DEPARTAMENTO DE CIÊNCIAS DA VIDA

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Molecular and biochemical characterization of a rare glucokinase with a cryptic function in environmental mycobacteria

Daniela Cristina Nunes Costa

2015

DEPARTAMENTO DE CIÊNCIAS DA VIDA

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Molecular and biochemical characterization of a rare glucokinase with a cryptic function in environmental mycobacteria

Dissertação apresentada à Universidade de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Bioquímica, realizada sob a orientação científica do Doutor Nuno Miguel da Silva Empadinhas (Centro de Neurociências e Biologia Celular e Instituto de Investigação Interdisciplinar, Universidade de Coimbra) e co-orientação institucional da Prof. Doutora Paula Maria de Melim Vasconcelos de Vitorino Morais (Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade de Coimbra)

Daniela Cristina Nunes Costa

2015

Acknowledgments

Os meus sinceros agradecimentos:

Ao Doutor Nuno Empadinhas por me ter aceitado no seu grupo, por ter orientado e constantemente incentivado o meu trabalho, e pela disponibilidade em partilhar conhecimentos e ideias e em debater novas hipóteses e estratégias durante todo o trabalho.

À Ana Maranha pela disponibilidade e apoio constantes e por todo o tempo dispensado a ensinar-me e a partilhar valiosas indicações durante a elaboração de todo o trabalho.

À Susana, à Mafalda, ao Diogo, à Rita e à Sónia por terem contribuído diariamente para criar um ambiente de trabalho saudável e de entreajuda e por todo o apoio e companheirismo ao longo deste ano.

À Professora Doutora Paula Morais por assumir responsabilidade interna pela Tese no Departamento de Ciências da Vida.

Ao Doutor Brian Goodfellow (Universidade de Aveiro) pelo apoio na obtenção e análise dos espectros de RMN. Os espectrómetros de RMN da Universidade de Aveiro pertencem à *Portuguese National NMR Network* (PTNMR) apoiada pela Fundação para a Ciência e a Tecnologia (FCT).

À Doutora Sandra Macedo Ribeiro (IBMC, Porto) pelo apoio com o *software* Aline.

Ao Doutor Hugo Osório (Ipatimup Proteomics Unit) pela análise por espectrometria de massa.

À FCT-Fundação para a Ciência e a Tecnologia, Portugal, através de fundos nacionais FCT/MCTES (PIDDAC) e do Fundo Europeu de Desenvolvimento Regional (FEDER) através do COMPETE–Programa Operacional Factores de Competitividade (POFC) pelos projectos FCOMP-01-0124-FEDER-028359 [PTDC/BIA-MIC/2779/2012] e projecto estratégico UID/NEU/04539/2013 que suportaram a execução experimental deste trabalho.

Aos meus pais, à minha família e aos meus amigos por todas as palavras de incentivo e motivação, mas também por todos os momentos de "distracção" e pelo empenho em me garantir sempre a dose certa de diversão e convívio mesmo (e sobretudo) nas fases de maior pressão e trabalho.

INDEX

Abstract

Mycobacteria are a diverse genus of actinobacteria better known for its most threatening members *Mycobacterium tuberculosis* and *Mycobacterium leprae*. However, the lifestyle of these strict pathogens is not representative of the genus since the majority of mycobacterial species are saprophytic environmental microorganisms and part of water and soil microbial communities. They are also opportunistic pathogens that pose novel and challenging therapeutic obstacles highlighted by the steady increase in infections for which they are responsible, which has been met with a renewed interest in studying their unique ecological and physiological traits.

A recent development in mycobacterial glycobiology was the discovery of a novel and essential pathway of α-glucan synthesis whereby trehalose is converted to maltose itself phosphorylated to maltose 1-phosphate, which is then added to growing glucan chains by a maltosyltransferase. The genes coding for the two first enzymes in this pathway, trehalose synthase (TreS) and maltokinase (Mak), are almost always associated in the same gene cluster reflecting their metabolic connection.

We identified a second gene annotated as a maltokinase in the genome of only a few strains of environmental mycobacteria and a few other actinomycetes, included in a different genomic context within a rare but conserved gene cluster. The widely-studied model organism *Mycobacterium smegmatis* was selected to investigate this Mak homologue. We cloned the gene and purified the recombinant enzyme, which we found to specifically phosphorylate glucose to glucose 6-phosphate at 37°C and pH 8 (optimal conditions) with a strict requirement for Mg^{2+}. Similarly to other kinases, this glucokinase (GluK) displayed a residual ATPase activity which was unexpectedly and strongly stimulated in the presence of Zn^{2+} or Mn^{2+}. The enzyme exhibited Michaelis-Menten kinetics for both glucokinase and ATPase activities.

GluK's physiological function remains elusive but the molecular and biochemical characterization reported here represents an important first step in elucidating its role and possible participation in a novel pathway in environmental mycobacteria.

Keywords: environmental mycobacteria, actinomycetes, *Mycobacterium smegmatis*, isoenzymes, glucokinase

Resumo

As micobactérias são um género diverso de actinobactérias cujos membros mais conhecidos são também os mais ameaçadores, *Mycobacterium tuberculosis* e *Mycobacterium leprae*. O modo de vida destes parasitas obrigatórios não é no entanto representativo do género, uma vez que na sua maioria as espécies de micobactérias são ambientais e saprófitas, fazendo parte das comunidades microbianas do solo e da água. Estas espécies ambientais são também capazes de causar infecções oportunistas que colocam importantes desafios terapêuticos pelo aumento generalizado de casos, o que tem estimulado um renovado interesse no estudo da sua ecologia e fisiologia únicas.

Um desenvolvimento recente no âmbito da glicobiologia micobacteriana foi a descoberta de uma via essencial de síntese de α-glucanos em que trealose é convertida em maltose e subsequentemente fosforilada a maltose 1-fosfato, que é adicionada a cadeias crescentes de glucanos por uma maltosiltransferase. Em micobactérias, os genes que codificam as duas primeiras enzimas nesta via, trealose sintase (TreS) e maltocinase (Mak), surgem associados reflectindo a sua ligação metabólica.

Um segundo gene anotado como maltocinase foi identificado no genoma de apenas algumas micobactérias ambientais e alguns outros actinomicetos, estando incluído num *cluster* de genes raro mas conservado e sem associação com TreS. O modelo experimental *Mycobacterium smegmatis* foi seleccionado para investigar este homólogo da Mak. Clonámos o gene e purificámos a proteína recombinante, que verificámos fosforilar glucose a glucose 6-fosfato a 37°C e pH 8 (condições óptimas) requerendo a presença de iões Mg^{2+}. Tal como outras cinases, esta glucocinase (GluK) apresenta uma actividade residual de ATPase que é no entanto exacerbada na presença de Zn^{2+} ou Mn^{2+}. A enzima segue uma cinética de Michaelis-Menten tanto para a actividade de glucocinase como para a de ATPase.

A função fisiológica desta GluK permanece obscura mas a caracterização molecular e bioquímica descritas neste trabalho contribuem para elucidar o seu papel e possível participação numa nova via metabólica em micobactérias ambientais.

Palavras-chave: micobactérias ambientais, actinomicetos, *Mycobacterium smegmatis*, isoenzimas, glucocinase

List of Abbreviations

ADPGK – ADP-dependent glucokinase

AG – arabinogalactan

APH – aminoglycoside phosphotransferase

ATPase – adenosine triphosphatase

ELK – eukaryotic-like kinase

FAS-I – fatty acid synthase

G6P – glucose 6-phosphate

GlgB – glycogen branching enzyme

GlgE – maltosyltransferase

GluK – glucokinase

GT – glycosyltransferase

HK – hexokinase

M1P – maltose 1-phosphate

MA – mycolic acid

Mak – maltokinase

MGLP – methylglucose lipopolysaccharide

MMP – methylmannose polysaccharide

MTRK – 5-methylthioribose kinase

NTM – nontuberculous mycobacteria

PG – peptidoglycan

PGM – phosphoglucomutase

PMP – polymethylated polysaccharide

PolyP – polyphosphate

PPase – pyrophosphatase

PPGK – polyphosphate-dependent glucokinase

PTS – phosphotransferase system

RGM – rapidly-growing mycobacteria

ROK – repressor open reading frame kinase

SGM – slowly-growing mycobacteria

SIS – sugar isomerase

TB – tuberculosis

TLC – thin-layer chromatography

TreS – trehalose synthase

UDPGP – UTP-glucose-1-phosphate uridylyltransferase

CHAPTER 1 – INTRODUCTION

1.1. The order *Actinomycetales*

The phylum *Actinobacteria* represents one of the largest and earliest prokaryotic lineages and accommodates great morphological, physiological and metabolic diversity among its various genera (Gao & Gupta 2012; Ventura et al. 2007). *Actinomycetales*, the largest order within *Actinobacteria*, includes families such as the *Actinomycetaceae*, *Corynebacteriaceae*, *Frankiaceae*, *Mycobacteriaceae*, *Nocardiaceae*, *Streptomycetaceae* and *Streptosporangiaceae* to name only a few (Zhi et al. 2009). Although generally cited as consisting of gram-positive bacteria with a high G+C content but recently enlarged to accommodate low G+C members (Ghai et al. 2011), this group also includes some prominent members, notably the *Mycobacteriaceae*, with remarkably complex and unique cell envelope structures that differ from both gram-positive and gram-negative cell walls (see section 1.3.1) (Angala et al. 2014; Lopez-Marin 2012).

Most species within *Actinomycetales* (also commonly referred to as "actinomycetes") are environmental saprophytes and this group is well represented among microbial soil and water communities in both terrestrial and marine environments (Goodfellow & Williams 1983; Stach & Bull 2005). It also includes notorious pathogens responsible for well-known diseases such as tuberculosis (*Mycobacterium tuberculosis*) and diphtheria (*Corynebacterium diphtheriae*), while simultaneously providing humanity with the best weapons against infectious diseases themselves through its numerous antibiotic-producing members (namely *Streptomyces*) (Doroghazi & Metcalf 2013; Ruiz et al. 2010). Standing as essential members of microbial ecosystems, as a key source of antibiotics and other important secondary metabolites, or as relevant agents of disease, the *Actinomycetales* are undoubtedly of great importance at many levels. Thus, there is an obvious interest in furthering our understanding of both pathogenic and environmental species, discovering new enzymes, deciphering unique metabolic pathways and gathering knowledge that may prove helpful in humanity's ongoing struggle with infectious diseases.

1.2. The genus *Mycobacterium*

Mycobacterium is the type and until very recently the only genus within the *Mycobacteriaceae* family, now joined by *Amycolicicoccus* (Wang et al. 2010). The first members of the genus to be discovered were obligate pathogenic species, namely *Mycobacterium tuberculosis* and *Mycobacterium leprae*, which for historical and clinical reasons remain the most well-known mycobacteria as the cause of tuberculosis (TB) and leprosy, respectively. These diseases plagued humankind for millennia and still persist nowadays as important health problems and, particularly in the case of tuberculosis, difficult to control infections (Galagan 2014; Legendre et al. 2012; Rodrigues & Lockwood 2011). *M. leprae*, the etiological agent of leprosy, was discovered in 1873 by Hansen and is an intracellular parasite with a downsized genome known to infect tissue macrophages of the dermis and Schwann cells of the peripheral nerves (Cole et al. 2001; Scollard et al. 2015). This leads to extensive skin and nerve damage, which may be permanent and result in severe disability, disfigurement and the associated social stigma (Suzuki et al. 2012). *M. tuberculosis* was discovered in 1882 by Koch and since then has been joined by other TB-causing species and variants, together forming the *M. tuberculosis* complex (*M. tuberculosis*, *M. bovis*, *M. africanum*, *M. canettii*, *M. microti*, *M. caprae*, *M. pinnipedii*, *M. mungi, M. orygis, M. suricattae* and the dassie bacillus) (Ingen et al. 2012; Mostowy et al. 2004; Parsons et al. 2013). The members of this complex diverge in terms of host species preference and geographical distribution, but all share a remarkable ability to survive and multiply within lung macrophages upon inhalation by ingeniously manipulating the cell's phagocytic and signaling pathways (Poirier & Av-Gay 2012; Vergne et al. 2004). This strategy allows *M. tuberculosis* to persist in the infected host for decades in a latent, asymptomatic stage, and it has been estimated that one third of the world population is at present in these circumstances (Parrish et al. 1998). When *M. tuberculosis* is able to overcome this dormant state, either for reasons related with the virulence of the strain or with the immune status of the host, active pulmonary TB develops, with consequent host lung damage and dissemination to new, uninfected hosts through cough (Ernst 2012; Gengenbacher & Kaufmann 2012; Lee et al. 2011). Although pulmonary TB is the most common presentation of the disease, *M. tuberculosis* has

tropism for other organs besides the lung and TB can develop anywhere in body, originating cases of extrapulmonary TB and, when dissemination through the bloodstream occurs, of a serious condition called miliary TB (van Crevel et al. 2002; Peto et al. 2009; Sharma & Mohan 2004).

The strictly parasitic lifestyle of *M. leprae* and *M. tuberculosis* has enormous implications for human health, but it is hardly representative of the *Mycobacterium* genus. In fact, of the one hundred and seventy-two species described to date, only a few are obligate pathogens (www.bacterio.net/mycobacterium.htμ). Rather, and much like is typical of actinomycetes, most mycobacteria are environmental saprophytes and inherent members of soil and water microbial communities which can nonetheless cause opportunistic infections (Falkinham 2009a; Primm et al. 2004). These environmental species are collectively called nontuberculous mycobacteria (NTM, see section 1.2.1) to distinguish them from TB-causing mycobacteria (Falkinham 2009a).

Despite the diversity of lifestyles and habitats found within the *Mycobacterium* genus, all mycobacteria share some general characteristics. They are aerobic, rod-shaped, non-sporulating microorganisms with a uniquely complex and lipid-rich cell envelope (Rastogi et al. 2001). The mycobacterial cell wall (see section 1.3.1) is indeed one their most distinctive features, being responsible for much of their resilience and adaptability to various environments including to the infected host, as well as for their abnormal gram stain and acid-fast properties (Falkinham 2009a; Rastogi et al. 2001; Trifiro et al. 1990).

Traditionally, mycobacteria are divided into two groups based on how quickly they form visible colonies on agar plates: slowly-growing mycobacteria (SGM), which take more than 7 days of incubation for visible colonies to form, and rapidly-growing mycobacteria (RGM), which form visible colonies in the first 7 days of incubation (Shinnick & Good 1994). Phylogenetic studies generally support this SGM/RGM division, with a few exceptions, and have shown that while slow growers are a strictly monophyletic group, rapid growers are a polyphyletic group and seem to evolutionarily precede slow growers (Devulder et al. 2005; Stahl & Urbance 1990). Furthermore, it is generally perceived that pathogenicity is more often associated with SGM than with RGM species, although there are important and increasingly notorious

exceptions (De Groote & Huitt 2006; Shinnick & Good 1994). It should be noted that even RGM still grow substantially slower than most bacteria (as an example, *Escherichia* and *Bacillus* have typical generation times of ≈20-30 min compared with 3 h for the RGM *M. smegmatis* and 24 h for the SGM *M. tuberculosis*), and this aspect of mycobacterial physiology greatly influences their ecology and some key features of the infections that they cause, such as long incubation times and latency (Beste et al. 2009; Falkinham 2009a; Suzuki et al. 2010). Slow growth in mycobacteria has been mainly attributed to a low number of 16S rRNA cistrons, cell wall impermeability to nutrients, and to the high cost of mycolic acid synthesis (Primm et al. 2004).

1.2.1. Nontuberculous mycobacteria (NTM)

Environmental species of mycobacteria, called nontuberculous mycobacteria (NTM), are ubiquitous inhabitants of natural waters and soils and of human-engineered environments, such as water distribution systems (Falkinham 2015). Their relative slow growth when compared with most environmental microbes could suggest that NTM are poor competitors, but their remarkable ability to withstand extreme stress conditions and adapt gives them a clear advantage over these more rapidly growing species when it comes to colonizing and thriving in harsher environments (Falkinham 2009b). Thus, NTM tend to preferentially occupy ecological niches where trading fast growth for increased resilience and endurance pays off.

Instances of NTM resilience are manifold and include their oligotrophic lifestyle, as they are able to grow at very low nutrient levels and display an astonishing starvation survival (Norton et al. 2004; Smeulders et al. 1999). Some species are able to metabolize several pollutants, including aromatic and halogenated hydrocarbons, allowing them to thrive on polluted waters and soils (Falkinham 2010; Wang et al. 2006). Generally, NTM have a high tolerance to pH and temperature extremes, to desiccation, and to low oxygen levels (Falkinham 2015; Harland et al. 2008; Kirschner et al. 1999; Schulze-Robbecke & Buchholtz 1992). Due to their very hydrophobic and impermeable cell envelope, NTM display an impressive resistance to heavy metals, disinfectants and antibiotics (Falkinham 2010). This surface hydrophobicity also makes

NTM excellent biofilm formers, allowing them to attach strongly to surfaces thus preventing them from being washed out and diluted in water flowing systems (Falkinham 2009b; Williams et al. 2009).

Natural habitats from which large numbers of NTM have been isolated include peat-rich boreal forest soils and waters, acidic brown-water swamps, estuaries, coastal swamps and acidic hot springs (Falkinham 2009b; Niva et al. 2006; Santos et al. 2007). These are strongly acidic and nutrient-poor environments, but generally rich in humic and fulvic acids, which have been shown to stimulate NTM growth, demonstrating that environmental mycobacteria are very well adapted to this type of habitat (Kirschner et al. 1999). Consequently, they are also present in commercially available peat-rich potting soils and that is one of the many routes of human exposure to NTM (De Groote et al. 2006). Association with plants has also been documented, with mycobacteria found among the endophytic microbial community of rice and wheat roots (Conn & Franco 2004; Mano et al. 2007), associated with barley roots (Child et al. 2007), and living within the shoots of pine trees and grasses (Koskimäki et al. 2010; Laukkanen et al. 2000).

The impact of human activity on NTM ecology has been substantial and many human-engineered environments have proven to be ideal niches for environmental mycobacteria to occupy (Falkinham 2010). Some mycobacteria seem to thrive on heavily polluted soils and have been proposed as having a role in natural attenuation of contaminated sites (Meissner & Falkinham 1984; Wang et al. 2006). The reckless discharge of antibiotics, disinfectants and other biocides in the environment is also thought to selectively favor NTM and lead to their enrichment in perturbed habitats. Indeed, not only do mycobacteria resist to these antimicrobial agents but they also benefit from the deleterious action of such compounds on their more susceptible microbial competitors (Falkinham 2009b). The most flagrant example of a human-engineered environment where these dynamics are at play is the water supply network. Chlorination is the preferred method of disinfection in water distribution systems and has been an invaluable tool in protecting human populations from waterborne diseases such as cholera and typhoid (Dawson & Sartory 2000). However, NTM are highly resistant to chlorine disinfection and are able to colonize, persist,

survive and grow in chlorinated tap water (Le Dantec et al. 2002). Moreover, NTM have been found to be enriched in showerheads, where they form biofilms that render them even more resistant to antimicrobials and less likely to be washed out by high water flow rates (Falkinham 2009b; Feazel et al. 2009). This allows them to stably colonize point-of-use water sources and represents an important route of human exposure to opportunistic mycobacteria (Decker & Palmore 2014).

The significant overlap between our environmental surroundings and mycobacteria's preferred habitats has greatly increased the opportunities for human exposure to NTM, which in turn raised awareness to their ability to cause serious opportunistic infections in humans (Falkinham 2009a; Primm et al. 2004). The link between NTM and disease is not new, but for many years it was mainly associated with a few well-known species and very specific situations (Grange & Yates 1986; Wolinsky 1992). Over the last two decades, however, the list of known NTM has grown immensely and many of those newly described species have been implicated in human disease (Tortoli 2014; Tortoli 2003).

Perhaps the most pathogenic and clinically significant NTM is *Mycobacterium ulcerans*, the causative agent of the third most common mycobacterial disease worldwide after tuberculosis and leprosy, Buruli ulcer (Silva et al. 2009). Still a largely neglected disease, Buruli ulcer is endemic to 33 countries in Africa, South America and the Western Pacific, but rare or nonexistent everywhere else (World Health Organization 2015). The most defining feature of the slowly-growing *M. ulcerans* is the synthesis of the exotoxin mycolactone, a polyketide-derived macrolide which mediates the necrotizing skin and bone lesions that characterize the disease (Walsh et al. 2008). *M. ulcerans* is an environmental pathogen that humans acquire from contact with its aquatic niches, namely rivers, ponds, swamps and lakes of endemic regions (Morris et al. 2014). A close relative of *M. ulcerans* is *M. marinum*, another aquatic-dwelling SGM species associated with skin infections in humans (Doig et al. 2012). *M. marinum* causes a chronic progressive disease in both fresh and salt water fish and is thus adapted to infect ectotherms, explaining its low optimal growth temperature of around 30°C (Stamm & Brown 2004). In humans exposed to infected fish, *M. marinum* can cause a granulomatous localized skin infection called "fish tank granuloma" and in

immunocompromised individuals it may develop into disseminated disease (Petrini 2006). Some RGM species are also frequently associated with skin and soft tissue infections, namely *M. abscessus*, *M. chelonae* and *M. fortuitum*, which have been the cause of many outbreaks of post-surgical infections (Tortoli 2009).

Besides infecting the skin and soft tissues, NTM may also cause lung infections, lymphadenitis, bone and joint infections and disseminated disease (Tortoli 2009). Pulmonary disease, caused by the inhalation of aerosolized mycobacteria, is a particularly common manifestation of NTM infection, with members of the *M. avium* complex, *M. kansasii*, *M. szulgai*, *M. xenopi*, *M. malmoense* and the three RGM species mentioned above among its most important etiological agents (Glassroth 2008; McGrath et al. 2010; Prevots & Marras 2015). NTM lymphadenitis, observed almost exclusively in children and HIV-infected patients, is mostly associated with *M. avium* complex species, *M. malmoense* and *M. scrofulaceum* (Griffith et al. 2007). Disseminated NTM disease may occur in severely immunocompromised patients, notably in association with AIDS and frequently with the involvement of *M. avium* complex and *M. genavense* (Tortoli 2009).

As mentioned previously, high intrinsic antibiotic resistance is a property common to all NTM, making these infections relatively difficult to treat. Some species, however, are more challenging than others: *M. kansasii* is often cited as one of the most easily treatable NTM pathogens, while *M. abscessus* has earned a reputation as the most drug-resistant mycobacteria, able to cause infections that are almost impossible to fight with available antibiotics (Griffith 2010; Koh et al. 2014; Nessar et al. 2012). Moreover, NTM species are not all equally virulent. Some, like the species mentioned through the present section, are established opportunistic pathogens and their isolation from clinical samples is almost always indicative of disease, while others such as *M. gordonae* are ubiquitous and very frequently isolated but only rarely linked with disease (Arend et al. 2009).

Evidence of human-to-human transmission of NTM is very limited and the surrounding environment remains the most likely source of human NTM infection and colonization (Bryant et al. 2013; Falkinham 2009a). Although exposure to NTM is widespread in the human population, it has been observed that some groups are at a

greater risk of developing disease than others. Indeed, some risk factors pre-disposing to NTM disease have been identified and are mainly associated with immune fragilities of the host (Falkinham 2009a). Moreover, convincing evidence that the incidence of NTM infections is on the rise in the human population is accumulating, confirming the status of opportunistic mycobacteria as emerging human pathogens (Bodle et al. 2008; De Groote & Huitt 2006; Kendall & Winthrop 2013; Wentworth et al. 2013).

1.2.1.1. *Mycobacterium smegmatis*

Mycobacterium smegmatis is a soil-dwelling NTM of the RGM group and one of the first mycobacteria to be discovered, shortly after *M. tuberculosis* (Gordon & Smith 1953). Since then, it has been extensively studied and widely used as a model organism for the study of mycobacteria's basic cellular processes and also as a surrogate model for *M. tuberculosis* research (Kim et al. 2011; Wang et al. 2005). Indeed, *M. smegmatis* shares many features with *M. tuberculosis* and the majority of *M. tuberculosis* proteins have homologues in *M. smegmatis* (Baloni et al. 2014; Tyagi & Sharma 2002). *M. smegmatis* is, however, much more practical and safe to culture in the laboratory owing to its faster growth rate and low pathogenicity. *M. smegmatis* is a Biosafety Level 1 organism with a generation time of 3-4 hours and only rarely reported as a cause of opportunistic infections, while *M. tuberculosis* is a Biosafety Level 3 pathogen with a generation time of 20-24 hours (Long et al. 2012; Singh & Reyrat 2009).

The genome of *M. smegmatis* is significantly larger than that of *M. tuberculosis* with an additional 2000 genes, probably reflecting its saprophytic, free-living lifestyle (Baloni et al. 2014; Long et al. 2012). Several efforts have been made to probe into *M. smegmatis'* intricate metabolic networks by integrating genomic, transcriptomic and proteomic data and important insights have been gained regarding the metabolic plasticity, redundancy and adaptability of this environmental microbe (Baloni et al. 2014; Chopra et al. 2014; Long et al. 2012; Titgemeyer et al. 2007; Wang et al. 2005). These studies rely heavily on sequence-based annotation of mycobacterial genes to infer and assign functions to them while biochemical confirmation of actual protein function is lacking in the majority of cases. This approach is often unreliable in

mycobacteria and such annotations do not always correspond to true biochemical function, underscoring the need for experimental validation and functional characterization of mycobacterial enzymes (Mendes et al. 2011). Such functional studies are not only crucial to guide the development of new strategies to fight mycobacterial diseases, but may also shed light on some still underappreciated aspects of the physiology and ecology of these unique microorganisms.

1.3. Mycobacterial glycobiology

Mycobacteria synthesize a wide array of unique carbohydrates and glycoconjugates and possess a remarkably diverse and complex glycobiology that renders this one of the most fascinating topics in mycobacterial research. The importance of carbohydrates and glycoconjugates in the physiology of these microorganisms is highlighted by the fact that the biosynthetic machinery behind these compounds is widely seen as a primary target for antimycobacterial drugs (Kalscheuer et al. 2010; Mendes et al. 2012; Mishra et al. 2011; Nobre et al. 2014). In this section, a brief overview of some relevant and unique aspects of mycobacterial glycobiology is given.

1.3.1. Cell envelope glycoconjugates

The cell envelope of mycobacteria (Figure 1) is unique in its composition and characteristics, and is a central aspect of mycobacterial glycobiology due to the many complex carbohydrates and glycoconjugates it harbors (Nobre et al. 2014).

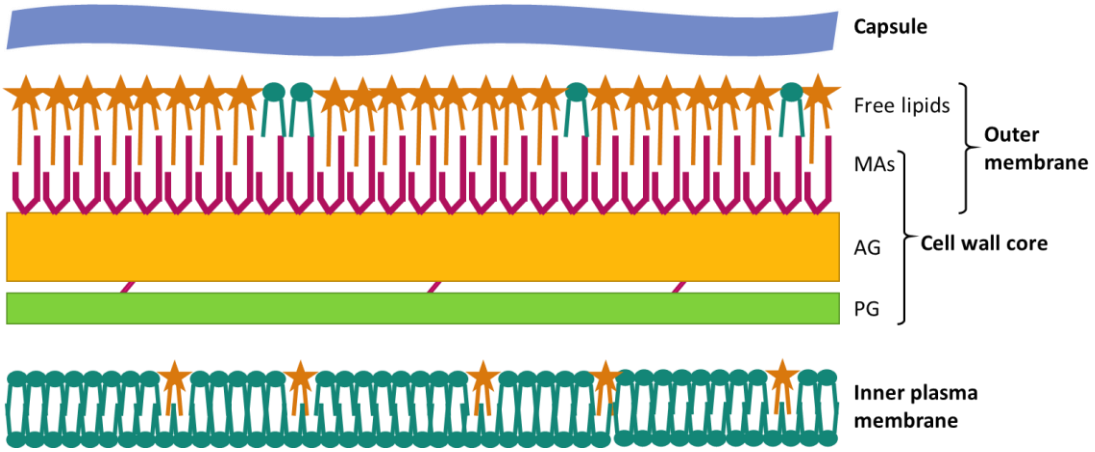

Figure 1. Schematic representation of the mycobacterial cell envelope and its layers. Figure is not drawn to scale. - phospholipids; - all classes of glycolipids; - mycolic acid chains. MAs – mycolic acids; AG – arabinogalactan; PG – peptidoglycan.

This intricate structure, schematically depicted in Figure 1, is formed by three layers surrounding the inner plasma membrane: 1) a cell wall core, 2) an outer membrane (also called “mycomembrane”) and 3) an outer capsule. The cell wall core is composed of a covalently linked complex of peptidoglycan (PG), arabinogalactan (AG), and mycolic acids (MAs), with phosphoryl-*N*-acetylglucosaminosyl-rhamnosyl linkage units attaching PG to AG. Arabinogalactan is in turn esterified to MAs, which are very hydrophobic and long fatty acids (60-90 carbons per chain) with short α-chains. The outer membrane is an asymmetrical bilayer whose inner leaflet is essentially formed by the mycolic acids’ chains, which interact and intercalate with the non-covalently attached lipids present in the outer leaflet. Many of these free lipids as well as some of the lipids of the inner plasma membrane are conjugated with sugar moieties, forming glycolipids that are characteristic of mycobacteria and closely related actinomycetes. These include phosphatidylinositol mannosides, lipoarabinomannan, lipomannan and various acyltrehaloses (including sulfolipids, the highly apolar phthiocerol dimycocerosate and trehalose mono and dimycolates) (Angala et al. 2014; Kaur et al.

2009; Nobre et al. 2014). In addition to these layers, some mycobacteria also have a loose capsule-like structure at the outermost layer of their cell envelope formed by cell-surface polysaccharides and some proteins. The most abundant of these capsular polysaccharides is a high molecular weight branched α-D-glucan with α-(1→4)-glycosidic linkages and α-(1→6) branches every 5 or 6 residues, very similar in structure to glycogen, but D-arabino-D-mannan and D-mannan are also found within this layer (Kaur et al. 2009).

In *M. tuberculosis*, the glycolipids and capsule components of the mycobacterial cell envelope are known to play important roles in host-pathogen interactions and to contribute to virulence and pathogenicity (Brennan 2003; Kaur et al. 2009; Lopez-Marin 2012; Mishra et al. 2011). Moreover, the whole envelope structure forms a remarkably impervious barrier that shields mycobacteria from several biotic and abiotic stress conditions, being responsible for their resilient character as discussed in section 1.2.1. Overall, the importance of the cell wall in mycobacteria's survival and adaptability renders the biosynthetic machinery involved in its assembly and maintenance one of their most attractive and explored drug targets (Mishra et al. 2011).

1.3.2. Cytosolic polysaccharides

In addition to the complex carbohydrates and glycoconjugates of the cell envelope, mycobacteria also synthesize cytosolic polysaccharides like glycogen and the rare polymethylated polysaccharides (PMPs). Glycogen is an important reserve of carbon and energy used by most bacteria, including mycobacteria, and its structure is almost identical to the capsular α-glucan mentioned above (Sambou et al. 2008). PMPs are exclusively found in the *Mycobacterium* genus and in a few strains of closely related genera such as *Nocardia* and *Streptomyces*, and are thought to play a crucial role in the assembly and biogenesis of the mycobacterial cell wall by modulating the synthesis of mycolic acid precursors (Mendes et al. 2012; Pommier & Michel 1986).

Two chemically different but structurally similar classes of PMPs are synthetized by mycobacteria: 6-*O*-methylglucose lipopolysaccharides (MGLPs, Figure 2A) and 3-*O*-

methylmannose polysaccharides (MMPs, Figure 2B). MGLPs are present in both slowly-growing and rapidly-growing mycobacteria, while MMPs occur mostly in RGM. Both are known to assume a helical conformation in solution, creating a hydrophobic cavity which allows formation of stable 1:1 complexes between PMPs and long-chain fatty acids or acyl-coenzyme A derivatives, and to modulate the activity of the enzyme fatty acid synthase I (FAS-I) *in vitro* (Bloch & Vance 1977). These observations have led to the suggestion that PMPs may be crucial regulators of fatty acid synthesis *in vivo* by sequestering the products of FAS-I and thereby promoting enzyme turnover (Jackson & Brennan, 2009). Fatty acids are the building blocks of mycolic acids, so PMPs are likely important for the assembly of the mycobacterial cell wall, rendering the enzymes involved in PMP synthesis potential targets for the development of new antimycobacterial drugs. Although their isolation and characterization was done over 40 years ago (Gray & Ballou 1971; Lee 1966), it was only more recently that progress began to be made regarding the identification of genes and enzymes involved in their biosynthesis (Mendes et al. 2012; Stadthagen et al. 2007).

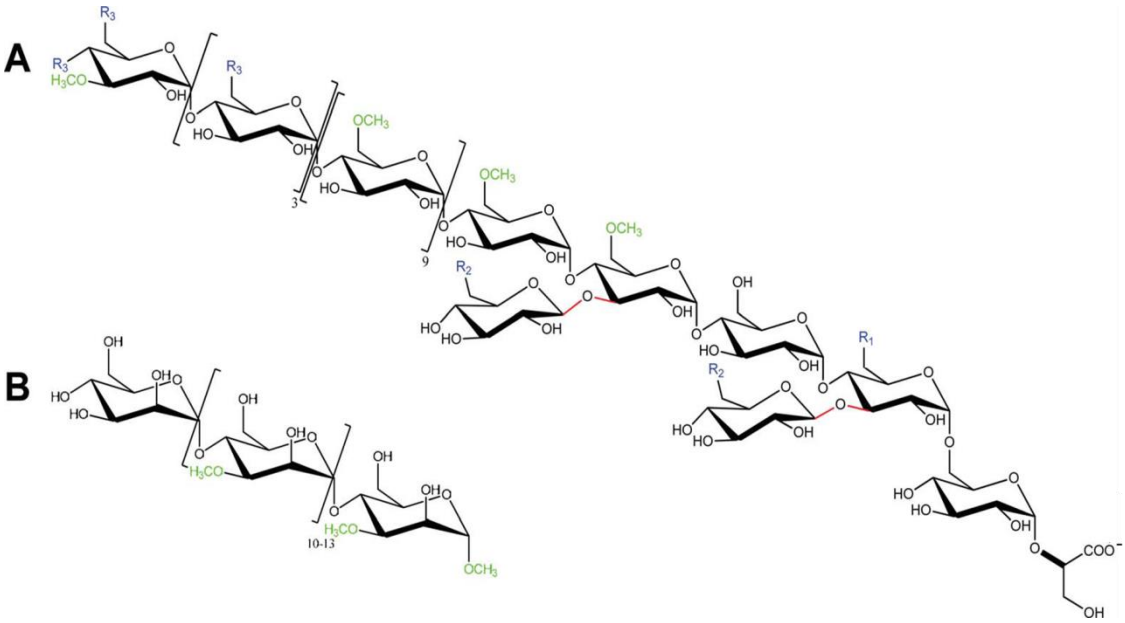

Figure 2. Structure of MGLP from *M. bovis* BCG (A) and of MMP from *M. smegmatis* (B). R1, octanoate; R2, succinate; R3, acetate, propionate or isobutyrate. (Mendes et al. 2012, Maranha et al. unpublished).

1.3.3. Trehalose and the TreS-Mak-GlgE pathway to α-glucans

Trehalose (α-D-glucose(1→1)α-D-glucose) is a ubiquitous and versatile disaccharide universally recognized as a protector molecule that shields proteins and biological membranes from desiccation, freezing, heat and oxidation (Elbein et al. 2003). Some microorganisms, including *Escherichia coli*, accumulate trehalose in their cytoplasm only under stressing conditions, but in mycobacteria trehalose is constitutively present in significant amounts and is essential for survival (De Smet et al. 2000; Woodruff et al. 2004). In mycobacteria and other related genera such as *Nocardia* and *Corynebacterium*, trehalose is not just a protectant and an energy store but also a precursor of many trehalose-containing cell wall components, while in *Streptomyces* it has been associated with spore germination (Nobre et al. 2014). Five different pathways for trehalose synthesis have been identified across the three life domains and while most trehalose-synthesizing organisms have only one, mycobacteria seem to have at least three at their disposal (Nobre et al. 2014; Woodruff et al. 2004). One of them, involving the enzyme trehalose synthase (TreS), has recently been implicated in α-glucan synthesis in mycobacteria (Elbein et al. 2010).

TreS is capable of interconverting maltose and trehalose *in vitro* by isomerizing between maltose's α-(1→4)-linkage and trehalose's α-(1→1)-linkage and was thus proposed to function as a pathway for trehalose biosynthesis. However, it was subsequently shown that *in vivo* the flux through this enzyme flowed in the opposite direction, from trehalose to maltose (Miah et al. 2013). This activity of TreS effectively links trehalose metabolism to the biosynthesis of glycogen and other α-glucans, a link that was initially suspected from the observation that *M. smegmatis* grown in high concentrations of trehalose accumulated higher levels of glycogen, but only if TreS was functional, suggesting that this enzyme was somehow involved in glycogen synthesis (Elbein et al. 2010). Indeed, TreS catalyzes the first step in a non-classical pathway of α-glucan synthesis by converting trehalose into maltose. This maltose is subsequently phosphorylated by Mak, a maltokinase that uses ATP to synthesize maltose 1-phosphate (M1P) (Mendes et al. 2010). M1P is a substrate for GlgE, a maltosyltransferase transfers maltose from M1P to nonreducing ends of glycogen, thus elongating α-(1→4)-glucan chains two glucoses at a time (Kalscheuer et al. 2010). This

key phosphorylated intermediate metabolite is toxic to mycobacteria, and GlgE inactivation has been found to cause self-poisoning due to M1P build-up (Kalscheuer et al. 2010). Finally, the glycogen branching enzyme GlgB transfers oligoglucans from the non-reducing end of linear α-(1→4)-chain to the C-6 hydroxyl group of a glucose residue within the chain, generating α-(1→6) branches. This yields branched glycogen as well as more nonreducing ends for GlgE to elongate.

The physiological functions of this pathway may be multifaceted: glycogen is not the only mycobacterial α-glucan that can be synthesized through the action of these four enzymes, since capsular α-glucan and MGLPs also have a main chain consisting of α-(1→4)-linked glucoses. Thus, the TreS-Mak-GlgE pathway (Figure 3) could be involved in the synthesis of at least three polysaccharides with extremely important physiological functions. Notably, all four enzymes in this pathway (TreS, Mak, GlgE and GlgB) have been proposed to be essential for *M. tuberculosis* growth, turning it into an attractive target for the development of new antimycobacterial drugs (Griffin et al. 2011). Recently, a GlgE inhibitor mimicking M1P was developed (Veleti et al. 2014).

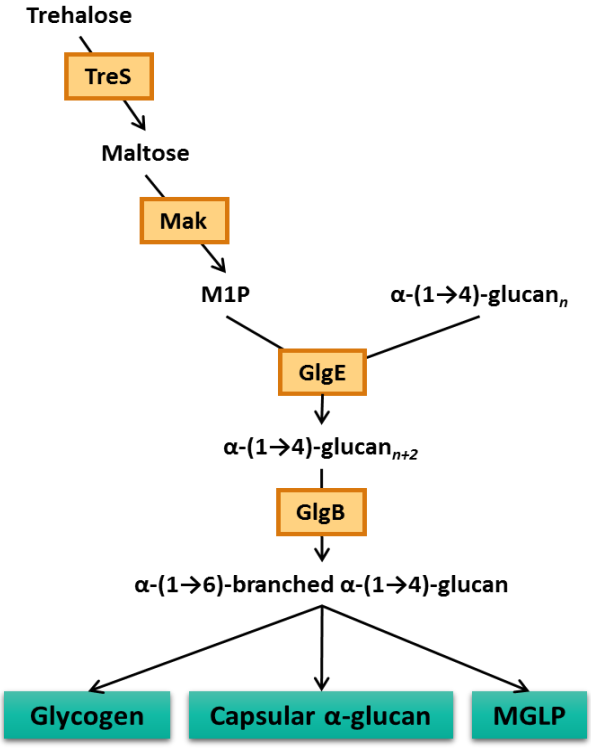

Figure 3. The TreS-Mak-GlgE pathway to α-(1→4)-glucans in mycobacteria. TreS, trehalose synthase; Mak, maltokinase; M1P, maltose 1-phosphate; GlgE, maltosyltransferase; GlgB, glycogen branching enzyme.

1.4. Kinases

Kinases are ubiquitous enzymes that catalyze phosphoryl transfer reactions from high-energy phosphate-donating molecules (e.g., ATP) to macromolecules or metabolites. In doing so, kinases participate in a wide variety of cellular pathways and play essential roles in central and secondary metabolism, signal transduction cascades, gene regulation, protein regulation, muscle contraction, antibiotic resistance, and other cellular processes (Cheek et al. 2002). In addition to the diversity and number of pathways in which they are involved, kinases also display remarkable diversity in terms of sequences, structural folds and substrate specificities. A comprehensive attempt to classify all known kinases based on sequence and structural relationships resulted in the recognition of 25 different families of homologous kinases grouped into 12 distinct fold groups (Table 1) (Cheek et al. 2005). Furthermore, it was concluded that some activities, such as the glucokinase activity, occur in more than one fold group/family, representing cases of convergent evolution in which different scaffolds evolved independently to carry out the same function (Cheek et al. 2002).

According to the IUBMB (International Union of Biochemistry and Molecular Biology) enzyme nomenclature (www.enzyme-database.org), kinases are included in the enzyme subclass EC 2.7 (phosphotransferases or "transferases transferring phosphorous-containing groups") and are subdivided based on substrate specificity, according to the acceptor group. Kinases may transfer phosphate to an alcohol group (EC 2.7.1), to a carboxyl group (EC 2.7.2), to a nitrogenous group (EC 2.7.3) or to a phosphate group (EC 2.7.4). Protein kinases, which phosphorylate proteins on specific amino acid residues, are classified separately and divided among the sub-subclasses EC 2.7.10 to EC 2.7.13 and EC 2.7.99, also according to their specificity.

Within this vast and diverse group of enzymes there are many cases of different catalytic scaffolds converging to catalyze the same reaction and also of the same scaffold diverging to expand its substrate specificity (Allen & Dunaway-Mariano 2004; Bork et al. 1993). Thus, there is no direct relationship between the classification based on substrate specificity and that based on structural and sequence relationships. While

the first is useful for nomenclature and categorizing purposes, only the latter can provide insights into the evolutionary relationships of kinases.

Table 1. Kinase fold groups and families (adapted from Cheek et al. 2005).

Fold Group	Families	Family representative PDB entry
Protein kinase-like	Protein kinase; Lipid kinase; ATP-grasp	1CDK; 1BO1; 1DIK
Rossmann-like	P-loop kinases; Phosphoenolpyruvate carboxykinase; Phosphoglycerate kinase; Aspartokinase; Phosphofructokinase-like; Ribokinase-like; Thiamin pyrophosphokinase; Glycerate kinase	1QF9; 1AQ2; 13PK; 1B7B; 4PFK; 1RKD; 1IG0; 1TO6
Ferredoxin-like	Nucleoside-diphosphate kinase; HPPK; Guanido kinases; Histidine kinase	2BEF; 1EQO; 1BG0; 1I59
Ribonuclease H-like	Ribonuclease H-like	1DGK
TIM β/α-barrel kinase	TIM β/α-barrel kinase	1A49
GHMP kinase	GHMP kinase	1H72
AIR synthetase-like	AIR synthetase-like	1CLI
Riboflavin kinase	Riboflavin kinase	1NB9
Dihydroxyacetone kinase	Dihydroxyacetone kinase	1UN9
Putative glycerate kinase	Putative glycerate kinase	1O0U
Polyphosphate kinase	Polyphosphate kinase	1XDO
Integral membrane kinases	Dolichol kinase	-

1.4.1. Microbial glucokinases

Glucose phosphorylation is a key and early step in nearly all glucose utilization pathways, regardless of whether glucose is to be used for catabolic or biosynthetic purposes and in central or secondary metabolism. Conversion of glucose into glucose 6-phosphate (G6P) is the first step in several important pathways for generation of energy and anabolic precursors such as the Embden-Meyerhof, Entner-Doudoroff and

the pentose phosphate pathways (Jurtshuk 1996). Furthermore, G6P can be isomerized to glucose 1-phosphate by the enzyme phosphoglucomutase and further combined with a nucleoside triphosphate (ATP, UTP, GTP, TTP or CTP) to yield a sugar nucleotide that can act as sugar donor in glycosylation reactions (He & Liu 2002; Samuel & Reeves 2003). In this way, glucose (or other sugar) can be transferred to other molecules by specific glycosyltransferases, involved in processes such as protein glycosylation, polysaccharide synthesis and elongation, assembly of some cell wall components and synthesis of sugar-containing antibiotics (such as macrolides, aminoglycosides and glycopeptide antibiotics) (Dell et al. 2010; Kaur et al. 2009; Liu & Rosazza 1998; Losey et al. 2001; Piepersberg & Distler 1997; Tuckman et al. 1997). In mycobacteria, intracellular G6P levels have been found to be remarkably high when compared with other genera, which in the case of *M. smegmatis* may build up to 130-fold higher levels than in *Escherichia coli* or *Bacillus megaterium* (Hasan et al. 2010). Although the significance of this observed G6P accumulation in mycobacteria is not yet known, it was proposed to function as a source of reducing power to fight oxidative stress. This would presumably involve the unusual F_{420}-dependent glucose 6-phosphate dehydrogenase, an enzyme that is mostly found in methanogenic archaea and in a few bacterial genera, including *Mycobacterium* and *Nocardia*, and which transfers electrons from G6P to the rare electron carrier F_{420} (instead of transferring them to NADP) (Hasan et al. 2010).

In many bacteria, phosphorylation of glucose is primarily done during its uptake in an ATP-independent manner via the phosphotransferase system (PTS) (Postma et al. 1993). Alternatively, glucose may be transported into cells by other transporters and permeases, in which case phosphorylation is mediated by cytoplasmic kinases (Lunin et al. 2004). In *M. smegmatis* there is evidence for both types of glucose uptake, while in *M. tuberculosis* the PTS seems to be absent (Baloni et al. 2014; Titgemeyer et al. 2007). In the cytoplasm, glucose can be phosphorylated by hexokinases (EC 2.7.1.1), which have broad substrate specificity and are also capable of phosphorylating other hexoses, or by glucokinases (EC 2.7.1.2), which are highly specific for glucose (Mendz & Hazell 1993).

Based on sequence comparisons, microbial glucokinases can be divided into three groups (Lunin et al. 2004; Romero-Rodríguez et al. 2015). Group I (Pfam accession number PF04587) consists of ADP-dependent glucokinases mainly found in archaea but which have also been identified in eukaryotes (Ronimus & Morgan 2004). They participate in a modified Embden-Meyerhof pathway in hyperthermophilic archaea and are structurally similar to *E. coli* ribokinase, which places them into the ribokinase-like family as per the classification outlined in Table 1 and into the Ribokinase Pfam clan (CL0118) (Ito et al. 2003; Sakuraba et al. 2004). Group II (Pfam accession number PF02685) comprises ATP-dependent glucokinases that do not have the classical ROK (repressor open reading frame kinase) sequence motif, including the well characterized *E. coli* glucokinase, and are similar in structure to yeast and human hexokinases (Lunin et al. 2004; Miyazono et al. 2012). Group III glucokinases (Pfam accession number PF00480), mainly found in Gram-positive bacteria and in archaea, are also ATP-dependent with a few exceptions and contain the ROK motif (Miyazono et al. 2011). *Streptomyces* and *Mycobacterium smegmatis* glucokinases are included in this group, as well as polyphosphate (PolyP)-dependent glucokinases, namely *M. tuberculosis* bifunctional ATP/PolyP-dependent and *Microlunatus phosphovorus* strictly PolyP-dependent glucokinases (Imriskova et al. 2005; Lunin et al. 2004; Pimentel-Schmitt et al. 2007). Both Group II and Group III (non-ROK and ROK, respectively) glucokinases share a Ribonuclease H-like fold (Table 1) and belong to the Actin_ATPase Pfam clan (CL0108) (Conejo et al. 2010).

1.4.2. The mycobacterial maltokinase

Maltokinase (Mak, EC 2.7.1.175) was first characterized in *Actinoplanes missouriensis* (order *Actinomycetales*) only 11 years ago, even though its product, maltose 1-phosphate, had already been detected back in the 1960's in *Mycobacterium bovis* BCG cell extracts (Jarling et al. 2004; Narumi & Tsumita 1967). Subsequently, the orthologues from *Streptomyces coelicolor*, *M. bovis* BCG, *M. smegmatis* and *M. vanbaalenii* have also been confirmed to be functional maltokinases (Fraga et al. 2015; Jarling et al. 2004; Mendes et al. 2010; Roy et al. 2013). Although the role of Mak in

bacterial physiology remains largely elusive, Mak encoding genes are widely distributed among the bacterial domain, being present in almost all phyla (Fraga et al. 2015). Mak's involvement in glycogen synthesis was suspected early on from genomic context analyses which revealed this gene to be almost invariably present in a glycogen metabolism gene cluster, often contiguous or fused with *treS*, giving the first hint into the existence of a pathway channeling trehalose to glycogen synthesis through a M1P intermediate (Jarling et al. 2004).

Sequence alignments showed from the beginning that Mak had some homology with aminoglycoside phosphotransferases and eukaryotic protein kinases (Jarling et al. 2004). Recently, the three-dimensional structures of the *M. tuberculosis* and *M. vanbaalenii* Mak proteins were reported, confirming this relationship (Fraga et al. 2015; Li et al. 2014). Like the prokaryotic aminoglycoside phosphotransferases (antibiotic resistance) and 5-methylthioribose kinases (bacterial methionine salvage pathway), maltokinase has a typical eukaryotic-like kinase (ELK) fold that places it within the Protein Kinase fold group and family (Table 1) (Fraga et al. 2015; Hon et al. 1997; Ku et al. 2007). The ELK fold has been demonstrated to be able to sustain remarkable sequence variation and divergence while conserving both fold and catalytic mechanism, explaining both the functional diversity and low sequence identity observed among its members (Kannan et al. 2007; Scheeff & Bourne 2005).

A distinguishing and striking characteristic of Mak that is not observed in other ELKs is its unique N-terminal subdomain, which has no sequence similarity to other known proteins but is topologically similar to protease inhibitors of the cystatin family (Fraga et al. 2015). It had been previously reported that *M. tuberculosis* Mak forms a multimeric enzymatic complex with TreS and that complex formation markedly accelerates maltokinase activity (Roy et al. 2013). Thus, the discovery of this novel subdomain at the N-terminal of Mak led to the proposal that it could have a regulatory function by mediating the interaction with TreS and tethering the enzymatic complex at the site of glycogen synthesis. This would, presumably, ensure adequate regulation of Mak activity and prevent uncontrolled production of the highly toxic M1P (Fraga et al. 2015).

1.5. Objective

A gene annotated as a putative maltokinase was detected in the genome of the model organism *M. smegmatis*, a NTM in which the canonical maltokinase has already been identified and functionally validated (Roy et al. 2013), raising the intriguing possibility of the presence of a second maltokinase in this organism. This gene, absent from the vast majority of mycobacterial species with sequenced genomes, was found to be included in a gene cluster (Figure 4) with other hypothetical genes whose annotations appeared to be unrelated to maltose metabolism or to the recently discovered TreS-Mak-GlgE pathway (Elbein et al. 2010). Given that bacterial genes participating in the same pathway are often grouped into gene clusters (Ma & Xu 2013) and knowing that digital annotation of mycobacterial genes has proven to be unreliable before (Mendes et al. 2011), the objective of this work was to investigate and functionally characterize the enzyme encoded by this enigmatic gene in order to begin understanding its physiological role in the very few actinobacterial species bearing it.

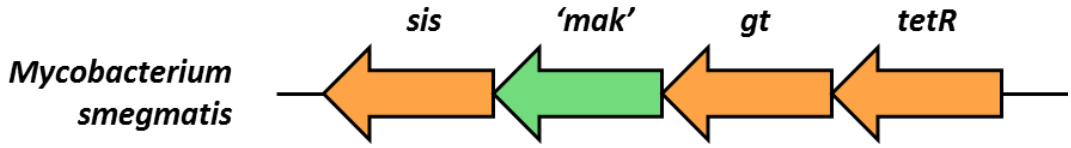

Figure 4. Genomic context of a hypothetical alternative maltokinase gene (shaded green) in *M. smegmatis*. ***sis***, putative sugar isomerase; '***mak***', putative maltokinase; ***gt***, putative glycosyltransferase; ***tetR***, putative TetR family transcriptional regulator.

CHAPTER 2 - MATERIALS AND METHODS

2.1. Sequence analysis, genomic context and phylogenetic tree

The hypothetical alternative maltokinase gene MSMEI_5186 was identified in the genome of *Mycobacterium smegmatis* mc^{2}155 by BLAST searches using the amino acid sequence of the characterized maltokinase gene MSMEI_6342 as query. The nucleotide sequence of the MSMEI_5186 gene and its corresponding amino acid sequence were retrieved from the KEGG database (www.genome.jp/kegg). The amino acid sequence was used to search for homologues in other species by BLAST searches at the National Center for Biotechnology Information database (NCBI, blast.ncbi.nlm.nih.gov/Blast.cgi). Genomic context analyses were performed using PATRIC (patricbrc.org/portal/portal/patric) and KEGG databases. Molecular weight and theoretical isoelectric point were predicted with ExPASy's ProtParam tool (web.expasy.org/protparam). Identification of conserved domains was performed at NCBI's Conserved Domain Database (Marchler-Bauer et al. 2014).

Multiple sequence alignments with the amino acid sequences of MSMEI_5186, its orthologues, and selected maltokinases were done with ClustalW 2.1 (www.ebi.ac.uk/Tools/msa/clustalw2). Phyre2 was used to predict a three-dimensional structural model of the *M. smegmatis* protein (Kelley et al. 2015), which was visualized and superposed with its closest structural homologue using UCSF Chimera software (www.cgl.ucsf.edu/chimera/, MatchMaker tool).

ClustalW (www.genome.jp/tools/clustalw/) was used to generate a phylogenetic tree with selected glucokinase and maltokinase amino acid sequences (UPGMA clustering method).

2.2. *M. smegmatis* mc^{2}155 growth conditions and DNA extraction

Mycobacterium smegmatis mc^{2}155 (ATCC 700084) was obtained from LGC Standards S.L.U. (Barcelona, Spain) and cultivated on agar plates for 5 days at 30°C in a glycerol-based medium containing, per liter: 20 g glycerol (Merck), 5 g casamino acids (Difco), 1 g fumaric acid (Sigma), 1 g K_2HPO_4, 0.3 g $MgSO_4$, 0.02 g $FeSO_4$ and 2 g Tween 80 (Sigma) at pH 7.0 (Brennan & Ballou 1967). Chromosomal DNA was extracted and

purified from bacterial cells with the SmartHelix DNAid Mycobacteria kit (Sekvenator, Slovenia).

2.3. Cloning and heterologous overexpression

2.3.1. PCR amplification

The 1200 bp MSMEI_5186 gene was amplified from the chromosomal DNA of *M. smegmatis* mc^2155 using the forward primer 5'-TAATATCATATGATCGAGCTCGACC-3' and the reverse primer 5'- TATAAGCTTCGTGCTTGTCCC-3', designed to include a *Nde I* and a *Hind III* restriction site, respectively (underlined). The start codon, originally the alternative and rarer GTG, was replaced by the standard ATG start codon in the forward primer which is also part of the *Nde I* restriction sequence. The stop codon was removed from the reverse primer to allow for the translation of a C-terminal 6xHis-tag from the expression vector. PCR amplification was carried out with proofreading KOD Hot Start DNA polymerase (Novagen) according to the manufacturer's instructions and with the following conditions: 2 min pre-incubation step at 95°C, followed by 35 cycles of denaturation at 95°C for 20 s, annealing at 54°C for 10 s and extension at 70°C for 36 s, and a final elongation step of 10 min at 70°C. The PCR product was analyzed by agarose gel electrophoresis (Annex I, section 1) and the 1200 bp band was excised from the gel and purified with JETquick Gel Extraction Spin Kit (Genomed) according to the suppliers' instructions.

2.3.2. Cloning and transformation of *Escherichia coli*

The amplified gene and the cloning/expression vector pET30a (Novagen) were digested with *Nde I* and *Hind III* restriction enzymes (New England BioLabs) at 37°C for 1 h, visualized by agarose gel electrophoresis and purified from the gel as described above. The linearized vector was further dephosphorylated with Alkaline Phosphatase Calf Intestinal (New England BioLabs) and purified with JETquick PCR Product Purification Spin Kit (Genomed). Both fragments were ligated with ExpressLink™ T4 DNA Ligase (Invitrogen) at room temperature for 15 minutes in a 20 µL reaction

mixture. The recombinant plasmids were transformed into *E. coli* BL21 Star (kind gift of Dr. Isaura Simões, CNC/UC-Biotech) competent cells (Annex I, Section 2) by adding 10 µL of ligation mixture to 100 µL of competent cells, incubating on ice for 30 min and performing a 42° heat shock for 45 s on a pre-heated water bath. The cells were incubated at 37°C for 1 h in LB liquid medium (Annex I, section 3) and subsequently plated on LB agar medium containing 30 µg/mL of kanamycin, followed by overnight incubation at 37°C. Resistant colonies were selected and cultured overnight in 10 mL of LB liquid medium with 30 µg/mL kanamycin in an orbital shaker at 37°C, 150 rpm. The cells were harvested by centrifugation and plasmid extraction and purification was carried out with ZR plasmid miniprep™-classic kit (Zymo Research). Positive clones were confirmed by restriction analysis, digesting extracted plasmids with *Nde I* and *Hind III* and visualizing the product by agarose gel electrophoresis. Recombinant plasmids were sequenced to confirm the identity of the inserts (GATC Biotech).

2.3.3. Overexpression

E. coli BL21 Star cells carrying a recombinant plasmid were grown in LB liquid medium (2.7 L) containing kanamycin (30 µg/mL) in an orbital shaker at 37°C, 140 rpm, to mid-exponential phase (OD_{610nm}=0.8). Growth temperature was then lowered to 25°C and gene expression was induced by adding IPTG to a final concentration of 0.5 mM. Cells were harvested 18 h later by centrifugation (8000 rpm, 10 min, 4°C), suspended in 20 mM sodium phosphate buffer pH 7.4 with 0.5 M NaCl and 20 mM imidazole (buffer A) containing 5 mM $MgCl_2$ and 10 µg/mL DNase I, and disrupted by sonication on ice with six 60 Hz pulses of 45 s (15 s interval between pulses) per each 5 mL of extract. The lysate was centrifuged to remove cell debris (14000 rpm, 30 min, 4°C) and protein expression was confirmed by SDS-PAGE (Annex I, section 4).

2.4. Purification of recombinant protein

His-tagged recombinant protein was purified in a Fast Protein Liquid Chromatography (FPLC) system with a Ni-Sepharose high-performance column (His-Prep FF 16/10, GE Healthcare) equilibrated with buffer A (section 2.3.3). The protein extract from the previous section was filtered through a 0.2 µm cellulose filter and loaded onto the column. A washing step was performed by increasing imidazole concentration to 45 mM, and elution of the bound recombinant protein was carried out with 200 mM imidazole. Fractions of 5 mL were collected and purity was assayed by SDS-PAGE. The purest fractions were pooled, diluted 10-fold with 20 mM Bis-tris propane buffer (BTP) pH 7.5, concentrated by ultrafiltration in 30 kDa cutoff centricons (Amicon) and equilibrated with the same buffer containing 50 mM NaCl. Protein content was quantified by the Bradford assay (BioRad). The identity of the purified protein was determined by peptide mass fingerprinting after excising the corresponding SDS-PAGE band.

2.5. Analysis of enzyme activity and substrate specificity by thin-layer chromatography

The enzyme activity and substrate specificity of the recombinant enzyme was tested in 50 µL mixtures containing pure enzyme (4 µg), 20 mM of phosphate acceptor and 5 mM of phosphate donor. Reactions were carried out in 50 mM BTP buffer pH 8.0 with 10 mM $MgCl_2$ at 37°C for 2 h or overnight. Equivalent mixtures without enzyme were used as negative controls. The following sugars, sugar derivatives and sugar-containing antibiotics were tested as possible phosphate acceptors: D-maltose, isomaltose, D-trehalose, D-turanose, D-leucrose, sucrose, maltotriose, raffinose, D-glucose, D-galactose, D-mannose, D-allose, D-tagatose, fructose, D-arabinose, glucosylglycerate, D-glucuronic acid, glucosamine, *N*-acetyl-D-glucosamine, *N*-acetyl-D-mannosamine, kanamycin, streptomycin, gentamycin, hygromycin B and clarithromycin (all from Sigma-Aldrich). The following phosphate donors were tested with glucose as the acceptor: ATP, UTP, GTP, ADP and inorganic phosphate (all from Sigma-Aldrich). ATPase activity was accessed in reaction mixtures containing 50 mM

BTP buffer pH 8.0 with 10 mM $MgCl_2$, 5 mM ATP and 3 µg of pure enzyme and incubated overnight at 37°C. The existence of phosphatase activity towards glucose 6-phosphate was also investigated in 50 mM BTP buffer pH 8.0 with 10 mM $MgCl_2$, 5 mM G6P and 3 µg of pure enzyme, in the presence or absence of 5 mM Zn^{2+}.

Product formation was examined by thin-layer chromatography (TLC) (Silica Gel 60, Merck). Reactions and respective negative controls were spotted onto a TLC plate and developed using one of the two following solvent systems: acetic acid/ethyl acetate/water/ammonia 25% (6:6:2:1, vol/vol) or ethanol/water (7:3, vol/vol). Substrates and products were visualized by spraying the plate with α-naphthol-sulfuric acid solution (Annex I, section 5) and charring it at 120°C for 5 to 10 minutes (Jacin & Mishkin 1965).

2.6. Purification and chemical identification of phosphorylated product

2.6.1. Glucose-phosphate purification

The glucose-phosphate enzymatic product obtained upon incubation of the recombinant enzyme with glucose and ATP was overproduced and purified for identification purposes. Ten reactions of 2 mL containing 50 mM BTP pH 8.0, 10 mM glucose, 5 mM ATP and 2.5 µg of recombinant enzyme were incubated at 37°C overnight to ensure total ATP consumption. The phosphorylated compound was subsequently purified using liquid chromatography techniques, following a protocol adapted from a process developed for purification of maltose 1-phosphate, which was the subject of a provisional patent application filed recently (Empadinhas, Maranha, Mendes, 2014).

2.6.2. Two-dimensional NMR spectroscopy

The phosphorylated enzymatic product was identified by two-dimensional nuclear magnetic resonance (2D NMR) spectroscopy, employing the ^{1}H-^{13}C heteronuclear single quantum correlation (^{1}H-^{13}C HSQC) experiment. The 500 MHz ^{1}H-^{13}C HSQC spectrum was acquired at 298 K and compared with standard spectra retrieved from the Biological Magnetic Resonance Bank database (www.bmrb.wisc.edu). The spectra were obtained at Dr. Brian Goodfellow's laboratory (http://www.ciceco.ua.pt/BrianGoodfellow) at the University of Aveiro.

2.7. Enzyme assays

Enzyme activity was measured using three different methods.

ADP formation was quantified spectrophotometrically by an indirect method as described previously (Mendes et al. 2010). Reactions and negative controls (50 µl) were performed in duplicate under the desired conditions and stopped after 5, 10 or 15 minutes by cooling on ethanol-ice followed by enzyme inactivation with 5 µl of 5N HCl and neutralization with 5 µl of 5N NaOH. The samples were brought to a final volume of 1 mL with 50 mM Tris-HCl buffer pH 7.4 containing 2.5 mM $MgCl_2$ and 2.5 mM KCl and incubated with 3 U of pyruvate kinase and lactate dehydrogenase, 0.3 mM NADH and 2.5 mM phosphoenolpyruvate (all from Sigma-Aldrich) at 30°C for 10 min. Absorbance at 340 nm was measured in 1 mL cuvettes to determine NADH consumption under each condition.

G6P formation was quantified spectrophotometrically by an indirect method adapted from Zhang *et al* (Zhang et al. 2011). Reactions and negative controls (50 µl) were run in duplicate under the desired conditions and stopped after 5, 10 or 15 minutes by cooling on ethanol-ice followed by enzyme inactivation as described above. The samples were brought to a final volume of 150 µL with 50 mM BTP buffer pH 8.0 containing 5 mM $MgCl_2$ and incubated with 0.3 U of glucose 6-phosphate dehydrogenase from *Leuconostoc mesenteroides* and 1.5 mM NAD (all from Sigma-

Aldrich) at 30°C for 10 min. Absorbance at 340 nm was measured in 96-well microtiter plates to determine NADH formation under each condition.

To follow G6P release in real-time, a coupled enzyme assay system adapted from Skarlatos and Dhal was used (Skarlatos & Dahl 1998). Assay mixtures (150 µL total reaction volume) containing a similar reaction mixture to the one previously mentioned, 50 mM BTP buffer pH 8.0, 20 mM $MgCl_2$, with varying glucose and ATP concentrations, plus 1.5 mM NAD and 0.3 U of glucose 6-phosphate dehydrogenase from *Leuconostoc mesenteroides* (all from Sigma-Aldrich), were prepared. Reactions were initiated by adding 4 µg of pure recombinant enzyme and incubated at 37°C for 15 min. Absorbance at 340 nm was monitored in 96-well microtiter plates during incubation at 0, 5, 10 and 15 min time points.

2.8. Biochemical and kinetic characterization

Temperature and pH profiles and kinetic parameters were determined for the glucokinase activity. The temperature profile was determined between 25 and 55°C in assay mixtures with 50 mM BTP pH 8.0, 20 mM $MgCl_2$, 5mM ATP and 20 mM glucose using the G6P terminal quantification method described above. The pH profile was determined at 37°C in 50 mM buffer BTP (6.5 to 9.5) or CAPSO (9.5 to 10) with 20 mM $MgCl_2$, 5 mM ATP and 20 mM glucose using the same G6P quantification method. The kinetic parameters were determined by following G6P release in real-time as described above at 37°C, in 50 mM BTP pH 8.0 with 20 mM of $MgCl_2$ by varying ATP concentrations with 40 mM glucose or varying glucose concentrations with 5 mM ATP.

The pH profile and kinetic parameters for the ATPase activity were determined using the ADP terminal quantification method described above. The pH profile was determined at 37°C in 50 mM buffer BTP (6.5 to 9.5) with 10 mM $MgCl_2$, 5 mM $ZnSO_4$ and 5 mM ATP. The kinetic parameters were determined at 37°C, 50 mM BTP pH 8.0 in the presence of 10 mM $MgCl_2$, 5 mM $ZnSO_4$ and varying ATP concentrations.

The effect of divalent cations on the enzymatic activity was examined at 37°C in reactions containing 50 mM BTP pH 8.0, 20 mM glucose, 5 mM ATP and varying

concentrations of the chloride salts of Mg^{2+}, Co^{2+}, Mn^{2+} or Ca^{2+} or the sulfate salts of Zn^{2+} or Cu^{2+}, or in the presence of 2 mM EDTA. Combinations of these cations were also tested. Under each condition, ATPase activity was measured by the ADP quantification method described above. For the glucokinase activity, the effect of different concentrations of Mg^{2+} was measured by the G6P quantification method as described above and the remaining conditions, including effect of other cations and combinations of cations, were examined by TLC.

K_m and V_{max} values for both the glucokinase and ATPase activities were obtained with GraphPad Prism (version 5.00 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com) using the Michaelis-Menten equation (nonlinear regression). All experiments were performed in duplicate with appropriate controls.

CHAPTER 3 - RESULTS

3.1. Sequence analysis, genomic context and phylogenetic tree

The genome of the strain *M. smegmatis* mc^2155 has been annotated three times (2006, 2012 and 2014), but BLAST searches using the amino acid sequence of the canonical maltokinase gene identified a putative second maltokinase gene with 26% sequence homology only in the two latest versions of the genome, corresponding to MSMEI_5186 in the 2012 version and to LJ00_26350 in the 2014 version (www.genome.jp/kegg/). In the first version, the matching nucleotide sequence is included in a region annotated as pseudogene and is assigned the code MSMEG_5331. Since 2006 several corrections have been made to the genome of *M. smegmatis* mc^2155 mainly as a result of the detection of mistakes arising from both sequencing errors and incorrect start codon predictions (Deshayes et al. 2007; Gallien et al. 2009). MSMEI_5186 is thus one of the protein encoding genes that were missed during the first genome annotation and subsequently recognized as such. Based on sequence similarities, it is currently annotated as either a putative maltokinase or aminoglycoside phosphotransferase gene.

The 1200 bp MSMEI_5186 gene encodes a 399 amino acid polypeptide chain with an estimated molecular mass of 42.3 kDa and a calculated isoelectric point of 5.16. It has a conserved catalytic domain of the Protein Kinase superfamily (PKc-like). BLAST analyses with its amino acid sequence revealed homologues in only a few species of mycobacteria, namely *M. iranicum* (61%), *M. rufum* (61%), *M. aromaticivorans* (60%), *M. cosmeticum* (58%), *M. vaccae* (57%), *M. obuense* (56%) and *M. neoaurum* (52%). More distantly related homologues were also identified in a few other actinomycetes, namely *Streptacidiphilus jiangxiensis* (44%), *Mumia flava* (43%), *Frankia* sp. EuI1c (41%) and *Streptosporangium roseum* (39%). The maltokinase from *M. smegmatis*, whose activity has been experimentally validated (Roy et al. 2013), shares 26% amino acid identity with the protein encoded by MSMEI_5186. BLAST searches also confirm the existence of canonical maltokinases in all species cited above.

This gene is included in a rare but conserved gene cluster (Figure 5), which in mycobacteria also includes open reading frames (*orfs*) predicted to encode a SIS

domain-containing putative phosphosugar isomerase, a putative glycosyltransferase and a putative TetR family protein, none of which have been characterized. The same scheme is found in *Mumia flava* and in *Streptacidiphilus jiangxiensis* where two additional genes can be found within the cluster, coding for a putative phosphoglucomutase and for a putative UTP-glucose 1-phosphate uridylyltransferase. The TetR family gene is absent from *Frankia*, and in *Streptosporangium* the four genes are present, but there are four hypothetical genes annotated as spermidine/putrescine ABC transporter components inserted in the middle of the gene cluster.

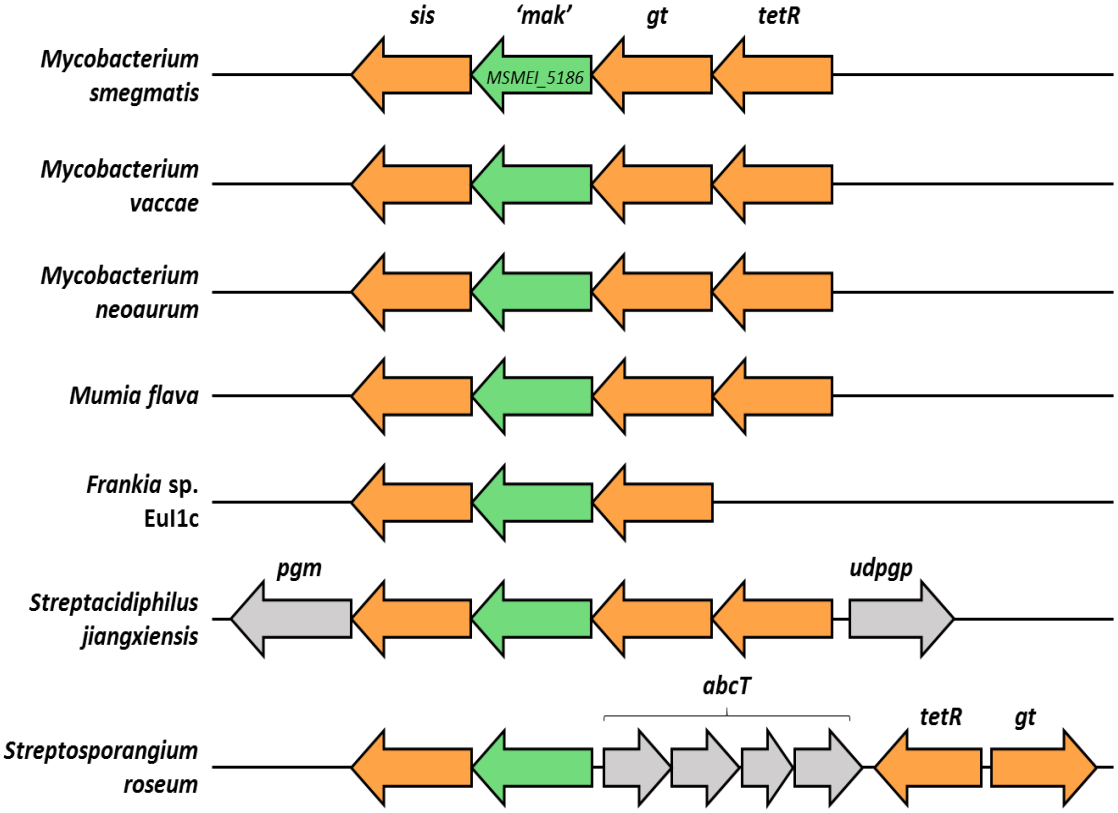

Figure 5. Genomic context of the MSMEI_5186 gene and of its orthologues (shaded green) in different actinobacteria. Shaded orange are neighboring genes which are also conserved across the different organisms. Arrows represent open reading frames and their orientation but are not to scale. ***sis***, putative sugar isomerase; ***'mak'***, putative maltokinase; ***gt***, putative glycosyltransferase; ***tetR***, putative TetR family transcriptional regulator; ***pgm***, putative phosphoglucomutase; ***udpgp***, putative UTP-glucose-1-phosphate uridylyltransferase; ***abcT***, putative ABC transporter components.

To investigate the relationship between MSMEI_5186 (GluK) and the related and well characterized canonical maltokinase, selected sequences of both proteins were

aligned (Figure 6) to search for differences and similarities in key conserved motifs and residues, particularly those known to be important for Mak catalytic activity. From the alignment it is evident that GluK is shorter than Mak mainly due to the fact that a significant part of Mak's N-terminal cap domain is missing, including the conserved "RWYAG" motif. When all GluK sequences are aligned (alignment not shown), though, two mostly conserved "WRRA" and "GDG" motifs are found at its shorter N-terminal region. The nucleotide-positioning loop (P-loop) (positions 132 to 138) is largely conserved between Mak and GluK except for the replacement of a conserved glutamate in Mak for a conserved aspartate in GluK and, similarly, of a threonine for a glutamate. The phosphate-binding lysine (148) is also conserved, although the motif in which it is included is different, with a strictly conserved "VVKW" in GluK that is not seen in Mak. The catalytic loop is conserved as well in the form of an HGDXHXGQ motif (319-326) that includes the catalytic base, an aspartate residue widely conserved across different ELK families (Kannan et al. 2007). The DFE-loop (336-342) is mostly conserved and the aspartate residue known to bind magnesium in Mak is also present in GluK, but the two glutamates conserved in Mak are replaced by an aspartate (making it a DFD-loop) and by an asparagine. In Mak, the histidine included in the AXVH motif (251-254) and the aspartate included in the DVA motif (359-361) are engaged in a hydrogen bond network that interlinks the nucleotide and sugar binding pockets and stabilizes the catalytic loop (Fraga et al. 2015). While the DVA motif is well conserved in GluK with the aspartate residue strictly conserved, the AXVH motif is overall poorly conserved and the histidine residue appears missing from two of the aligned sequences. Closer to the C-terminal, Mak has a DKAVYE conserved motif that is absent from GluK, while GluK has a HLPRW motif that is not found in Mak. Also of note, seven of the conserved residues lining the probable maltose binding pocket in Mak (Fraga et al. 2015) are replaced by different residues, equally conserved, in GluK. Namely, three arginines are replaced by leucine/valine (261, 358, and 437), a glutamate by asparagine (342), a tyrosine by histidine (369), a lysine by glutamine (425) and a tyrosine by arginine (428). Overall, the motifs involved in catalysis and in nucleotide binding remain largely conserved in GluK, while those involved in maltose binding show significant disparities.

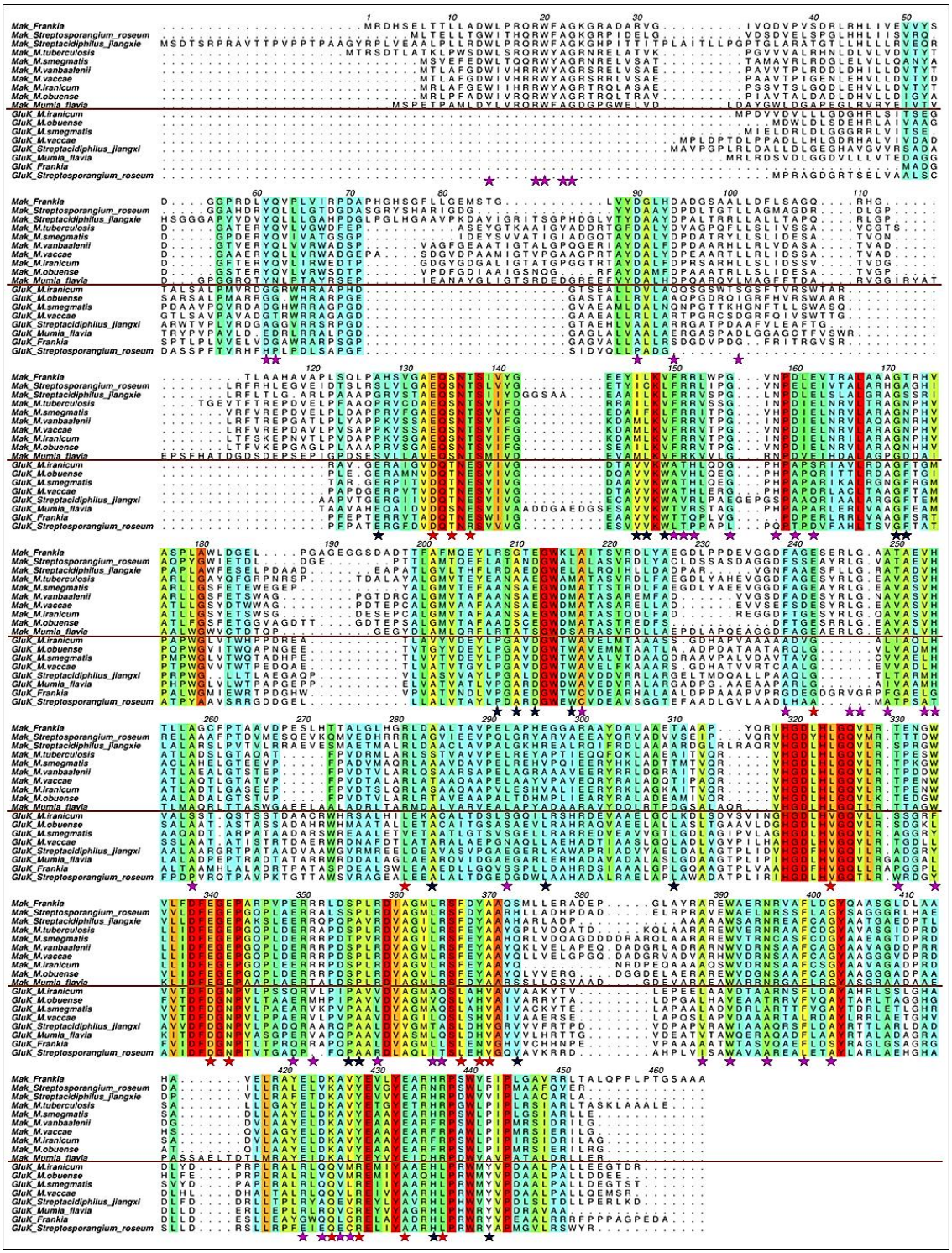

Figure 6. Amino acid sequence alignment of representative actinobacterial maltokinases (Mak) and MSMEI_5186 orthologues (GluK). Residues are colored based on a residue conservation scale (red: identical residues; orange to blue: decreasing conservation of amino acid properties; white: dissimilar residues). Stars signal residues with different conservation patterns between Mak and GluK (pink stars: conserved in Mak but not in GluK; dark blue stars: conserved in GluK but not in Mak; red stars: conserved differently in Mak and GluK. Figure prepared with Aline (Bond & Schüttelkopf 2009).

A predictive model of the three-dimensional structure of the *M. smegmatis* protein based on homology modeling (Figure 7) suggests a very high structural homology with the two mycobacterial maltokinases whose structures have been determined (PDB codes 4U94 and 4O7O), with 91% of the sequence modelled with 100% confidence when these are used as templates.

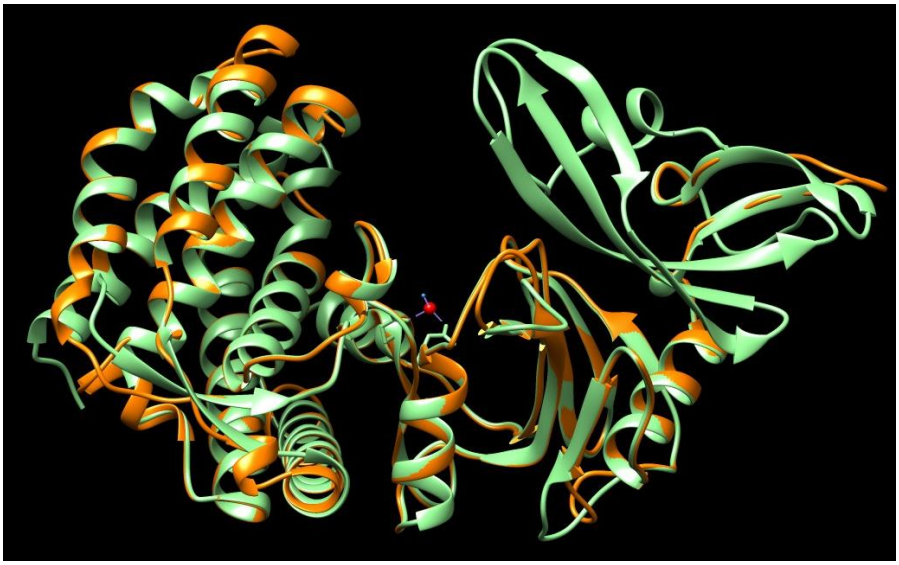

Figure 7. Superposition of the three-dimensional homology model of the *M. smegmatis* GluK protein obtained with Phyre2 and the *M. vanbaalenii* maltokinase structure (PDB code 4U94), evidencing the typical bilobal fold of eukaryotic-like kinases (ELKs). *M. vanbaalenii* Mak structure is colored light green and GluK model is colored orange. Red sphere – magnesium ion; blue sphere – magnesium-coordinating water molecule. Figure prepared with Chimera (http://www.cgl.ucsf.edu/chimera/).

A simple clustering method was used to generate a phylogenetic tree (Figure 8) from the aligned sequences of glucokinases belonging to different families, the glucokinase characterized in this work and some maltokinases. This tree evidences the separation of known microbial glucokinases into three distinct clusters, corresponding to Group I, Group II and Group III glucokinases (Lunin et al. 2004). It is also clear that the glucokinase encoded by MSMEI_5186 does not belong to any of these groups, clustering instead with maltokinases as expected.

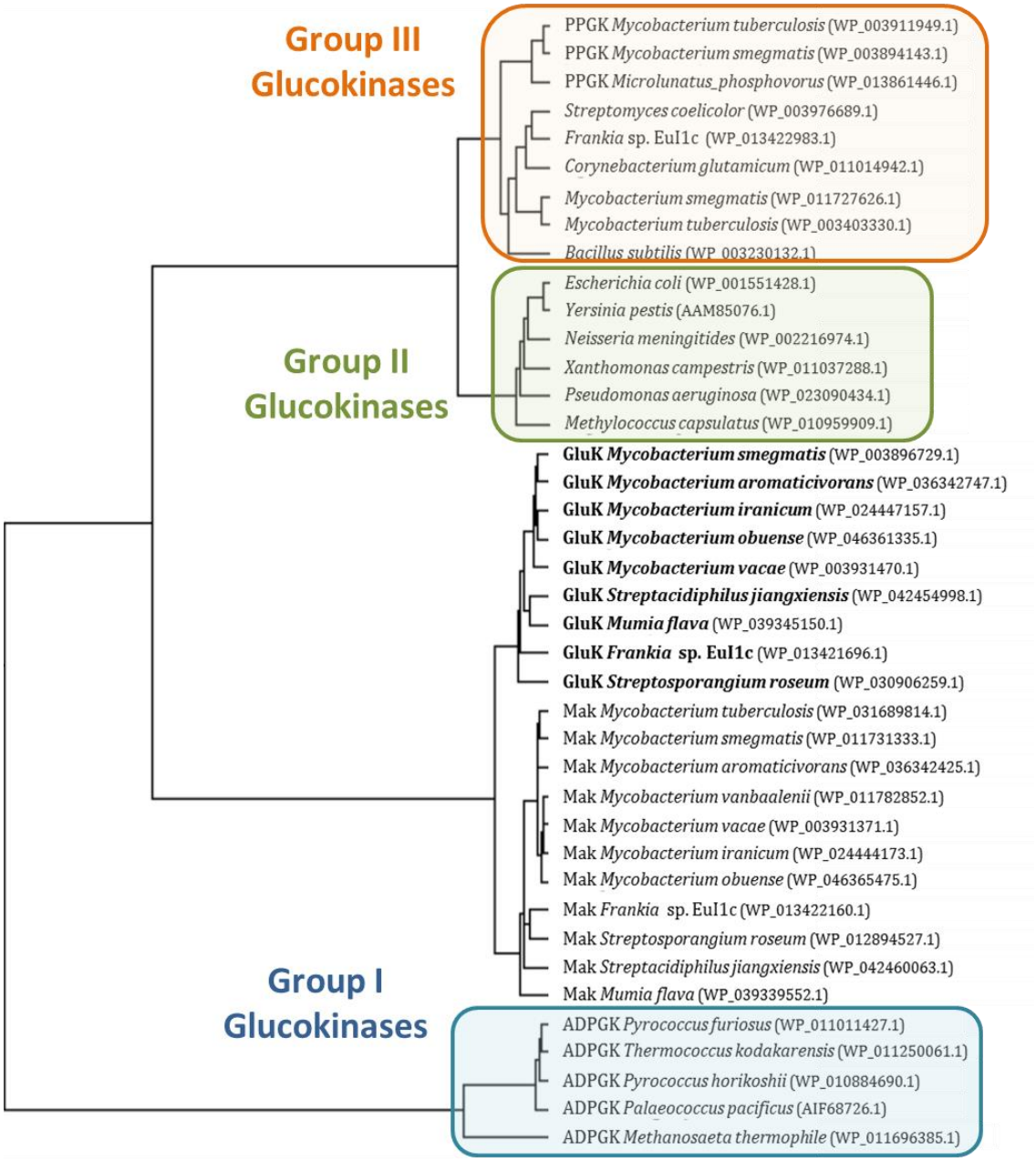

Figure 8. Phylogenetic tree of selected microbial glucokinases and maltokinases based on amino acid sequence. MSMEI_5186 and orthologues are in bold. The three groups of microbial glucokinases (I, II and III shaded blue, green and orange, respectively) are clearly separated from each other and from the GluK/Mak cluster. PPGK, polyphosphate-dependent glucokinase; ADPGK, ADP-dependent glucokinase. RefSeq accession numbers are indicated in parenthesis. The tree was generated by ClustalW using the UPGMA algorithm.

3.2. Expression and purification of the recombinant protein

The His-tagged 43.9 kDa recombinant protein was expressed in *E. coli* in soluble form and purified to homogeneity in one chromatographic step using a Ni-Sepharose high-performance column (Figure 9), as verified by SDS-PAGE (Figure 10), and with a yield of approximately 11 mg of pure protein per liter of culture.

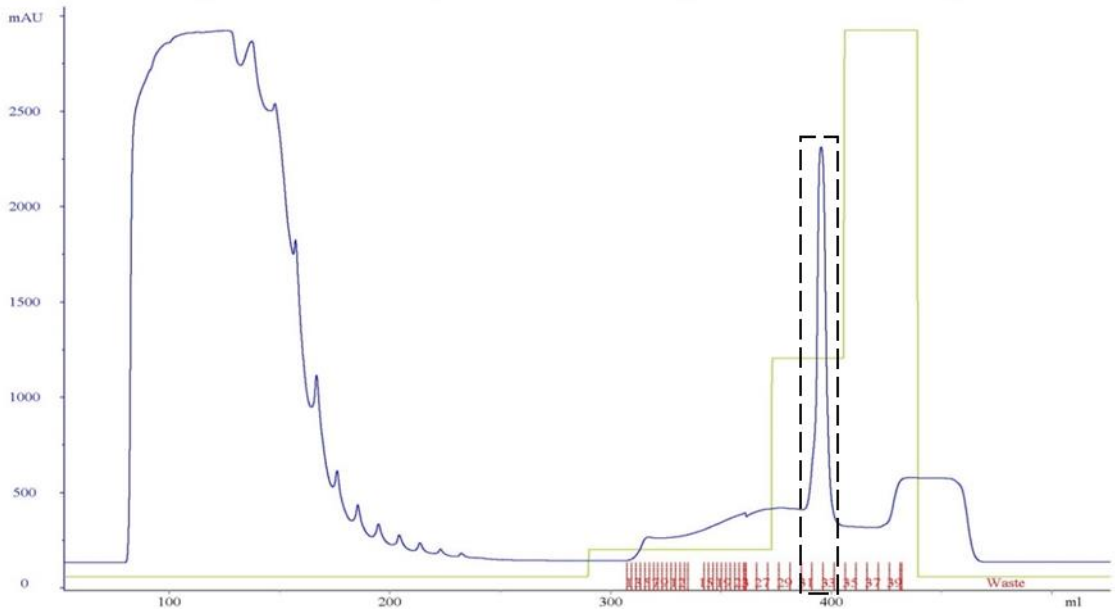

Figure 9. FPLC chromatogram of His-tagged recombinant protein purification using a Ni-Sepharose column. Dashed area represents protein elution and pooled fractions. Blue line – UV absorbance; Green line – buffer B concentration (steps of 0, 5, 40 and 100%).

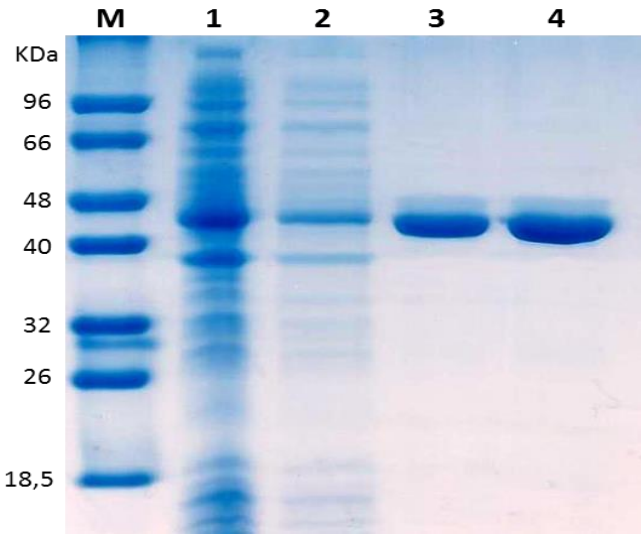

Figure 10. SDS-PAGE of the pure recombinant protein following Ni^{2+}-affinity chromatography. **Lane M** – molecular mass marker. **Lane 1** – recombinant *E. coli* cell extract after sonication; **Lane 2** - soluble fraction of cell extract as loaded on the Ni-column; **Lanes 3 and 4** – 4 µg and 6 µg of purified His-tagged recombinant protein with a predicted MW of 44 kDa.

The purified protein was confirmed as the product of MSMEI_5186 by peptide mass fingerprinting (data not shown).

3.3. Kinase substrate specificity and ATPase activity

Among the 25 substrates tested, the enzyme was only able to phosphorylate glucose using ATP as the phosphate donor (Figure 11), forming a product identified as glucose 6-phosphate by 2D-NMR (see section 3.4).

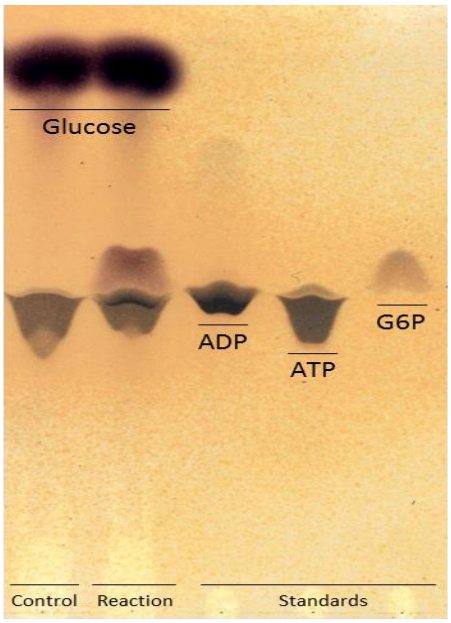

Figure 11. TLC analysis of the glucokinase reaction catalyzed by the pure recombinant protein, using an ethanol/water (7:3, v/v) solvent system. From left to right: control mixture without enzyme, reaction mixture with enzyme, ADP standard, ATP standard, G6P standard. Reaction and control were both incubated for 2h.

Other phosphate donors tested besides ATP were not utilized by the enzyme. Very low trace activity was detected with other two hexoses, allose and mannose, after overnight incubation at 37°C (Figure 12). Despite being annotated as a maltokinase, the enzyme was unable to phosphorylate maltose even after an overnight incubation in 50 mM BTP buffer pH 8.0 with 10 mM $MgCl_2$ at 37°C (Figure 12). It is thus better described as a glucokinase (GluK).

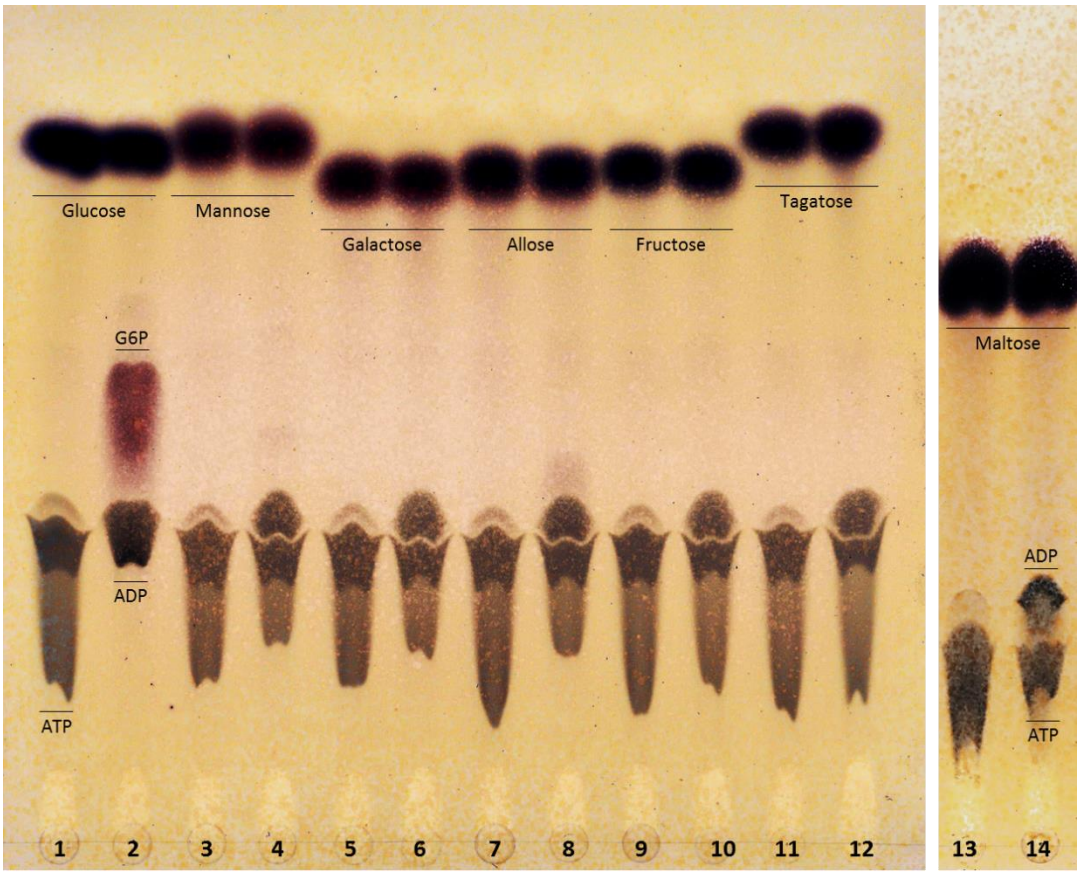

Figure 12. TLC analysis of GluK's substrate specificity using different sugars as potential phosphate acceptors and ATP as donor, using an ethanol/water (7:3, v/v) solvent system. **Odd lanes (1 to 13)** – control mixtures without enzyme. **Even lanes (2 to 14)** – reactions with recombinant enzyme. All reactions and controls were incubated overnight. In this solvent, spots corresponding to phosphorylated sugars, both mono and disaccharides, appear purple above ADP.

ATP consumption with resulting ADP formation was observed with all the substrates tested in the presence of the enzyme, but not in negative controls. However, except in the case of glucose and to a much lesser extent mannose and allose, this did not correspond to the formation of any phosphorylated sugar as visualized by TLC (Figure 12). The existence of an ATPase activity independent of the substrate utilized was later confirmed by incubating the enzyme with ATP only, in the absence of any sugar acceptor (Figure 13). A similar result was not observed when the maltokinase from *M. vanbaalenii* was likewise incubated with ATP overnight, confirming that Mak does not have ATPase activity (data not shown). It was also investigated if GluK had any phosphatase activity towards glucose 6-phosphate, and the results were negative (data not shown).

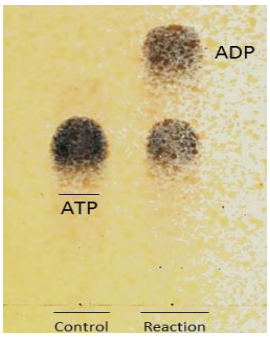

Figure 13. TLC analysis of the ATPase activity catalyzed by GluK in the presence of ATP and no sugar acceptor, using the acetic acid/ethyl acetate/water/ammonia 25% (6:6:2:1, vol/vol) solvent system.

3.4. Identification of glucose 6-phosphate by NMR

The phosphorylated product obtained after incubation of the enzyme with glucose and ATP was identified as glucose 6-phosphate by two-dimensional ^{1}H-^{13}C HSQC NMR spectroscopy (Figure 14).

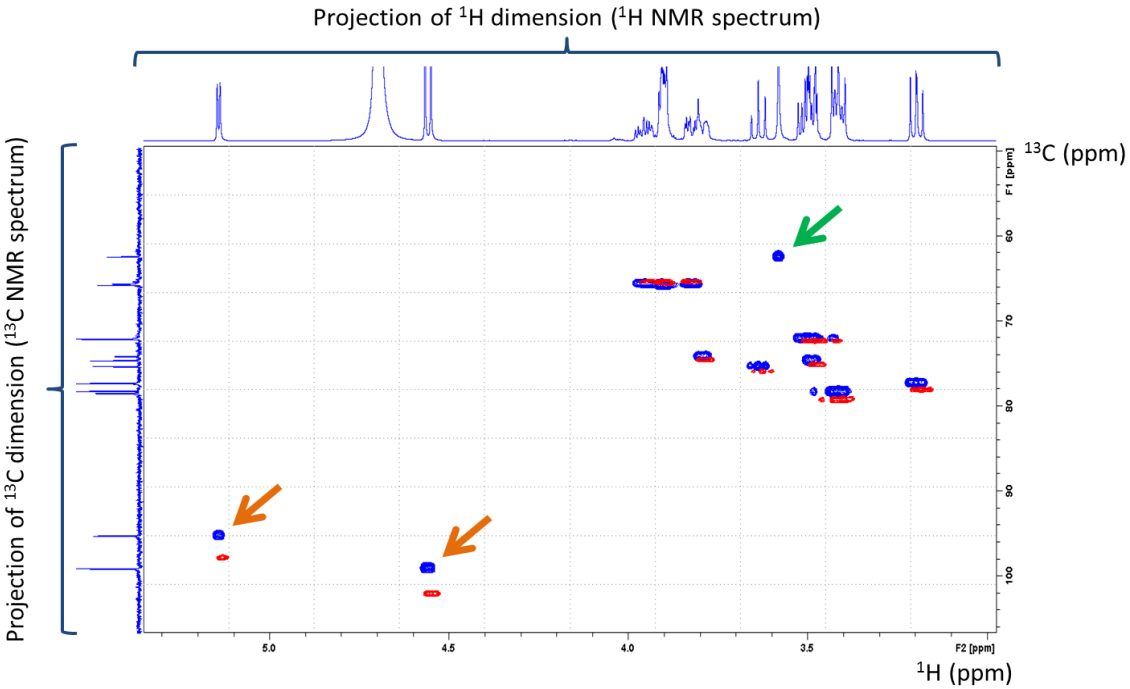

Figure 14. Comparison of the 500 MHz ^{1}H-^{13}C HSQC NMR spectrum of the glucose-phosphate compound produced by GluK (blue resonances) and of a glucose 6-phosphate standard spectrum retrieved from the Biological Magnetic Resonance Bank database (red resonances). Green arrow indicates an unknown resonance in the sample that is not observed in the standard, probably due to an unknown contaminant. Orange arrows indicate the two anomeric signals of H1-C1, corresponding to the alpha and beta conformations of the compound. A slight shift between sample and standard matching peaks may be attributed to small differences in pH and sample preparation between the two spectra.

3.5. Biochemical characterization

3.5.1. Activity modulation by divalent cations

Among the chloride or sulfate salts of cations tested, Co^{2+}, Cu^{2+} and Ca^{2+} produced no measurable activity. The same result was obtained in the absence of any divalent cation and in the presence of the chelating agent EDTA, demonstrating the absolute requirement of cations for enzymatic activity (Figure 15). On the other hand, Mg^{2+}, Mn^{2+} and Zn^{2+} were found to stimulate the release of ADP by the enzyme in the presence of ATP and glucose (Figure 15A). In relation to Mg^{2+}, the addition of 20 mM of cation to the reaction mixture showed the most pronounced stimulatory effect, activity was barely measurable below 5 mM and it took concentrations of above 100 mM for any enzyme inhibition to be evident. With Zn^{2+}, 5 mM was the most stimulatory concentration and 10 mM was already inhibitory. Due to interferences with the ADP detection spectrophotometric assay caused by manganese ion's color in solution, it was difficult to accurately measure activity and for this reason further concentrations were not tested for this cation.

However, TLC analysis of reactions carried out under selected conditions unequivocally demonstrated that both Zn^{2+} and Mn^{2+} were specifically stimulating the ATPase activity of the enzyme (and thus ADP release) and inhibiting the production of glucose 6-phosphate, while only the seemingly moderate activity measured with Mg^{2+} alone was in fact glucokinase activity (Figure 15B). The effect of different concentrations of Mg^{2+} (5-100 mM) on glucokinase activity was measured spectrophotometrically by following G6P release, revealing a profile similar to that observed for ADP release, with 20 mM as the most stimulatory concentration (Figure 15B).

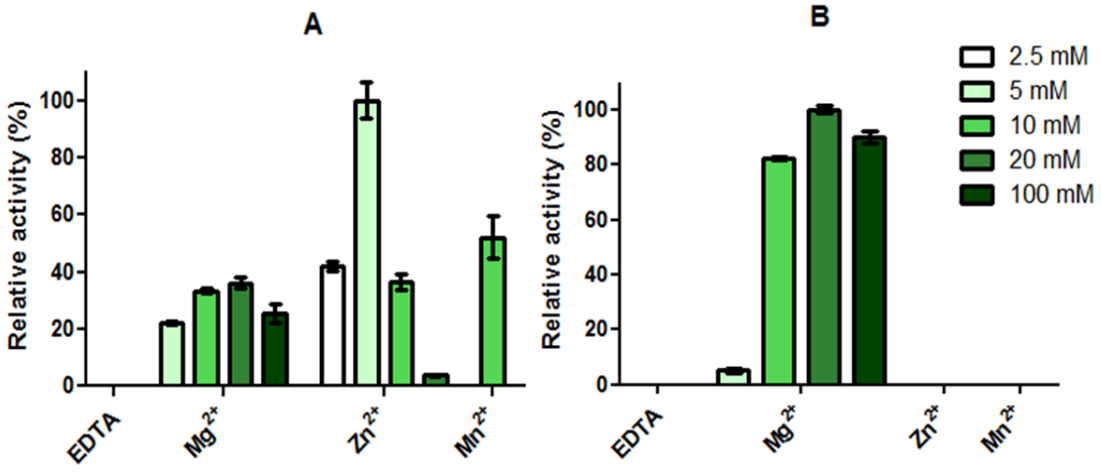

Figure 15. Enzyme activity measured in the presence of different concentrations of stimulatory cations. A – ADP release; B – G6P release.

Moreover, combinations of either Mg^{2+}+Zn^{2+} or Mn^{2+}+Zn^{2+} were found to maximally stimulate ADP release, surpassing the effect of each cation on its own (Figure 16). These combinations were also shown by TLC analysis to inhibit glucokinase activity. In the presence of 10 mM Mg^{2+}, addition of 5 mM Zn^{2+} to the reaction resulted in inhibition of the glucokinase activity as compared to the maximal activity obtained with Mg^{2+} alone, while favoring non-productive ATP hydrolysis.

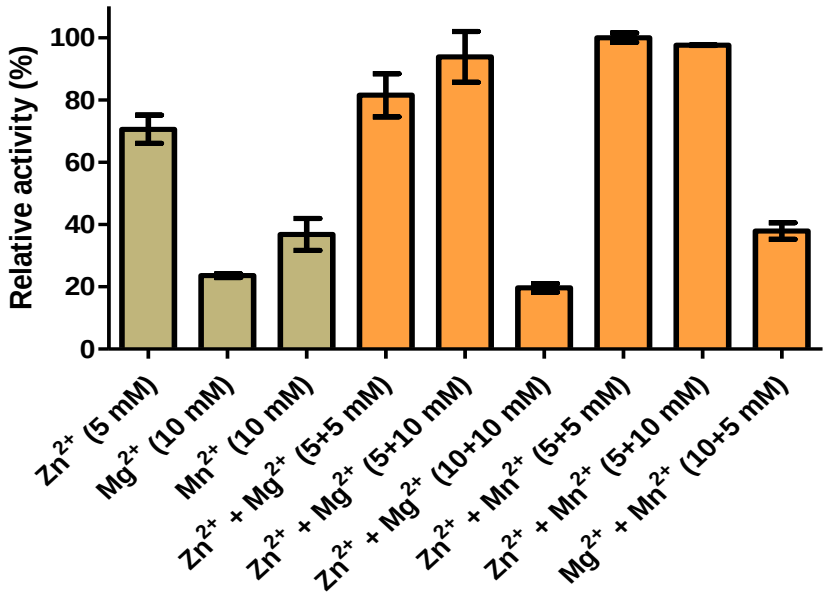

Figure 16. Effect of different combinations of the three stimulating cations (orange bars) on the ATPase activity of GluK and comparison with each cation on its own at its optimal concentration (light brown bars).

Thus, biochemical characterization and kinetic studies of the glucokinase activity were carried out in the presence of 10 mM Mg^{2+}, while for the ATPase activity a combination of 10 mM Mg^{2+} plus 5 mM Zn^{2+} was used.

3.5.2. Temperature and pH profiles of glucokinase activity

The glucokinase was maximally active between 37 and 40°C and progressively lost activity for lower and higher temperatures (Figure 17). At the lowest temperature tested, 25°C, the enzyme still retained more than 50% of maximum activity. Above 55°C, activity was barely detectable.

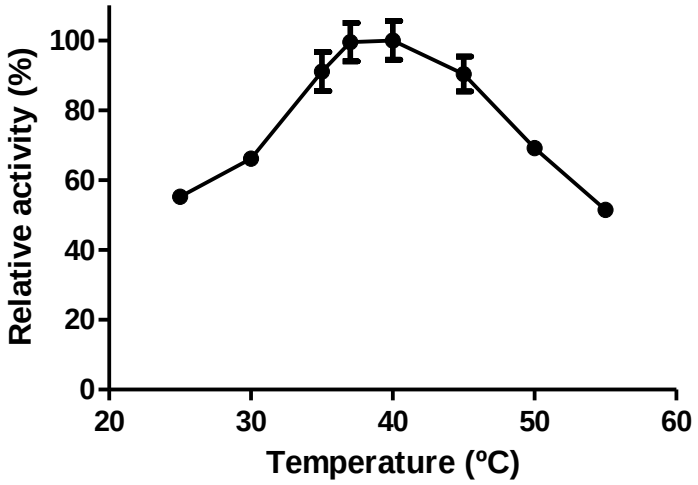

Figure 17. Temperature profile of the glucokinase activity of GluK at pH 8.0 in the presence of 10 mM Mg^{2+}.

The optimum pH of the recombinant enzyme for the glucokinase activity was 8.0 and more than 90% of maximal activity was retained up to pH 9.0 (Figure 18). A slight drop in activity was observed when transitioning from BTP buffer to CAPSO buffer at pH 9.5.

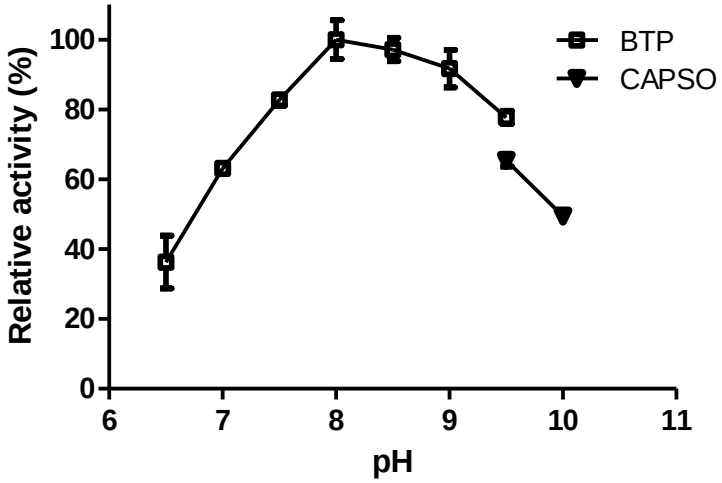

Figure 18. pH profile of the glucokinase activity of GluK at 37°C in the presence of 10 mM Mg^{2+}.

3.5.3. pH profile of ATPase activity

For the ATPase activity, maximal activity was observed between pH 7.5-8.5, with 8.5 as the optimum pH (Figure 19). Outside of this optimal range, enzyme activity dropped sharply and more markedly than what was observed for the glucokinase activity (Figure 18).

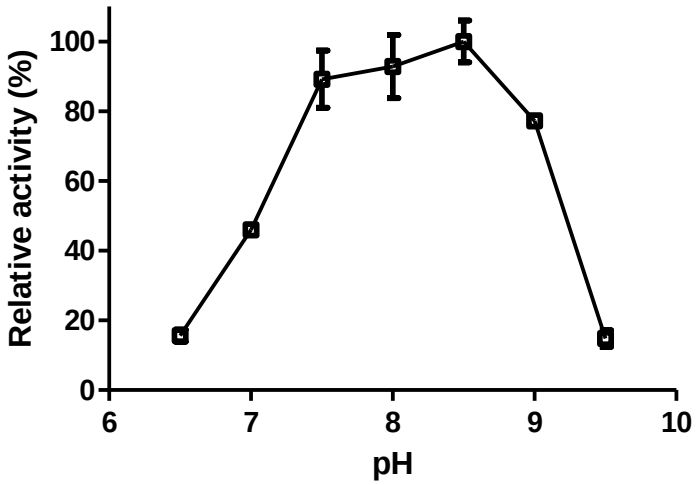

Figure 19. pH profile of the ATPase activity of GluK at 37°C in the presence of 10 mM Mg^{2+} and 5 mM Zn^{2+}.

3.6. Kinetic studies

The *M. smegmatis* recombinant GluK exhibited Michaelis-Menten kinetics at 37°C with ATP up to 5 mM (Figure 20) and with glucose up to 120 mM (Figure 21). The ATPase activity of the enzyme also followed Michaelis-Menten kinetics at 37°C with ATP up to 15 mM (Figure 22). Kinetic parameters are indicated in Table 2.

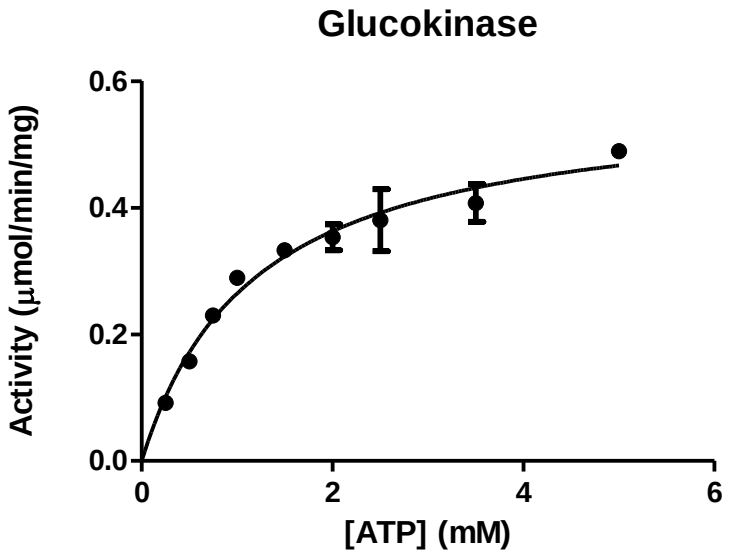

Figure 20. Effect of ATP concentration on the glucokinase activity of GluK at 37°C with glucose concentration held constant at 40 mM and in the presence of 10 mM Mg^{2+}.

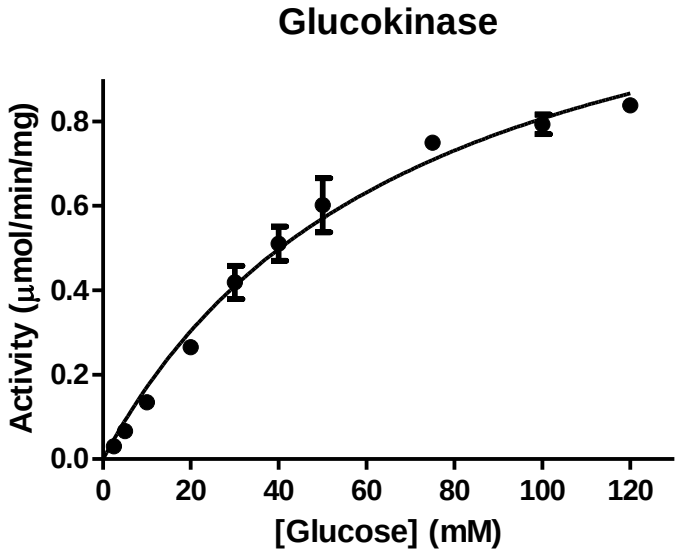

Figure 21. Effect of glucose concentration on the glucokinase activity of GluK at 37°C with ATP concentration held constant at 5 mM and in the presence of 10 mM Mg^{2+}.

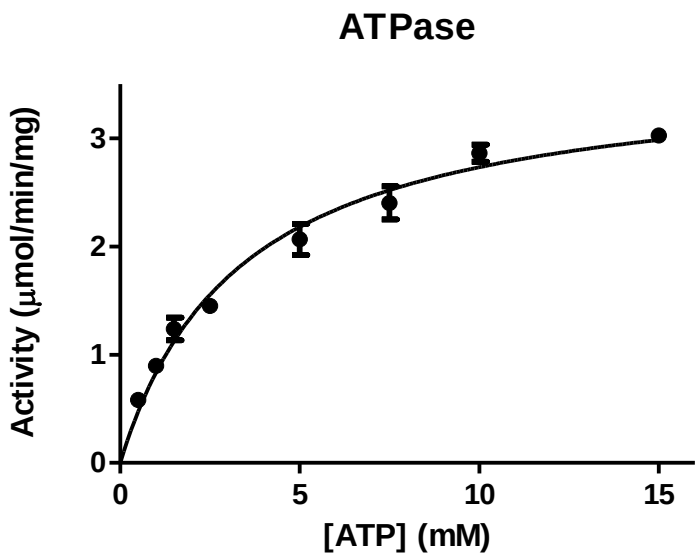

Figure 22. Effect of ATP concentration on the ATPase activity of GluK at 37°C and in the presence of 10 mM Mg^{2+} and 5 mM Zn^{2+}.

Table 2. Kinetic parameters of the recombinant GluK from *M. smegmatis*.

Activity	Substrate	K_m (mM)	V_{max} (μmol/min/mg)
Glucokinase	ATP	1.18±0.16	0.58±0.03
	Glucose	70.80±9.98	1.38±0.10
ATPase	ATP	3.38±0.40	3.66±0.15

CHAPTER 4 – DISCUSSION

The worldwide emergence of opportunistic infections caused by environmental mycobacteria is an increasing cause of concern, further aggravated by their intrinsic antibiotic resistance and by their remarkable adaptability and ability to persist in human-made environments (Falkinham 2009a; Prevots & Marras 2015). From causing severe pulmonary infections with symptoms reminiscent of TB but unresponsive to antituberculous treatment, to being the source of major healthcare-associated outbreaks of post-surgical infections, NTM have undoubtedly earned a reputation as serious opportunistic pathogens (Duarte et al. 2009; Glassroth 2008). The therapeutic challenges posed by NTM are as unparalleled as their ecological, physiological and metabolic traits, and will not be met without intensifying research into the molecular and biochemical foundations of their unique characteristics. Much of mycobacteria's adaptive success, resilience and uniqueness can be attributed to their remarkable cell envelope, a thick, hydrophobic and impermeable barrier that not only shields the cell from outside stressors but also mediates its interaction with the surrounding environment, be it an inanimate surface or an infected host (Falkinham 2009b; Nobre et al. 2014). NTM also display great metabolic versatility and plasticity, which likewise contributes to their ability to succeed and thrive in diverse and often harsh environments (Baloni et al. 2014).

Understanding what makes mycobacteria so successful and unique at the molecular level is an important but challenging task. The growing availability of complete and annotated genome sequences for several mycobacterial species, as well as the development of genetic manipulation tools for these organisms, have proved to be invaluable tools in mycobacterial research and have helped to identify some of the genetic and molecular bases of many of their unique features (Kaur et al. 2009; Mendes et al. 2012). However, interpretation of the large amounts of data generated from whole-genome sequencing is inevitably limited by the fact that many mycobacterial genes have no functional prediction (i.e., no annotation) and, even among those which do, it has been observed that annotation of mycobacterial genes based on sequence similarities alone is often unreliable (Chopra et al. 2014; Mendes et al. 2011). In many cases, experimental investigation is the only way to validly assign

biochemical functions to predicted genes and to obtain novel insights and clues into yet unexplored aspects of mycobacterial physiology.

The recently discovered TreS-Mak-GlgE pathway, crucial in mycobacteria and validated as a promising target for drug development (Elbein et al. 2010; Kalscheuer et al. 2010), also illustrates this point. Based on sequence similarities with α-amylases and genomic context, GlgE was initially annotated as a glucanase and believed to be involved in glycogen degradation (Belanger & Hatfull 1999), until experimental data revealed it to be a maltosyltransferase which elongates α-glucans using maltose 1-phosphate as a substrate (Kalscheuer et al. 2010). Maltokinase orthologues are often found annotated on genome databases as aminoglycoside phosphotransferases based on sequence homology, and in some cases as trehalose synthases (or as trehalose synthase-fused maltokinases). The latter assignment may look puzzling at first, but it's a result of automated transitive annotations from bifunctional TreS-Mak proteins found in some organisms where the two genes are fused together (Kannan et al. 2007). Because these fused genes are annotated as "trehalose synthase-fused maltokinase" or simply as "trehalose synthase" and automated algorithms find significant homology between *mak* genes and the C-terminal of *treS-mak* genes, the first automatically receive the same annotation as the latter, despite coding for independent maltokinases.

Mak is widespread across the bacterial domain and is invariably found lying next to or fused with TreS, reflecting the close association between these two key enzymes of the TreS-Mak-GlgE pathway (Fraga et al. 2015; Mendes et al. 2010). This association is further supported by the finding that TreS and Mak interact with each other forming an enzymatic complex which boosts Mak activity (Roy et al. 2013). Genomic association of genes involved in the same pathway is indeed a common occurrence in bacterial genomes and genetic context is often used as a guide to predict gene function and involvement in a given pathway (Ma & Xu 2013; Mavromatis et al. 2009).

We identified a gene annotated as a maltokinase in the genome of *M. smegmatis*, sharing 26% amino acid identity with the *M. smegmatis* canonical and characterized maltokinase and in a radically different genetic context, with no *treS*-like gene next to it. Unexpectedly, this gene was now found to code for a highly specific

glucokinase (GluK), with no traces of maltokinase activity. It should be noted that sequence conservation among maltokinases is known to be very low, dropping as low as 10% among distantly related taxa and below 60% within the same genus (Fraga et al. 2015), so an amino acid identity of 26% was not in itself taken as a strong hint that this enzyme had a different biochemical function from Mak. The genetic context, on the other hand, was highly suggestive of a different biological role for this enzyme and for its involvement in a different pathway, a suspicion strengthened by the experimental results reported here.

Mak and this newly identified GluK show some sequence similarity with aminoglycoside phosphotransferases (APH), antibiotic resistance bacterial enzymes responsible for inactivating aminoglycoside antibiotics through phosphorylation (Hon et al. 1997). For this reason, the aminoglycoside antibiotics kanamycin, streptomycin, gentamycin and hygromycin B were tested as possible substrates for this newly discovered enzyme, but none were phosphorylated by it. Several disaccharides (including maltose) and trisaccharides were tested and were also not phosphorylated. Thus, the substrate specificity of this enzyme is markedly different from its two closest relatives, Mak and APH. Interestingly, this is in accordance with the observation that the most significant differences in terms of residue conservation between Mak and GluK occur in the residues that, in Mak's three-dimensional structure, line the probable maltose binding pocket. While no definitive conclusions can be drawn before determining GluK's structure, these conserved residue replacements could explain the different substrate specificity observed experimentally.

Glucose can be phosphorylated by hexokinases or by glucokinases, with the distinction being mainly made based on the substrate specificity of the enzyme (Mendz & Hazell 1993). Given that of the six hexoses tested glucose was the only to undergo appreciable phosphorylation under the experimental conditions employed, the enzyme appears to be a *bona fide* glucokinase. This renders it the second glucokinase to be experimentally identified in *M. smegmatis*, since Pimentel-Schmitt *et al* cloned and identified the Group III glucokinase (see section 1.4.1) $GlkA^{Msm}$, homologous to the well characterized GlkA from *Streptomyces coelicolor* known to be involved in central carbon metabolism in this related actinomycete (Pimentel-Schmitt et al. 2007). It is

also an atypical and novel glucokinase in the sense that it does not fit within any of the three groups of microbial glucokinases recognized to date (Romero-Rodríguez et al. 2015).

Unlike the mycobacterial maltokinases characterized to date, which can use GTP or UTP as a phosphate donors in replacement of ATP (Fraga et al. 2015; Mendes et al. 2010), GluK was only able to use ATP. This is similar to what was reported for the maltokinases of *Actinoplanes missouriensis* and *S. coelicolor*, which also use ATP as their sole phosphate donor (Jarling et al. 2004). The best characterized mycobacterial glucokinase, *M. tuberculosis* PolyP/ATP-dependent glucokinase (which also has a homologue in *M smegmatis* distinct from both the GluK reported here and from the above-mentioned GlkA), is able to use other nucleoside triphosphates such as GTP and UTP in addition to ATP and PolyP (Hsieh et al. 1993).

The recombinant GluK was maximally active within the 8-9 pH range, consistent with what is observed for most kinases, including Mak (Niehues et al. 2003). With the exception of *S. coelicolor* Mak, all maltokinases characterized to date have an optimal temperature of 60°C (Fraga et al. 2015; Jarling et al. 2004; Mendes et al. 2010). GluK, however, displays maximal activity in the 35-45°C range and has no measurable activity at temperatures above 55°C, resembling the temperature profile of *S. coelicolor* Mak. Although no pH or temperature profiles have been reported for $GlkA^{Msm}$, its *S. coelicolor* homologue exhibits maximal activity at 42°C and at pH 7.6 (Imriskova et al. 2005).

Kinases and other ATP-utilizing enzymes require divalent cations for activity (most commonly Mg^{2+}) which form complexes with the negatively charged phosphate donor, activating the attacked phosphoryl group by charge neutralization, electron withdrawal, and conformational adjustments (Matte et al. 1998; Mildvan 1987). Accordingly, this GluK strictly requires Mg^{2+} for glucokinase activity and unlike the case for *M. bovis* BCG Mak other cations could not partially replace Mg^{2+}. The divalent cations Mn^{2+} and Zn^{2+} had an unexpected effect on the activity of GluK. In the presence of Mg^{2+}, GluK was found to display a residual ATPase activity in which some ATP was hydrolyzed to ADP without the phosphoryl group being transferred to glucose. This activity was always present to some extent throughout the characterization of the

enzyme, as none of the conditions tested resulted in complete repression of GluK's residual ATPase activity while simultaneously preserving its glucokinase activity. In the presence of Zn^{2+} or Mn^{2+}, however, ATPase activity was markedly stimulated and became predominant, relegating the enzyme's glucokinase activity to residual and barely detectable levels. Appropriate negative controls ruled out the possibility of the observed hydrolysis being due to nonenzymatic metal catalysis. In the absence of glucose and presence of Mg^{2+} plus Zn^{2+}, GluK effectively behaved as an efficient ATPase.

No such behavior has been reported for any of the maltokinases characterized to date and we experimentally confirmed that *M. vanbaalenii* Mak has no ATPase activity in the conditions tested (results not shown). However, kinases with appreciable ATPase activities are not unprecedented even among Mak and GluK structural homologues. Indeed, significant ATP hydrolysis has been observed with aminoglycoside 3'-phosphotransferases, representing an ATPase activity which competes directly with phosphate transfer to aminoglycoside antibiotics in these enzymes (Kim et al. 2006). Residual ATPase activities, resulting from phosphate transfer to a water molecule rather than to the enzyme's substrate, have also been reported and described in other kinases including the eukaryotic phosphorylase kinase (a protein kinase which also belongs to the same fold group as Mak), yeast hexokinase, and *E. coli* glycerol kinase (Cleland & Hengge 2006; Kaji & Colowick 1965; Paudel & Carlson 1991; Pettigrew et al. 1990). Moreover, there is evidence suggesting that 5-methylthioribose kinase (MTRK), another ELK closely related to Mak and to GluK, is at least mechanistically capable of hydrolyzing ATP in the absence methylthioribose (Ku et al. 2007). This evidence, however, comes from structural studies that give insight into MTRK's catalytic mechanism and not from activity studies, and they also suggest that the inorganic phosphate likely stays bound to the enzyme, presumably until an accepting methylthioribose molecule binds (Ku et al. 2007). In this case no ATPase activity would be measured because even though ATP is hydrolyzed, in the absence of methylthioribose the products of its hydrolysis are not released from the enzyme's catalytic chamber and there is no ATP turnover. Ku and colleagues further propose that retaining ADP+P_i in place works as a strategy to minimize the abortive ATPase

reaction observed in other kinases. Assuming that MTRK and GluK share a similar catalytic mechanism, it appears that GluK would differ in its inability to minimize said abortive reaction, but further conclusions cannot be drawn at this point given that GluK's three-dimensional structure has not been solved yet.

The best characterized ATPase activity of a kinase is by far that of the yeast hexokinase (HK). Its weak ATPase activity is markedly increased in the presence of the non-phosphorylable substrate analogues xylose and lyxose, which bind HK and induce a conformational change that facilitates ATP hydrolysis (DelaFuente 1970; DelaFuente et al. 1970). The pH profiles of HK's hexokinase and ATPase activities have also been investigated and compared with each other (Kaji & Colowick 1965). The observed trend closely resembles the results obtained for GluK in that the pH optimum for the ATPase activity is slightly more basic and the drop in activity outside the optimal pH range is sharper than for the glucokinase activity. The effect of zinc and manganese ions was also investigated for HK, but unlike what is observed with GluK, these two cations were shown to inhibit both the ATPase and hexokinase activities of HK (Kaji & Colowick 1965). Indeed, stimulation of ATPase activity by zinc ions has not to our knowledge been reported before for a kinase, but the effect is reminiscent of what is observed for pyrophosphatases (PPases). The physiological cofactor of PPases is Mg^{2+}, and in its presence they display nearly absolute substrate specificity towards the hydrolysis of pyrophosphate, but in the presence of Zn^{2+} or Mn^{2+} this specificity is lost and they exhibit appreciable ATPase activity (Zyryanov et al. 2002).

The physiological relevance, if any, of the Zn^{2+}-induced shift in GluK's activity is not readily apparent. Such high levels of futile ATP hydrolysis would have an obvious deleterious effect on the cell with no apparent benefit resulting from it. While it will be interesting to further investigate the mechanism behind zinc's effect on GluK, particularly if it acts at the structural level by inducing a conformational change or if it directly participates in catalysis, this is most likely a nonphysiological event, as is zinc's *in vitro* effect on PPases. Zinc homeostasis is stringently controlled in bacteria, including mycobacteria, because even though many enzymes use it as an essential catalytic or structural cofactor, zinc is also highly toxic (Riccardi et al. 2008). Thus, under physiological conditions, free cytoplasmic zinc is maintained at picomolar levels,

much lower than the milimolar concentrations employed in GluK's characterization (Cerasi et al. 2013).

But even without Zn^{2+}, when Mg^{2+} is the sole cofactor and glucose is present in the reaction mixture, the ATPase activity of GluK is far from being negligible and directly competes with its glucokinase activity. This is akin to what has been reported for APHs and would logically seem to imply a significant energetic burden on the cells expressing this enzyme. The costs and consequences of expressing an enzyme with such characteristics have been studied *in vivo* for APHs (Kim et al. 2006). Bacteria constitutively expressing APH were shown to absorb the cost of increased ATP turnover remarkably well, with no visible impairments at the phenotypic level and no reduced performance in competition experiments. There was, however, a strong selective pressure for loss of the plasmid carrying the gene in the absence of aminoglycoside antibiotics, 5-fold greater than observed for a different antibiotic/resistance gene pair (Kim et al. 2006). It is not known at this point whether GluK's expression is constitutive or regulated, but these considerations seem to argue for a tightly regulated expression that would minimize futile ATP consumption and restrict it to whenever it is advantageous to activate the pathway GluK putatively participates in. The presence of a TetR-family transcriptional regulator in GluK's gene cluster further supports this assumption (Cuthbertson & Nodwell 2013). The results with APH also demonstrate that the costs of harboring an enzyme with significant ATPase activity are well tolerated by bacterial cells provided that the enzyme confers them a selective advantage of other sort.

Another puzzling result came from the kinetic studies with GluK, which revealed an unexpectedly high K_m for glucose (≈70 mM), suggesting that the recombinant enzyme is binding this sugar with a very low affinity under the experimental conditions employed. For comparison, *M. vanbaalenii* and *M. bovis* maltokinases show a K_m for maltose of 7 and 2.5 mM, respectively, and the other *M. smegmatis* characterized glucokinase, $GlkA^{Msm}$, has a K_m of 9 mM for glucose (Fraga et al. 2015; Pimentel-Schmitt et al. 2007). Much lower values have been reported for most bacterial glucokinases, often below 1 mM (Romero-Rodríguez et al. 2015). Although we tested a battery of 25 possible substrates including common glucose isomers and glucose

derivatives, we cannot at this point completely rule out the possibility that glucose isn't the physiological substrate of this kinase, which would explain the high K_m towards this sugar. Another possibility is that the experimental conditions employed are less than suboptimal and that these *in vitro* kinetic parameters of a recombinant form of GluK do not truly reflect or correlate with the real efficiency of the native enzyme in the crowded and rich environment within the cell (Zimmerman & Trach 1991). For example, a missing cofactor can be enough to radically alter the affinity of an enzyme for its substrate, and the participation of an enzyme in a multicatalytic complex or just its interaction with other macromolecules can significantly affect its kinetic parameters (Price et al. 2004; Sanli & Blaber 2001; Zhang 2011). Building on this latter observation and considering that Mak does interact with at least one enzyme, TreS, through its N-terminal domain and has its catalytic efficiency significantly altered by that interaction (Roy et al. 2013), it is tempting to speculate a similar situation taking place with GluK and its putative pathway partners. Indeed, the N-terminal of GluK is different from that of Mak and slightly shorter, but it has some unique conserved motifs that could theoretically be involved in mediating such hypothetical interaction. However, only the structural characterization of GluK as well as the cloning and characterization of its neighboring genes will generate evidence for or against this hypothesis.

Similarly, characterizing the enzymes encoded by the other genes in MSMEI_5186 operon will be crucial to elucidate the biological role of GluK and of the putative novel pathway it participates in, something that at this point can only be speculated. Analyzing the genomic context can give some clues, but conclusions are necessarily limited by the problems inherent to sequence-based functional annotations as already discussed. The adjacent MSMEI_5185 is annotated as either a glutamine-fructose-6-phosphate transaminase or as a sigma factor regulator, but a closer analysis of its sequence shows that it has neither the glutaminase domain needed for the transaminase reaction nor the DNA binding domain of sigma factor regulators. It does have a N-terminal phosphosugar-binding domain (SIS domain) with sequence homology to one of the SIS domains found in glutamine-fructose-6-phosphate transaminase and which is common to many other phosphosugar

isomerases (Bateman 1999), but the C-terminal of this hypothetical protein has no homology with any domain deposited in NCBI's conserved domain database. It likely binds a phosphosugar, maybe GluK's product, and probably isomerizes it, but experimental investigation will be required to identify its actual biochemical function. The contiguous MSMEI_5187 is annotated as a glycosyltransferase of the GT1 family with a conserved GT1_Gtf_like domain sequence, being closely related to glycosyltransferases involved in the production of glycopeptide and macrolide antibiotics (Bolam et al. 2007; Mulichak et al. 2003) and in the synthesis of plant secondary metabolites and plant xenobiotic metabolism (Brazier-Hicks & Edwards 2005; Shao et al. 2005). It is interesting to note that the two additional genes found in the gene cluster of *Streptacidiphilus*, encoding a putative phosphoglucomutase-like protein and a putative UTP-glucose-1-phosphate uridylyltransferase, would allow the G6P produced by GluK to be converted to G1P and nucleotide-activated to become a substrate for the glycosyltransferase. In this hypothetical pathway, the role of the putative phosphosugar isomerase is not clear, but it could theoretically act on G6P and convert it into a different phosphorylated hexose that could well be actual substrate of the uncharacterized phosphohexomutase. The TetR-family protein (MSMEI_5188) is probably involved in gene regulation and its orientation and proximity to the other genes in the cluster suggest that it is cotranscribed with them (Cuthbertson & Nodwell 2013). Because most TetR-family regulators are known to be autoregulatory, this transcriptional regulator may be predicted to control the expression of its cotranscribed genes, but it may also be postulated to act on other regions of the genome, possibly recruiting other enzymes for this putative pathway (Cuthbertson & Nodwell 2013).

The genomic context and the experimentally observed activity of GluK suggest the involvement of this gene cluster in a glucose utilization biosynthetic pathway, but trying to pinpoint its biological role proves even more difficult and speculative than the above attempt at conjecturing a possible step-by-step pathway. The fact that this cluster is exclusively found in only a few strains of soil-dwelling actinomycetes seems to indicate a nonessential function, such as secondary metabolite biosynthesis. Moreover, the close relationship between GluK's adjacent glycosyltransferase and

some glycosyltransferases used by *Streptomyces* for antibiotic production strengthens this suspicion and raises the intriguing possibility that this cluster could take part in the synthesis and transfer of an unusual sugar unit to a secondary metabolite such as a glycopeptide antibiotic or a macrolide. In fact, glucose 6-phosphate is the precursor of most sugar units found in macrolide antibiotics as well as in other secondary metabolites (Liu & Rosazza 1998; Piepersberg & Distler 1997). But on the other hand, dedicated glucokinases lying within antibiotic biosynthetic gene clusters have not been reported and typical clusters are usually much larger and include a wide variety of specialized genes not found in the vicinity of this cluster, although homologues of some of these specialized genes (e.g., a putative macrocin *O*-methyltransferase implicated in tylosin synthesis) could be identified in another region of the genome of *M. smegmatis* (Karray et al. 2007; Li et al. 2004; Mast et al. 2011; Zotchev et al. 2000).

The real function of this rare gene cluster and of its atypical glucokinase remains elusive, but these findings still leave open the neglected but intriguing question of whether NTM are able to synthesize antibiotics. This question has been raised before but there is a tremendous lack of research into it, and it is still unclear if the absence of mycobacteria from the catalogues of antibiotic-producing microorganisms is due to their inability to synthesize these compounds or to the fact that they are slow growers and poor experimental organisms that require genus-specific isolation techniques and usually get overlooked during the screening of environmental samples for antibiotic activities (Falkinham 2009b). In principle, antibiotic production could give these environmental mycobacteria a competitive advantage in mixed microbial populations, especially important given that their competitors have higher growth rates (Falkinham 2009b). Antibiotic production also seems to be more important for biofilm-forming bacteria than for free-living bacteria, and NTM belong to the former group (Falkinham 2009b; Long & Azam 2001). Moreover, probing the genomes of some environmental mycobacteria reveals that they contain the potential for antibiotic synthesis. In spite of this, a literature search yielded only three old reports of possible antibiotic production by mycobacteria (Kozlovskaia et al. 1998; Obgol'tseva et al. 1991; Uesaka et al. 1954).

The alarming rise of multidrug resistance among important human pathogens coupled with the steep decline in the discovery and development of new antibiotics

constitutes a true health emergency and turns all avenues of research into this area worth pursuing, including those that consist of mining underexplored potential sources of novel compounds (Fischbach & Walsh 2010). Soil-dwelling bacteria, particularly actinomycetes, have long been a leading source of antibiotics and other medically or industrially useful secondary metabolites, but the focus has naturally been placed on readily cultured bacteria and their easily induced secondary metabolism pathways (Milshteyn et al. 2014). As this pool becomes depleted, though, there is a need to turn to other less obvious sources, including difficult-to-culture bacteria and "silent" biosynthetic gene clusters which are not readily expressed under standard laboratory conditions (Milshteyn et al. 2014).

The molecular and biochemical characterization of this rare and atypical glucokinase is the first step in trying to understand the biological role of an intriguing biosynthetic gene cluster that possesses all the hallmarks of secondary metabolism but whose real function is still obscure. Many interesting questions have been raised by the results presented in this work. What are the molecular and structural bases for the different substrate specificity between GluK and Mak? Does GluK interact with another enzyme through its N-terminal domain in a manner analogous to Mak with TreS? What is the biochemical function of the neighboring SIS-domain containing protein? What is the substrate specificity of the glycosyltransferase? Are there specific conditions or stimuli that trigger the expression of these genes? Are there other operons elsewhere in the genome of *M. smegmatis* coding for enzymes that will complete a putative novel biosynthetic pathway in which these genes may participate? Why is this cluster so rare and scattered across the microbial world and what is the significance of the two additional neighboring genes that are found in *Streptacidiphilus* but not in *Mycobacterium*? Is this putative pathway fully functional in all microorganisms displaying this rare glucokinase? Does it confer the bacteria bearing it any competitive advantage?

Further work will be required to begin answering these questions and to assign a function to this cryptic gene cluster, which will undoubtedly contribute to improve our knowledge of the ecology and physiology of nontuberculous mycobacteria and related actinomycetes.

CHAPTER 5 - CONCLUSIONS

Over the past two decades the number of described species of environmental mycobacteria has grown significantly and so has the awareness of their importance as opportunistic pathogens. The rising availability of whole-genome sequences for many of these species has been crucial for improving our knowledge of both NTM and their strictly pathogenic counterparts. Sequence fidelity as well as gene prediction and annotation accuracy are critical for meaningful interpretation of these large amounts of data, but remain a source of occasional errors that can only be truly dissipated through functional characterization studies.

A gene missed during the first annotation of *M. smegmatis* mc^2155 genome and later misannotated as a maltokinase was experimentally shown to be a *bona fide* glucokinase with a cryptic metabolic function and an extremely scattered distribution among soil-dwelling actinomycetes. It represents a rare and novel isofunctional glucokinase that does not fit within the traditional classification of microbial glucokinases and displays high homology with eukaryotic-like kinases.

Genomic context suggests a role in a glucose utilization metabolic pathway and GluK's inclusion in a rare but conserved gene cluster present in only a few strains of actinomycetes from different families evokes a probably nonessential but advantageous function in secondary metabolism that for now remains elusive.

The biochemical properties reported here for this rare GluK represent a preliminary step towards understanding its importance and physiological function in the few actinomycetes bearing it. GluK possesses some peculiar characteristics that require further investigation, namely its zinc-induced activity shift and seemingly low *in vitro* catalytic efficiency.

The work presented in this thesis paves the way for the elucidation of the biological role of a rare and intriguing gene cluster whose study may shed light on yet unexplored aspects of mycobacterial ecology, physiology and metabolism that may drive important biomedical applications.

REFERENCES

Allen, K. N., & Dunaway-Mariano, D. (2004). Phosphoryl group transfer: evolution of a catalytic scaffold. *Trends in Biochemical Sciences*, *29*(9), 495–503.

Angala, S. K., Belardinelli, J. M., Huc-Claustre, E., Wheat, W. H., & Jackson, M. (2014). The cell envelope glycoconjugates of *Mycobacterium tuberculosis*. *Critical Reviews in Biochemistry and Molecular Biology*, *49*(5), 361–399.

Arend, S. M., van Soolingen, D., & Ottenhoff, T. H. M. (2009). Diagnosis and treatment of lung infection with nontuberculous mycobacteria. *Current Opinion in Pulmonary Medicine*, *15*(3), 201–208.

Baloni, P., Padiadpu, J., Singh, A., Gupta, K. R., & Chandra, N. (2014). Identifying feasible metabolic routes in *Mycobacterium smegmatis* and possible alterations under diverse nutrient conditions. *BMC Microbiology*, *14*, 276.

Bateman, A. (1999). The SIS domain: a phosphosugar-binding domain. *Trends Biochem Sci*, *24*(3), 94–95.

Belanger, A., & Hatfull, G. (1999). Exponential-phase glycogen recycling is essential for growth of *Mycobacterium smegmatis*. *Journal of Bacteriology*, *181*(21), 6670–6678.

Beste, D. J. V, Espasa, M., Bonde, B., Kierzek, A. M., Stewart, G. R., & McFadden, J. (2009). The genetic requirements for fast and slow growth in mycobacteria. *PLoS ONE*, *4*(4), e5349.

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

Bodle, E. E., Cunningham, J. A., Della-Latta, P., Schluger, N. W., & Saiman, L. (2008). Epidemiology of nontuberculous mycobacteria in patients without HIV infection, New York City. *Emerging Infectious Diseases*, *14*(3), 390–396.

Bolam, D. N., Roberts, S., Proctor, M. R., Turkenburg, J. P., Dodson, E. J., Martinez-Fleites, C., … Gilbert, H. J. (2007). The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(13), 5336–5341.

Bond, C., & Schüttelkopf, A. (2009). ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr D Biol Crystallogr*, *65*(Pt 5), 510–512.

Bork, P., Sander, C., & Valencia, A. (1993). Covergent evolution of similar enzymatic function on diffeerent protein folds: the hexokianse, ribokinase and galactokinase families of sugar kinases. *Protein Science*, *2*(1), 31–40.

Brazier-Hicks, M., & Edwards, R. (2005). Functional importance of the family 1 glucosyltransferase UGT72B1 in the metabolism of xenobiotics in *Arabidopsis thaliana*. *Plant Journal*, *42*(4), 556–566.

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

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

Bryant, J. M., Grogono, D. M., Greaves, D., Foweraker, J., Roddick, I., Inns, T., … Floto, R. A. (2013). Whole-genome sequencing to identify transmission of *Mycobacterium abscessus* between patients with cystic fibrosis: A retrospective cohort study. *The Lancet*, *381*(9877), 1551–1560.

Cerasi, M., Ammendola, S., & Battistoni, A. (2013). Competition for zinc binding in the host-pathogen interaction. *Frontiers in Cellular and Infection Microbiology*, *3*, 108.

Cheek, S., Ginalski, K., Zhang, H., & Grishin, N. V. (2005). A comprehensive update of the sequence and structure classification of kinases. *BMC Structural Biology*, *5*, 6.

Cheek, S., Zhang, H., & Grishin, N. V. (2002). Sequence and Structure Classification of Kinases. *Journal of Molecular Biology*, *320*(4), 855–881.

Child, R., Miller, C. D., Liang, Y., Narasimham, G., Chatterton, J., Harrison, P., … Anderson, A. J. (2007). Polycyclic aromatic hydrocarbon-degrading *Mycobacterium* isolates: Their association with plant roots. *Applied Microbiology and Biotechnology*, *75*(3), 655–663.

Chopra, T., Hamelin, R., Armand, F., Chiappe, D., Moniatte, M., & McKinney, J. D. (2014). Quantitative mass spectrometry reveals plasticity of metabolic networks in *Mycobacterium smegmatis*. *Molecular & Cellular Proteomics*, *13*(11), 3014–28.

Cleland, W. W., & Hengge, A. C. (2006). Enzymatic mechanisms of phosphate and sulfate transfer. *Chemical Reviews*, *106*(8), 3252–3278.

Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., … Barrell, B. G. (2001). Massive gene decay in the leprosy bacillus. *Nature*, *409*(6823), 1007–1011.

Conejo, M. S., Thompson, S. M., & Miller, B. G. (2010). Evolutionary bases of carbohydrate recognition and substrate discrimination in the ROK protein family. *Journal of Molecular Evolution*, *70*(6), 545–556.

Conn, V., & Franco, C. (2004). Analysis of endophytic actinobacterial population in the roots of wheat (*Triticum aestivum* L.) by terminal restriction fragment length polymorphism and sequencing of 16S rRNA clones. *Applied and Environmental Microbiology*, *70*(3), 1787–1794.

Cuthbertson, L., & Nodwell, J. R. (2013). The TetR family of regulators. *Microbiology and Molecular Biology Reviews*, *77*(3), 440–475.

Dawson, D. J., & Sartory, D. P. (2000). Microbiological safety of water. *British Medical Bulletin*, *56*(1), 74–83.

De Groote, M. A., & Huitt, G. (2006). Infections due to rapidly growing mycobacteria. *Clinical Infectious Diseases*, *42*(12), 1756–1763.

De Groote, M. A., Pace, N. R., Fulton, K., & Falkinham, J. O. (2006). Relationships between *Mycobacterium* isolates from patients with pulmonary mycobacterial infection and potting soils. *Applied and Environmental Microbiology*, *72*(12), 7602–7606.

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

Decker, B. K., & Palmore, T. N. (2014). Hospital water and opportunities for infection prevention. *Current Infectious Disease Reports*, *16*(10), 432.

DelaFuente, G. (1970). Specific inactivation of yeast hexokinase induced by xylose in the presence of a phosphoryl donor substrate. *European Journal of Biochemistry*, *16*(2), 240–243.

DelaFuente, G., Lagunas, R., & Sols, A. (1970). Induced fit in yeast hexokinase. *European Journal of Biochemistry*, *16*(2), 226–233.

Dell, A., Galadari, A., Sastre, F., & Hitchen, P. (2010). Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *International Journal of Microbiology*, *2010*, 148178.

Deshayes, C., Perrodou, E., Gallien, S., Euphrasie, D., Schaeffer, C., Van-Dorsselaer, A., … Reyrat, J.-M. (2007). Interrupted coding sequences in *Mycobacterium smegmatis*: authentic mutations or sequencing errors? *Genome Biology*, *8*(2), R20.

Devulder, G., Montclos, M., Flandrois, J. P., Pérouse de Montclos, M., & Flandrois, J. P. (2005). A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model. *International Journal of Systematic and Evolutionary Microbiology*, *55*(Pt 1), 293–302.

Doig, K. D., Holt, K. E., Fyfe, J. A., Lavender, C. J., Eddyani, M., Portaels, F., … Stinear, T. P. (2012). On the origin of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. *BMC Genomics*, *13*, 258.

Doroghazi, J. R., & Metcalf, W. W. (2013). Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics*, *14*, 611.

Duarte, R. S., Lourenço, M. C. S., Fonseca, L. D. S., Leão, S. C., Amorim, E. D. L. T., Rocha, I. L. L., … Sampaio, J. L. M. (2009). Epidemic of postsurgical infections caused by *Mycobacterium massiliense*. *Journal of Clinical Microbiology*, *47*(7), 2149–55.

Elbein, A. D., Pan, Y. T., Pastuszak, I., & Carroll, D. (2003). New insights on trehalose: A multifunctional molecule. *Glycobiology*, *13*(4), 17R–27R.

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

Ernst, J. D. (2012). The immunological life cycle of tuberculosis. *Nature Reviews Immunology*, *12*(8), 581–591.

Falkinham, J. O. (2009a). Surrounded by mycobacteria: nontuberculous mycobacteria in the human environment. *Journal of Applied Microbiology*, *107*(2), 356–67.

Falkinham, J. O. (2009b). The biology of environmental mycobacteria. *Environmental Microbiology Reports*, *1*(6), 477–87.

Falkinham, J. O. (2010). Impact of Human Activities on the Ecology of Nontuberculous Mycobacteria. *Future Microbiology*, *5*(6), 951–60.

Falkinham, J. O. (2015). Environmental Sources of Nontuberculous Mycobacteria. *Clinics in Chest Medicine*, *36*(1), 35–41.

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

Fischbach, M. A., & Walsh, C. T. (2010). Antibiotics for Emerging Pathogens. *Science*, *325*(5944), 1089–1093.

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

Galagan, J. E. (2014). Genomic insights into tuberculosis. *Nature Reviews. Genetics*, *15*(5), 307–20.

Gallien, S., Perrodou, E., Carapito, C., Deshayes, C., Reyrat, J. M., Van Dorsselaer, A., … Lecompte, O. (2009). Ortho-proteogenomics: Multiple proteomes investigation through orthology and a new MS-based protocol. *Genome Research*, *19*(1), 128–135.

Gao, B., & Gupta, R. S. (2012). Phylogenetic Framework and Molecular Signatures for the Main Clades of the Phylum Actinobacteria. *Microbiology and Molecular Biology Reviews*, *76*(1), 66–112.

Gengenbacher, M., & Kaufmann, S. H. E. (2012). *Mycobacterium tuberculosis*: Success through dormancy. *FEMS Microbiology Reviews*, *36*(3), 514–532.

Ghai, R., Pašić, L., Fernández, A. B., Martin-Cuadrado, A.-B., Mizuno, C. M., McMahon, K. D., … Rodríguez-Valera, F. (2011). New Abundant Microbial Groups in Aquatic Hypersaline Environments. *Scientific Reports*, *1*, 135.

Glassroth, J. (2008). Pulmonary disease due to nontuberculous mycobacteria. *Chest*, *133*(1), 243–251.

Goodfellow, M., & Williams, S. T. (1983). Ecology of actinomycetes. *Annual Review of Microbiology*, *37*(41), 189–216.

Gordon, R. E., & Smith, M. M. (1953). Rapidly Growing, Acid Fast Bacteria I. Species' Descriptions of *Mycobacterium phlei* Lehman and Neumann and *Mycobacterium smegmatis* (Trevisan) Lehman and Neumann. *Journal of Bacteriology*, *66*(1), 41–48.

Grange, J. M., & Yates, M. D. (1986). Infections caused by opportunist mycobacteria: a review. *Journal of the Royal Society of Medicine*, *79*(4), 226–229.

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

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

Griffith, D. E. (2010). Nontuberculous mycobacterial lung disease. *Current Opinion in Infectious Diseases*, *23*, 185–190.

Griffith, D. E., Aksamit, T., Brown-Elliott, B. A., Catanzaro, A., Daley, C., Gordin, F., … Winthrop, K. (2007). An official ATS/IDSA statement: Diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *American Journal of Respiratory and Critical Care Medicine*, *175*(4), 367–416.

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

Hasan, M. R., Rahman, M., Jaques, S., Purwantini, E., & Daniels, L. (2010). Glucose 6-phosphate accumulation in mycobacteria: implications for a novel F420-dependent anti-oxidant defense system. *The Journal of Biological Chemistry*, *285*(25), 19135–44.

He, X. M., & Liu, H.-W. (2002). Formation of unusual sugars: mechanistic studies and biosynthetic applications. *Annual Review of Biochemistry*, *71*, 701–754.

Hon, W. C., McKay, G. a, Thompson, P. R., Sweet, R. M., Yang, D. S., Wright, G. D., & Berghuis, a M. (1997). Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases. *Cell*, *89*(6), 887–895.

Hsieh, P.-C., Shenoy, B. C., Jentoft, J. E., & Phillips, N. F. B. (1993). Purification of Polyphosphate and ATP Glucose Phosphotransferase from *Mycobacterium tuberculosis* H37Ra: Evidence that Poly(P) and ATP Glucokinase activities are catalyzed by the same enzyme. *Protein Expression and Purification*, *4*(1), 76-84.

Imriskova, I., Arreguín-Espinosa, R., Guzmán, S., Rodriguez-Sanoja, R., Langley, E., & Sanchez, S. (2005). Biochemical characterization of the glucose kinase from *Streptomyces coelicolor* compared to *Streptomyces peucetius* var. *caesius*. *Research in Microbiology*, *156*(3), 361–366.

Ingen, J. Van, Rahim, Z., Mulder, A., Boeree, M. J., Simeone, R., Brosch, R., & Soolingen, D. van. (2012). Characterization of *Mycobacterium orygis* as *M. tuberculosis* complex subspecies. *Emerging Infectious Diseases*, *18*(4), 653–655.

Ito, S., Fushinobu, S., Jeong, J. J., Yoshioka, I., Koga, S., Shoun, H., & Wakagi, T. (2003). Crystal structure of an ADP-dependent glucokinase from *Pyrococcus furiosus*: Implications for a sugar-induced conformational change in ADP-dependent kinase. *Journal of Molecular Biology*, *331*(4), 871–883.

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

Jarling, M., Cauvet, T., Grundmeier, M., Kuhnert, K., & 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), 360–373.

Jurtshuk, P. J. (1996). Bacterial Metabolism. In S. Baron (Ed.), *Medical Microbiology* (4th editio.). Galveston (TX): University of Texas Medical Branch at Galveston.

Kaji, A., & Colowick, S. (1965). Adenosine triphosphatase activity of yeast hexokinase and its relation to the mechanism of the hexokinase reaction. *Journal of Biological Chemistry*, *240*(11), 4454–4462.

Kalscheuer, R., Syson, K., Veeraraghavan, U., Weinrick, B., Biermann, K. E., Liu, Z., … Jacobs, W. R. (2010). Self-poisoning of *Mycobacterium tuberculosis* by targeting GlgE in an alpha-glucan pathway. *Nature Chemical Biology*, *6*(5), 376–84.

Kannan, N., Taylor, S. S., Zhai, Y., Venter, J. C., & Manning, G. (2007). Structural and functional diversity of the microbial kinome. *PLoS Biology*, *5*(3), 0467–0478.

Karray, F., Darbon, E., Oestreicher, N., Dominguez, H., Tuphile, K., Gagnat, J., … Pernodet, J.-L. (2007). Organization of the biosynthetic gene cluster for the macrolide antibiotic spiramycin in *Streptomyces ambofaciens*. *Microbiology*, *153*(Pt 12), 4111–22.

Kaur, D., Guerin, M. E., Škovierová, H., Brennan, P. J., & Jackson, M. (2009). Biogenesis of the Cell Wall and Other Glycoconjugates of *Mycobacterium tuberculosis*. *Advances in Applied Microbiology*, *69*, 23–78.

Kelley, L. a, Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling , prediction and analysis. *Nature Protocols*, *10*(6), 845–858.

Kendall, B., & Winthrop, K. (2013). Update on the epidemiology of pulmonary nontuberculous mycobacterial infections. *Semin Respir Crit Care Med*, *34*(1), 87–94.

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

Kim, S. Y., Sohn, H., Choi, G. E., Cho, S. N., Oh, T., Kim, H. J., … Shin, S. J. (2011). Conversion of *Mycobacterium smegmatis* to a pathogenic phenotype via passage of epithelial cells during macrophage infection. *Medical Microbiology and Immunology*, *200*(3), 177–191.

Kirschner, R. A., Parker, B. C., & Falkinham, J. O. (1999). Humic and fulvic acids stimulate the growth of *Mycobacterium avium*. *FEMS Microbiology Ecology*, *30*(4), 327–332.

Koh, W., Stout, J. E., & Yew, W. (2014). Advances in the management of pulmonary disease due to *Mycobacterium abscessus* complex. *International Journal of Tuberculosis and Lung Disease*, *18*(10), 1141–1148.

Koskimäki, J. J., Hankala, E., Suorsa, M., Nylund, S., & Pirttilä, A. M. (2010). Mycobacteria are hidden endophytes in the shoots of rock plant [*Pogonatherum paniceum* (Lam.) Hack.] (Poaceae). *Environmental Microbiology Reports*, *2*(4), 619–624.

Kozlovskaia, N., Obgol'tseva, I., & Iakovleva, E. (1998). Isolation and identification of *Mycobacterium* cultures antagonistic to phytopathogenic microflora [Article in Russian]. *Antibiot Khimioter*, *43*(6), 20–23.

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

Laukkanen, H., Soini, H., Kontunen-Soppela, S., Hohtola, a, & Viljanen, M. (2000). A mycobacterium isolated from tissue cultures of mature *Pinus sylvestris* interferes with growth of Scots pine seedlings. *Tree Physiology*, *20*(13), 915–920.

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

Lee, J., Repasy, T., Papavinasasundaram, K., Sassetti, C., & Kornfeld, H. (2011). *Mycobacterium tuberculosis* induces an atypical cell death mode to escape from infected macrophages. *PLoS ONE*, *6*(3), 1–13.

Lee, Y. C. (1966). Isolation and characterization of lipopolysaccharides containing 6-O-methyl-D-glucose from *Mycobacterium* species. *Journal of Biological Chemistry*, *241*(8), 1899–1908.

Legendre, D. P., Muzny, C. A., & Swiatlo, E. (2012). Hansen's disease (Leprosy): Current and future pharmacotherapy and treatment of disease-related immunologic reactions. *Pharmacotherapy*, *32*(1), 27–37.

Li, J., Guan, X., Shaw, N., Chen, W., Dong, Y., Xu, X., … Rao, Z. (2014). Homotypic dimerization of a maltose kinase for molecular scaffolding. *Scientific Reports*, *4*, 6418.

Li, T., Huang, F., Haydock, S., Mironenko, T., Leadlay, P., & Spencer, J. (2004). Biosynthetic Gene Cluster of the Glycopeptide Antibiotic Teicoplanin: Characterization of Two Glycosyltransferases and the Key Acyltransferase. *Chemistry & Biology*, *11*(1), 107–119.

Liu, S., & Rosazza, J. (1998). Enzymatic conversion of glucose to UDP-4-keto-6-deoxyglucose in *Streptomyces* spp. *Applied and Environmental Microbiology*, *64*(10), 3972.

Long, Q., Zhou, Q., Ji, L., Wu, J., Wang, W., & Xie, J. (2012). *Mycobacterium smegmatis* genomic characteristics associated with its saprophyte lifestyle. *Journal of Cellular Biochemistry*, *113*(10), 3051–3055.

Long, R. A., & Azam, F. (2001). Antagonistic Interactions among Marine Pelagic Bacteria. *Applied and Environmental Microbiology*, *67*(11), 4975–4983.

Lopez-Marin, L. M. (2012). Nonprotein structures from mycobacteria: emerging actors for tuberculosis control. *Clinical & Developmental Immunology*, *2012*, 917860.

Losey, H. C., Peczuh, M. W., Chen, Z., Eggert, U. S., Dong, S. D., Pelczer, I., … Walsh, C. T. (2001). Tandem action of glycosyltransferases in the maturation of vancomycin and teicoplanin aglycones: Novel glycopeptides. *Biochemistry*, *40*(15), 4745–4755.

Lunin, V. V, Li, Y., Schrag, J. D., Iannuzzi, P., Cygler, M., & Matte, A. (2004). Crystal Structures of *Escherichia coli* ATP-Dependent Glucokinase and Its Complex with Glucose. *Structure*, *186*(20), 6915–6927.

Ma, Q., & Xu, Y. (2013). Global Genomic Arrangement of Bacterial Genes Is Closely Tied with the Total Transcriptional Efficiency. *Genomics, Proteomics and Bioinformatics*, *11*(1), 66–71.

Mano, H., Tanaka, F., Nakamura, C., Kaga, H., & Morisaki, H. (2007). Culturable Endophytic Bacterial Flora of the Maturing Leaves and Roots of Rice Plants (*Oryza sativa*) Cultivated in a Paddy Field. *Microbes and Environments*, *22*(2), 175–185.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., … Bryant, S. H. (2014). CDD: NCBI's conserved domain database. *Nucleic Acids Research*, *43*, D222–D226.

Mast, Y., Weber, T., Gölz, M., Ort-Winklbauer, R., Gondran, A., Wohlleben, W., & Schinko, E. (2011). Characterization of the "pristinamycin supercluster" of *Streptomyces pristinaespiralis*. *Microbial Biotechnology*, *4*(2), 192–206.

Matte, A., Tari, L. W., & Delbaere, L. T. (1998). How do kinases transfer phosphoryl groups? *Structure*, *6*(4), 413–419.

Mavromatis, K., Chu, K., Ivanova, N., Hooper, S. D., Markowitz, V. M., & Kyrpides, N. C. (2009). Gene context analysis in the Integrated Microbial Genomes (IMG) data management system. *PLoS ONE*, *4*(11), e7979.

McGrath, E. E., Blades, Z., McCabe, J., Jarry, H., & Anderson, P. B. (2010). Nontuberculous mycobacteria and the lung: from suspicion to treatment. *Lung*, *188*(4), 269–282.

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

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

Mendes, V., Maranha, A., Alarico, S., & Empadinhas, N. (2012). Biosynthesis of mycobacterial methylglucose lipopolysaccharides. *Natural Product Reports*, *29*(8), 834–44.

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

Mendz, G. L., & Hazell, S. L. (1993). Glucose phosphorylation in *Helicobacter pylori*. *Archives of Biochemistry and Biophysics*.

Miah, F., Koliwer-Brandl, H., Rejzek, M., Field, R. A., Kalscheuer, R., & Bornemann, S. (2013). Flux through trehalose synthase flows from trehalose to the alpha anomer of maltose in mycobacteria. *Chemistry & Biology*, *20*(4), 487–93.

Mildvan, A. (1987). Role of magnesium and other divalent cations in ATP-utilizing enzymes. *Magnesium*, *6*(1), 28–33.

Milshteyn, A., Schneider, J. S., & Brady, S. F. (2014). Mining the Metabiome: Identifying Novel Natural Products from Microbial Communities. *Chemistry & Biology*, *21*(9), 1211–1223.

Mishra, A. K., Driessen, N. N., Appelmelk, B. J., & Besra, G. S. (2011). Lipoarabinomannan and related glycoconjugates: structure, biogenesis and role in *Mycobacterium tuberculosis* physiology and host-pathogen interaction. *FEMS Microbiology Reviews*, *35*(6), 1126–57.

Miyazono, K. I., Tabei, N., Marushima, K., Ohnishi, Y., Horinouchi, S., & Tanokura, M. (2011). Purification, crystallization and preliminary X-ray analysis of glucokinase from *Streptomyces griseus* in complex with glucose. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, *67*(Pt 8), 914–916.

Miyazono, K. I., Tabei, N., Morita, S., Ohnishi, Y., Horinouchi, S., & Tanokura, M. (2012). Substrate recognition mechanism and substrate-dependent conformational changes of an ROK family glucokinase from *Streptomyces griseus*. *Journal of Bacteriology*, *194*(3), 607–616.

Morris, A., Gozlan, R., Marion, E., Marsollier, L., Andreou, D., Sanhueza, D., … Guégan, J. F. (2014). First Detection of *Mycobacterium ulcerans* DNA in Environmental Samples from South America. *PLoS Neglected Tropical Diseases*, *8*(1), 10.

Mostowy, S., Cousins, D., & Behr, M. A. (2004). Genomic Interrogation of the Dassie Bacillus Reveals It as a Unique RD1 Mutant within the *Mycobacterium tuberculosis* Complex. *Journal of Bacteriology*, *186*(1), 104–109.

Mulichak, A. M., Losey, H. C., Lu, W., Wawrzak, Z., Walsh, C. T., & Garavito, R. M. (2003). Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(16), 9238–9243.

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

Nessar, R., Cambau, E., Reyrat, J. M., Murray, A., & Gicquel, B. (2012). *Mycobacterium abscessus*: a new antibiotic nightmare. *The Journal of Antimicrobial Chemotherapy*, *67*(4), 810–8.

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

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

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

Norton, C. D., LeChevallier, M. W., & Falkinham, J. O. (2004). Survival of *Mycobacterium avium* in a model distribution system. *Water Research*, *38*(6), 1457–1466.

Obgol'tseva, I., Kuznetsova, O., & Iakovleva, E. (1991). Various methodological approaches in controlled screening of antibiotic producers among the group of Coryneform bacteria [Article in Russian]. *Antibiot Khimioter*, *36*(8), 6–10.

Parrish, N. M., Dick, J. D., & Bishai, W. R. (1998). Mechanisms of latency in *Mycobacterium tuberculosis*. *Trends in Microbiology*, *6*(3), 107–112.

Parsons, S. D. C., Drewe, J. A., van Pittius, N. C. G., Warren, R. M., & van Helden, P. D. (2013). Novel cause of tuberculosis in meerkats, South Africa. *Emerging Infectious Diseases*, *19*(12), 2004–2007.

Paudel, H. K., & Carlson, G. M. (1991). The ATPase activity of phosphorylase kinase is regulated in parallel with its protein kinase activity. *Journal of Biological Chemistry*, *266*(25), 16524–16529.

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

Petrini, B. (2006). *Mycobacterium marinum*: ubiquitous agent of waterborne granulomatous skin infections. *Eur J Clin Microbiol Infect Dis*, *25*(10), 609–613.

Pettigrew, D. W., Yu, G. J., & Liu, Y. (1990). Nucleotide regulation of *Escherichia coli* glycerol kinase: Initial-velocity and substrate binding studies. *Biochemistry*, *29*(37), 8620–8627.

Piepersberg, W., & Distler, J. (1997). Aminoglycosides and Sugar Components in Other Secondary Metabolites. In H.-G. Rehm & G. Reed (Eds.), *Biotechnology*.

Pimentel-Schmitt, E. F., Thomae, A. W., Amon, J., Klieber, M. A., Roth, H.-M., Muller, Y. A., … Titgemeyer, F. (2007). A glucose kinase from *Mycobacterium smegmatis*. *Journal of Molecular Microbiology and Biotechnology*, *12*(1-2), 75–81.

Poirier, V., & Av-Gay, Y. (2012). *Mycobacterium tuberculosis* modulators of the macrophage's cellular events. *Microbes and Infection*, *14*(13), 1211–9.

Pommier, M., & Michel, G. (1986). Isolation and characterization of an O-methylglucose-containing lipopolysaccharide produced by *Nocardia otitidis-caviarum*. *Journal of General Microbiology*, *132*(9), 2433–2441.

Postma, P. W., Lengeler, J. W., & Jacobson, G. R. (1993). Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. *Microbiological Reviews*, *57*(3), 543–594.

Prevots, D. R., & Marras, T. K. (2015). Epidemiology of Human Pulmonary Infection with Nontuberculous Mycobacteria. *Clinics in Chest Medicine*, *36*(1), 13–34.

Price, A. C., Zhang, Y. M., Rock, C. O., & White, S. W. (2004). Cofactor-induced conformational rearrangements establish a catalytically competent active site and a proton relay conduit in FabG. *Structure*, *12*(3), 417–428.

Primm, T. P., Lucero, C. A., & Falkinham, J. O. (2004). Health Impacts of Environmental Mycobacteria. *Clinical Microbiology Reviews*.

Rastogi, N., Legrand, E., & Sola, C. (2001). The mycobacteria: an introduction to nomenclature and pathogenesis. *Rev Sci Tech*, *20*(1), 21–54.

Riccardi, G., Milano, A., Pasca, M. R., & Nies, D. H. (2008). Genomic analysis of zinc homeostasis in *Mycobacterium tuberculosis*. *FEMS Microbiology Letters*, *287*(1), 1–7.

Rodrigues, L. C., & Lockwood, D. N. J. (2011). Leprosy now: Epidemiology, progress, challenges, and research gaps. *The Lancet Infectious Diseases*, *11*(6), 464–470.

Romero-Rodríguez, A., Ruiz-Villafán, B., Rocha-Mendoza, D., Manzo-Ruiz, M., & Sánchez, S. (2015). Biochemistry and regulatory functions of bacterial glucose kinases. *Archives of Biochemistry and Biophysics*, *577-578C*, 1–10.

Ronimus, R. S., & Morgan, H. W. (2004). Cloning and biochemical characterization of a novel mouse ADP-dependent glucokinase. *Biochemical and Biophysical Research Communications*, *315*(3), 652–658.

Roy, R., Usha, V., Kermani, A., Scott, D. J., Hyde, E. I., Besra, G. S., … Fütterer, K. (2013). Synthesis of α-glucan in mycobacteria involves a hetero-octameric complex of trehalose synthase TreS and Maltokinase Pep2. *ACS Chemical Biology*, *8*(10), 2245–55.

Ruiz, B., Chávez, A., Forero, A., García-Huante, Y., Romero, A., Sánchez, M., … Langley, E. (2010). Production of microbial secondary metabolites: regulation by the carbon source. *Critical Reviews in Microbiology*, *36*(2), 146–167.

Sakuraba, H., Goda, S., & Ohshima, T. (2004). Unique sugar metabolism and novel enzymes of hyperthermophilic archaea. *Chemical Record*, *3*(5), 281–287.

Sambou, T., Dinadayala, P., Stadthagen, G., Barilone, N., Bordat, Y., Constant, P., … Jackson, M. (2008). Capsular glucan and intracellular glycogen of *Mycobacterium tuberculosis*: biosynthesis and impact on the persistence in mice. *Molecular Microbiology*, *70*(3), 762–774.

Samuel, G., & Reeves, P. (2003). Biosynthesis of O-antigens: Genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydrate Research*, *338*(23), 2503–2519.

Sanli, G., & Blaber, M. (2001). Structural assembly of the active site in an aldo-keto reductase by NADPH cofactor. *Journal of Molecular Biology*, *309*(5), 1209–1218.

Santos, R., Fernandes, J., Fernandes, N., Oliveira, F., & Cadete, M. (2007). *Mycobacterium parascrofulaceum* in acidic hot springs in Yellowstone National Park. *Applied and Environmental Microbiology*, *73*(15), 5071–5073.

Scheeff, E. D., & Bourne, P. E. (2005). Structural evolution of the protein kinase-like superfamily. *PLoS Computational Biology*, *1*(5), e49.

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

Scollard, D. M., Truman, R. W., & Ebenezer, G. J. (2015). Mechanisms of nerve injury in leprosy. *Clinics in Dermatology*, *33*(1), 46–54.

Shao, H., He, X., Achnine, L., Blount, J. W., Dixon, R. A., Wang, X., … Foundation, N. (2005). Crystal Structures of a Multifunctional Triterpene/Flavonoid Glycosyltransferase from *Medicago truncatula*. *The Plant Cell*, *17*(11), 3141–3154.

Sharma, S. K., & Mohan, A. (2004). Extrapulmonary tuberculosis. *Indian J Med Res*, *120*(4), 316–353.

Shinnick, T. M., & Good, R. C. (1994). Mycobacterial taxonomy. *European Journal of Clinical Microbiology & Infectious Diseases*, *13*(11), 884–901.

Silva, M. T., Portaels, F., & Pedrosa, J. (2009). Pathogenetic mechanisms of the intracellular parasite *Mycobacterium ulcerans* leading to Buruli ulcer. *The Lancet Infectious Diseases*, *9*(11), 699–710.

Singh, A. K., & Reyrat, J.-M. (2009). Laboratory Maintenance of *Mycobacterium smegmatis*. In *Current Protocols in Microbiology* (pp. 10C.1.1–10C.1.12).

Skarlatos, P., & Dahl, M. K. (1998). The glucose kinase of *Bacillus subtilis*. *Journal of Bacteriology*, *180*(12), 3222–3226.

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

Stach, J. E. M., & Bull, A. T. (2005). Estimating and comparing the diversity of marine actinobacteria. *Antonie van Leeuwenhoek*, *87*(1), 3–9. doi:10.1007/s10482-004-6524-1

Stadthagen, G., Sambou, T., Guerin, M., Barilone, N., Boudou, F., Korduláková, J., … Jackson, M. (2007). Genetic basis for the biosynthesis of methylglucose lipopolysaccharides in *Mycobacterium tuberculosis*. *The Journal of Biological Chemistry*, *282*(37), 27270–6.

Stahl, D. A., & Urbance, J. W. (1990). The division between fast- and slow-growing species corresponds to natural relationships among the mycobacteria. *Journal of Bacteriology*, *172*(1), 116–124.

Stamm, L. M., & Brown, E. J. (2004). *Mycobacterium marinum*: The generalization and specialization of a pathogenic mycobacterium. *Microbes and Infection*, *6*(15), 1418–1428.

Suzuki, K., Akama, T., Kawashima, A., Yoshihara, A., Yotsu, R. R., & Ishii, N. (2012). Current status of leprosy: epidemiology, basic science and clinical perspectives. *The Journal of Dermatology*, *39*(2), 121–9.

Suzuki, K., Udono, T., Fujisawa, M., Tanigawa, K., Idani, G., & Ishii, N. (2010). Infection during infancy and long incubation period of leprosy suggested in a case of a chimpanzee used for medical research. *Journal of Clinical Microbiology*, *48*(9), 3432–3434.

Titgemeyer, F., Amon, J., Parche, S., Mahfoud, M., Bail, J., Schlicht, M., … Niederweis, M. (2007). A genomic view of sugar transport in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *Journal of Bacteriology*, *189*(16), 5903–5915.

Tortoli, E. (2003). Impact of genotypic studies on mycobacterial taxonomy: The new mycobacteria of the 1990s. *Clinical Microbiology Reviews*, *16*(2), 319–354.

Tortoli, E. (2009). Clinical Manifestations of Nontuberculous Mycobacteria Infections. *Clinical Microbiology and Infection*, *15*(10), 906–910.

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

Trifiro, S., Bourgault, A. M., Lebel, F., & Rene, P. (1990). Ghost mycobacteria on Gram stain. *Journal of Clinical Microbiology*, *28*(1), 146–147.

Tuckman, D., Donnelly, R. J., Zhao, F. X., Jacobs, W. R., & Connell, N. D. (1997). Interruption of the phosphoglucose isomerase gene results in glucose auxotrophy in *Mycobacterium smegmatis*. *Journal of Bacteriology*, *179*(8), 2724–2730.

Tyagi, J. S., & Sharma, D. (2002). *Mycobacterium smegmatis* and tuberculosis. *Trends in Microbiology*, *10*(2), 68–69.

Uesaka, I., Yoshioka, H., & Wakuda, E. (1954). On the Antibiotic Activity of Mycobacteria and Nocardia: Preliminary Report on a Tuberculostatic Substance contained in Bacteria Cell. *Acta Tuberculosea Japonica*, *4*(1), 1–6.

Van Crevel, R., Ottenhoff, T. H. M., & van der Meer, J. W. M. (2002). Innate immunity to *Mycobacterium tuberculosis*. *Clinical Microbiology Reviews*, *15*(2), 294–309.

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

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F., & van Sinderen, D. (2007). Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiology and Molecular Biology Reviews*, *71*(3), 495–548.

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

Walsh, D. S., Portaels, F., & Meyers, W. M. (2008). Buruli ulcer (*Mycobacterium ulcerans* infection). *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *102*(10), 969–978.

Wang, R., Prince, J. T., & Marcotte, E. M. (2005). Mass spectrometry of the *M.smegmatis* proteome: Protein expression levels correlate with function,operons and codon bias. *Genome Research*, *15*(512), 1118–1126.

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

Wang, Y.-N., Chi, C. Q., Cai, M., Lou, Z. Y., Tang, Y. Q., Zhi, X. Y., … Du, X. (2010). *Amycolicicoccus subflavus* gen. nov., sp. nov., an actinomycete isolated from a saline soil contaminated by crude oil. *International Journal of Systematic and Evolutionary Microbiology*, *60*(3), 638–643.

Wentworth, A. B., Drage, L. A., Wengenack, N. L., Wilson, J. W., & Lohse, C. M. (2013). Increase incidence of cutaneos nontuberculous mycobacterial infection, 1980 to 2009: A population-based study. *Mayo Clin Proc*, *88*(1), 38–45.

Williams, M. M., Yakrus, M. A., Arduino, M. J., Cooksey, R. C., Crane, C. B., Banerjee, S. N., … Donlan, R. M. (2009). Structural analysis of biofilm formation by rapidly and slowly growing nontuberculous mycobacteria. *Applied and Environmental Microbiology*, *75*(7), 2091–2098.

Wolinsky, E. (1992). Mycobacterial diseases other than tuberculosis. *Clinical Infectious Diseases*, *15*(1), 1–10.

Woodruff, P. J., Carlson, B. L., Siridechadilok, B., Pratt, M. R., Senaratne, R. H., Mougous, J. D., … Bertozzi, C. R. (2004). Trehalose is required for growth of *Mycobacterium smegmatis*. *The Journal of Biological Chemistry*, *279*(28), 28835–28843.

World Health Organization, 2015. Buruli ulcer (*Mycobacterium ulcerans* infection): Fact sheet N°199. Available at: www.who.int/mediacentre/factsheets/fs199/en/. Acessed 22 June 2015.

Zhang, Q., Zhang, Y., Zhong, Y., Ma, J., Peng, N., Cao, X., … Zhao, G. (2011). *Leptospira interrogans* encodes an ROK family glucokinase involved in a cryptic glucose utilization pathway. *Acta Biochimica et Biophysica Sinica*, *43*(8), 618–629.

Zhang, Y. H. P. (2011). Substrate channeling and enzyme complexes for biotechnological applications. *Biotechnology Advances*, *29*(6), 715–725.

Zhi, X. Y., Li, W. J., & Stackebrandt, E. (2009). An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *International Journal of Systematic and Evolutionary Microbiology*, *59*(3), 589–608.

Zimmerman, S. B., & Trach, S. O. (1991). Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *Journal of Molecular Biology*, *222*(3), 599–620.

Zotchev, S., Haugan, K., Sekurova, O., Sletta, H., Ellingsen, T. E., & Valla, S. (2000). Identification of a gene cluster for antibacterial polyketide-derived antibiotic biosynthesis in the nystatin producer *Streptomyces noursei* ATCC 11455. *Microbiology*, *146*(3), 611–619.

Zyryanov, A. B., Shestakov, A. S., Lahti, R., & Baykov, A. a. (2002). Mechanism by which metal cofactors control substrate specificity in pyrophosphatase. *The Biochemical Journal*, *367*(Pt 3), 901–906. doi:10.1042/BJ20020880

ANNEX I

(PROTOCOLS AND SOLUTIONS)

1) Agarose Gel Electrophoresis

Prepare agarose gel 1%. Load the gel with samples containing loading buffer (Takara) and with a DNA molecular weight marker (NZYDNA Ladder III, NZYTech). Run in TAE buffer 1x at 90 V for about 30 min.

TAE buffer 50x

Reagent	Amount per L	Final Concentration
Tris base	242 g	2 M
Acetic acid	57.1 mL	2 M
0.5 M EDTA pH 8	100 mL	50 mM

Dissolve Tris base in the aqueous solution of EDTA. Mix gently and heat the solution if necessary. Add the acetic acid and adjust to pH 8 with NaOH. Add distilled water up to a final volume of 1 L.

Agarose Gel 1%

Reagent	Amount per 100 mL
Agarose	1 g
TAE buffer 1x	100 mL
RedSafe™ 20000x (Intron Biotechnology)	5 µL

Add agarose to TAE buffer 1x and dissolve by heating. Cool to 40-50°C and add the RedSafe™ DNA staining solution.

2) *E. coli* BL21 Star Competent Cells

Grow *E. coli* BL21 Star cells in LB medium (see below) at 37°C and 150 rpm up to DO_{610nm}=0.3-0.4. Harvest 20 mL of cells by centrifugation in sterilized tubes for 15 min at 3000 rpm, 4°C. Gently suspend cells in 8 mL of RF1 solution (see below) and chill on ice for 15 min. Centrifuge 15 min at 3000 rpm, 4°C. Gently suspend cells in 2 mL of RF2 solution (see below) and chill on ice for another 15 min. Prepare aliquots of 100 µL of cells and store at -80°C.

RF1 Solution

Reagent	Amount per 100 mL	Final Concentration
CH_3CO_2K	296 mg	30 mM
$CaCl_2$	112 mg	10 mM
$MnCl_2$	632 mg	50 mM
Glycerol	15 mL	15%

Add reagents in this order and adjust to pH 5.8 with acetic acid to prevent precipitation. Sterilize by filtration.

RF2 Solution

Reagent	Amount per 100 mL	Final Concentration
MOPS	200 mg	10 mM
$CaCl_2$	840 mg	75 mM
Glycerol	15 mL	15%

Adjust to pH 6.8-7 to prevent precipitation and sterilize by filtration.

3) LB medium

Reagent	Amount per L
Tryptone	10 g
Yeast Extract	5 g
NaCl	5 g

For LB agar medium, add 20 g of agar per liter. Sterilize by autoclaving at 120°C, 1 atm pressure for 20 min.

4) Sodium Dodecylsulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Use a glass cassette with 1.0 mm of thickness. Prepare the Resolving Gel solution and let polymerize. Prepare the Stacking Gel solution, insert the comb and let polymerize. Remove the comb and assemble the tank. Add loading buffer (5x SDS-PAGE Sample Loading Buffer, NZYTech) to the samples and incubate at 98°C for 5 min. Load the gel with samples and with a low molecular weight protein marker (NZYTech). Add Running Buffer and run the electrophoresis at 200 V for 45 min. Stain the gel with Coomassie Staining Solution during 20 min and destain with Destaining Solution by slow shaking.

Resolving Gel 12%

Reagent	Amount
Acrylamide/bis-Acrylamide, solution 29:1 (40%) (NZYTech)	1.688 mL
H_2O	2.445 mL
1.5 M Tris-HCl pH 8.8	1.41 mL
SDS 10% (w/v)	56.3 µL
Ammonium persulfate (APS) 10% (w/v)	56.3 µL
TEMED	7.5 µL

Stacking Gel 4%

Reagent	Amount
Acrylamide/bis-Acrylamide, solution 29:1 (40%) (NZYTech)	179 µL
H_2O	1.21 mL
0.5 M Tris-HCl pH 6.8	470 µL
SDS 10% (w/v)	18.8 µL
Ammonium persulfate (APS) 10% (w/v)	18.8 µL
TEMED	3.8 µL

Running Buffer 10x

Reagent	Amount per L	Final Concentration
Tris base	30.3 g	250 mM
Glycine	144 g	1.92 M
SDS	10 g	10%

Coomassie Staining Solution

Reagent	Amount per L	Final Concentration
Coomassie R-250	1 g	0.1%
Glacial acetic acid	100 mL	10% (v/v)
Methanol	400 mL	40% (v/v)
H_2O	500 mL	-

Add the glacial acetic acid to the ultrapure water followed by methanol and the coomassie dye. Stir to homogeneity and filter to remove the debris.

Destaining Solution

Reagent	Amount per L	Final Concentration
Glacial acetic acid	75 mL	10% (v/v)
Methanol	250 mL	40% (v/v)
H_2O	675 mL	-

Add glacial acetic acid to 675 mL ultrapure water followed by methanol.

5) α-naphthol-sulfuric acid solution

Reagent	Amount
α-naphthol solution at 15% in ethanol	10.5 mL
Concentrated sulfuric acid	6.5 mL
Ethanol	40.5 mL
H_2O	4 mL

Add reagents in this order while cooling on ice.