium tuberculosis* physiology and host-pathogen interaction. *FEMS Microbiology Reviews*, *35*(6), 1126–1157. https://doi.org/10.1111/j.1574-6976.2011.00276.x

Miyamoto, Y., Mukai, T., Naka, T., Fujiwara, N., Maeda, Y., Kai, M., Mizuno, S., Yano, I., & Makino, M. (2010). Novel rhamnosyltransferase involved in biosynthesis of serovar 4-specific glycopeptidolipid from *Mycobacterium avium* complex. *Journal of Bacteriology*, *192*(21), 5700–5708. https://doi.org/10.1128/JB.00554-10

Moraes, G. L., Gomes, G. C., Monteiro De Sousa, P. R., Alves, C. N., Govender, T., Kruger, H. G., Maguire, G. E. M., Lamichhane, G., & Lameira, J. (2015). Structural and functional features of enzymes of *Mycobacterium tuberculosis* peptidoglycan biosynthesis as targets for drug development. *Tuberculosis*, *95*(2), 95–111. https://doi.org/10.1016/j.tube.2015.01.006

Moremen, K. W., & Haltiwanger, R. S. (2019). Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. *Nature Chemical Biology*, *15*(9), 853–864. https://doi.org/10.1038/s41589-019-0350-2

Moynihan, P. J., Cadby, I. T., Veerapen, N., Jankute, M., Crosatti, M., Mukamolova, G. V., Lovering, A. L., & Besra, G. S. (2019). The hydrolase LpqI primes mycobacterial peptidoglycan recycling. *Nature Communications*, *10*, 2647. https://doi.org/10.1038/s41467-019-10586-2

Mukherjee, R., & Chatterji, D. (2012). Glycopeptidolipids: Immuno-modulators in greasy mycobacterial cell envelope. *IUBMB Life*, *64*(3), 215–225. https://doi.org/10.1002/iub.602

Murphy, H. N., Stewart, G. R., Mischenko, V. V., Apt, A. S., Harris, R., McAlister, M. S. B., Driscoll, P. C., Young, D. B., & Robertson, B. D. (2005). The OtsAB pathway is essential for trehalose biosynthesis in *Mycobacterium tuberculosis*. *Journal of Biological Chemistry*, *280*(15), 14524–14529. https://doi.org/10.1074/jbc.M414232200

Nah, H. J., Pyeon, H. R., Kang, S. H., Choi, S. S., & Kim, E. S. (2017). Cloning and heterologous expression of a large-sized natural product biosynthetic gene cluster in *Streptomyces* species. *Frontiers in Microbiology*, *8*, 394. https://doi.org/10.3389/fmicb.2017.00394

Nasiri, M. J., Haeili, M., Ghazi, M., Goudarzi, H., Pormohammad, A., Imani Fooladi, A. A., & Feizabadi, M. M. (2017). New Insights in to the Intrinsic and Acquired Drug Resistance Mechanisms in Mycobacteria. *Frontiers in Microbiology*, *8*, 681. https://doi.org/10.3389/fmicb.2017.00681

Nedal, A., & Zotchev, S. B. (2004). Biosynthesis of deoxyaminosugars in antibiotic-producing bacteria. *Applied Microbiology and Biotechnology*, *64*(1), 7–15. https://doi.org/10.1007/s00253-003-1535-9

Negatu, D. A., Liu, J. J. J., Zimmerman, M., Kaya, F., Dartois, V., Aldrich, C. C., Gengenbacher, M., & Dicka, T. (2018). Whole-cell screen of fragment library identifies gut microbiota metabolite indole propionic acid as antitubercular. *Antimicrobial Agents and Chemotherapy*, *62*(3), e01571-17. https://doi.org/10.1128/AAC.01571-17

Negatu, D. A., Gengenbacher, M., Dartois, V., & Dick, T. (2020). Indole Propionic Acid, an Unusual Antibiotic Produced by the Gut Microbiota, With Anti-inflammatory and Antioxidant Properties. *Frontiers in*

Microbiology, *11*, 575586. https://doi.org/10.3389/fmicb.2020.575586

Nessar, R., Cambau, E., Reyrat, J. M., Murray, A., & Gicquel, B. (2012). *Mycobacterium abscessus*: a new antibiotic nightmare. *The Journal of Antimicrobial Chemotherapy*, *67*(4), 810–818. https://doi.org/10.1093/jac/dkr578

Nett, M., Ikeda, H., & Moore, B. S. (2009). Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Natural Product Reports*, *26*(11), 1362–1384. https://doi.org/10.1039/b817069j

Neukamm, J., Pfrengle, S., Molak, M., Seitz, A., Francken, M., Eppenberger, P., Avanzi, C., Reiter, E., Urban, C., Welte, B., Stockhammer, P. W., Teßmann, B., Herbig, A., Harvati, K., Nieselt, K., Krause, J., & Schuenemann, V. J. (2020). 2000-year-old pathogen genomes reconstructed from metagenomic analysis of Egyptian mummified individuals. *BMC Biology*, *18*(1), 108. https://doi.org/10.1186/s12915-020-00839-8

Newton, G. L., Buchmeier, N., & Fahey, R. C. (2008). Biosynthesis and functions of mycothiol, the unique protective thiol of Actinobacteria. *Microbiology and Molecular Biology Reviews*, *72*(3), 471–494. https://doi.org/10.1128/MMBR.00008-08

Niva, M., Hernesmaa, A., Haahtela, K., Salkinoja-Salonen, M., Sivonen, K., & Haukka, K. (2006). Actinobacterial communities of boreal forest soil and lake water are rich in mycobacteria. *Boreal Environment Research*, *11*(1), 45–53.

Nobre, A., Alarico, S., Maranha, A., Mendes, V., & Empadinhas, N. (2014). The molecular biology of mycobacterial trehalose in the quest for advanced tuberculosis therapies. *Microbiology (United Kingdom)*, *160*(Pt 8), 1547–1570. https://doi.org/10.1099/mic.0.075895-0

Norton, C. D., LeChevallier, M. W., & Falkinham, J. O. (2004). Survival of *Mycobacterium avium* in a model distribution system. *Water Research*, *38*(6), 1457–1466. https://doi.org/10.1016/j.watres.2003.07.008

Nouioui, I., Carro, L., García-López, M., Meier-Kolthoff, J. P., Woyke, T., Kyrpides, N. C., Pukall, R., Klenk, H. P., Goodfellow, M., & Göker, M. (2018). Genome-based taxonomic classification of the phylum actinobacteria. *Frontiers in Microbiology*, *9*, 2007. https://doi.org/10.3389/fmicb.2018.02007

Nouioui, I., Klenk, H.-P., Igual, M., Gulvik, C. A., Lasker, B. A., & McQuiston, J. R. (2019). *Streptacidiphilus bronchialis* sp. nov., a ciprofloxacin-resistant bacterium from a human clinical specimen; reclassification of *Streptomyces griseoplanus* as *Streptacidiphilus griseoplanus* comb. nov. and emended description of the genus *Streptacidiphilus*. *International Journal of Systematic and Evolutionary Microbiology*, *69*(4), 1047–1056. https://doi.org/10.1099/ijsem.0.003267

Nunes-Costa, D. (2015). *Molecular and biochemical characterization of a rare glucokinase with a cryptic function in environmental mycobacteria* [University of Coimbra, Master's thesis]. Available at: http://hdl.handle.net/10316/32158

Nunes-Costa, D., Alarico, S., Dalcolmo, M. P., Correia-Neves, M., & Empadinhas, N. (2016). The looming tide of nontuberculous mycobacterial infections in Portugal and Brazil. *Tuberculosis*, *96*, 107–119. https://doi.org/10.1016/j.tube.2015.09.006

Nunes-Costa, D., Maranha, A., Costa, M., Alarico, S., & Empadinhas, N. (2017). Glucosylglycerate metabolism, bioversatility and mycobacterial survival. *Glycobiology*, *27*(3), 213–227. https://doi.org/10.1093/glycob/cww132

O'Callaghan, A., & van Sinderen, D. (2016). Bifidobacteria and their role as members of the human gut microbiota. *Frontiers in Microbiology*, *7*, 925. https://doi.org/10.3389/fmicb.2016.00925

Pardeshi, P., Rao, K. K., & Balaji, P. V. (2017). Rv3634c from *Mycobacterium tuberculosis* H37Rv encodes an enzyme with UDP-Gal/Glc and UDP-GalNAc 4-epimerase activities. *PLoS ONE*, *12*(4), e0175193. https://doi.org/10.1371/journal.pone.0175193

Park, J. K., Wang, L. X., & Roseman, S. (2002). Isolation of a glucosamine-specific kinase, a unique enzyme of *Vibrio cholerae*. *Journal of Biological Chemistry*, *277*(18), 15573–15578. https://doi.org/10.1074/jbc.M107953200

Park, J. W., Park, S. R., Nepal, K. K., Han, A. R., Ban, Y. H., Yoo, Y. J., Kim, E. J., Kim, E. M., Kim, D., Sohng, J. K., & Yoon, Y. J. (2011). Discovery of parallel pathways of kanamycin biosynthesis allows antibiotic manipulation. *Nature Chemical Biology*, *7*(11), 843–852. https://doi.org/10.1038/nchembio.671

Peng, W., Zou, L., Bhamidi, S., McNeil, M. R., & Lowary, T. L. (2012). The galactosamine residue in mycobacterial arabinogalactan is α-linked. *Journal of Organic Chemistry*, *77*(21), 9826–9832. https://doi.org/10.1021/jo301393s

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

https://doi.org/10.1371/journal.pone.0003748

Pérez, E., Constant, P., Lemassu, A., Laval, F., Daffé, M., & Guilhot, C. (2004). Characterization of three glycosyltransferases involved in the biosynthesis of the phenolic glycolipid antigens from the *Mycobacterium tuberculosis* complex. *Journal of Biological Chemistry*, *279*(41), 42574–42583. https://doi.org/10.1074/jbc.M406246200

Pérez, J., Castañeda-García, A., Jenke-Kodama, H., Müller, R., & Muñoz-Dorado, J. (2008). Eukaryotic-like protein kinases in the prokaryotes and the myxobacterial kinome. *PNAS*, *105*(41), 15950–15955. https://doi.org/10.1073/pnas.0806851105

Peterz, M., Butot, S., Jagadeesan, B., Bakker, D., & Donaghy, J. (2016). Thermal Inactivation of *Mycobacterium avium* subsp. *paratuberculosis* in Artificially Contaminated Milk by Direct Steam Injection. *Applied and Environmental Microbiology*, *82*(9), 2800–2808. https://doi.org/10.1128/AEM.04042-15.Editor

Peto, H. M., Pratt, R. H., Harrington, T. a, LoBue, P. a, & Armstrong, L. R. (2009). Epidemiology of extrapulmonary tuberculosis in the United States, 1993-2006. *Clinical Infectious Diseases*, *49*(9), 1350–1357. https://doi.org/10.1086/605559

Petsko, G. A. (2001). Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat." *Methods in Enzymology*, *334*, 469–478. https://doi.org/10.1016/S0076-6879(01)34486-5

Polsfuss, S., Hofmann-Thiel, S., Merker, M., Krieger, D., Niemann, S., Rüssmann, H., Schönfeld, N., Hoffmann, H., & Kranzer, K. (2019). Emergence of Low-level Delamanid and Bedaquiline Resistance During Extremely Drug-resistant Tuberculosis Treatment. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *69*(7), 1229–1231. https://doi.org/10.1093/cid/ciz074

Prevots, D. R., & Marras, T. K. (2015). Epidemiology of Human Pulmonary Infection with Nontuberculous Mycobacteria. *Clinics in Chest Medicine*, *36*(1), 13–34. https://doi.org/10.1016/j.ccm.2014.10.002

Primm, T. P., Lucero, C. A., & Falkinham, J. O. (2004). Health Impacts of Environmental Mycobacteria. In *Clinical Microbiology Reviews* (Vol. 17, Issue 1, pp. 98–106). https://doi.org/10.1128/CMR.17.1.98-106.2004

Qasba, P. K., Ramakrishnan, B., & Boeggeman, E. (2005). Substrate-induced conformational changes in glycosyltransferases. *Trends in Biochemical Sciences*, *30*(1), 53–62. https://doi.org/10.1016/j.tibs.2004.11.005

Quigley, J., Hughitt, V. K., Velikovsky, C. A., Mariuzza, R. A., El-Sayed, N. M., & Briken, V. (2017). The cell wall lipid PDIM contributes to phagosomal escape and host cell exit of *Mycobacterium tuberculosis*. *MBio*, *8*(2), e00148-17. https://doi.org/10.1128/mBio.00148-17

Rabkin, C. S., Galaid, E. I., Hollis, D. G., Weaver, R. E., Dees, S. B., Kai, A., Moss, C. W., Sandhu, K. K., & Broome, C. V. (1985). Thermophilic bacteria: a new cause of human disease. *Journal of Clinical Microbiology*, *21*(4), 553–557. https://doi.org/10.1128/jcm.21.4.553-557.1985

Radestock, S., & Gohlke, H. (2011). Protein rigidity and thermophilic adaptation. *Proteins*, *79*(4), 1089–1108. https://doi.org/10.1002/prot.22946

Rahlwes, K. C., Sparks, I. L., & Morita, Y. S. (2019). Cell Walls and Membranes of Actinobacteria. In: Kuhn, A. (Ed.) *Bacterial Cell Walls and Membranes*. Subcellular Biochemistry, vol 92. Springer, Cham. doi:10.1007/978-3-030-18768-2_13

Rawat, M., & Av-Gay, Y. (2007). Mycothiol-dependent proteins in actinomycetes. *FEMS Microbiology Reviews*, *31*(3), 278–292. https://doi.org/10.1111/j.1574-6976.2006.00062.x

Raymond, J. B., Mahapatra, S., Crick, D. C., & Pavelka, M. S. (2005). Identification of the *namH* gene, encoding the hydroxylase responsible for the N-glycolylation of the mycobacterial peptidoglycan. *Journal of Biological Chemistry*, *280*(1), 326–333. https://doi.org/10.1074/jbc.M411006200

Rigali, S., Nothaft, H., Noens, E. E. E., Schlicht, M., Colson, S., Müller, M., Joris, B., Koerten, H. K., Hopwood, D. A., Titgemeyer, F., & Van Wezel, G. P. (2006). The sugar phosphotransferase system of *Streptomyces coelicolor* is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development. *Molecular Microbiology*, *61*(5), 1237–1251. https://doi.org/10.1111/j.1365-2958.2006.05319.x

Rigali, S., Titgemeyer, F., Barends, S., Mulder, S., Thomae, A. W., Hopwood, D. A., & van Wezel, G. P. (2008). Feast or famine: The global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. *EMBO Reports*, *9*(7), 670–675. https://doi.org/10.1038/embor.2008.83

Ripoll-Rozada, J., Costa, M., Manso, J. A., Maranha, A., Miranda, V., Sequeira, A., Ventura, M. R., Macedo-Ribeiro, S., Pereira, P. J. B., & Empadinhas, N. (2019). Biosynthesis of mycobacterial methylmannose

polysaccharides requires a unique 1-O-methyltransferase specific for 3-O-methylated mannosides. *PNAS*, *116*(3), 835 LP – 844. https://doi.org/10.1073/pnas.1813450116

Rodrigo-Unzueta, A., Ghirardello, M., Urresti, S., Delso, I., Giganti, D., Anso, I., Trastoy, B., Comino, N., Tersa, M., D'Angelo, C., Cifuente, J. O., Marina, A., Liebau, J., Mäler, L., Chenal, A., Albesa-Jové, D., Merino, P., & Guerin, M. E. (2020). Dissecting the Structural and Chemical Determinants of the "Open-To-Closed" Motion in the Mannosyltransferase PimA from Mycobacteria. *Biochemistry*, *59*(32), 2934–2945. https://doi.org/10.1021/acs.biochem.0c00376

Romero-García, J., Francisco, C., Biarnés, X., & Planas, A. (2013). Structure-function features of a mycoplasma glycolipid synthase derived from structural data integration, molecular simulations, and mutational analysis. *PLoS ONE*, *8*(12), e81990. https://doi.org/10.1371/journal.pone.0081990

Rothschild, B. M., Martin, L. D., Lev, G., Bercovier, H., Bar-Gal, G. K., Greenblatt, C., Donoghue, H., Spigelman, M., & Brittain, D. (2001). *Mycobacterium tuberculosis* complex DNA from an extinct bison dated 17,000 years before the present. *Clinical Infectious Diseases*, *33*(3), 305–311. https://doi.org/10.1086/321886

Roy, R., Usha, V., Kermani, A., Scott, D. J., Hyde, E. I., Besra, G. S., Alderwick, L. J., & Fütterer, K. (2013). Synthesis of α-glucan in mycobacteria involves a hetero-octameric complex of trehalose synthase TreS and Maltokinase Pep2. *ACS Chemical Biology*, *8*(10), 2245–2255. https://doi.org/10.1021/cb400508k

Ruis, C., Bryant, J. M., Bell, S. C., Thomson, R., Davidson, R. M., Hasan, N. A., van Ingen, J., Strong, M., Floto, R. A., & Parkhill, J. (2021). Dissemination of *Mycobacterium abscessus* via global transmission networks. *Nature Microbiology*, *6*(10), 1279–1288. https://doi.org/10.1038/s41564-021-00963-3

Runyon, E. H. (1959). Anonymous mycobacteria in pulmonary disease. *The Medical Clinics of North America*, *43*(1), 273–290. https://doi.org/10.1016/S0025-7125(16)34193-1

Rutala, W. A., & Weber, D. J. (2013). Disinfection and sterilization: An overview. *American Journal of Infection Control*, *41*, S2–S5. https://doi.org/10.1002/9781118637272.ch42

Sajed, T., Marcu, A., Ramirez, M., Pon, A., Guo, A. C., Knox, C., Wilson, M., Grant, J. R., Djoumbou, Y., & Wishart, D. S. (2016). ECMDB 2.0: A richer resource for understanding the biochemistry of *E. coli*. *Nucleic Acids Research*, *44*(D1), D495-501. https://doi.org/10.1093/nar/gkv1060

Salam, N., Jiao, J. Y., Zhang, X. T., & Li, W. J. (2020). Update on the classification of higher ranks in the phylum Actinobacteria. *International Journal of Systematic and Evolutionary Microbiology*, *70*(2), 1331–1355. https://doi.org/10.1099/ijsem.0.003920

Salzer, H. J. F., Chitechi, B., Hillemann, D., Mandl, M., Paar, C., Mitterhumer, M., Lamprecht, B., & Maurer, F. P. (2020). Nontuberculous mycobacterial pulmonary disease from *Mycobacterium hassiacum*, Austria. *Emerging Infectious Diseases*, *26*(11), 2776–2778. https://doi.org/10.3201/eid2611.191718

Samanovic, M. I., Tu, S., Novák, O. ej, Iyer, L. M., McAllister, F. E., Aravind, L., Gygi, S. P., Hubbard, S. R., Strnad, M., & Darwin, K. H. (2015). Proteasomal Control of Cytokinin Synthesis Protects *Mycobacterium tuberculosis* against Nitric Oxide. *Molecular Cell*, *57*(6), 984–994. https://doi.org/10.1016/j.molcel.2015.01.024

Sambamoorthy, G., & Raman, K. (2018). Understanding the evolution of functional redundancy in metabolic networks. *Bioinformatics*, *34*(17), i981–i987. https://doi.org/10.1093/bioinformatics/bty604

Sambou, T., Dinadayala, P., Stadthagen, G., Barilone, N., Bordat, Y., Constant, P., Levillain, F., Neyrolles, O., Gicquel, B., Lemassu, A., Daffé, M., & Jackson, M. (2008). Capsular glucan and intracellular glycogen of *Mycobacterium tuberculosis*: biosynthesis and impact on the persistence in mice. *Molecular Microbiology*, *70*(3), 762–774. https://doi.org/10.1111/j.1365-2958.2008.06445.x

Santos, R., de Carvalho, C. C. C. R., Stevenson, A., Grant, I. R., & Hallsworth, J. E. (2015). Extraordinary solute-stress tolerance contributes to the environmental tenacity of mycobacteria. *Environmental Microbiology Reports*, *7*(5), 746–764. https://doi.org/10.1111/1758-2229.12306

Santos, R., Fernandes, J., Fernandes, N., Oliveira, F., & Cadete, M. (2007). *Mycobacterium parascrofulaceum* in acidic hot springs in Yellowstone National Park. *Applied and Environmental Microbiology*, *73*(15), 5071–5073. https://doi.org/10.1128/AEM.00353-07

Sarkar, D., Sidhu, M., Singh, A., Chen, J., Lammas, D. A., van der Sar, A. M., Besra, G. S., & Bhatt, A. (2011). Identification of a glycosyltransferase from *Mycobacterium marinum* involved in addition of a caryophyllose moiety in lipooligosaccharides. *Journal of Bacteriology*, *193*(9), 2336–2340. https://doi.org/10.1128/JB.00065-11

Sato, S., Kitamoto, D., & Habe, H. (2014). *In vitro* evaluation of glyceric acid and its glucosyl derivative, α-glucosylglyceric acid, as cell proliferation inducers and protective solutes. *Bioscience, Biotechnology and Biochemistry*, *78*(7), 1183–1186. https://doi.org/10.1080/09168451.2014.885823

Sawangwan, T., Goedl, C., & Nidetzky, B. (2010). Glucosylglycerol and glucosylglycerate as enzyme stabilizers. *Biotechnology Journal*, *5*(2), 187–191. https://doi.org/10.1002/biot.200900197

Scheeff, E. D., & Bourne, P. E. (2005). Structural evolution of the protein kinase-like superfamily. *PLoS Computational Biology*, *1*(5), e49. https://doi.org/10.1371/journal.pcbi.0010049

Schildkraut, J. A., Gallagher, J., Morimoto, K., Lange, C., Haworth, C., Floto, R. A., Hoefsloot, W., Griffith, D. E., Wagner, D., & Ingen, J. van. (2020). Epidemiology of nontuberculous mycobacterial pulmonary disease in Europe and Japan by Delphi estimation. *Respiratory Medicine*, *173*, 106164. https://doi.org/10.1016/j.rmed.2020.106164

Schneider, M. (2004). A rational approach to maximize success rate in target discovery. *Archiv Der Pharmazie*, *337*(12), 625–633. https://doi.org/10.1002/ardp.200400913

Schrey, S. D., & Tarkka, M. T. (2008). Friends and foes: *Streptomycetes* as modulators of plant disease and symbiosis. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *94*(1), 11–19. https://doi.org/10.1007/s10482-008-9241-3

Schröder, K. H., Naumann, L., Kroppenstedt, R. M., & Reischl, U. (1997). *Mycobacterium hassiacum* sp. nov., a new rapidly growing thermophilic mycobacterium. *International Journal of Systematic Bacteriology*, *47*(12), 86–91. https://doi.org/10.1099/00207713-47-1-86

Schuldt, L., Weyand, S., Kefala, G., & Weiss, M. S. (2009). The three-dimensional Structure of a mycobacterial DapD provides insights into DapD diversity and reveals unexpected particulars about the enzymatic mechanism. *Journal of Molecular Biology*, *389*(5), 863–879. https://doi.org/10.1016/j.jmb.2009.04.046

Schulze-Robbecke, R., & Buchholtz, K. (1992). Heat susceptibility of aquatic mycobacteria. *Applied and Environmental Microbiology*, *58*(6), 1869–1873.

Scollard, D. M., Truman, R. W., & Ebenezer, G. J. (2015). Mechanisms of nerve injury in leprosy. *Clinics in Dermatology*, *33*(1), 46–54. https://doi.org/10.1016/j.clindermatol.2014.07.008

Scossa, F., & Fernie, A. R. (2020). The evolution of metabolism: How to test evolutionary hypotheses at the genomic level. *Computational and Structural Biotechnology Journal*, *18*, 482–500. https://doi.org/10.1016/j.csbj.2020.02.009

Senhaji-Kacha, A., Esteban, J., & Garcia-Quintanilla, M. (2021). Considerations for Phage Therapy Against *Mycobacterium abscessus*. *Frontiers in Microbiology*, *11*, 609017. https://doi.org/10.3389/fmicb.2020.609017

Setia, M. S., Steinmaus, C., Ho, C. S., & Rutherford, G. W. (2006). The role of BCG in prevention of leprosy: a meta-analysis. *Lancet Infectious Diseases*, *6*(3), 162–170.

Shapiro, A. B., Walkup, G. K., & Keating, T. A. (2009). Correction for interference by test samples in high-throughput assays. *Journal of Biomolecular Screening*, *14*(8), 1008–1016. https://doi.org/10.1177/1087057109341768

Sharma, S. K., & Mohan, A. (2004). Extrapulmonary tuberculosis. *Indian J Med Res*, *120*(4), 316–353. https://doi.org/10.1016/j.mpmed.2013.10.008

Shepard, C. C. (1960). The experimental disease that follows the injection of human leprosy bacilli into foot-pads of mice. *The Journal of Experimental Medicine*, *112*(3), 445–454. https://doi.org/10.1084/jem.112.3.445

Shleeva, M. O., Trutneva, K. A., Demina, G. R., Zinin, A. I., Sorokoumova, G. M., Laptinskaya, P. K., Shumkova, E. S., & Kaprelyants, A. S. (2017). Free trehalose accumulation in dormant *Mycobacterium smegmatis* cells and its breakdown in early resuscitation phase. *Frontiers in Microbiology*, *8*, 524. https://doi.org/10.3389/fmicb.2017.00524

Sivalingam, P., Hong, K., Pote, J., & Prabakar, K. (2019). Extreme environment streptomyces: Potential sources for new antibacterial and anticancer drug leads? *International Journal of Microbiology*, *2019*, 5283948. https://doi.org/10.1155/2019/5283948

Skinnider, M. A., Johnston, C. W., Merwin, N. J., Dejong, C. A., & Magarvey, N. A. (2018). Global analysis of prokaryotic tRNA-derived cyclodipeptide biosynthesis. *BMC Genomics*, *19*(1), 45. https://doi.org/10.1186/s12864-018-4435-1

Škovierová, H., Larrouy-Maumus, G., Pham, H., Belanová, M., Barilone, N., DasGupta, A., Mikušová, K., Gicquel, B., Gilleron, M., Brennan, P. J., Puzo, G., Nigou, J., & Jackson, M. (2010). Biosynthetic origin of the galactosamine substituent of arabinogalactan in *Mycobacterium tuberculosis*. *Journal of Biological Chemistry*, *285*(53), 41348–41355. https://doi.org/10.1074/jbc.M110.188110

Smeulders, M. J., Keer, J., Speight, R. a, & Williams, H. D. (1999). Adaptation of *Mycobacterium smegmatis* to Stationary Phase. *Journal of Bacteriology*, *181*(1), 270–283.

Smith, C. S., Aerts, A., Saunderson, P., Kawuma, J., Kita, E., & Virmond, M. (2017). Multidrug therapy for leprosy: a game changer on the path to elimination. *The Lancet Infectious Diseases*, *17*(9), e293–e297. https://doi.org/10.1016/S1473-3099(17)30418-8

Smith, W. L., & Ballou, C. E. (1973). The 6- O -Methylglucose-containing Lipopolysaccharides of *Mycobacterium phlei*: Locations of the Neutral and Acidic Acyl Groups. *The Journal of Biological Chemistry*, *248*(20), 7118–7125.

Soni, V., Upadhayay, S., Suryadevara, P., Samla, G., Singh, A., Yogeeswari, P., Sriram, D., & Nandicoori, V. K. (2015). Depletion of *M. tuberculosis* GlmU from Infected Murine Lungs Effects the Clearance of the Pathogen. *PLoS Pathogens*, *11*(10), e1005235. https://doi.org/10.1371/journal.ppat.1005235

Stadthagen, G., Sambou, T., Guerin, M., Barilone, N., Boudou, F., Korduláková, J., Charles, P., Alzari, P. M., Lemassu, A., Daffé, M., Puzo, G., Gicquel, B., Rivière, M., & Jackson, M. (2007). Genetic basis for the biosynthesis of methylglucose lipopolysaccharides in *Mycobacterium tuberculosis*. *Journal of Biological Chemistry*, *282*(37), 27270–27276. https://doi.org/10.1074/jbc.M702676200

Stanford, J., Rook, G. A. W., Bahr, G. M., Dowlati, Y., Ganapati, R., Ghazi Saidi, K., Lucas, S., Ramu, G., Torres, P., Minh Ly, H., & Anstey, N. (1990). *Mycobacterium vaccae* in immunoprophylaxis and immunotherapy of leprosy and tuberculosis. *Vaccine*, *8*(6), 525–530. https://doi.org/https://doi.org/10.1016/0264-410X(90)90002-4

Stanford, J., & Stanford, C. (2012). Mycobacteria and their world. *International Journal of Mycobacteriology*, *1*(1), 3–12. https://doi.org/10.1016/j.ijmyco.2012.01.001

Storrs, E. E., Walsh, G. P., Burchfield, H. P., & Binford, C. H. (1974). Leprosy in the armadillo: new model for biomedical research. *Science*, *183*(4127), 851–852. https://doi.org/10.1126/science.183.4127.851

Strack, R. (2020). Cryo-EM goes atomic. *Nature Methods*, *17*(12), 1175. https://doi.org/10.1038/s41592-020-01014-1

Subramani, R., & Aalbersberg, W. (2012). Marine actinomycetes: An ongoing source of novel bioactive metabolites. *Microbiological Research*, *167*(10), 571–580. https://doi.org/10.1016/j.micres.2012.06.005

Suy, F., Carricajo, A., Grattard, F., Cazorla, C., Denis, C., Girardin, P., Lucht, F., & Botelho-Nevers, E. (2013). Infection due to *Mycobacterium thermoresistibile*: A case associated with an orthopedic device. *Journal of Clinical Microbiology*, *51*(9), 3154–3156. https://doi.org/10.1128/JCM.00925-13

Suzuki, K., Akama, T., Kawashima, A., Yoshihara, A., Yotsu, R. R., & Ishii, N. (2012). Current status of leprosy: epidemiology, basic science and clinical perspectives. *The Journal of Dermatology*, *39*(2), 121–129. https://doi.org/10.1111/j.1346-8138.2011.01370.x

Swenson, C., Zerbe, C. S., & Fennelly, K. (2018). Host Variability in NTM Disease: Implications for Research Needs. *Frontiers in Microbiology*, *9*, 2901. https://doi.org/10.3389/fmicb.2018.02901

Świątek-Połatyńska, M. A., Bucca, G., Laing, E., Gubbens, J., Titgemeyer, F., Smith, C. P., Rigali, S., & Van Wezel, G. P. (2015). Genome-wide analysis of *in vivo* binding of the master regulator DasR in *Streptomyces coelicolor* identifies novel non-canonical targets. *PLoS ONE*, *10*(4), e0122479. https://doi.org/10.1371/journal.pone.0122479

Świątek, M. A., Tenconi, E., Rigali, S., & Van Wezel, G. P. (2012). Functional analysis of the *N*-acetylglucosamine metabolic genes of *Streptomyces coelicolor* and role in control of development and antibiotic production. *Journal of Bacteriology*, *194*(5), 1136–1144. https://doi.org/10.1128/JB.06370-11

Syson, K., Stevenson, C. E. M., Lawson, D. M., & Bornemann, S. (2020). Structure of the *Mycobacterium smegmatis* α-maltose-1-phosphate synthase GlgM. *Acta Crystallographica Section F: Structural Biology Communications*, *76*(Pt 4), 175–181. https://doi.org/10.1107/S2053230X20004343

Tabouret, G., Astarie-Dequeker, C., Demangel, C., Malaga, W., Constant, P., Ray, A., Honoré, N., Bello, N. F., Perez, E., Daffé, M., & Guilhot, C. (2010). *Mycobacterium leprae* Phenolglycolipid-1 Expressed by Engineered *M. bovis* BCG modulates early interaction with human phagocytes. *PLoS Pathogens*, *6*(10), e1001159. https://doi.org/10.1371/journal.ppat.1001159

Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D. L., Pulcini, C., Kahlmeter, G., Kluytmans, J., Carmeli, Y., Ouellette, M., Outterson, K., Patel, J., Cavaleri, M., Cox, E. M., Houchens, C. R., Grayson, M. L., Hansen, P., Singh, N., … Zorzet, A. (2018). Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The Lancet Infectious Diseases*, *18*(3), 318–327. https://doi.org/10.1016/S1473-3099(17)30753-3

Tatham, E., Sundaram Chavadi, S., Mohandas, P., Edupuganti, U. R., Angala, S. K., Chatterjee, D., & Quadri, L. E. N. (2012). Production of mycobacterial cell wall glycopeptidolipids requires a member of the MbtH-like

protein family. *BMC Microbiology*, *12*(1), 118. https://doi.org/10.1186/1471-2180-12-118

Taujale, R., Venkat, A., Huang, L. C., Zhou, Z., Yeung, W., Rasheed, K. M., Li, S., Edison, A. S., Moremen, K. W., & Kannan, N. (2020). Deep evolutionary analysis reveals the design principles of fold a glycosyltransferases. *ELife*, *9*, e54532. https://doi.org/10.7554/eLife.54532

Theuretzbacher, U., Bush, K., Harbarth, S., Paul, M., Rex, J. H., Tacconelli, E., & Thwaites, G. E. (2020). Critical analysis of antibacterial agents in clinical development. *Nature Reviews Microbiology*, *18*(5), 286–298. https://doi.org/10.1038/s41579-020-0340-0

Tiago, I., Maranha, A., Mendes, V., Alarico, S., Moynihan, P. J., Clarke, A. J., Macedo-Ribeiro, S., Pereira, P. J. B., & Empadinhas, N. (2012). Genome sequence of *Mycobacterium hassiacum* DSM 44199, a rare source of heat-stable mycobacterial proteins. *Journal of Bacteriology*, *194*(24), 7010–7011. https://doi.org/10.1128/JB.01880-12

Tindall, B. J. (1999). Misunderstanding the Bacteriological Code. *International Journal of Systematic Bacteriology*, *49*(3), 1313–1316. https://doi.org/10.1099/00207713-49-3-1313

Titgemeyer, F., Amon, J., Parche, S., Mahfoud, M., Bail, J., Schlicht, M., Rehm, N., Hillmann, D., Stephan, J., Walter, B., Burkovski, A., & Niederweis, M. (2007). A genomic view of sugar transport in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *Journal of Bacteriology*, *189*(16), 5903–5915. https://doi.org/10.1128/JB.00257-07

Toriyama, S., Yano, I., Masui, M., Kusunose, E., Kusunose, M., & Akimori, N. (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. (2009). Clinical Manifestations of Nontuberculous Mycobacteria Infections. *Clinical Microbiology and Infection*, *15*(10), 906–910.

Tortoli, E., Brown-Elliott, B. A., Chalmers, J. D., Cirillo, D. M., Daley, C. L., Emler, S., Andres Floto, R., Garcia, M. J., Hoefsloot, W., Koh, W. J., Lange, C., Loebinger, M., Maurer, F. P., Morimoto, K., Niemann, S., Richter, E., Turenne, C. Y., Vasireddy, R., Vasireddy, S., … Van Ingen, J. (2019). Same meat, different gravy: Ignore the new names of mycobacteria. *European Respiratory Journal*, *54*(1), 19–21. https://doi.org/10.1183/13993003.00795-2019

Tortoli, E., Fedrizzi, T., Meehan, C. J., Trovato, A., Grottola, A., Giacobazzi, E., Serpini, G. F., Tagliazucchi, S., Fabio, A., Bettua, C., Bertorelli, R., Frascaro, F., De Sanctis, V., Pecorari, M., Jousson, O., Segata, N., & Cirillo, D. M. (2017). The new phylogeny of the genus *Mycobacterium*: The old and the news. *Infection, Genetics and Evolution*, *56*, 19–25. https://doi.org/10.1016/j.meegid.2017.10.013

Tortoli, E., Kohl, T. A., Trovato, A., Baldan, R., Campana, S., Cariani, L., Colombo, C., Costa, D., Cristadoro, S., Di Serio, M. C., Manca, A., Pizzamiglio, G., Rancoita, P. M. V., Rossolini, G. M., Taccetti, G., Teri, A., Niemann, S., & Cirillo, D. M. (2017). *Mycobacterium abscessus* in patients with cystic fibrosis: Low impact of inter-human transmission in Italy. *European Respiratory Journal*, *50*(1), 6–9. https://doi.org/10.1183/13993003.02525-2016

Tortoli, E., Reischl, U., Besozzi, G., & Emler, S. (1998). Characterization of an isolate belonging to the newly described species *Mycobacterium hassiacum*. *Diagnostic Microbiology and Infectious Disease*, *30*(3), 193–196. https://doi.org/10.1016/S0732-8893(97)00242-3

Trujillo, M. E., Velázquez, E., Kroppenstedt, R. M., Schumann, P., Rivas, R., Mateos, P. F., & Martínez-Molina, E. (2004). *Mycobacterium psychrotolerans* sp. nov., isolated from pond water near a uranium mine. *International Journal of Systematic and Evolutionary Microbiology*, *54*(5), 1459–1463. https://doi.org/10.1099/ijs.0.02938-0

Tyson, J., & Sockett, R. E. (2017). Predatory Bacteria: Moving from Curiosity Towards Curative. *Trends in Microbiology*, *25*(2), 90–91. https://doi.org/10.1016/j.tim.2016.12.011

Urresti, S., Albesa-Jové, D., Schaeffer, F., Pham, H. T., Kaur, D., Gest, P., Van Der Woerd, M. J., Carreras-González, A., López-Fernández, S., Alzari, P. M., Brennan, P. J., Jackson, M., & Guerin, M. E. (2012). Mechanistic insights into the retaining glucosyl-3-phosphoglycerate synthase from mycobacteria. *Journal of Biological Chemistry*, *287*(29), 24649–24661. https://doi.org/10.1074/jbc.M112.368191

Usha, V., Lloyd, A. J., Roper, D. I., Dowson, C. G., Kozlov, G., Gehring, K., Chauhan, S., Imam, H. T., Blindauer, C. A., & Besra, G. S. (2016). Reconstruction of diaminopimelic acid biosynthesis allows characterisation of *Mycobacterium tuberculosis N*-succinyl-L,L-diaminopimelic acid desuccinylase. *Scientific Reports*, *6*, 23191. https://doi.org/10.1038/srep23191

van Bergeijk, D. A., Terlouw, B. R., Medema, M. H., & van Wezel, G. P. (2020). Ecology and genomics of Actinobacteria: new concepts for natural product discovery. *Nature Reviews Microbiology*, *18*(10), 546–558.

https://doi.org/10.1038/s41579-020-0379-y

van Crevel, R., Ottenhoff, T. H. M., & van der Meer, J. W. M. (2002). Innate immunity to *Mycobacterium tuberculosis*. *Clinical Microbiology Reviews*, *15*(2), 294–309. https://doi.org/10.1128/CMR.15.2.294

van der Heul, H., Bilyk, B., & McDowall, K. (2018). Regulation of antibiotic production in Actinobacteria: new perspectives from the post-genomic era. *Natural Product Reports*, *35*(6), 575–604. https://doi.org/https://doi.org/10.1039/c8np00012c

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Zídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., … Velankar, S. (2022). AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, *50*(D1), D439–D444. https://doi.org/10.1093/nar/gkab1061

Veleti, S. K., Lindenberger, J. J., Ronning, D. R., & Sucheck, S. J. (2014). Synthesis of a C-phosphonate mimic of maltose-1-phosphate and inhibition studies on *Mycobacterium tuberculosis* GlgE. *Bioorganic and Medicinal Chemistry*, *22*(4), 1404–1411. https://doi.org/10.1016/j.bmc.2013.12.058

Vergne, I., Fratti, R. A. R., Hill, P. J. P., Chua, J., Belisle, J., & Deretic, V. (2004). *Mycobacterium tuberculosis* phagosome maturation arrest: mycobacterial phosphatidylinositol analog phosphatidylinositol mannoside stimulates early endosomal fusion. *Molecular Biology of the Cell*, *15*(2), 751–760. https://doi.org/10.1091/mbc.E03

Vieille, C., & Zeikus, G. J. (2001). Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. *Microbiology and Molecular Biology Reviews*, *65*(1), 1–43. https://doi.org/10.1128/MMBR.65.1.1

Viens, P., Dubeau, M. P., Kimura, A., Desaki, Y., Shinya, T., Shibuya, N., Saito, A., & Brzezinski, R. (2015). Uptake of chitosan-derived D-glucosamine oligosaccharides in *Streptomyces coelicolor* A3(2). *FEMS Microbiology Letters*, *362*(9), fnv048. https://doi.org/10.1093/femsle/fnv048

Wagner, D., Lipman, M., Cooray, S., Ringshausen, F. C., Morimoto, K., Koh, W.-J., & Thomson, R. (2019). Global Epidemiology of NTM Disease (Except Northern America). In D. E. Griffith (Ed.), *Nontuberculous Mycobacterial Disease: A Comprehensive Approach to Diagnosis and Management* (pp. 163–260). Springer International Publishing. https://doi.org/10.1007/978-3-319-93473-0_8

Walker, E. H., Perisic, O., Ried, C., Stephens, L., & Williams, R. L. (1999). Structural insights into phosphoinositide 3-kinase catalysis and signalling. *Nature*, *402*(6759), 313–320.

Walsh, D. S., Portaels, F., & Meyers, W. M. (2011). Buruli ulcer: Advances in understanding *Mycobacterium ulcerans* infection. *Dermatologic Clinics*, *29*(1), 1–8. https://doi.org/10.1016/j.det.2010.09.006

Wang, R., Li, K., Yu, J., Deng, J., & Chen, Y. (2021). Mutations of folC cause increased susceptibility to sulfamethoxazole in *Mycobacterium tuberculosis*. *Scientific Reports*, *11*, 1352. https://doi.org/10.1038/s41598-020-80213-4

Wang, Y., Ogawa, M., Fukuda, K., Miyamoto, H., & Taniguchi, H. (2006). Isolation and identification of mycobacteria from soils at an illegal dumping site and landfills in Japan. *Microbiology and Immunology*, *50*(7), 513–524.

Ward, A. C., & Allenby, N. E. (2018). Genome mining for the search and discovery of bioactive compounds: The *Streptomyces* paradigm. *FEMS Microbiology Letters*, *365*(24), fny240. https://doi.org/10.1093/femsle/fny240

Welsh, K. J., Hunter, R. L., & Actor, J. K. (2013). Trehalose 6,6′-dimycolate - A coat to regulate tuberculosis immunopathogenesis. *Tuberculosis*, *93*, S3–S9. https://doi.org/10.1016/S1472-9792(13)70003-9

Weymouth-Wilson, A. C. (1997). The role of carbohydrates in biologically active natural products. *Natural Product Reports*, *14*(2), 99–110. https://doi.org/10.1039/np9971400099

Whiley, H., Bentham, R., & Brown, M. H. (2017). Legionella persistence in manufactured water systems: Pasteurization potentially selecting for thermal tolerance. *Frontiers in Microbiology*, *8*, 1330. https://doi.org/10.3389/fmicb.2017.01330

WHO. (2015). Global Action Plan on Antimicrobial Resistance. *World Health Organization*, Geneva Available at: https://www.who.int/publications/i/item/9789241509763

WHO. (2017). Guidelines for the diagnosis, treatment and prevention of leprosy. *World Health Organization*, New Dehli. Available at: https://apps.who.int/iris/bitstream/handle/10665/274127/9789290226383-eng.pdf

WHO. (2020a). Global leprosy (Hansen disease) update, 2019: time to step-up prevention initiatives. *Weekly Epidemiological Record*, *95*(36), 417–440. Available at: http://www.who.int/wer

WHO. (2020b). Global Tuberculosis Report 2020. *World Health Organization*, Geneva. Available at: https://www.who.int/publications/i/item/9789240013131

Wiliams, S. T. (1990). Actinomycetes - The ray fungi. *Mycologist*, *4*(3), 110–114. https://doi.org/10.1016/S0269-915X(09)80036-0

Williams, M. M., Yakrus, M. A., Arduino, M. J., Cooksey, R. C., Crane, C. B., Banerjee, S. N., Hilborn, E. D., & Donlan, R. M. (2009). Structural analysis of biofilm formation by rapidly and slowly growing nontuberculous mycobacteria. *Applied and Environmental Microbiology*, *75*(7), 2091–2098. https://doi.org/10.1128/AEM.00166-09

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica. Section D, Biological Crystallography*, *67*(Pt 4), 235–242. https://doi.org/10.1107/S0907444910045749

Wolfe, J. M., & Moore, D. F. (1992). Isolation of *Mycobacterium thermoresistibile* following augmentation mammaplasty. *Journal of Clinical Microbiology*, *30*(4), 1036–1038.

Woodruff, H. B. (2014). Selman A. Waksman, winner of the 1952 nobel prize for physiology or medicine. *Applied and Environmental Microbiology*, *80*(1), 2–8. https://doi.org/10.1128/AEM.01143-13

Woodruff, P. J., Carlson, B. L., Siridechadilok, B., Pratt, M. R., Senaratne, R. H., Mougous, J. D., Riley, L. W., Williams, S. J., & Bertozzi, C. R. (2004). Trehalose is required for growth of Mycobacterium smegmatis. *The Journal of Biological Chemistry*, *279*(28), 28835–28843. https://doi.org/10.1074/jbc.M313103200

Wu, M. L., Aziz, D. B., Dartois, V., & Dick, T. (2018). NTM drug discovery: status, gaps and the way forward. *Drug Discovery Today*, *23*(8), 1502–1519. https://doi.org/10.1016/j.drudis.2018.04.001

Yakovlieva, L., & Walvoort, M. T. C. (2020). Processivity in Bacterial Glycosyltransferases. *ACS Chemical Biology*, *15*(1), 3–16. https://doi.org/10.1021/acschembio.9b00619

Yamada, H., Chikamatsu, K., Aono, A., Murata, K., Miyazaki, N., Kayama, Y., Bhatt, A., Fujiwara, N., Maeda, S., & Mitarai, S. (2020). Fundamental Cell Morphologies Examined With Cryo-TEM of the Species in the Novel Five Genera Robustly Correlate With New Classification in Family *Mycobacteriaceae*. *Frontiers in Microbiology*, *11*, 562395. https://doi.org/10.3389/fmicb.2020.562395

Yan, J., Kevat, A., Martinez, E., Teese, N., Johnson, K., Ranganathan, S., Harrison, J., Massie, J., & Daley, A. (2020). Investigating transmission of *Mycobacterium abscessus* amongst children in an Australian cystic fibrosis centre. *Journal of Cystic Fibrosis*, *19*(2), 219–224. https://doi.org/10.1016/j.jcf.2019.02.011

Yan, L., & Liu, Y. (2020). The Retaining Mechanism of Xylose Transfer Catalyzed by Xyloside α-1,3-Xylosyltransferase (XXYLT1): A Quantum Mechanics/Molecular Mechanics Study. *Journal of Chemical Information and Modeling*, *60*(3), 1585–1594. https://doi.org/10.1021/acs.jcim.9b00976

Yaqub, O. (2018). Serendipity: Towards a taxonomy and a theory. *Research Policy*, *47*(1), 169–179. https://doi.org/10.1016/j.respol.2017.10.007

Yip, M. J., Porter, J. L., Fyfe, J. A. M., Lavender, C. J., Portaels, F., Rhodes, M., Kator, H., Colorni, A., Jenkin, G. A., & Stinear, T. (2007). Evolution of *Mycobacterium ulcerans* and other mycolactone-producing mycobacteria from a common *Mycobacterium marinum* progenitor. *Journal of Bacteriology*, *189*(5), 2021–2029. https://doi.org/10.1128/JB.01442-06

Young, P. G., Walanj, R., Lakshmi, V., Byrnes, L. J., Metcalf, P., Baker, E. N., Vakulenko, S. B., & Smith, C. A. (2009). The crystal structures of substrate and nucleotide complexes of *Enterococcus faecium* aminoglycoside-2″-phosphotransferase-IIa [APH(2″)-IIa] provide insights into substrate selectivity in the APH(2″) subfamily. *Journal of Bacteriology*, *191*(13), 4133–4143. https://doi.org/10.1128/JB.00149-09

Yu, H., Takeuchi, M., LeBarron, J., Kantharia, J., London, E., Bakker, H., Haltiwanger, R. S., Li, H., & Takeuchi, H. (2015). Notch-modifying xylosyltransferase-substrate complexes support an SNi-like retaining mechanism. *Nature Chemical Biology*, *11*(11), 847–854. https://doi.org/10.1038/nchembio.1927.Notch-modifying

Yuan, Y., Barrett, D., Zhang, Y., Kahne, D., Sliz, P., & Walker, S. (2007). Crystal structure of a peptidoglycan glycosyltransferase suggests a model for processive glycan chain synthesis Yanqiu. *PNAS*, *104*(13), 5348–5353. https://doi.org/10.1073/pnas.0701160104

Zallot, R., Oberg, N., & Gerlt, J. A. (2019). The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry*, *58*(41), 4169–4182. https://doi.org/10.1021/acs.biochem.9b00735

Zegzouti, H., Zdanovskaia, M., Hsiao, K., & Goueli, S. A. (2009). ADP-Glo: A Bioluminescent and homogeneous ADP monitoring assay for kinases. *Assay and Drug Development Technologies*, *7*(6), 560–572. https://doi.org/10.1089/adt.2009.0222

Zhang, C., & Anderson, A. J. (2013). Utilization of pyrene and benzoate in *Mycobacterium* isolate KMS is regulated differentially by catabolic repression. *Journal of Basic Microbiology*, *53*(1), 81–92. https://doi.org/10.1002/jobm.201100480

Zhang, H., Zhu, F., Yang, T., Ding, L., Zhou, M., Li, J., Haslam, S. M., Dell, A., Erlandsen, H., & Wu, H. (2014). The highly conserved domain of unknown function 1792 has a distinct glycosyltransferase fold. *Nature Communications*, *5*, 4339. https://doi.org/10.1038/ncomms5339

Zhang, P., Zhang, Z., Zhang, L., Wang, J., & Wu, C. (2020). Glycosyltransferase GT1 family: Phylogenetic distribution, substrates coverage, and representative structural features. *Computational and Structural Biotechnology Journal*, *18*, 1383–1390. https://doi.org/10.1016/j.csbj.2020.06.003

Zhang, Z., Bulloch, E. M. M., Bunker, R. D., Baker, E. N., & Squire, C. J. (2009). Structure and function of GlmU from *Mycobacterium tuberculosis*. *Acta Crystallographica Section D: Biological Crystallography*, *65*(3), 275–283. https://doi.org/10.1107/S0907444909001036

Zhao, Y., Park, R. D., & Muzzarelli, R. A. A. (2010). Chitin deacetylases: Properties and applications. *Marine Drugs*, *8*(1), 24–46. https://doi.org/10.3390/md8010024

Zheng, Q., Jiang, D., Zhang, W., Zhang, Q., Zhao, Q., Jin, J., Li, X., Yang, H., Bartlam, M., Shaw, N., Zhou, W., & Rao, Z. (2014). Mechanism of dephosphorylation of glucosyl-3-phosphoglycerate by a histidine phosphatase. *Journal of Biological Chemistry*, *289*(31), 21242–21251. https://doi.org/10.1074/jbc.M114.569913

Zheng, R. B., Jégouzo, S. A. F., Joe, M., Bai, Y., Tran, H. A., Shen, K., Saupe, J., Xia, L., Ahmed, M. F., Liu, Y. H., Patil, P. S., Tripathi, A., Hung, S. C., Taylor, M. E., Lowary, T. L., & Drickamer, K. (2017). Insights into Interactions of Mycobacteria with the Host Innate Immune System from a Novel Array of Synthetic Mycobacterial Glycans. *ACS Chemical Biology*, *12*(12), 2990–3002. https://doi.org/10.1021/acschembio.7b00797

Zuber, B., Chami, M., Houssin, C., Dubochet, J., Griffiths, G., & Daffé, M. (2008). Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. *Journal of Bacteriology*, *190*(16), 5672–5680. https://doi.org/10.1128/JB.01919-07

Zwerling, A., Behr, M. A., Verma, A., Brewer, T. F., Menzies, D., & Pai, M. (2011). The BCG world atlas: A database of global BCG vaccination policies and practices. *PLoS Medicine*, *8*(3), e1001012. https://doi.org/10.1371/journal.pmed.1001012