Ana Branco Maranha Tiago

MOLECULAR AND BIOCHEMICAL STUDIES OF A MYCOBACTERIAL MALTOKINASE AND A UNIQUE OCTANOYLTRANSFERASE: TOWARDS RECONSTRUCTION OF THE PATHWAY FOR METHYLGLUCOSE LIPOPOLYSACCHARIDES BIOSYNTHESIS

Tese de Doutoramento em Biociências, ramo de especialização em Microbiologia, realizada sob orientação científica do Doutor Nuno Miguel da Silva Empadinhas e apresentada ao Departamento de Ciências da Vida da Faculdade de Ciências e Tecnologia da Universidade de Coimbra

Setembro 2015

· U C ·

Universidade de Coimbra

Ana Branco Maranha Tiago

Molecular and biochemical studies of a mycobacterial maltokinase and a unique octanoyltransferase: Towards reconstruction of the pathway for methylglucose lipopolysaccharides biosynthesis

Tese de Doutoramento em Biociências, ramo de especialização em Microbiologia, realizada sob orientação científica do Doutor Nuno Miguel da Silva Empadinhas e apresentada ao Departamento de Ciências da Vida da Faculdade de Ciências e Tecnologia da Universidade de Coimbra

2015

· U C ·

UNIVERSIDADE DE COIMBRA

Cover image:

Illustration of a Ziehl-Neelsen-stained photomicrograph sputum smear sample containing *Mycobacterium tuberculosis*.

Courtesy of Centers for Disease Control Public Health Image Library (CDC PHIL).

Agradecimentos

Gostaria de agradecer em primeiro lugar ao Doutor Nuno Empadinhas por me ter acolhido no grupo e pela confiança que depositou em mim. Por contrariar as minhas dúvidas com uma constante atitude positiva e, ao longo destes anos, sempre me ter encorajado, orientado, ensinado e incentivado a aprender. Especificamente, gostaria de agradecer as muitas horas dispendidas a rever este manuscrito.

A todos os meus companheiros no laboratório de micobacteriologia, os que me acompanham no dia-a-dia e aos que já me acompanharam. Ao Vítor Mendes, com quem iniciei o trabalho de bancada, que me transmitiu o entusiasmo pela investigação e que tanto me ensinou. Um obrigada muito especial à Susana Alarico pela sua infindável paciência e carinho. À Mafalda Costa minha "consultora de imagem", à Daniela Costa, à Rita Pereira, à Andreia Lamaroso e ao Diogo Reis, pela ajuda, boa disposição e brincadeiras.

I am profoundly grateful to Prof. Anthony Clarke for opening his lab to me, for sharing his expertise and making me feel welcome. To everyone at the Clarke Lab (U of G, Canada) my deepest thanks for taking such good care of me during my stay. A special thanks to Patrick Moynihan for all I learned from him.

À Doutora Sandra de Macedo-Ribeiro e ao Doutor Pedro Pereira e aos seus grupos de trabalho no IBMC no Porto, que sempre tão bem me receberam, que me incluiram no quotidiano do laboratório como se estivesse "em casa." Um especial obrigada à Joana Fraga que alinha sempre nas minhas "loucuras laboratoriais".

Não poderia deixar de expressar a minha enorme gratidão à Doutora Rita Ventura e a toda a sua equipa no ITQB, Oeiras, que foram sempre incansáveis tanto a esclarecer qualquer dúvida que eu como bióloga, não-química, pudesse ter, como a tentar satisfazer a minha insaciável necessidade de substratos.

À Fundação para a Ciência e a Tecnologia por me ter concedido o apoio financeiro através de uma Bolsa de Doutoramento (SFRH/BD/74845/2010) e pelo suporte financeiro para realização do trabalho experimental através dos projetos EU-FEDER Programa COMPETE FCOMP-01-0124-FEDER-014321 (PTDC/BIA-PRO/110523/2009), FCOMP-01-0124-FEDER-028359 (PTDC/BIA-MIC/2779/2012), FCOMP-01-0124-FEDER-037276 [PEst-C/SAU/LA0001/2013-2014] e UID/NEU/04539/2013.

I would also like to acknowledge the support of the Mizutani Foundation for Glycoscience, Japan, through grant 120123 to Dr. Nuno Empadinhas and also the Strategic Programme InovC-2014 and DITS (Divisão de Inovação e Transferências do Saber), University of Coimbra.

Uma especial nota de agradecimento à Prof. Doutora Paula Veríssimo do Departamento de Ciências da Vida da UC, pela sua constante disponibilidade e pelas profícuas discussões científicas sobre proteínas e enzimologia. Gostaria também de agradecer ao Doutor Tiago Faria, que tantas vezes me ajudou e ao Doutor Hugo Osório pelas análises por MALDI-TOF. Assinalo também a ajuda preciosa da equipa técnica do CNC.

Não poderia deixar de agradecer a todos os colegas dos diversos laboratórios por onde passei. Aos colegas do laboratório de microbiologia no patronato, obrigada por me terem ensinado que a ciência não é uma actividade solitária. Aos colegas da microbiologia do 3º piso da FMUC, obrigada por terem tornado a nossa transição tão fácil. Um obrigada especial à Lisa, minha companheira de "infortúnio".

Ao "pessoal do almoço" pelas conversas estimulantes, as conversas tontas, a amizade e por tantas vezes me guiarem neste labirinto que é a vida científica.

Aos meus amigos que têm uma paciência infinita para me aturarem, que me perguntam pelo meu trabalho, mesmo quando não percebem metade do que lhes estou a explicar, que me incitam sempre a continuar e que me ajudam a manter sã.

À minha família, especialmente à minha mãe e aos meus irmãos, o meu grande sistema de apoio, que seguram as pontas do meu mundo.

A todos os que de alguma forma contribuiram para a elaboração deste trabalho o meu muito obrigada.

Ana

Coimbra, Agosto 2015

Table of contents

Abstract

Mycobacteria belong to an extremely diversified bacterial genus that contains both saprophytic organisms and dangerous pathogens such as *Mycobacterium leprae* and *Mycobacterium tuberculosis*, the etiological agents of leprosy and tuberculosis, respectively. Tuberculosis (TB) is one of the deadliest diseases caused by a single infectious agent and its eradication is considered a major global health priority. Nonetheless the lack of proper access to medical care in developing countries, the burden of chronic illnesses in developed countries and the emergence of multidrug-resistant strains has hindered the process towards that goal.

Mycobacteria possess a distinctive lipid-rich cell envelope to which they owe much of their pathogenicity. Among the unique molecules synthesized by mycobacteria there are two intracellular polymethylated polysaccharides (PMPSs) that likely assume a helical conformation in the cytoplasm and are postulated to interact with fatty acid synthase I (FAS-I) to modulate fatty acid synthesis. PMPSs can be polysaccharides of 3-*O*-methylmannose (MMPs) or lipopolysaccharides of 6-*O*-methylglucose (MGLPs). These molecules were discovered in the 1960's but only recently with the unveiling of the mycobacterial genome was it possible to uncover more aspects of their biogenesis. MGLPs in particular have been the target of renewed attention because they probably represent the only type of PMPSs present in the slowly-growing phenotype frequently associated with pathogenic mycobacteria such as *M. tuberculosis,* while the distribution of MMP appears to be almost exclusively restricted to rapidly-growing mycobacteria. In addition, some genes involved in the MGLP biosynthetic pathway have been considered essential for *M. tuberculosis* growth, making their products potential drug targets.

Recently a novel metabolic pathway (GlgE pathway) has been uncovered, connecting trehalose and glycogen through the intermediate metabolite maltose-1-phosphate. This pathway may also represent an alternative route for polymerization of the seemingly vital mycobacterial MGLP that has a glycogen-like arrangement of glucose units in the main chain.

We pursued the identification and characterization of enzymes with probable involvement in the MGLP pathway and simultaneously proposed to be essential for *M. tuberculosis* growth. As a consequence of functional and genomic context surveys two genes were selected for study. Gene *Rv0127* encoding a maltokinase for the synthesis of maltose-1-phosphate and *Rv2418c* of unknown function, but located in an operon containing the atypical glucosyl-3-phosphoglycerate phosphatase responsible for the production of the MGLP primer, glucosylglycerate.

The work presented in this thesis contributes for the elucidation of the maltokinase structure and the functional characterization of a novel sugar octanoyltransferase adding new insights into an intricate and vital pathway of mycobacterial glycobiology.

Resumo

As micobactérias pertencem a um género bacteriano extremamente diverso que contém bactérias saprófitas e patogénicas. Das patogénicas destacam-se *Mycobacterium leprae* e *Mycobacterium tuberculosis,* os agentes etiológicos da lepra e da tuberculose, respectivamente. A tuberculose encontra-se entre as doenças com maior taxa de mortalidade causadas por um único agente infeccioso e a sua erradicação é considerada uma prioridade. No entanto, a falta de acesso a cuidados de saúde primários em países em desenvolvimento, o aumento da incidência de doenças crónicas em países desenvolvidos e o aparecimento de estirpes multirresistentes têm causado sérios obstáculos à sua erradicação.

As micobactérias possuem uma parede celular distinta que lhes confere muitas das suas características patogénicas e que é bastante rica em componentes lipídicos. Entre as moléculas singulares sintetizadas por micobactérias encontram-se dois polissacáridos polimetilados (PMPSs) intracelulares cuja hipotética interacção com a sintetase de ácidos gordos I (FAS-I) modula a síntese de ácidos gordos. Os PMPSs dividem-se em polissacáridos de 3-*O*-metilmanose (MMPs) ou lipopolissacáridos de 6-*O*-metilglucose (MGLPs). Embora a sua descoberta remonte à década de 1960, apenas recentemente, com o descodificar do genoma micobacteriano, foi possível compreender melhor a sua biogénese. Os MGLPs foram alvo de particular atenção por terem sido os únicos PMPS detectados em micobactérias de crescimento lento, geralmente associadas a maior patogenicidade, como *M. tuberculosis*. Os MMPs parecem ser exclusivos de micobactérias de crescimento rápido. Alguns do genes envolvidos na via biossintética do MGLP estão descritos como sendo essenciais para o crescimento de *M. tuberculosis*, tornando-os potenciais alvos terapêuticos.

Recentemente foi descoberta em micobactérias uma nova via biossintética (via GlgE) que converte trealose em glicogénio através do intermediário maltose-1-fosfato. Esta via representa uma possível alternativa para a polimerização de MGLP, cuja estrutura base é semelhante à cadeia principal de glicogénio no que respeita à ligação entre glucoses.

Como uma das prioridades no combate à tuberculose é a identificação de novos alvos terapêuticos, considerámos importante identificar a função e caracterizar enzimas com provável envolvimento na via do MGLP e que estivessem simultaneamente descritas como essenciais para o crescimento de *M. tuberculosis*. Por reunirem estas condições, o gene *Rv0127,* responsável pela síntese de maltose-1-fosfato, e o gene *Rv2418c,* localizado num operão que contém a enzima glucosil-3-fosfoglicerato fosfatase, responsável pela produção do substrato inicial para o MGLP, foram os alvos selecionados para este trabalho.

O trabalho aqui apresentado contribui para elucidar a estrutura da maltocinase e para a caracterização funcional de uma nova e atípica octanoiltransferase acrescentado novas prespectivas a uma via intrincada e vital da glicobiologia micobacteriana.

Abbreviation list

AG	arabinogalactan
AppCp	non-hydrolysable ATP analog, adenosine-5'- [(β,γ)-methyleno] triphosphate
Ara6	hexa-arabinoside
CrOT	carnitine *O*-octanoyltransferase
DAT	di-acyltrehaloses
DGG	diglucosylglycerate
DGG-Oct	octanoylated diglucosylgycerate (see 6-Oct-DGG **1** and DGG-3-Oct **2** bellow)
DTNB	5,5'-dithiobis-2-nitrobenzoic acid, Ellman's Reagent
ELK	eukaryotic-like protein kinase
ePKs	eukaryotic protein kinases
FA	fatty acid
FAS	fatty acid synthase system; FAS-I or FAS-II
GA	glyceric acid
GG	glucosylglycerate
GOAT	mammalian ghrelin *O*-acyltransferase
GPG	glucosyl-3-phosphoglycerate
Hex-CoA	hexanoyl-CoA
LOS	lipooligosaccharides
M1P	maltose-1-phosphate
MA	mycolic acids
mAGP	mycolyl-arabinogalactan-peptidoglycan complex or PG-AG-MA complex
Mak^{Mtb}	maltokinase from *M. tuberculosis*
Mak^{Mvan}	maltokinase from *M. vanbaalenii*
ManT	α-(1→4)-mannosyltransferase
MDR-TB	multidrug-resistant tuberculosis
MG	mannosylglycerate
MGG	mannosylglucosylglycerate
MGLPs	6-*O*-methyl-α-D-glucose lypopolysaccharides

MMPs	3-*O*-methyl-D-mannose polysaccharides
MS	Mass Spectrometry
MTBC	*Mycobacterium tuberculosis* complex
MTRK	bacterial 5-methylthioribose kinase
NTM	nontuberculous mycobacteria
Oct-CoA	octanoyl-CoA
Oct-DGG	octanoylated diglucosylgycerate (see 6-Oct-DGG **1** and DGG-3-Oct **2** bellow)
Oct-*p*NP	octanoyl-*p*-nitrophenol
OctT	octanoyltransferase
OMetT	3-*O*-methyltransferase
PAT	poly-acyltrehaloses
PatB	peptidoglycan *O*-acetyltransferase B
PG	peptidoglycan
PKL	protein kinase–like
PMPSs	polymethylated polysaccharides
RGM	rapidly-growing mycobacteria
SGM	slowly-growing mycobacteria
T6P	trehalose-6-phosphate
TB	tuberculosis
TDM	trehalose dimycolate
TLC	thin-layer chromatography
WT	wild-type
XDR-TB	extensively drug resistant tuberculosis
3-PGA	D-3-phosphoglyceric acid
6-Oct-DGG **1** and DGG-3-Oct **2**	octanoylated diglucosylgycerate in positions C6 OH of the second glucose and C3 of glyceric acid, respectively

List of genes and enzymes

AcT	putative acyltransferase involved in MGLP biosynthesis
DggS	putative diglucosylglycerate synthase involved in MGLP biosynthesis
GgH	glucosylglycerate hydrolase (almost exclusively detected in RGM and absent from *M. tuberculosis*)
GlgA (*Rv1212c*)	α-(1→4)-glycosyltransferase involved in glycogen and capsular glucan biosynthesis
GlgB (*Rv1326c*)	α-(1→4)-glucan branching enzyme involved in glycogen and capsular glucan metabolism as a branching enzyme as well as in the GlgE pathway
GlgC (*Rv1213*)	ADP-glucose synthase involved in glycogen and capsular glucan biosynthesis
GlgE (*Rv1327c*)	maltose-1-phosphate maltosyltransferase using the maltosyl unit in maltose-1-phosphate to elongate linar α-(1→4)-glucans
GpgP (*Rv2419c*)	glucosyl-3-phosphoglycerate phosphatase catalyzing the dephosphorylation of glucosyl-3-phosphoglycerate to glucosylglycerate, considered the primer for MGLP biosynthesis
GpgS (*Rv1208*)	glucosyl-3-phosphoglycerate synthase catalyzing the condensation of NDP-glucose and 3-phosphoglycerate into glucosyl-3-phosphoglycerate, the first commited step in MGLP biosynthesis
GT	glycosyltransferase α-(1→4) (see Rv3032, GlgA) or glycosyltransferase β-(1→3)
Mak/Pep2 (*Rv0127*)	maltokinase phosphorylating maltose into maltose-1-phosphate, involved in the GlgE pathway
MeT (*Rv3030*/*Rv3037c*)	methyltransferases probably responsible for methylation of MGLP
OctT (*Rv2418C*)	octanoyltransferase identified in this study
OtsA (*Rv3490*)	trehalose-6-phosphate synthase (Tps) transfering glucose from UDP-glucose to glucose-6-phosphate originating trehalose-6-phosphate, involved in trehalose biosynthesis
OtsB (*Rv2006*)	trehalose-6-phosphate phosphatase (Tpp) for dephosphorylation of trehalose-6-phosphate to free trehalose

OtsB2 (*Rv3372*)	in mycobacteria, trehalose-6-phosphate phosphatase activity is encoded by this gene
Rv3032	α-(1→4) glycosyltransferase involved in MGLP and glycogen biosynthesis
TreS (*Rv0126*)	trehalose synthase isomerizing maltose into trehalose; involved in trehalose metabolism and in the GlgE pathway
TreX (*Rv1564c*)	putative glycogen debranching enzyme possibly involved in glycogen and trehalose metabolism
TreY (*Rv1563c*)	maltooligosyltrehalose synthase isomerizing the terminal disaccharide of an α-(1→4)-glucose polymer into trehalose, involved in trehalose biosynthesis
TreZ (*Rv1562c*)	maltooligosyltrehalose trehalohydrolase hydrolyzing the product of TreY isomerization to free trehalose

Gene nomenclature of *M. tuberculosis* H37Rv

List of figures

List of tables

Chapter 1 General Introduction

Chapter 1- table of contents

1. The genus *Mycobacterium*

Mycobacteria are a diverse group of bacteria classified in a genus comprising over 170 validated species, according to the List of Prokaryotic names with Standing in Nomenclature (http://www.bacterio.net/), most of which with a saprophytic lifestyle. The genus *Mycobacterium* belongs to the suborder Corynebacterineae of the phylum/class Actinobacteria and groups alongside several closely related genera of industrial, clinical or environmental importance such as *Corynebacterium, Nocardia, Gordonia* and *Rhodococcus* (Hartmans *et al.*, 2006).

Traditionally, mycobacteria can be separated into two major groups based on their growth rate, reflected by the celerity with which they produce visible colonies on solid media (Sidders and Stoker, 2001). Rapidly-growing mycobacteria (RGM) form visible colonies within seven days while colonies of slowly-growing mycobacteria (SGM) appear later. Pathogenicity has been often associated with SGM such as *Mycobacterium tuberculosis,* whereas rapid growth tends to be considered an archetypal feature of environmental species that usually have a saprophytic lifestyle. For example, *Mycobacterium smegmatis* is a saprophytic RGM that has been widely used as a model to study mycobacterial biology, it forms colonies on agar plates within 3-4 days and lacks the virulent properties described for *M. tuberculosis* (Reyrat and Kahn, 2001; Bashiri and Baker, 2015). Nevertheless, the growing numbers of valid species in this genus unveiled many exceptions to this empirical association of pathogenicity and growth rate, for example the rapidly-growing pathogen *Mycobacterium abscessus* that causes serious pulmonary infections against which very few antibiotics are effective (Medjahed *et al.*, 2010; Nessar *et al.*, 2012).

The most infamous members of this genus are *M. tuberculosis* and *M. leprae*, the etiological agents of tuberculosis (TB) and leprosy, respectively, two of the most ancient diseases known to mankind (Glickman and Jacobs, 2001). TB has claimed more lives than any other microbial pathogen (Roberts and Buikstra, 2003) and remains today one of the leading causes of morbidity and mortality throughout the world (WHO, 2014). Moreover, in recent years, the once considered innocuous saprophytic mycobacteria, designated nontuberculous mycobacteria (NTM), started to get rampant attention because of their newly acknowledged relevance as agents of both nosocomial and atypical infections, having higher expression in patients with co-morbidities or in immunocompromised patients, but nevertheless also affecting immunocompetent individuals (Tortoli, 2009; Tortoli, 2014).

There are a few species closely related to *M. tuberculosis* that despite differing widely in terms of their host tropism and phenotype are all capable of causing animal or human tuberculosis, namely *M. africanum* (responsible for human TB outbreaks in Africa)*, M. bovis*, *M. canettii*, *M. microti, M. caprae, M. pinnipedii,* as well as the recently added members *M. mungi*, *M. orygis, and M. suricattae* (Tortoli, 2014). Their pathogenic characteristics as well as a genomic similarity allowed the inclusion of these species into a group designated the *M. tuberculosis* complex (MTBC) (Brosch *et al.*, 2002; Tortoli, 2014). Mycobacteria capable of causing TB are primarily transmitted via aerosols disseminated by patients with active pulmonary TB (Glaziou *et al.*,

2015). Recently, *M. tuberculosis* and other MTBC members have also been found to maintain its virulence for extended periods of time in soil samples (Ghodbane *et al.*, 2014a).

The outcome of TB infection is highly variable as exposure can be followed by rapid clearance through innate immunity or establishment of active or latent infection, which can be re-activated at a later time in life (O'Garra *et al.*, 2013; Millet *et al.*, 2013). The unpredictability in disease outcome is not only associated with variability in the host responses, associated to the host's genetic susceptibility to infection, but also to known risk factors that include underlying co-morbidities or immune susceptibilities associated to HIV, diabetes, or malnutrition (O'Garra *et al.*, 2013; Bates *et al.*, 2015). The establishment and progression of mycobacterial infection depends primarily on the pathogen's ability to subvert the host's physiological and immune responses and successfully survive within an infected cell (Dey and Bishai, 2014). As such, tuberculous mycobacteria evolved to prevent the maturation of phagosomes into phagolysosomes, allowing their replication within professional macrophages, and creating a niche for survival (O'Garra *et al.*, 2013). Active TB disease comprises a range of presentations, including the classical pulmonary TB, and various forms of extra pulmonary disease that can affect different body sites and can progress to a multisystemic disseminated form of TB designated miliary TB (O'Garra *et al.*, 2013).

1. 1 Tuberculosis: a brief historical prespective

There are two main theories for the origin of human tuberculosis (Brites and Gagneux, 2015) both defending that the pathogen evolved from an ancestral environmental bacterium very similar to *M. canetti* (Supply *et al.*, 2013). One theory argues for an animal origin of TB and subsequent transfer to humans (Minnikin *et al.*, 2015) whereas the other defends *M. tuberculosis* co-evolution with humans and its emergence as a human pathogen in East Africa around 70.000 years ago (Comas *et al.*, 2013; Brites and Gagneux, 2015). The disease may then have spread out of the continent with the first 'Out-of-Africa' migrations and diversified following the migrations of human populations to form the geographically restricted linages plaguing humanity until today (Comas *et al.*, 2013).

Concrete proof of TB can be traced back to the Neolithic (Formicola *et al.*, 1987; Hershkovitz *et al.*, 2008) and evidence is widespread in ancient Egypt through the analysis of the art and mummies presenting morphological skeletal changes consistent with TB (Zink *et al.*, 2007). Although precise data before the 19th century is scarce it is known that mortality rates were extremely high and that the disease was an enigma to physicians. In the 1850's TB was one of leading causes of death, particularly in urban areas, with a mortality rate ranging from 0.4% in England (McKeown, 1976) to 1% at the peak of the epidemic in some European metropolis (Lönnroth *et al.*, 2009). In 1882, Robert Koch gave his historic lecture "Die Aetiologie der Tuberculose" enlightening the cause of the disease after identifying the tubercle bacillus and demonstrating unequivocally TB infectious etiology (Koch, 1982). In the late 19th century, in Western Europe and North America, there was a noticeable decline in TB death rates which can be partly attributed to better living conditions, such as improved nutrition and introduction of

sanitary measures, including mechanisms to isolate infected patients in sanatoria, possibly in addition to epidemiological phenomena unrelated to human intervention, for instance acquired immunity (Lönnroth *et al.*, 2009). In 1921, the only vaccine that is currently in use to prevent TB infection was implemented, the BCG vaccine (Zwerling *et al.*, 2011). Developed from an attenuated *M. bovis* strain the vaccine is nevertheless not totally effective in protecting inoculated individuals, especially adults and certain populations exposed to environmental mycobacterial antigens since early age (Zwerling *et al.*, 2011). Introduction of isoniazid (1952), pyrazinamide (1954), ethambutol (1961) and rifampicin (1963) as anti-TB agents (Zumla *et al.*, 2013) dramatically improved TB control and after implementation of a combination therapy with a minimum duration of six months (Glaziou *et al.*, 2015) eradicating TB appeared a possible scenario (Chakroborty, 2011). The aminoglycoside antibiotics as well as most available 2nd line drugs were also introduced in the 1950's and 1960's. Shortly afterwards the plight of TB was in steady decline and around 1980 it was considered a disease of the past (Zumla *et al.*, 2013). However, emergence of the HIV-AIDS epidemics, the lack of compliance with the long chemotherapeutic courses, inadequate access to proper medical care in developing countries, and in most recent years, the aging of the population with the escalation of chronic diseases such as diabetes and cardiovascular disease, made it possible for TB to reappear, especially in the considerably more dangerous forms of multidrug resistant TB (MDR-TB) and extensively drug resistant TB (XDR-TB) (Comas and Gagneux, 2009).

In 2013 there were 9 million new TB case notifications worldwide, 5% of which were estimated to be MDR-TB including about 9% XDR-TB cases among the MDR-TB forms (WHO, 2014). The incidence of atypical infections by nontuberculous mycobacteria (NTM), which were uncommonly observed prior to the onset of the AIDS epidemic, has also increased with a frightening surge in recent years possibly as a consequence of immune fragilities in modern societies and an ageing population with chronic illnesses (van der Werf *et al.*, 2014). NTM are also a concerning source of nosocomial infections to hospitalized patients which relates to their environmental ubiquity and poor disinfection strategies of water distribution systems, which are powerful dissemination sources of the generally disinfectant-resistant NTM (Falkinham III, 2015).

Regardless of pressing need, there has only been one introduction of a new class of antibiotics (bedaquiline) for TB since the 1960's but, due to its toxicity, it is only used in combination therapy for treatment of adults with MDR-TB when standard therapies become ineffective (Zumla *et al.*, 2013). Most recently, attention has turned to uncultured bacteria as a potential source of antibacterial compounds and some authors developed several innovative methods to grow uncultured organisms, which enabled them to isolate promising new antibacterial compounds (Nichols *et al.*, 2010). Usually *M. tuberculosis* slow growth makes clinical diagnosis and antibiotic susceptibility testing difficult. However, the development of an improved medium in combination with optimized culture conditions allowed a dramatic reduction in culture time from over 7 days to 3 days. The subsequent development of rapid antibiotic susceptibility testing could detect resistance within a reasonable timeframe (Ghodbane *et al.*, 2014b). Despite these distant promises, the currently available therapies target only a few cell functions and the

growing number of drug-resistant strains urges for the development of shorter and more effective TB treatments and the identification of new drug targets, in addition to an improved understanding of mycobacterial physiology (Zumla *et al.*, 2013).

2. The mycobacterial cell envelope

Mycobacteria possess a unique cell envelope which is a thick lipid-rich structure with complex carbohydrates, globally providing the bacteria with the ability to survive in harsh environments, also contributing to these organisms' pathogenicity and survival within macrophage phagolysosomes as well as impermeability to diverse antimicrobial compounds (Daffé *et al.*, 2014). The most significant studies on mycobacterial cell envelope architecture were conducted in the 1960's, 1970's and early 1980's when David E. Minnikin proposed the currently accepted model for this structure (Minnikin, 1982). However it was only recently that developments in analytical techniques and ultrastructural studies, combined with the deciphering of the *M. tuberculosis* genome, allowed a better understanding of this remarkable structure and many details of its biosynthesis (Brennan, 2003).

The cell envelope is a multilayered structure composed by a plasma membrane surrounded by an intermediate heterogeneous cell wall, and a loosely associated outer capsule (Figure 1.1) (Hoffmann *et al.*, 2008; Sani *et al.*, 2010). The innermost leaflet of the cell wall consists of peptidoglycan (PG) convalently linked to arabinogalactan (AG) and esterified to long-chain (C60–C90) fatty acids called mycolic acids (MA), forming a complex known as the PG-AG-MA (or mAGP, mycolyl-arabinogalactan-peptidoglycan complex) that extends from the plasma membrane outward in layers (Jankute *et al.*, 2014). This is the essential core of the cell wall responsible for survival under harsh conditions being an insoluble complex targeted by most drugs used against TB (Hett and Rubin, 2008; Daffé *et al.*, 2014). The outermost leaflet of the cell wall is the mycomembrane, an important permeability barrier formed by a densely packed lipid bilayer that consists mainly of mycolic acids. These mycolic acids are present in two main forms: partly linked to the cell wall AG and partly as free esters of trehalose in the form of trehalose dimycolate (TDM), the latter are intercalated with an assortment of free glycolipids, lipoglycans and proteins (Zuber *et al.*, 2008; Daffé *et al.*, 2014). Finally, the capsule is an ill-defined layer found in both pathogenic and non-pathogenic mycobacteria and dominated by lipoglycans and proteins that contribute to modulation of host immune responses (Sani *et al.*, 2010) (Figure 1.1).

2. 1 Cell wall core

The cell wall core includes the peptidoglycan (PG) covalently linked to the reducing end of arabinogalactan (AG) through a disaccharide linker essential for its structural integrity, and finally mycolic acids linked to the majority of the non-reducing end of the AG. This structure is the most invariable arrangement across the *Mycobacterium* genus (Jackson *et al.*, 2013).

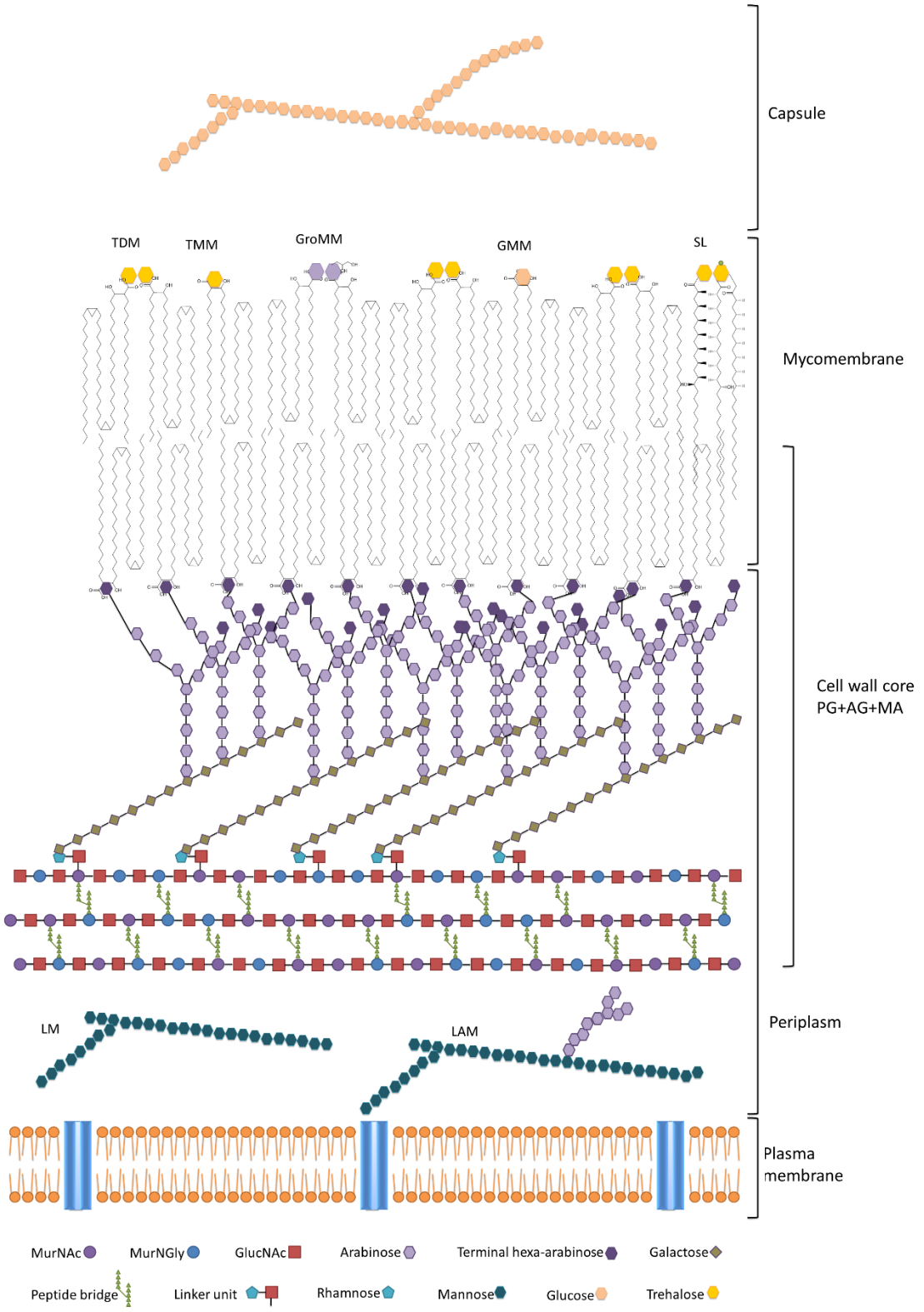


Figure 1.1 - Schematic representation of the mycobacterial cell envelope highlighting its major structural features consisting of a inner plasma membrane, an intermediate cell wall core, the mycolic acids-containing mycomembrane and, according to recent studies, a capsule and a periplasmic region (Hoffmann *et al.*, 2008). The overall schematic and individual structures are not drawn to scale. Dimensions suggested by electron microscopy propose a plasma membrane with ~7 nm, a periplasm with ~16-19 nm, PG+AG with ~7.5 nm, the mycomembrane with ~ 8 nm and a capsule up to 40 nm thick (Hoffmann *et al.*, 2008; Zuber *et al.*, 2008; Sani *et al.*, 2010). Both the MAs connected to AG and the extractable esters are presented in folded conformation in order to fit into the outer membrane of the cell wall (Zuber *et al.*, 2008) and are shown in the folded "w" compact conformation as suggested by Groenewald *et al.* (2014). For simplicity only α-MA with double cyclopropanation are represented. MurNAc, N-acetylmuramic acid; MurNGly, N-glycolylmuramic acid; GlcNAc, N-acetylglucosamine. Adapted from (Jackson *et al.*, 2013; Daffé *et al.*, 2014; Kieser and Rubin, 2014; Nobre *et al.*, 2014).

The PG layer forms the backbone of the cell envelope conferring rigidity and integrity to the cell and allowing it to withstand changes in osmotic pressure (Mahapatra *et al.*, 2008; Pavelka *et al.*, 2014). Mycobacterial PG is composed of a glycan chain of alternating β-(1→4)-linked N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) residues, some of which modified as follows (Kieser and Rubin, 2014; Pavelka *et al.*, 2014). The muramic acid residues contain a mixture of N-acetyl and N-glycolyl derivatives forming MurNGly. This modification is present in mycobacteria and closely related organisms with the exception of *M. leprae* (Mahapatra *et al.*, 2008) and is a hallmark of the genera (Moynihan *et al.*, 2014). These additional glycolyl-containing residues result in extra hydrogen bonding, which should strengthen the mesh-like structure and protect the organism from lysozyme degradation (Hett and Rubin, 2008) also contributing to enhanced immunogenicity (Hansen *et al.*, 2014). Mycobacterial PG has 12% of its muramic acid residues covalently attached to the disaccharide linker which is in turn connected to the reducing end of AG (Jankute *et al.*, 2014) (Figure 1.1).

Arabinogalactan is composed of D-arabinofuranosyl (Ara*f*) and D-galactofuranosyl (Gal*f*) residues consisting of a linear galactan chain composed of 30 linear alternating β-(1→5) and β-(1→6) linked Gal*f* residues anchored into the the linker unit. Three D-arabinan chains comprised of 22 or 23 Ara*f* residues are affixed to the C-5 of the 8th, 10th and 12th Gal*f* residue of this linear chain (Figure 1.1). The arabinan domain is present as a highly branched network that has a distinct motif of hexa-arabinoside (Ara6) at its non-reducing end, the place of mycolic acid attachment, with roughly two-thirds of the Ara6 motifs mycolated (Jankute *et al.*, 2014). The AG layer is quite flexible and it presents in a very densely packed configuration (Jackson *et al.*, 2013).

2. 1. 1 Mycomembrane structure

The presence of high molecular weight α-alkyl, β-hydroxy fatty acids, called mycolic acids (MAs) is a hallmark of mycobacteria and some members of the suborder Corynebacterineae (Takayama *et al.*, 2005). These are long-chain fatty acids (FAs) that have two-branches, the short saturated alkyl chain is called the α-branch, because it is in the α (or 2) position in relation to the carboxylic acid group, and the longer chain is named meroaldehyde which is in the β (or 3) position relative to the carboxylic group (Pawełczyk and Kremer, 2014; Marrakchi *et al.*, 2014) (Figure 1.2).

Mycolic acids differ in length across genera and a recent discovery “dethroned” *Mycobacterium* as the producer of the longest mycolate chains, a title now earned by *Segniliparus rotundus*, a rapidly-growing bacterium isolated from the human respiratory tract that produces mycolates (termed segnilomycolates) ranging from 58-100 carbons, in contrast to mycobacterial MAs that contain 60-90 carbons (Hong *et al.*, 2012). MAs are very diverse as well as genus and species-specific molecules (Marrakchi *et al.*, 2014) that vary not only in chain lengths but also in chemical functions, which depend on enzymatically introduced chemical modifications in the meromycolate chain (Barry III *et al.*, 1998). Chemical modifications occur in only two positions of the meromycolate chain: proximal and distal, with respect to the β-hydroxy group (Figure 1.2). They can include polar modifications, such as the addiction of keto, epoxy, and methoxy functional groups (usually restricted to the distal position) or nonpolar modifications such as

double bonds or cyclopropane rings (Pawełczyk and Kremer, 2014). Each mycobacterial species is characterized by a specific mycolic acid profile (Pawełczyk and Kremer, 2014; Barry III *et al.*, 1998) which generally can be related to growth rate and pathogenicity (Laval *et al.*, 2001). The functional groups present in the meromycolate chain affect the packing of the long hydrocarbon chains, thereby influencing the intrinsic physiological functions of the cell envelope (Watanabe *et al.*, 2001; Barry III *et al.*, 1998).

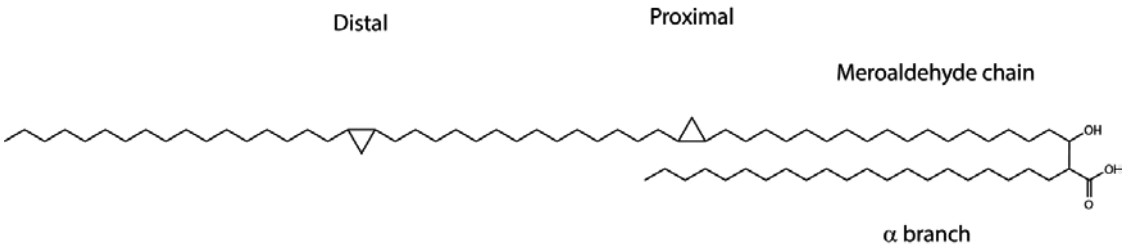


Figure 1.2 - General scheme of an α cis/cis cyclopropanated mycolic acid typical of mycobacteria.

MAs constitute about 60% of the mycobacterial cell wall dry weight and can be found mainly in three different forms fulfiling different functions, (i) bound to the arabinogalactan of the cell wall in parallel arrangement (the MAs that compose the cell wall core are the majority of MAs found in mycobacteria), (ii) in cell wall extractable esters such as TDM, (iii) or secreted as free MAs (Pawełczyk and Kremer, 2014; Marrakchi *et al.*, 2014). The most important characteristics conferred by MAs include low permeability to chemicals (including antibiotics) and resistance to chemical injury and dehydration, which allow mycobacteria to survive within macrophage phagolysosomes and other inhospitable environments such as dust particles or soil (Barry III *et al.*, 1998; Nobre *et al.*, 2014).

(i) Cell wall bound MAs

Several studies suggest that the mycobacterial cell wall is asymmetric, with the outer leaflet being moderately fluid in addition to the inner leaflet having very low fluidity (Hong and Hopfinger, 2004; Liu *et al.*, 1995). This may be the result of the unique chemical structures of MAs with their mostly saturated hydrocarbon chains, since lipids containing longer hydrocarbon chains and fewer unsaturations tend to become packed tightly against each other (Liu *et al.*, 1995). Also the exceptionally flexible structure of the terminal hexa-arabinose in the AG cell wall core likely facilitates the lateral packing of mycolic acid chains (Hong and Hopfinger, 2004). However, recent cryo-electron microscopy studies suggested that the mycobacterial outer membrane is morphologically symmetrical and that free lipids, with heterogeneous head groups, are distributed over both mycomembrane leaflets and not restricted to the outer leaflet alone, in contrast to the previous presented models (Hoffmann *et al.*, 2008) where the inner leaflet very low fluidity was thought to account for the overall low permeability of the cell wall (Barry III *et al.*, 1998).

An important property of biological membranes is to maintain a stable cell function, to that end controlling fluidity and preserving the function of integral membrane proteins is pivotal. All

organisms can modify membrane fluidity in response to environmental temperature changes which is generally achieved by inducing variation in lipid structure (Mansilla *et al.*, 2004). In the case of *E. coli* the mechanisms of adaptation are known to occur mostly in membrane phospholipids (Mansilla *et al.*, 2004). In the particular case of mycobacteria the mechanisms by which they adapt to changes in temperature are less known however they appear to adjust by varying envelop lipid composition, especially at MA level, by altering: (a) the length of the meromycolate chain (increasing chain length in response to rising temperatures) (Toriyama *et al.*, 1980; Baba *et al.*, 1989; Liu *et al.*, 1996), (b) the ratio of cis/trans geometry at the proximal position and (Liu *et al.*, 1996) (c) the ratio the of the different classes of MAs (Baba *et al.*, 1989; Kremer *et al.*, 2002). These modifications are important to maintain the viability and low permeability of the cell envelope.

(ii) Cell wall extractable esters

Mycolic acids in the extractable lipid fraction exist predominantly linked to trehalose in the form of trehalose dimycolate (TDM) or more rarely linked to glycerol (Rombouts *et al.*, 2012; Layre *et al.*, 2009) or glucose (Matsunaga *et al.*, 2008) (Figure 1.1). TDM exists manly in the outer layer of the mycomembrane, which contains various other free lipids intercalated with the MA esters. These comprise, for example, phenolic glycolipids (which are dimycocerosate esters of phenolphthiocerol) and a myriad of other trehalose-esters, including di- and poly-acyltrehaloses (DATs, and PATs), sulfolipids (SLs), and lipooligosaccharides (LOSs) (Nobre *et al.*, 2014) all of these implicated in pathogen-host crosstalk or virulence. Some of these trehalose esters have a species-specific distribution (for example DAT and PAT are unique to some species of the *M. tuberculosis* complex) (Nobre *et al.*, 2014).

TDM, also called cord factor, was associated with the rough morphology of some mycobacterial colonies although it was later demonstrated that 'non-cording' mycobacteria also produce TDM, confirming that this structure could not be solely responsible for cord formation (Hunter *et al.*, 2006). TDM represents the most abundant antigenic surface-exposed extractable lipid of *M. tuberculosis* and actively participates in blocking phagosome maturation during infection, also contributing to the pathogenesis of tuberculosis (Axelrod *et al.*, 2008). Trehalose monomycolate (TMM) is also a major component of the envelope and the TDM precursor (Pawełczyk and Kremer, 2014). Mycobacteria possess other minor mycolate esters, namely dimycolyl diarabinoglycerol (DMAG), which was postulated to share proinflammatory cytokine production and granuloma formation properties with TDM. DMAG is present in various mycobacterial species and particularly in large quantities in slowly growing species such as *M. tuberculosis* (Rombouts *et al.*, 2012). Other mycolyl esters such as glycerol monomycolate (GroMM) (Layre *et al.*, 2009) and glucose mono mycolate (GMM) are also mycoloyl-based lipid antigens (Matsunaga *et al.*, 2008).

In some mycobacteria there are also a small number of proteins such as porins intercalated with the MA esters (Niederweis *et al.*, 2010). Interestingly, when compared to other bacteria, namely *E. coli*, *M. tuberculosis* presents an extremely low abundance of porins (Niederweis *et al.*, 2010) which, combined with a high number of efflux pumps (da Silva *et al.*, 2011) and the inherent

hydrophobicity of the mycobacterial cell wall, contribute to the typically slow nutrient uptake and the intrinsic resistance of the organism to many drugs (Niederweis *et al.*, 2010).

(iii) Secreted as free mycolic acids

Mycobacteria form naturally-occurring biofilms during growth to protect the cells against exogenous threats and secreted free mycolates are pivotal to that end (Richards and Ojha, 2014; Ojha *et al.*, 2010). Many NTM, including opportunistic pathogens, have been found in polymicrobial biofilms obtained from water distribution systems which were implicated in resistance to antibiotics (Richards and Ojha, 2014; Ojha *et al.*, 2010). Biofilms may also play a role in the transmission and pathogenesis of *M. ulcerans* and have been associated with the pathogenic mechanism of *M. avium* (Richards and Ojha, 2014).

In *M. smegmatis* and *M. tuberculosis* secreted free mycolic acids are produced through enzymatic release from newly synthesized TDM, which is cleaved by a TDM-specific serine esterase. A mutation in this *M. smegmatis* esterase showed only a partial reduction in free mycolate synthesis, as well as retarded biofilm formation, suggesting that there are other pathways for free MAs production (Ojha *et al.*, 2010).

2. 1. 1. 1 Biosynthesis of fatty acids and processing to mycolic acids

The biosynthesis of fatty acids (FAs) involves different fatty acid synthase systems (FASs): FAS-I a single multifunctional enzyme typical of eukaryotes (except plants) and FAS-II, usually found in prokaryotes and plants, which is composed of a series of discrete enzymes (Pawełczyk and Kremer, 2014; Marrakchi *et al.*, 2014). Mycobacteria are atypical with respect to FAS systems in that they possess both FAS-I and FAS-II systems (Bloch, 1975). In these organisms FAS-I produces FAs with a bimodal distribution, the two dominant products being saturated C16 FAs (that can be supplied to the second complex) and saturated C24 (that will correspond to the α-branch found in MAs). FAS-II is responsible for the elongation of FAs to yield the very long meromycolic chains, precursors of MAs (Barry III *et al.*, 1998). Although the structural organization of FAS-I and FAS-II is markedly different, the chemical reactions and the catalytic mechanisms for FA biosynthesis that occur in the two systems are very similar (Marrakchi *et al.*, 2014).

Mycobacteria have a single FAS-I-encoding gene (*fas,* Rv2524c), shown to be essential in *M. smegmatis* (Zimhony *et al.*, 2004) and proposed to be essential for *M. tuberculosis* growth (Griffin *et al.*, 2011b). FAS-I carries all the functional domains required for *de novo* FA synthesis and all intermediates that are generated during elongation remain enzyme-bound, undergoing transacylation to other catalytic sites within the enzyme (Pawełczyk and Kremer, 2014). FAS-I catalyzes *de novo* FA synthesis by utilizing acetyl-CoA as primer and malonyl-CoA to elongate the acetyl group by two carbons in repetitive series of four reactions in each cycle (Pawełczyk and Kremer, 2014). When FAS-I elongation yelds C16/ C18 enzyme-bound thioesters the cells' metabolic needs determine their fate. The thioesters can be further elongated to produce C26-enzyme bound thioesters, which after conversion to CoA derivatives will participate in MA α-branch synthesis (the dominant products being C24-CoA in *M. smegmatis* and C26-CoA in *M.*

tuberculosis). On the contrary, elongation can cease and after convertion to C16/C18 acyl-CoAs these can be used for the synthesis of membrane phospholipids (Zimhony *et al.*, 2004) or released into the FAS-II system (Pawełczyk and Kremer, 2014) (Figure 1.3). Because the mycobacterial FAS-II system is incapable of *de novo* synthesis the C16/C18-CoAs synthesized by FAS-I are fed to FAS-II and further elongated into the very long meromycolate chain of MAs (Takayama *et al.*, 2005). The link between the two systems is the FabH (β-ketoacyl-ACP synthase III) enzyme that catalyzes the condensation between malonyl covalently linked to mycobacterial acyl carrier protein (ACP) and the C16/ C18-CoA (Choi *et al.*, 2000). These derivatives are then reduced to acyl-ACP and extended by two carbons by iterative cycles each comprising four steps. The product of the last reaction undergoes the next cycle of elongation as the ACP derivative on another FAS-II module (Marrakchi *et al.*, 2014). The introduction of functional groups is proposed to occur while the mero chain is being synthesized (Marrakchi *et al.*, 2014; Pawełczyk and Kremer, 2014).

The final steps of MAs synthesis involve the condensation of the two chains (Takayama *et al.*, 2005). The C26-CoA released from FAS-I is carboxylated by acyl-CoA carboxylases, yielding α-carboxyl-C26-CoA. The meromycolyl-S-ACP from the FAS-II pathway is converted into meromycolyl-AMP by FadD32, a specific fatty acyl-AMP ligase. The substrates are then condensed by the multidomain Pks13 (the type I polyketide synthase family protein) in a four step reaction culminating in a Claisen-type condensation, generating a mature MA which is transported to a second unknown acceptor (Takayama *et al.*, 2005). The processing of newly synthesized mycolic acids and their subsequent transport for deposition in the cell wall remains poorly understood (Marrakchi *et al.*, 2014; Pawełczyk and Kremer, 2014). The proposed hypothetical pathway involves the transfer of the mature mycolate to an isoprenoid carrier, forming MycPL (6-*O*-mycolyl-β-D-mannopyranosyl-mono phospho- heptaprenol) (Takayama *et al.*, 2005), and subsequently to trehalose inside the cell to yield TMM, the indispensable esterified form under which mycolates are translocated across the membrane (Takayama *et al.*, 2005). TMM is proposed to be transported by MmpL3 (and possibly other transporters of the Mmp family) (Grzegorzewicz *et al.*, 2012) where it is used as a substrate by the mycolyltransferases of the Ag85 complex, to transfer the mycolate chain to (i) either the AG complex to form the cell core bound mycolates, or (ii) to another TMM resulting in formation of TDM (Varela *et al.*, 2012; Grzegorzewicz *et al.*, 2012) or even (iii) to be the substrate of TDM-specific serine esterase resulting in release of free mycolates (Ojha *et al.*, 2010) (Figure 1.3).

Although the mechanisms by which FAs uptake into the cell occurs are unclear, FAS-II can use exogenous FA, or alkanes, for the elongation process, bypassing *de novo* FAS-I synthesis (Niederweis, 2008; Ascenzi and Vestal, 1979). During *M. tuberculosis* dormancy in macrophages host-derived fatty acids have been shown to be critically important energy sources (Peyron *et al.*, 2008). Also required for the pathogen's persistence are host derived lipids and cholesterol, since mycobacteria have the unusual ability to utilize steroids like cholesterol as an energy source (Griffin *et al.*, 2011b).

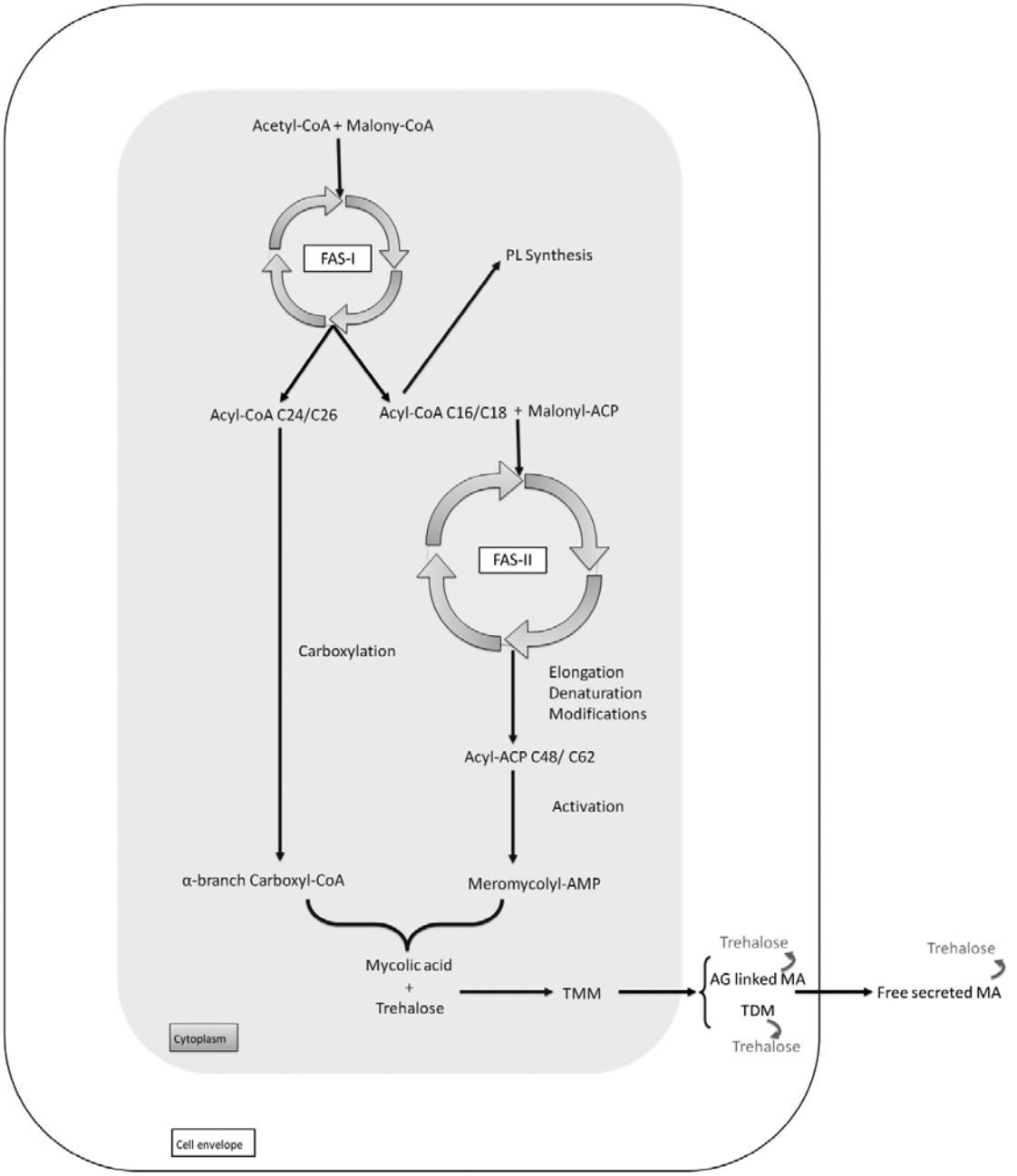


Figure 1.3 - Schematic representation of key steps in the biosynthesis of MAs. Initiated with *de novo* synthesis (FAS-I) and elongation (FAS-II) of FAs. In both systems, chain elongation steps consist of an iterative series of reactions adding a two-carbon unit (provided by malonyl-CoA/ACP) to a nascent acyl group. The carboxylation of FAS-I longer products provides the MA α-branch. The FAS-II synthase products undergo further elongation and modifications/decorations to produce the long mero-chain precursors (C48/C62). Condensation of the activated mero-chain with the α-branch occurs in the Pks13 protein yielding the α-alkyl β-ketoacyl chain that, upon reduction, leads to the mature MA. After processing of newly synthesized MA to TMM it can be translocated across the inner membrane and be further processed into its different forms. PL, phospholipid; TMM, trehalose monomycolate; TDM, trehalose dimycolate; AG, arabinogalactan. Adapted from (Marrakchi *et al.*, 2014; Pawełczyk and Kremer, 2014).

The activity of FabH in *M. tuberculosis* has been shown to be regulated *in vivo* by Ser/Thr phosphorylation (Veyron-Churlet *et al.*, 2009) and a possible post-translational form of regulation of the FAS-II complex condensing enzymes also by Ser/Thr phosphorylation has been recently proposed (Pawełczyk and Kremer, 2014; Molle and Kremer, 2010). The coordination and regulation of FA synthesis is crucial in deciding if the FAS-I products will originate phospholipids, MAs, or the storage compound TAG (triacyl-glycerol), a vital fine-tuning to maintain membrane homeostasis (Mondino *et al.*, 2013), ensuring plasticity and survival under variable conditions.

2. 2 The capsule

In contrast to the long lasting paradigm that only pathogenic mycobacteria had capsules, recent advances in electron microscopy allowed the recognition that all *Mycobacterium* species studied to date have capsule-like structures that primarily consist of polysaccharides and proteins, with generally minor amounts of species-specific lipids (Sani *et al.*, 2010). The overall capsule composition is also species specific, while in *M. tuberculosis* the major surface capsular constituents are polysaccharides, in *M. phlei* and *M. smegmatis* the main components are proteins (Daffé *et al.*, 2014). The same microscopy techniques also allowed visualization of the mycobacterial cell envelope in its native state and proved the existence of a periplasmic space which was traditionally a Gram-negative attribute (Hoffmann *et al.*, 2008).
The capsule main component is a high-molecular-weight α-D-glucan composed of a α-(1→4)-glucopyranose-linked core, branched α-(1→6)-every five or six residues (Daffé *et al.*, 2014) and structurally very similar to the intracellular glycogen of *M. tuberculosis* and *M. bovis* BCG, although its three-dimensional structure appears to be more compact and its molecular mass slightly higher (Dinadayala *et al.*, 2008). An important role for capsular glucan is the regulation of phagocytosis (Stokes *et al.*, 2004), modulation of the immune response (Gagliardi *et al.*, 2007; Geurtsen *et al.*, 2009) and a possible influence in the persistence phase of infection (Sambou *et al.*, 2008).

3. Intracellular glucans in mycobacteria

In addition to the extracellular sugars mentioned above, mycobacteria also produce a number of intracellular carbohydrates and derivatives that include glycogen, trehalose and a group of unique polymethylated polysaccharides (PMPSs) with a proposed crucial role on fatty acids metabolism. These intracellular glucans have received less attention compared with the structures at the interface with the host and most of their physiological functions remain somewhat mysterious.

3. 1 Glycogen biosynthesis

Glycogen is a soluble glucose polymer composed of approximately 90% α-(1→4)-linked glucopyranose residues as the core structure that branch at C6 OH positions. It is a common energy source in many organisms including mycobacteria (Henrissat *et al.*, 2002; Chandra *et al.*, 2011) and lacks an overall defined structure or molecular weight (Chandra *et al.*, 2011).
The structural similarity between capsular α-D-glucan and glycogen (see section 2.2) allowed the identification of most of the mycobacterial genes involved in their biosynthesis (Sambou *et al.*, 2008). In *M. tuberculosis* H37Rv the involvement of two distinct α-(1→4)-glucosyltransferases, Rv3032 and GlgA (*Rv1212c*), of ADP-glucose synthase GlgC (*Rv1213*) and of the branching enzyme GlgB (*Rv1326c*) responsible for introducing α-(1→6)-linked branches into linear α-(1→4)-glucans, was confirmed by the analysis of mutants and complementation studies (Sambou *et al.*, 2008). Despite the *glgA* mutant being impaired in its ability to persist in mice and showing diminished glucan production, a partial redundancy between GlgA and Rv3032 (α-(1→4)-glucosyltransferase) was observed because glucan production in the *glgA*-deleted strain could be restored after complementation with a wild-type copy of the *Rv3032* gene (Sambou *et al.*, 2008). However, while abolition of GlgA function affected mainly glucan synthesis, inactivation of *Rv3032* affected glycogen synthesis, without affecting capsular glucan production (Sambou *et al.*, 2008). The authors concluded that a functional copy of at least one of these genes is required for growth as they were unable to delete both genes simultaneously. Whether this physiological requirement is related to the synthesis of glycogen, capsular glucan or even the chemically related mycobacterial methylglucose lipopolysaccharides (MGLPs) has not yet been elucidated (Sambou *et al.*, 2008; Jackson and Brennan, 2009; Kalscheuer *et al.*, 2010a). GlgA (*Rv1212c*) and the glycosyltransferase Rv3032 have also been found to be involved in the biosynthesis of unique intracellular MGLPs, which were postulated to modulate the synthesis of fatty acids and whose main chain is also a oligomer of α-(1→4)-linked glucoses and methylglucoses with two branching β-(1→3)-glucoses (Stadthagen *et al.*, 2007) (see section 3.4). Furthermore mycobacterial α-(1→4)-glucans can also be synthesized from trehalose through a four-step pathway comprising trehalose synthase (TreS), maltokinase (Mak/Pep2), maltose-1-phosphate maltosyltransferase (GlgE), and glycogen branching enzyme (GlgB) (Kalscheuer *et al.*, 2010a) (see section 3.3).

3. 2 Trehalose roles and metabolism

Trehalose is a ubiquitous compound synthesized by many different organisms including bacteria, yeasts, fungi, plants and invertebrates (Paul *et al.*, 2008). Free trehalose has multiple roles: it can regulate energetic supplies and intracellular signaling, serve as a stress bioprotectant and it can also stabilize macromolecular structures such as lipid membranes and proteins (Paul *et al.*, 2008). Trehalose is a non-reducing disaccharide of α-(1→1)-linked glucose that is an essential metabolite in mycobacteria (Murphy *et al.*, 2005; Woodruff *et al.*, 2004) with functions ranging from intracellular carbon storage and stress protectant to being the precursor of other

important molecules involved in carbohydrate metabolism, cell wall architecture , especially being an essential component of various cell wall glycolipids, or mycobacteria-host interactions (Nobre *et al.*, 2014) (see section 2.1.1). Since trehalose is not synthesized in mammals (although trehalases are present) the mycobacterial trehalose biosynthetic machinery represents an attractive target for the development of new antimycobacterial agents and has been the focus of much attention (Nobre *et al.*, 2014).

There are five known different pathways for trehalose biosynthesis and, while most bacteria possess only one *de novo* biosynthetic pathway, in mycobacteria three different pathways are involved in trehalose synthesis namely the OtsA-OtsB, TreY-TreZ, and TreS pathways (Paul *et al.*, 2008; Nobre *et al.*, 2014). The essentiality of this sugar was established after the observation that *M. smegmatis* triple mutants were unable to grow unless trehalose was exogenously supplied (Woodruff *et al.*, 2004). At elevated temperatures the mutant was unable to proliferate, even in the presence of trehalose, which suggested a dual role for trehalose as both a thermoprotectant and a precursor of critical metabolites (Woodruff *et al.*, 2004). In *M. tuberculosis* the OtsA/OtsB pathway (Ots, osmotic trehalose synthesis, was the designation that emerged from studies in *E. coli*) is the dominant pathway for trehalose formation (Murphy *et al.*, 2005). OtsA (or Tps, trehalose-6-phosphate synthase) condenses NDP-glucose and glucose-6-phosphate into the phosphorylated intermediate, trehalose-6-phosphate (T6P), which is subsequently dephosphorylated to trehalose by a specific trehalose-6-phosphate phosphatase (OtsB or Tpp) (Murphy *et al.*, 2005) a function that in *M. tuberculosis* relies on OtsB2 (*Rv3372*) (Murphy *et al.*, 2005). This gene could not be inactivated even in the presence of exogenously supplied trehalose, proving its essentiality in *M. tuberculosis* and *M. bovis* BCG (Murphy *et al.*, 2005). One explanation for this lethal phenotype might be the potential toxic effect of an accumulation of the intermediate trehalose-6-phosphate (Kalscheuer and Koliwer-Brandl, 2014).

In the alternative TreY/TreZ (*Rv1563c* and *Rv1562c*) pathway a reducing end disaccharide unit in a glycogen chain [α-(1→4) linkage] is isomerized to trehalose [α-(1→1) linkage] by a maltooligosyltrehalose synthase (TreY) and subsequently hydrolyzed by a maltooligosyltrehalose trehalohydrolase (TreZ) to yield one molecule of free trehalose and a glycogen chain two glucose residues shorter (De Smet *et al.*, 2000; Paul *et al.*, 2008). In contrast to the essentiality of the OtsAB route, this pathway was shown to be non-essential in *M. tuberculosis* as inactivation of TreY and TreZ had no detectable effects on *in vitro* growth or mice infection (Murphy *et al.*, 2005). Paradoxically this operon-like cluster contains an additional *TreX* gene (*Rv1564c*) encoding a putative glycogen-debranching enzyme which is proposed to be essential for *M. tuberculosis* growth (De Smet *et al.*, 2000; Griffin *et al.*, 2011b). However, experimental validation of TreX involvement in mycobacterial trehalose metabolism is still missing. While the TreYZ pathway may not be essential for *M. tuberculosis* survival it can still play important roles in trehalose metabolism namely in sugar recycling (see section 3.3).

The enzyme trehalose synthase (TreS) catalyses the reversible isomerization of the α-(1→4) linkage of maltose into an α-(1→1) linkage in trehalose (De Smet *et al.*, 2000). *In vitro* and *in vivo* studies initially suggested that the enzyme converted free maltose to free trehalose (De

Smet *et al.*, 2000; Pan *et al.*, 2004; Woodruff *et al.*, 2004). However, a recent *in vivo* study has demonstrated that the direction of flux occurs in fact from trehalose to maltose and that the TreS pathway does not contribute to the *de novo* biosynthesis of mycobacterial trehalose (Miah *et al.*, 2013), arguing against the notion that in *M. smegmatis* all three trehalose pathways are functionally redundant (Woodruff *et al.*, 2004). The trehalose-maltose flux is driven, rapidly and irreversibly, by the ATP-dependent phosphorylation of the newly formed maltose to maltose-1-phosphate, via maltokinase/Pep2 (*Rv0127*) (Miah *et al.*, 2013) (Figure 1.4). However, TreS conversion of maltose to trehalose in organisms that have access to sufficient cytosolic maltose, from intracellular or extracellular sources, cannot be completely ruled out (Miah *et al.*, 2013).

3. 3 Maltose-1-phosphate synthesis and functions

Maltose-1-phosphate (M1P) was initially identified in 1967 in *M. bovis* BCG cell extracts (Narumi and Tsumita, 1967) and almost thirty years later Drepper and colleagues identified an enzyme in the bacterium *Actinoplanes missouriensis* capable of catalyzing an ATP-dependent phosphorylation of maltose to maltose-1-phosphate which they named maltokinase (Mak) (Drepper *et al.*, 1996). *Escherichia coli* was also suggested to be able to produce M1P via TreC (trehalose-6-phosphate hydrolase) in an ATP-independent manner and it was speculated that the phosphorylated disaccharide might play a role as an inducer of the maltose operon (Decker *et al.*, 1999). Since then several Mak homologues have been functionally characterized as maltokinases namely the enzymes from *A. missouriensis* (Niehues *et al.*, 2003), *Streptomyces coelicolor* (Jarling *et al.*, 2004), *M. tuberculosis/M. bovis* BCG (Mendes *et al.*, 2010; Roy *et al.*, 2013) and *M. smegmatis* (Elbein *et al.*, 2010).

Homologues of maltokinases (EC 2.7.1.175) are disseminated throughout bacterial phyla but absent from the genomes of eukarya. The Mak protein presents a highly conserved function despite having low amino acid identity even within the same genus. In mycobacteria the *mak* gene is usually located next to the *treS* gene and, in many other organisms, is very common to find both genes fused in a bifunctional unit (Mendes *et al.*, 2010) underlining the functional link between the two corresponding proteins. Interestingly, *M. abscessus* and *M. leprae* lack the traditional Mak (Mendes *et al.*, 2010) and while the reason for this absence from *M. abscessus* is still elusive, it is not surprising for *M. leprae* that lacks most enzymes for glycogen metabolism, a consequence of a severe genomic reductive evolution evolution determining its strictly parasitic lifestlyle (Henrissat *et al.*, 2002).

Recently, a link between trehalose, maltose and glycogen metabolism has been established in a novel four step pathway involving TreS, Mak, GlgE and GlgB (Elbein *et al.*, 2010; Kalscheuer *et al.*, 2010a) (Figure 1.4). In this pathway trehalose is converted to maltose by TreS, and subsequently phosphorylated by Mak to M1P, which is the substrate GlgE utilizes to elongate the nonreducing end of an α-(1→4)-glucan acceptor molecule by a maltosyl moiety (Kalscheuer *et al.*, 2010a; Elbein *et al.*, 2010). Finally, GlgB introduces α-(1→6)-linked branches into the linear α-(1→4) chain (Kalscheuer *et al.*, 2010a), ultimately producing glycogen (Elbein *et al.*, 2010) (Figure 1.4). Trehalose for this GlgE pathway may arise *de novo* via the OtsA/OtsB route, but the

alternative TreY-TreZ pathway may allow for glycogen and trehalose salvaging, in accordance with genetic studies that suggested a role for GlgE in glycogen recycling (Chandra *et al.*, 2011). Alternatively, trehalose may be a byproduct of TDM hydrolysis and transported across the cell wall into the cytoplasm, a recycling process that was shown to be critical for *M. tuberculosis* to establish infection in mice (Kalscheuer *et al.*, 2010b). It was postulated that after a stress signal cellular trehalose levels would be high and ATP levels would be low, thus the GlgE pathway was suggested to be active mainly under heat stress conditions, preventing GlgE inhibition by physiological ATP levels (> 1 mM) (Elbein *et al.*, 2010) to which the enzyme is particularly sensitive. Under these conditions, the GlgE pathway would be activated and the excess trehalose could be converted to glycogen and stored for future use in metabolism (Elbein *et al.*, 2010). Structural studies with *M. tuberculosis* TreS showed that the protein formed tetramers in solution and what seemed like an unusually large hetero-octameric complex with Mak, which was proposed to markedly accelerate maltokinase activity (Roy *et al.*, 2013).

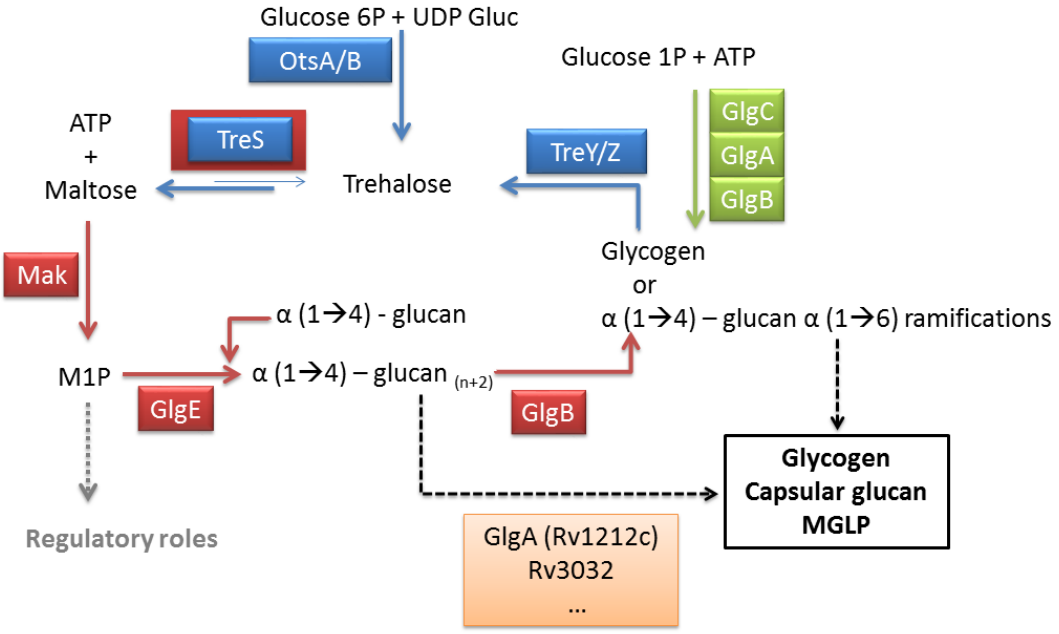


Figure 1.4 - Pathway for synthesis and recycling of trehalose and glycogen to α-glucans. Enzymes involved in the trehalose pathways are boxed in blue and enzymes involved in the newly discovered GlgE pathway are boxed in red. Glycogen pathway is highlighted in green. Dashed lines indicate hypothetical functions. Adapted from (Nobre *et al.*, 2014; Miah *et al.*, 2013; Chandra *et al.*, 2011; Kalscheuer *et al.*, 2010a).

The GlgE pathway can also contribute to the formation of molecules other than glycogen, namely capsular glucan, although this remains to be experimentally confirmed (Kalscheuer *et al.*, 2010a). A possible role was also suggested for the GlgE pathway in the biosynthesis of the unique mycobacterial intracellular lipopolysaccharides (MGLPs). Genetic studies implied that the GlgE pathway and the glycosyltransferase Rv3032, both associated with glycogen synthesis, are also possibly involved in the production of the structurally related MGLP (Kalscheuer *et al.*, 2010a). This would explain why the GlgE pathway is dispensable for the viability and virulence of *M. tuberculosis*, as long as the redundant Rv3032 pathway remains functional and also why the simultaneous inhibition of both pathways is lethal (Kalscheuer *et al.*, 2010a).

The genes *mak, glgE* and *glgB* were all proposed to be essential for *M. tuberculosis* growth by saturation transposon mutagenesis (Griffin *et al.*, 2011b) and *glgB* essentiality was further confirmed by allelic exchange mutagenesis (Sambou *et al.*, 2008). The gene *glgE* was also validated as an essential gene based on the lethal accumulation of M1P, which caused rapid death of *M. tuberculosis* both *in vitro* and *in vivo* through a phenomenon designated synthetic lethality (Kalscheuer *et al.*, 2010a). In the absence of GlgB branching activity GlgE produces linear α-(1→4)-linked glucans that have fewer non-reducing ends available for maltosyl transfer, which can lead to a toxic buildup of M1P (Kalscheuer *et al.*, 2010a). Therefore it was postulated that *glgB* essentiality might be related to this M1P self-poisoning and not to direct accumulation of insoluble α-glucans as believed (Sambou *et al.*, 2008).

3. 4 The polymethylated polysaccharides MGLP and MMP

Polymethylated polysaccharides (PMPSs) are a unique class of glucans found in mycobacteria and a restricted group of related bacteria (Table 1-I) that are postulated to modulate FAs synthesis by sequestering fatty acyl-CoAs (Mendes *et al.*, 2012). In 1964, Lee and Ballou reported finding in *M. phlei* and *M. tuberculosis* 6-*O*-methyl-α-D-glucose precursors of a novel polysaccharide (Lee and Ballou, 1964), later shown to be acylated and named MGLP (methylglucose lipopolysaccharide) (Lee, 1966; Keller and Ballou, 1968). In 1966 a polysaccharide containing 3-*O*-methyl-D-mannose residues was isolated from *Streptomyces griseus* (Candy and Baddiley, 1966), followed by the isolation and characterization of a very similar polymer of 3-*O*-methyl-D-mannose from *M. phlei* which was designated MMP (methylmannose polysaccharide) (Gray and Ballou, 1971). Much of the information available about these polysaccharides arose from early studies conducted in the 1960's in the Ballou and Bloch laboratories.

Table 1-I - Distribution of PMPSs in *Mycobacterium* species and in strains of related genera.

Genera	Species	MGLP	MMP	Growth	Reference
Mycobacterium	*M. smegmatis*	✓	✓	Rapidly-growing	(Kamisango *et al.*, 1987),(Bergeron *et al.*,
	M. phlei	✓	✓		(Lee, 1966), (Gray and Ballou, 1971), (Weisman and Ballou, 1984b)
	M. parafortuitum	✓	✓		(Weisman and Ballou, 1984b)
	M. aurum	✓	✓		(Weisman and Ballou, 1984b)
	M. chitae	✓	✓		(Weisman and Ballou, 1984b)
	M. vacae		✓		(Tian *et al.*, 2000)
	M. tuberculosis	✓		Slowly-growing	(Lee, 1966; Stadthagen *et al.*, 2007)
	M. bovis	✓			(Tuffal *et al.*, 1998)
	M. leprae	✓			(Hunter *et al.*, 1986)
	M. xenopi	✓			(Tuffal *et al.*, 1995b)
Nocardia	*N. otitidis-caviarurn*	✓			(Pommier and Michel, 1986)
	N. brasiliensis	✓			(Pommier and Michel, 1986)
	N. farcinica	✓			(Pommier and Michel, 1986)
	N. kirovani	✓			(Pommier and Michel, 1986)
Streptomyces	*S. griseus*		✓		(Candy and Baddiley, 1966)

The strain numbers are not indicated because several laboratory isolates were utilized in some studies. Adapted from (Mendes et al., 2012; Alarico et al., 2014).

Mycobacterium smegmatis MMPs exist as a mixture of four homologues that have a linear chain of 11 to 14 α-(1→4)-linked 3-*O*-methylmannoses with an α-(1→4)-linked unmethylated mannose at the non-reducing end, whereas the reducing end is blocked by a methyl aglycon (Maitra and Ballou, 1977) (Figure 1.5). MMPs appear to be restricted to rapidly-growing mycobacteria and *S. griseus* (Table 1-I). S*treptomyce griseus* MMP is named AMMP due to the presence of additional acetyl residues at C6 OH position of some methylmannoses (Harris and Gray, 1977). In contrast, MGLPs seem to be the only form of PMPS present in slowly-growing mycobacteria (Table 1-I).

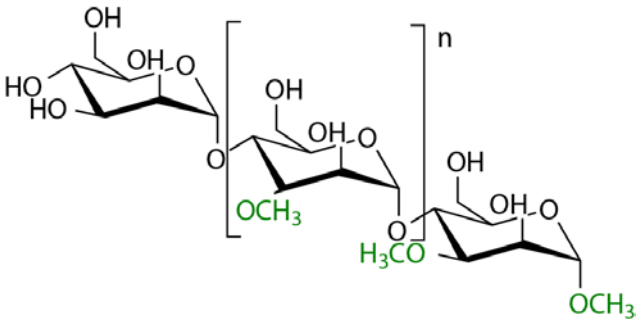


Figure 1.5 - Scheme depicting the general structure of mycobacterial MMPs from *M. smegmatis*. (n) indicates 9 to 12 residues according to (Maitra and Ballou, 1977). Methyl groups highlighted in green.

Following the isolation of 6-*O*-methyl-α-D-glucose by Lee and Ballou in 1964 (Lee and Ballou, 1964) further studies established that the molecule was in fact a lipopolysaccharide (Lee, 1966). After exhaustive structural studies, the discovery of 3-*O*-methyl glucose in the MGP (core structure) non-reducing end and of glyceric acid in the reducing end culminated with a first model of the complete structure of the *M. phlei* polysaccharide backbone (Saier and Ballou, 1968a, b). In the same year Keller and Ballou determined the acyl substitutions of *M. phlei* MGLP and concluded that the molecule had esterified acetate, propionate, isobutyrate and octanoate, in the molar ratios of 3: 1: 1: 1. The level of heterogeneity in MGLP allowed a separation in four forms (which they named MGLP-I to MGLP-IV) due to differences in charge owing to the presence of 0 to 3 monoesterified succinate residues (Keller and Ballou, 1968). Later, in 1972, Gray and Ballou identified eight specific locations of acyl groups in the MGLP structure and concluded that the non-reducing terminus 3-*O*-methylglucose residue was substituted at positions 4 and 6 as well as the subsequent glucose residue at position 6 (Gray and Ballou, 1972). Furthermore one or both of the neighboring glucoses were substituted at position 6. On the other extremity, the second and third glucose residues from the glyceric acid end contained acyl substituents at position 6 and the glyceric acid was found to be esterified on its primary hydroxyl group (Gray and Ballou, 1972) (Figure 1.6 panel 1). In 1973, the specific acyl groups at each location were defined by Smith and Ballou, after determining the location of the neutral and acidic groups (succinyl). They found that five of the six neutral groups were located at specific sites at the non-reducing end of the molecule and that the location of the some monobasic acids was somewhat heterogenous (Smith and Ballou, 1973). These authors also concluded that the branching glucose residue was succinylated and that succinylation also occurred on the glucose residues near the glyceric acid. From the analysis of un-succinylated MGLP it was suggested that

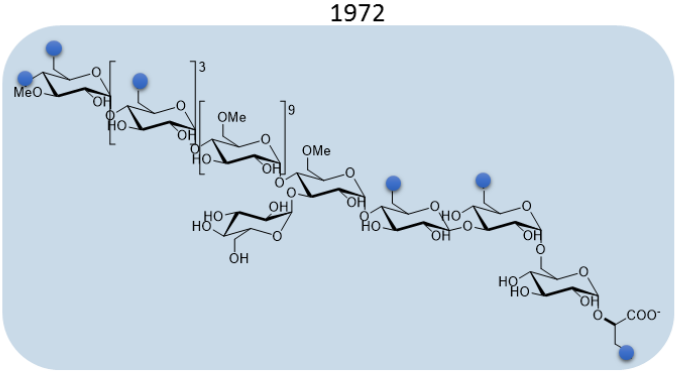
the octanoyl group would be esterified to position 3 of glyceric acid (Smith and Ballou, 1973) (Figure 1.6, panel 2). Since the work of Tung and Ballou the MGP (core) structural model has been corrected twice, namely for an additional side chain glucose unit in 1982 (Forsberg *et al.*, 1982) (Figure 1.6, panel 3) and then in 1998 when the β-(1→3) branching linkages were estabished (Tuffal *et al.*, 1998) (Figure 1.6, panel 4). However, acylation positions were not re-examined since and the most recent analysis on this context was conducted in 1983 by fast-atom-bombardment mass spectrometry (FABMS) by Dell and Ballou confirming the results of Tung and Ballou (Figure 1.6, panel 3) (Dell and Ballou, 1983).

MGLP's currently accepted structure is based on early studies with *M*. *phlei* acylated forms (Dell and Ballou, 1983) and in the major mature form present in *M. bovis* BGC (Tuffal *et al.*, 1998; Jackson and Brennan, 2009). MGLP is postulated to be a linear chain of 15-20 α-(1→4)-linked glucoses, nine of which are 6-*O*-methylglucose units and with a non-reducing end of 3-*O*-methylglucose. The reducing end is composed of glyceric acid α-(1→2)-linked to a D-α glucopyranose (glucosylglycerate) in turn linked to a second glucose through an α-(1→6) linkage that initiates the main MGLP chain, which has two additional β-(1→3)-linked branching glucoses. In the mature structure the short-chain fatty acids, acetate, propionate or isobutyrate, esterify the glucoses near the polysaccharide's non-reducing end while the β-(1→3)-branching glucoses are esterified with succinate. The octanoate moiety in MGLP was proposed to esterify the primary hydroxyl group of glyceric acid (Forsberg *et al.*, 1982; Tuffal *et al.*, 1998; Dell and Ballou, 1983) (Figure 1.6 last panel). *In vivo,* MGLPs exist as a mixture of structurally complex molecules that vary in the number of glucose units and methyl groups, as well as in the number and nature of acyl group modifications (Tuffal *et al.*, 1995a; Tuffal *et al.*, 1995b).

3. 4. 1 Physiological role

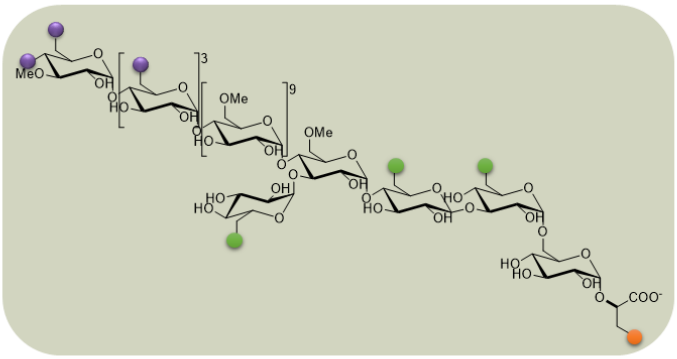
Due to stereochemical constraints arising from the α-(1→4)-linked PMPSs main chains, they are likely to adopt a coiled conformation in solution, forming a cylindrical hydrophobic cavity with the methyl groups facing innward and where a long-chain fatty acid derivative (acyl-CoA) can be accomodated (Bergeron *et al.*, 1975; Yabusaki and Ballou, 1978). The complexation ability of PMPSs appears to heighten after two complete turns of a tightly coiled helix and the length of such a helix is roughly equal to that of the extended palmitic acid (C16) or stearic acid (C18) chain (Kiho and Ballou, 1988). Thus, it is reasonable to assume that these fatty acid lengths provide optimum binding (K_d 0.1 μM) (Yabusaki and Ballou, 1978) whereas smaller chains form weaker complexes (Kiho and Ballou, 1988; Yabusaki and Ballou, 1978). When an acyl-CoA derivative binds to the polysaccharide it changes its conformation from a random coil to a more ordered helix (Maggio, 1980). Consequently, the complexed acyl-CoA may fold over to assume a hairpin-like structure, where some charged moieties of the coenzyme can interact with the hydrophilic polysaccharide exterior, stabilizing the complex (Bergeron *et al.*, 1975).

1972



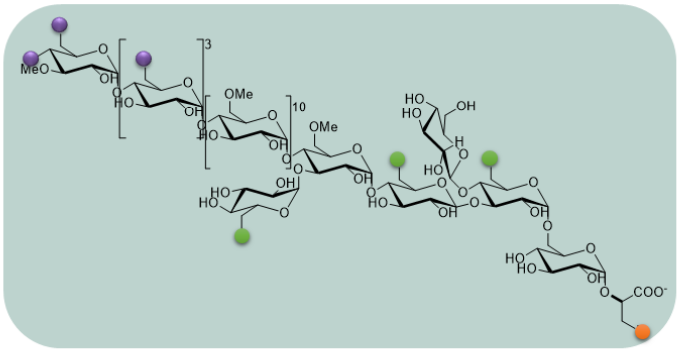
Gray and Ballou, 1972

1973



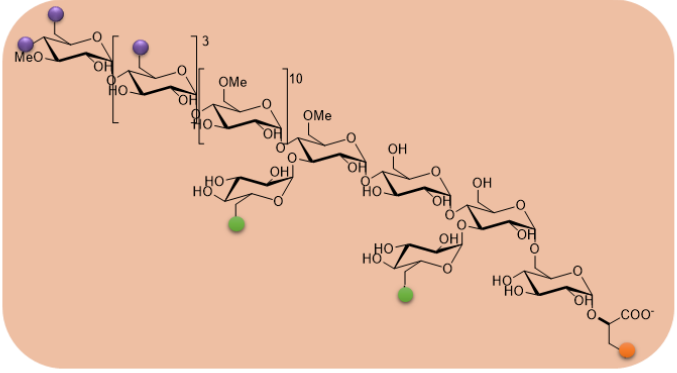
Smith and Ballou, 1973

1983



Forsberg and Ballou, 1982; Dell and Ballou, 1983

2009



Tuffal *et al.*, 1998; Jackson and Brennan, 2009

Figure 1.6 - Evolution of MGLP structure and acylation position over time. Blue circles represent any of the acyl groups detected in MGLP (acetate, propionate, isobutyrate, succinate or octanoate); purple circles represent acetate, propionate or isobutyrate; green circles represent succinate and orange circles represent octanoate. (Gray and Ballou, 1972; Smith and Ballou, 1973; Forsberg *et al.*, 1982; Dell and Ballou, 1983; Tuffal *et al.*, 1998). The bottom panel corresponds to the currently accepted structure of MGLP from *M. bovis* BCG (Tuffal *et al.*, 1998; Jackson and Brennan, 2009).

This binding capability provides PMPSs the ability to form stable 1:1 complexes with long acyl-coAs (Machida and Bloch, 1973) which is the core property of their proposed regulatory role in FAs synthesis simultaneously allowing (i) protection of FAS systems against end product inhibition by accelerating the rate of the over-all elongation process and the rate of diffusion of neo-synthesized acyl chains (ii) protection of acyl-CoA chains from degradation by cytoplasmic lipolytic enzymes while also preventing disruption of metabolism (Bloch and Vance, 1977).

The diffusion of CoA derivatives is rate-limiting for FAS-I and, consequently, the ability of PMPSs to sequester FAS-I end products is essential for enzyme turnover dramatically increasing the rate of fatty acid synthesis, also accelerating the otherwise slow diffusion of C24-C26 acyl-CoAs (Banis *et al.*, 1977). As a consequence of this accelerating effect there is an expected change on the bimodal chain distribution, shifting the pattern from C24-26 to C16-18 acyl-CoAs (Wood *et al.*, 1977). These shorter chain fatty acids can enter the synthesis of the conventional membrane phospholipids or be fed to FAS-II to be further elongated. Both these longer chains and the C24-26 CoAs serve as precursors for components of the more complex MAs (Figure 1.3 / Figure 1.7).

FAS-I from *M. phlei* was discovered to be dependent on a heat-stable fraction for normal FAs synthesis at low acetyl-CoA concentrations (Ilton *et al.*, 1971). This heat-stable fraction was shown to correspond to PMPSs (BSA could substitute to some extent), which markedly lowered the enzyme's K_m for acyl-CoA, the primer for FAs elongation (from 200 µM in the absence of PMPS to 4 µM in the presence of PMPSs) (Ilton *et al.*, 1971). The presence of purified PMPSs *in vitro* also had the ability to shift the pattern of synthesis of FAs chain length (Flick and Bloch, 1974; Papaioannou *et al.*, 2007).

In a FAS-I *M. smegmatis* purified fraction, at low acetyl-CoA concentrations (20 µM), the elongation process was fast but the termination reaction was slow. Because the amount and length of acyl chains released depends on the relative rates of elongation and termination this resulted in a C24-C26 acyl-CoA distribution pattern. In the presence of MMP the termination rates increased because MMP forms a complex with neo-synthesized acyl-CoAs thus aiding the ordinarily slow rate-limiting release of product, irrespective of chain length. Nontheless, the effect on chain length was unnoticeable because elongation rates were still greater than termination rates. At higher acetyl-CoA concentrations (300 µM) the elongation rates were reduced because acyl-CoA competes for active sites and the effect of MMP on termination became sufficient to yield a shift in pattern length (Wood *et al.*, 1977). The bimodal change in chain length is influenced by an array of factors, including acetyl-CoA to malonyl-CoA ratios, and temperature (Flick and Bloch, 1975; Bloch, 1977). Raising the acetyl-CoA to malonyl-CoA ratio from 0.4 (8 µM acetyl-CoA) to 150 (3 mM acetyl-CoA) increased the percentage of shorter chain fatty acids from 12% to 87% and also caused an overall 5 to 10-fold increase in FAs synthesis. The addition of PMPSs (at 300 µM acetyl-CoA) caused a similar shift in the bimodal distribution, from a low to a high (25% to 85%) proportion of shorter FA chain, and in the over-all FA synthetic rate (Flick and Bloch, 1974). Notably, in the presence of PMPSs this shift occured at a lower acyl-CoA concentration enabling fatty acid synthesis to proceed at low and more physiological substrate concentrations (Figure 1.7).

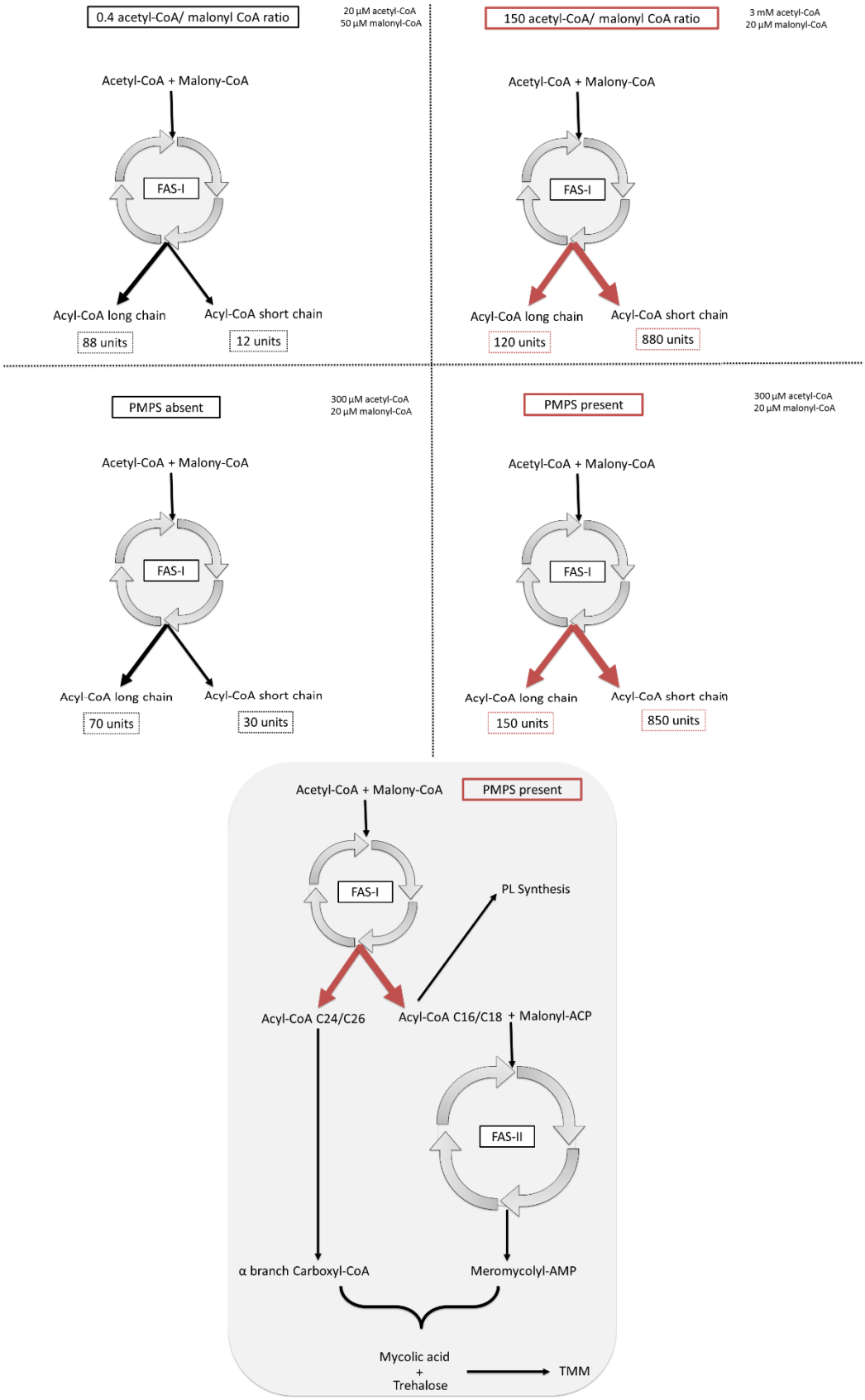


Figure 1.7 - Schematic representation of PMPS and acetyl-CoA/malonyl-CoA ratio effect on bimodal pattern of FAS-I and FA synthetic rate for *M. smegmatis*. Units were calculated based on percentages presented in Flick and Bloch (1974) assuming a 10 fold increase in FA synthetic rate. Short chains are a mixture of C14-C16-C18 the dominant products being C16 and C18. Long chains are a mixture of C20-C22-C24 where the dominant product is C24. Boxed in grey is represented the integration of FAS-I with FAS-II system and MA biosynthesis in mycobacteria, in the presence of PMPS.

Raising mycobacterial growth temperature from 37 to 45°C tends to increase the average chain length for both FAS systems (Odriozola *et al.*, 1977). In a fraction of purified *M. smegmatis* FAS-I, the ratio C16-C18 to C24 acyl-CoAs was stable at 25°C, while at 45°C the proportion of C24 acyl-CoA was thrice the proportion of the shorter chain FAs (Bloch, 1977). The FA distribution pattern is highly flexible but persistently bimodal and changes in distribution in response to environmental and physiological changes illustrate the unusual flexibility of the mycobacterial FAS-I (Bloch, 1977).

The newly synthetized fatty acyl chains are likely sequestered into the inner cavity of the coiled polysaccharide (Yabusaki and Ballou, 1979). The complexating ability is key when protecting FA chains as demonstrated by a *M. smegmatis* MGLP analog lacking the four non reducing end glucoses (after amylase digestion), which lost the ability to bind the substrate and could not prevent subsequent acyl-CoA hydrolysis (Yabusaki and Ballou, 1979). Because the intracellular concentration of long chain acyl-CoA is about 0.3 mM and the concentration of PMPSs approaches 1 mM, considering an estimated K_d of 0.1 µM, all the long-chain-CoAs in the cytosol should form complexes, which would not only protect acyl-CoAs from degradation but could also confer PMPSs a role as lipid carriers (Yabusaki and Ballou, 1978).

The actual importance of PMPSs for regulation of fatty acid metabolism remains to be experimentally demonstrated. The first attempt to study their *in vivo* function was carried out in a *M. smegmatis* spontaneous mutant with reduced levels of PMPSs (50% of MGLP and 7% of MMP when compared to WT). However fatty acid synthesis was not dramatically altered, except for the accumulation of higher levels of short-chain and less long-chain unsaturated FAs relative to WT cells (Maloney and Ballou, 1980). The lack of any severe anomalies, except for the slower growth rate when compared to WT, could possibly be explained by the appreciable levels of MGLP still being produced by the mutant (Maloney and Ballou, 1980).

In order to further investigate PMPSs physiological functions, genetic studies regarding MGLP biosynthetic genes were conducted. The α-(1→4)-glucosyltransferase gene, responsible for MGLP elongation (*Rv3032*), and the methyltransferase for 6-*O*-methylation (*Rv3030* orthologue *MSMEG_2349*) were inactivated in *M. tuberculosis* and *M. smegmatis,* respectively (Stadthagen *et al.*, 2007). The *M. tuberculosis Rv3032* mutant presented a slower growth rate at 37°C and growth was abolished for both mutants at temperatures higher than 39°C (ΔRv3032) or 42°C (ΔMSMEG_2349) (Stadthagen *et al.*, 2007). When grown at 37°C, they both displayed WT fatty acid content albeit with diminished MGLP synthesis and less glycosylated (shorter) when compared to WT form (Jackson and Brennan, 2009; Stadthagen *et al.*, 2007). It was noted however a compensatory activity not only for the glucosyltransferase but also for the *O*-methyltransferase (Stadthagen *et al.*, 2007) which frustrated total abolition of MGLP synthesis. Another study analyzed the disruption of the *gpgS* gene responsible for initiation of MGLP biosynthesis (*Rv1208* orthologue *MSMEG_5084*). Similarly to the *M. tuberculosis* Δ*Rv3032* the *M. smegmatis* Δ*gpgS* showed a significantly reduced growth rate (Kaur *et al.*, 2009) with impared *de novo* synthesis of MGLP (80% of WT) but, once again, the production of the polysaccharide was not totally abolished and neither FA metabolism nor their levels were affected (Kaur *et al.*, 2009). This could be due to residual amounts of MGLP proposed to be the result of unidentified

GpgS or GG synthase activities, leading to the synthesis of the MGLP primer GG (see below) or, most probably, to the compensatory effect of the related polysaccharide MMP present in *M. smegmatis,* whose levels were unaffected in this Δ*gpgS* mutant (Kaur *et al.*, 2009).
The results from the abovementioned genetic studies in addition to the proposed function of PMPSs as regulators of FAs metabolism strongly indicate that the polysaccharides may also be critically involved in mycobacterial adaptation to thermal stress, directly or indirectly, through interference with FAs metabolism and supply to the cell envelope (Jackson and Brennan, 2009; Stadthagen *et al.*, 2007). As stated in section 2.1.1, when growing at higher temperatures mycobacteria synthesize longer mycolic acids chains to maintain cell wall integrity (Toriyama *et al.*, 1980; Baba *et al.*, 1989; Liu *et al.*, 1996). The ability of PMPSs to modulate fatty acid synthase together with the temperature sensitivity of the MGLPs-defective mutants, strongly suggest a vital role in adaptation to thermal stress (Stadthagen *et al.*, 2007; Mendes *et al.*, 2012).
The specific role of MGLP acylation is unknown although it has been suggested to regulate its own methylation levels(Grellert and Ballou, 1972) (see section 3.4.2). In addition to and similarly to methylation, the short chain neutral acyl groups are also expected to enhance the nonpolar interactions of the polysaccharide with FAs adding hydrophobicity to the interior of the coil (Hindsgaul and Ballou, 1984). Moreover, it has been suggested that acylation further stabilizes the polysaccharide in helical conformation, even in the absence of bound lipid, thus contributing to a more ordered configuration. The octanoyl group specifically has been proposed to behave as an axis around which MGLP could stably fold in the absence of longer acyl-CoAs (Hindsgaul and Ballou, 1984), also raising the interaction energy required for complexation with longer fatty acyl-CoAs hence conferring MGLP further discriminatory capability when compared to MMP (Hindsgaul and Ballou, 1984). The octanoate moiety was further suggested to attach MGLP intermediates to the inner leaflet of the cytoplasmic membrane acting as a stabilization anchor during during the elongation steps (Smith and Ballou, 1973). The acidic succinyl groups could possibly be involved in interactions with the proteins or with cations because they were reported to facilitate the transport of calcium ions when the lipopolysaccharide was incorporated into artificial lipid bilayers (Smith and Ballou, 1973) or even help to increase the solubility of MGLP-lipid complexes (Hindsgaul and Ballou, 1984).

3. 4. 2 Biosynthesis of MGLP

The steps towards identifying the enzymes involved in PMPSs biosynthesis were initiated in the Ballou laboratory with the isolation and characterization of a native enzyme for the transfer of methyl groups from S-adenosylmethionine to glucose (Ferguson and Ballou, 1970). However, further elucidation of the pathway only became possible with the growing availability of sequenced mycobacterial genomes and the advances in molecular biology techniques. The general aspects of MMP biosynthesis were initially elucidated by Weisman and Ballou, who detected activities corresponding to an α-(1→4)-mannosyltransferase (ManT) (Weisman and Ballou, 1984b) and a 3-*O*-methyltransferase (OMetT) in *M. smegmatis* cell-free extracts (Weisman and Ballou, 1984a). A model for MMP biosynthesis was proposed in which elongation

would occur by alternating mannosylation followed by methylation and in which termination of elongation would take place when the chain reached 10–14 monosaccharide residues in length, which is compatible with fatty acid-binding (Weisman and Ballou, 1984a). This model proposed that MantT elongation starts with a mannose tetrasacharide but did not explain how the biosynthesis is initiated (Weisman and Ballou, 1984b).

More recently Xia and co-workers showed that MMP elongation does not require the sequential activity of ManT and OMetT but that instead ManT can independently produce a linear polymer polymer using an α-(1→4)-linked tetramannoside as substrate. In this alternative model the unmethylated mannose chains (or MMP fragments) are first extended by ManT and subsequently methylated at O-3 of each mannose by OMetT, implicating that ManT could introduce more than one mannose, or even construct the whole chain, prior to methylation (Xia *et al.*, 2012). These authors also identified, overexpressed in *E. coli* and confirmed in cell extracts the ManT activity responsible for chain elongation (Xia, 2013). To date no other genes proposed to be involved in the biosynthesis of MMPs have been identified.

MGLP has received substantially more attention that MMP because it is the only PMPS found in slowly-growing mycobacteria such as *M. tuberculosis* (Jackson and Brennan, 2009; Mendes *et al.*, 2012). As such, some of the genes and enzymes involved in its biogenesis have been identified and characterized in recent years (Stadthagen *et al.*, 2007; Empadinhas *et al.*, 2008; Pereira *et al.*, 2008; Kaur *et al.*, 2009; Mendes *et al.*, 2011; Alarico *et al.*, 2014) (Figure 1.8).

The first committed step in MGLPs biosynthesis is the synthesis of GG, initially reported to be the primer for their biogenesis in *M. smegmatis* (Kamisango *et al.*, 1987). Many years later, the enzyme glucosyl-3-phosphoglycerate synthase (GpgS/*Rv1208*), catalyzing the synthesis of the phosphorylated intermediate glucosyl-3-phosphoglycerate (GPG) was characterized (Empadinhas *et al.*, 2008) and genetically confirmed to be linked to the initiation of MGLP synthesis (Kaur *et al.*, 2009). The subsequent step was found to encompass the dephosphorylation of GPG by a atypical glucosyl-3-phosphoglycerate phosphatase (GpgP/*Rv2419c*) yielding GG (Mendes *et al.*, 2011).

A GG hydrolase with an important role in the recovery from nitrogen stress was recently identified in rapidly-growing mycobacteria but its involvement in the regulation of MGLP biosynthesis is still hypothetical (Alarico *et al.*, 2014) . Although experimental confirmation is still missing, the predicted third biosynthetic enzyme likely catalyzes the transfer of an additional glucose unit to GG to yield glucosyl-α-(1→6)-glucosylglycerate (DGG). Based on genomic context analysis the gene predicted to coordinate this step is *Rv3031*, coding for a putative glycoside hydrolase of the GH-57 family of branching enzymes and a predicted α-(1→6)-branching enzyme (Stadthagen *et al.*, 2007). Moreover *Rv3031* is part of an operon (Figure 1.8) already demonstrated to be involved in MGLP biosynthesis, including its neighboring genes *Rv3030* and *Rv3032* for the 6-*O*-methyltransferase and glycosyltransferase for MGLP methylation and extension, respectively (Stadthagen *et al.*, 2007).

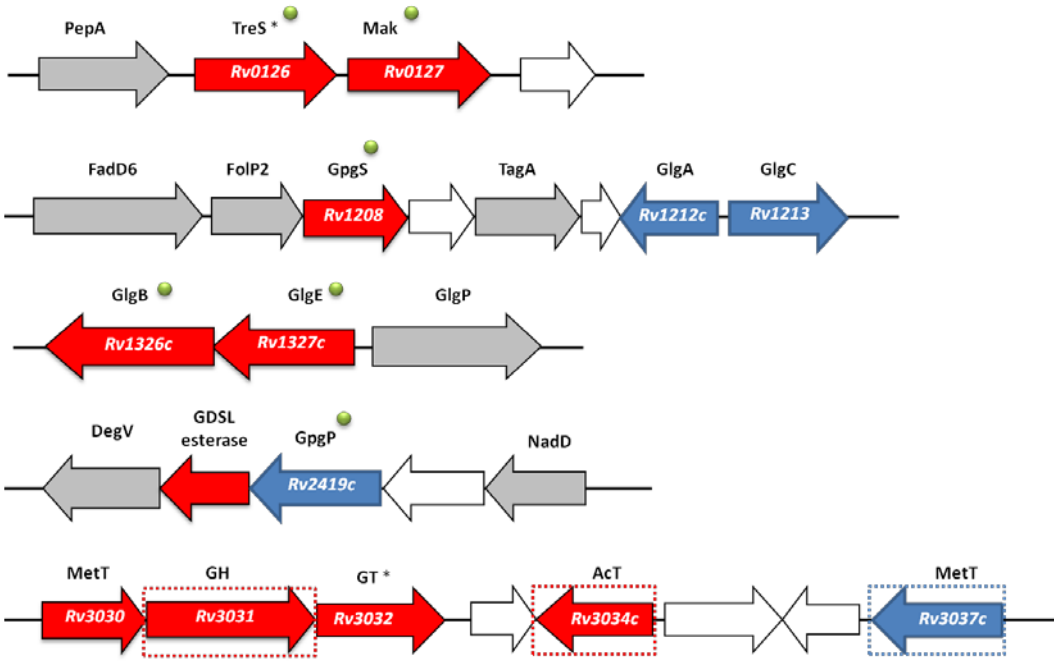


Figure 1.8 - *Mycobacterium tuberculosis* H37Rv gene clusters proposed to participate in MGLPs biosynthesis. Genes of unknown or hypothetical function are shaded white and grey, respectively. Genes considered essential for *M. tuberculosis* growth by saturation transposon mutagenesis (Griffin *et al.*, 2011b) are shaded red or blue if nonessential. *Although proposed to be essential for growth deletion of these genes produced viable mutants in *M. tuberculosis* H37Rv (Murphy *et al.*, 2005; Stadthagen *et al.*, 2007). Genes whose confirmation of involvement in MGLP pathway is still missing are boxed. Green spheres represent characterized proteins. PepA, serine protease; FadD6, long-chain-acyl-CoA synthetase; FolP2, dihydropteroate synthase; TagA, DNA-3-methyladenine glycosylase; NadD, nicotinate-nucleotide adenylyltransferase; GH, glucoside hydrolase.

Kamisango and colleagues proposed a model for MGLP biosynthesis wherein the elongation of the chain would proceed stepwise after DGG formation, from the reducing toward the non-reducing end, through sequential glucosylation and methylation reactions (Kamisango *et al.*, 1987). Gene disruption studies in *M. tuberculosis* revealed the involvement of *Rv3032* in glucosylation activity and that this gene is ultimately responsible for the elongation of the MGLP main chain (Stadthagen *et al.*, 2007). Following *Rv3032* disruption in *M. tuberculosis* a decrease in MGLP content was observed, with a pronounced stimulatory effect on MGLP production after overexpression of *Rv3032* in *M. smegmatis* which strongly suggests that Rv3032 is indeed the main α-(1→4)-glucosyltransferase committed to the elongation of these lipopolysaccharides (Stadthagen *et al.*, 2007). The effect of inactivating *Rv1212c* was not directly observed in MGLP biosynthesis leading to the conclusion that while both genes are involved in MGLP, glycogen and α-glucan synthesis, Rv3032 is mainly involved in MGLP synthesis whereas, in the absence of stress, Rv1212c is responsible for the synthesis of the other glucans (Stadthagen *et al.*, 2007; Sambou *et al.*, 2008). Interestingly, *M. leprae*, which produces MGLPs and a capsule but not glycogen (Berg *et al.*, 2007) carries an orthologue of *Rv3032* in its genome although the *Rv1212c* orthologue is a pseudogene. Conversely, a frameshift mutation at the 5' end of the *Rv3032* homologue in *M. smegmatis* apparently abolished the activity of this gene product. However, the genome carries an orthologue of *Rv1212c*, which most likely accounts for the production of MGLPs in this species (Jackson and Brennan, 2009). The gene responsible for the

methyltransferase activity was identified based on a similar genetic approach. The disruption of the *Rv3030* orthologue (MSMEG_2349) in *M. smegmatis*, resulted in a phenotype with decreased MGLP levels, confirming its involvement in MGLP biosynthesis (Stadthagen *et al.*, 2007). However a genetic context analysis revealed the existence of another methyltransferase (*Rv3037c*) which could explain the residual MGLP content found. Although functional confirmation for this hypothesis is missing this gene was proposed to encode a putative SAM-dependent 3-*O*-methyltransferase and could be responsible for the 3-*O*-methylation of the MGLP non-reducing end, or have interchangeable activity with Rv3030 (Stadthagen *et al.*, 2007; Jackson and Brennan, 2009) (Figure 1.9).

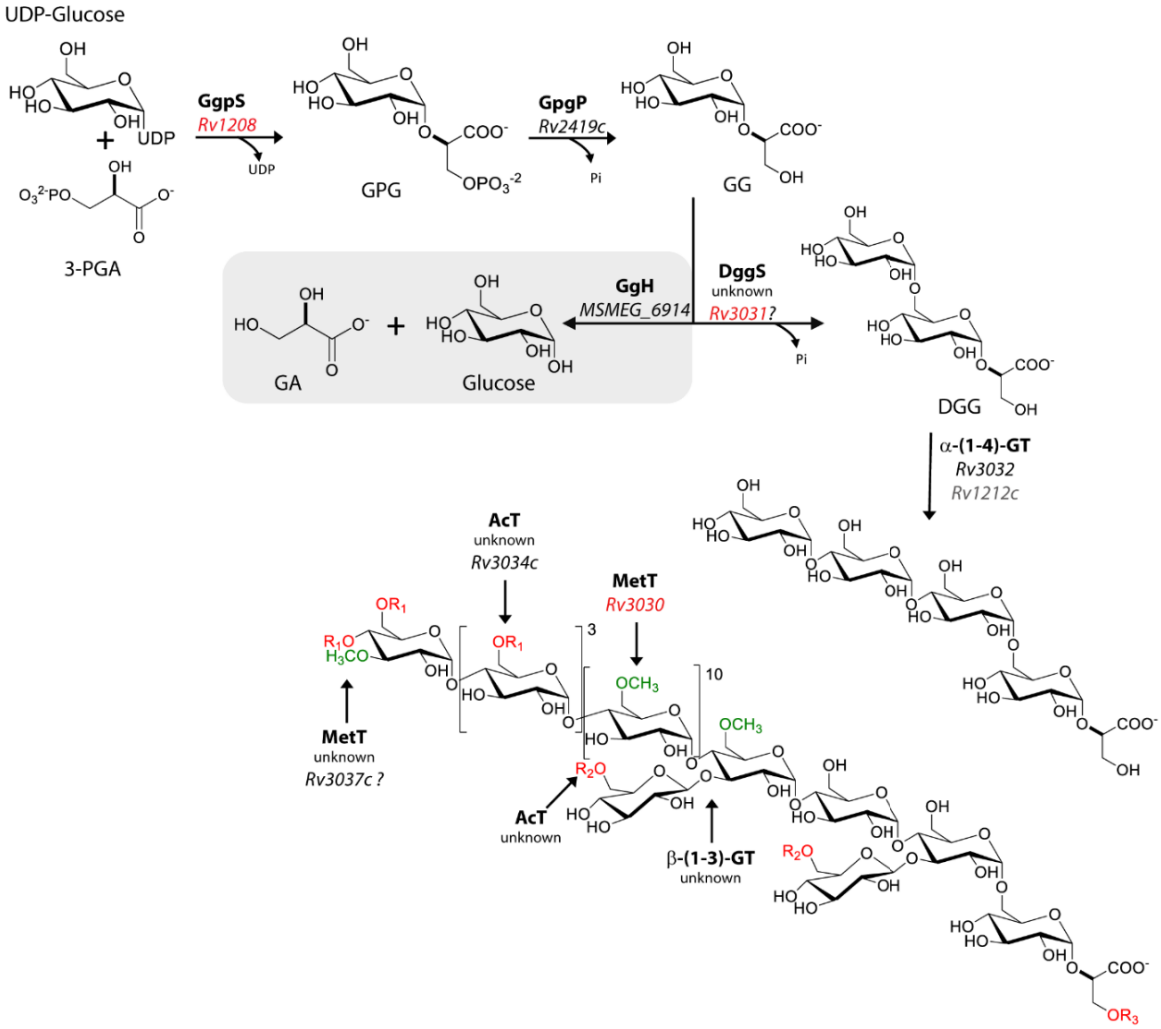


Figure 1.9 - Schematic representation of the proposed biosynthetic pathway for MGLP. Highlighted in red are genes proposed to be essential for *M. tuberculosis* growth (Griffin *et al.*, 2011b). Question marks represent tentative assignment of gene function. The enzyme GgH possible regulatory function (only present in rapidly-growing mycobacteria) is boxed in grey (Alarico *et al.*, 2014). The postulated compensatory function of the *Rv1212c* gene is highlighted in grey (Stadthagen *et al.*, 2007; Sambou *et al.*, 2008). R_1, short chain acyl groups (acetate, propionate or isobutyrate); R_2, succinate. R_3 represents octanoate and its position in the mature MGLP is represented according to (Smith and Ballou, 1973; Dell and Ballou, 1983). β-(1→3) glycosylations, methyltransferase and acyltransferase reactions occur at unknown stages during the biosynthetic pathway. Adapted from (Mendes *et al.*, 2012).

The model proposed by Kamisango where glycosylation and methylation proceed sequentially is so far believed to be correct (Kamisango *et al.*, 1987). Methylation was shown to be dependent on acylation levels in *M. phlei* extracts where, unless the glucooligosaccharide acceptors were partially acetylated, no methylation was observed (Grellert and Ballou, 1972). Also methylation at C-3 of the terminal glucose is apparently dependent on high levels of acetylation since at low levels of acetylation (0.2 acetyl groups per glucose), methylation occurred only at position 6 of the glucose. At high levels of acetylation (1.6 acetyl groups per glucose), methylation at position 6 was strongly depressed and occurred instead at position 3. It was then suggested that acylation and methylation occurred together, with the former process controlling the latter (Grellert and Ballou, 1972).

Early studies on MGLP acylation preformed with an enzyme preparation from *M. phlei* cell extracts showed acyltransferase activity transferring acyl groups from acyl-CoA to the lipopolysaccharide, as well as to similar artificial acceptors, but displaying a preference for hydrophobic acceptors like the methylated polysaccharide chain. (Tung and Ballou, 1973). The enzyme was capable of transferring acetate, propionate and isobutyrate, with comparable efficiencies, which suggested that a single enzyme would be involved in non-reducing end acylation. However, very low transferase activity was detected with succinyl and octanoyl which in the authors' opinion was consistent with these groups being in a different position, closer to the reducing end, and was indicative of the involvement of more than one acyltransferase (Tung and Ballou, 1973). Nonetheless, the gene or genes responsible for acylation are still unknown. Based on genetic context analysis the putative acetyltransferase gene *Rv3034c*, belonging to the cluster of *Rv3030* and *Rv3032,* seemed to be a good candidate (Stadthagen *et al.*, 2007) but experimental evidence to support this theory is missing. (Figure 1.9). The enzyme or enzymes responsible for introduction of β-(1→3)-ramifications into the MGLP second and fourth glucose residues are so far unknown, as is the timing at which the reactions occur.

The termination reaction is supposed to involve a similar mechanism to that of MMP. When the chain reaches a sufficient length it will readily associate with a long-chain fatty acid and form an inclusion complex in which the polymer assumes a tightly coiled conformation conformation, which was also postulated to inhibit further 6-*O*-methylation and promote 3-*O*-methylation (Kamisango *et al.*, 1987) thus favoring the termination reaction.

Objectives

The mycobacterial polysaccharide MGLP was discovered over half-century ago but its biosynthetic pathway has only recently begun to be unraveled showing that many of the genes involved have been proposed to be essential for *M. tuberculosis* growth. A novel four-step pathway (GlgE pathway), diverting maltose units from trehalose into glycogen, has been recently identified and presumed to represent a parallel pathway for polymerization of the seemingly vital mycobacterial MGLP, a molecule with a glycogen-like arrangement of glucose units in the main chain. To contribute to the quest for potential targets aiming at the discovery of new drugs for TB and other mycobacterial infections, we sought characterization of enzymes with anticipated involvement in the MGLP pathway and simultaneously proposed to be essential for *M. tuberculosis* growth. *Rv0127* and *Rv2418c* were selected for the study as a consequence of functional and genomic context surveys. The former was responsible for the synthesis of the critical metabolite maltose-1-phosphate, the supplier of maltose blocks to α-(1→4) polymers extension, namely those dependent on the GlgE pathway, and the latter was located in an operon containing a novel glucosyl-3-phosphoglycerate phosphatase responsible for the production of the MGLP primer, glucosylglycerate. Three layers of investigation were initially projected to cover the disclosure of gene identification and functional characterization, from chemical synthesis of intermediates to the inspection of molecular details arising from three-dimensional structures that would lay the foundations for ensuing drug discovery efforts. The intersection of microbiology, enzymology, synthetic chemistry and crystallography was established from the early stage of this project and successfully implemented in the development of the work in this thesis.

Chapter 2 Structural and biochemical characterization of *Mycobacterium vanbaalenii* maltokinase

This chapter is published in:

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

*These authors contributed equally to the work

Chapter 2- table of contents

Abstract

A novel four-step pathway identified recently in mycobacteria channels trehalose to glycogen synthesis and is also presumably involved in the biosynthesis of two other crucial polymers: intracellular methylglucose lipopolysaccharides (MGLPs) and exposed capsular glucan. The three-dimensional structures of three of the intervening enzymes - GlgB, GlgE, and TreS - were recently reported, providing the first templates for rational drug design. Here we describe the biochemichal and structural characterization of the fourth enzyme of the pathway, mycobacterial maltokinase (Mak), uncovering a new family of eukaryotic-like kinases (ELK). Remarkably, this new family of ELKs has a novel N-terminal domain which by interfacing with and restraining the mobility of the phosphate-binding region of the N-terminal lobe might regulate its phosphotransfer activity and also possibly represent a likely anchoring point for TreS, the upstream enzyme in the pathway. By completing the gallery of atomic-detail models of an essential pathway, this structure opens new avenues for the rational design of alternative anti-tubercular compounds.

Introduction

Glycogen is a central energy storage molecule in bacteria and the metabolic pathways associated with its biosynthesis and degradation are crucial for maintaining cellular energy homeostasis. The classical pathway for glycogen synthesis involves the enzymes GlgC-GlgA-GlgB (Sambou *et al.*, 2008; Chandra *et al.*, 2011). However, a novel four-step α-glucan biosynthetic pathway, the GlgE pathway, has been recently identified (Kalscheuer *et al.*, 2010a; Elbein *et al.*, 2010). In mycobacteria, the GlgE pathway involves the combined action of trehalose synthase (TreS), which produces maltose, then phosphorylated by maltokinase (Mak) into maltose-1-phosphate (M1P), and maltosyltransferase (GlgE) (Elbein *et al.*, 2010; Kalscheuer and Jacobs, 2010; Miah *et al.*, 2013) for the synthesis of linear α-(1→4)-glucans, which are substrates for the α-(1→6)-branching enzyme GlgB ultimately producing glycogen(Garg *et al.*, 2007). Although these four enzymes have been proposed to be essential in *Mycobacterium tuberculosis* by saturation mutagenesis experiments, pointing to critical roles in the pathogen, targeted disruption of the *mak* gene remains to be performed (Griffin *et al.*, 2011b; Murphy *et al.*, 2005). Together with TreY/TreZ, this pathway forms a cycle in which trehalose is recycled into and from glycogen (Chandra *et al.*, 2011).

While M1P was originally identified in *Mycobacterium bovis* BCG cell extracts (Narumi and Tsumita, 1967), the enzyme responsible for its biosynthesis, Mak (EC 2.7.1.175), was only discovered thirty years later in members of the related genus *Actinoplanes* (Drepper *et al.*, 1996; Niehues *et al.*, 2003). It has been also suggested that M1P could serve as substrate for the elongation of the important glycogen-related capsular α-glucan and intracellular methylglucose lipopolysaccharides, hypotheses still lacking experimental support (Sambou *et al.*, 2008;

Kalscheuer *et al.*, 2010a; Mendes *et al.*, 2012). The GlgE-dependent α-glucan biosynthetic pathway was genetically validated as essential and its potential as drug target discussed extensively (Kalscheuer *et al.*, 2010a; Kalscheuer and Jacobs, 2010). Novel inhibitors mimicking the GlgE substrate, M1P, were recently developed (Veleti *et al.*, 2014). Remarkably, GlgE inhibition elicits the build-up of M1P, shown to be toxic for mycobacteria (Kalscheuer *et al.*, 2010a). It has been proposed that metabolite flux through the GlgE pathway is strongly influenced by the ATP-driven M1P synthesis by Mak (Kalscheuer *et al.*, 2010a; Syson *et al.*, 2011), whose encoding gene (*Rv0127*) in *M. tuberculosis* H37Rv was also proposed to be essential for growth (Griffin *et al.*, 2011b).

The structural, biochemical and mechanistic analysis of the glycogen synthesis-associated GlgE pathway enzymes is therefore instrumental for designing new molecules with potential application in anti-tuberculosis therapies. In fact, with the notable exception of Mak, the three-dimensional structures of all other enzymes in this pathway have been determined recently (Syson *et al.*, 2011; Caner *et al.*, 2013; Pal *et al.*, 2010; Roy *et al.*, 2013; Syson *et al.*, 2014). The three dimensional structures of *M. smegmatis* and *M. tuberculosis* TreS revealed a C-terminal carbohydrate-binding domain, which was proposed to be relevant for glycogen recognition and to provide anchorage of this enzyme to the site of glycogen polymerization (Caner *et al.*, 2013; Roy *et al.*, 2013). Structural and biochemical data also suggested that TreS and Mak form a hetero-octameric complex, enhancing Mak catalytic activity in M1P biosynthesis (Roy *et al.*, 2013), possibly by favoring substrate channeling. This finding is in good agreement with the identification of numerous Mak orthologues in members of the actinobacteria and in distantly related phyla, where they frequently occur as trehalose synthase/maltokinase (TreS-Mak) bifunctional units (Chandra *et al.*, 2011; Mendes *et al.*, 2012).

The maltokinase from *Mycobacterium bovis* BCG, identical to the *M. tuberculosis* orthologue (Mak^{Mtb}), was characterized in detail, identifying ATP as the preferential phosphate donor and the requirement of magnesium ions (Syson *et al.*, 2011) for maximal enzyme activity(Mendes *et al.*, 2012). However, given the instability of Mak^{Mtb}, we have expressed and purified the orthologue from *M. vanbaalenii* (Mak^{Mvan}) for biochemical characterization and solved the three-dimensional structures of its free and non-hydrolysable ATP-bound forms. The structure of Mak^{Mvan}, which shares 59% amino acid sequence identity with Mak^{Mtb}, revealed a typical bilobal eukaryotic protein kinase-like (ELK) fold, conserving the main structural motifs required for the phosphotransfer reaction. Strikingly, Mak^{Mvan} displays a novel N-terminal domain, unique to maltokinases and conserved in the bifunctional TreS-Mak proteins, and without sequence similarity to other known proteins. This novel domain, topologically similar to protease inhibitors of the cystatin family (Turk *et al.*, 2008), is proposed to act as an anchoring point tethering maltokinase and trehalose isomerase activities to the site of glycogen biosynthesis, ensuring correct regulation of Mak activity and possibly preventing excessive accumulation of M1P. The maltokinase structures described here also provide the first structural insight into a subfamily of ELKs, as well as the framework for the discovery of new antimycobacterial drugs, contributing towards better solutions to one of the most insidious and re-emerging infectious diseases in the world.

Methods

Sequence analysis, strains and culture conditions

Mak sequences were obtained from the NCBI database using the *M. tuberculosis* sequence (GenPept accession NP_214641.1) as template. Genome context analysis were performed at PATRIC (http:// patricbrc.org/portal/portal/patric) and KEGG (http://www.genome.jp/kegg) databases. Sequence alignments were generated with T-COFFEE (http://tcoffee.grg. cat) and manually curated in MEGA6 (Tamura *et al.*, 2013).
Mycobacterium vanbaalenii PYR-1 (DSM 7251), obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (Germany), were cultivated in agar plates for 5 days at 35°C in glycerol-based medium at pH 7.0 (20 g/L glycerol, 5 g/L casaminoacids (Difco), 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).

Identification and cloning of the maltokinase gene

Sequence similarity searches with the amino acid sequence of *M. bovis* BCG maltokinase against the NCBI database allowed identification of the *M. vanbaalenii* maltokinase (Mak^{Mvan}) sequence. The corresponding *mak* gene (Mvan_5735) was amplified by PCR using KOD hot start DNA polymerase (Novagen) from *M. vanbaalenii* chromosomal DNA using the MakF and MakR primers (Table 2-I) and cloned between the *Nde* I and *Hind* III sites of pET30a (Novagen).

Site-directed mutagenesis of *M. tuberculosis* and *M. vanballenii* Mak

The N145A and S144A mutants were generated with a two-round megaprimer PCR-based approach as previously described (Empadinhas *et al.*, 2011). The first amplification round was performed with primers A/B1 or A/B2 or C1/D or C2/D (Table 2-I) using the wild-type *M. bovis* BCG/*M. tuberculosis mak* as template for insertion of each independent mutation. The purified amplification products were used as templates in the second round of PCR with primers A/D (Table 2-I). The final products carrying the desired mutations were cloned into pET30a, sequenced (LGC Genomics) and transformed into *E. coli* BL21 for expression. The D339N and R351A mutants were generated by inserting a synthetic 751-base pair fragment (GenScript) carrying the desired mutation into the natural *Bsi* WI and *Hind* III restriction sites of the pET30a-based construct carrying wild-type *M. tuberculosis* mak. Site-directed mutagenesis with the QuikChange II Site-Directed Mutagenesis Kit (Agilent) was used to generate Mak *M. vanbaalenii* mutants K413A, Y416A, Y416F, Y420A and Y420F. Oligonucleotides used as PCR primers are listed in Table 2-I.

Table 2-I - Oligonucleotides used as PCR primers

Code	Name	Sequence(5'-3')
MakF	MakF	ATTGATCAACATATGACGCTGGCATTCGGCGATTG
MakR	MakR	ATCAAGCTTGCCCAGGATGAGGCTGATCGATC
A	WT_NdeF	CTTACATATGACTCGGTCGGACACGC
B1	S144A_F	GACGCCGAACAGGCCAACACCAGTG
C1	S144A_R	CACTGGTGTTGGCCTGTTCGGCGTC
D	WT_HindR	ATTAAGCTTGCTAGCGGTCAGGCGGG
B2	N145A_F	GACGCCGAACAGAGCGCCACCAGTG
C2	N145A_R	CACTGGTGGCGCTCTGTTCGGCGTC
	K413A_F	GCCTACGAGCTCGACGCGGCGGTGTACGAAGC
	K413A_R	GCTTCGTACACCGCCGCGTCGAGCTCGTAGGC
	Y416A_F	CTCGACAAGGCGGTGGCCGAAGCCGCTTACGA
	Y416A_R	TCGTAAGCGGCTTCGGCCACCGCCTTGTCGAG
	Y416F_F	CTCGACAAGGCGGTGTTCGAAGCCGCTTA
	Y416F_R	TAAGCGGCTTCGAACACCGCCTTGTCGAG
	Y420A_F	GGTGTACGAAGCCGCTGCCGAGGCCCGTTTCC
	Y420A_R	GGAAACGGGCCTCGGCAGCGGCTTCGTACACC
	Y420F_F	TGTACGAAGCCGCTTTCGAGGCCCGTTTC
	Y420F_R	GAAACGGGCCTCGAAAGCGGCTTCGTACA

Nde I and *Hind* III restriction sites added to the primers are underlined. Mutated codons are boxed in grey

Protein production and purification

Recombinant wild-type or the mutants of Mak^{Mvan} and Mak^{Mtb} were overexpressed as previously described (Mendes *et al.*, 2010). Recombinant proteins were purified in a HisPrep FF 16/10 column (GE Healthcare) equilibrated with 20 mM sodium phosphate pH 7.4, 0.5 M NaCl, 20 mM imidazole and eluted with 200 mM imidazole in the same buffer. The purity of the fractions was evaluated by SDS-PAGE. The purest active fractions were pooled, diluted 10-fold with 20 mM Bis-Tris propane buffer (BTP) pH 7.4, loaded onto a 6 mL Resource Q column (GE Healthcare) equilibrated with the same buffer and eluted with a NaCl linear gradient (0–500 mM). Mak-containing fractions with highest activity and purity, as assessed by SDS-PAGE and enzymatic assays, were pooled and concentrated in 30 kDa cut-off ultrafiltration devices (Amicon) with concomitant buffer exchange for 50 mM BTP pH 7.4, 50 mM NaCl. The selected fractions correspond to monomeric enzyme as judged by size exclusion chromatography on a Superdex 12 10/300 GL column (GE Healthcare). Mak activity was monitored during purification by thin-layer chromatography (TLC) using as standard maltose-1-phosphate (M1P) synthesized with recombinant Mak^{Mvan} and purified with methods described elsewhere (provisional patent application). Reaction mixtures (50 mL) containing 25 mL cell-free extract or 15 mL of each chromatography fraction in 50 mM BTP pH 8.0, 3 mM ATP, 5 mM maltose, 10 mM $MgCl_2$ were incubated at 37°C for 15 min prior to separation by TLC, developed with acetic acid/ethyl

acetate/water/25% ammonia (6:6:2:1) and stained with α-naphthol, followed by charring at 120°C (Jacin and Mishkin, 1965). Cell-free extracts from *E. coli* BL21 carrying an empty pET30a vector were used as negative controls.

Characterization of Mak^{Mvan} and Mak^{Mtb} mutants

Biochemical and kinetic data of Mak^{Mvan} and Mak^{Mtb} mutants were determined as previously described (Mendes *et al.*, 2010). Temperature and pH profiles, effect of cations and substrate specificity were determined by addition of 0.25 µg Mak^{Mvan} to 50 µL mixtures containing the appropriate buffer, 5 mM maltose, 3 mM NTP, 5–15 mM divalent cation. Reactions were stopped by cooling on an ethanol-ice bath (-10°C), followed by Mak inactivation with 5 µL of 5 N HCl and neutralization with 5 µL of 5 N NaOH. Controls were performed to account for NTP degradation following acid treatment. The amount of NDP released was determined by measuring the absorption at 340 nm after incubation of the sample with 3 U each of pyruvate kinase and lactate dehydrogenase, 0.3 mM NADH, 2.5 mM phosphoenolpyruvate for 10 min at 30°C (Ornston and Ornston, 1969). ATP, CTP, GTP, TTP and UTP were tested as phosphate donors, glucose, trehalose, maltose, maltotriose, and maltotetraose as sugar substrate acceptors, and M1P was also tested as substrate for the Mak reverse reaction. As described above M1P was synthesized with recombinant Mak^{Mvan} and purified with methods described elsewhere (provisional patent application). The temperature profile was determined between 20 and 65°C in 50 mM BTP pH 7.5, 10 mM $MgCl_2$. The effect of pH was determined at 30°C in 50 mM BTP (pH 6.0–9.0) and in 50 mM CAPS (pH 9.0–11.0), with 10 mM $MgCl_2$. The effect of cations was examined by incubating the reaction mixture with the appropriate substrates and the chloride salts of Mg^{2+}, Mn^{2+} or Co^{2+} at 30°C. Kinetic parameters for *M. vanbaalenii* Mak were calculated with Prism 5 (GraphPad). The K_m values for ATP, GTP, and maltose were determined at 30°C and pH 7.5. K_m and V_{max} values for Mak^{Mtb} mutants were determined for maltose and ATP at 37°C and pH 8.0. All experiments were performed in triplicate. The specific activity of Mak^{Mvan} mutants was determined at 30°C by measuring the amount of ADP released as described above and expressed as percentage of wild-type Mak^{Mvan} maximal activity. Reactions were carried out in mixtures containing 0.25 µg protein, 5 mM ATP and 20 mM maltose in 20 mM BTP pH 8.0, 10 mM $MgCl_2$.

Crystallization

All crystallization procedures and structure determination and analysis are described in:
Fraga, J., Maranha, A., Mendes, V., Pereira, P.J., Empadinhas, N., and Macedo-Ribeiro, S. (2015). *Structure of mycobacterial maltokinase, the missing link in the essential GlgE-pathway*. Scientific Reports **5**: p. 8026

Results

Biochemical characterization and kinetic parameters

After the kinetic characterization of Mak^{Mtb} attempts were made towards characterizing the molecular details underlying maltose phosphorylation in mycobacteria and the protein was screened for crystallization. However, given the instability of Mak^{Mtb} and the unsuccessful attempts at its crystallization the more stable orthologue from *M. vanbaalenii,* which is 59% identical to Mak^{Mtb}, was selected for biochemical and structural characterization. We have expressed, purified and characterized this protein and solved the three-dimensional structures of its free and non-hydrolysable ATP-bound forms.

The Mak^{Mvan} enzyme was active between 20 and 65°C, with maximal activity at 60°C (Figure 2.1 A). At 37°C, Mak^{Mvan} was active between pH 6 and 11 with optimum between pH 8.5 and 10 (Figure 2.1 B). The enzyme was strictly dependent on divalent cations and Mg^{2+} (10 mM) had the most pronounced stimulatory effect, being inhibitory at higher concentrations (Figure 2.1 C). Other divalent cations like Mn^{2+} and Co^{2+} stimulated Mak activity to a lower extent but never surpassing 25% of maximal activity (Figure 2.1 C). The recombinant Mak used ATP and GTP (39% of maximal activity) as phosphate donors with decreasing efficiency while UTP was a poor donor (3% of maximal activity). Mak^{Mvan} was unable to use TTP or CTP as phosphate donors to phosphorylate maltose (Figure 2.1 D).

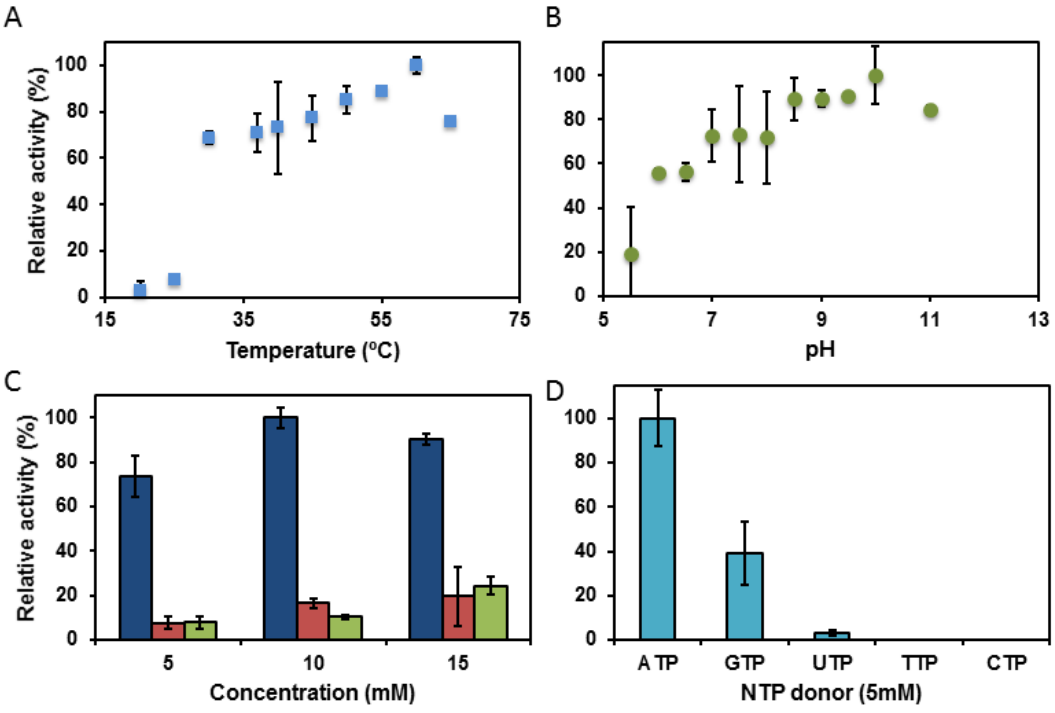


Figure 2.1 - Biochemical characterization of Mak from *M. vanbaalenii*. (A) Influence of temperature and (B) buffer pH and on Mak^{Mvan} enzymatic activity. (C) Effect of divalent cations on enzyme activity (Mg^{2+}, blue bars; Co^{2+}, red; Mn^{2+}, green). (D) Specificity of Mak^{Mvan} for different phosphate donor substrates. Plotted data are the mean of three independent experiments.

Kinetic parameters showed that Mak exhibited Michaelis-Menten kinetics at 37°C with ATP and GTP up to 5 mM since higher concentrations of these donors were progressively inhibitory. Maximum activity was achieved with 5 mM ATP, dropping to 85% and 55% in the presence of 10 mM and 20 mM ATP respectively, with GTP exhibiting a similar inhibitory effect on Mak^{Mvan} activity. The activity with UTP was negligible, unlike the Mak from *M. bovis* BCG/*M. tuberculosis* that could use GTP and UTP with comparable efficiency (Mendes *et al.*, 2010). Mak^{Mvan} displayed maltokinase activity and showed over 10-fold higher catalytic efficiency than Mak^{Mtb} (Table 2-II). Maltose was the only sugar acceptor used by this enzyme and upon incubation with pure maltose-1-phosphate (6 or 18 mM) no reverse reaction was detected, even in the presence of ADP as a putative acceptor (data not shown).

Table 2-II - Comparison of kinetic parameters of recombinant Mak^{Mvan} and Mak^{Mtb}

	Enzyme	Maltose	ATP	GTP	UTP
		(with 5mM ATP)	(with 20 mM maltose)		
K_m (mM)	Mak ^{Mtb*}	2.5±0.4	0.7±0.1	1.0±0.2	1.3±0.1
	Mak ^{Mvan}	7.3±0.9	2.6±0.5	5.8±1.4	ND
V_{max} ($\mu mol.min^{-1}.mg^{-1}$)	Mak ^{Mtb*}	21±1	21±1	19±1	7.1±0.3
	Mak ^{Mvan}	301±14	209±19	62±8	ND
k_{cat} (min^{-1})	Mak ^{Mtb*}	52±3	54±2	ND	ND
	Mak ^{Mvan}	3006±139	2091±188	616±78	ND
k_{cat}/K_m ($mM^{-1}.min^{-1}$)	Mak ^{Mtb*}	21±4	72±12	ND	ND
	Mak ^{Mvan}	414±54	810±163	105±28	ND

*Values from the identical Mak from *M. bovis* BCG (Mendes *et al.*, 2010). ND - Not determined.

Overall architecture

Mak^{Mvan} crystallized in the orthorhombic space group $P222_1$ and diffracted to 1.47 A˚ resolution. The protein displays an elongated concave-shaped structure that can be divided into two half-lobes with a central narrow acidic channel (Figure 2.2 A). The N-terminal lobe can be divided into two subdomains: (i) a cap N-terminal subdomain comprising the first 88 amino acids and composed by a curved β-sheet that encloses the N-terminal αA*-helix and a short two-stranded β-sheet (ii) and an intermediate subdomain (residues 89–200) composed of an anti-parallel β-sheet flanked by two helices. A nine-residue linker (residues 201–209) containing a short β-strand connects the intermediate subdomain and the C-terminal lobe, which is mostly α-helical (Figure 2.2 B and C).

The cap N-terminal subdomain is a novel N-terminal domain, unique to maltokinases and conserved both in these proteins and in the bifunctional TreS-Mak proteins, and without sequence similarity to other known proteins. A structural search with only the first 88 amino acid residues of Mak^{Mvan}, corresponding to this unique N-terminal cap subdomain, unveiled an

unforeseen topologically resemblance with proteins displaying the cystatin fold and a remote similarity with the N-terminal domain of the serine/threonine protein kinase.

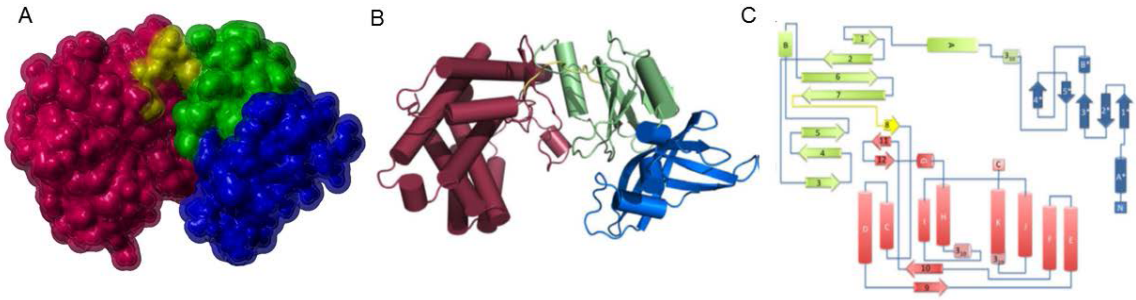


Figure 2.2 - Overall structure of Mak^{Mvan}. (A) Surface representation of Mak^{Mvan} showing the distinct domains in different colors (cap domain: blue; intermediate domain: green; C-terminal domain: pink; linker: yellow). (B) Cartoon representation of Mak^{Mvan} three-dimensional structure, color-coded as in panel A. (C) Topology diagram of Mak^{Mvan}, color-coded as in panel A. Figure in panel A was prepared with YASARA view. Figure in panel B was prepared with PyMOL (http://www.pymol.org). Panels B and C were prepared by Sandra de Macedo-Ribeiro and Pedro J. B. Pereira at Instituto de Biologia Molecular e Celular (IBMC, Porto) and adapted from from Fraga *et al.* (2015) with permission.

The two N-terminal subdomains (cap and intermediate domain) are tightly interconnected through a wide interface area, predominantly by polar contacts and a significant number of salt bridges, some of them involving the side chains of conserved amino acid residues. On both edges of the interface, strong polar contacts are observed that staple these two regions and limit the interdomain flexibility. One such interaction occurs between the invariant Arg13 in the cap subdomain which forms polar contacts with the main chain carbonyls of Glu134 and Asn137 of the intermediate subdomain, while the conserved Trp14 stacks with Arg152 (strictly conserved), whose side chain is involved in a salt bridge with Asp88 (strictly conserved) (Figure 2.3). The extent and strength of the interactions between the two subdomains indicates that their relative motion is highly interdependent. Analysis of the B-factor distribution reveals that the cap subdomain has the highest values for the atomic temperature factors and the C-terminal domain the lowest, suggesting that the N-terminal lobe, in particular the cap region, is more dynamic. Mak^{Mvan} revealed a typical bilobal eukaryotic protein kinase-like (ELK) fold, conserving the main structural motifs required for the phosphotransfer reaction. Many of the canonical structural motifs associated with nucleotide binding and enzymatic activity are conserved in Mak^{Mvan}. Particularly, the catalytic loop, that contains the putative catalytic base (Asp305), and the ^{322} DFE ^{324} motif with its magnesium-binding Asp322 (Figure 2.3). This latter pattern is reminiscent of the DXD motif commonly found in glycosyltransferases that, together with the nucleotide sugar donor, coordinates the active site divalent ion (Pereira *et al.*, 2008). The P-loop (^{133} AEQSNTSV ^{140} nucleotide-positioning loop) is the structural equivalent of the glycine-rich G-loop found in eukaryotic protein kinases that typically contains the residues interacting with the terminal γ-phosphate of the bound nucleotide (Scheeff and Bourne, 2005; Grant *et al.*, 1998). In Mak^{Mvan} the region is devoid of glycine residues, being also shorter than the structurally equivalent in ELKs.

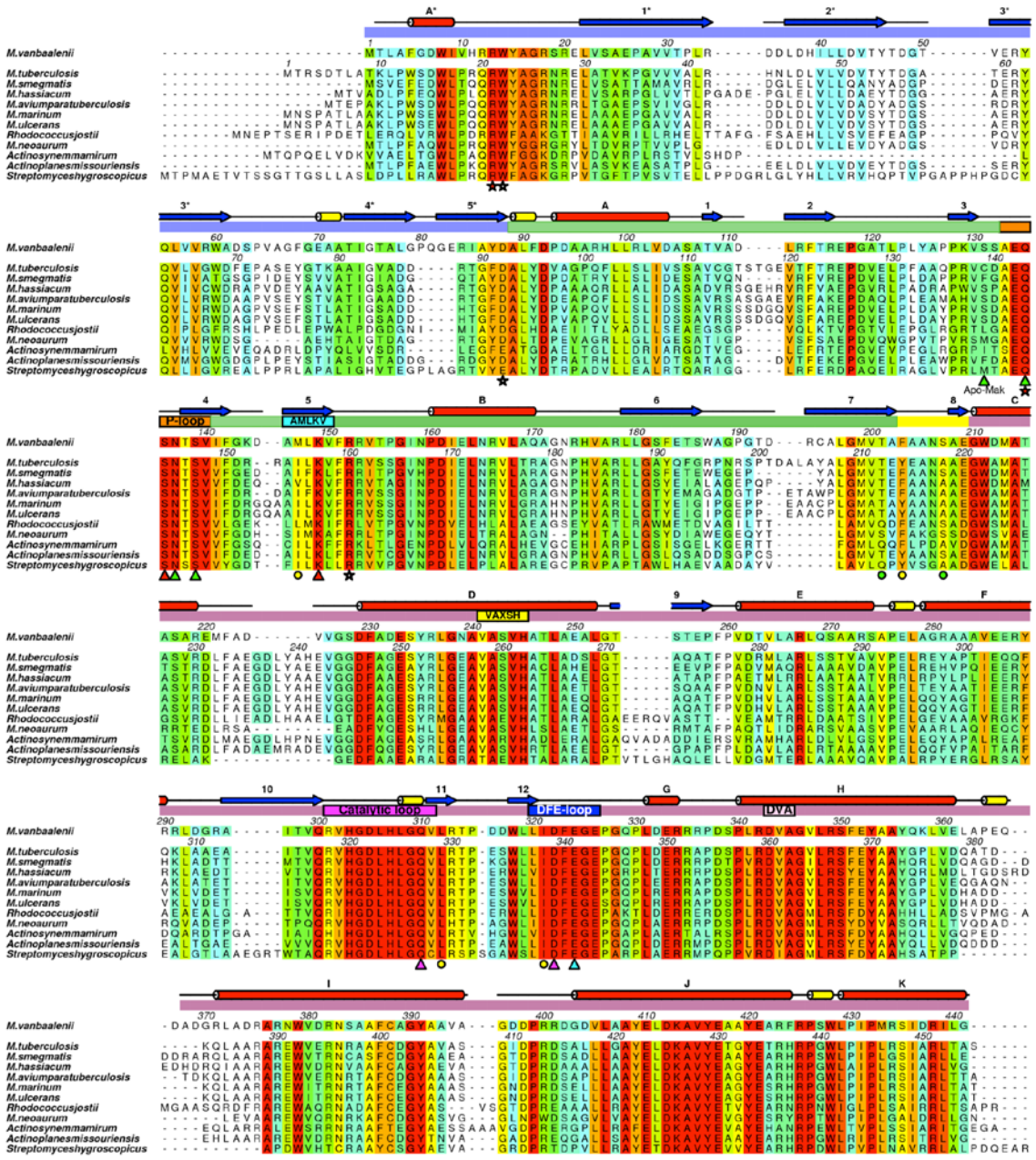


Figure 2.3 - Amino acid sequence alignment of representative maltokinases from actinobacteria. Residues are colored according to conservation (red: identical residues, orange to blue: decreasing conservation; white: dissimilar residues). Amino acid numbers are given above the *M. vanbaalenii* and *M. tuberculosis* sequences. Secondary structure elements are based on the Mak^{Mvan} 3D structure (red cylinders, α-helices; yellow cylinders, 3_{10}-helices; blue arrows, β-sheets). Colored lines above the alignment indicate the domains identified in the 3D structure (blue: cap subdomain; green: intermediate domain; pink: C-terminal domain) and the main structural motifs associated with phosphotransfer activity are highlighted. Residues involved in interactions with the nucleotide are marked below the alignment (stars: residues involved in polar contacts at the interface between the cap and the intermediate domain; yellow circles: residues involved in Van der Waals interactions with the adenine base; green circles: residues involved in water-mediated interactions with the adenine base; red triangles: residues involved in direct hydrogen bonds with the γ-phosphoryl group; green triangles: residues involved in water-mediated polar contacts with the phosphoryl groups; blue triangles: residues directly stabilizing magnesium-coordinating water molecules; magenta triangles: residues involved in magnesium ion coordination). Figure prepared with Aline2 by Sandra de Macedo-Ribeiro at Instituto de Biologia Molecular e Celular (IBMC, Porto) and adapted from Fraga *et al.* (2015) with permission.

Nucleotide binding site

The structure of Mak^{Mvan} in complex with the non-hydrolysable ATP analog, adenosine-5'- [(β,γ)-methyleno] triphosphate (AppCp), was determined 1.15 A˚ resolution. The nucleotide binds with two alternative conformations within the deep pocket between the intermediate subdomain and the C-terminal domain.

Only localized structural changes occur to accommodate the AppCp molecule, particularly in the linker between the intermediate and the C-terminal domains and its close neighboring regions. The most significant conformational changes occur in the region preceding α-helix G and in the loop connecting strands β3 and β4 in the intermediate domain, in order for the residues to establish contact with the γ-phosphate group. On the C-terminal neighboring side, the side chain of Asp322 shifts to coordinate the magnesium ions. The γ-phosphate group is then stabilized by a direct interaction with the P-loop residue Ser136 and further stabilized through water-mediated contacts to Asn137 and Lys149.

In the Mak^{Mvan}:AppCp complex there are two magnesium binding sites and while the Glu324 residue (DFE motif) does not participate in magnesium coordination, it shifts to interact with the second magnesium ion, coordinating a water molecule in the Mak^{Mvan}:AppCp complex structure. Overall, there is conservation of the residues within the nucleotide-binding site in actinobacteria, in particular those establishing either direct or water-mediated contacts with the phosphate groups on both sides of the binding cleft.

Maltose binding site - Mutational analysis of *M. vanbaalenii* Mak

Despite repeated attempts no diffracting crystals of Mak^{Mvan} in complex with its sugar substrate could be obtained. Therefore, in order to identify the maltose binding site, a search for suitable cavities was performed with fpocket (Le Guilloux *et al.*, 2009)

In the structure of free Mak^{Mvan} the largest pocket identified (pocket 1 blue in Figure 2.4) is located to the left of the exposed catalytic cleft. This pocket is predominantly hydrophobic with an overall positive charge and it is delineated by the invariant residues Trp211, Tyr353, Arg425, Tyr420, Lys413, Tyr416 within the C-terminal lobe, as well as by residues from the catalytic loop region such as His307 and Asp305 and from the DFE motif (Figure 2.4 B). At the bottom of the pocket, two positively charged residues, Arg267 and Arg342 (stabilized by a salt bridge with Glu326) probably play a role in directing the negatively charged reaction product away from the enzyme active site. Pocket 1 is close to the active site HGD motif and includes Asp305, which is topologically equivalent to the catalytic base in analogous enzymes, and globally coincident with the substrate-binding region in structurally related ELKs.

Two additional large pockets could be identified in the vicinity of the enzyme's active site. One of them corresponds to the nucleotide-binding site (pocket 3, Figure 2.4). The other pocket (pocket 2, Figure 2.4), also large enough for accommodating maltose, encompasses the γ-phosphate binding region and is framed by residues from the loop linking strand β5 and helix αB from the P-loop and from the DFE motif.

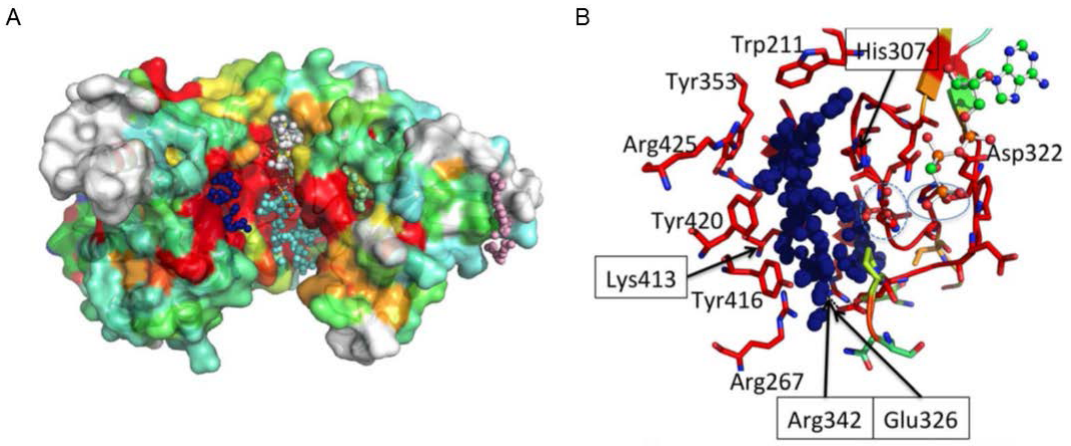


Figure 2.4 - Identification of surface cavities in Mak^{Mvan} (A) Surface representation of free Mak^{Mvan} colored according to residue conservation (consistent with the multiple sequence alignment shown in Figure 2.3). Cavities identified with the software fpocket are represented as dotted spheres (pocket 1: blue; pocket 2: cyan; pocket 3: white; pocket 4: light green; pocket 5: light pink). (B) Close view of the residues lining pocket 1 colored as in panel A. Important residues are shown as sticks according to sequence conservation (protein) or with oxygen atoms in red, nitrogen in blue, sulfur in yellow, phosphorus in orange and carbon in green (nucleotide). Residue Asp305 (ball-and-stick) is highlighted by a dashed blue line; the γ-phosphate of AppCp (ball-and-stick representation) is highlighted by a blue line. Figure prepared with PyMOL (http://www.pymol.org) by Sandra de Macedo-Ribeiro and Pedro J. B. Pereira at Instituto de Biologia Molecular e Celular (IBMC, Porto) and adapted from Fraga *et al.* (2015) with permission.

However, considering the overall higher conservation of pocket 1 in Mak^{Mvan} three-dimensional structure, the topological equivalent to the substrate binding region in other ELKs, this cavity most likely represents the maltose-binding site. A mutational analysis of key residues was carried out to corroborate this hypothesis. Accordingly, mutations of the invariant aromatic residues Tyr416 and Tyr420 and the basic residue Lys413 practically abolished maltokinase activity (Figure 2.5). The observation that mutation of residues 416 and 420 to phenylalanine was sufficient to drastically reduce Mak^{Mvan} catalytic activity suggests that maltose binding involves the establishment of polar interactions with Tyr416 and Tyr420 side chains.

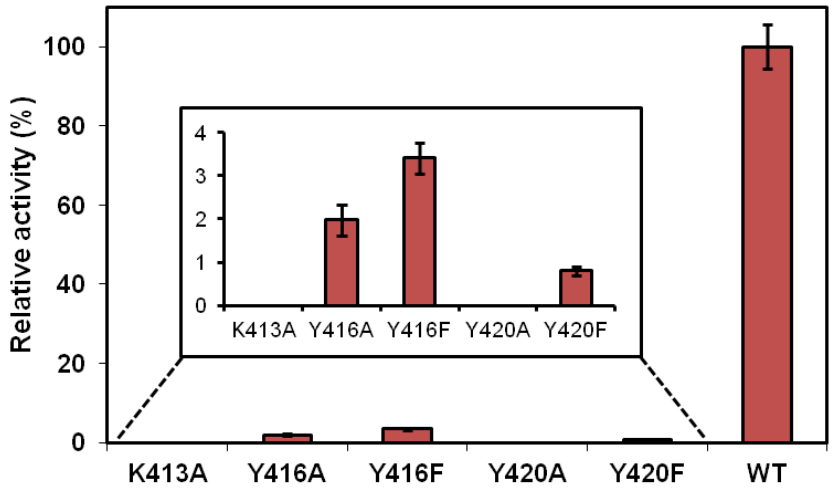


Figure 2.5 - Mutational analysis of Mak^{Mvan} maltose binding site. Analysis of the effect of point mutations in key residues for Mak^{Mvan} catalytic activity.

Conservation of key active site residues - Mutational analysis of *M. tuberculosis* Mak

Mak^{Mtb} shares with Mak^{Mvan} conservation of most residues within the active site cleft. In the active site most differences can be identified at the adenine base-binding site, namely in the linker region and in strand β5 (Figure 2.3). Variations in this region were previously correlated with nucleoside recognition specificity in aminoglycoside phosphotransferases (Shi and Berghuis, 2012) and could justify the experimentally observed small differences in nucleotide base specificity between these two mycobacterial enzymes (Figure 2.1 and Table 2-II).
A mutational analysis was performed to analyze the overall conservation of the main structural features observed in Mak^{Mvan} in the *M. tuberculosis* enzyme regarding key residues. Overall, the P-loop, catalytic loop and the DFE motif remain largely conserved, as well as the residues forming the identified maltose-binding pocket (Figure 2.6). Mutation of the invariant magnesium-binding Asp339 of Mak^{Mtb} (Asp322 in Mak^{Mvan}) to asparagine completely abolished the enzymatic activity, while the mutation of the γ-phosphate interacting residue Ser144 (Ser136 in Mak^{Mvan}) drastically reduced maltose phosphorylation. Also, destabilization of the DFE-loop conformation by mutating the Glu340-stabilizing (Glu324 in Mak^{Mvan}) arginine side chain (mutant R351A; Arg334 in Mak^{Mvan}) resulted in decreased activity. Taken together, this mutational analysis underscores the overall conservation of the main structural features observed in Mak^{Mvan} in the *M. tuberculosis* enzyme.

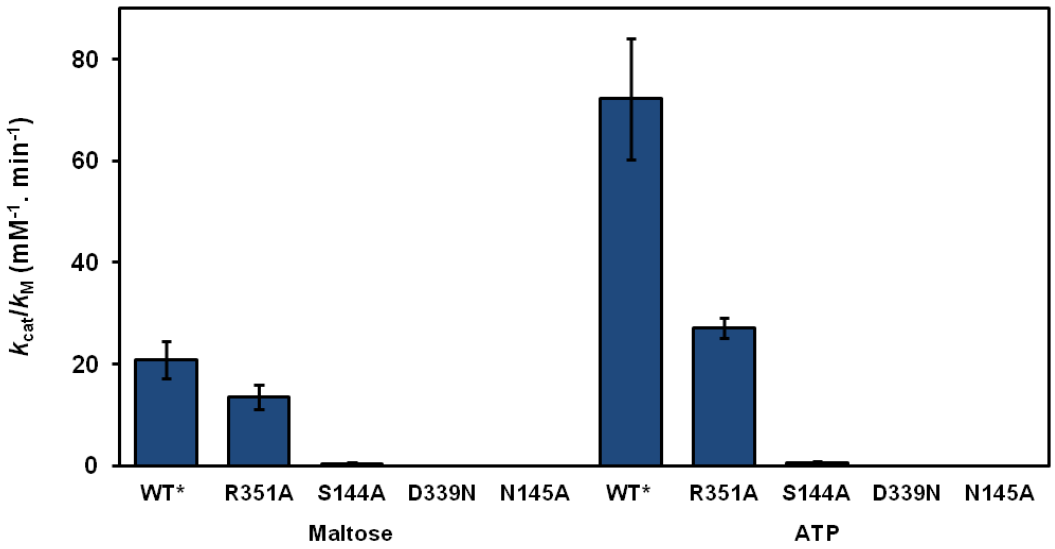


Figure 2.6 - Mutational analysis of *M. tuberculosis* Mak P-loop residues and DFE Motif. Analysis of the effect of point mutations in key residues for Mak^{Mtb} catalytic efficiency (k_{cat}/K_m). *WT, Adapted from (Mendes *et al.*, 2010).

The relevance of the tight interaction between the γ-phosphate binding loop and the cap subdomain is highlighted by the finding that mutation of Asn145 (Asn137 in Mak^{Mvan}) completely inactivated Mak^{Mtb}. This asparagine residue does not interact directly with the nucleotide in the Mak^{Mvan}:AppCp complex, but stabilizes an ordered solvent network interlinking the P-loop, the nucleotide, and the DFE motif and helps maintaining the main-chain conformation of Arg160 (Arg152 in Mak^{Mvan}), a crucial residue at the interface between the intermediate and the cap subdomains that stacks with the invariant Trp14. These invariant residues, tightly cross-linked

to the γ-phosphate binding P-loop, form part of one of the small cavities identified in the structure of Mak^{Mvan} (pocket 4, Figure 2.4). This conserved interface region is stabilized predominantly by a network of polar interactions, suggesting that structural changes in the flexible N-terminal cap subdomain could be easily conveyed to the nucleotide-binding region, including the P-loop, and interfere with phosphate donor stabilization and phosphoryl group transfer.

Discussion

Mak (*Rv0127*) from *M. tuberculosis* is a bona fide maltokinase (Mendes *et al.*, 2010) and recently this activity was also confirmed for the *M. smegmatis* orthologue (Roy *et al.*, 2013). Here we demonstrate that the *M. vanbaalenii* Mak orthologue (Mak^{Mvan}) is also a maltokinase whose novel structure reveals the molecular details of this family of enzymes, identifying the conserved structural motifs associated with its phosphotransferase activity. Despite low sequence identity with aminoglycoside phosphotransferases, maltokinases display significant structural homology and conserve the main eukaryotic protein kinase-like (ELK) motifs (Jarling *et al.*, 2004; Kannan *et al.*, 2007; Scheeff and Bourne, 2005).

The concept of protein phosphorylation as a mechanism for regulation dates to the 1950's and was the foundation for the family of eukaryotic protein kinases (ePKs), one of the largest gene families, accounting for 2-4% of all genes expressed and regulating most of the biological processes in the cell (Taylor *et al.*, 2012). The protein kinase–like (PKL) is a superfamily of enzymes that can be divided into the eukaryotic protein kinases (ePKs), the controllers of eukaryotic cell biology, and their prokaryotic counterparts, the eukaryotic-like protein kinases (ELK) (Meharena *et al.*, 2013). PKL kinases are scattered across phyla and are usually regulatory proteins, although they can have several different functions. They share a common fold and the biochemical function of ATP-dependent phosphorylation. ELKs generally display very low sequence identity (7%–17%) to ePKs and to each other (Kannan *et al.*, 2007). The prototypical kinase structure consists of two lobes joined by a short loop, known as the hinge segment. The smaller N-terminal lobe is composed of five-stranded anti-parallel β-sheet and the larger C-terminal lobe almost exclusively α-helixes. Catalysis occurs at the interface of the two lobes. The deep cleft between the two lobes forms the active site, ATP binding pocket, where phosphorylation occurs. The residues involved in the phosphotransfer reaction are highly conserved. Both lobes participate in the binding of ATP with the metal ions necessary for the reaction to occur. The C-terminal lobe binds the substrate, bringing it in close proximity to ATP, resulting in the phosphorylation. The N-terminal lobe is mostly involved in orienting nucleotide binding (Pérez *et al.*, 2008; Taylor *et al.*, 2012). Kinases share a highly similar core cassette and numerous sequence motifs, that are extremely conserved, but because they are a very diverse group of proteins phosphorylating a vast array of substrates, their overall structure translates such diversity. The C-terminal is especially diverse and, in some cases, structural changes have

allowed the kinases in the superfamily to partner with accessory domains important to activity and/or regulation (Scheeff and Bourne, 2005). Overall, Mak^{Mvan} structure is very similar to ELK structure with its bilobal arrangement with an exposed catalytic cleft. However, in contrast to ELKs, the N-terminal lobe has an additional N-terminal subdomain, the cap subdomain, which is unique to maltokinases and unveiled an unforeseen resemblance with proteins displaying the cystatin fold. Cystatins are a protein superfamily of naturally occurring cysteine protease inhibitors whose first solved structure was extracted from chicken and revealed a new fold, named the cystatin fold, which is based on a five stranded anti-parallel β-sheet wrapped around the central N-terminal helix (Turk *et al.*, 2008; Rao *et al.*, 1998).

ELKs are abundant only in specific groups of prokaryotes (Pérez *et al.*, 2008). Mycobacteria are especially rich in eukaryotic-like Ser/Thr protein kinases (STPKs) involved in regulation cell growth and division, stress responses and pathogenicity. Other examples of ELKs are 5-methylthioribose (MTR) kinase involved in bacterial methionine salvage pathway (Ku *et al.*, 2007) and aminoglycoside phosphotransferases bacterial proteins which *O*-phosphorylate in a ATP-dependent fashion and confer resistance to various aminoglycosidic antibiotics (Ramirez and Tolmasky, 2010).

Structural analysis of free Mak and of the Mak:AppCp complex suggests that unlike protein kinases, no large lobe movements occur upon nucleotide binding, a common feature in the structurally related ELKs such as MTRK32 and aminoglycoside phosphotransferases (Shi and Berghuis, 2012; Burk *et al.*, 2001; Hon *et al.*).

The P-loop (a.k.a. nucleotide-positioning loop) is the structural equivalent of the glycine-rich G-loop found in eukaryotic protein kinases and MTRK, and anchors the terminal phosphate groups of the nucleotide (Scheeff and Bourne, 2005). Curiously, in the Mak^{Mvan}:AppCp complex the γ-phosphate does not seem to be ideally oriented for phosphoryl transfer as it faces away from the putative catalytic base (Asp305) and the proposed maltose-binding pocket. Nevertheless, the relevance of the observed interaction between the γ-phosphate and the invariant Ser136 in the P-loop is underscored by the 97% decrease of enzymatic activity upon mutation of this residue to alanine. We propose that repositioning of the terminal phosphoryl group is required for optimum phosphate transfer to the maltose acceptor, and that either substrate binding or a conformational change in the proximity of the P-loop would poise the terminal phosphoryl group for catalysis.

For the kinases to function properly their catalytic activity must be tightly controlled. The catalytic site varies between conformational states. Regulatory mechanics include (i) binding of allosteric effectors and subsequent conformational changes in the catalytic domain (ii) autophosphorylation mechanisms. (iii) phosphorylation at Ser/Thr/Tyr conserved sites activating signal transduction pathways (Pérez *et al.*, 2008; Meharena *et al.*, 2013). Conservation of the cap subdomain in maltokinases (including the bifunctional TreS-Mak enzymes), in particular of the residues in the proximity of the P-loop, together with the potential flexibility of this region as indicated by its high B-values in the Mak^{Mvan} structure, are compatible with regulatory functions for this subdomain. In fact, regulatory non-catalytic domains have been identified in eukaryotic protein kinases, modulating enzyme activity through interaction with

other macromolecules (Bai *et al.*, 2014; Leonard *et al.*, 2014). It has been recently observed that *M. smegmatis* TreS directly interacts with maltokinase also stimulating its enzymatic activity (Roy *et al.*, 2013). Considering that in many species Mak is fused to the C-terminus of TreS, suggesting spatial proximity between the N-terminal region of Mak and the C-terminal carbohydrate-binding domain of TreS (Caner *et al.*, 2013), it is possible to hypothesize that the N-terminal cap subdomain plays a central role in modulation of Mak enzymatic activity. In fact, a narrow pocket that could accommodate an extended molecule (e.g. a linear peptide) (Figure 2.4, pink spheres) can be identified in the cap subdomain. Macromolecular interactions in this region could result in a concerted motion at the interface between the cap and the intermediate subdomains, particularly affecting the dynamics of the phosphate-binding P-loop, and providing a possible explanation for the observed activation of maltokinase activity upon binding to TreS. Although no TreS homologue could so far be identified in the genome of *M. vanbaalenii*, sequence homology searches pinpointed a gene coding for a non-homologous putative TreS-like protein, Mvan_5178. In this candidate TreS, whose functional relevance in this process remains to be elucidated, relevant catalytic residues are conserved despite low overall amino acid sequence identity.

A regulated activation of maltokinase activity is in agreement with the fact that M1P is highly toxic to the mycobacterial cell and its production needs to be tightly controlled. In a possible model for the flow of substrates for α-glucan biosynthesis, TreS is recruited to the appropriate reservoir of glucose (glycogen) through its C-terminal domain. Upon binding to this site TreS produces maltose and associates with Mak, thereby activating the synthesis of M1P, itself channelled through GlgE for glycogen extension. The selection of the pathway leading to glycogen degradation or biosynthesis could be crucial for homeostasis of the cellular energy levels in mycobacteria and it is tempting to speculate that the enzymes involved in these pathways are associated in multifunctional protein complexes. Maltokinase's proposed essentiality in *M. tuberculosis* (Griffin *et al.*, 2011b) was hypothesized to stem from a constitutive regulatory role in sugar metabolism, in line with similar suggestions for *E. coli* (Mendes *et al.*, 2010; Decker *et al.*, 1999). The structural characterization and the identification of the molecular features associated with substrate recognition and catalytic activity of the enzymes in the essential GlgE pathway is crucial for the rational design of novel specific antimicrobial compounds. Although GlgE is a validated target for the design of specific inhibitors such as M1P analogs, a maltose transporter is lacking in *M. tuberculosis* presumably hampering the uptake of maltose analogs (Syson *et al.*, 2014). This argues in favour of designing pro-drugs that can be enzymatically converted by endogenous TreS and/or Mak into the biologically active GlgE inhibitors (Swarts *et al.*, 2012; Chopra and Brennan, 1998). The high-resolution three-dimensional structure of mycobaterial maltokinase is a cornerstone in the structural roadmap of the essential GlgE pathway and completes the necessary experimental framework for the rational design of mycobacterial-targeting compounds that may act as narrow spectrum antibiotics (Chopra and Brennan, 1998).

Acknowledgements and work contributions

All crystallography procedures (with corresponding data collection and processing, structure determination and figure preparation) were performed by Joana S. Fraga, Sandra de Macedo-Ribeiro and Pedro J. B. Pereira at Instituto de Biologia Molecular e Celular (IBMC, Porto). DNA manipulations with QuikChange II Site-Directed Mutagenesis Kit (Agilent) were performed by Joana S. Fraga.

This work was funded by national funds through Fundação para a Ciência e a Tecnologia (FCT) and by EU-FEDER through the Operational Competitiveness Programme – COMPETE with grants FCOMP-01-0124-FEDER-014187 (PTDC/BIA-BCM/112459/2009) and FCOMP-01-0124-FEDER-028359 (PTDC/BIA-MIC/2779/2012).

Chapter 3 Octanoylation of early intermediates of mycobacterial methylglucose lipopolysaccharides

This chapter is published in:

Maranha, A., Moynihan, P.J., Miranda, V., Lourenço, E.C., 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**: p. 13610.

Chapter 3- table of contents

Abstract

Mycobacteria synthesize unique intracellular methylglucose lipopolysaccharides (MGLPs) proposed to modulate fatty acid metabolism. In addition to the partial esterification of glucose or methylglucose units with short-chain fatty acids, octanoate was invariably detected on the MGLP reducing end. We have identified a novel sugar octanoyltransferase (OctT) that efficiently transfers octanoate to glucosylglycerate (GG) and diglucosylglycerate (DGG), the earliest intermediates in MGLP biosynthesis. Enzymatic studies, synthetic chemistry, NMR spectroscopy and mass spectrometry approaches suggest that, in contrast to the prevailing consensus, octanoate is not esterified to the primary hydroxyl group of glycerate but instead to the C6 OH of the second glucose in DGG. These observations raise important new questions about the MGLP reducing end architecture and about subsequent biosynthetic steps. Functional characterization of this atypical octanoyltransferase, whose gene has been proposed to be essential for *M. tuberculosis* growth, adds new insights into a vital mycobacterial pathway, which may inspire new drug discovery strategies.

Introduction

The genus *Mycobacterium* includes several successful pathogens two of which, *M. leprae* and *M. tuberculosis,* cause the particularly challenging and hard-to-eradicate diseases leprosy and tuberculosis (TB).

Mycobacterial pathogenesis is intimately associated to a unique lipid-rich cell wall (Daffé *et al.*, 2014), to several distinctive metabolic pathways and to rare macromolecules that include intracellular α–(1→4)-linked polymethylated polysaccharides (PMPSs) of 3–*O*–methylmannose (MMP) and of 6–*O*–methylglucose (MGLP) (Mendes *et al.*, 2012; Xia *et al.*, 2012; Jackson and Brennan, 2009). PMPSs were proven to form stable 1:1 complexes with long-chain fatty acids *in vitro* modulating and stimulating their synthesis through interaction with fatty acid synthase I (FAS-I) although their *in vivo* function remains to be demonstrated (Bloch and Vance, 1977; Yabusaki and Ballou, 1979). MGLP is composed of 15-20 α–(1→4)–linked glucose some of which methylated (see Chapter 1, Figure 1.6) (Jackson and Brennan, 2009). In the mature MGLP structure, short-chain fatty acids (namely acetate, propionate or isobutyrate) are often esterified to the glucose units near the non-reducing end (Smith and Ballou, 1973; Dell and Ballou, 1983). The polysaccharide's reducing end is composed of diglucosylglycerate (DGG) and the second glucose of DGG is linked to the MGLP main chain, which contains two additional branching glucoses to which succinate may be esterified (see Chapter 1, Figure 1.9) (Tuffal *et al.*, 1998; Mendes *et al.*, 2012). An octanoate moiety was initially detected at the glycerate unit and suggested to anchor MGLP intermediates to the cytoplasmic membrane during the elongation steps (Smith and Ballou, 1973). Octanoate has also been proposed to raise the interaction energy required for complexation with longer acyl-CoAs, which might confer further discriminatory ability to MGLP (Hindsgaul and Ballou, 1984). In addition, octanoate was suggested to contribute to stabilize MGLP's helical conformation by forming an axis around which MGLP could fold in the absence of longer acyl-CoAs (Hindsgaul and Ballou, 1984). In *M. phlei*, MGLP acylation levels were shown to regulate methylation, which could only be observed after partial acetylation of glucoligosaccharide acceptors (Grellert and Ballou, 1972). This could explain octanoylation as the basal acylation required for 6–*O* methylation to occur.

Although MGLP was identified in the 1960s (Lee and Ballou, 1964) and its chemical composition, structure and interactions with fatty acids have been thoroughly examined, its biosynthetic pathway remained largely unknown (Mendes *et al.*, 2012). Glucosylglycerate (GG), the putative primer for MGLP synthesis, is now known to arise from the consecutive action of the enzymes glucosyl-3-phosphoglycerate synthase (GpgS) and glucosyl-3-phosphoglycerate phosphatase (GpgP), encoded in *M. tuberculosis* H37Rv by *Rv1208* and *Rv2419c*, respectively (see Chapter 1, Figure 1.9) (Kamisango *et al.*, 1987; Empadinhas *et al.*, 2008; Kaur *et al.*, 2009; Mendes *et al.*, 2011). A GG hydrolase was recently identified in rapidly-growing mycobacteria but its involvement in the regulation of MGLP biosynthesis has not been examined (Alarico *et al.*, 2014). The third MGLP biosynthetic step was predicted to involve the transfer of an additional glucose to GG to yield DGG (Stadthagen *et al.*, 2007). Given that bacterial genes involved in one pathway

are often grouped into operons (Ma and Xu, 2013), confirmation of the genetic implication of Rv3032 (α–(1→4)-glycosyltransferase) and Rv3030 (6–*O*–methyltransferase) in MGLP extension and methylation, respectively, linked Rv3031 (a putative glycoside hydrolase similar to α–(1→6)-branching enzymes) to DGG synthesis (Stadthagen *et al.*, 2007; Jackson and Brennan, 2009) (Figure 3.1). After assigning GpgP activity to Rv2419c we deemed important to investigate the function of the neighboring gene considered essential for *M tuberculosis* H37Rv growth (Mendes *et al.*, 2011; Sassetti *et al.*, 2003; Griffin *et al.*, 2011a) and annotated as a protein with motifs of the GDSL hydrolase family (Akoh *et al.*, 2004). In addition to *Rv2418c,* we cloned the *M. smegmatis* orthologue and that from the thermophilic *M. hassiacum*, for protein stability constraints often curbing functional and structural studies (Tiago *et al.*, 2012).

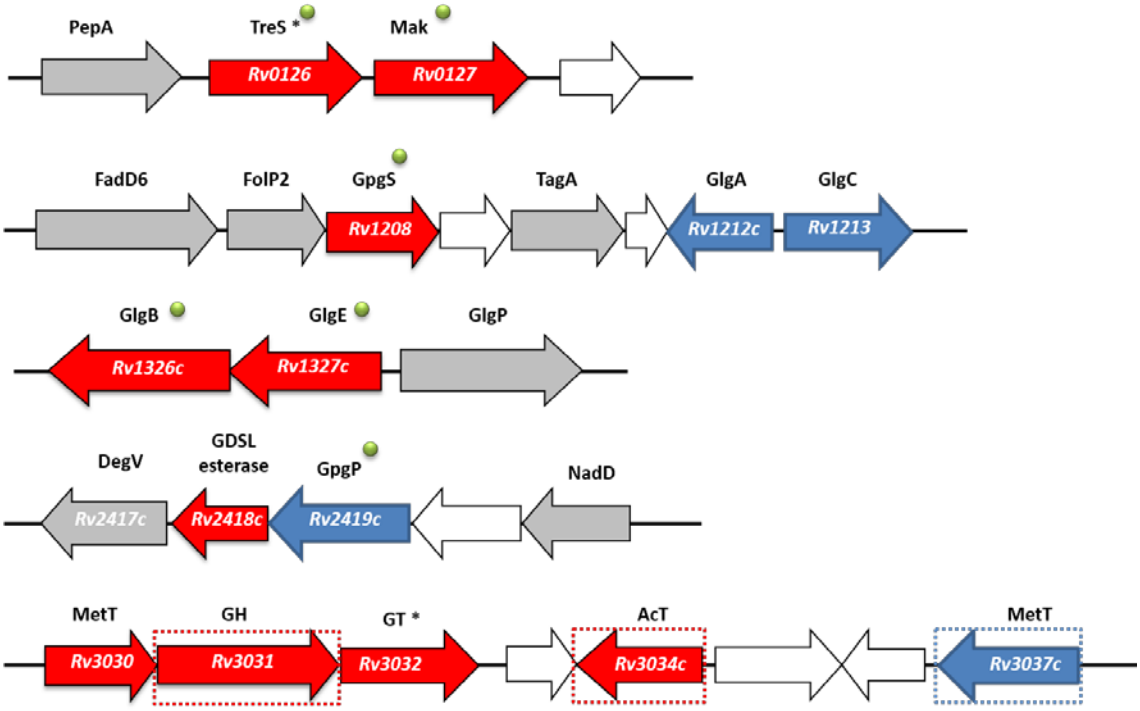


Figure 3.1 - *Mycobacterium tuberculosis* H37Rv gene clusters proposed to participate in MGLPs biosynthesis. Genes of unknown or hypothetical function are shaded white and grey, respectively. Genes considered essential for *M. tuberculosis* growth by saturation transposon mutagenesis (Griffin *et al.*, 2011b) are shaded red or blue if nonessential. *Although proposed to be essential for growth deletion of these genes produced viable mutants in *M. tuberculosis* H37Rv (Murphy *et al.*, 2005; Stadthagen *et al.*, 2007). Genes whose confirmation of involvement in MGLP pathway is still missing are boxed. Green spheres represent characterized proteins: PepA, serine protease; FadD6, long-chain-acyl-CoA synthetase; FolP2, dihydropteroate synthase; TagA, DNA-3-methyladenine glycosylase; NadD, nicotinate-nucleotide adenylyltransferase; GH, glucoside hydrolase.

Herein, we show that Rv2418c catalyzes the transfer of octanoate and to a lesser extent, some short-chain fatty acids, to DGG and GG *in vitro*. Surprisingly, we found that octanoate is not transferred to glycerate as previously reported but instead to the C6 OH of the second glucose in DGG. This was confirmed by comparison of the enzymatic product with both chemically synthesized versions of octanoylated DGG. The implication of this novel octanoyltransferase (OctT) in the biosynthetic pathway to MGLP furthers our knowledge on the molecular glycobiology of mycobacteria and offers additional tools for drug discovery.

Methods

Genomic context, sequence analysis and structure prediction

After assignment of the GpgP function to *Rv2419c* (Mendes *et al.*, 2011) we sought to investigate the function of the neighboring gene *Rv2418c* (Figure 3.1). The Rv2418c amino acid sequence was retrieved from the TubercuList database (http://tuberculist.epfl.ch/index.html) (Lew *et al.*, 2011) and used as template for BLAST searches in available mycobacterial genomes. Parameters such as predicted molecular weight and isoelectric point were generated using the ProtParam tool (http://web.expasy.org/protparam). The protein sequence alignment editor Aline (Bond and Schuttelkopf, 2009) was used for preparing the sequence alignment in (Figure 3.2). A structural model prediction was generated using a combined approach with Swiss model (Biasini *et al.*, 2014) and the Phyre2 prediction server (Kelley *et al.*, 2015) .

Molecular biology and recombinant gene expression

For recombinant expression we selected the gene from *Mycobacterium tuberculosis* H37Rv (*Rv2418c*) and the orthologues from *M. smegmatis* mc^2155 (MSMEG_4578 or MSMEI_4466 or LJ00_22655) and *M. hassiacum* DSM44199 (GenPept accession WP_026213345.1). Each of the *M. tuberculosis* and *M. hassiacum* sequences were optimized for expression in *E. coli*, synthesized (GenScript) and cloned between the *Nde* I and *Hind* III restriction sites of the expression vector pET30a (Novagen). MSMEI_4466 was amplified from chromosomal DNA purified from *M. smegmatis* (SmartHelix DNAid Mycobacteria kit, Sekvenator, Slovenia). Amplification primers were designed from the sequence retrieved from the SmegmaList database (http://mycobrowser.epfl.ch/smegmalist.html). *Nde* I and *Hind* III restriction sites (underlined) were added to the forward 5'-TACGATCATATGTCCTCTGAGACATCCTCG) and the reverse (5'-TATAAGCTTGGTTGCGGAATCCCGTTG) primers, respectively. Stop codons were removed from reverse primers to allow translation of C-terminal 6×His-tags. Amplification was carried out in 50 µL mixtures with KOD hot start DNA polymerase (Novagen) using 150 ng of template DNA and 68°C as annealing temperature. PCR products were purified from agarose gels (Jetquick, Genomed) digested with the appropriate restriction enzymes, ligated to the corresponding sites of pET30a (Life Technologies) and transformed into *E. coli* DH5α with standard procedures (Sambrook and Russell, 2001). Recombinant plasmids were purified and sequenced to confirm the identity of inserts (Macrogen Europe). All constructs were transformed into *E. coli* BL21 and growth was carried out at 37°C in LB medium containing kanamycin (30 µg/mL) to early exponential phase of growth (OD_{610}=0.6) in an orbital shaker at 150 rpm. Growth temperature was progressively decreased to 20°C until mid-exponential phase of growth (OD_{610}=0.9) and IPTG was added to a final concentration of 0.5 mM to induce gene expression. Cells carrying the recombinant OctT were harvested 18 h later by centrifugation (9000 *× g*, 10 min, 4°C), suspended in 20 mM sodium phosphate buffer pH 8.0 with 0.3 M NaCl, 10 mM imidazole, 7.5 mM β-mercaptoethanol and 0.1% Tween 20 (buffer A) containing 10

μg/mL DNAse I, and disrupted by sonication on ice with three 40 Hz pulses of 45 s (30 s pause between pulses) per 7 mL of lysate, followed by centrifugation to remove debris (15000 × *g*, 4°C, 30 min). Protein production was analyzed in 12% SDS-PAGE gels followed by Western immunoblot confirmation. Detection of His-tagged proteins was performed with a 1:1000 dilution (in 3% BSA in TBS) of a mouse anti-His_6 antibody (Santa Cruz Biotechnology) and a 1:2000 dilution (in 3% (wt/vol) BSA in TBS) of the secondary goat anti-mouse IgG + IgM alkaline phosphatase-conjugated antibody (Bio-Rad). The signal was revealed with Pierce NBT/BCIP 1-Step Solution (Thermo Scientific) and visualized on a Geldoc equipment (Bio-Rad).

Protein purification

His-tagged recombinant OctTs were purified with a Ni-Sepharose high-performance column (His-Prep FF 16/10, GE Healthcare) equilibrated with buffer A. Elution was carried out in the same buffer containing 200 mM imidazole and the purity of the fractions was evaluated by SDS-PAGE. The purest active fractions pooled after visualization of product formation by thin-layer chromatography (see below) were diluted five-fold with 20 mM Bis-tris propane buffer (BTP) pH 8.0, 200 mM NaCl, 7.5 mM β-mercaptoethanol, 0.1% Tween 20, concentrated by ultrafiltration in 10 kDa molecular weight cutoff centrifugal devices (Amicon) and loaded onto a HiPrep 16/60 Sephacryl S-200 HR column (GE Healthcare) equilibrated with the same buffer and eluted isocratically. The purity of the fractions was assessed by SDS-PAGE and the purest fractions were pooled, concentrated as described above, and equilibrated in 20 mM BTP pH 8.0, 50 mM NaCl. Protein content was determined by the Bradford assay (Bio-Rad). The molecular mass of the recombinant OctTs was estimated by gel filtration using a HiPrep 16/60 Sephacryl S-200 HR column equilibrated with 20 mM BTP pH 8.0, 200 mM NaCl, 7.5 mM β-mercaptoethanol, 0.1% Tween 20 and using ribonuclease A (13.7 kDa), carbonic anhydrase (29 kDa), ovalbumin (43 kDa) and conalbumin (75 kDa) as molecular mass standards. Blue dextran 2000 (GE Healthcare) was used to determine the void volume.

Chemical synthesis of GG; DGG; (2*R*)-2-*O*-[6-*O*-octanoyl-(α-D-glucopyranosyl-(1→6)-α-D-glucopyranosyl]-2,3-dihydroxypropanoic acid 1 and of (2*R*)-2-*O*-(α-D-glucopyranosyl-(1→6)-α- D-glucopyranosyl)-3-*O*-octanoyl-2,3-dihydroxypropanoic acid 2

The substrates glucosylglycerate (GG) diglucosylglycerate (DGG) were chemically synthesized as previously described (Lourenço *et al.*, 2009). ^{1}H NMR spectra were obtained at 400 MHz in $CDCl_3$, DMSO-d_6 or D_2O. ^{13}C NMR spectra were obtained at 100.61 MHz in the same deuterated solvents. Assignments are supported by 2D correlation NMR studies. Medium pressure preparative column chromatography: Silica Gel Merck 60 H. Analytical TLC: Aluminum-backed Silica Gel Merck 60 F254. Reagents and solvents were purified and dried according to Armarego and Chai (Armarego and Chai, 2003). All reactions were carried out under an inert atmosphere (argon) except when the solvents were not dried. The methods for synthesis and coordinates of

all chemically synthesized compounds which were the subject of a provisional patent application are indicated in:

Maranha, A., Moynihan, P.J., Miranda, V., Lourenço, E.C., Nunes-Costa, D., Fraga, J.S., José Barbosa Pereira, P., Macedo-Ribeiro, S., Ventura, M.R., Clarke, A.J*., et al.* (2015). *Octanoylation of early intermediates of mycobacterial methylglucose lipopolysaccharides*. Scientific Reports **5**: p. 13610.

Analysis of enzyme activity by thin-layer chromatography

Enzyme activity was monitored by thin-layer chromatography (TLC) on Silica 60 gel plates (Merck) with a solvent system composed of acetic acid/ethyl acetate/water/ammonia 25% (6:6:2:1, vol/vol). Sugars, sugar derivatives and esterified sugars were visualized by spraying with α-naphtol-sulfuric acid solution followed by charring at 120°C (Jacin and Mishkin, 1965). Standards of each compound and of synthetic octanoylated DGG (Oct-DGG) were used for comparative purposes.

Enzyme assays

The recombinant *M. hassiacum* OctT was selected for biochemical characterization due to its greater stability while the recombinant protein from *M. smegmatis* was used to calculate relevant kinetic constants. Activity was measured with both discontinuous and continuous methods as described below.

For routine detection of activity (continuous assays) in 96-well microtiter plates, the synthetic substrate octanoyl-*p*-nitrophenol (Oct-*p*NP; 1 mM) was added to reaction mixtures containing 50 mM BTP pH 8.0 and 12 mM DGG acceptor. Reactions (100 µL) were initiated by the addition of enzyme (4 µg) and product formation was followed by monitoring the increase of absorbance at 348 nm, indicative of release of *p*-nitrophenolate. For quantification of reaction rates with CoA-activated substrates, 2.5 mM of DTNB (5,5'-dithiobis-2-nitrobenzoic acid, Ellman's Reagent) was added to reaction mixtures prior to enzyme addition, and the release of CoASH was monitored at 412 nm and 0.5 mM Oct-CoA was utilized. Control reactions were performed to account for possible substrate degradation and to ensure that ethanol-solubilized DTNB did not influence enzyme activity at the concentrations tested.

Mycobacterium hassiacum OctT temperature and pH profiles were determined with a discontinuous assay in reactions initiated by the addition of 4 µg OctT to mixtures (100 µL) containing the appropriate buffer, 1 mM Oct-*p*NP and 12 mM DGG. Cooling on an ethanol-ice bath stopped reactions and the enzyme was inactivated by the addition of 2 µL of 5 N HCl. The mixture was neutralized with 2 µL of 5 N NaOH followed by the addition of 100 µL of 100 mM BTP pH 8.0 to stabilize the pH. The amount of *p*-nitrophenolate released was estimated by measuring the absorbance of the reaction mixtures at 348 nm. To probe for a possible esterase activity, OctT (4 µg in 100 µL 50 mM BTP pH 8.0) was incubated for 30 min at 37°C with pure

synthetic Oct-DGG (200 μM, 500 μM or 1 mM) with appropriate controls to account for possible degradation.

Enzyme characterization

Substrate specificity was assessed using glucose, mannose, galactose, sucrose, laminaribiose (Dextra), trehalose, isomaltose maltose, maltotriose, maltotetraose, maltopentaose, maltohexaose, maltoheptaose (all maltooligosaccharides from Carbosynth), glucosamine, kanamycin and D,L-glyceric acid (sugars were all tested both independently and in combination with D,L-glyceric acid). GG (α-D-glucosyl-(1→2)-D-glycerate), GPG (α-D-glucosyl-3-phospho-D-glycerate), DGG (α-D-glucosyl-(1→6)-α-D-glucosyl-(1→2)-D-glycerate) (Lourenço *et al.*, 2009), MG (α-D-mannosyl-(1→2)-D-glycerate), MPG (α-D-mannosyl-3-phospho-D-glycerate) (Empadinhas *et al.*, 2011), MGG (α-D-mannosyl-(1→2)-α-D-glucosyl-(1→2)-D-glycerate) were also tested as possible acceptors (Lourenço and Ventura, 2011). The CoA derivatives palmitoyl-CoA (C16), tetradecanoyl-CoA (C14), dodecanoyl-CoA (C12), decanoyl-CoA (C10), octanoyl-CoA (C8, Oct-CoA), hexanoyl-CoA (C6, Hex-CoA), succinyl-CoA, butyryl-CoA (C4), propionyl-CoA (C3) and acetyl-CoA (C2) (last three from Larodan Fine Chemicals) were tested as possible acyl donors. Octanoic acid (isolated and in combination with ATP), octanoyl-*p*NP, hexanoyl-*p*NP (TCI), butyryl-*p*NP and acetyl-*p*NP were also tested as possible substrates. All chemicals were from Sigma-Aldrich unless otherwise stated. Pentanoyl-CoA (C5) and heptanoyl-CoA (C7) were not commercially available. The substrate combinations to be tested were selected according to their relevance in the context of MGLP structure and acylation pattern. Product formation was examined by TLC (as described above) and confirmed by MS/MS analysis as described below.
The effect of divalent cations on enzyme activity was examined by incubating the reaction mixture with the chloride salts of Mg^{2+}, Mn^{2+}, Co^{2+}, Zn^{2+} (5 to 20 mM), without cations or in the presence of 10 mM EDTA, at 37°C. Because the CoA esters solutions utilized contained lithium salt the effect of this cation on enzyme activity was also tested within the range of 1 μM-3 μM (equivalent to the concentration in the CoA esters solution) to an excess of 1mM.
The temperature profile was determined between 25 and 55°C in 50 mM BTP pH 8.0. The effect of pH was determined at 37°C in 50 mM buffer [MES (pH 5.5 to 6.5), BTP (pH 6.5 to 9.5) or CAPS (pH 9.5 to 10.0)].

Kinetic parameters

The kinetic parameters for the recombinant OctTs were determined by measuring the amount of CoA released at 412 nm in 96-well microtiter plates as described above. Reactions were performed in 50 mM BTP buffer containing the appropriate substrates at 37°C and initiated by the addition of *M. hassiacum* OctT (4 μg) or *M. smegmatis* OctT (2.5 μg). The K_m values for DGG, GG, Oct-CoA or Hex-CoA were determined using fixed saturating concentrations of either donor Oct-CoA or Hex-CoA (100 μM), and acceptor DGG or GG (20 mM). The inhibitory effect of free CoA, free octanoate and of the Oct-DGG product on the reaction rate was tested by the addition

of known amounts of these products to mixtures containing fixed concentrations of DGG (20 mM) and Oct-CoA (80 µM). All experiments were performed in triplicate with appropriate controls. For each method, standard curves of *p*-nitrophenol and free CoA were generated to normalize the effect of sugar acceptor and buffer on the absorbance.

Mass spectrometry

Products of enzymatic reactions (100 µL) containing 50 mM BTP pH 8.0 and 0.5 mM or 1 mM ester donor (CoA or *p*NP), 12 mM sugar acceptor and 4 µg OctT were analyzed by MS or MS/MS. Sample purification was performed with Alltech® Extract-Clean Carbograph columns washed with 50% (vol/vol) and 100% acetonitrile and eluted with 3:1 isopropanol:acetonitrile, 0.1% (vol/vol) trifluoroacetic acid. Eluted reaction products were analyzed by ESI-MS on an Amazon SL ion trap mass spectrometer (Bruker Daltonics, Billerica, MA) at the Mass Spectrometry Facility of the Advanced Analysis Centre (University of Guelph, Canada). Samples were applied by direct infusion at a flow rate of 5 µL/min. The mass spectrometer electrospray capillary voltage was maintained at 4.5 kV and the drying temperature at 300°C with a flow rate of 8 L/min. Nitrogen was used as both nebulizing (40 psi) and drying gas with helium (60 psi) as collision gas. The mass-to-charge ratio was scanned across the *m/z* range 15-3000 in enhanced resolution negative-ion mode. The resulting mass spectra were analyzed using the open-source mMass 3.0 software package (Strohalm *et al.*, 2010).
Glyceric acid-containing samples were analyzed by matrix-assisted laser desorption ionization (MALDI) in a time-of-flight (TOF) analyzer at the Proteomics and Mass Spectrometry unit of Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP).

Purification of Oct-DGG and analysis by nuclear magnetic resonance (NMR)

For the analysis of the natural Oct-DGG produced by the *M. tuberculosis*, *M. hassiacum* and *M. smegmatis* recombinant OctT enzymes the product was separated from the reaction components by TLC. Oct-DGG was produced in 20 mL reactions with 50 mM BTP pH 8.0, 7 mM DGG, 1.5 mM Oct-*p*NP or 1.5 mM Oct-CoA and 800 µg of enzyme, for 3 hours at 37°C. The reactions were spotted on TLC plates and developed with a solvent system composed of 2-propanol/ethyl acetate/water/ammonia 25% (5:1:3:1 vol/vol). After staining the marginal lanes of the TLC plate and identifying the spot corresponding to Oct-DGG, preparative scale purification of the product was carried out by scraping the corresponding region in the inner unstained lanes of the TLC plate followed by extraction of the product from the silica gel with ultrapure water. The soluble product was then further purified on a Sephadex G10 column (GE Healthcare) with a water flow of 1 mL/min, lyophilized and analyzed by NMR. To confirm the structure of Oct-DGG, with special focus on the position of the octanoate esterification, the ^{1}H NMR and ^{13}C NMR spectra of synthetic compounds **1** and **2** and of the enzymatically produced compound were compared and recorded on a NMR spectrometer Bruker AVANCE II+ 400 MHz using D_2O as solvent.

Results

Identification, sequence analysis, and structure prediction

The genes adjacent to the recently identified mycobacterial glucosyl-3-phosphoglycerate phosphatase (GpgP, *Rv2419c*) (Mendes *et al.*, 2011) encoded enzymes of unknown function as well as the versatile acetyltransferase Rv2416c (Chen *et al.*, 2011) (Figure 3.1). *Rv2417c*, one of the genes with unknown function in this region showed 24% amino acid sequence identity with TM841, a *Thermotoga maritima* protein of unknown function belonging to the DEGV family (Schulze-Gahmen *et al.*, 2003). The crystal structure of TM841 has been determined (PDB entry: 1MGP) and a fatty acid molecule was identified at a non-conserved interface between the two α/β domains, suggesting a role as fatty acid carrier protein or in fatty acid metabolism (Schulze-Gahmen *et al.*, 2003). The contiguous Rv2418c was annotated as a possible lysophospholipase belonging to the SGNH_hydrolase (or GDSL_hydrolase, PF00657) superfamily, whose members contain a conserved GDSL motif characteristic of lipases, acylhydrolases and esterases (Akoh *et al.*, 2004). BLAST analysis with the Rv2418c amino acid sequence (OctT) revealed hypothetical homologues in most sequenced mycobacterial genomes (79-100% amino acid identity). Strains of the species *M. abscessus* had the most divergent sequences among the mycobacterial genomes analyzed (71% identity) and, so far, only *M. leprae* lacked a detectable OctT homologue. Members of closely related genera of the suborder Corynebacterineae such as *Rhodococcus*, *Nocardia* and others, but not of the genus *Corynebacterium*, also possess hypothetical OctT orthologues with significant amino acid identity (>50%). Among these, only a few strains of *Nocardia* are known to synthesize MGLP (Pommier and Michel, 1986). OctT orthologues have also been detected in strains of *Streptomyces* (WP_009081052). However, these strains' taxonomical classification has not been clearly defined. The closely related *Streptomyces griseus* has been reported to produce a form of acetylated methylmannose polysaccharide (AMMP) although limited details about its structure are available (Harris and Gray, 1977). No *octT* orthologues could be detected in bacterial genomes outside the class Actinobacteria.

An alignment (Figure 3.2) of selected putative actinobacterial OctTs shows that, despite the annotation as GDSL/SGNH esterase family member, and possessing the four characteristic and invariant residues (Ser, Gly, Asn, and His) of the SGNH subfamily as well as the postulated Ser-His-Asp catalytic triad required for activity, they do not possess the highly conserved and distinct GDSL region (Gly, Asp, Ser, Leu) (Akoh *et al.*, 2004), which is here limited to DSL. Interestingly, the peptidoglycan *O*-Acetyltransferase B (PatB), initially annotated as a GDSL hydrolase (with the typical catalytic triad of serine esterases), was proven to be a bona-fide acetyltransferase with only weak esterase activity (Moynihan and Clarke, 2010). The SGNH family presents overall low sequence similarity, but the limited number of known 3D structures revealed significant structural homology (Leščić Ašler *et al.*, 2010) so an attempt at a structural prediction was made with the Rv2418c amino acid sequence homologue in *M. hassiacum*.

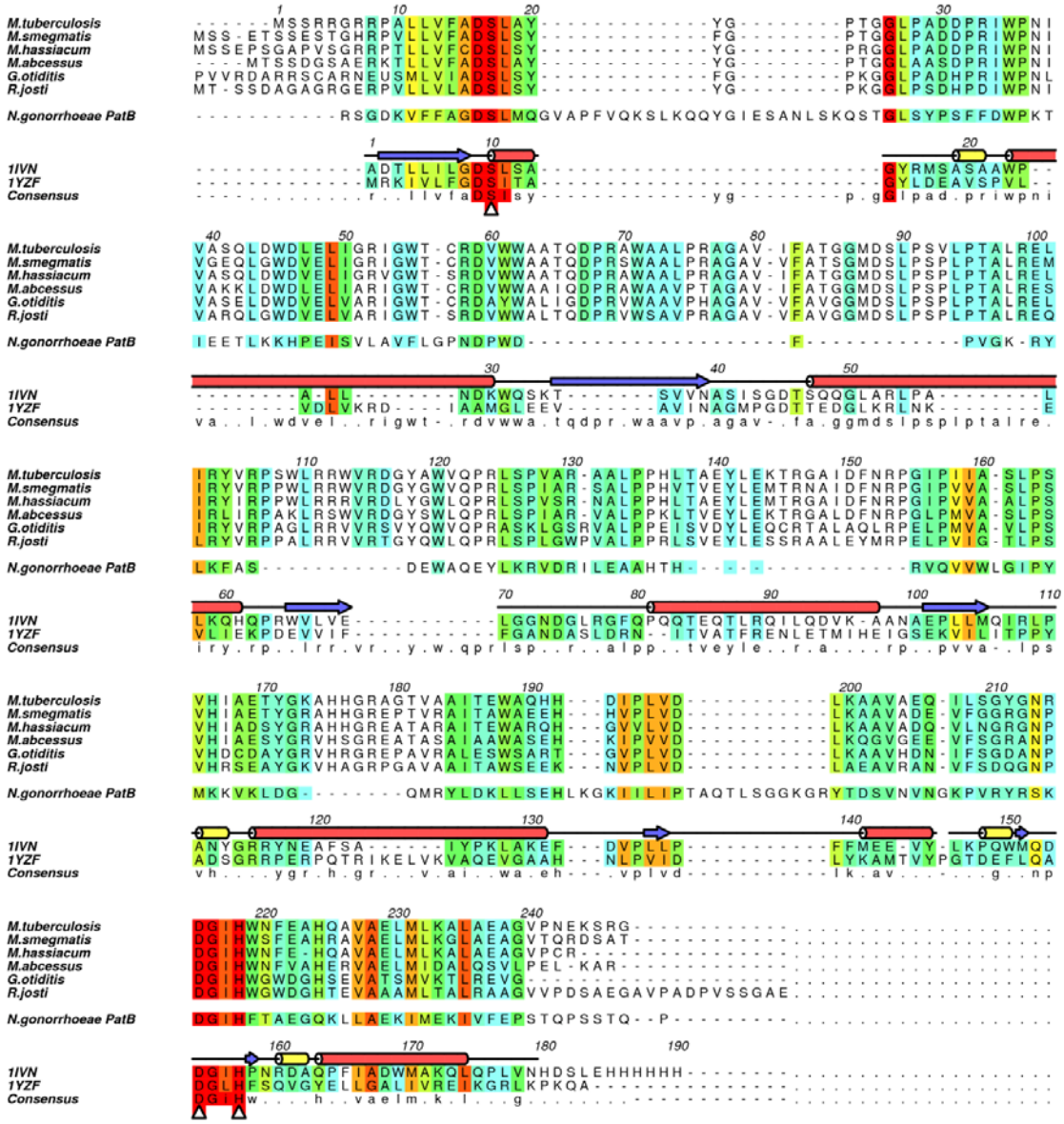


Figure 3.2 - Amino acid sequence alignment of putative acyltransferases of the GDSL family. Selected sequences include the octanoyltransferases (OctT) from *M. tuberculosis* H37Rv (Rv2418c), *M. smegmatis* mc^{2}155 (MSMEI_4466), *M. hassiacum* (accession number EKF23139), *M. abscessus* (WP_005074557) and homologues from the closely related actinobacteria *Rhodococcus jostii* (ABG93118) and *Gordonia otitidis* (WP_007238992) as well as from PatB, the peptidoglycan *O*-Acetyltransferase B from *Neisseria gonorrhoeae* (YP_207683) and two predicted structural homologues from *Escherichia coli* and *Enterococcus faecalis* (PDB entries 1IVN and 1YZF). Strictly conserved residues are boxed in red; the conserved regions are boxed from dark orange (highly conserved) to blue (partially conserved) according to the decreasing number of conserved amino acids in the aligned column. Numbers refer to the *M. tuberculosis* OctT sequence. The secondary structure elements derived from the three-dimensional structure of predicted structural homologues are shown above the alignment (red cylinders: α-helices; yellow cylinders: 3_{10} helices; blue arrows: β-sheets). The consensus sequence was calculated for the whole sequence alignment. Possible catalytic residues are indicated by white triangles below the alignment. Figure prepared with Aline2 by Sandra de Macedo-Ribeiro at Instituto de Biologia Molecular e Celular (IBMC, Porto).

The highest scoring sequence using the SWISS MODEL template search presented low homology hence the target protein was subjected to another threading method, using a different algorithm, with Phyre2, and two models were constructed. The two top scoring sequences retrieved with the two different methods were also used in the sequence alignment.

The construction of the 3D model and the alignment allowed us to estimate the location of the predicted cataltityc triad. The structural prediction is consistent with the alignment in predicting the following catalytic residues Ser23, His224 and Asp221 (

Figure 3.3). However it is possible that Ser95 also is involved in the catalytic mechanism being part of a highly conserved region in mycobacteria homologues (Figure 3.2). Site-directed mutagenesis on PatB confirmed the identity of the catalytic residues Ser133, His305 and Asp302 (Moynihan and Clarke, 2014) the His and Asp residues coincide on alignment with the ones predict for *M. hassiacum*.

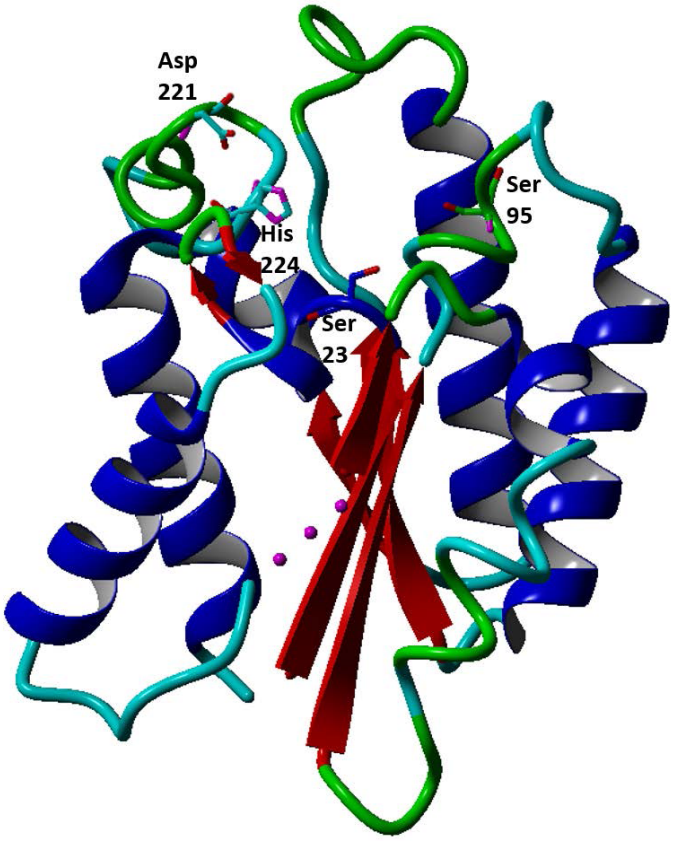


Figure 3.3 - Model of three-dimensional structure of the *M. hassiacum* homologue of Rv2418c as predicted by Phyre2. The protein was threaded onto the SGNH hydrolase lipase/acylhydrolase from *Enterococcus faecalis* (PDB entry: 1YZF). Possible catalytic residues are indicated and shown as sticks (with oxygen atoms in red, nitrogen in magenta). Represented in blue are predicted α-helices, in red β-strands, in green 3_{10}-helices and in cyan random coils, gaps in amino acid sequence which were not threaded are represented in purple. Figure prepared with YASARA view (Krieger and Vriend, 2014).

Recombinant expression and purification of OctT

The *M. tuberculosis*, *M. smegmatis* and *M. hassiacum* homologues were cloned and expressed in *E. coli*. Purification of soluble and bioactive recombinant OctTs required β-mercapthoethanol (7.5 mM) and Tween 20 (0.1% vol/vol) in all the buffers throughout the purification procedure. Recombinant OctT production was confirmed by Western immunoblot analysis and peptide mass fingerprinting (IPATIMUP Proteomics Unit), which indicated the presence of two forms of His-tagged OctT, one with the expected mass and another with much higher mass (Figure 3.4). Gel filtration experiments indicated that recombinant His-tagged OctTs could be separated into two populations: a subpopulation of molecules with a molecular mass of 111.1±6.9 kDa (*M. tuberculosis*), 126.6±6.9 kDa (*M. smegmatis*) or 125.2±6.9 kDa (*M. hassiacum*), compatible with a tetrameric architecture, and another fraction corresponding to a higher order oligomer, later shown to be inactive (see below). The three recombinant enzymes were purified to apparent homogeneity as determined by SDS-PAGE analysis (Figure 3.4) with yields of 1.2 mg (*M. tuberculosis*), 10 mg (*M. smegmatis*) or 3 mg (*M. hassiacum*) of OctT per liter of culture.

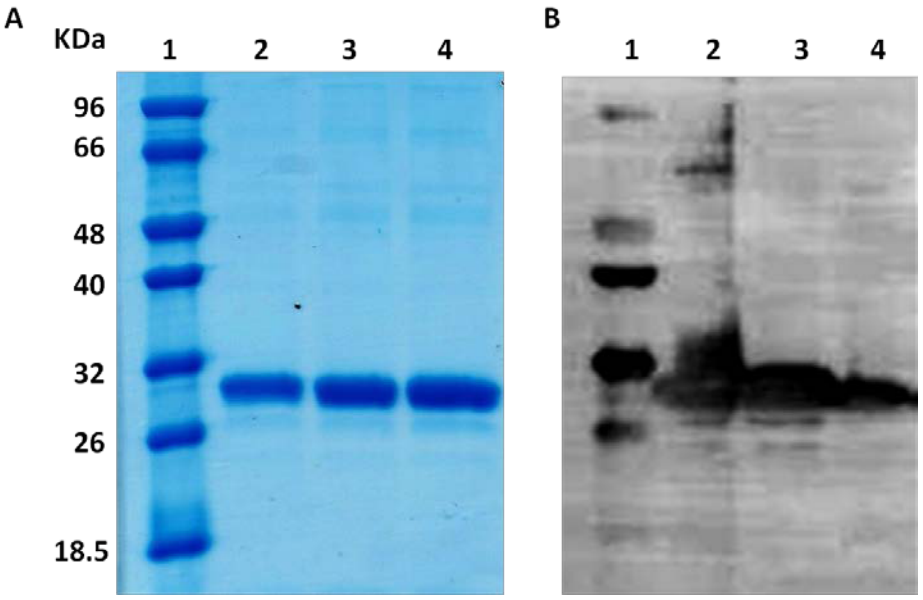


Figure 3.4 - Purification of mycobacterial OctTs. (A) SDS-PAGE analysis of purified recombinant proteins. Lane 1: molecular weight marker; lanes 2-4: purified recombinant OctT from *M. tuberculosis*, *M. smegmatis* and *M. hassiacum*, respectively. (B) Western immunoblot analysis using an anti-His_6 specific antibody.

Properties of recombinant OctT

The recombinant OctT from *M. hassiacum*, the most thermophilic species among known mycobacteria, was selected for characterization given its expectably higher stability. In accordance, it was observed that the OctT from *M. hassiacum* remained stable in solution and active even after 3 weeks on ice while the OctT from *M. smegmatis* progressively lost its activity after 2 weeks on ice. In contrast, *M. tuberculosis* OctT was highly unstable rapidly losing activity and precipitating only after 3 days.

Despite being native to the moderate thermophile *M. hassiacum* with an optimal growth temperature around 50°C in rich medium (Tiago *et al.*, 2012), the optimal temperature for OctT activity *in vitro* ranged from 37 to 45°C with maximal activity at 45°C and sharply decreasing above this temperature (Figure 3.5 A). At 37°C the enzyme was maximally active between pH 7.0 and 8.5 (Figure 3.5 B) and in the absence of added cations although EDTA slightly inhibited the activity (27% decrease), similar to the effect observed for most cations at 5 mM (Zn^{2+}> Mg^{2+}>Co^{2+}) or Li^{2+} at 1 mM (Figure 3.5 C).

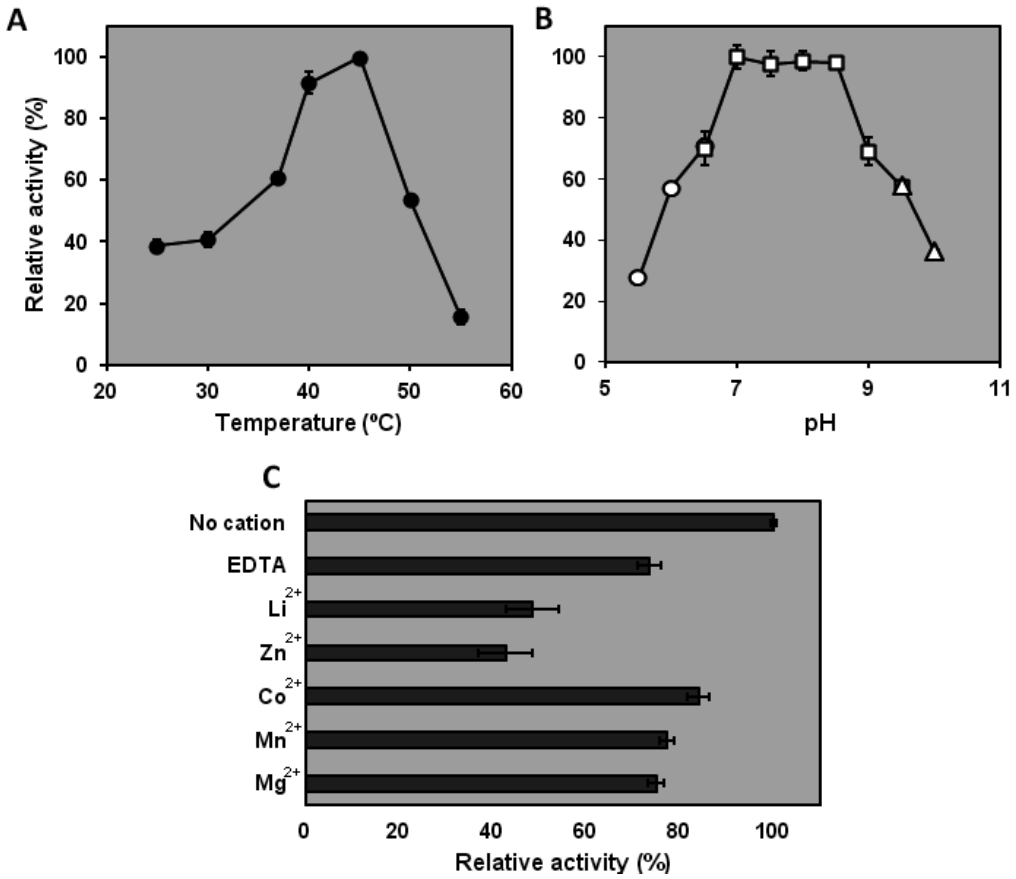


Figure 3.5 - *M. hassiacum* OctT properties. (A) Temperature profile. (B) pH dependence (circles, MES buffer; squares, BTP buffer; triangles, CAPS buffer). (C) Effect of divalent cations on OctT activity.

The *Mycobacterium hassiacum* OctT was very selective towards the acceptor sugars and from the array of compounds tested it could only efficiently use GG and DGG (Table 3-I A), with a detectable preference toward DGG (Table 3-II). We could only measure a significant transfer of ester groups from CoA- or *p*NP-activated octanoate and hexanoate while vestigial transfer of other acyl-CoAs could only be detected by mass spectrometry (described below) (Table 3-I A). For convenience, the synthetic *p*NP esters were used to characterize *M. hassiacum* OctT temperature profile, pH dependence and the effect of divalent cations while the kinetic parameters were examined with CoA-activated substrates hexanoyl-CoA (Hex-CoA) and octanoyl-CoA (Oct-CoA) as donors.

Table 3-I - Substrate specificity of *Mycobacterium hassiacum* OctT towards (A) acyl donor (B) acceptor substrate (C) acyl donors in combination with oligosaccharides mimicking specific motifs found in the mature MGLP structure.

A

Donor	Substrate	
	GG	**DGG**
Acetyl-CoA	-	**+**
Propionyl-CoA	-	**+**
Butyryl-CoA	-	**++**
Succinyl-CoA	-	ND
Hexanoyl-CoA	**++**	**+++**
Octanoyl-CoA	**++**	**+++**
Decanoyl-CoA	-	+
Dodecanoyl-CoA	-	ND
Tetradecanoyl-CoA	-	ND
Palmitoyl-CoA	-	ND
Octanoyl-OH *	-	ND

B

Substrate	Donor	Substrate	Donor
	Octanoyl-CoA		**Octanoyl-CoA**
DGG	**+++**	Trehalose§	+
GG	**++**	Maltose§	+
MGG	**+**	Isomaltose§	+
MG	ND	Sucrose§	ND
Glyceric acid	ND	Laminaribiose§	ND
GPG	ND	Maltotriose§	ND
MPG	ND	Maltotetraose	ND
Glucose§	ND	Maltopentaose	ND
Mannose§	ND	Glucosamine	ND
Galactose§	ND	Kanamycin	ND

C

Substrate	Donor				
	Acetyl-CoA	**Butyryl-CoA**	**Succinyl-CoA**	**Hexanoyl-CoA**	**Octanoyl-CoA**
DGG	**+**	**+**	ND	**+++**	**+++**
Trehalose	-	-	-	-	**+**
Maltose	ND	-	ND	-	**+**
Laminaribiose	-	-	ND	-	ND
Maltotetraose	ND	-	-	ND	ND
Maltopentaose	ND	.ND	-	ND	ND
Maltoheptaose	-	ND	-	ND	ND

ND, Not Detected. +, ++ and +++ indicate increasing amounts of product detected. Product levels + were detected by MS but were not quantifiable by enzymatic assays; Product levels ++ could be spectrophotometrically detected (-) Not tested. When available, *p*NP substrates were also used.

*Octanoyl-OH was also tested in combination with ATP with similar results.

§ Substrates were also tested in combination with glyceric acid with similar results.

However, because the kinetic constants for the *M. hassiacum* OctT determined at 37°C were very low (Table 3-II) it was deemed essential to assess if this was a characteristic of this particular enzyme or a property shared by other OctTs. Thus, relevant kinetic parameters were also calculated for the OctT from *M. smegmatis* (optimal growth at 37°C), which revealed significantly higher turnover and catalytic efficiencies (Table 3-II).

Table 3-II - Apparent kinetic parameters of *M. hassiacum* and *M. smegmatis* OctTs.

OctT	Substrate	Co-substrate fixed concentration	K_m (mM)	V_{max} (nmol.min^{-1}.mg^{-1})	k_{cat} (min^{-1})	k_{cat}/K_m (mM^{-1}. min^{-1})
Mh	DGG	Oct-CoA 100 µM	9.5±0.8	134±5	$0.1±3\times10^{-3}$	$0.01±10^{-3}$
Ms			20.0±3.6	1939±166	2.3±0.2	0.11±0.05
Mh	Oct-CoA	DGG 20 mM	0.06±0.01	162±14	0.12±0.01	1.9±0.4
Ms		DGG 45 mM	0.03±0.01	1898±427	2.2±0.5	86±29
Mh	GG	Oct-CoA 100 µM	6.6±0.78	32±2	$0.02±10^{-3}$	$3\times10^{-3}±4\times10^{-4}$
Ms			19.0±8.6	20±5	$0.02±6\times10^{-4}$	$0.12±2\times10^{-5}$
Mh	Oct-CoA	GG 20 mM	$0.01±2\times10^{-3}$	21±1	$0.01±7\times10^{-4}$	1.1±0.2
Ms		GG 45 mM	$0.02±2\times10^{-3}$	19±3	$0.02±3\times10^{-3}$	1.5±0.2
Mh	Hex-CoA	DGG 20 mM	$0.02±2\times10^{-3}$	59±2	0.04±0.01	2.1±0.6
	DGG	Hex-CoA 100 µM	9.8±2.3	41±4	$0.03±3\times10^{-3}$	$3\times10^{-3}±7\times10^{-4}$
	Hex-CoA	GG 20 mM	$0.01±2\times10^{-3}$	19±1	$0.01±6\times10^{-4}$	1.3±0.2

Mh and Ms represent *M. hassiacum* OctT and *M. smegmatis* OctT, respectively.

Both recombinant OctTs exhibited Michaelis-Menten kinetic behavior at 37°C with Oct-CoA or Hex-CoA up to a concentration of 100 µM (Figure 3.6). Higher concentrations of these ester donors in mixtures containing 20 mM or 45 mM GG or DGG were progressively inhibitory of enzyme activity (Figure 3.7).

Possible inhibitory effects of free CoA, free octanoate, Li^{2+} or of Oct-DGG product on enzyme activity were tested. Free CoA and free octanoate showed no effect on the reaction rate or amount of product formed while the addition of DGG-Oct or Li^{2+} was slightly inhibitory. However, the inhibition within the range of concentrations utilized for kinetic measurements was inconsequential which is illustrated with a 22% loss of activity in the presence of 400 µM of added product and 21% loss of activity in the precence of 3 µM Li^{2+} (equivalent to adding 280 µM Oct-CoA to the reaction, Figure 3.7 C-D). The OctTs studied did not exhibit detectable esterase activity towards Oct-DGG under the conditions tested (see Methods).

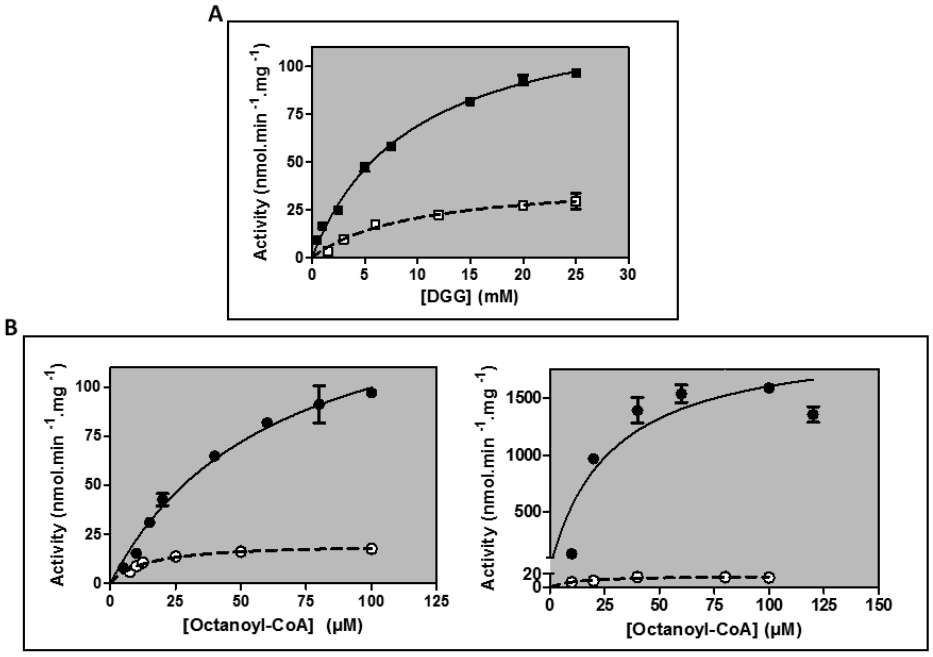


Figure 3.6 - Kinetic properties of recombinant OctT. (A) Plot of *M. hassiacum* OctT activity as a function of acceptor DGG concentration (Oct-CoA, filled squares and Hex-CoA, open squares). (B) Plot of *M. hassiacum* (left panel) or *M. smegmatis* OctT (right panel) activity as a function of donor Oct-CoA concentration (DGG, filled circles and GG, open circles).

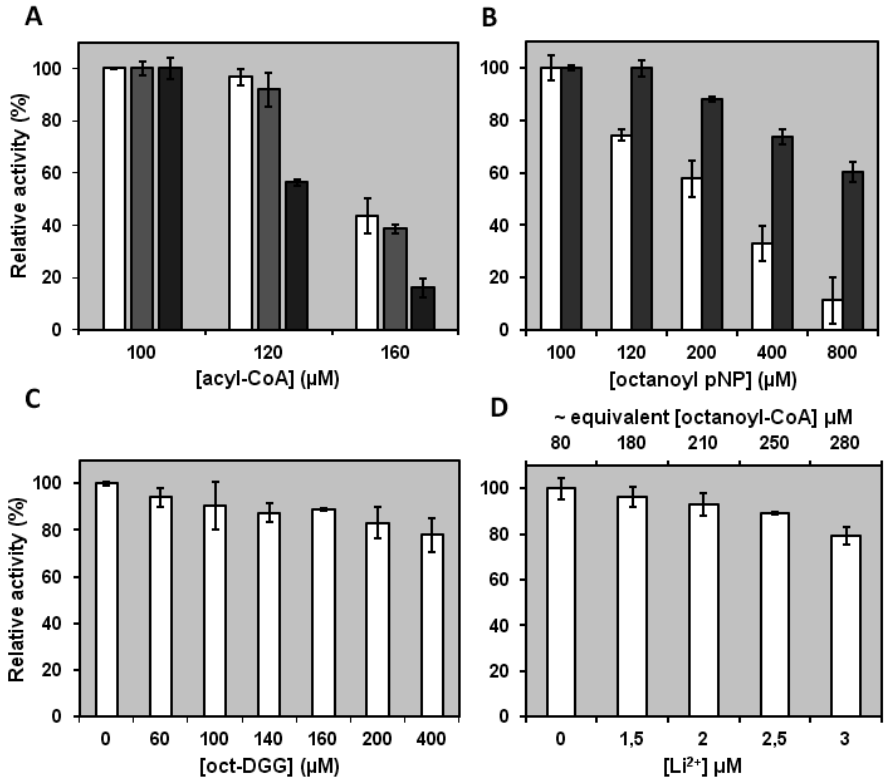


Figure 3.7 - Recombinant OctT concentration dependent inhibition by acyl ester donor or product. (A) Inhibition by Oct-CoA (white and grey bars) in *M. hassiacum* and *M. smegmatis*, respectively and Hex-CoA in *M. hassiacum* (black bars) in the presence of DGG (20 mM or 45 mM). (B) Inhibition by Oct-*p*NP in the presence of each of the preferred glycoside acceptors GG (white bars) and DGG (grey bars) (both substrates at 20 mM). (C) Inhibition by product DGG-Oct. (D) Inhibition by Li^{2+}.

Enzymatic and chemical product synthesis

The enzymatic synthesis of Oct-DGG was carried out at supra optimal concentrations of acyl donor (1.5 mM) in order to obtain substantial amounts of product for NMR data acquisition and product characterization (Figure 3.8). Efforts to purify the related OctT product octanoyl-GG were unsuccessful as only trace amounts could be obtained enzymatically. When using GG at higher acyl donor concentrations, the enzyme was inhibited and yielded insignificant product levels (Figure 3.7).

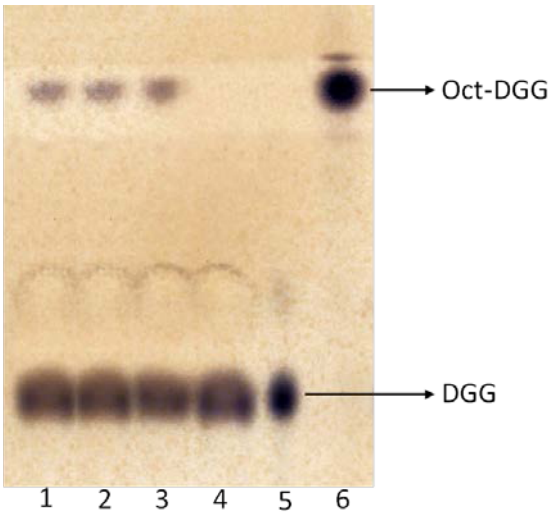


Figure 3.8 – Thin layer chromatography analysis of OctT reaction products. Lanes 1-3: Reaction mixtures with OctT recombinant from *M. tuberculosis, M. hassiacum* and *M. smegmatis*, respectively (50 mM BTP pH 8.0, 7 mM DGG, 1.5 mM Oct-activated donor incubated for 3 hours at 37°C, see methods). Lane 4: control reaction mixture without enzyme; lane 5: DGG standard lane 6: chemically synthesized Oct-DGG standard.

The synthetic route for 6-Oct-DGG **1** ((2*R*)-2-*O*-[6-*O*-octanoyl-(α-D-glucopyranosyl-(1→6)-α-D-glucopyranosyl]-2,3-dihydroxypropanoic acid), identical to the enzymatically obtained product and for DGG-3-Oct **2** ((2*R*)-2-*O*-(α-D-glucopyranosyl-(1→6)-α-D-glucopyranosyl)-3-*O*-octanoyl-2,3-dihydroxypropanoic acid) mimicking the previously reported MGLP reducing end, is described and illustrated in Maranha *et al.* (2015).

From the results obtained, it can be concluded that the proton NMR spectrum of the enzymatically obtained compound is identical to that of compound **1** (due to its purification process, the NMR spectrum of the enzymatic Oct-DGG has some impurities that appear as extra peaks) (Figure 3.9). We have also compared these NMR spectra to that obtained for the enzyme substrate DGG (Lourenço *et al.*, 2009). In DGG and compound **1**, the easily identifiable H7 chemical shift is very similar, whereas in compound **2** this signal appears downfield, indicating that an electron withdrawing group, such as the octanoate ester, is esterifying the vicinal hydroxyl group. Furthermore, the CH_2 group of the glycerate (two H8 signals – AB system) has shifted considerably downfield, confirming this assumption. With compound **1**, and because the octanoate is esterifying the free C-6 hydroxyl group, these signals (two H6 signals – AB system) are significantly shifted downfield.

The chemical synthesis of both compounds **1** and **2** was a fundamental resource to unequivocally confirm the structure of the natural Oct-DGG.

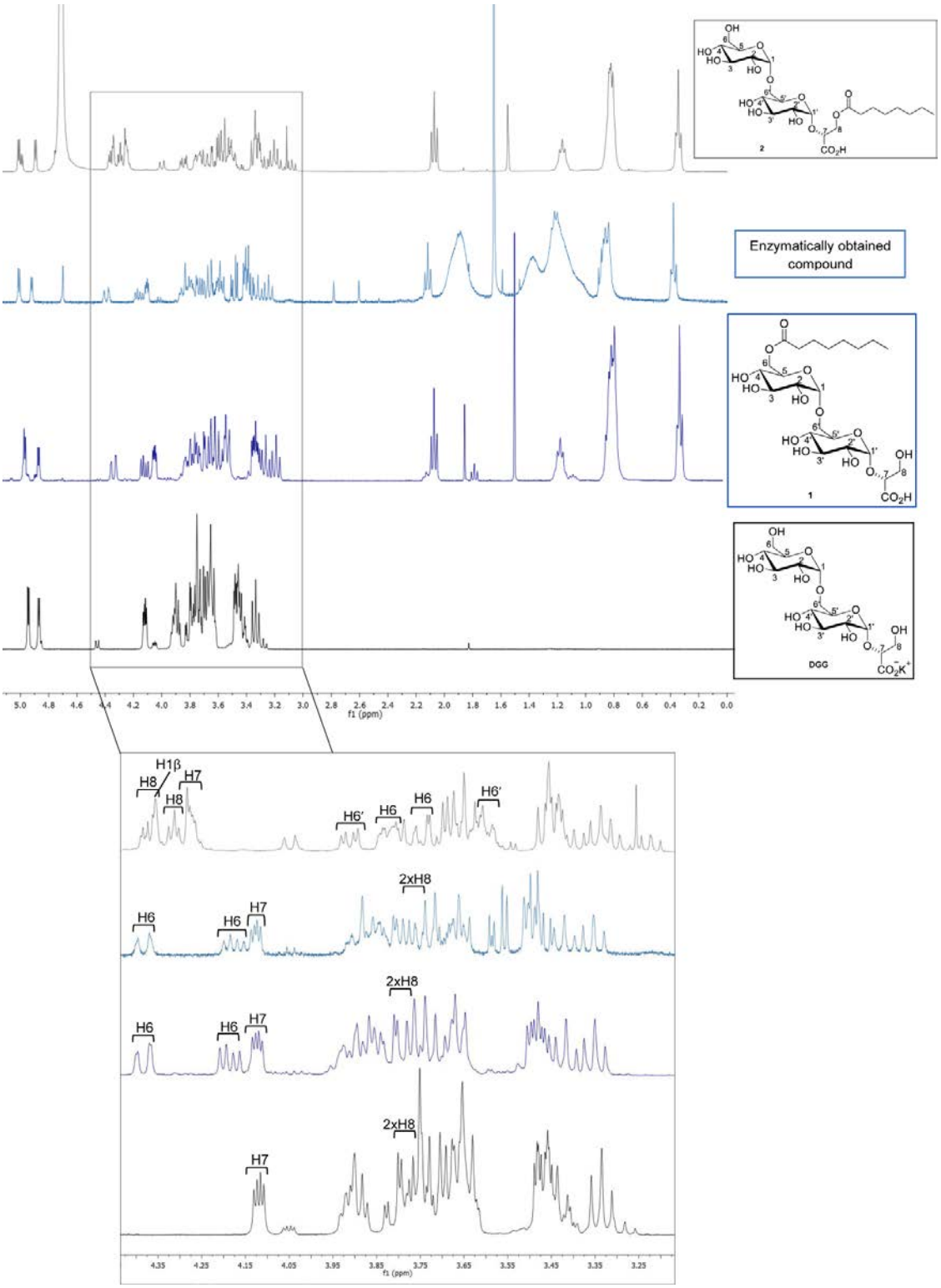


Figure 3.9 - ^{1}H-NMR spectra of natural Oct-DGG, chemical synthesized compounds **1** and **2** and DGG**.** Peak assignment was supported by 2D correlation NMR experiments (COSY and HMQC). Figure prepared by M. Rita Ventura at Instituto de Tecnologia Química Biológica (ITQB, Oeiras).

Mass spectrometry

To overcome the constraints inherent to monitoring product formation using a colorimetric reporter (either *p*NP or DTNB) and due to the limited sensitivity of enzymatic quantification with several of the acyl activated donors, direct observation of the product was sought using mass spectrometry (Figure 3.10). We found that product formation is consistent with those reactions where a rate of *p*NP or CoA release was quantifiable and for which enzyme kinetics could be analyzed (Table 3-II).

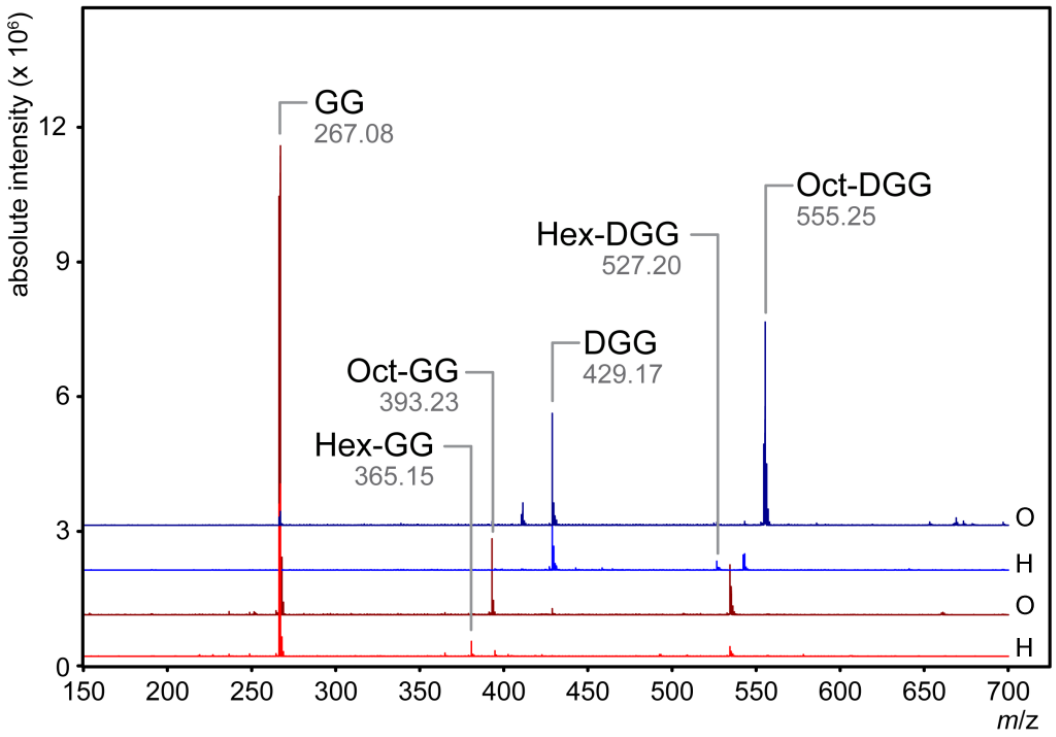


Figure 3.10 - Mass spectra of *M. hassiacum* OctT reaction products for which enzyme kinetics could be followed. DGG (blue spectra) and GG (red spectra) incubated in the presence of either Oct-CoA (O) or Hex-CoA (H). All unique peaks not present in control reactions are identified. Spectra were analyzed using the open-source mMass software package (Strohalm *et al.*, 2010) and all identified masses are the $[M-H]^{-1}$ ions. Figure prepared by Patrick J. Moynihan at University of Guelph, Canada.

Confirmation of the site of acyl modification was initially sought using tandem MS of enzymatically derived products. Unfortunately, relatively few peaks were observed that are unique to modified products and therefore a reconstruction of the product was not possible. With this in mind an NMR analysis was used to interrogate the major enzymatically derived product, which indicated it was consistent with compound **1** (Oct-DGG) that has the acyl modification on the C6 OH of the second glucose (Figure 3.9). The site of modification on several of the OctT reaction products was further validated by comparing the MS/MS fragmentation of the enzymatically-derived material with the fragmentation of the synthetic material (Figure 3.11) and found to be consistent with acyl modification on the C6 OH.

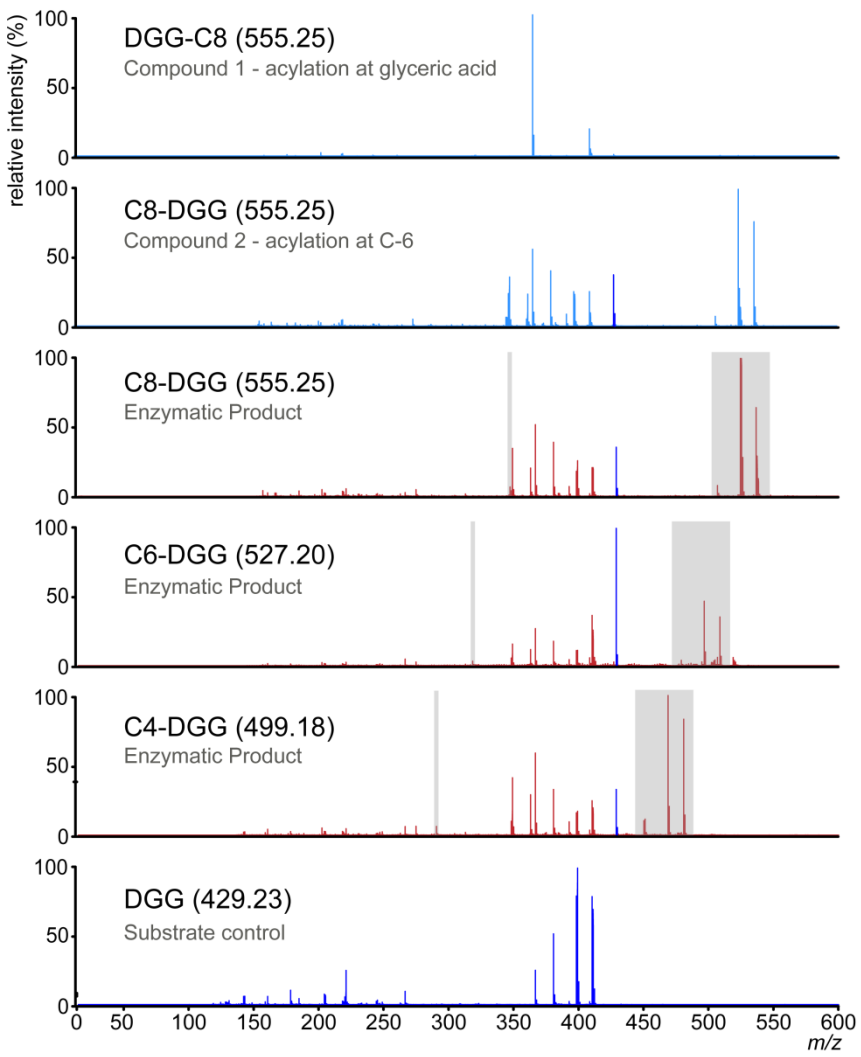


Figure 3.11 - MS/MS analysis of DGG and of its modified variants. A limited number of ions were observed which are unique to each modified reaction product and are shifted by 28 Da with the corresponding reduction in acyl-donor length (highlighted grey). Unmodified DGG, dark blue; compounds **1** and **2**, light blue; butyryl-CoA (C4),hexanoyl-CoA (C6) and octanoyl-CoA (C8) reaction products (red).Spectra were normalized and analyzed using the open-source mMass software (Strohalm *et al.*, 2010). Parental ions are indicated in parenthesis and are the $[M-H]^{-1}$ ions. Figure prepared by Patrick J Moynihan at University of Guelph, Canada.

Product could be detected with donors from acetyl-CoA (C2) to decanoyl-CoA (C10) but absent when donors with an acyl chain longer than C10 were used (Figure 3.12 and Table 3-I A). Relative ionization is consistent with the enzymatic rates observed using the kinetic assay, however quantitation of reaction products by MS was not possible without suitable standards. The acidic acyl substrate succinyl-CoA was also tested and product formation could not be detected with any acceptor utilized (Table 3-I A-C). Once a reconstruction of Oct-DGG product was possible by comparison to NMR analysis, we examined the site of modification for the most relevant acyl donors with DGG as acceptor and concluded that regardless of acyl chain size the modification is invariably present on the second glucose C6 OH position (Figure 3.11).

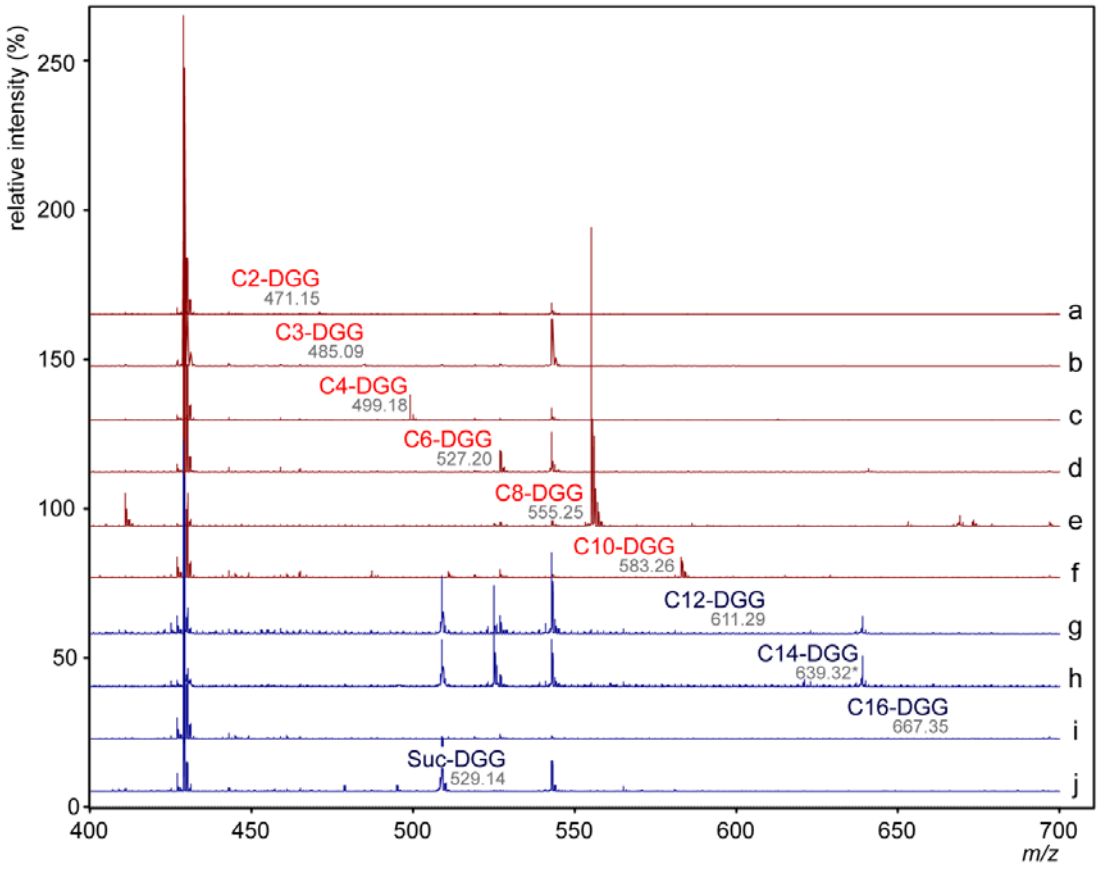


Figure 3.12 - OctT acyl-donor promiscuity. Reaction products after DGG incubation with acyl-CoA donnor including acetyl-CoA (C2, a); propionyl-CoA (C3, b); butyryl-CoA (C4, c); hexanoyl-CoA (C6, d); octanoyl-CoA (C8, e); decanoyl-CoA (C10, f); dodecanoyl-CoA (C12, g); tetradecanoyl-CoA (C14, h) and pamitoyl-CoA (C16, i) in addition to succinyl-CoA (j) in the presence of OctT from *M. hassiacum*. Mass spectra from reactions where transferase activity was observed are colored red while spectra lacking product are colored blue. Spectra were normalized and analyzed using the open-source mMass software package (Strohalm *et al.*, 2010) and all identified masses are the $[M-H]^{-1}$ ions. *In all cases except for C14 the spectra from control reactions lacked the observed product peaks. A peak at the expected size of the C14 reaction product was observed in both control reactions and the C12 spectra. Its presence in those spectra and the lack of product in the C12 and C16 reactions makes it unlikely that this ion represents an authentic DGG-C14 reaction product. Figure prepared by Patrick J. Moynihan at University of Guelph, Canada.

Previous studies anticipated that octanoylation position should occur in the primary hydroxyl group of the glyceric acid (Smith and Ballou, 1973; Dell and Ballou, 1983). After assigning the acylation position to the second glucose of DGG we wondered about the importance of glyceric acid (GA) in the acyl transfer reaction. A combination of substrates mimicking the preferred substrates and analogs were tested, including the combination isomaltose + GA mimicking DGG or glucose + GA mimicking GG. The disaccharides trehalose, maltose and isomaltose, isolated or in combination with glyceric acid, were substrates for the enzyme but no discernible difference was detected in the presence or absence of GA (Table 3-I B/C). Although product was detected with these disaccharides the reaction rates were extremely low and below the detection limit of the enzymatic assay.

In order to test donor specificity, an array of substrates were analyzed as shown in Table 3-I. Substrates analogous to the preferred acceptors [DGG (α–D–glucosyl–(1→6)–α–D–glucosyl–

(1→2)–D–glycerate) and GG (α–D–glucosyl–(1→2)–D–glycerate)] were examined and product formation was detected with MGG (α–D–mannopyranosyl–(1→2)–α–D–glucopyranosyl–(1→2)–D-glycerate) but not MG (α-D-mannosyl-(1→2)- D-glycerate) (Table 3-I A). Although we cannot exclude MGG as substrate this vestigial utilization is most likely due to trace amounts of GG in the MGG sample in result of degradation and not to real mannosyl substrate usage. MGG is an extremely rare compatible solute that so far has been identified in very few organisms, none of which Actinobacteria (Jorge *et al.*, 2007; Fernandes *et al.*, 2010).

The proposed involvement of OctT in the pathway for mature MGLP (polyacylated) led us to also test several possible acceptor substrates mimicking distinct parts of the polysaccharide that are often acylated with short-chain fatty acids (Mendes *et al.*, 2012). Maltooligosaccharides mimicking the MGLP non reducing end (maltotriose, maltotetraose, maltopentaose, maltoheptaose) were tested as acceptors in combination with the preferred octanoyl substrate as well as with activated acetyl and butyryl typically found in this MGLP region, but no product could be detected for any of the reactions (Table 3-I C). The disaccharide laminaribiose (Lam) mimicking the β-(1→3) branching units attached to the MGLP main chain was also tested to but no product could be detected with either Oct-CoA or Suc-CoA (succinyl is commonly found in these branching units) as donors (Table 3-I C). These results strongly indicate that the purified OctT can only efficiently use a very narrow range of glucosyl acceptors *in vitro*, which strengthens the hypothesis of a role in the specific acylation of the MGLP reducing end *in vivo* (Table 3-I).

Discussion

In this study, we identified a novel mycobacterial acyltransferase that efficiently transfers octanoate (comparable efficiency for hexanoate) to the sugar derivatives glucosylglycerate (GG) and diglucosylglycerate (DGG), the two earliest intermediates in methylglucose lipopolysaccharide (MGLP) biosynthesis (Kamisango *et al.*, 1987) forming its reducing end structure, reported to be octanoylated in the glycerate moiety (Smith and Ballou, 1973; Dell and Ballou, 1983). Since our preliminary MS results indicated that OctT did not transfer octanoate to the primary hydroxyl group of glycerate (Mendes *et al.*, 2012), we deemed essential to perform an exhaustive characterization of the product and found that octanoate is not transferred to the glycerate moiety as described for preparations obtained from *M. phlei* (Smith and Ballou, 1973), but instead to the C6 OH of the glucose moiety of DGG (Figure 3.9).

The primer for MGLP biosynthesis, glucosyl-3-phosphoglycerate (GPG), is formed from NDP-glucose and 3-phosphoglycerate (3-PGA) by glucosyl-3-phosphoglycerate synthase (GpgS) and dephosphorylated to GG by glucosyl-3-phosphoglycerate phosphatase (GpgP) (Empadinhas *et al.*, 2008; Mendes *et al.*, 2011) (Figure 3.13). Upon assignment of the second step of the MGLP

biosynthesis (GG formation) to the GpgP gene *Rv2419c*, its neighboring *Rv2417c* and *Rv2418c*, with predicted motifs of lipid-modifying enzymes, were hypothesized to participate in the pathway (Figure 3.1/ Figure 3.14) (Mendes *et al.*, 2011). The moderate level of amino acid identity (24%) between Rv2417c and the *Thermotoga maritima* DegV protein (TM841), whose three-dimensional structure was solved with an endogenously co-purified fatty acid chain, could suggest a role in the transport of activated acyl groups or involvement in acyl-CoA synthesis.

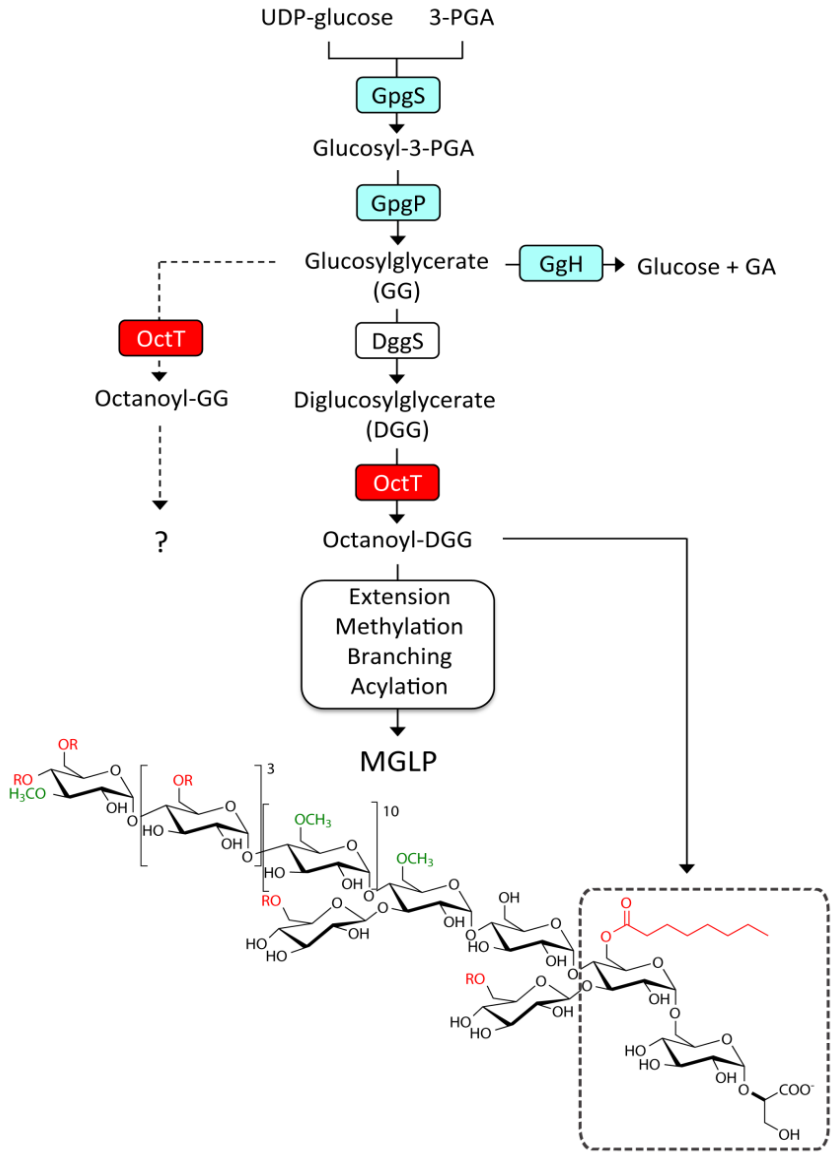


Figure 3.13 - Early steps of the proposed pathway for MGLP biosynthesis. Newly identified octanoyltransferase (OctT) is highlighted red. The dashed line indicates a hypothetical regulatory role of OctT. Experimentally validated functions are shaded blue. Open boxes represent unknown functions or those lacking biochemical confirmation. *R* groups (red) on the structure indicate acyl chains (acetate, propionate, isobutyrate or succinate) and methyl groups are in green. 3-PGA, D-3-phosphoglyceric acid. GpgS, glucosyl-3-phosphoglycerate synthase (Empadinhas *et al.*, 2008); GpgP, glucosyl-3-phosphoglycerate phosphatase (Mendes *et al.*, 2011); DggS, putative diglucosylglycerate synthase (Stadthagen *et al.*, 2007) GgH, glucosylglycerate hydrolase (almost exclusively detected in rapidly-growing mycobacteria) (Alarico *et al.*, 2014).

A high number of mycobacterial genes are incorrectly annotated, mainly due to the obstacles in performing functional studies arising from the often-difficult purification of the corresponding enzymes from native hosts or from recombinant sources sources in soluble bioactive form (Bashiri *et al.*, 2012; Mendes *et al.*, 2011). Although *Rv2418c* was annotated as a member of the GDSL hydrolase family (Akoh *et al.*, 2004) based on sequence similarities, this approach has previously proven unreliable for function assignment (Mendes *et al.*, 2010). GDSL esterases and lipases are hydrolytic multifunctional enzymes with a very flexible active site and broad substrate specificity. In contrast to other lipases, which have the highly conserved motif Gly-X-Ser-X-Gly in the middle of the aminoacid sequence, they posses a distinct GDSL sequence motif near de N-terminus (this particular sequence is not present in OctT, only DSL) and they share five blocks of highly conserved residues (I to V), which are important for their classification (Akoh *et al.*, 2004). The SGNH hydrolases are a subfamily of the GDSL esterases whose name derives from the peculiarity of having four signature invariant residues (Ser-Gly-Asn-His) found in the separate conserved blocks. Additionally, in the last conserved block (block V) there is usually an Asp residue three amino acids ahead of an His (i.e., DxxH) forming with Ser of block I the postulated catalytic triad required for activity in serine esterases (Ser/Asp /His). Because the catalytic aspartate in block V is not entirely conserved throughout the family (glutamate being one other residue found) it is not included in the family name (Akoh *et al.*, 2004).

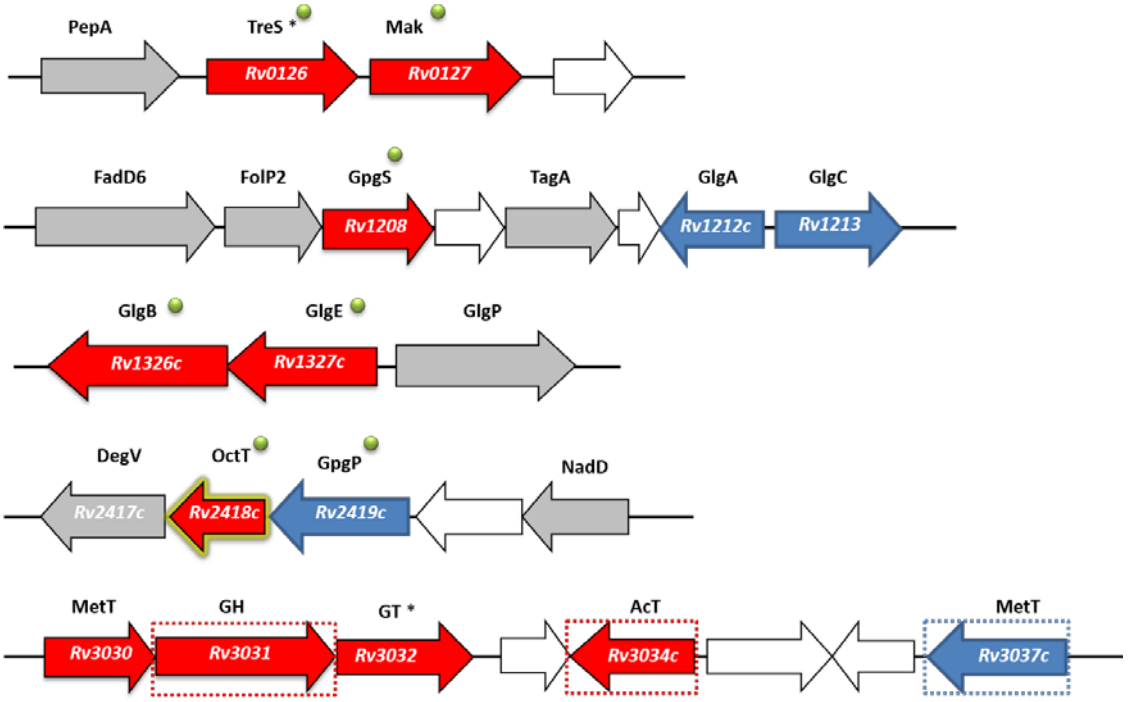


Figure 3.14 - *Mycobacterium tuberculosis* H37Rv gene clusters proposed to participate in MGLPs biosynthesis. The gene *Rv2418c* studied in this work is highlighted. Genes of unknown or hypothetical function are shaded white and grey, respectively. Genes considered essential for *M. tuberculosis* growth by saturation transposon mutagenesis (Griffin *et al.*, 2011b) are shaded red or blue if nonessential. *Although proposed to be essential for growth deletion of these genes produced viable mutants in *M. tuberculosis* H37Rv (Murphy *et al.*, 2005; Stadthagen *et al.*, 2007). Genes whose confirmation of involvement in MGLP pathway is still missing are boxed. Green spheres represent characterized proteins.

A peptidoglycan *O*-acetyltransferase (PatB) with the archetypal catalytic triad Asp-His-Ser of serine esterases annotated as a GDSL hydrolase displayed only weak esterase activity and was shown instead to be a bona-fide acetyltransferase (Moynihan and Clarke, 2010). Herein, we show that Rv2418c possesses acyltransferase activity, establishing a previously unrecognized function in mycobacteria, and displays no hydrolytic activity dispite having the postulated catalytic triad of serine esterases (as predicted by structural modeling Ser23, Asp221 and His224). The enzyme is very specific in the acceptor substrates it utilized, but presents some flexibility in the acyl donners used (Table 3-I) demonstrating some affinity with GDSL hydrolase which are notably promiscuous (Leščić Ašler *et al.*, 2010).

Based on genetic context and conserved domain analyses, the third step in the MGLP pathway (formation of DGG) was proposed to be catalyzed by Rv3031 (Stadthagen *et al.*, 2007). However, if the observed substrate ambiguity of OctT toward GG and DGG *in vitro* also occurs *in vivo*, octanoylation of GG in position 6 OH of glucose is expected to curb the progression of MGLP synthesis since this is the site of attachment of the second glucose in DGG and the first unit of the α-(1$\rightarrow$4)-linked MGLP main chain (Stadthagen *et al.*, 2007). Hypothetically, GG octanoylation *in vivo* could play a regulatory node in the pathway (Figure 3.13), possibly related to GG levels accumulated under certain conditions, for example nitrogen starvation. Interestingly, rapidly growing mycobacteria possess a specific GG hydrolase (GgH) activated upon replenishment of assimilable nitrogen to nitrogen-starved cells (Alarico *et al.*, 2014). However, the mechanism through which mycobacteria tune GG levels while regulating MGLP synthesis and function during exposure to stress remains undisclosed.

Octanoylation is a rare process in nature and several of the octanoyltransferases known to date are involved in fatty acid metabolism. LipB (EC 2.3.1.181) participates in the synthesis of lipoic acid, a highly conserved organosulfur cofactor derived from octanoic acid, which is essential in aerobic bacteria and eukaryotes (Spalding and Prigge, 2010). In *E. coli*, lipoate synthesis precedes through two reactions: the one catalyzed by the octanoyl transferase, LipB, and a second one catalyzed by a lipoate synthase, LipA. LipB transfers an octanoyl group from octanoyl-ACP to the apo form of multi enzyme complexes turning the octanoylated subunit into a substrate for the lipoate synthase, LipA, which subsequentially inserts two sulfur atoms to form the dithiolane ring of lipoate. LipB does not efficiently use free octanoate as a substrate and consequently is dependent on octanoyl-ACP however LipB can be bypassed by the lipoate ligase, LplA, which can use free octanoate (Spalding and Prigge, 2010). Mycobacteria possesses a similar system for lipoate synthesis except that it lacks the exogenous LplA pathway being entirely dependent on LipB (Ma *et al.*, 2006). This dependence combined with the fact that LipB was found to be up-regulated in patients with multidrug-resistant *M. tuberculosis* (Rachman *et al.*, 2006) as well as that it was considered essential for growth of the pathogen (Griffin *et al.*, 2011b) make it an attractive drug target. LipB crystal structure was solved in complex with decanoic acid. Such an interaction *in vivo* would lead to an irreversible inhibition of the enzyme with toxic, or even lethal, effects for *M. tuberculosis*, therefore it was hypothesized that this adduct could be an artifact of heterologous expression in E. coli, whose fatty acid synthesis pathway markedly differs from that of mycobacteria, this was confirmed by expressing the LipB protein in *M.*

smegmatis and observing no adduct formation (Ma *et al.*, 2006). Lip B is the only case known of spefic octanoyltransferases in mycobacteria.

In eukaria there are a few examples of octanoyltransferases, carnitine acyltransferases mediate the transport of fatty acid chains across different cellular compartments by catalyzing the reversible transfer of acyl groups between carnitine and coenzyme A (CoA). The family includes the peroxisomal carnitine *O*-octanoyltransferase (CrOT) (EC 2.3.1.137) responsible for the transfer of medium chain fatty acids to carnitine (Jogl *et al.*, 2004). Mammalian ghrelin *O*-acyltransferase (GOAT) is a membrane-bound enzyme that transfers octanoate to serine-3 of ghrelin, an appetite-stimulating peptide hormone that only displays activity upon acylation (Yang *et al.*, 2008). In yeast, two ethanol *O*-acyltransferases capable of transferring medium-chain fatty acids from CoA to ethanol have been implicated in the biosynthesis of flavor-active ethyl esters, including ethyl octanoate (Saerens *et al.*, 2006; Saerens *et al.*, 2008). Although these are all acyltransferases (EC 2.3.1.x), the specificity for the acceptor group greatly varies among them. This group of enzymes catalyzes the transfer of acyl groups to the hydroxyl, amino, or thiol group of a variety of acceptor molecules, ranging from protein modification to lipid, or sugar, modification, to yield an acyl ester derivative. They use high energy activated acyl-donors such as acylated-acyl carrier protein or most commonly acyl Coenzyme A, forming the different acyltransferase classes. (Röttig and Steinbüchel, 2013). In accordance, GOAT and LipB transfer the octanoyl moiety to a specific amino acid residue within a conserved sequence motif or protein domain, respectively, CrOT acylates an ammonium quaternary compound, and ethanol *O*-acyltransferases acylate an alcohol.

The OctT described in this work is, to the best of our knowledge, unique in its substrate specificity, since it is able to acylate the sugar derivatives GG and DGG, and to transfer short-to-medium chain acyl groups ranging from acetyl to decanoyl although products synthesized from the smaller (C2, C3) and larger (C10) acyl chains occurred only at minimal levels (Table 3-I).

The *M. smegmatis* OctT had a much higher apparent k_{cat} and k_{cat}/K_m than the *M. hassiacum* enzyme (Table 3-II). Very low turnover values (up to 6×10^{-4} min^{-1}) are known for enzymes referred to as “molecular switches” that signal and regulate shifts between states of activity for transient but defined periods (seconds, hours or longer) favoring accuracy over speed (Traut, 2008). While it is extremely premature to attribute a signaling role for OctT in the context of MGLP biosynthesis, it is also premature to exclude the contribution of unknown stabilizing factors, post-translational modifications, co-factors or interacting proteins that may translate into higher turnover and catalytic efficiency *in vivo*. One possible explanation for the apparent underperformance of *M. hassiacum* OctT in comparison to the *M. smegmatis* enzyme may reside on the temperature at which the kinetic constants were determined since we decided to study *M. hassiacum* OctT at 37°C, which is the human host temperature and the optimum for *M. smegmatis* growth but well below the native host’s optimal growth temperature (50°C) (Schröder *et al.*, 1997).

Despite the similar k_{cat}/K_m values for octanoyl and hexanoyl donors (Table 3-II), the latter has never been detected as a naturally occurring group in MGLP (Tung and Ballou, 1973; Keller and Ballou, 1968). Similar promiscuity *in vitro* was reported for GOAT, which is able to use a range

of donor acyl chains from acetate to tetradecanoic acid although octanoylated ghrelin is the predominant form of the enzyme *in vivo* followed by decanoylated ghrelin (Ohgusu *et al.*, 2012; Gutierrez *et al.*, 2008). *In vitro,* GOAT prefers hexanoyl-CoA over octanoyl-CoA while decanoyl-CoA is a poor substrate (Nishi *et al.*, 2013). CrOT also displays fairly broad donor substrate specificity but prefers C6 and C10 substrates (Jogl *et al.*, 2005). We propose that despite the low-level and apparently broad utilization of acyl-CoAs *in vitro*, OctT may be able to discriminate towards octanoyl groups *in vivo* possibly driven by Oct-CoA availability in the cytoplasm. Narumi and co-workers studied [^{14}C]acyl incorporation in *M. phlei* whole cells and found that when [^{14}C]hexanoate was supplied to the growth medium it was readily incorporated into MGLP (Narumi *et al.*, 1973). This allowed them to suggest that the acyl groups found in MGLP likely reflect their intracellular availability in mycobacterial cells at the site of MGLP acylation reactions and also that endogenous hexanoyl-CoA may not be available in the cytoplasm.

Early results by Tung and colleagues with *M. phlei* extracts suggested that distinct MGLP acylating reactions could be performed by a single enzyme, proposed to be capable of transferring acyl groups from different acyl-CoAs to MGLP and similar artificial acceptors (Tung and Ballou, 1973). However, the rate of incorporation from succinyl-CoA and octanoyl-CoA was significantly lower than from the acyl-CoAs carrying fatty acids normally esterified to the MGLP non-reducing end (acetate, propionate and isobutyrate), suggesting the involvement of more than one acyltransferase. In support of this hypothesis, we found that succinyl-CoA was not a substrate for OctT, nor did the enzyme efficiently use very short-chain fatty acids (C2-C4) found in the MGLP non-reducing terminus. Although none of the anticipated acyltransferase genes have been so far identified in mycobacterial genomes, genetic context analyses suggest that *Rv3034c* in the cluster containing the 6-*O*-methyltransferase (*Rv3030*) and the 4-*O*-glycosyltransferase (*Rv3032*) genes in *M. tuberculosis* H37Rv might be involved (Stadthagen *et al.*, 2007; Jackson and Brennan, 2009). Moreover, oligosaccharides mimicking regions of MGLP where acylations have been previously detected could also not be acylated by OctT using different-sized acyl-CoA donors (Table 3-I). Our results support differential acylation of the non-reducing end glucoses and of the branching glucoses in MGLP by enzymes other than OctT, which is also in agreement with the argument that acylation in such precise locations likely require highly specific acyltransferases (Gray and Ballou, 1972). Two theories regarding MGLP methylation have been put forward: one supports the existence of two different enzymes for differential 6-*O* and 3-*O* methylation of non-reducing end glucoses (Stadthagen *et al.*, 2007; Jackson and Brennan, 2009). Alternatively, 3-*O*-methylation was suggested to depend on previous acylation (Grellert and Ballou, 1972). Either because 3-*O*-methylation could depend on previous acylation or on a separate enzyme, an oligosaccharide with (four) unmethylated glucoses could be substrate for an acyltransferase acting on the non-reducing end segment of MGLP. However, and although several unmethylated glucoligosacharides were tested as possible OctT acceptor substrates, acylation from different donors was not detected.

The mycobacterial envelope possesses different surface-exposed acyltrehaloses, namely lipooligosaccharides (LOS) implicated in pathogen-host immune interactions (Nobre *et al.*, 2014). LOS often differ in the number and structure of sugar residues and the frequently

complex acyl moieties attached to trehalose (Kalscheuer and Koliwer-Brandl, 2014). Evidence suggests that LOS and other antigenic acyltrehaloses may also be esterified with octanoate (Daffé *et al.*, 2014; Besra *et al.*, 1994; McNeil *et al.*, 1987). However, and although we cannot fully exclude an involvement of OctT in these reactions, the insignificant transfer of octanoate to trehalose by OctT renders this a remote possibility (Table 3-I).

Herein we report the unprecedented OctT-catalyzed transfer of octanoate to DGG and GG *in vitro* and, to lesser extent, of other short-chain fatty acids. Despite the fact that we cannot fully exclude the existence of an enzyme capable to octanoylate the primary hydroxyl of glyceric acid estimated from earlier observations, OctT transfers octanoate to the second glucose in DGG, which may encompass important consequences during MGLP assembly. Since both MGLP and glycogen share a similar α-(1$\rightarrow$4)-linked backbone, the formation of octanoate-DGG may play a role in discriminating between homofunctional glycosyltransferases and recruiting Rv3032 (the glycosyltransferase proposed to be committed to elongation (Stadthagen *et al.*, 2007)) for the MGLP pathway, or the unknown glycosyltransferase for attachment of the first β-(1$\rightarrow$3) branching glucose to DGG. However, this remains hypothetical. Although it is premature to physiologically interpret the results reported here, the proposed essentiality of *Rv2418c* in *M. tuberculosis* renders the knowledge of OctT unique properties valuable in the quest for innovative strategies to fight mycobacterial infections.

Acknowledgements and work contributions

Chemical synthesis procedures and NMR analysis of the compounds were performed by Vanessa Miranda, Eva C. Lourenço and M. Rita Ventura at Instituto de Tecnologia Química Biológica, (ITQB, Oeiras).

Patrick J. Moynihan and Anthony J. Clarke provided extensive support with mass spectrometry methodologies, MS data analysis and designed the MS figures. Dyanne Brewer and Armen Charchoglyan at the Mass Spectrometry Facility of the Advanced Analysis Centre (University of Guelph, Canada) provided technical support with mass spectrometry. David Sychantha kindly analyzed a few samples.

MALDI mass spectrometry analysis were performed by Hugo Osório at the Proteomics Unit of the Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP).

This work was funded by national funds through Fundação para a Ciência e a Tecnologia (FCT) and by EU-FEDER through the Operational Competitiveness Programme – COMPETE with grants FCOMP-01-0124-FEDER-014321 (PTDC/BIA-PRO/110523/2009) and FCOMP-01-0124-FEDER-028359 (PTDC/BIA-MIC/2779/2012).

Chapter 4 Concluding remarks and future directions

Mycobacteria uniqueness derives to great extent from their extraordinary cell envelope rich in complex lipoglycans, glycolipids and other exceptional lipids, especially mycolic acids, which originate from fatty acids (FAs) (Daffé *et al.*, 2014; Pawełczyk and Kremer, 2014). Furthermore, mycobacteria also produce several distinctive intracellular polysaccharides, including two unique polymethylated polysaccharides (PMPS) of 3-*O*-methylmannose (MMPs) and of 6-*O*-methylglucose (MGLPs) that likely assume a helical conformation in the cytoplasm and are capable of forming 1:1 complexes with the products of Fatty Acid Synthase I (FAS-I) (Mendes *et al.*, 2012; Jackson and Brennan, 2009). This sequestering ability is believed to play a role in regulating FA metabolism (Bloch, 1977) although experimental evidence for this hypothesis is still missing.

Studies on MGLP started in the 1960's when much of the work towards its structural elucidation was preformed (Saier and Ballou, 1968b; Gray and Ballou, 1972; Smith and Ballou, 1973; Forsberg *et al.*, 1982). However, studies regarding its biogenesis were limited (Grellert and Ballou, 1972; Tung and Ballou, 1973; Kamisango *et al.*, 1987). The sequencing of *M. tuberculosis* genome (Cole *et al.*, 1998) and molecular biotechnology advances paved the way to deciphering the identity of the genes involved in this biosynthetic pathway. Recently MGLPs have been the focus of renewed attention because, given the apparently restricted distribution of the related polysaccharides MMPs to RGMs, they represent the only type of PMPS present in the slowly-growing mycobacteria usually associated with pathogens such as *M. tuberculosis* (Mendes *et al.*, 2012; Alarico *et al.*, 2014). In addition, some genes involved in its biosynthetic pathway have been considered essential for *M. tuberculosis* growth, (Griffin *et al.*, 2011b) making them potential drug targets.

In recent years a lot has been revealed about the biosynthetic pathway of MGLP. The essentiality of GpgS (glucosyl-3-phosphoglycerate synthase) for initiation of MGLP biosynthesis has been confirmed (Empadinhas *et al.*, 2008; Kaur *et al.*, 2009) and the enzyme GpgP (glucosyl-3-phosphoglycerate phosphatase) involved in the second step of the pathway has been uncovered (Mendes *et al.*, 2011). The identity of the enzymes responsible for elongation and methylation (Stadthagen *et al.*, 2007) of the polysaccharide have been established, alongside with the intricate intertwine of the MGLP pathway with other glucan pathways, namely for glycogen and capsular glucan (Stadthagen *et al.*, 2007; Sambou *et al.*, 2008). These three structures share a similar backbone of α-(1→4)-linked glucoses and the genes responsible for the formation of glycogen or capsular glucan backbone also seem to play a role in MGLP elongation (Stadthagen *et al.*, 2007; Sambou *et al.*, 2008).

Moreover, in addition to the suspected essentiality of MGLP in SGM, a link with the newly discovered TreS-Mak-GlgE-GlgB pathway, validated as essential in *M. tuberculosis*, has recently emerged (Kalscheuer *et al.*, 2010a) (

Figure 4.1). The essentiality of this pathway is related to the toxic effects in mycobacterial cells caused by excessive accumulation of the product of maltokinase (Mak), maltose-1-phosphate (M1P) (Kalscheuer *et al.*, 2010a) and the hypothetical synergy with MGLP biosynthesis may expose new strategies for therapeutic intervention.

Of the four main enzymes involved in the abovementioned pathway, three (TreS-GlgE-GlgB) had their three dimensional structures solved, Mak being the conspicuously missing structure, reason why we deemed essential its elucidation. Recombinant Mak^{Mtb} was produced, purified and unsuccessfully screened for crystallization. As *M. tuberculosis* proteins are often notably unstable (Edwards *et al.*, 2012) the *M. vanbaalenii* orthologue (59% identical to Mak^{Mtb}), shown to have maltokinase activity, was proved to be active and more stable and thus it was selected for a comprehensive biochemical and structural characterization.

Surprisingly, Mak^{Mvan} displays a typical eukaryotic protein kinase-like fold (ELK) uncovering a new family of ELKs with an additional regulatory N-terminal subdomain unique to maltokinases. The protein revealed an elongated concave-shaped bilobal structure with an exposed catalytic cleft. Site-directed mutagenesis in key conserved residues predicted the existence of highly conserved regions between Mak^{Mvan} and Mak^{Mtb} namely among the P-loop (containing the residues interacting with the terminal γ-phosphate), the catalytic loop and the DFE motif (magnesium-coordinating motif), as well as in the key residues forming the maltose-binding pocket. In accordance, the structure of *M. tuberculosis* Mak has been recently solved (Li *et al.*, 2014) and overall the key catalytic elements and main structural features are conserved when compared to *M. vanbaalenii* Mak and the unique N-terminal domain was also a shared feature, as predicted by the alignment of mycobacterial Mak sequences.

A hetero-octameric complex has been proposed to occur between TreS/Mak and to significantly enhance the activity of Mak^{Mtb} (Roy *et al.*, 2013). Because Mak's unusual cap N-terminal has a predicted flexibility compatible with regulatory functions it might regulate its phosphotransfer activity, providing a possible explanation for the observed enhancement of maltokinase activity upon binding to TreS.

Mak^{Mtb} was the first mycobacterial Mak to be characterized and like the Mak^{Mvan} studied in this work it was reported to be a monomer (Mendes *et al.*, 2010). During the elaboration of this thesis, Mak^{Mtb} structure was determined and found to occur in a dimeric form, representing only one of the conformations assumed by the protein, which was also found as bioactive monomers and tetrameric oligomers (Li *et al.*, 2014). The authors hypothesized that Mak^{Mtb} undergoes signal-independent homotypic dimerization mediated by the N-lobe and they built a *in silico* model for the interaction of MaK with TreS with the N-terminus of Mak placed in proximity of the C-terminus of TreS, as it would be in the organisms that present the the naturally fused bifunctional gene. In this model, a dimer of Mak fits on each face of a TreS tetramer and the TreS-MaK complex presents aligned active sites in a manner that facilitates the flow of product. Hypothetically, the accumulation of toxic M1P could be prevented through the assemblage of a complex between TreS, MaK and GlgE in which a dimer of GlgE could possibly fit on the unoccupied face of the MaK dimer giving rise to a super complex with the putative arrangement GlgE dimer – MaK dimer – Tres tetramer – MaK dimer – GlgE dimer. However, biochemical evidence for such a TreS-MaK-GlgE complex in any oligomeric organization is still missing. Auspiciously, the elucidation of Mak structure takes us one step closer to the rational design of new anti-tuberculous compounds thus contributing towards better tools to fight one of the most problematic re-emerging infectious diseases in the world.

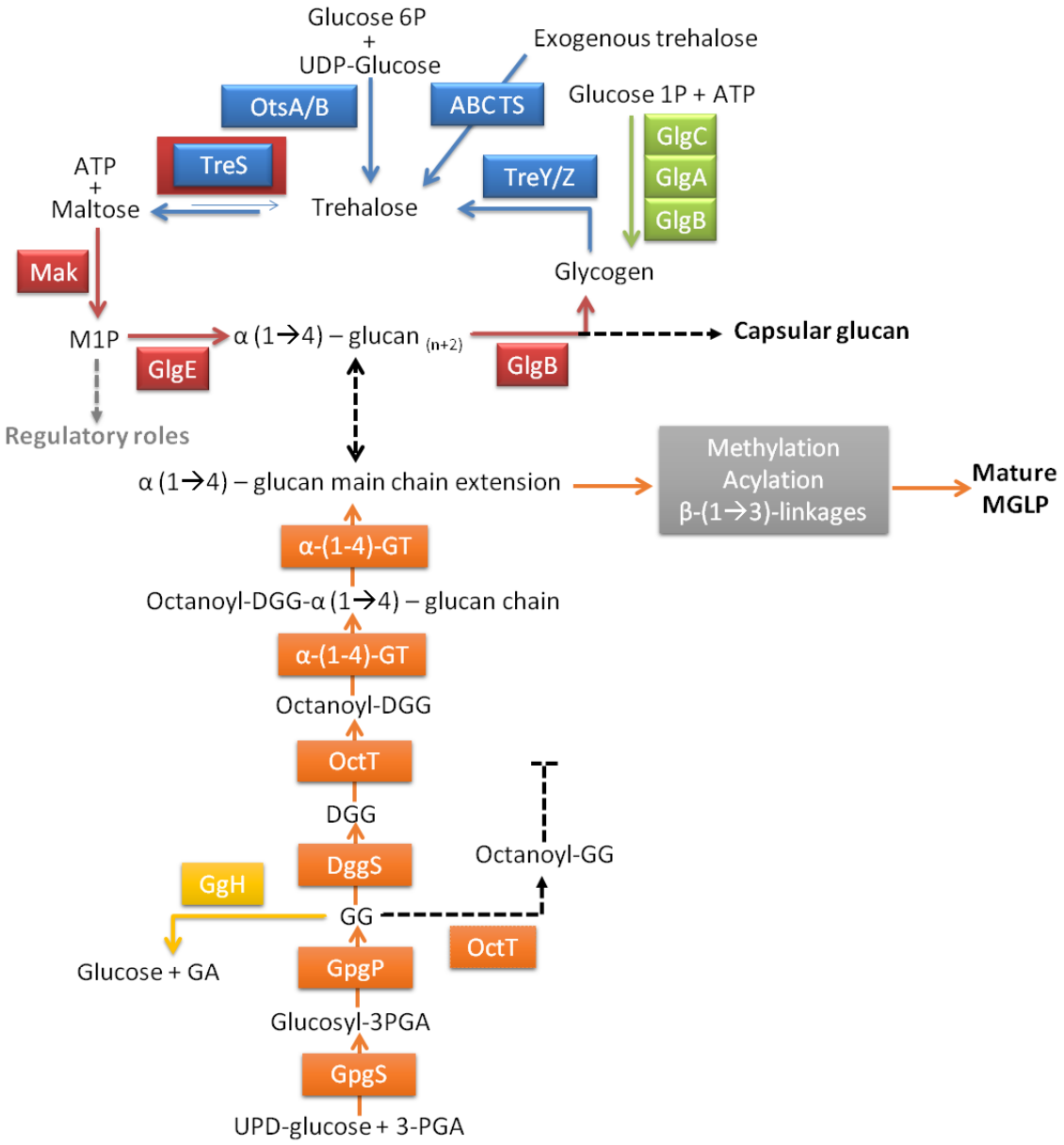


Figure 4.1 - Intersection of pathways involved in α-glucan biosynthesis in mycobacteria. Enzymes involved in the trehalose pathways are boxed in blue, enzymes involved in the newly discovered GlgE pathway are boxed in red. Glycogen pathway in green. MGLP pathway in orange. GgH pathway only present in RGM represented in yellow. Tentative assignment of OctT regulatory function and other postulated functions represent by dashed lines. The extension and methylation reactions are proposed to occur concomitantly (Kamisango *et al.*, 1987) catalyzed respectively by a α-(1→4) glycosyltransferase (*Rv3032* or *Rv1212c*) and by SAM-dependent methyltransferases (Stadthagen *et al.*, 2007). Branching and acylation steps are unknown. 3-PGA, D-3-phosphoglyceric acid; GA, glyceric acid; ABC TS, ABC transporter system.

In recent years, the increasing numbers of TB incidence and the growing resistance to the available anti-TB therapies rendered MGLPs a target of renewed interest uncovering new details regarding the biosynthesis of these important lipopolysaccharides. After assigning the glucosyl-3-phosphoglycerate phosphatase (GpgS) function to the *Rv2419c* gene the functional identification of gene *Rv2418c* as an acyltransferase involved in MGLP biosynthesis was particularly striking because of its annotation as a putative GDSL esterase. With this work we further unveil MGLP complex biogenesis through the identification of an enzyme we believe to be responsible for the 4th biosynthetic step in the pathway. The enzyme is capable of octanoylating the primers for MGLP assembly, GG and DGG, tough, in contrast to what would be expected from previous studies, the transfer of octonoate *in vitro* does not occur to the primary OH in the glyceric acid of these glycosides (Tung and Ballou, 1973) but to the C6-OH of glucose Figure 4.2 panels 2/4) . Given that the linkage of the second glucose in DGG is α-(1→6), in case octanoylation of the intermediate GG occurs *in vivo* it would thwart elongation of the α-(1→4) MGLP main chain. This hypothetical reaction could represent a regulatory function of MGLP assembly *in vivo* that might play a role in mycobacteria lacking GG hydrolase (GgH) which was also recently shown to specifically hydrolyze the MGLP primer GG and hypothesized to have a role in fine-tunning MGLP assembly during specific growth conditions or stress (Alarico *et al.*, 2014). However, the regulatory events coordinating MGLP biosynthesis have not been investigated.

Based on the work presented here and the studies of Tung and Ballou (Tung and Ballou, 1973) we propose that at least three enzymes are involved in MGLP acylation, one for the non-reducing end acylation (transferring acetate, propionate and isobutyrate), another for the addition of succinyl groups to the β-(1→3) glucose branches and a third for octanoylation of the reducing end MGLP, catalyzed by Rv2418c as proposed in this work. Genetic context analyses suggests that the putative acetyltransferase gene *Rv3034c* located in close proximity of a cluster containing *Rv3030* and *Rv3032* is a suitable candidate to carry out one of the remaining MGLP acylation steps (Stadthagen *et al.*, 2007).

The role of acylation is largely unknown but the short chain neutral acyl groups are expected, like methylation, to enhance the nonpolar interaction of the polysaccharide with its host fatty acid adding hydrophobicity to the inner coil cavity making them pivotal in the polysaccharide's complexing ability (Hindsgaul and Ballou, 1984). Acylation has also been suggested to regulate the levels and occurrence of either 6-*O* or 3-*O* methylation (Grellert and Ballou, 1972). The roles suggested for octanoylation include stabilizing the polysaccharide in helical conformation. In the absence of bound fatty acyl-CoA octanoate can form an axis around which MGLP folds, raising the interaction energy required for the displacement of the octanoyl group and complexation with longer fatty acyl-CoAs thus conferring to MGLP further discriminatory capability when compared to MMP (Hindsgaul and Ballou, 1984). The fact that octanoate is linked to the second glucose of DGG could confer to this ester group a privileged position around which the α-(1→4) main chain could be extended. The octanoate group was also suggested to anchor MGLP intermediates to the cytoplasmic membrane during the elongation steps (Smith and Ballou, 1973).

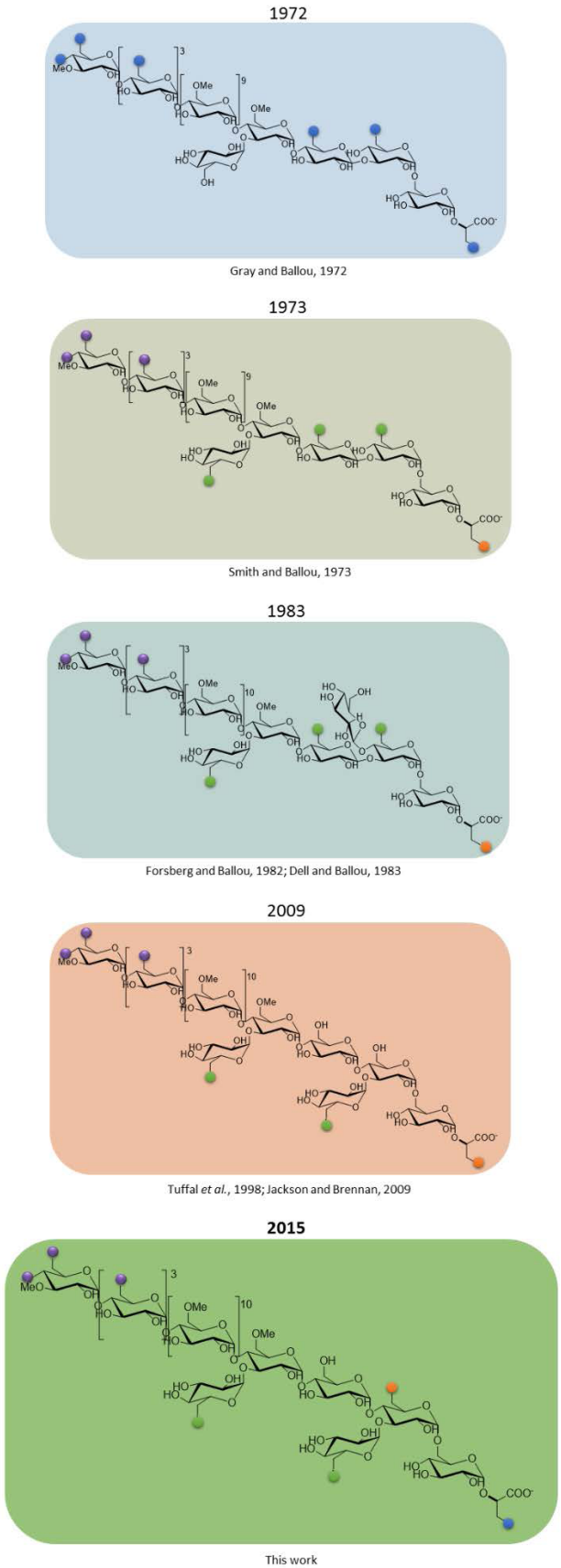


Figure 4.2 - Evolution of MGLP structure and acylation position over time. Blue circles represent any of the acyl groups detected in MGLP (acetate, propionate, isobutyrate, succinate or octanoate); purple circles represent acetate, propionate or isobutyrate; green circles represent succinate and orange circles represent octanoate. (Gray and Ballou, 1972; Smith and Ballou, 1973; Forsberg *et al.*, 1982; Dell and Ballou, 1983; Tuffal *et al.*, 1998). The forth panel represents the classically accepted structure of MGLP from *M. bovis* BCG (Tuffal *et al.*, 1998; Jackson and Brennan, 2009). The bottom panel represents the structure of MGLP with the proposed octanoylation position as indicated by the present study.

However, the octanoyl group is a much smaller chain when compared to the ester groups that normally anchor proteins or other compound to the membrane. The most common outcome of lipid modification in a protein is an increased affinity for membranes, nonetheless when proteins are modified with myristate (C14, the smallest ester that modifies proteins) its major function is only to assist in directing the modified protein to membranes but not being sufficiently hydrophobic to stably anchor it to a lipid bilayer (Resh, 2013). Consequently, this hypothetical anchorage of MGLP-bound octanoate to the membrane would be very transient, progressively detaching with the growing of the MGLP chain.

Since the early work conducted at Clinton Ballou's laboratory on MGLP structure (Saier and Ballou, 1968b; Gray and Ballou, 1972; Smith and Ballou, 1973; Forsberg *et al.*, 1982) the core structure has been revised twice. Once for the addition of a side chain glucose unit (Forsberg *et al.*, 1982) (Figure 4.2, panel 3) and latter, in 1998, leading to the currently accepted structure that defines the side chain glucose linkage to β-(1→3) (Figure 4.2, panel 4) (Tuffal *et al.*, 1998). However, in these studies the acylation positions were not re-analyzed and the most recent information about acylation dates back to studies conducted in 1983 (Figure 4.2 panel 3) (Dell and Ballou, 1983; Tung and Ballou, 1973). Based on previous studies and in our *in vitro* results we propose that MGLP is likely octanoylated in the second glucose of the MGLP backbone *in vivo* but we cannot exclude a possible additional acylation site occurring in the glyceric acid moiety (Figure 4.2, last panel). To establish the acylation positions on the naturally occurring and heterogeneously acylated forms of MGLP additional studies are mandatory.

Although the association of a novel gene to the biosynthetic route of MGLP discloses an additional function in the *M. tuberculosis* genome and provides new clues into MGLP acylation, the preliminary studies on this enzyme reported here must be transposed to a physiological context to confirm OctT involvement in MGLP biosynthesis *in vivo*. Because the *Rv2418c* gene has been proposed to be essential for *M. tuberculosis* growth (Griffin *et al.*, 2011b) and homologues are absent from eukaryotes, it represents a good candidate for future drug discovery efforts.

By presumably forming complexes with neo synthesized fatty acyl-CoAs and interacting with FAS-I, PMPSs may serve as storage and FAs carriers allowing mycobacteria to produce the levels required to build different cellular components, namely phospholipids, glycolipds and the cell envelope MAs indispensable for mycobacterial pathogenesis (Barry III *et al.*, 1998). Although disruption of the MGLP pathway has so far been unsuccessful in unsettling FAs or MAs metabolism, such mutagenesis results were nonetheless crucial to uncover important clues linking MGLPs biosynthesis and mycobacterial adaptation to thermal stress (Stadthagen *et al.*, 2007; Jackson and Brennan, 2009). The precise mechanism by which MGLP aids thermal adaptation remains unclear but hypothetically they could provide a certain degree of plasticity necessary to maintain the cell envelope homeostasis essential for survival under challenging environmental conditions, with expected implications also for pathogenesis. Hence, the elucidation of additional biosynthetic steps in this important pathway and its interconnection

with other metabolic routes warrants further understanding of those mechanisms placing us closer to devise new strategies to fight mycobacterial infections.

Several questions remain unanswered not only about the MGLP biosynthetic pathway but also in its interconnection with the GlgE pathway and regarding the mechanistic flow of intermediates. Nevertheless, the findings reported here on the elucidation of Mak structure and the functional characterization of a novel octanoyltransferase add new insights into a intricate and vital mycobacterial pathway, which may inspire new drug discovery strategies.

References

References

Akoh, C.C., Lee, G.C., Liaw, Y.C., Huang, T.H., and Shaw, J.F. (2004). *GDSL family of serine esterases/lipases*. Progress in lipid research **43** (6): p. 534-552.

Alarico, S., Costa, M., Sousa, M.S., Maranha, A., Lourenco, E.C., Faria, T.Q., Ventura, M.R., and 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**: p. 6766.

Armarego, W.L.F., and Chai, C.L.L. (2003). *Purification of Laboratory Chemicals*, 5th edn (Butterworth-Heinemann: Burlington).

Ascenzi, J.M., and Vestal, J.R. (1979). *Regulation of fatty acid biosynthesis by hydrocarbon substrates in Mycobacterium convolutum*. Journal of Bacteriology **137** (1): p. 384-390.

Axelrod, S., Oschkinat, H., Enders, J., Schlegel, B., Brinkmann, V., Kaufmann, S.H.E., Haas, A., and Schaible, U.E. (2008). *Delay of phagosome maturation by a mycobacterial lipid is reversed by nitric oxide*. Cellular Microbiology **10** (7): p. 1530-1545.

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

Bai, M., Ni, J., Shen, S., Wu, J., Huang, Q., Le, Y., and Yu, L. (2014). *Two newly identified sites in the N-terminal regulatory domain of Aurora-A are essential for auto-inhibition*. Biotechnology Letters **36** (8): p. 1595-1604.

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

Barry III, C.E., Lee, R.E., Mdluli, K., Sampson, A.E., Schroeder, B.G., Slayden, R.A., and Yuan, Y. (1998). *Mycolic acids: structure, biosynthesis and physiological functions*. Progress in Lipid Research **37** (2–3): p. 143-179.

Bashiri, G., and Baker, E.N. (2015). *Production of recombinant proteins in Mycobacterium smegmatis for structural and functional studies*. Protein Science **24** (1): p. 1-10.

Bashiri, G., Perkowski, E.F., Turner, A.P., Feltcher, M.E., Braunstein, M., and Baker, E.N. (2012). *Tat–Dependent Translocation of an F_{420}–Binding Protein of Mycobacterium tuberculosis*. PLoS ONE **7** (10): p. e45003.

Bates, M., Marais, B.J., and Zumla, A. (2015). *Tuberculosis Comorbidity with Communicable and Noncommunicable Diseases*. Cold Spring Harbor Perspectives in Medicine: p. a017889.

Berg, S., Kaur, D., Jackson, M., and Brennan, P.J. (2007). *The glycosyltransferases of Mycobacterium tuberculosis—roles in the synthesis of arabinogalactan, lipoarabinomannan, and other glycoconjugates*. Glycobiology **17** (6): p. 35R-56R.

Bergeron, R., Machida, Y., and Bloch, K. (1975). *Complex formation between mycobacterial polysaccharides or cyclodextrins and palmitoyl coenzyme A*. Journal of Biological Chemistry **250** (4): p. 1223-1230.

Besra, G.S., Khoo, K.-H., Belisle, J.T., McNeil, M.R., Morris, H.R., Dell, A., and Brennan, P.J. (1994). *New pyruvylated, glycosylated acyltrehaloses from Mycobacterium smegmatis strains, and their implications for phage resistance in mycobacteria*. Carbohydrate Research **251**: p. 99-114.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L*., et al.* (2014). *SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information*. Nucleic Acids Research **42** (W1): p. W252-W258.

Bloch, K. (1975). *Fatty acid synthases from Mycobacterium phlei*. In Methods in Enzymology Lipids Part B, M.L. John, Editor Academic Press: p. 84-90.

Bloch, K. (1977). *Control Mechanisms for Fatty Acid Synthesis in Mycobacterium Smegmatis*. In Advances in Enzymology and Related Areas of Molecular Biology, A. Meister, Editor John Wiley & Sons, Inc.: Hoboken, NJ p. 1-84.

Bloch, K., and Vance, D. (1977). *Control mechanisms in the synthesis of saturated fatty acids*. Annual Review of Biochemistry **46**: p. 263-298.

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

Brennan, P.J. (2003). *Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis*. Tuberculosis **83** (1–3): p. 91-97.

Brites, D., and Gagneux, S. (2015). *Co-evolution of Mycobacterium tuberculosis and Homo sapiens*. Immunological Reviews **264** (1): p. 6-24.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K*., et al.* (2002). *A new evolutionary scenario for the Mycobacterium tuberculosis complex*. Proceedings of the National Academy of Sciences **99** (6): p. 3684-3689.

Burk, D.L., Hon, W.C., Leung, A.K.W., and Berghuis, A.M. (2001). *Structural Analyses of Nucleotide Binding to an Aminoglycoside Phosphotransferase*. Biochemistry **40** (30): p. 8756-8764.

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

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

Chakroborty, A. (2011). *Drug-resistant tuberculosis: an insurmountable epidemic?* Inflammopharmacology **19** (3): p. 131-137.

Chandra, G., Chater, K.F., and Bornemann, S. (2011). *Unexpected and widespread connections between bacterial glycogen and trehalose metabolism*. Microbiology **157** (Pt 6): p. 1565-1572.

Chen, W., Biswas, T., Porter, V.R., Tsodikov, O.V., and Garneau-Tsodikova, S. (2011). *Unusual regioversatility of acetyltransferase Eis, a cause of drug resistance in XDR-TB*. Proceedings of the National Academy of Sciences **108** (24): p. 9804-9808.

Choi, K.-H., Kremer, L., Besra, G.S., and Rock, C.O. (2000). *Identification and Substrate Specificity of β-Ketoacyl (Acyl Carrier Protein) Synthase III (mtFabH) from Mycobacterium tuberculosis*. Journal of Biological Chemistry **275** (36): p. 28201-28207.

Chopra, I., and Brennan, P. (1998). *Molecular action of anti-mycobacterial agents*. Tubercle and Lung Disease **78** (2): p. 89-98.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd*, et al.* (1998). *Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence*. Nature **393** (6685): p. 537-544.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G*., et al.* (2013). *Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans*. Nature Genetics **45** (10): p. 1176-1182.

Comas, I., and Gagneux, S. (2009). *The Past and Future of Tuberculosis Research*. PLoS Pathogens **5** (10): p. e1000600.

da Silva, P.E.A., Von Groll, A., Martin, A., and Palomino, J.C. (2011). *Efflux as a mechanism for drug resistance in Mycobacterium tuberculosis*. FEMS Immunology and Medical Microbiology **63** (1): p. 1-9.

Daffé, M., Crick, D.C., and Jackson, M. (2014). *Genetics of Capsular Polysaccharides and Cell Envelope (Glyco)lipids*. In Molecular Genetics of Mycobacteria, Second Edition, G. Hatfull, and W.R. Jacobs, Jr. , Editors. American Society of Microbiology Press: Washington DC p. 559-609.

De Smet, K.A., Weston, A., Brown, I.N., Young, D.B., and Robertson, B.D. (2000). *Three pathways for trehalose biosynthesis in mycobacteria*. Microbiology **146** (Pt 1): p. 199-208.

Decker, K., Gerhardt, F., and Boos, W. (1999). *The role of the trehalose system in regulating the maltose regulon of Escherichia coli*. Molecular Microbiology **32** (4): p. 777-788.
Dell, A., and Ballou, C.E. (1983). *Fast-atom-bombardment, negative-ion mass spectrometry of the mycobacterial O-methyl-D-glucose polysaccharide and lipopolysaccharides*. Carbohydrate Research **120**: p. 95-111.
Dey, B., and Bishai, W.R. (2014). *Crosstalk between Mycobacterium tuberculosis and the host cell*. Seminars in Immunology **26** (6): p. 486-496.
Dinadayala, P., Sambou, T., Daffé, M., and Lemassu, A. (2008). *Comparative structural analyses of the α-glucan and glycogen from Mycobacterium bovis*. Glycobiology **18** (7): p. 502-508.
Drepper, A., Peitzmann, R., and Pape, H. (1996). *Maltokinase (ATP:maltose 1-phosphotransferase) from Actinoplanes sp.: demonstration of enzyme activity and characterization of the reaction product*. FEBS Letters **388** (2-3): p. 177-179.
Edwards, T.E., Liao, R., Phan, I., Myler, P.J., and Grundner, C. (2012). *Mycobacterium thermoresistibile as a source of thermostable orthologs of Mycobacterium tuberculosis proteins*. Protein Science **21** (7): p. 1093-1096.
Elbein, A.D., Pastuszak, I., Tackett, A.J., Wilson, T., and 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*. Journal of Biological Chemistry **285** (13): p. 9803-9812.
Empadinhas, N., Albuquerque, L., Mendes, V., Macedo-Ribeiro, S., and da Costa, M.S. (2008). *Identification of the mycobacterial glucosyl-3-phosphoglycerate synthase*. FEMS Microbiology Letters **280** (2): p. 195-202.
Empadinhas, N., Pereira, P.J., Albuquerque, L., Costa, J., Sa-Moura, B., Marques, A.T., Macedo-Ribeiro, S., and 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): p. 76-93.
Falkinham III, J.O. (2015). *Environmental Sources of Nontuberculous Mycobacteria*. Clinics in Chest Medicine **36** (1): p. 35-41.
Ferguson, J.A., and Ballou, C.E. (1970). *Biosynthesis of a Mycobacterial Lipopolysaccharide: properties of the polysaccharide methyltransferase*. Journal of Biological Chemistry **245** (16): p. 4213-4223.
Fernandes, C., Mendes, V., Costa, J., Empadinhas, N., Jorge, C., Lamosa, P., Santos, H., and 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): p. 1624-1633.
Flick, P.K., and Bloch, K. (1974). *In Vitro Alterations of the Product Distribution of the Fatty Acid Synthetase from Mycobacterium phlei*. Journal of Biological Chemistry **249** (4): p. 1031-1036.
Flick, P.K., and Bloch, K. (1975). *Reversible inhibition of the fatty acid synthetase complex from Mycobacterium smegmatis by palmitoyl-coenzyme A*. Journal of Biological Chemistry **250** (9): p. 3348-3351.
Formicola, V., Milanesi, Q., and Scarsini, C. (1987). *Evidence of spinal tuberculosis at the beginning of the fourth millennium BC from Arene Candide cave (Liguria, Italy)*. American Journal of Physical Anthropology **72** (1): p. 1-6.
Forsberg, L.S., Dell, A., Walton, D.J., and Ballou, C.E. (1982). *Revised structure for the 6-O-methylglucose polysaccharide of Mycobacterium smegmatis*. Journal of Biological Chemistry **257** (7): p. 3555-3563.
Fraga, J., Maranha, A., Mendes, V., Pereira, P.J., Empadinhas, N., and Macedo-Ribeiro, S. (2015). *Structure of mycobacterial maltokinase, the missing link in the essential GlgE-pathway*. Scientific reports **5**: p. 8026.
Gagliardi, M.C., Lemassu, A., Teloni, R., Mariotti, S., Sargentini, V., Pardini, M., Daffé, M., and Nisini, R. (2007). *Cell wall-associated alpha-glucan is instrumental for Mycobacterium tuberculosis to block CD1 molecule expression and disable the function of dendritic cell derived from infected monocyte*. Cellular Microbiology **9** (8): p. 2081-2092.

Garg, S.K., Alam, M.S., Kishan, K.V.R., and Agrawal, P. (2007). *Expression and characterization of α-(1,4)-glucan branching enzyme Rv1326c of Mycobacterium tuberculosis H37Rv*. Protein Expression and Purification **51** (2): p. 198-208.

Geurtsen, J., Chedammi, S., Mesters, J., Cot, M., Driessen, N.N., Sambou, T., Kakutani, R., Ummels, R., Maaskant, J., Takata, H*., et al.* (2009). *Identification of Mycobacterial α-Glucan As a Novel Ligand for DC-SIGN: Involvement of Mycobacterial Capsular Polysaccharides in Host Immune Modulation*. The Journal of Immunology **183** (8): p. 5221-5231.

Ghodbane, R., Medie, F.M., Lepidi, H., Nappez, C., and Drancourt, M. (2014a). *Long-term survival of tuberculosis complex mycobacteria in soil*. Microbiology **160** (3): p. 496-501.

Ghodbane, R., Raoult, D., and Drancourt, M. (2014b). *Dramatic reduction of culture time of Mycobacterium tuberculosis*. Scientific reports **4**: p. 4236.

Glaziou, P., Sismanidis, C., Floyd, K., and Raviglione, M. (2015). *Global epidemiology of tuberculosis*. Cold Spring Harbor Perspectives in Medicine **5** (2): p. a017798.

Glickman, M.S., and Jacobs, W.R., Jr. (2001). *Microbial pathogenesis of Mycobacterium tuberculosis: dawn of a discipline*. Cell **104** (4): p. 477-485.

Grant, B.D., Hemmer, W., Tsigelny, I., Adams, J.A., and Taylor, S.S. (1998). *Kinetic Analyses of Mutations in the Glycine-Rich Loop of cAMP-Dependent Protein Kinase*. Biochemistry **37** (21): p. 7708-7715.

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

Gray, G.R., and Ballou, C.E. (1972). *The 6-O-Methylglucose-containing Lipopolysaccharides of Mycobacterium phlei : locations of the acyl groups*. Journal of Biological Chemistry **247** (24): p. 8129-8135.

Grellert, E., and Ballou, C.E. (1972). *Biosynthesis of a Mycobacterial Lipopolysaccharide: evidence for an acylpolysaccharide methyltransferase*. Journal of Biological Chemistry **247** (10): p. 3236-3241.

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

Griffin, J.E., Gawronski, J.D., DeJesus, M.A., Ioerger, T.R., Akerley, B.J., and Sassetti, C.M. (2011b). *High-Resolution Phenotypic Profiling Defines Genes Essential for Mycobacterial Growth and Cholesterol Catabolism*. PLoS Pathogens **7** (9): p. e1002251.

Groenewald, W., Baird, M.S., Verschoor, J.A., Minnikin, D.E., and Croft, A.K. (2014). *Differential spontaneous folding of mycolic acids from Mycobacterium tuberculosis*. Chemistry and physics of lipids **180**: p. 15-22.

Grzegorzewicz, A.E., Pham, H., Gundi, V.A.K.B., Scherman, M.S., North, E.J., Hess, T., Jones, V., Gruppo, V., Born, S.E.M., Korduláková, J*., et al.* (2012). *Inhibition of mycolic acid transport across the Mycobacterium tuberculosis plasma membrane*. Nature Chemical Biology **8** (4): p. 334-341.

Gutierrez, J.A., Solenberg, P.J., Perkins, D.R., Willency, J.A., Knierman, M.D., Jin, Z., Witcher, D.R., Luo, S., Onyia, J.E., and Hale, J.E. (2008). *Ghrelin octanoylation mediated by an orphan lipid transferase*. Proceedings of the National Academy of Sciences of the United States of America **105** (17): p. 6320-6325.

Hansen, J.M., Golchin, S.A., Veyrier, F.J., Domenech, P., Boneca, I.G., Azad, A.K., Rajaram, M.V.S., Schlesinger, L.S., Divangahi, M., Reed, M.B*., et al.* (2014). *N-Glycolylated Peptidoglycan Contributes to the Immunogenicity but Not Pathogenicity of Mycobacterium tuberculosis*. Journal of Infectious Diseases **209** (7): p. 1045-1054.

Harris, L.S., and Gray, G.R. (1977). *Acetylated methylmannose polysaccharide of Streptomyces*. Journal of Biological Chemistry **252** (8): p. 2470-2477.

Hartmans, S., de Bont, J.M., and Stackebrandt, E. (2006). *The Genus Mycobacterium--Nonmedical*. In The Prokaryotes, M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt, Editors. Springer New York: p. 889-918.

Henrissat, B., Deleury, E., and Coutinho, P.M. (2002). *Glycogen metabolism loss: a common marker of parasitic behaviour in bacteria?* Trends in Genetics **18** (9): p. 437-440.
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*., et al.* (2008). *Detection and Molecular Characterization of 9000-Year-Old Mycobacterium tuberculosis from a Neolithic Settlement in the Eastern Mediterranean*. PLoS ONE **3** (10): p. e3426.
Hett, E.C., and Rubin, E.J. (2008). *Bacterial growth and cell division: a mycobacterial perspective*. Microbiology and Molecular Biology Reviews **72** (1): p. 126-156.
Hindsgaul, O., and Ballou, C.E. (1984). *Affinity purification of mycobacterial polymethyl polysaccharides and a study of polysaccharide-lipid interactions by ^{1}H NMR*. Biochemistry **23** (3): p. 577-584.
Hoffmann, C., Leis, A., Niederweis, M., Plitzko, J.M., and Engelhardt, H. (2008). *Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure*. Proceedings of the National Academy of Sciences of the United States of America **105** (10): p. 3963-3967.
Hon, W.-C., McKay, G.A., Thompson, P.R., Sweet, R.M., Yang, D.S.C., Wright, G.D., and Berghuis, A.M. *Structure of an Enzyme Required for Aminoglycoside Antibiotic Resistance Reveals Homology to Eukaryotic Protein Kinases*. Cell **89** (6): p. 887-895.
Hong, S., Cheng, T.-Y., Layre, E., Sweet, L., Young, D.C., Posey, J.E., Butler, W.R., and Moody, D.B. (2012). *Ultralong C100 Mycolic Acids Support the Assignment of Segniliparus as a New Bacterial Genus*. PLoS ONE **7** (6): p. e39017.
Hong, X., and Hopfinger, A.J. (2004). *Construction, Molecular Modeling, and Simulation of Mycobacterium tuberculosis Cell Walls*. Biomacromolecules **5** (3): p. 1052-1065.
Hunter, R.L., Olsen, M.R., Jagannath, C., and Actor, J.K. (2006). *Multiple Roles of Cord Factor in the Pathogenesis of Primary, Secondary, and Cavitary Tuberculosis, Including a Revised Description of the Pathology of Secondary Disease*. Annals of Clinical & Laboratory Science **36** (4): p. 371-386.
Hunter, S.W., Gaylord, H., and Brennan, P.J. (1986). *Structure and antigenicity of the phosphorylated lipopolysaccharide antigens from the leprosy and tubercle bacilli*. Journal of Biological Chemistry **261** (26): p. 12345-12351.
Ilton, M., Jevans, A.W., McCarthy, E.D., Vance, D., White, H.B., 3rd, and Bloch, K. (1971). *Fatty acid synthetase activity in Mycobacterium phlei: regulation by polysaccharides*. Proceedings of the National Academy of Sciences of the United States of America **68** (1): p. 87-91.
Jacin, H., and Mishkin, A.R. (1965). *Separation of carbonhydrates on borate-impregnated silica gel G plates*. Journal of Chromatography A **18**: p. 170-173.
Jackson, M., and Brennan, P.J. (2009). *Polymethylated polysaccharides from Mycobacterium species revisited*. Journal of Biological Chemistry **284** (4): p. 1949-1953.
Jackson, M., McNeil, M.R., and Brennan, P.J. (2013). *Progress in targeting cell envelope biogenesis in Mycobacterium tuberculosis*. Future Microbiology **8** (7): p. 855-875.
Jankute, M., Grover, S., Birch, H.L., and Besra, G.S. (2014). *Genetics of Mycobacterial Arabinogalactan and Lipoarabinomannan Assembly*. In Molecular Genetics of Mycobacteria, Second Edition, G. Hatfull, and W.R. Jacobs, Jr., Editors. American Society of Microbiology Press: Washington DC p. 535-557.
Jarling, M., Cauvet, T., Grundmeier, M., Kuhnert, K., and Pape, H. (2004). *Isolation of mak1 from Actinoplanes missouriensis and evidence that Pep2 from Streptomyces coelicolor is a maltokinase*. Journal of Basic Microbiology **44** (5): p. 360-373.
Jogl, G., Hsiao, Y.-S., and Tong, L. (2004). *Structure and Function of Carnitine Acyltransferases*. Annals of the New York Academy of Sciences **1033** (1): p. 17-29.
Jogl, G., Hsiao, Y.-S., and Tong, L. (2005). *Crystal Structure of Mouse Carnitine Octanoyltransferase and Molecular Determinants of Substrate Selectivity*. Journal of Biological Chemistry **280** (1): p. 738-744.

Jorge, C.D., Lamosa, P., and Santos, H. (2007). *α-d-Mannopyranosyl-(1→2)-α-d-glucopyranosyl-(1→2)-glycerate in the thermophilic bacterium Petrotoga miotherma − structure, cellular content and function*. FEBS Journal **274** (12): p. 3120-3127.

Kalscheuer, R., and Jacobs, W.R., Jr. (2010). *The significance of GlgE as a new target for tuberculosis*. Drug News Perspect **23** (10): p. 619-624.

Kalscheuer, R., and Koliwer-Brandl, H. (2014). *Genetics of Mycobacterial Trehalose Metabolism*. In Molecular Genetics of Mycobacteria, Second Edition, G. Hatfull, and W.R. Jacobs, Jr., Editors. American Society of Microbiology Press: Washington DC p. 361-375.

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

Kalscheuer, R., Weinrick, B., Veeraraghavan, U., Besra, G.S., and Jacobs, W.R. (2010b). *Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of Mycobacterium tuberculosis*. Proceedings of the National Academy of Sciences **107** (50): p. 21761-21766.

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

Kannan, N., Taylor, S.S., Zhai, Y., Venter, J.C., and Manning, G. (2007). *Structural and Functional Diversity of the Microbial Kinome*. PLoS Biology **5** (3): p. e17.

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

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

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). *The Phyre2 web portal for protein modeling, prediction and analysis*. Nat Protocols **10** (6): p. 845-858.

Kieser, K.J., and Rubin, E.J. (2014). *How sisters grow apart: mycobacterial growth and division*. Nat Rev Micro **12** (8): p. 550-562.

Kiho, T., and Ballou, C.E. (1988). *Thermodynamic parameters and shape of the mycobacterial polymethylpolysaccharide-fatty acid complex*. Biochemistry **27** (15): p. 5824-5828.

Koch, R. (1982). *I. Die Aetiologie der Tuberculose: Nach einem in der physiologischen Gesellschaft zu Berlin am 24. März cr. gehaltenen Vortrage*. Zentralblatt für Bakteriologie, Mikrobiologie und Hygiene 1 Abt Originale A, Medizinische Mikrobiologie, Infektionskrankheiten und Parasitologie **251** (3): p. 287-296.

Kremer, L., Guerardel, Y., Gurcha, S.S., Locht, C., and Besra, G.S. (2002). *Temperature-induced changes in the cell-wall components of Mycobacterium thermoresistibile*. Microbiology **148** (Pt 10): p. 3145-3154.

Krieger, E., and Vriend, G. (2014). *YASARA View—molecular graphics for all devices—from smartphones to workstations*. Bioinformatics **30** (20): p. 2981-2982.

Ku, S.-Y., Yip, P., Cornell, K.A., Riscoe, M.K., Behr, J.-B., Guillerm, G., and Howell, P.L. (2007). *Structures of 5-Methylthioribose Kinase Reveal Substrate Specificity and Unusual Mode of Nucleotide Binding*. Journal of Biological Chemistry **282** (30): p. 22195-22206.

Laval, F., Lanéelle, M.-A., Déon, C., Monsarrat, B., and Daffé, M. (2001). *Accurate Molecular Mass Determination of Mycolic Acids by MALDI-TOF Mass Spectrometry*. Analytical Chemistry **73** (18): p. 4537-4544.

Layre, E., Collmann, A., Bastian, M., Mariotti, S., Czaplicki, J., Prandi, J., Mori, L., Stenger, S., De Libero, G., Puzo, G.*, et al.* (2009). *Mycolic Acids Constitute a Scaffold for Mycobacterial Lipid Antigens Stimulating CD1-Restricted T Cells*. Chemistry & Biology **16** (1): p. 82-92.

Le Guilloux, V., Schmidtke, P., and Tuffery, P. (2009). *Fpocket: An open source platform for ligand pocket detection*. BMC Bioinformatics **10** (1): p. 168.

Lee, Y.C. (1966). *Isolation and characterization of lipopolysaccharides containing 6-O-methyl-D-glucose from Mycobacterium species*. Journal of Biological Chemistry **241** (8): p. 1899-1908.
Lee, Y.C., and Ballou, C.E. (1964). *6-O-Methyl-D-Glucose from Mycobacteria*. Journal of Biological Chemistry **239**: p. PC3602-3603.
Leonard, S.E., Register, A.C., Krishnamurty, R., Brighty, G.J., and Maly, D.J. (2014). *Divergent Modulation of Src-Family Kinase Regulatory Interactions with ATP-Competitive Inhibitors*. ACS Chemical Biology **9** (8): p. 1894-1905.
Leščić Ašler, I., Ivić, N., Kovačić, F., Schell, S., Knorr, J., Krauss, U., Wilhelm, S., Kojić-Prodić, B., and Jaeger, K.-E. (2010). *Probing Enzyme Promiscuity of SGNH Hydrolases*. ChemBioChem **11** (15): p. 2158-2167.
Lew, J.M., Kapopoulou, A., Jones, L.M., and Cole, S.T. (2011). *TubercuList – 10 years after*. Tuberculosis **91** (1): p. 1-7.
Li, J., Guan, X., Shaw, N., Chen, W., Dong, Y., Xu, X., Li, X., and Rao, Z. (2014). *Homotypic dimerization of a maltose kinase for molecular scaffolding*. Scientific reports **4**: p. 6418.
Liu, J., Barry, C.E., 3rd, Besra, G.S., and Nikaido, H. (1996). *Mycolic acid structure determines the fluidity of the mycobacterial cell wall*. Journal of Biological Chemistry **271** (47): p. 29545-29551.
Liu, J., Rosenberg, E.Y., and Nikaido, H. (1995). *Fluidity of the lipid domain of cell wall from Mycobacterium chelonae*. Proceedings of the National Academy of Sciences of the United States of America **92** (24): p. 11254-11258.
Lönnroth, K., Jaramillo, E., Williams, B.G., Dye, C., and Raviglione, M. (2009). *Drivers of tuberculosis epidemics: The role of risk factors and social determinants*. Social Science & Medicine **68** (12): p. 2240-2246.
Lourenço, E.C., Maycock, C.D., and Rita Ventura, M. (2009). *Synthesis of potassium (2R)-2-O-α-d-glucopyranosyl-(1→6)-α-d-glucopyranosyl-2,3-dihydroxypropanoate a natural compatible solute*. Carbohydrate Research **344** (15): p. 2073-2078.
Lourenço, E.C., and Ventura, M.R. (2011). *Synthesis of Potassium (2R)-2-O-α-D-Mannopyranosyl-(1→2)-α-D-glucopyranosyl-2,3-dihydroxypropanoate: A Naturally Compatible Solute*. European Journal of Organic Chemistry **2011** (33): p. 6698-6703.
Ma, Q., and Xu, Y. (2013). *Global Genomic Arrangement of Bacterial Genes Is Closely Tied with the Total Transcriptional Efficiency*. Genomics, Proteomics & Bioinformatics **11** (1): p. 66-71.
Ma, Q., Zhao, X., Eddine, A.N., Geerlof, A., Li, X., Cronan, J.E., Kaufmann, S.H.E., and Wilmanns, M. (2006). *The Mycobacterium tuberculosis LipB enzyme functions as a cysteine/lysine dyad acyltransferase*. Proceedings of the National Academy of Sciences **103** (23): p. 8662-8667.
Machida, Y., and 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): p. 1146-1148.
Maggio, J.E. (1980). *Structure of a mycobacterial polysaccharide-fatty acyl-CoA complex: nuclear magnetic resonance studies*. Proceedings of the National Academy of Sciences of the United States of America **77** (5): p. 2582-2586.
Mahapatra, S., Crick, D.C., McNeil, M.R., and Brennan, P.J. (2008). *Unique Structural Features of the Peptidoglycan of Mycobacterium leprae*. Journal of Bacteriology **190** (2): p. 655-661.
Maitra, S.K., and Ballou, C.E. (1977). *Heterogeneity and refined structtures of 3-O-methyl-D-mannose polysaccharides from Mycobacterium smegmatis*. Journal of Biological Chemistry **252** (8): p. 2459-2469.
Maloney, D.H., and Ballou, C.E. (1980). *Polymethylpolysaccharide synthesis in an ethionine-resistant mutant of Mycobacterium smegmatis*. Journal of Bacteriology **141** (3): p. 1217-1221.
Mansilla, M.C., Cybulski, L.E., Albanesi, D., and de Mendoza, D. (2004). *Control of Membrane Lipid Fluidity by Molecular Thermosensors*. Journal of Bacteriology **186** (20): p. 6681-6688.
Maranha, A., Moynihan, P.J., Miranda, V., Lourenço, E.C., Nunes-Costa, D., Fraga, J.S., José Barbosa Pereira, P., Macedo-Ribeiro, S., Ventura, M.R., Clarke, A.J*., et al.* (2015). *Octanoylation of early intermediates of mycobacterial methylglucose lipopolysaccharides*. Scientific reports **5**: p. 13610.

Marrakchi, H., Lanéelle, M.-A., and Daffé, M. (2014). *Mycolic Acids: Structures, Biosynthesis, and Beyond*. Chemistry & Biology **21** (1): p. 67-85.
Matsunaga, I., Naka, T., Talekar, R.S., McConnell, M.J., Katoh, K., Nakao, H., Otsuka, A., Behar, S.M., Yano, I., Moody, D.B.*, et al.* (2008). *Mycolyltransferase-mediated Glycolipid Exchange in Mycobacteria*. Journal of Biological Chemistry **283** (43): p. 28835-28841.
McKeown, T. (1976). *The Role of Medicine: Dream, Mirage, or Nemesis?*, edn (The Nuffield Provincial Hospitals Trust: London).
McNeil, M., Tsang, A.Y., McClatchy, J.K., Stewart, C., Jardine, I., and Brennan, P.J. (1987). *Definition of the surface antigens of Mycobacterium malmoense and use in studying the etiology of a form of mycobacteriosis*. Journal of Bacteriology **169** (7): p. 3312-3320.
Medjahed, H., Gaillard, J.L., and Reyrat, J.M. (2010). *Mycobacterium abscessus: a new player in the mycobacterial field*. Trends in Microbiology **18** (3): p. 117-123.
Meharena, H.S., Chang, P., Keshwani, M.M., Oruganty, K., Nene, A.K., Kannan, N., Taylor, S.S., and Kornev, A.P. (2013). *Deciphering the Structural Basis of Eukaryotic Protein Kinase Regulation*. PLoS Biology **11** (10): p. e1001680.
Mendes, V., Maranha, A., Alarico, S., da Costa, M.S., and Empadinhas, N. (2011). *Mycobacterium tuberculosis Rv2419c, the missing glucosyl-3-phosphoglycerate phosphatase for the second step in methylglucose lipopolysaccharide biosynthesis*. Scientific reports **1**: p. 177.
Mendes, V., Maranha, A., Alarico, S., and Empadinhas, N. (2012). *Biosynthesis of mycobacterial methylglucose lipopolysaccharides*. Natural product reports **29** (8): p. 834-844.
Mendes, V., Maranha, A., Lamosa, P., da Costa, M.S., and Empadinhas, N. (2010). *Biochemical characterization of the maltokinase from Mycobacterium bovis BCG*. BMC Biochemistry **11**: p. 21.
Miah, F., Koliwer-Brandl, H., Rejzek, M., Field, R.A., Kalscheuer, R., and Bornemann, S. (2013). *Flux through trehalose synthase flows from trehalose to the alpha anomer of maltose in mycobacteria*. Chemistry & Biology **20** (4): p. 487-493.
Millet, J.-P., Moreno, A., Fina, L., del Baño, L., Orcau, A., de Olalla, P., and Caylà, J. (2013). *Factors that influence current tuberculosis epidemiology*. European Spine Journal **22** (4): p. 539-548.
Minnikin, D.E. (1982). *Lipids: Complex lipids, their chemistry, biosynthesis and role*. In The Biology of Mycobacteria, C. Ratledge, and J. Stanford, Editors. Academic Press: London p. 94-184.
Minnikin, D.E., Lee, O.Y., Wu, H.H., Besra, G.S., Bhatt, A., Nataraj, V., Rothschild, B.M., Spigelman, M., and Donoghue, H.D. (2015). *Ancient mycobacterial lipids: Key reference biomarkers in charting the evolution of tuberculosis*. Tuberculosis (Edinb) **95** (Suppl 1): p. S133-139.
Molle, V., and Kremer, L. (2010). *Division and cell envelope regulation by Ser/Thr phosphorylation: Mycobacterium shows the way*. Molecular Microbiology **75** (5): p. 1064-1077.
Mondino, S., Gago, G., and Gramajo, H. (2013). *Transcriptional regulation of fatty acid biosynthesis in mycobacteria*. Molecular Microbiology **89** (2): p. 372-387.
Moynihan, P.J., and Clarke, A.J. (2010). *O-acetylation of peptidoglycan in gram-negative bacteria: identification and characterization of peptidoglycan O-acetyltransferase in Neisseria gonorrhoeae*. Journal of Biological Chemistry **285** (17): p. 13264-13273.
Moynihan, P.J., and Clarke, A.J. (2014). *Mechanism of action of peptidoglycan O-acetyltransferase B involves a Ser-His-Asp catalytic triad*. Biochemistry **53** (39): p. 6243-6251.
Moynihan, P.J., Sychantha, D., and Clarke, A.J. (2014). *Chemical biology of peptidoglycan acetylation and deacetylation*. Bioorganic Chemistry **54**: p. 44-50.
Murphy, H.N., Stewart, G.R., Mischenko, V.V., Apt, A.S., Harris, R., McAlister, M.S., Driscoll, P.C., Young, D.B., and Robertson, B.D. (2005). *The OtsAB pathway is essential for trehalose biosynthesis in Mycobacterium tuberculosis*. Journal of Biological Chemistry **280** (15): p. 14524-14529.
Narumi, K., Keller, J.M., and Ballou, C.E. (1973). *Biosynthesis of a mycobacterial lipopolysaccharide. Incorporation of (14C)acyl groups by whole cells in vivo*. Biochemical Journal **132** (2): p. 329-340.

Narumi, K., and Tsumita, T. (1967). *Identification of alpha,alpha-trehalose 6,6'-dimannosylphosphate and alpha-maltose 1-phosphate of Mycobacteria*. Journal of Biological Chemistry **242** (9): p. 2233-2239.

Nessar, R., Cambau, E., Reyrat, J.M., Murray, A., and Gicquel, B. (2012). *Mycobacterium abscessus: a new antibiotic nightmare*. Journal of Antimicrobial Chemotherapy **67** (4): p. 810-818.

Nichols, D., Cahoon, N., Trakhtenberg, E.M., Pham, L., Mehta, A., Belanger, A., Kanigan, T., Lewis, K., and Epstein, S.S. (2010). *Use of Ichip for High-Throughput In Situ Cultivation of "Uncultivable" Microbial Species*. Applied and Environmental Microbiology **76** (8): p. 2445-2450.

Niederweis, M. (2008). *Nutrient acquisition by mycobacteria*. Microbiology **154** (3): p. 679-692.

Niederweis, M., Danilchanka, O., Huff, J., Hoffmann, C., and Engelhardt, H. (2010). *Mycobacterial outer membranes: in search of proteins*. Trends in Microbiology **18** (3): p. 109-116.

Niehues, B., Jossek, R., Kramer, U., Koch, A., Jarling, M., Schroder, W., and Pape, H. (2003). *Isolation and characterization of maltokinase (ATP:maltose 1-phosphotransferase) from Actinoplanes missouriensis*. Archives of Microbiology **180** (4): p. 233-239.

Nishi, Y., Mifune, H., Yabuki, A., Tajiri, Y., Hirata, R., Tanaka, E., Hosoda, H., Kangawa, K., and Kojima, M. (2013). *Changes in Subcellular Distribution of n-Octanoyl or n-Decanoyl Ghrelin in Ghrelin-Producing Cells*. Frontiers in endocrinology (Lausanne) **4**: p. 84.

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

O'Garra, A., Redford, P.S., McNab, F.W., Bloom, C.I., Wilkinson, R.J., and Berry, M.P.R. (2013). *The Immune Response in Tuberculosis*. Annual Review of Immunology **31** (1): p. 475-527.

Odriozola, J.M., Ramos, J.A., and Bloch, K. (1977). *Fatty acid synthetase activity in Mycobacterium smegmatis Characterization of the acyl carrier protein-dependent elongating system*. Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism **488** (2): p. 207-217.

Ohgusu, H., Takahashi, T., and Kojima, M. (2012). *Enzymatic Characterization of GOAT, ghrelin O-acyltransferase*. In Methods in Enzymology Volume 514, K. Masayasu, and K. Kenji, Editors. Academic Press: p. 147-163.

Ojha, A.K., Trivelli, X., Guerardel, Y., Kremer, L., and Hatfull, G.F. (2010). *Enzymatic Hydrolysis of Trehalose Dimycolate Releases Free Mycolic Acids during Mycobacterial Growth in Biofilms*. Journal of Biological Chemistry **285** (23): p. 17380-17389.

Ornston, M.K., and Ornston, L.N. (1969). *Two forms of D-glycerate kinase in Escherichia coli*. Journal of Bacteriology **97** (3): p. 1227-1233.

Pal, K., Kumar, S., Sharma, S., Garg, S.K., Alam, M.S., Xu, H.E., Agrawal, P., and Swaminathan, K. (2010). *Crystal Structure of Full-length Mycobacterium tuberculosis H37Rv Glycogen Branching Enzyme: insights of N-terminal β-sandwich in substrate specificity and enzymatic activity*. Journal of Biological Chemistry **285** (27): p. 20897-20903.

Pan, Y.T., Koroth Edavana, V., Jourdian, W.J., Edmondson, R., Carroll, J.D., Pastuszak, I., and Elbein, A.D. (2004). *Trehalose synthase of Mycobacterium smegmatis: purification, cloning, expression, and properties of the enzyme*. European Journal of Biochemistry **271** (21): p. 4259-4269.

Papaioannou, N., Cheon, H.-S., Lian, Y., and Kishi, Y. (2007). *Product-Regulation Mechanisms for Fatty Acid Biosynthesis Catalyzed by Mycobacterium smegmatis FAS I*. ChemBioChem **8** (15): p. 1775-1780.

Paul, M.J., Primavesi, L.F., Jhurreea, D., and Zhang, Y. (2008). *Trehalose metabolism and signaling*. Annual Review of Plant Biology **59**: p. 417-441.

Pavelka, M.S., Mahapatra, S., and Crick, D.C. (2014). *Genetics of Peptidoglycan Biosynthesis*. In Molecular Genetics of Mycobacteria, Second Edition, G. Hatfull, and W.R. Jacobs, Jr., Editors. American Society of Microbiology Press: Washington DC p. 513-533.

Pawełczyk, J., and Kremer, L. (2014). *The Molecular Genetics of Mycolic Acid Biosynthesis*. In Molecular Genetics of Mycobacteria, Second Edition, G. Hatfull, and W.R. Jacobs, Jr., Editors. American Society of Microbiology Press: Washington DC p. 611-631.

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

Pérez, J., Castañeda-García, A., Jenke-Kodama, H., Müller, R., and Muñoz-Dorado, J. (2008). *Eukaryotic-like protein kinases in the prokaryotes and the myxobacterial kinome*. Proceedings of the National Academy of Sciences **105** (41): p. 15950-15955.

Peyron, P., Vaubourgeix, J., Poquet, Y., Levillain, F., Botanch, C., Bardou, F., Daffe, M., Emile, J.F., Marchou, B., Cardona, P.J*., et al.* (2008). *Foamy macrophages from tuberculous patients' granulomas constitute a nutrient-rich reservoir for M. tuberculosis persistence*. PLoS Pathogens **4** (11): p. e1000204.

Pommier, M.T., and Michel, G. (1986). *Isolation and characterization of an O-methylglucose-containing lipopolysaccharide produced by Nocardia otitidis-caviarum*. Journal of general microbiology **132** (9): p. 2433-2441.

Rachman, H., Strong, M., Ulrichs, T., Grode, L., Schuchhardt, J., Mollenkopf, H., Kosmiadi, G.A., Eisenberg, D., and Kaufmann, S.H.E. (2006). *Unique Transcriptome Signature of Mycobacterium tuberculosis in Pulmonary Tuberculosis*. Infection and Immunity **74** (2): p. 1233-1242.

Ramirez, M.S., and Tolmasky, M.E. (2010). *Aminoglycoside modifying enzymes*. Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy **13** (6): p. 151-171.

Rao, M.B., Tanksale, A.M., Ghatge, M.S., and Deshpande, V.V. (1998). *Molecular and Biotechnological Aspects of Microbial Proteases*. Microbiology and Molecular Biology Reviews **62** (3): p. 597-635.

Resh, M.D. (2013). *Covalent lipid modifications of proteins*. Current Biology **23** (10): p. R431-435.

Reyrat, J.-M., and Kahn, D. (2001). *Mycobacterium smegmatis: an absurd model for tuberculosis?* Trends in Microbiology **9** (10): p. 472-473.

Richards, J.P., and Ojha, A.K. (2014). *Mycobacterial Biofilms*. In Molecular Genetics of Mycobacteria, Second Edition, G. Hatfull, and W.R. Jacobs, Jr., Editors. American Society of Microbiology Press: Washington DC p. 773-784.

Roberts, C.A., and Buikstra, J.E. (2003). *The bioarchaeology of tuberculosis : a global perspective on a re-emerging disease*, (University Press of Florida: Gainesville, Florida).

Rombouts, Y., Brust, B., Ojha, A.K., Maes, E., Coddeville, B., Elass-Rochard, E., Kremer, L., and Guerardel, Y. (2012). *Exposure of Mycobacteria to Cell Wall-inhibitory Drugs Decreases Production of Arabinoglycerolipid Related to Mycolyl-arabinogalactan-peptidoglycan Metabolism*. Journal of Biological Chemistry **287** (14): p. 11060-11069.

Röttig, A., and Steinbüchel, A. (2013). *Acyltransferases in Bacteria*. Microbiology and Molecular Biology Reviews **77** (2): p. 277-321.

Roy, R., Usha, V., Kermani, A., Scott, D.J., Hyde, E.I., Besra, G.S., Alderwick, L.J., and 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): p. 2245-2255.

Saerens, S.M.G., Delvaux, F., Verstrepen, K.J., Van Dijck, P., Thevelein, J.M., and Delvaux, F.R. (2008). *Parameters Affecting Ethyl Ester Production by Saccharomyces cerevisiae during Fermentation*. Applied and Environmental Microbiology **74** (2): p. 454-461.

Saerens, S.M.G., Verstrepen, K.J., Van Laere, S.D.M., Voet, A.R.D., Van Dijck, P., Delvaux, F.R., and Thevelein, J.M. (2006). *The Saccharomyces cerevisiae EHT1 and EEB1 Genes Encode Novel Enzymes with Medium-chain Fatty Acid Ethyl Ester Synthesis and Hydrolysis Capacity*. Journal of Biological Chemistry **281** (7): p. 4446-4456.

Saier, M.H., Jr., and Ballou, C.E. (1968a). *The 6-O-methylglucose-containig lipopolysaccharide of Mycobacterium phlei. Complet structure of the polysaccharide*. Journal of Biological Chemistry **243** (16): p. 4332-4341.

Saier, M.H., Jr., and Ballou, C.E. (1968b). *The 6-O-methylglucose-containing lipopolysaccharide of Mycobacterium phlei. Identification of D-glyceric acid and 3-O-methyl-D-glucose in the polysaccharide*. Journal of Biological Chemistry **243** (5): p. 992-1005.

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

Sambrook, J., and Russell, D.W. (2001). *Molecular Cloning: A laboratory manual*, 3rd edn (Cold Spring Harbor Laboratory Press: New York).

Sani, M., Houben, E.N.G., Geurtsen, J., Pierson, J., de Punder, K., van Zon, M., Wever, B., Piersma, S.R., Jiménez, C.R., Daffé, M*., et al.* (2010). *Direct Visualization by Cryo-EM of the Mycobacterial Capsular Layer: A Labile Structure Containing ESX-1-Secreted Proteins*. PLoS Pathogens **6** (3): p. e1000794.

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

Scheeff, E.D., and Bourne, P.E. (2005). *Structural Evolution of the Protein Kinase–Like Superfamily*. PLoS Computational Biology **1** (5): p. e49.

Schröder, K.H., Naumann, L., Kroppenstedt, R.M., and Reischl, U. (1997). *Mycobacterium hassiacum sp. nov., a New Rapidly Growing Thermophilic Mycobacterium*. International Journal of Systematic and Evolutionary Microbiology **47** (1): p. 86-91.

Schulze-Gahmen, U., Pelaschier, J., Yokota, H., Kim, R., and Kim, S.-H. (2003). *Crystal structure of a hypothetical protein, TM841 of Thermotoga maritima, reveals its function as a fatty acid–binding protein*. Proteins: Structure, Function, and Bioinformatics **50** (4): p. 526-530.

Shi, K., and Berghuis, A.M. (2012). *Structural Basis for Dual Nucleotide Selectivity of Aminoglycoside 2"-Phosphotransferase IVa Provides Insight on Determinants of Nucleotide Specificity of Aminoglycoside Kinases*. The Journal of Biological Chemistry **287** (16): p. 13094-13102.

Sidders, B., and Stoker, N.G. (2001). *Mycobacteria: Biology*. In eLS, John Wiley & Sons, Ltd: p.

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

Spalding, M.D., and Prigge, S.T. (2010). *Lipoic Acid Metabolism in Microbial Pathogens*. Microbiology and Molecular Biology Reviews **74** (2): p. 200-228.

Stadthagen, G., Sambou, T., Guerin, M., Barilone, N., Boudou, F., Kordulakova, J., Charles, P., Alzari, P.M., Lemassu, A., Daffe, M*., et al.* (2007). *Genetic basis for the biosynthesis of methylglucose lipopolysaccharides in Mycobacterium tuberculosis*. Journal of Biological Chemistry **282** (37): p. 27270-27276.

Stokes, R.W., Norris-Jones, R., Brooks, D.E., Beveridge, T.J., Doxsee, D., and Thorson, L.M. (2004). *The Glycan-Rich Outer Layer of the Cell Wall of Mycobacterium tuberculosis Acts as an Antiphagocytic Capsule Limiting the Association of the Bacterium with Macrophages*. Infection and Immunity **72** (10): p. 5676-5686.

Strohalm, M., Kavan, D., Novak, P., Volny, M., and Havlicek, V. (2010). *mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data*. Analytical chemistry **82** (11): p. 4648-4651.

Supply, P., Marceau, M., Mangenot, S., Roche, D., Rouanet, C., Khanna, V., Majlessi, L., Criscuolo, A., Tap, J., Pawlik, A*., et al.* (2013). *Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis*. Nature Genetics **45** (2): p. 172-179.

Swarts, B.M., Holsclaw, C.M., Jewett, J.C., Alber, M., Fox, D.M., Siegrist, M.S., Leary, J.A., Kalscheuer, R., and Bertozzi, C.R. (2012). *Probing the Mycobacterial Trehalome with Bioorthogonal Chemistry*. Journal of the American Chemical Society **134** (39): p. 16123-16126.

Syson, K., Stevenson, C.E.M., Rashid, A.M., Saalbach, G., Tang, M., Tuukkanen, A., Svergun, D.I., Withers, S.G., Lawson, D.M., and Bornemann, S. (2014). *Structural Insight into How Streptomyces coelicolor Maltosyl Transferase GlgE Binds α-Maltose 1-Phosphate and Forms a Maltosyl-enzyme Intermediate*. Biochemistry **53** (15): p. 2494-2504.

Syson, K., Stevenson, C.E.M., Rejzek, M., Fairhurst, S.A., Nair, A., Bruton, C.J., Field, R.A., Chater, K.F., Lawson, D.M., and Bornemann, S. (2011). *Structure of Streptomyces Maltosyltransferase GlgE, a Homologue of a Genetically Validated Anti-tuberculosis Target*. Journal of Biological Chemistry **286** (44): p. 38298-38310.

Takayama, K., Wang, C., and Besra, G.S. (2005). *Pathway to synthesis and processing of mycolic acids in Mycobacterium tuberculosis*. Clinical Microbiology Reviews **18** (1): p. 81-101.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). *MEGA6: Molecular Evolutionary Genetics Analysis version 6.0*. Molecular Biology and Evolution: p.

Taylor, S.S., Keshwani, M.M., Steichen, J.M., and Kornev, A.P. (2012). *Evolution of the eukaryotic protein kinases as dynamic molecular switches*. Philosophical Transactions of the Royal Society B: Biological **367** (1602): p. 2517-2528.

Tiago, I., Maranha, A., Mendes, V., Alarico, S., Moynihan, P.J., Clarke, A.J., Macedo-Ribeiro, S., Pereira, P.J., and Empadinhas, N. (2012). *Genome Sequence of Mycobacterium hassiacum DSM 44199, a Rare Source of Heat-Stable Mycobacterial Proteins*. Journal of Bacteriology **194** (24): p. 7010-7011.

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

Toriyama, S., Yano, I., Masui, M., Kusunose, E., Kusunose, M., and 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*. Journal of Biochemistry **88** (1): p. 211-221.

Tortoli, E. (2009). *Clinical manifestations of nontuberculous mycobacteria infections*. Clinical Microbiology and Infection **15** (10): p. 906-910.

Tortoli, E. (2014). *Microbiological Features and Clinical Relevance of New Species of the Genus Mycobacterium*. Clinical Microbiology Reviews **27** (4): p. 727-752.

Traut, T. (2008). *The Limits For Life Define The Limits For Enzymes*. In Allosteric Regulatory Enzymes, Springer US: p. 29-48.

Tuffal, G., Albigot, R., Monsarrat, B., Ponthus, C., Picard, C., Riviere, M., and Puzo, G. (1995a). *Purification and LSIMS analysis of methyl glucose polysaccharides from Mycobacterium xenopi, a slow-growing Mycobacterium*. Journal of Carbohydrate Chemistry **14** (4-5): p. 631-642.

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

Tuffal, G., Ponthus, C., Picard, C., Riviere, M., and Puzo, G. (1995b). *Structural elucidation of novel methylglucose-containing polysaccharides from Mycobacterium xenopi*. European Journal of Biochemistry **233** (1): p. 377-383.

Tung, K.K., and Ballou, C.E. (1973). *Biosynthesis of a mycobacterial lipopolysaccharide. Properties of the polysaccharide: acyl coenzyme A acyltransferase reaction*. Journal of Biological Chemistry **248** (20): p. 7126-7133.

Turk, V., Stoka, V., and Turk, D. (2008). *Cystatins: biochemical and structural properties, and medical relevance*. Frontiers in bioscience : a journal and virtual library **13**: p. 5406-5420.

van der Werf, M., Kodmon, C., Katalinic-Jankovic, V., Kummik, T., Soini, H., Richter, E., Papaventsis, D., Tortoli, E., Perrin, M., van Soolingen, D*., et al.* (2014). *Inventory study of non-tuberculous mycobacteria in the European Union*. BMC Infectious Diseases **14** (1): p. 62.

Varela, C., Rittmann, D., Singh, A., Krumbach, K., Bhatt, K., Eggeling, L., Besra, G.S., and Bhatt, A. (2012). *MmpL genes are associated with mycolic acid metabolism in mycobacteria and corynebacteria*. Chemistry & Biology **19** (4): p. 498-506.

Veleti, S.K., Lindenberger, J.J., Ronning, D.R., and 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): p. 1404-1411.

Veyron-Churlet, R., Molle, V., Taylor, R.C., Brown, A.K., Besra, G.S., Zanella-Cléon, I., Fütterer, K., and Kremer, L. (2009). *The Mycobacterium tuberculosis β-Ketoacyl-Acyl Carrier Protein Synthase III Activity Is Inhibited by Phosphorylation on a Single Threonine Residue*. Journal of Biological Chemistry **284** (10): p. 6414-6424.

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

Weisman, L.S., and Ballou, C.E. (1984a). *Biosynthesis of the mycobacterial methylmannose polysaccharide. Identification of a 3-O-methyltransferase*. Journal of Biological Chemistry **259** (6): p. 3464-3469.

Weisman, L.S., and Ballou, C.E. (1984b). *Biosynthesis of the mycobacterial methylmannose polysaccharide. Identification of an alpha 1-->4-mannosyltransferase*. Journal of Biological Chemistry **259** (6): p. 3457-3463.

WHO (2014). *Global tuberculosis report*. In, Editor^Editors. (World Health Organization: Geneva), p.

Wood, W.I., Peterson, D.O., and Bloch, K. (1977). *Mycobacterium smegmatis fatty acid synthetase. A mechanism based on steady state rates and product distributions*. Journal of Biological Chemistry **252** (16): p. 5745-5749.

Woodruff, P.J., Carlson, B.L., Siridechadilok, B., Pratt, M.R., Senaratne, R.H., Mougous, J.D., Riley, L.W., Williams, S.J., and Bertozzi, C.R. (2004). *Trehalose is required for growth of Mycobacterium smegmatis*. Journal of Biological Chemistry **279** (28): p. 28835-28843.

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

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

Yabusaki, K.K., and Ballou, C.E. (1978). *Interaction of mycobacterial polymethylpolysaccharides with paranaric acid and palmitoyl-coenzyme A: structural specificity and monomeric dissociation constants*. Proceedings of the National Academy of Sciences of the United States of America **75** (2): p. 691-695.

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

Yang, J., Brown, M.S., Liang, G., Grishin, N.V., and Goldstein, J.L. (2008). *Identification of the Acyltransferase that Octanoylates Ghrelin, an Appetite-Stimulating Peptide Hormone*. Cell **132** (3): p. 387-396.

Zimhony, O., Vilchèze, C., and Jacobs, W.R. (2004). *Characterization of Mycobacterium smegmatis Expressing the Mycobacterium tuberculosis Fatty Acid Synthase I (fas1) Gene*. Journal of Bacteriology **186** (13): p. 4051-4055.

Zink, A.R., Molnár, E., Motamedi, N., Pálfy, G., Marcsik, A., and Nerlich, A.G. (2007). *Molecular history of tuberculosis from ancient mummies and skeletons*. International Journal of Osteoarchaeology **17** (4): p. 380-391.

Zuber, B., Chami, M., Houssin, C., Dubochet, J., Griffiths, G., and Daffé, M. (2008). *Direct Visualization of the Outer Membrane of Mycobacteria and Corynebacteria in Their Native State*. Journal of Bacteriology **190** (16): p. 5672-5680.

Zumla, A., Nahid, P., and Cole, S.T. (2013). *Advances in the development of new tuberculosis drugs and treatment regimens*. Nature Reviews: Drug Discovery **12** (5): p. 388-404.

Zwerling, A., Behr, M.A., Verma, A., Brewer, T.F., Menzies, D., and Pai, M. (2011). *The BCG World Atlas: A Database of Global BCG Vaccination Policies and Practices*. PLoS Medicine **8** (3): p. e1001012.