DISSERTATION

CHARACTERIZATION OF GRCC1 AND GRCC2 PRENYL DIPHOSPHATE SYNTHASES POTENTIALLY INVOLVED IN MENAQUINONE SYNTHESIS IN *MYCOBACTERIUM TUBERCULOSIS*, AND A HOMOLOGOUS ENZYME (MS1133) IN *MYCOBACTERIUM SMEGMATIS*

Submitted by

Hana Bashir Gatlawi

Microbiology, Immunology, and Pathology Department

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2021

Doctoral Committee:

Advisor: Dean C. Crick

Sandra L. Quackenbush
Laurie A. Stargell
Rushika Perera

ABSTRACT

CHARACTERIZATION OF GRCC1 AND GRCC2 PRENYL DIPHOSPHATE SYNTHASES POTENTIALLY INVOLVED IN MENAQUINONE SYNTHESIS IN *MYCOBACTERIUM TUBERCULOSIS*, AND A HOMOLOGOUS ENZYME (MS1133) IN *MYCOBACTERIUM SMEGMATIS*

Biosynthesis pathways provide attractive drug targets in *Mycobacterium tuberculosis*. Understanding the biochemistry of the enzymes that are involved in those pathways is very important to make an achievement in this field. Menaquinones are the major lipoquinones found in *M. tuberculosis* and their synthesis has been suggested to be a valid drug target in this bacterium. Menaquinone is a key component of the *M. tuberculosis* respiratory chain and a major electron carrier during aerobic growth with many electron acceptors. It has also been reported that the isoprenoid side chain is an important antioxidant.

The genes *grcC1* and *grcC2* in *M. tuberculosis* encode proteins, probably polyprenyl diphosphate synthases, but little information is known about these enzymes although previous studies suggested that polyprenyl diphosphate synthases could be involved in the menaquinone biosynthesis. In this study, we cloned, expressed, purified and biochemically characterized the enzymatic activity of enzymes encoded by *grcC1*, *grcC2* in *M. tuberculosis* strain H37Rv, and the homologue polyprenyl diphosphate synthase *ms1133* in *Mycobacterium smegmatis*. The enzymes were active and catalyzed the condensation reaction that added [^{14}C]IPP to allylic substrates of varying chain-lengths, including dimethylallyl diphosphate (DMAPP), geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP). The purified expressed

proteins generally preferred the longer chains allylic substrates of 10 or more carbon atoms. GPP is likely the preferred substrate of GrcC1, whereas FPP is likely the preferred substrate of GrcC2 and MS1133. The values of V_{max} , K_m and K_{cat} have been calculated and the final products of these enzymes were determined. The final products of all of these enzymes are long chain polyprenyl diphosphates. GrcC1 catalyzed the formation of solanesyl diphosphate with nine isoprene units (C45) which is needed for the menaquinone biosynthesis in *M. tuberculosis*. Since the role of the enzyme in the menaquinone biosynthesis pathway has been suggested elucidation of the properties of this enzyme and obtaining more information on the biological role helps to demonstrate the involvement of the enzyme in menaquinone biosynthesis in *M. tuberculosis*. Our results support the hypothesis that the enzyme is integral part of menaquinone synthesis in mycobacteria. This dissertation also illustrates the essentiality of the studied genes. Our constructed knock-out mutations tested the essentiality of the genes *grcC1*, *grcC2* and *ms1133* for the survival and growth of the pathogen *M. tuberculosis* and its lab surrogate *M. smegmatis*. The results demonstrated that *grcC1* is an essential gene for *M. tuberculosis* survival. *grcC2* and *ms1133* both are not essential for survival, but the absence of these genes delays bacterial growth in *M. tuberculosis* and *M. smegmatis*.

ACKNOWLEDGEMENTS

No words can express how thankful I am for my advisor Dr. Dean Crick whose patience, support and kindness guided me through this journey. I am thankful to everyone who helped me in all of my struggles to complete my PhD. Also, I would like to acknowledge Dr. Rakesh Dhiman and Dr. Venugopal Pujari, and appreciate their assistance. Many thanks to my family, siblings, friends, and my husband Majed. Thank you for the greatest gift I have ever gotten, who made me better, stronger and fulfilled; my daughters, Aya, Ganah and Mennah.

DEDICATION

To the woman who taught me to believe in myself, Sabria Ahmad, my mother.

To the soul of the man who taught me how to be strong and independent,

Basheer Aljatlawi, my father.

TABLE OF CONTENTS

Chapter 1

Literature review

1-1 Mycobacteria

Tuberculosis (TB) is one of the oldest and deadliest recorded diseases (Smith, 2003). Studies of *Mycobacterium tuberculosis* (*Mtb*) genomic DNA from remains in southern Peruvian and Egyptian mummies show that human tuberculosis existed 5000-9000 years ago (Hershkovitz, 2008; Smith, 2003; Daniel, 2006). This infectious disease mainly affects the lungs, but it can also affect other organs such as the kidney, spine and bones (Smith, 2003; Daniel, 2006).

Historically, TB has not been an easy research subject due to the rapid changes in its characteristics and its ability to encounter and interact with the host cells. For example, mycobacteria intermediary metabolism can rapidly change from an aerobic carbohydrate mode to an anaerobic-lipid one (Dormandy, 2000; Smith, 2003). *Mycobacterium tuberculosis* the causative agent of tuberculosis was first stained and visualized in 1882 by Robert Koch (Daniel, 2006; Keshavjee, 2012). Streptomycin was discovered in 1943, but it was not long before *Mtb* isolates developed resistance to this first anti-tuberculosis drug (Crofton, 1948). In the 1950's, TB control was threatened as multi-drug resistant tuberculosis (MDR-TB) emerged. MDR-TB is caused by the same *Mtb* agent but is resistant to rifampicin and isoniazid, two of the first-line antibiotics for treating TB (Keshavjee, 2012; Barrow and Brennan, 1982).

About 23 percent of the world's population is estimated to have latent TB, and these individuals have a 5-10 percent lifetime risk of developing active disease (WHO, 2015). TB is the leading cause of death from a single infectious agent and the tenth leading cause of death worldwide (WHO, 2018). In 2017, an estimated 1.3 million deaths were due to TB among HIV-negative persons, and another 300,000 deaths were due to TB among HIV-positive persons (Gandhi, 2006). The World Health Organization estimates that 10 million people developed TB in 2017; which is equivalent to 133 cases per 100,000 people. For 2017, WHO estimated 550,000 new cases of rifampicin resistant TB, with 82 percent of these being MDR-TB (WHO, 2018). Millions of people have died of TB, and *Mtb* has caused chronic infectious disease all around the world (Sacchettini, 1997; Cole, 1998). Recent estimates indicate a new person is infected with *Mtb* every second, especially in developing countries (Sacchettini, 1997). A single *Mtb* bacterium is able to cause a new infection and can grow and survive in the host infected cell, yet it can survive in the environment and in the presence of disinfectants (Sacchettini, 1997; Smith, 2003).

MDR-TB is considered a major health problem worldwide, even in the presence of large numbers of antibiotics because of rapid development of drug resistance. Mycobacterial drug resistance menaces the remaining, available effective drugs used to treat MDR-TB (Pedro 2011; Scherman 2012). Since 2011, extensively drug-resistant tuberculosis (XDR-TB) has become a global matter and threatening alert because it travels easily around the world (American Lung Association, 2013; Shamaei, 2009).

Mycobacteria can not only survive through environmental challenges, but are known for their broad resistance to different kinds of antibiotics and chemotherapeutic agents. Studies have shown *Mtb's* unique structures have an important role in pathogenicity, antibiotic resistance, and

virulence. Previous studies have resolved many of the *Mtb* cell wall components and current studies are focusing on understanding the physiological aspects and components (Cole, 1998; Eoh, 2009; Brown, 2010; Pedro, 2011). MDR-TB increases the need for new drug targets rather than new drugs because *Mtb* can develop drug resistance and extensive drug resistance rapidly. To overcome bacterial persistence and resistance, the critical step is to understand the metabolic process and requirements for survival. *Mtb* resists drugs by suppressing the drug target and finding an alternative one or modifying the drug target to develop drug resistance (Gygli, 2017; Palomino, 2014). This activity causes the failure of many drugs to treat or cure TB disease, thus specific drug targets for TB must be identified (Palomino, 2014). *Mtb* is an intracellular pathogen that can persist for decades even with the presence of host immune response (Pandey, 2008) and TB drug candidates need to target *Mtb* persistence and metabolic state (Mdluli 2015, Watanabe 2001).

New drug target discovery is the principal aim of pharmaceutical companies and many researchers (Rask-Andersen, 2011). Development or discovery of a new drug is expensive, difficult, and time consuming. Each year the drug industry spends more than $50 billion on drug development with scant knowledge of the specific molecular targets within pathways (Overington, 2006). MDR-TB requires a combination of many antibiotics at the same time over a long treatment period, and entails side effects and a considerable chance of resistance (Drews 1996). The main theory behind the therapeutic effect of drugs is these drugs bind and regulate the activity of particular protein targets in the bacterial cell. Many drug targets have been explored and other novel drug targets need to be discovered, especially with the increase in drug-resistant tuberculosis and its co-infection with other diseases (Zhen 2006). The main focus of some of the current research in this field is to find new drug targets more than to find new compounds (which

the bacteria can counteract easily and against which it can develop resistance). It has been recognized in some research fields, such as cholesterol control, that targeting essential enzymes is a powerful approach that results in discovery of promising and new antimicrobials (Scherman 2012). The results of the present study are expected to be significant in discovering essential proteins for *Mtb* survival and, based on this, in developing new drug targets for existing drug agents or new drug targets against *Mtb* specifically and Gram-positive bacilli generally. It is important to survey genes in the *Mtb* genome in order to obtain some gene inhibitors or suppressors because of the possibility that an inhibitor may be effective with some genes of the same strain and not with others.

Highlighting the aforementioned difficulty in controlling tuberculosis, and continued, urgent search for new drug targets, our lab is interested in isoprenoid biosynthesis and has long conducted research in this field. The results of this research provided evidence that lipoquinones in general and menaquinone in specific are promising drug targets. (Crick et al, 2000; Kaur et al, 2004; Dhiman,et al, 2005; Dhiman et al, 2009; Eoh, et al, 2009; Narayanasamy et al, 2010; Li et al, 2014; Upadhyay et al, 2018; Dhiman et al, 2019; Kumar et al, 2020). Accordingly, inhibitors of menaquinone biosynthesis and enzymes involved in electron transport chain (ETC) become a potential target in the drug development process against MDR-TB.

This study will mainly characterize the *Mtb* genes *grcC1*and *grcC2* which encode prenyltransferases that catalyze isoprenoid synthesis. It also, demonstrates the essentiality of these genes for the growth and survival of Mtb. In the upcoming sections of this thesis, we will give a brief introduction to the roles and importance of mycobacterial electron transport chain, menaquinone, isoprenoid biosynthesis and prenyltransferases to bacterial growth and survival.

1-2 Electron transport chain

In 1884, Charles Alexander MacMunn investigated what he called histohematin, an intracellular respiratory pigment found widely in nature and having a role in cell respiration. The respiratory chain or cytochrome system was reported as a respiratory pigment (composed of three hemoproteins), which has a role in aerobic respiration (Anraku, 1988). Later, oxidative phosphorylation was elucidated, and the search continued to investigate energy transformation to determine proton flux and the chemiosmotic coupling theory (Mitchell 1979). Cytochrome properties and respiratory chain structures have been studied; and their components have been investigated (Anraku 1988, Mitchell 2011). These investigations highlighted the importance of the bacterial ETC as a genetic model to study cytochrome biosynthesis and biochemical development of respiratory metabolism (Anraku 1988). The ETC main function is to translocate protons out of the cytoplasmic membrane to generate proton motive force which drives ATP synthesis (Anraku 1988). This system allows the bacterial cell to overcome challenges in the growth environment and survive. Mainly, the bacterial respiratory system or ETC includes varied dehydrogenases whose function is to transfer electrons and protons to the menaquinone (MK) or ubiquinone (UQ) pool which will be re-oxidized by the final electron acceptor (Nowicka, 2010; Soballe, 1999).

Since electron transport chain components and ATP synthesis are needed for mycobacteria to survive; mycobacteria have a respiratory system that allows them to have effective energy production in different environments (Figure 1) (Berney, and Cook. 2010). Even though, mycobacterial strains can produce ATP by using both oxidative phosphorylation and

fermentable carbon substrate sources, the oxidative phosphorylation pathway is essential and critical for growth and survival of all mycobacterial strains and it is validated drug target (Tran, 2005; Cook, 2014; Bald, 2017; Li, 2014).

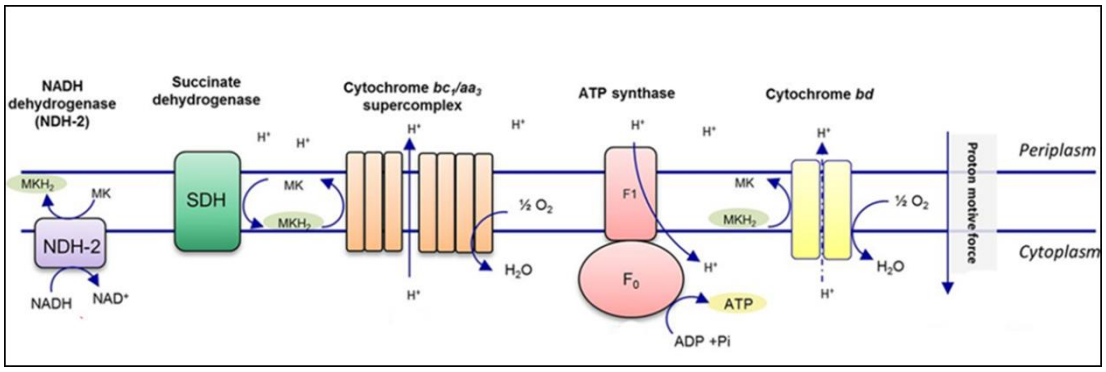


Figure 1: Membrane electron transport system compartments in mycobacteria (Bald, 2017). *Mtb* has two copies of type II NADH dehydrogenase (NDH2) and two succinate dehydrogenase enzymes (Sdh1 &2), and a central menaquinone pool (MK/MKH_2 , highlighted with green ovals).

In most of bacterial species, the electron transport chain's main proteins or components are: complex I (NADH dehydrogenase); complex II (succinate dehydrogenase); and complex III (cytochrome super complex and oxidase). The *Mtb* electron transport chain has type I and II NADH dehydrogenases (type II NADH dehydrogenase presented in 2 copies), two succinate dehydrogenases Sdh-1 and Sdh-2, fumarate reductase and the super-complex (cytochromes bc1 and aa3). Menaquinone is a mobile part of this process which moves around inside the lipid layer of the membrane between these complexes (Matsoso, 2005; Debnath, 2012; Bald, 2017). It is a selective electron acceptor and mainly accepts electrons from NADH, formate and hydrogen (Soballe, 1999).

In the *Mtb* ETC, the NADH dehydrogenase feeds the electrons that come from the electron donor NADH into the ETC (Figure 2). As a result, the menaquinone pool gets reduced (MK/MKH_2) and then the electrons are transferred to the cytochrome supercomplex which

transfers the electrons onto the final electron acceptor (oxygen) (Matsoso, 2005; Bald 2017). The electrons move from an electron carrier with high reducing potential to an electron carrier with low reducing potential and during the transportation of the electrons in the respiratory chain, the protons get pumped to the other side of the membrane which results in a proton motive force (PMF). ATP synthase complex uses the PMF energy to generate ATP (Upadhyay, 2015). This process makes the cytoplasm an alkaline environment with high OH^- and low H^+ content and the continuation of the oxidative phosphorylation process neutralizes the external acid pool and cell cytoplasm, which makes the continuation of electron transport vital for the maintenance of pH gradient (Mitchell, 2011; Richardson, 2002; Bald 2017).

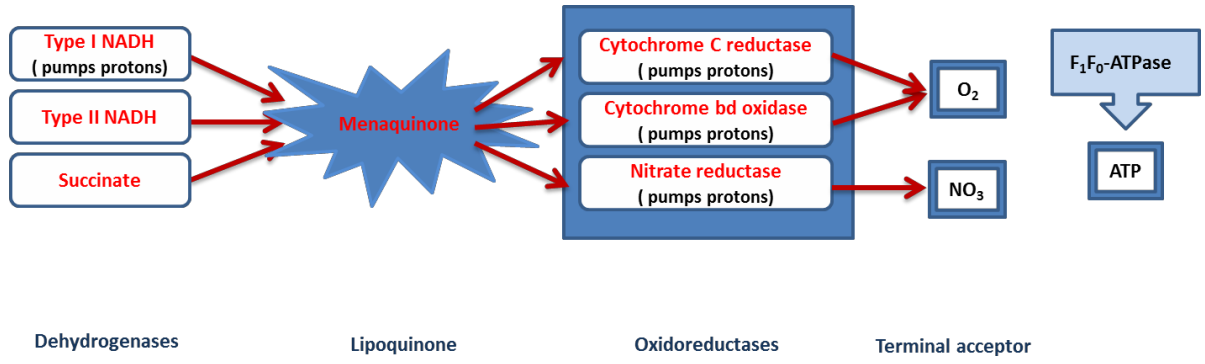


Figure 2 **:** The electron flow in selected respiratory pathways in mycobacteria showing the role of mycobacterial menaquinone in oxidative phosphorylation and respiration (Kurosu, 2010; Dhiman, 2009)

Inhibition of the enzymes NADH type II dehydrogenase and F_1F_0-ATP synthase have been reported to kill MDR-TB (Andries et al, 2004 and Edward, et al 2005). Also, targeting menaquinone synthesis may pull the ETC apart and result in reduction of ATP synthesis in both persistent and active *Mtb* (Eoh, 2009). Mycobacteria maintain ATP synthesis and PMF during both respiration and hypoxia by the alteration of the terminal electron acceptor (Mdluli,2015,

Watanabe, 2011). The electrons flux from the menaquinone pool to the cytochrome complex and are transferred to the oxidase system where oxygen is used as the terminal electron acceptor during respiration and succinate is the acceptor in the absence of respiration (Mdluli,2015, Watanabe, 2011). In mycobacteria, menaquinone biosynthesis is considered to be critical under all conditions (aerobic and hypoxic) (Dhiman, 2009; Kurosu, 2009).

1-3 Menaquinone

Lipoquinones are lipid-soluble components of the bacterial membrane bound electron transport chain (Dhiman, 2009; Soballe, 1999; Bentley, 1982). These membrane bound compounds function principally as electron and proton shuttles. Isoprenoid quinones, are present in all organisms, and have been divided into two major taxonomic types: benzoquinones and naphthoquinones (known as bacterial respiratory quinones). The benzoquinone group contains UQ or coenzyme Q and the naphthoquinone group contains menaquinone (MK or K_2) (Nowicka, 2010; Soballe, 1999). According to the basic structural consideration, the naphthoquinone group is divided further into two major families: phylloquinones or vitamin K_1 and menaquinone or vitamin K2 (Collins, 1981; Kindberg,1989).

Vitamin K or Koagulating vitamin was discovered and chemical nature determined in 1936 by Henrick Dam and Edward Doisy (Dam, 1936; Halder, 2019). The German word “Koagulation”, from which the “K” is derived, refers to the formation of blood clot (Halder, 2019). Vitamin K_1 is found commonly in, and isolated, from plants. Bacteria do not synthesize vitamin K_1 but they can synthesize other forms of vitamin K. MK-7 (C35) was found in 1948 by Tishler and Sampson in *Bacillus brevis* (soil bacteria) (Collins 1981; Kurosu 2010). In 1949, a different menaquinone compound with a 45-carbon side chain (solanesyl) was discovered by Francis et al in *Mtb* (Collins

& Jones,1981). Variations in molecular structure and different distribution in bacteria give the menaquinone molecules an important role in bacterial taxonomy, besides their main role in bacterial electron transport chain, active transport, and oxidative phosphorylation (Kurosu, 2010; Collins, 1981). Menaquinone is required for bacterial respiration, photosynthetic electron transport chain and cell wall synthesis (Nowicka, 2010; Soballe 1999).

Menaquinones (2-methyl-3-polyprenyl-1,4-naphthoquinones) are common isoprenoid respiratory microbial quinones, found in Gram-positive and anaerobic Gram-negative bacteria, and are the main lipoquinone compounds in mycobacteria (Eoh, 2009; Nowicka 2010). Some Gram-positive bacteria and *Mycobacterium* spp. use only menaquinone in their electron transport chain and the biosynthesis of menaquinone is necessary for their survival (Collins, 1981; Dhiman, 2009). Menaquinone's chemical structure has the head naphthoquinone ring and isoprenoid side chain (tail) which varies in length based on the organism (Figure 3) and is composed mostly of 6-14 prenyl units (Soballe, 1999; Nowicka, 2010). Mycobacteria synthesize MK-9 (nine is the number of isoprene units in the side chain) and Gram-negative bacteria *E. coli* synthesize MK-8, UQ-8 and MK-7. The existence of menaquinone's lipid soluble side chain and its ability of reversible reduction make it the perfect compound to function as electrons shuttle between the proteins complexes of biological structures (Lichtenthaler 1977, Kellogg 1997, Nowicka 2010). The menaquinone side chain is usually unsaturated although sometimes it is saturated or partially saturated, depending on the species growing environment (Nowicka 2010).

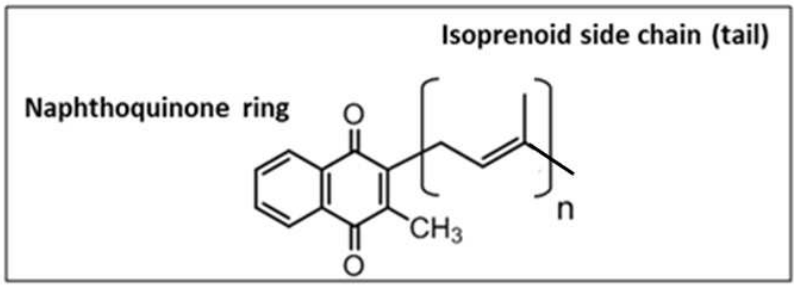


Figure 3: The structure of Menaquinone: *M. tuberculosis* – MK (n=9) (Eoh, 2009). n, refers to the number of isoprene units.

Menaquinone's role in electron transport chain

In Gram-positive bacteria in general and in *Mtb* specifically, MK is the only quinone used in the electron transport chain and its unique biosynthesis pathway is essential for bacteria (Kurosu, 2010; Nowicka, 2010; Soballe 1999; Dhiman, 2009). This combined with the absence of the synthesis pathway in humans' makes menaquinone an attractive target for anti-tuberculosis drug discovery (Dunphy 1968; Chen, 2013; Kurosu, 2010; Meganathan, 2009). Menaquinone plays roles in aerobic and anaerobic respiration processes (Black 2014). It transfers two electrons from electron donors (NDH-1, NDH-2 and succinate dehydrogenase) to the terminal electron acceptor as illustrated in Figure 2 (Kainou, 2001; Black 2014). The electron transport chain is considered a central component in ATP production and bacterial growth, which gives the menaquinone involved in ETC a central and important role in ATP synthesis (Yasuhiro,1988; Haddock,1977).

Menaquinone synthesis / Menaquinone biosynthesis pathway

Biosynthesis pathways contain attractive drug targets in *Mtb* and understanding the biochemistry of the enzymes involved in those pathways is critical to advancement in this field (Dhiman 2009; Mdluli, 2015; Upadhyay, 2015; Debnath, 2012). The nine isoprenoid unit menaquinones are the major lipoquinones found in *Mtb* and their synthesis has been suggested to be a valid drug target in this bacteria (Mdluli, 2015; Upadhyay, 2015; Debnath, 2012). Understanding mechanisms of the pathway and characterizing the enzymes involved would contribute to identification and validation of mechanisms as a biological target in mycobacteria and to the search for better inhibitors. Previous studies proved the ability of some compounds to inhibit aerobic growth and persistence in *Mtb* and suggested that menaquinone biosynthesis inhibitors are promising compounds for treating *Mtb* infections (Dhiman 2009; Li, 2014; Mdluli 2015).

Recently, the biosynthetic pathways in *Mtb* have become potential targets for new tuberculosis drugs (Dhiman 2009; Black 2017; Li 2017). Characterizing the enzymatic activity of mycobacteria genes could contribute to understanding the interaction between their function and the function of other drug targeted genes (Dhiman, 2009; Morand, et al 1997; Collins, 1981). The synthesis of menaquinone is a new target of drugs and drug derivatives as shown in recent research studies (Kurosu, 2010, Dhiman2009, Debnath 2012).

The naphthoquinone ring and the isoprene units in menaquinone are synthesized separately and then condensed together by prenyltransferases enzymes (Nowicka 2010). The biosynthesis of menaquinone's molecule is controlled by groups of enzymes encoded by 2 genes clusters; the first cluster of *men* genes includes: *menB, C*,*D*,*E* and *F*; and the second cluster includes *menA*, *H*, *G, J* and *I* (Shineberg, 1976; Nowicka 2010; Black 2014; Mdluli 2015; Upadhyay 2015;

Cook,2017).

The synthesis of menaquinone's head group precursor (1,4-dihydroxy-2-naphthoate (DHNA)) starts with the chorismate via the shikimate pathway. In the shikimate pathway, shikimate is converted to chorismate which is the precursor for the formation of the naphthoquinone ring. S-adenosylmethionine and prenyl diphosphate are the precursor of methyl and isoprenoid side chains. The benzenoid derivative o-succinylbenzoate (OSB) and naphthalenoid compound are the major intermediates in this reaction (Bentley, 1982, Kurosu, 2010). Chorismate is isomerized to isochorismate by the enzyme MenF and the addition of α-ketoglutarate with isochorismate is catalyzed by MenD to form 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexadiene-1-carboxylate. The enzyme MenH removes the pyruvate portion of this compound to produce 2-succinyl-6-hydroxy-2,4-cyclo-hexadiiene-1-carboxylate, which forms o-succinylbenzoate in a process catalyzed by MenC. MenE converts o-succinylbenzoate to o-succinylbenzoate-CoA by MenE, and MenB catalyses the formation of 1,4-dihydroxy-2- naphthoyl-CoA (condensation of o-succinylbenzoate-CoA). The enzyme thioeterase hydrolyses 1,4-dihydroxy-2-naphthoyl-CoA to form 1,4-dihydroxy-2-naphthoate, which is prenylated by MenA and methylated by MenG as illustrated in Figure (4), the dominant menaquinone form is reduced by MenJ (Kurosu, 2010; Upadhyay, 2015). The menaquinone side chain (prenyl diphosphate), with appropriate length of isoprene units (n=9) in MK-9 in *Mtb* is biosynthesized by chain elongation reaction (described below), with the size of the menaquinone side chain varying in the same organism and between different species (Nowicka 2010, Kurosu, 2010).

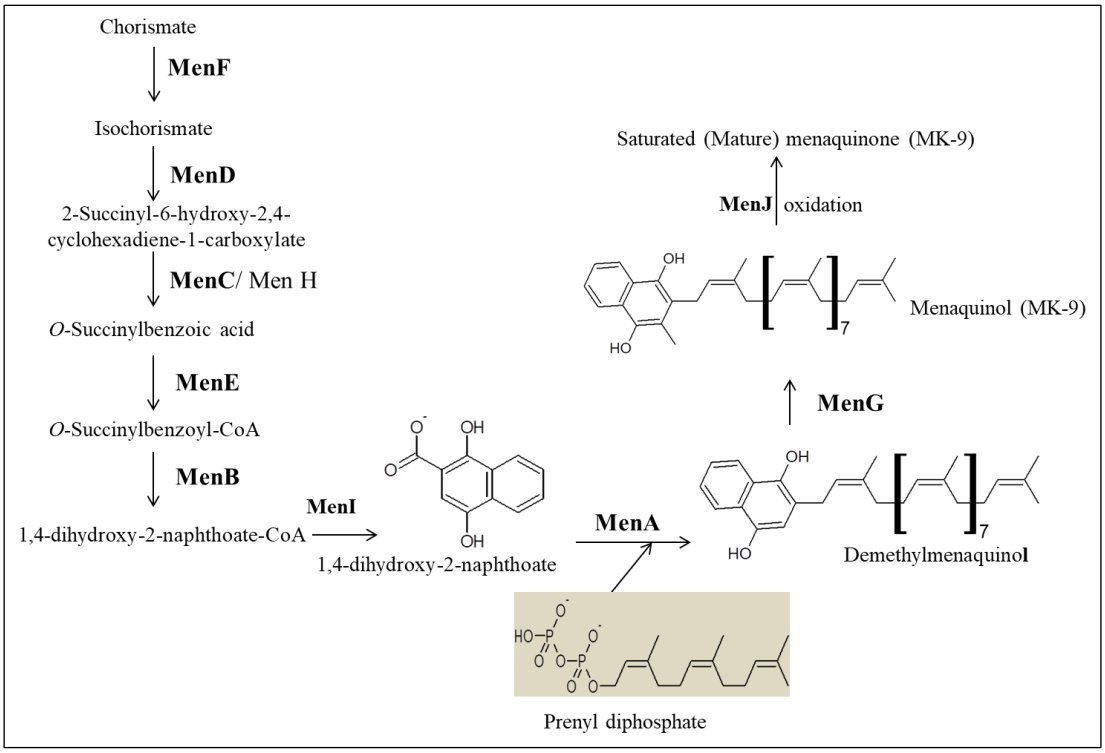


Figure 4 : Menaquinone biosynthesis pathway in mycobacteria. (Upadhyay 2015; Crick unpublished work).

1-4 Prenyl diphosphate synthases

Isoprenoids are a family of chemical compounds that have diverse chemical structures and serve different functions in all living creatures and biotechnology (Takahashi 1998, Camposa 2001). Isoprenoids are the main building unit in *Mtb* diterpene virulence factor and the virulence factor is synthesized by the condensation reaction of isopentenyl diphosphate (IPP) with dimethylallyl diphosphate (DMAPP), which is catalyzed by a group of *E*-prenyl synthases (Oldfield, 2015; Layre 2014). Isoprene derived products are involved in various processes, such as respiration, vitamin K, energy utilization and membrane structure (Rip,1985). Targeting this series of enzymes is not only a potential drug target in the menaquinone biosynthesis pathway but for some *Mtb* virulence factor synthesis (Oldfield, 2015).

Isoprenoid quinones exist in most bacteria, and are composed mainly of hydrophilic head group (polar) and hydrophobic isoprenoid side chain (Collins, 1981). Many isoprenoids have been characterized in *Mtb*, such as the prenyl side chain of menaquinones (Figure 3) (Nowicka, 2010; Eoh, 2009; Hemmi, 2001). They are generated from the condensation of the two main precursors, IPP and its isomer DMAPP. The condensation is catalyzed by a group of enzymes named prenyl diphosphate synthases (also known as prenyltransferases) (Eoh, 2009; Takahashi, 1998; Wang, 2000). These two precursors are biosynthesized principally via two pathways: 2C-methyl-erythritol 4-phosphate pathway (MEP) and the mevalonate pathway (MVA) (Eoh, 2009; Oldfield 2012). The majority of organisms use only one of the two pathways to synthesize the required precursors. *M. tuberculosis* uses only the MEP pathway to biosynthesize these two precursors and it is an essential pathway for Mtb (Eoh, 2009; Collins, 1981, Rodriguez, 2002, Wang, 2018).

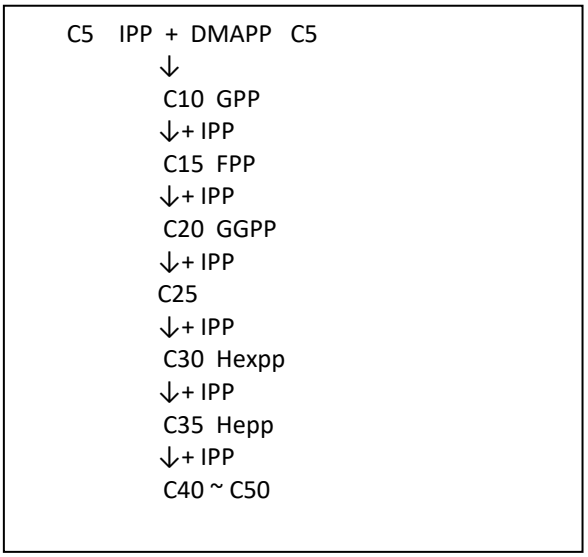


Figure 5: Isoprenoid Biosynthetic pathway (Muramatsu, 1997).

Prenyl diphosphate synthesis or chain elongation

All isoprenoid compounds are synthesized from the five carbon precursor isopentenyl IPP (Dewick, 1995; Kellogg 1997), and their biosynthesis is essential in all living organisms (Campos, 2001; Wang, 2002; Hemmi, 2002). The main biosynthetic mechanism of the isoprenoid chain starts with the condensation reactions of IPP and DMAPP to produce geranyl diphosphate (GPP) (Schulbach, 2000; Oldfield 2012), or until the products reach the required chain length (Kaur, 2004, Liang 2002). This condensation reaction is the basic building unit for chain elongation of the biosynthesis reaction not only for bacterial isoprenoids but for many diverse products such as sterols, dolichols, carotene and the side chain of respiratory coenzymes (Rip, 1985, Wang, 2018).

1. The synthesis of reaction precursor IPP and DMAPP:

The C_5 precursors DMAPP and IPP are mainly made by two pathways, namely, mevalonate (MVA) and methyl-erythritol phosphate (MEP) pathways (Rodriguez, 2002). The MEP pathway is the only pathway used by many eubacteria including *Mtb*, to synthesize the two common precursors IPP and DMAPP, making it an attractive drug target (Eoh 2007, Oldfield 2012, Campos, 2001).

The MEP pathway, starts with formation of 1-deoxy-D-xylulose 5-phosphohate (DXP) from pyruvate and glyceraldehyde 3-phosphate which is catalyzed by 1-deoxy-D-xylulose 5-phosphate synthase (DXS). The enzyme 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR) catalyzes the conversion of the formed DOXP to MEP. By enzyme mediation, 4-diphosphocytidyl-2-C-methyl-D-erythritol (CDP-ME) is formed from the condensation of MEP with CTP, which is phosphorylated to CDP-ME2P. 2-C-methyl-D- erythritol 2,4-

cyclodiphosphate is the last intermediate molecule. The isopentenyl monophosphate (IP) is phosphorylated to IPP and DMAPP (Campos, 2001, Nowicka 2010, Boucher 2002, Wang, 2018). The MEP precursor's synthesis pathway is illustrated and summarized in Figure (6).

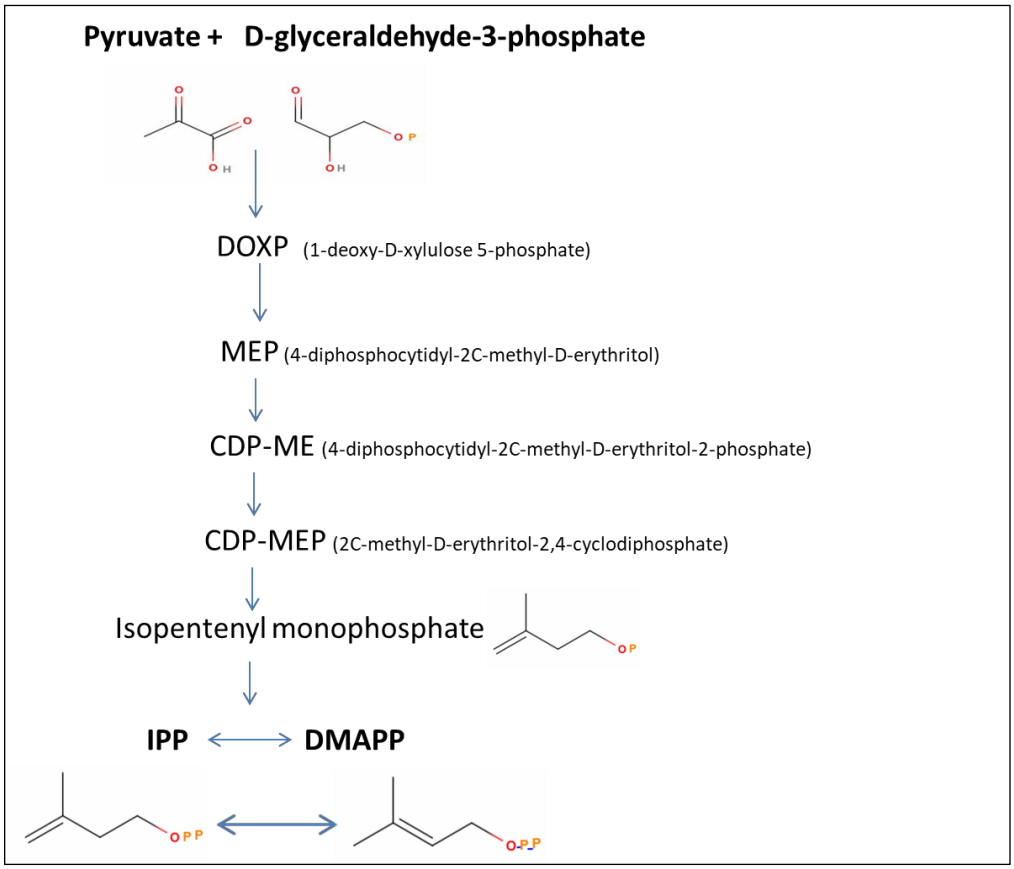


Figure 6: MEP pathway for the synthesis of precursor IPP. (Boucher, 2002; Kollas,2002)

2. Isoprenyl diphosphate chain elongation

Starting with the precursor (IPP), the *E*-prenyl synthases synthesize varied chain length linear isoprenyl diphosphates (C10, C15, C20, C25, C30, C35, C40, C45 and C50) and generally these are restricted to specific chain lengths of linear isoprene units. The enzymes that synthesize the linear prenyl-chain are *Z*- and *E*-prenyl diphosphate synthases. All types of *E*-prenyl diphosphates are synthesized by enzyme type *E*-polyprenyl pyrophosphate synthases (Rip, 1985, Liang 2002, Takahashi, 1998).

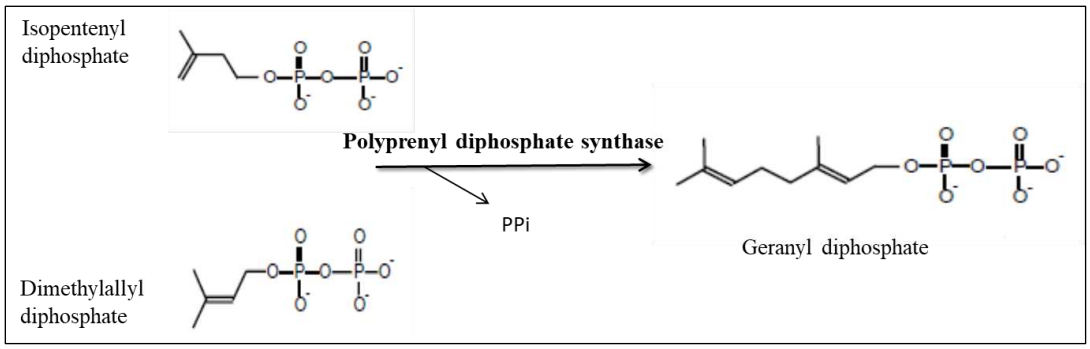


Figure 7: The fundamental condensation reaction between C4 of isopentenyl and C1 of allylic diphosphate (Poulter, 1976 ; Ogura1998).

The substrates DMAPP and IPP are condensed to synthesize all *E*-prenoid diphosphates GPP, FPP and GGPP. These reactions are catalyzed by the specific enzymes geranyl diphosphate synthase (GPS), farnesyl diphosphate synthase (FPS), and geranylgeranyl diphosphate synthase (GGPS). The reaction starts by head-to-tail condensation reaction of IPP and DMAPP after they bind to the enzyme (Fujii 1982, Crick, 2000). In the allylic substrate, the carbocation forms at the position C1' and attacks the C4' of IPP (this process requires divalent cations such as Mg^{++}) leading to the formation of C-C bond between the allylic substrate and IPP (Heider, 2014). Prenyltransferases or prenyl diphosphate synthases are the key enzymes in the isoprenoid biosynthesis and they are able to synthesize prenyl diphosphates with specific chain length and

stereochemistry by the repetition of this condensation reaction between IPP and allylic product, see Figure 7 (Ogura 1998; Kharel, 2006; Ye, 2007). Members of this enzyme family have the same catalytic reaction, but they have different properties that lead to their classification into different groups. The classification categories include the prenyl chain length of the final product "catalytic function", the formed double bond of the outcome stereochemistry, and the mode of subunit composition (Ogura 1998; Ohnuma, 1996; Poulter 1976).

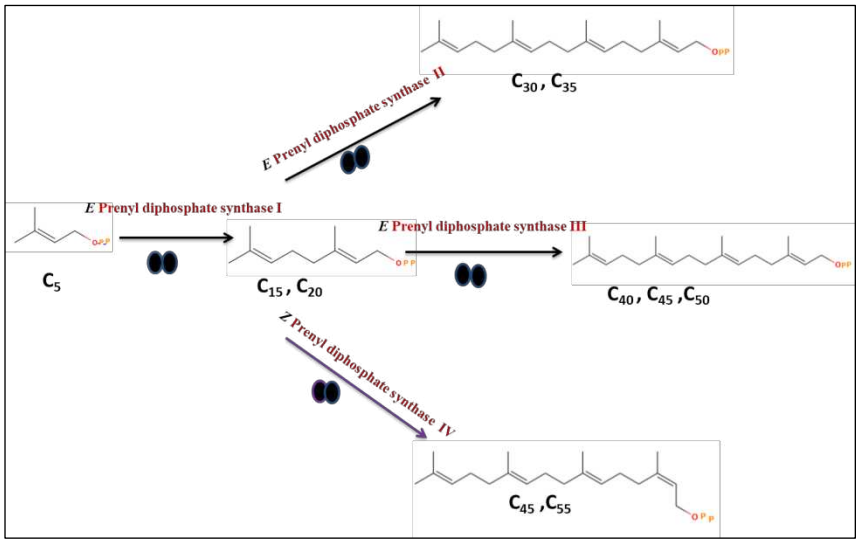


Figure 8: Prenylsynthase classification (Ogura, 1998): Although the reactions that prenylsynthases (polyprenyl diphosphate synthases) catalyze are similar, their properties are very contrasting, that they can be classified into four primary groups depending on the subunit composition, stereochemistry and prenylchain length of the final product. The colored ovals represent the enzyme subunit composition.

The classification depending on stereochemistry of the double bond formed during chain elongation, *Z*- and *E*- prenyl diphosphate synthase. The structures of *E*- and *Z*-prenyl diphosphate synthases are different but they catalyze similar reactions. Generally, all *E*-prenyl diphosphate synthases are derived from a standard precursor and catalyze the synthesis of products with (*E*-) double bonds with chain length up to C_{50} (Vandermoten 2009; Liang 2002). These prenyl

diphosphate synthases produce products with the specific chain length required for functions and commonly produce the *Z*- prenyl diphosphate synthase synthesis long-chain final products (Liang, 2002). *Z*-Prenyltransferases are considered as membrane enzymes and are isolated and characterized from some bacterial systems (Ericsson 1992).

According to their catalytic function, they are mainly classified into three aspects: Isoprenyl diphosphate synthases; protein prenyltransferases; and terpenoid cyclases (Liang, 2002). The enzyme's final product chain length is species-specific; in each organism, the polyprenyl diphosphate synthases are responsible for determining the specific chain length (Ohara, 2010, Hirooka 2000, Wang, 2000). Depending on the chain length of the final product, prenyltransferases are divided into four major groups: 1. Short-chain prenyl diphosphate synthases (Prenyltransferase I); 2. Medium-chain prenyl diphosphate synthases (Prenyltransferase II); 3. Long-chain *E*-prenyl diphosphate synthases (Prenyltransferase III); and 4. Long-chain (Z)-polyprenyl diphosphate synthases (Prenyltransferase IV). Classification Figure (8) (Ogura 1998):

1. Short-Chain Prenyl Diphosphate Synthases (C10 – C25):

Prenyltransferases in this group are geranyl diphosphate synthase (GPS), farnesyl diphosphate synthase (FPS), and geranylgeranyl diphosphate synthase (GGPS). All enzymes in this group have homodimer structures, a molecular mass range of 32-44 kDa, and require divalent cations for their reactions (Kellogg, 1998; Liang 2002; Vandermoten 2009). Short-chain prenyl diphosphates are the precursors for the synthesis of other isoprenoids.

2. Medium-chain prenyl diphosphate synthases (C30 – C40) :

These prenyltransferases elongate FPP to produce *E*- polyprenyl diphosphate in various chain lengths between C30 and- C40. Hexaprenyl diphosphate synthase (HexPS) is a bacterial

prenyltransferase that has a novel subunit structure and catalyzes the C30 chain elongation. This enzyme found in *Micrococcus luteus* synthesizes menaquinone-6 (MK-6). Heptaprenyl diphosphate synthase catalyzes the production of MK-7 in the Gram positive bacteria *Bacillus subtilis* (Ogura 1998). This transferase is only specific in producing medium chain products and is unable to catalyze the synthesis of short chains (C_5 -C_{10}-C_{15}). Heptaprenyl diphosphate synthase catalyzes the reaction by adding four IPP units to the substrate FPP to synthesize (*all-E*)-HepPP (Ogura 1998; Takahashi 1980).

3. Long-chain (*E*)-prenyl diphosphate synthases (C40 – C50):

The molecules of this category are the substrates for menaquinone and ubiquinone biosynthesis and are usually peripheral membrane proteins (Kellogg 1997). Since bacteria produce ubiquinones or menaquinones, they are considered the main source of long-chain prenyl diphosphate. This group of enzymes are active homodimer proteins that catalyze chain elongation up to C45–PP and more starting from GPP (C10). The example enzyme for this group is (*E*)-nonaprenyl (solanesyl) diphosphate synthase (SPP) in *M. luteus*, and their homologues in *E. coli*, namely *all-E*-octaprenyl diphosphate synthase (C_{40}) and *all-E*-decaprenyl diphosphate synthase (C_{50}) (Ogura 1998).

4. Long-chain (*Z*)-polyprenyl diphosphate synthases (C40 – C60):

This group of enzymes is connected to periplasmic membranes, which are not active themselves and require phospholipids or detergents to be activated. Some plants contain *Z*-polyprenyl diphosphate synthases, which catalyzes the same condensation reaction of IPP with allylic primer and produces *Z*-polyprenyl diphosphate with chain length C_{40}- C_{60} (Kellogg 1997; Ogura 1998; Takahashi, 1982; Ohnuma 1996).

Additionally, since the *E*-isoprenyl diphosphate synthases were classified according to their product chain length and the double bond stereochemistry, some studies classified the *E*-isoprenyl diphosphate synthase FPP into two subgroups and GGPP into three subgroups according to the similarities in their amino acid sequences and the mechanism of chain length determination. (Wang, 2000; Ohnuma,1996). All cloned *E*-isoprenyl diphosphate synthases encode FPP, GGPP, HexPP, HepPP, OPP and SPP, and have two conserved asparate rich motifs; the first aspartate rich motif (FARM) and second aspartate rich motif (SARM), which may provide product specificity. The fifth amino acid before the FARM is a major key for the determination of product final chain length, and any mutation in this position prevent the enzyme from producing longer chain products (block the chain elongation) (Ohnuma, 1998). The amino acid sequence around long chain *E*-prenyl diphosphate synthase's FARM is different from any other prenyl diphosphate synthases (Wang and Ohnuma, 2000). Due to these conserved regions, these enzymes are supposed to be derived from the same ancestor (Ohnuma, 1998).

Aims of the study

Aim I: Identify prenyl diphosphate synthases that could be involved in menaquinone biosynthesis by cloning, expression and biochemical characterization of polyprenyldiphosphate synthases (GrcC1 & GrcC2) in *Mtb* strain H37Rv.

Aim II: Demonstrate the essentiality or non-essentiality of GrcC1, GrcC2 and the homologous in *Mycobacterium smegmatis* (MS1133).

Aim III: Characterization of a long chain polyprenyl diphosphate synthase of *Mycobacterium smegmatis* homologous to GrcC1 and GrcC2 in *Mtb.*

Chapter 2

Identifying polyprenyl diphosphate synthases in *Mycobacterium tuberculosis*

Mtb H37Rv is the most characterized strain of mycobacteria. This strain has the complete genome sequenced and analyzed. The genome consists of approximately 4000 genes (Cole, 1998). In the DNA database, where the whole genomic sequences of *Mtb* and other Gram-positive bacteria are registered, the genes encoding the homologues of known long-chain prenyltransferases were identified. Basic Local Alignment Search Tool (BLAST) search of the inquiry genome sequence (homology search program), gene analysis showed the presence of 2 homologues encoding putative long-chain polyprenyl diphosphate synthases (Figure 9).

Mtb gene, *grcC1*, is annotated as a prenyltransferase that may have a role in the MK biosynthesis. Also, the gene, *grcC2*, is predicted to be a *trans*-hexaprenyltransferase. Thus both *grcC1* and *grcC2* encode proteins, probably polyprenyl diphosphate synthase (Tuberculist data base), found in the gene cluster under the category of cell process and metabolism (Tuberculist data base; Camacho, 1999). In *Mtb* genomic region, *grcC1* is located near previously identified genes involved in MK biosynthesis (*menC, D , H* and *J*) (Figure 10). Little information is known about these putative prenyl diphosphate synthases and previous studies suggested that they could be involved in the menaquinone biosynthesis as intermediate enzymes. The library of transposon mutations of *M. tuberculosis* predicted that *grcC1* is an essential gene for bacterial growth and *grcC2* is non-essential (Tuberculist data base; Sun 2001).

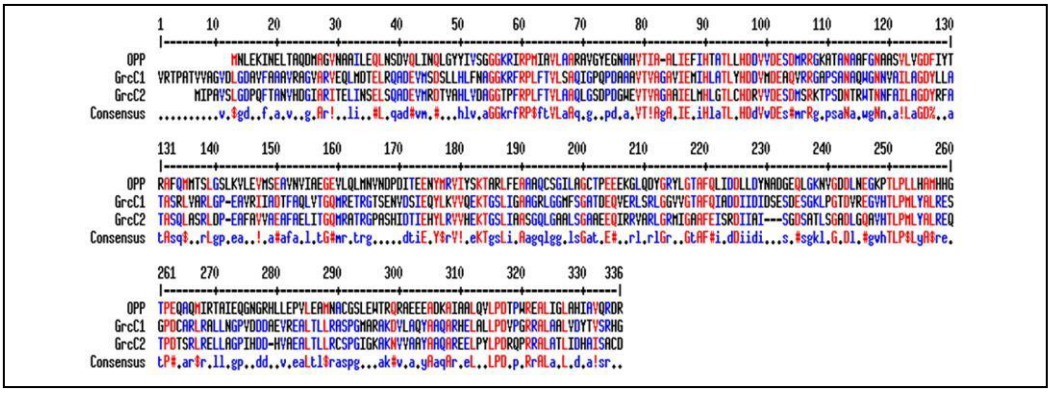


Figure 9: Amino acid sequence alignment GrcC1 & GrcC2 with known long-chain polyprenyl diphosphate synthase (OPS) from *Bacillus subtilis*. The red color represents highly conserved amino acids, and blue color represents the weakly conserved ones.

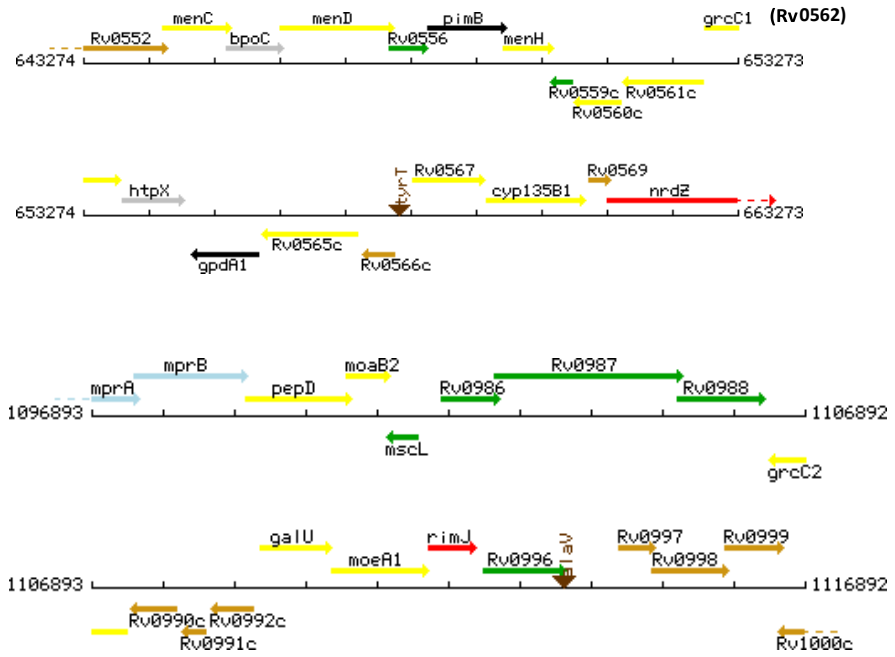


Figure 10: Organization of genomic region in *M. tuberculosis* showing genes *grcC1* and *grcC2*. *grcC1* is located near previously identified genes *menC, D and H* , and upstream the gene *menJ* [Rv0561], which are all involved in MK biosynthesis pathway. *grcC2* is not located near MK synthesis genes.

Protein-protein interaction networks were identified (Figure 11) using the website www.STRING-db.org. This identification functionally connected a group of protiens (prenylsynthases/transferases). According to the functional analysis; all of these proteins (idsB, MenA, GrcC1 and GrcC2) may be involved in the same pathway since they are predicted functional partners. Often, bacterial genes with related functions are organized together in operons (Campos, 2001, Sassett, 2003). And a study by Fu and Fu-Liu, 2007 classified the mycobacterial genes into 4 major functional clusters and identified a group of regulatory genes linked with these functional clusters and regulating genes in the same cluster. Also, the sequence alignments comparison indicated 13 different isoprenyl diphosphate synthases which emerged from the same ancestor into 3 clusters. These isoprenyl diphosphates included farnesyl-, geranylgeranyl-, and hexaprenyl diphosphate synthases (Chen, 1994).

The gene *grcC1* is annotated as a prenyl diphosphate synthase and is found in the gene cluster with genes known to have roles in menaquinone biosynthesis (Sun, 2001; STRING functional analysis). In *Mtb* genomic organization region, it is annotated as heptaprenyl diphosphate synthase (heptaprenyl diphosphate is the precursor of the isoprenoid quinone menaquinone side chain) and located upstream of the *Mtb* genes encoding geranylgeranyl hydrogenase and transferase (Sun, 2001). Therefore, it is hypothesized that this gene encodes the protein that synthesize the prenyl diphosphate intermediate in menaquinone synthesis. However, a close homolog, GrcC2, has been reported by (Mann, 2011) to be a geranyl diphosphate synthase, which is not consistent with the known amino acid sequence classifications of prenyl diphosphate synthases. In addition, transposon mutagenesis suggested that GrcC1 is essential for bacterial survival but GrcC2 is not (Tuberculist; Mycobrowser; DeJesus, 2017). The genes *grcC1* and *grcC2* are the closest related genes among *Mtb* polyprenyl synthases (Nagel, 2018), not only by sharing the genetic

nomenclature but also by sharing more than 56% of the amino acid sequence identity. One of the aims in this study is to characterize GrcC1 and GrcC2 and test their essentiality for bacterium growth, and this will help finding new drug targets which then can be used for the search for the best inhibitors.

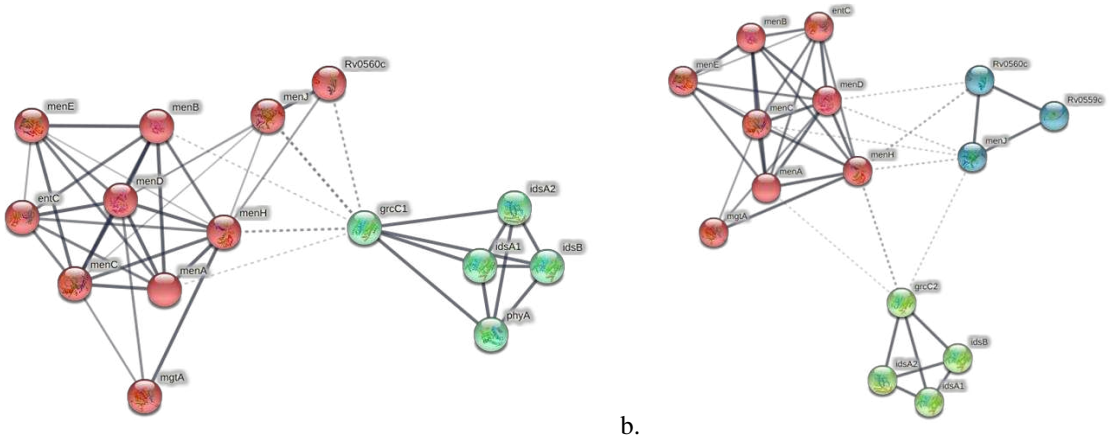


Figure 11: **Functional interaction networks of *Mtb* obtained from string-db.org (2021).** The functional network shows potential gene interaction links with *grcC1* (a) and *grcC2* (b) to functionally related menaquinone genes in the genome (*menC*, *menD*, *menH*, *menJ*) and also, with other known or predicted prenyl diphosphate synthases. The thicker lines mean the genes are more confidently associated "functionally". In these analysis, the query genes were *menC, D, H, J* and *grcC1* or *grcC2*. Settings used were full network, medium confidence, maximum interaction first shell 10, second shell none, MCL clustering parameter 3. *grcC1* enrichment p-value was 3.55e-15 while that for *grcC2* was 2.46e-12.

Many enzymes are involved in MK biosynthesis. Studies by Dhiman, et al (2009; 2019) suggested that the enzyme MenA is a target of a compound which inhibited *Mtb* growth. The gene *grcC1* is classified in the gene cluster zone under the category of cell wall, cell process and metabolism (Camacho, et al 1999) with the genes in MK pathway.This category includes membrane proteins and enzymes involved in intermediate metabolism which will identify novel drug targets and vaccines development. The same cluster of genes in *Mtb* genome are functionally related and likely regulate other genes in the same cluster (Wilson, 1999), these include genes that

are appointed to essential pathways as well as many of unknown function. According to those functional genes clustering, the genes *grcC1* and *grcC2* are located close to other genes that are annotated or/and confirmed to be involved in menaquinone synthesis, so *grcC1* and *grcC2* may also have a role in *Mtb*'s menaquinone biosynthesis pathway. Here we will identify and characterize genes in *Mtb* which are clustered with known genes of menaquinone pathway.

Identifying polyprenyl diphosphate synthases in *Mycobacterium smegmatis*

In the DNA database, where the genomic sequences of *M. smegmatis* and other gram-positive bacteria are registered, the genes encoding homologues to *Mtb* prenyltransferases (GrcC1 and GrcC2) were identified. By BLAST search of the inquiry genome sequence (homology search program), the analysis showed the presence of 2 homologues encoding putative prenyl diphosphate synthase (Figure 12). The gene *ms1133* showed 74% identity with *grcC1* and 63% with *grcC2*. The function of this gene annotated as unknown function.

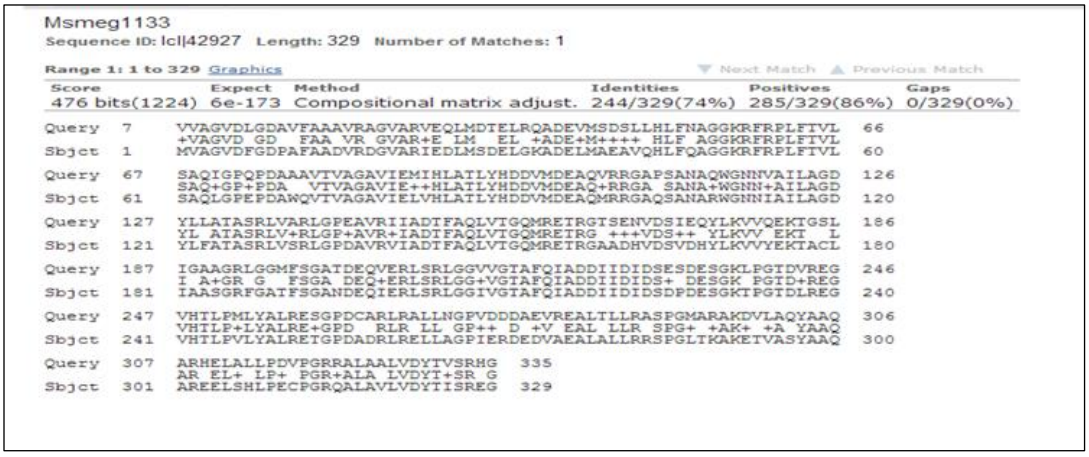


Figure 12: Amino acid sequence alignment of different long-chain prenyl diphosphate synthase: *M. tuberculosis* GrcC1 (query) with unknown homolog (Msmeg 1133) in *M. smegmatis.*

Looking at the importance of the isoprene long chain in MK structure and the importance of MK to *Mtb*, altogether mycobacteria use prenyl diphosphate synthases in many of its critical and essential biosynthetic pathways. They are also considered as a vital component to the *Mtb* growth and survival. This makes prenyl diphosphate synthases involved in MK synthesis a potential drug target and studying and identifying these enzymes in *Mtb* and its laboratory surrogate will expand the knowledge to support the drug design and the research of new targets.

Chapter 3

Identifying prenyl diphosphate synthases that could be involved in menaquinone biosynthesis: Cloning, expression, and biochemical characterization of GrcC1 of *Mtb* H37Rv.

GrcC1 (Rv0562) is a homolog to the medium-chain polyprenyl diphosphate synthases (Schulbach, 2000), and depending on the previous information it can be suggested to have a role in the menaquinone biosynthesis pathway, likely as an essential intermediate. The sequence alignment suggested that GrcC1 seems to be involved in the polyprenyl diphosphate synthesis as long chain prenyl diphosphate synthase (*Mtb* database). Previous research (Dhiman, 2009) has reported the ability of some inhibitor compounds to inhibit *Mtb*. For instance, these compounds inhibit MenA in the menaquinone biosynthesis pathway and thus inhibit bacterial growth. The *Mtb* gene *grcC1* is clustered by functional analysis (previously described in chapter 2) in the same gene cluster with *menA* and other genes involved in MK pathway, where they predicted to have related function. Genes included in this category are thus candidate new targets for these inhibitors. The proposed hypothesis is that the gene *grcC1* encodes the protein that synthesizes the polyprenyl diphosphate intermediate in the menaquinone synthesis pathway, which is an essential gene for *Mtb* growth.

Here we characterize the gene *grcC1* (polyprenyl diphosphate synthase annotated to be medium-chain prenyl diphosphate synthase involved in metabolic pathway) (MTB gene bank) to identify the final product of this enzyme and determine the isoprene chain length.

Cloning, expression and chemical characterization of the enzymatic activity of GrcC1 polyprenyl diphosphate synthases in H37Rv.

Bacterial strains, plasmids and materials, used in this study:

The bacterial strain *E.coli* DH5α (from Invitrogen) was used as the cloning host; and the BL21 strain was used as the expression host. PCR kit and the plasmid vector pGEM-T Easy Vector System I 3000bp (from Promega) was used. The expression vectors were pET28a 5369 bp, and pET29b 5371 bp (from Novagen). The pBlueScript (pBSK) 3000 bp was purchased from Stratagene. pUC4K 3914bp was purchased from Amersham.

The Expand High Fidelity PCR system and dNTPs were purchased from Roche Diagnostics and restriction endonucleases were purchased from New England Biolabs. FastDigest enzyme were from Fermentas life science. DNA ligase from BioLabs Inc. Potato acid phosphatase was from Sigma-Aldrich. The media for growing the bacterial strains was Lysogeny broth (LB) was purchased from Invitrogen. B-PER Bacterial protein extraction reagent was purchased from Thermo Science. Monoclonal anti-polyHistidine antibody produced in mouse, Anti-Mouse IgG (whole molecule) and alkaline phosphatase antibody produced in goat were purchased from Sigma-Aldrich. QIAprep Spin Miniprep (QIAGEN) and QIAquick Gel Extraction Microcentrifuge Kits were purchased from QIAGEN. PD-10 Desalting columns were purchased from GE Healthcare Bio-Sciences. Amicon Ultra-15 30K centrifugal filter devices were from Millipore. Geraniol (GOH), farnesol (FOH), and geranylgeraniol (GGOH) were purchased from SigmaAldrich.

Experimental Methods

Amplification of *M. tuberculosis* H37Rv (*Rv0562*) *grcC1*

To facilitate gene amplification, we used the PCR approach, which depends on annealing of two specific primers in a specific thermal profile cycle. The two primers were designed based on a homology search of the *M. tuberculosis* genome sequence, conducted with the webserver of the mycobacteria database (Tuberculist). The designed primers were: atata **cat atg** agg act ccg gcg acg gtg (forward) and atata **aag ct**t cta gcc gtg ccg gct cac ggt (reverse). The restriction sites for *NdeI* and *HindIII* (bold) were introduced to the forward and reverse primers and genomic DNA of *M. tuberculosis* was used as the template. The primers were synthesized and confirmed by Integrated DNA Technologies "XXIDT". Reaction mixes contained the enzyme *Taq* polymerase (1 unit), dNTPs (nucleotide mix 25%), DMSO (10%), and $MgCl_2$ buffer as well as the primers. The PCR reaction was performed on a Perkin Elmer PCR system 2400.

The amplified 1008 bp DNA fragment (PCR product) was extracted and gel purified using QIAquick Gel Extraction Kit following the manufacturer's protocol and digested with the restriction enzymes *NdeI* and *HindIII*. Both the PCR purified fragment and the ligation vector were digested with the restriction enzymes. The purified fragment was ligated into the pGEM-T Easy vector (Promega). The ligation standard reaction in the kit protocol was followed. The ligation reaction was used to transform Subcloning Efficiency $DH5_\alpha$ Competent Cells (Invitrogen) following the manufacturer's guidelines. The cells containing plasmid were selected on Lennox agar containing Ampicillin (100mg/ml), X-Gal (50 mg/ml) and 1 M IPTG (isopropyl-β-D-thiogalactopyranoside). Single colonies were picked and used to inoculate 5ml of Lennox/Ampicillin broth and the plasmid DNA was purified using QIAprep Spin Miniprep Kit

following the manufacturer's procedure. The purified plasmid DNA was digested using the restriction enzyme *SacI* to confirm the correct size of the product (~4000bp) using electrophoresis 1%TBE agarose gel. The pGEMT- GrcC1 construct was double digested with restriction enzymes *Hind III* and *NdeI* and subjected to electrophoresis on 1% agarose gel. The resulting 1000 bp fragment was purified and ligated into the expression vector pET28a+ and transformed into DH5α competent cells. The plasmids were purified using Qiagen mini-prep kit and subjected to restriction enzyme *Hind III* and *NdeI* and gel purified. The constructed pET28-GrcC1, was used to transform into *E.coli* BL21(DE3)pLysS cells. The empty vector (pET28) was used to transform cells as a negative control. The transformed cells were cultivated in 100 ml of LB broth supplemented with kanamycin (50µg/ml) overnight. For over-expression, the grown overnight transformed cells were transferred into 1 liter of LB broth and incubated with shaking at 37˚C to an optical density 600 (OD_{600}) of 0.5; and; the culture was induced by the addition of 0.5 mM of isopropyl β-D- thiogalactopyranoside (IPTG) and incubated for an additional 8 hours at 20˚C. The induced bacterial cells were harvested by centrifuging at 3000 rpm for 10 minutes, the pellets were lysed using B-PER bacterial protein extract solution and protein was estimated using the BCA kit. Western blot analysis and sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were used to confirm protein expression. Primary (anti-His antibodies) and secondary (polyclonal goat anti-mouse) monoclonal anti-polyhistidine antibodies (Sigma-Aldrich) were used for Western blot analysis.

Purification of the recombinant *grcC1*:

The harvested cells were disrupted by sonication on ice (12 cycles of 30 seconds) in 50 mM MOPS (pH 7.9) lysis buffer containing: 200 mM $NaCl_2$, 10 mM $MgCl_2$, protease inhibitor, 5 mM β-mercaptoethanol (BME), and 10% glycerol. The homogenized cell fractions were centrifuged at

15,000 xg for 15 minutes at 4˚C. The precipitate was suspended with buffer containing 50 mM MOPS, 200 mM NaCl, 1mM $MgCl_2$, 1mM BME, and 10% glycerol. Affinity purification of the recombinant protein was done using Bio-RAD poly-prep chromatography nickel columns and washed with constant increased concentrations of imidazole buffer. The protein was eluted in 1 ml volumes, with increasing imidazole concentrations (25, 50, 100, 150 and 200 mM). The level of protein purification was determined by Western blot and 10 % sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS- PAGE) on a series of different concentrations of imidazole. Western blot is a sensitive method that can detect purified proteins, probed and analyzed with monoclonal anti-polyhistidine tag produced in mouse. The purified protein was desalted using PD-10 desalting columns (from GE Healthcare Bio-sciences). The desalted protein (3.5ml) was concentrated using centrifugal filter units for concentration and purification of biological solutions (30,000 MWCO) Amicon Ultra15. BCA protein estimation kit was used to determine protein concentration. The protein quantified from a calibration curve created using BSA (bovine serum albumin), and the measurements done at 562 nm.

Protein activity assay for GrcC1 (Enzyme assay):

The activity of the resulting recombinant protein was determined by estimating the amount of radiolabeled [^{14}C]isopentyl pyrophosphate (IPP) incorporated into butanol-extracted polyprenyl diphosphate, which is an accurate and sensitive detection method. The 50 µl standard reaction assay mix contained 50 mM MOPS buffer pH 7.9, 2 mM $MgCl_2$, 2 % Triton X-100, 2 mM DTT, 35 µM [^{14}C]IPP, 50 µM allylic substrate [geranyl diphosphate salt (GPP)] and 35 ng purified protein. Starting with this reaction mix the optimal reaction conditions for the assays were identified. The activity of the recombinant protein was compared with the activity of known active crude protein fractions from *Msmeg*.

Enzymatic assays and reaction requirements:

Assay conditions required for optimal activity of the enzyme were standardized for protein concentration, time, divalent cations and pH. To determine the effect of protein concentration, the assay was conducted to adjust the optimal protein concentration in the previous reaction mix using 6 different protein concentrations in the mixes (50, 100, 200, 400, 800 and 1600 ng). The reaction mixes were incubated in a water bath at 37˚C for 30 minutes. The reaction was stopped by adding 1 ml of water saturated with NaCl. The [^{14}C]radiolabeled products were extracted with butanol saturated with water. Aliquots of the extracted product (100 µl) were used for liquid scintillation spectrometry.

Optimization of reaction time:

For reaction time optimization, the enzyme activity was tested using previous assay mix and incubated in water bath for time intervals (0, 10, 20, 30 and 50 minutes). The reaction was stopped by adding 1 ml of water saturated with NaCl. The [^{14}C]radiolabeled products were extracted with butanol saturated with water. Aliquots of the extracted product (100 µl) were used for liquid scintillation spectrometry.

Effect of metal divalent cation on the activity of GrcC1:

The rate of radioactivity incorporated into product was tested in the absence and presence of divalent cations. The recombinant protein was incubated for 15 minutes with Bio-Rex 70 cation exchange resin to remove endogenous cations. The incubated protein was tested in the previous reaction mix with and without divalent cation ($MgCl_2$) at 10 mM concentration, with a control that containing recombinant protein not incubated with Bio-Rex resin. To determine which divalent cation increases enzyme activity, the effect of different divalent cations was studied. The different cations, namely $CaCl_2$, $MgCl_2$, $MnCl_2$ and $ZnCl_2$, were added to the reaction mix at 2 mM

concentration and the radioactivity was determined by scintillation counter.

pH optimization:

The enzymatic assay was conducted to adjust the optimal reaction pH in the reaction mixes using 3 different buffers (MES, MOPS and Tris) with pH ranges from 5.0 to 9.0. The reaction mixes were incubated in a water bath at 37˚C for 30 minutes. The reaction was stopped by adding 1 ml of water saturated with NaCl. The [^{14}C]radiolabeled products were extracted with butanol saturated with water. Aliquots of the extracted product (100 µl) were used for liquid scintillation spectrometry.

Detergent effect:

The effect of detergents on GrcC1 activity was tested as well; the rate of radioactivity incorporated into product was analyzed in the absence and presence of detergent in the reaction mix. The enzymatic reaction was conducted as described previously with the addition of nonionic and zwitterionic detergents to test if they enhance the enzyme activity in the assay. The effect of the detergents Triton X-100, CHAPS and Tween-20 were tested at 0.1 % concentration.

Specificity of the allylic substrates:

By using the standardized reaction mix, the enzymatic activity of the recombinant protein was tested with different allylic substrates which are DMAPP, GPP, FPP, and GGPP. All four substrates were tested at the concentration of 50 µM, with 35 µM of IPP in the reaction mix. The different products of each allelic substrate were analyzed and identified by determining the radioactivity using scintillation counter. Also, the effect of different substrates concentrations on the enzyme activity was determined in broad range of substrates concentrations ranging from 0 µM to 250 µM.

The characterization of kinetic parameters:

The kinetic parameters for GrcC1 were determined under conditions which were linear for protein concentration and time, with different concentrations of the four substrates (GPP, DMAPP, FPP or GGPP) with constant concentration of [^{14}C]IPP and using Michaelis-Menten assumptions. The values of V_{max}, K_{cat} and K_m for each substrate were calculated. The data were analyzed by nonlinear regression using SigmaPlot 11.0.

The reaction product analysis:

The final products of the reaction assays with different allylic diphosphate substrates were characterized using thin layer chromatography (TLC) after the dephosphorylation of the synthesized radiolabeled compounds.The radiolabeled products (after the removal of butanol under nitrogen) were subjected to the dephosphorylation process prior to analysis of the product chain length. The enzyme potato acid phosphatase was used for the dephosphorylation of the final products. Starting with 100 µl aliquot of butanol extracted products; the butanol was removed under nitrogen stream and the radiolabeled products were dissolved in 5 ml of buffer (100 mM sodium acetate pH 4.8, 0.1% triton-X, 60% methanol) and bath sonicated for 2 minutes. Then, 20 units of the potato acid phosphatase were added to the mix and incubated overnight at 25˚C. The dephosphorylated products were extracted with 1 ml of n-hexane (3 times). The n-hexane pooled extracts were washed with water (1 ml) and the solvent evaporated under stream of nitrogen. The extracts were dissolved in 200 µl of chloroform: methanol (2:1, v/v) and aliquots were used for scintillation spectrometry (20 µl) and TLC analysis. The radiolabeled product aliquots were spotted on reversed-phase C_{18} TLC plates which were developed in methanol:acetone (8:2, v/v) and visualized by the Typhoon phosphoimager. Geraniol (GOH), farnesol (FOH), geranylgeraniol (GGOH) and solanesol were used as standard polyprenols for the TLC plate and these standards

were located and visualized using an anisaldehyde spray reagent (SigmaAldrich). All data presented is representatives of multiple experiments conducted in triplicate.

Results

Amplification of *M. tuberculosis* H37Rv (*Rv0562*) *grcC1*:

The PCR reaction was performed as previously described, and the amplified 1008 bp sized band of *grcC1* was extracted and gel purified (Figure 13) and ligated into the expression vectors as previously described. Western blot analysis and 10% SDS-PAGE were used to confirm protein expression and size. Primary (anti-His antibodies) and secondary (polyclonal goat anti-mouse) monoclonal anti-polyhistidine antibodies were used for Western blot analysis.

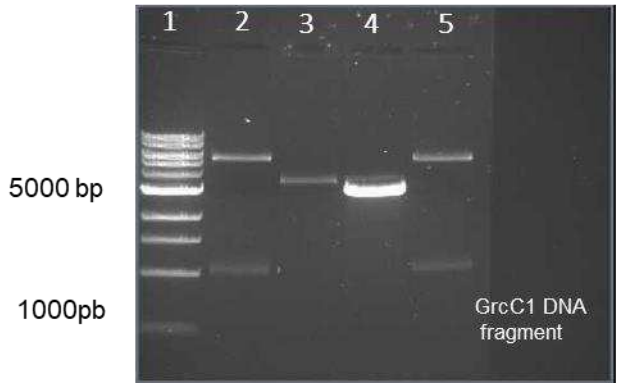
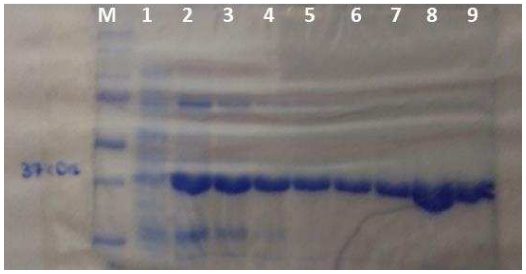


Figure 13: PCR amplified *grcC1* DNA fragment after double digestion using *NdeI* and *HindIII*. Reaction mixes contained the enzyme *Taq* polymerase (1 unit), dNTPs (nucleotide mix 25%), DMSO (10%), and $MgCl_2$ buffer as well as the primers. The PCR reaction was performed on a Perkin Elmer PCR system 2400. Lane 1: marker, lanes 2,3,4 &5 indicate 4 PCR reactions; lanes 2 and 5 show clones containing the *grcC1* gene.

The affinity purification of the recombinant protein GrcC1 was done using Bio-RAD poly-prep chromatography nickel columns and washed with constant increased concentrations of imidazole buffer. The protein was eluted with increasing imidazole concentrations (25, 50, 100, 150 and 200

mM). The level of protein purification was determined by Western blot and 10 % SDS- PAGE gel electrophoresis, (Figure 14 a, and b). The desalted protein (3.5ml) was concentrated and BCA protein estimation kit was used to determine protein concentration. The protein quantified from a calibration curve created using BSA (bovine serum albumin), and the measurements done at 562 nm.

a. 

b. 

Figure 14: a. Purification of *grcC1*:10% SDS-PAGE analysis of extracted showing the purification of GrcC1. The 10% SDS-PAGE were visualized using Coomassie Brilliant Blue G-250 stain. Lanes 1-6: 50 mM imidazole concentration. Lanes 7-9: 100 mM imidazole concentration. b. Western blot for purified protein (GrcC1): Western blot analysis of protein purified after transformation in *E.coli*. Lane 0, molecular weight marker, lane 1 and 2, uninduced protein. Lane 3 contains purified fractions of GrcC1 (Rv0562) with expected molecular weight of 37KDa.

Protein activity assay for GrcC1

The purified GrcC1 protein concentration was estimated to be ~ 300 µg/ml. The activity of the resulting recombinant GrcC1 was determined by estimating the amount of radiolabeled [^{14}C]IPP incorporated into butanol-extracted polyprenyl diphosphate. The 50 µl standard reaction assay mix contained 50 mM MOPS buffer pH 7.9, 2 mM $MgCl_2$, 2 % Triton X-100, 2 mM DTT, 35 µM [^{14}C]IPP, 50 µM allylic substrate [geranyl diphosphate salt (GPP)] and 35 ng purified protein. A known active crude protein fractions from *M. smegmatis* used as a control for this reaction. The purified recombinant protein was found to be active (Figure 15).

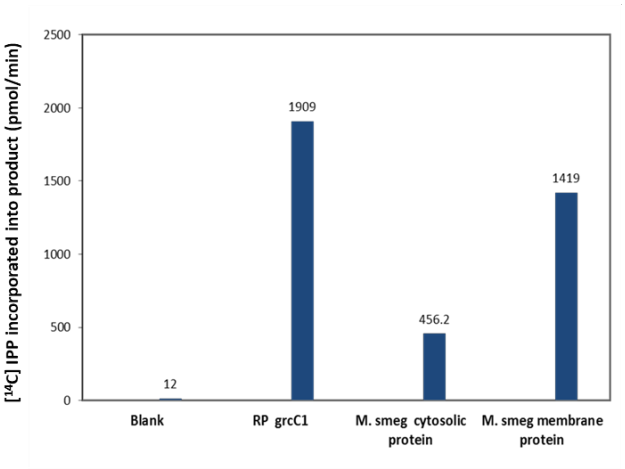


Figure 15: Recombinant protein GrcC1 (RP GrcC1) activity test. The purified protein was assayed in the reaction mix contained 50mM MOPS, 2 mM $MgCl_2$, 2mM DTT, 0.2% Triton X-100, 50 µM GPP, 35 ng protein and 35 µM [^{14}C]IPP and the activity was compared to activity from crude protein fractions from *M. smegmatis*.

Enzymatic assays and reaction requirements for GrcC1:

The purified enzyme was active and the assay conditions required for optimal activity of the enzyme were standardized for protein concentration, time, divalent cations and pH. Reaction mix contained: 50 mM MOPS, 2 mM $MgCl_2$, 2mM DTT, 0.2% Triton X-100, 50 µM GPP and 35 µM [^{14}C]IPP was used to standardize the enzymatic assay requirements.

The protein concentration in the reaction:

To determine the effect of protein concentration in the reaction, the assay was conducted as previously mentioned in Methods. Six different protein concentrations (50, 100, 200, 400, 800 and 1600 ng) were used (Figure 16). Depending on the analyzed [^{14}C]radiolabeled products, the maximum enzyme activity was at about 200 ng of protein concentration in the reaction. The reaction was inhibited at higher protein concentrations. The 100 ng/assay was chosen for further experiments (Figure 16).

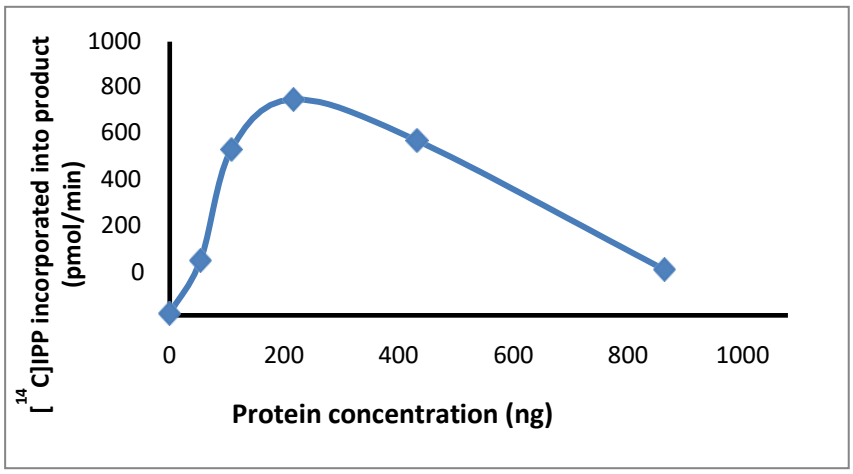


Figure 16: Protein concentration curve shows the effect of protein concentration on reaction activity. Protein concentrations in the reaction mixe were (50, 100, 200, 400, and 800 ng). The reaction was incubated at 37˚C for 30 minutes. The [^{14}C]radiolabeled products were extracted and analyzed using liquid scintillation spectrometry.

Optimization of reaction time:

For reaction time optimization, the enzyme activity was tested using previous assay mix and conditions, and incubated for time points of 0, 10, 20, 30 and 50 minutes. The reaction was linear for 30 minutes (Figure 17). Further experiments were conducted with 15 minutes incubation time.

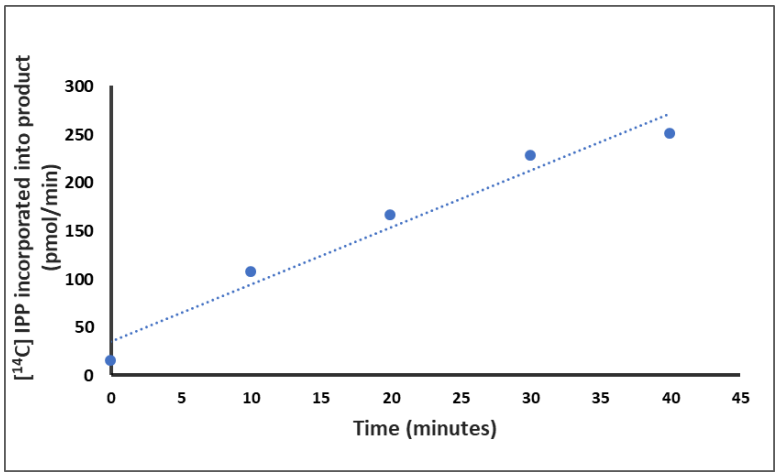


Figure 17: Reaction time optimization for GrcC1.The time range started at 0 minutes to 40 minutes. The enzyme activity was linear for approximately 30 minutes. The dotted line represents the best fit determined by linear regression analysis.

Effect of divalent cation on the activity of GrcC1:

Divalent cation requirement for the enzyme was tested in the absence and presence of the divalent cation $MgCl_2$ at 10 mM concentration. The protein was tested in the previous reaction mix with and without ($MgCl_2$), with a control that containing recombinant protein not incubated with Bio-Rex 70 resin. GrcC1 activity definitely depends on the present of divalent cations in the reaction (Figure 18). To determine which divalent cation supports enzyme activity optimally, the effect of different divalent cations was studied. The endogenous divalent cations were removed by using Bio-Rex 70 cation exchange resin mesh and the cation, $CaCl_2$, $MgCl_2$, $MnCl_2$ or $ZnCl_2$, was added to the reaction mix. The enzyme optimal activity was shown in the present of $MgCl_2$ at 2mM of concentration (Figures 19 & 20).

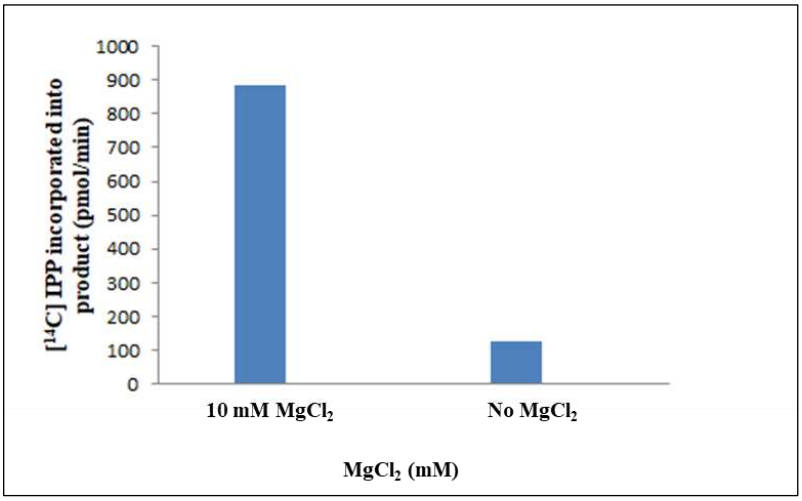


Figure 18: Requirement of divalent cation for the activity of GrcC1. The rate of incorporated [^{14}C]IPP into product was tested in the absence and presence of divalent cations. The recombinant protein was incubated for 15 minutes with Bio-Rex 70 cation exchange resin. The incubated protein was tested in the reaction mix with and without divalent cation ($MgCl_2$) and, with control containing recombinant protein not incubated with Bio-Rex resin.

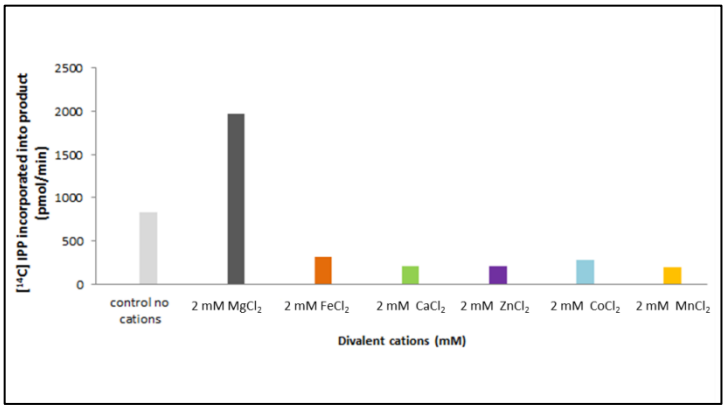


Figure 19: The enzyme GrcC1 requires the divalent cations to support activity and $MgCl_2$ is the preferred cation. The effect of different divalent cations was studied. The endogenous divalent cations were removed with Bio-Rex 70 cation exchange mesh, and the cations, $CaCl_2$, $MgCl_2$, $MnCl_2$ and $ZnCl_2$, were added to the reaction mix at 2 mM concentration. The enzyme optimal activity was shown in the presence of $MgCl_2$.

Since GrcC1 activity requires the presence of divalent cations and $MgCl_2$ was the preferred cation. The effect of different concentrations of $MgCl_2$ on the protein activity was analyzed. To optimize the Mg^{++} concentration in the reaction, concentration range of 0 mM to 10 mM were tested and 2 mM is optimal and was used in subsequent assays (Figure 20).

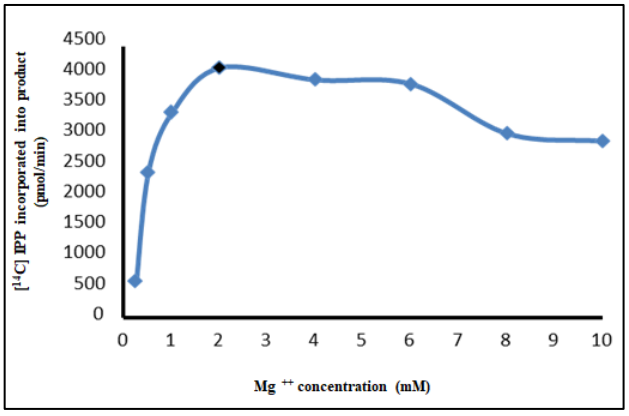


Figure 20: Optimization of Mg^{++} concentration for GrcC1 activity. 2 mM Mg^{++} concentration gives the highest reaction activity.

pH optimization:

The enzymatic assay was conducted to adjust the reaction pH in the reaction mixes using 3 different buffers (MES, MOPS and Tris) with pH ranges start from 5.0 to 9.0. The optimal pH for GrcC1 activity was determined between 6.0 and 6.5 (Figure 21).

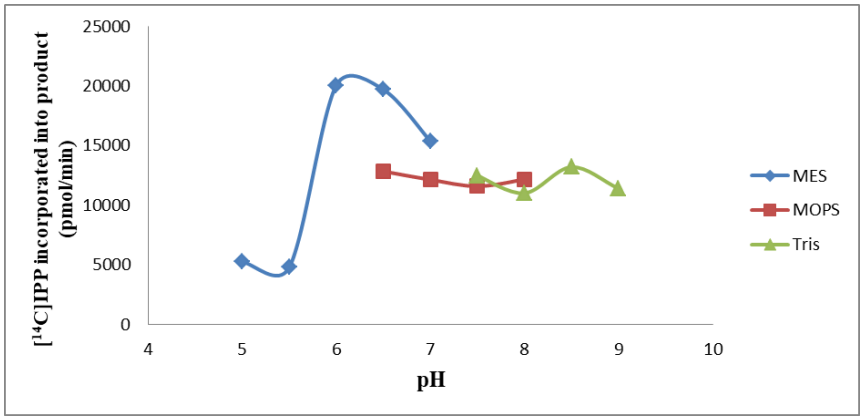
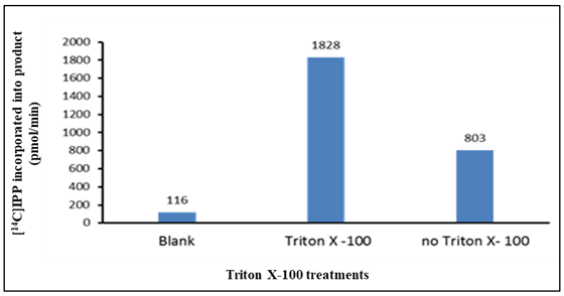


Figure 21: Effect of pH on GrcC1 activity. The enzymatic activity was tested in three different buffering systems: MES, MOPS, and Tris with pH ranged from 5.0 to 9.0. The optimum pH for the reaction is 6.5 (MES buffer) which was used for the further assays.

Detergent effect:

To test the effect of detergents on GrcC1 activity, the rate of radioactivity incorporated into product was analyzed in the absence and presence of 0.1 % detergent (Triton X-100). The presence of the detergent in the reaction mix enhanced the enzyme activity approximately by 40% compared with no detergent (Figure 22a.). The effect of different detergents including Triton X-100, CHAPS and Tween-20 was also tested at 0.1% concentration. GrcC1 showed an increase in its enzymatic activity in the presence of detergents and optimal activity was shown in the presence of Triton X-100 (Figure 22b.). The final concentration of Triton X-100 in the reaction was estimated by testing different concentrations starting from 0.1 to 0.8 %. The optimal concentration of Triton X-100 in the reaction was 0.2% (Figure 23), which was used in further assays.

a.



b.

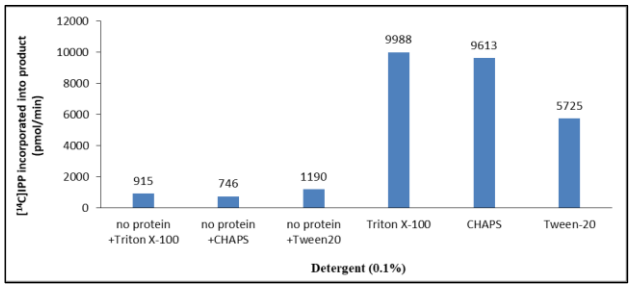


Figure 22: The effect of (0.1%) detergents on GrcC1 activity. a. The rate of incorporated [^{14}C]IPP into product was tested in the absence and presence of detergents (Triton X-100). b. shows the effect of the detergents: Triton X-100, CHAPS and Tween-20 on the enzymatic activity, which increased in the present of detergents and optimal activity was shown in the present of Triton X-100.

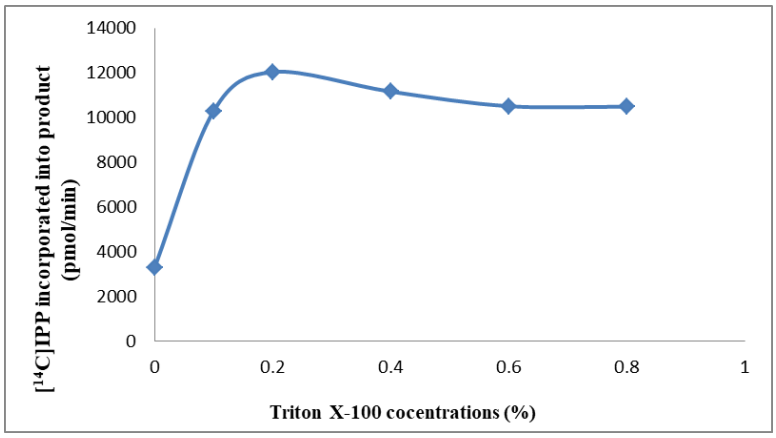


Figure 23: The effect of different concentrations of Triton X-100 on protein activity.The concentration of Triton X-100 in the reaction was optimized by adding the concentrations 0.1, 0.2, 0.4, 0.6 and 0.8 %. The optimal concentration of Triton X-100 in the reaction was 0.2%.

The results of the optimization of the reaction contents showed that the enzyme GrcC1 demonstrated its optimal activity in 50 µl reaction total volume and the assay mix contained 50 mM MES buffer pH 6.5, 2 mM $MgCl_2$, 0.2 % Triton X-100, 100 ng purified protein, 2 mM DTT, 35 µM [^{14}C]IPP, 50 µM allylic substrate [geranyl diphosphate salt (GPP)]. The optimized assay conditions were used to specify the enzyme substrates, characterize the enzyme kinetic parameters and final product.

Specificity of the allylic substrates:

Using the optimized reaction mix, the enzymatic activity of the recombinant protein was tested with the allylic substrates: DMAPP, GPP, FPP, and GGPP. The concentration for each substrate in the reaction was 50 µM, and 35 µM of IPP. The different products of each substrate were analyzed (Figure 24). The enzyme demonstrated high activity with the substrates GPP and FPP and the lowest activity using DMAPP. GPP is likely to be the natural substrate of this enzyme (Figure 24). The effect of substrates different concentrations on the enzyme activity was determined for the four substrates DMAPP, GPP, FPP, and GGPP in varying concentrations ranging from 0 µM to 20 µM (Figure 25). In general, GrcC1 enzymatic reaction was linear at the low ranges of substrate concentrations (~35 – 50 µM); the enzyme maximum activity was shown at the higher substrate concentration (100 – 150 µM). Also, the effect of different concentrations of IPP on the enzyme activity was tested at the concentration range from 0 to 100 µM and the enzyme showed optimal activity at the concentration 20-30 µM of IPP.

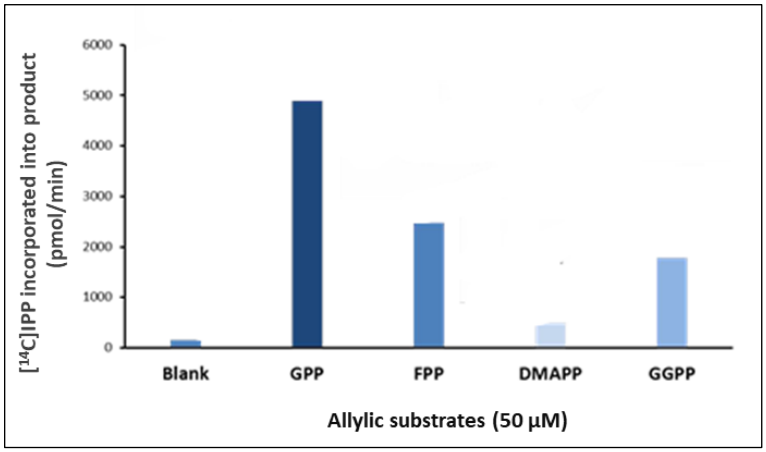


Figure 24: Activity of GrcC1 measured in the presence of the indicated allylic substrates as described in Methods section. The optimized reaction mix contains: 50 mM MES buffer pH 6.5, 2 mM $MgCl_2$, 0.2 % Triton X-100, 100 ng purified protein, 2 mM DTT, 35 µM [^{14}C]IPP, 50 µM allylic substrate was used. GrcC1 showed its maximum activity in the presence of the substrates GPP and FPP, and GPP is likely to be the preferred substrate of this enzyme. Blank bar contains no protein added to the reaction.

The characterization of kinetic parameters of GrcC1:

The optimized assays were used to determine the enzyme kinetic parameters. The values of K_m, V_{max} and K_{cat} were calculated using non-linear regression (SigmaPlot 11). The results are shown in Table (1). To reflect the enzyme's affinity to the used substrates, we calculated the K_m for GrcC1with different substrates. The low K_m value indicates that the enzyme has greater affinity (0.4 µM) for the longer chain substrates FPP and GGPP (see Table 1).

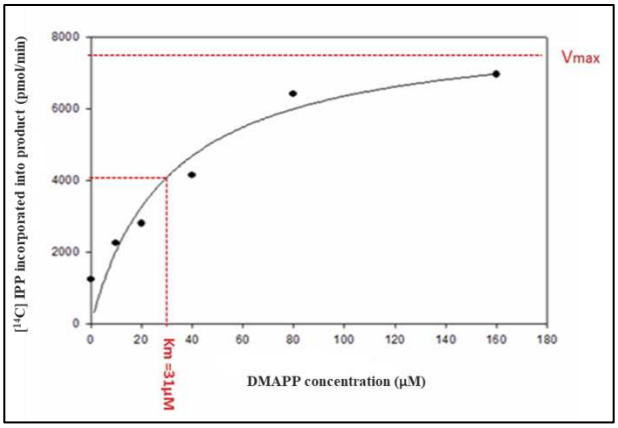


Figure 25: A representative activity curve of the effect of allylic substrate (DMAPP) varying concentrations on GrcC1 activity. Similar curves were obtained for the substrates GPP, FPP and GGPP. These data were used to calculate the kinetic parameters of GrcC1; V_{max} , K_m and K_{cat} values shown in table (1) using Michaellis-Menten assumptions.

Table 1: Kinetic parameters of GrcC1. The kinetic parameters were determined using the assay conditions as described for Figure (24). The data was analyzed using nonlinear regression (Sigma Plot 11.0); **V_{max}** represents the maximum rate achieved by the system, at maximum (saturating) substrate concentrations. **K_m** indicates the enzyme affinity for substrate. **K_{cat}** value gives direct measure of the catalytic production of product. K_{cat}/K_m measures enzyme efficiency.

	Substrates			
	DMAPP (C5)	**GPP (C10)**	**FPP (C15)**	**GGPP (C20)**
V_{max} (pmol/min)	89	12000	140	54
K_m (µM)	31	46	0.4	0.4
K_{cat} (1/min)	0.01	1.0	0.01	0.01
K_{cat} / K_m	0.0003	0.02	0.03	0.03

Final product analysis:

The chain length was determined by detecting the dephosphorylated radiolabeled compounds on reverse phase TLC plate. Multiple products were observed under the standardized assay conditions (C15, C20, C30, and C45). The results showed that GrcC1 catalyzes the enzyme reaction that produces multiple products with different chain lengths, the main products were GGPP (C20) and SPP (solanesyl diphosphate) (C45) in the presence of the substrate GPP. Also with the substrates GPP and DMAPP the final products were mainly GGPP and SPP, but with the presence of GGPP as a substrate, the final product was SPP (C45) (Figure 26). The enzyme synthesizes a long polyprenyl chain (length C45) which is needed for MK synthesis in mycobacteria.

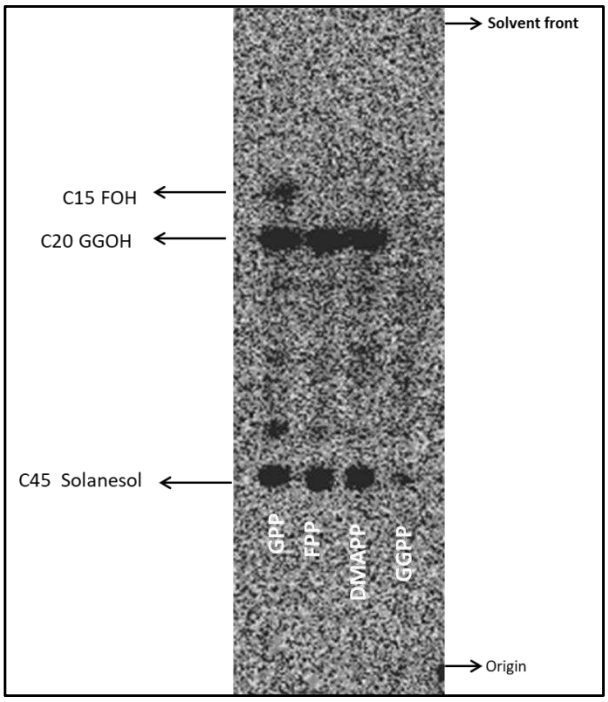


Figure 26: Reverse-phase TLC of reaction products of GrcC1. Equal amounts of radioactivity were spotted on reverse-phase TLC plates and the plates were developed in methanol:acetone (8:2, v/v). Products labeled with [^{14}C]IPP were dephosphorylated for TLC analysis and visualized using a Amersham Typhoon trio phosphoimager. Standard polyprenols were located with an anisaldehyde spray reagent. Migration of non-radioactive standards (GGOH, geranylgeraniol: FOH, farnesol) are indicated with arrows (compared to standards on the same plate).

Chapter 4

Identifying prenyl diphosphate synthases that could be involved in MK biosynthesis: Cloning, expression, and biochemical characterization of GrcC2 of *Mtb* H37Rv.

The gene *grcC2* (Rv0989) in *M. tuberculosis* genome encodes a protein, reported to be a polyprenyl diphosphate synthase, also found in the gene cluster under the category of cell process and metabolism. It is a close homolog of GrcC1, but little is known about this polyprenyl diphosphate synthase. Although the GrcC2 amino acid sequence is consistent with a long-chain prenyl diphosphate synthase, it has been reported to be a geranyldiphosphate synthase (Mann 2011; Mann 2012). The library of transposon mutations of *Mtb* suggested that *grcC2* is non-essential (Tuberculist). Therefore GrcC2 is hypothesized to be a geranyldiphosphate synthase and it may be non-essential for *Mtb* growth. Here we characterized the protein encoded in the gene *grcC2* and demonstrate that it is a long chain polyprenyl diphosphate synthase as predicted by its amino acid sequence.

Cloning, expression and chemical characterization of the enzymatic activity of GrcC2 polyprenyl diphosphate synthases in H37Rv.

Bacterial strains, plasmids and materials, used in this study:

The bacterial strains used as cloning, expression hosts, plasmid vectors, PCR system and all materials were same as mentioned in Chapter 3. The expression vector pET28a was replaced with PVV2 (supplied from Jackson's lab at CSU).

Experimental methods

Amplification of *M. tuberculosis* H37Rv (Rv0989) *grcC2*

To facilitate gene amplification, we used the same methods that described in Chapter 3. For PCR reaction, the two primers were designed based on a homology search of the *Mtb* genome sequence. The primers were: atata **catatg**atcccggcagtcagcctg (forward) and atata**aagctt**ctagtcgcacgcagatatcgc (reverse). The restriction sites for *NdeI* and *HindIII* (bold) were introduced to the forward and reverse primers and genomic DNA of *Mtb* was used as the template.

The amplified 978 bp DNA fragment (PCR product) was extracted and gel purified using the same protocols described in chapter 3, the purified fragment was ligated into the pGEM-T Easy vector (Promega). The pGEMT- *grcC2* construct was double digested with restriction enzymes Hind III and NdeI and subjected to electrophoresis on 1% agarose gels. For fragment purification, ligation and expression, the general procedure was described in chapter 3. Rv0989 initially was cloned in the expression vector pET28a+ and transformed into DH5α competent cells but later the PCR product was cloned into the mycobacterial expression vector (PVV2) to be expressed in

competent *M. smegmatis*. Western blot analysis and sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were also used to confirm protein expression.

Purification of the recombinant GrcC2:

The harvested cells were disrupted, and protein purified as described in Chapter 3. As well, the purified protein was desalted, concentrated and quantified using the same methods and procedures.

Protein activity assay for GrcC2 (Enzyme assay):

The activity of the resulting recombinant protein was determined by estimating the amount of radiolabeled [^{14}C]IPP incorporated into butanol-extracted polyprenyl diphosphate, following the same procedure that described in Chapter 3. The activity of the recombinant protein was compared with the activity of known active purified proteins GrcC1.

Enzymatic assays and reaction requirements:

Assay conditions required for optimal activity of the enzyme were standardized for protein concentration, time, divalent cations and pH. To determine the protein concentration, the assay was conducted to adjust the optimal protein concentration in the previous reaction mix using 6 different protein concentrations in the mixes (50, 100, 200, 400, 800 and 1600 ng). The reaction mixes were incubated in a water bath at 37˚C for 30 minutes. The reaction was stopped by adding 1 ml of water saturated with NaCl. The [^{14}C]radiolabeled products were extracted with butanol saturated with water. Aliquots of the extracted product (100 µl) were used for liquid scintillation spectrometry.

Optimization of reaction time:

For reaction time optimization, the enzyme activity was tested using previous assay mix and incubated in the water bath for 0, 10, 20, 30 and 50 minutes. The reaction was stopped by adding 1 ml of water saturated with NaCl. The [^{14}C]radiolabeled products were extracted with butanol saturated with water. Aliquots of the extracted product (100 µl) were used for liquid scintillation spectrometry.

Effect of divalent cation on the activity of GrcC2:

The rate of radioactivity incorporated into product was tested in the absence and presence of divalent cations (as described in Chapter3). Also to determine which divalent cation increases enzyme activity, the effect of different divalent cations was studied. The endogenous divalent cations were removed and cations, namely $CaCl_2$, $MgCl_2$, $MnCl_2$ and $ZnCl_2$, were added to the reaction mix at 2 mM concentration and the radioactivity incorporated into product was determined.

pH optimization:

The enzymatic assay was conducted to adjust the optimal reaction pH in the reaction mixes using 3 different buffers (MES, MOPS and Tris) with pH ranges start from 5.0 to 9.0. The reaction mixes were incubated in a water bath at 37˚C for 30 minutes. The reaction was stopped by adding 1 ml of water saturated with NaCl. The [^{14}C]radiolabeled products were extracted with butanol saturated with water. Aliquots of the extracted product (100 µl) were used for liquid scintillation spectrometry.

Detergent effect:

The effect of detergents on GrcC2 activity was tested as well; the rate of radioactivity incorporated into product was analyzed in the absence and presence of detergent in the reaction mix. The enzymatic reaction was conducted as described previously with the addition of nonionic and zwitterionic detergents to test if they enhance the enzyme activity in the assay. The effect of the detergents Triton X-100, CHAPS and Tween-20 were initially tested at 0.1 % concentration.

Specificity of the allylic substrates

Using the standardized reaction mix, the enzymatic activity of the recombinant protein was tested with different allylic substrates including, DMAPP, GPP, FPP, and GGPP. All four substrates were tested at the concentration of 50 µM, and 35 µM of IPP in the reaction mix.The products of each allelic substrate were analyzed and identified by determining the radioactivity using scintillation counter. Also, the effect of substrates different concentrations on the enzyme activity was determined in broad range of substrates concentrations ranging from 0 µM to 250 µM.

The characterization of kinetic parameters:

The kinetic parameters for GrcC2 were determined under conditions which were linear for protein concentration and time, with different concentrations of the four substrates (GPP, DMAPP, FPP or GGPP) with constant concentration of [^{14}C]IPP and using Michaelis-Menten assumptions. The values of V_{max}, K_{cat} and K_m for each substrate were calculated. The data were analyzed by nonlinear regression using the SigmaPlot 11.0.

The reaction product analysis:

The final products of the reaction assays with different allylic diphosphate substrates were characterized using the thin layer chromatography (TLC) after the dephosphorylation of the synthesized radiolabeled compounds. The radiolabeled products (after the removal of butanol under nitrogen) were subjected to the dephosphorylation process to analyze the product chain length. The enzyme potato acid phosphatase was used for the dephosphorylation of the final products. Starting with 100 µl aliquot of butanol extracted products; the butanol was removed under nitrogen stream and the radiolabeled products were dissolved in 5 ml of buffer (100 mM sodium acetate pH 4.8, 0.1% triton X-100, 60% methanol) and bath sonicated for 2 minutes. Then, 20 units of the potato acid phosphatase were added to the mix and incubated overnight at 25˚C. The dephosphorylated products were extracted with 1 ml of n-hexane (3 times). The n-hexane pooled extracts were washed with water (1 ml) and the solvent evaporated under stream of nitrogen. The extracts were dissolved in 200 µl of chloroform : methanol (2:1, v/v) and aliquots were used for scintillation spectrometry (20 µl) and TLC analysis. The radiolabeled product aliquots were spotted on reverse-phase C18 TLC plates which were developed in methanol:acetone (8:2, v/v) and visualized by the phosphoimager (Bioscan system imaging scanner). Geraniol (GOH), farnesol (FOH), geranylgeraniol (GGOH) and long-chain polyprenols were used as standard polyprenols for the TLC plate and these standards were located and visualized using an anisaldehyde spray reagent.

Also, since GrcC2 was previously described as GPP synthase (Mann 2011), we also analyzed the final products following the assay conditions that were used in the study reported by Mann, et al 2011 and 2012, with two assays as described above at the same time and conditions using GrcC2 proteins that expressed in Peters lab and in Crick lab and GrcC1 was used as a control. The assay

that Peters's group used to characterize the purified enzyme contained: 50 mM sodium phosphate buffer pH 7, 10 % glycerol, 5 mM $MgCl_2$, 2 mM DTT, 20 μM IPP, and DMAPP, GPP or FPP. The reaction final volume = 1 ml. Assays were stopped and dephosphorylated by adding 9:1 methanol: 3N hydrochloric acid. The final products were detected using TLC analysis as described above. All data presented is representatives of multiple experiments conducted in triplicate.

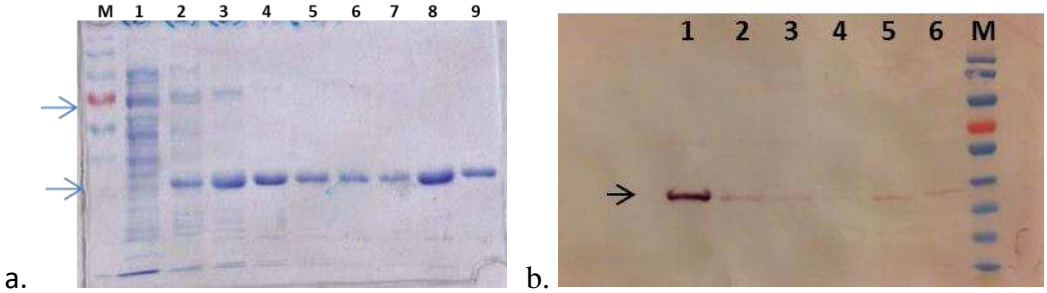
Results

Amplification of *M. tuberculosis* H37Rv (Rv0989) *grcC2*:

The PCR reaction was performed as previously described, and the amplified 978 bp sized band of *grcC2* was extracted and gel purified (Figure 27) and ligated into the expression vectors as previously described. Western blot analysis and 10% SDS-PAGE were used to confirm protein expression and size. Primary (anti-His antibodies) and secondary (polyclonal goat anti-mouse) monoclonal anti-polyhistidine antibodies were used for Western blot analysis.

Figure 27: PCR amplified GrcC2 DNA fragment after double digestion using *NdeI* and *HindIII*. Reaction mixes contained the enzyme *Taq* polymerase (1 unit), dNTPs (nucleotide mix 25%), DMSO (10%), and $MgCl_2$ buffer as well as the primers. The PCR reaction was performed on a Perkin Elmer PCR system 2400. Lanes: 1,2,3,4,5 and 6 indicate 6 PCR reactions; lane 4 shows the confirmed clone.

The gene *grcC2* (Rv0989) initially was cloned in the expression vector pET28a+ and used to transform in DH5α competent cells but the resulting purified protein was inactive; then the gene was cloned into the mycobacterial expression vector (PVV2) and used to transform competent *M. smegmatis*. Western blot analysis and sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were also used to confirm protein expression.The affinity purification of the recombinant protein GrcC2 was done using Bio-RAD poly-prep chromatography nickel columns washed with constant increased concentrations of imidazole buffer. The protein was eluted with increasing imidazole concentrations (25, 50, 100, 150 and 200 mM). The level of protein purification was determined by Western blot and 10 % sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS- PAGE), (figure 28 a, and b). The desalted protein was concentrated and BCA protein estimation kit was used to determine protein concentration.



a. b.

Figure 28: a. Purification of *grcC2*:10% SDS-PAGE analysis of extracted protein showing the purification of GrcC2 at the imidazole concentrations of 50 and 100 mM. The 10% SDS-PAGE were visualized using Coomassie Brilliant Blue G-250 stain. Lanes: 1- 6 show the 50 mM imidazole concentration, and lanes: 7-9 show the 100 mM. b. Western blot of purified protein (GrcC2): Western blot analysis of protein purified after transformation in *E. coli*. Lanes 1, 2, 3, 4, 5 and 6 contain purified fractions of GrcC2 with expected molecular weight of ~37KDa, M: molecular weight marker.

Protein activity assay for GrcC2

The purified GrcC2 protein concentration was estimated to be ~ 400µg/ml. The activity of the resulting recombinant GrcC2 was determined by estimating the amount of radiolabeled [^{14}C]IPP incorporated into butanol-extracted polyprenyl diphosphate. The 50 µl standard reaction assay mix contained 50 mM MOPS buffer pH 7.9, 2 mM $MgCl_2$, 2 % Triton X-100, 2 mM DTT, 35 µM [^{14}C]IPP, 50 µM allylic substrate [geranyl diphosphate salt (GPP)] and 35 ng purified protein. A known active purified protein, GrcC1, was used as a control for this reaction. The purified recombinant protein was found to be active (Figure 29).

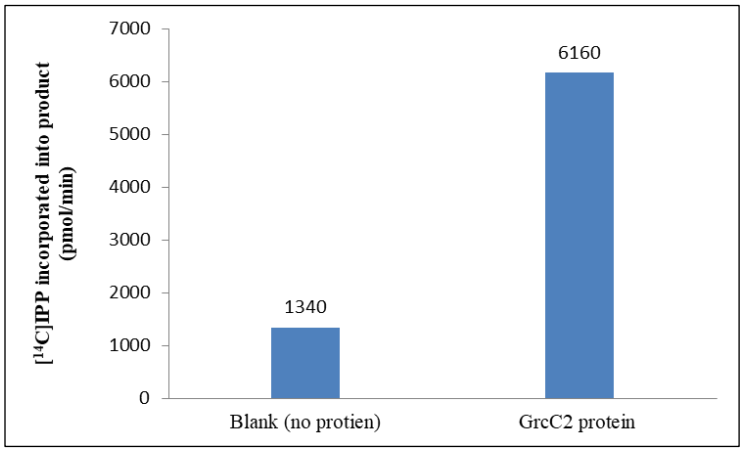


Figure 29: Recombinant protein GrcC2 activity test. The purified protein assayed in the reaction mix contained 50 mM MOPS, 2 mM $MgCl_2$, 2mM DTT, 0.2% Triton X-100, 50 µM GPP, 35 ng protein and 35 µM [^{14}C]IPP.

Enzymatic assays and reaction requirements for GrcC2:

The purified enzyme was active and the assay conditions required for optimal activity of the enzyme were standardized for protein concentration, time, divalent cations and pH. Reaction mix contained: 50 mM MOPS, 2 mM $MgCl_2$, 2mM DTT, 0.2% Triton X-100, 50 µM GPP and 35 µM [^{14}C]IPP was used to standardize the enzymatic assay requirements. To determine the

protein concentration in the reaction, the assay was conducted as previously mentioned in methods part. Six different protein concentrations (50, 100, 200, 400, 800 and 1800 ng) were used. Depending on the analyzed [^{14}C]-radiolabeled products, the maximum enzyme activity was at about 800 ng of protein concentration in the reaction.The concentration 400 ng was chosen for further assays (Figure 30).

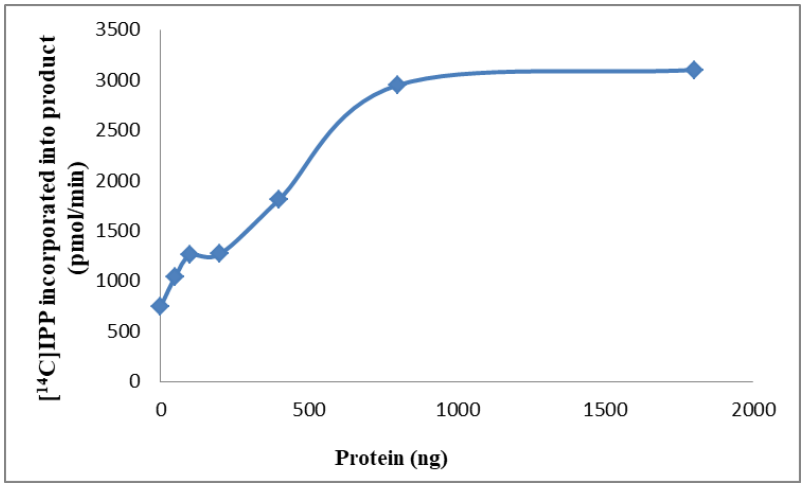


Figure 30: GrcC2 protein concentration curve shows the effect of protein concentration on reaction activity. Protein concentrations in the reaction mixes were (50, 100, 200, 400, 800 and 1800 ng). The reaction was incubated at 37˚C for 30 minutes. The [^{14}C]radiolabeled products were extracted and analyzed using liquid scintillation spectrometry.

Optimization of reaction time:

The enzyme activity was tested using previous assay mix and conditions, and incubated for time points of 0, 10, 20, 30 and 40 minutes. The reaction was linear for 20 minutes (Figure 31). Further experiments were conducted with 15 minutes incubation time.

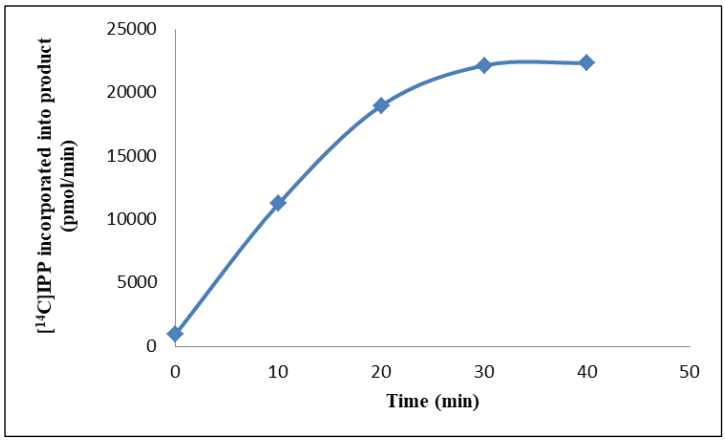


Figure 31: Reaction time optimization for GrcC2. The time range started at 0 minutes to 40 minutes. The enzyme activity was linear for approximately 20 minutes.

Effect of divalent cation on the activity of GrcC2:

Divalent cation requirement for the enzyme was tested in the absence or presence of divalent cation $MgCl_2$ at 10 mM concentration. The protein was tested in the previous reaction mix with and without ($MgCl_2$), with a control that containe recombinant protein not incubated with Bio-Rex 70 resin. GrcC2 activity definitely depends on the present of divalent cations in the reaction (Figure 32). To determine which divalent cation increases enzyme activity, the effects of different divalent cations were studied. The endogenous divalent cations were removed using Bio-Rex 70 cation exchange resin mesh and the cations, $CaCl_2$, $MgCl_2$, $MnCl_2$ or $ZnCl_2$, were added to the reaction mix. The enzyme optimal activity was shown in the presence of $MgCl_2$ at 2 mM of concentration (Figures 32 & 33).

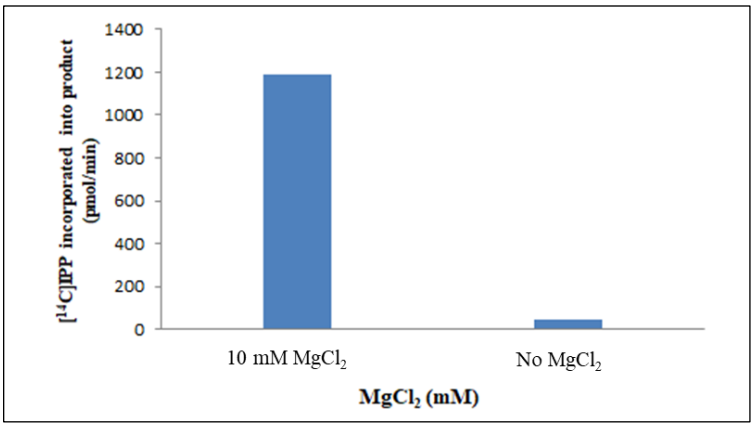


Figure 32: Requirement of divalent cation for the activity of GrcC2. The rate of incorporated [^{14}C]IPP into product was tested in the absence and presence of $MgCl_2$. The recombinant protein was incubated for 15 minutes with Bio-Rex 70 cation exchange resin.

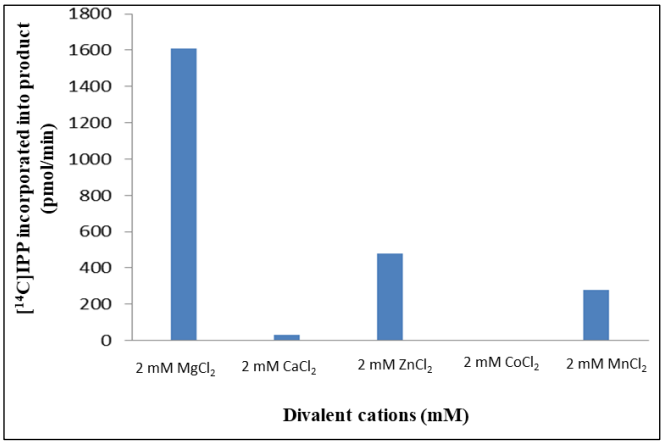


Figure 33: The enzyme GrcC2 requires the divalent cations to support activity and $MgCl_2$ is the preferred cation. The effect of different divalent cations was tested. The endogenous divalent cations were removed by using Bio-Rex 70 cation exchange resin mesh and the cations, $CaCl_2$, $MgCl_2$, $MnCl_2$ or $ZnCl_2$, was added to the reaction mix at 2 mM concentration. The enzyme optimal activity was shown in the present of $MgCl_2$.

Since GrcC2 activity requires the presence of divalent cations and $MgCl_2$ was the preferred cation. The effect of different concentrations of $MgCl_2$ on the protein activity was analyzed. To optimize the Mg^{++} concentration in the reaction, concentration range of 0 mM to 10 mM were tested and 2 mM is optimal and was used in subsequent assays (Figure 34).

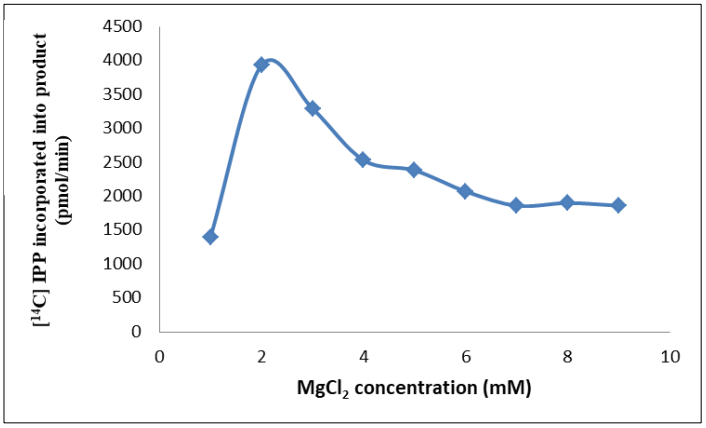


Figure 34: Effect of Mg^{++} concentration on GrcC2 optimal activity. 2mM Mg^{++} concentration gives the highest reaction activity.

pH optimization:

The pH in the reaction mixes was adjusted using 3 different buffers (MES, MOPS and Tris) with pH ranges start from 5.0 to 9.0. The optimal pH for GrcC2 activity was determined between 7.5 and 8.5 (Figure 35).

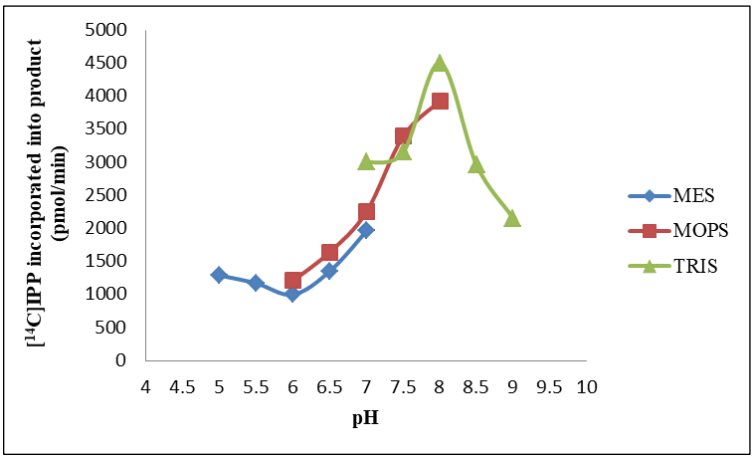


Figure 35: Effect of pH on GrcC2 activity. The enzymatic activity was tested in three different buffering systems: MES, MOPS, and Tris, with pH ranged from 5.0 to ~9.0. The optimum pH for the reaction is 8 (Tris buffer) which was used for the further assays.

Detergent effect:

The effects of different detergents including Triton X-100, CHAPS and Tween-20 were tested at 0.2% concentration. GrcC2 showed an increase in enzymatic activity in the presence of CHAPS (Figure 36). Also, the final concentration of CHAPS in the reaction was estimated by testing different concentrations starting from 0.1 to 0.8 %. The optimal concentration of CHAPS in the reaction was 0.2% (Figure 37), which was used in the further assays.

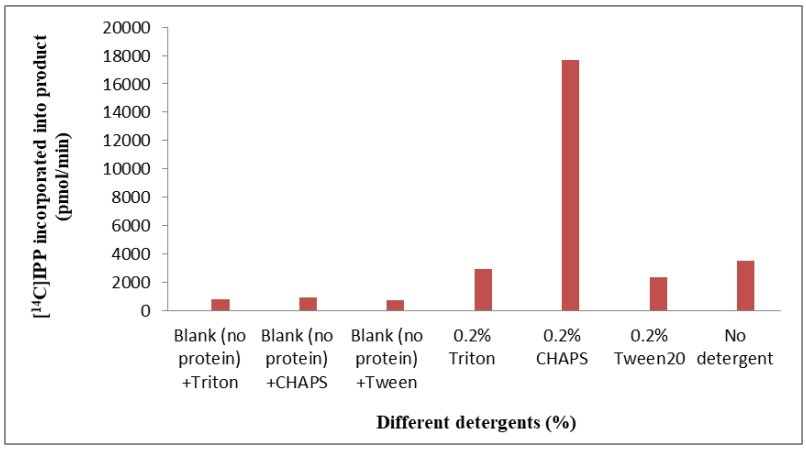


Figure 36: The effect of (0.2%) detergents on GrcC2 protein activity. The rate of incorporated [^{14}C]IPP into product was tested in the presence of the detergents: Triton X-100, CHAPS and Tween-20, which increased in the present of detergents and optimal activity was shown in the present of CHAPS.

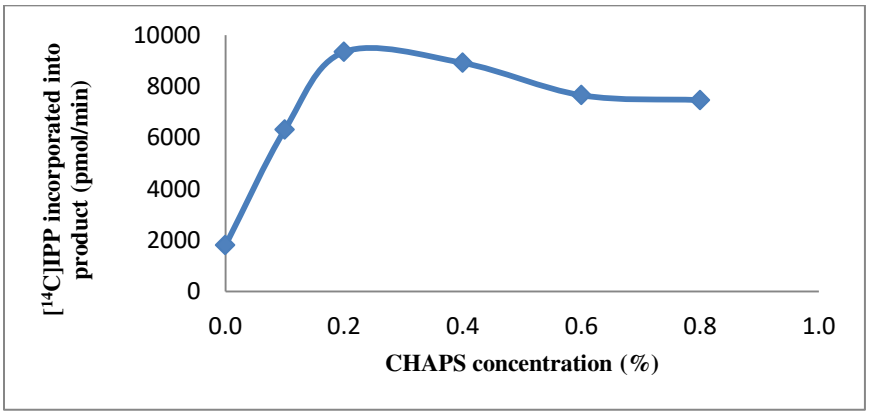


Figure 37: The effect of different concentrations of CHAPS on protein activity. The concentration of CHAPS in the reaction was optimized by adding the concentrations 0.1, 0.2, 0.4, 0.6 and 0.8 %. The optimal concentration of CHAPS in the reaction was 0.2%.

The results of the optimization of the reaction contents showed that the enzyme GrcC2 demonstrated its optimal activity in 50 µl reaction total volume and the assay mix contained 50 mM Tris buffer pH 8, 2 mM $MgCl_2$, 0.2 % CHAPS, 400 ng purified protein, 2 mM DTT, 35 µM [^{14}C]IPP, 50 µM allylic substrate [geranyl diphosphate salt (GPP)] for 15 minutes incubation time. The optimized assay conditions were used to specify the enzyme substrates, characterize the enzyme kinetic parameters and final product.

Specificity of the allylic substrates

Using the optimized reaction mix, the enzymatic activity of the recombinant protein was tested with the allylic substrates: DMAPP, GPP, FPP, and GGPP. The different products of each allylic substrate were analyzed (Figure 38). The enzyme demonstrated high activity with the substrates GPP and FPP and the lowest activity was with using DMAPP. GPP and FPP were likely to be the preferred substrate of this enzyme (Figure 38). There was no significant different between GPP and FPP as enzyme's preferred or natural substrate. The effect different concentrations of substrates on the enzyme activity was determined for the four substrates DMAPP, GPP, FPP, and GGPP in varying concentrations ranging from 0 µM to 100 µM (Figure 39). In general, the GrcC2 enzymatic reaction was linear at the low ranges of substrate concentrations for the substrates GPP, FPP and GGPP; the enzyme maximum activity was shown at lower substrate concentration (0.2-0.4µM for GPP) and for DMAPP the enzymes maximum activity was at the substrate concentration 80 µM.

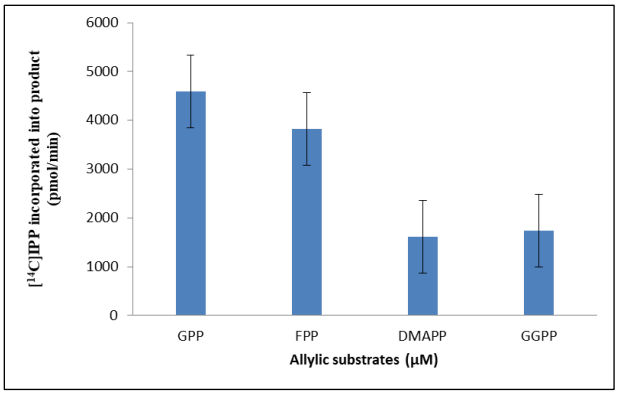


Figure 38: Activity of GrcC2 measured in the presence of the indicated allylic substrates as described in the methods section. The optimized reaction mix contains: 50 mM Tris buffer pH 8, 2 mM MgCl2, 0.2 % CHAPS, 400 ng purified protein, 2 mM DTT, 35 µM [^{14}C]IPP, 50 µM allylic substrate was used. GrcC2 showed its maximum activity in the presence of the substrates GPP and FPP, and GPP is likely to be the preferred substrate of this enzyme.

The characterization of kinetic parameters of GrcC2:

The standardized assays were used to determine the enzyme kinetic parameters. The values of K_m, V_{max} and K_{cat} were calculated using non-linear regression (SigmaPlot 11). The results are shown in table (2). The high K_m value indicates that the enzyme has lower affinity for the short chain substrate DMAPP (see Table 2). Also, according to the values of V_{max} and K_{cat}, there is little difference between the kinetic parameters calculated for GrcC2 in the presence of GPP or FPP.

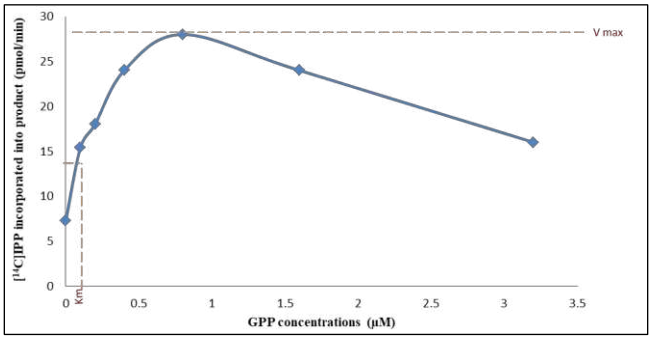


Figure 39: A representative activity curve of the effect of allylic substrate (GPP) varying concentrations on GrcC2 activity. Similar curves were obtained for the substrates DMAPP, FPP and GGPP. These data were used to calculate the kinetic parameters of GrcC2; V_{max} , K_m and K_{cat} values shown in table (2) using Michaellis-Menten assumptions.

Table 2: Kinetic parameters of GrcC2. The kinetic parameters were determined using the assay conditions as described for figure (39). The data were analyzed using nonlinear regression (Sigma Plot 11.0); V_{max} represents the maximum rate achieved by the system, at maximum (saturating) substrate concentrations. *K_m indicates the enzyme affinity for substrate. K_{cat} value gives direct measure of the catalytic production of product. K_{cat}/K_m measures enzyme efficiency.

	Substrates			
	DMAPP (C5)	GPP (C10)	FPP (C15)	GGPP (C20)
V_{max} (pmol/min)	25	28	43	25
K_m (µM)	43	0.1	0.2	0.18
K_{cat} (1/min)	0.00006	0.00007	0.0001	0.00006
K_{cat}/K_m	0.000001	0.0009	0.0006	0.0003

Final product analysis:

The chain length was determined by detecting the dephosphorylated radiolabeled compounds on reverse phase TLC plate. Multiple products were observed under the standardized assay conditions (C15, C20, C40, and C45). The results showed that GrcC2 catalyzes the enzyme reaction that produces multiple products with different chain lengths. The main products were FPP (C15), GGPP (C20), octaprenyl diphosphate (OPP) (C40) and solanesyl diphosphate (SPP) (C45) in the presence of the substrate GPP. Also with the substrates FPP and DMAPP the final products were mainly GGPP, OPP and SPP, but with the presence of GGPP as a substrate, the final product was primarily OPP (C40) (Figure 40). Thus, enzyme synthesizes a long polyprenyl chain length C45 which is needed for MK synthesis in mycobacteria, but primarily synthesizes a long polyprenyl chain length C40, which confirms the lake of essentiality and suggesting it may not be involved in MK synthesis (see Chapter 6).

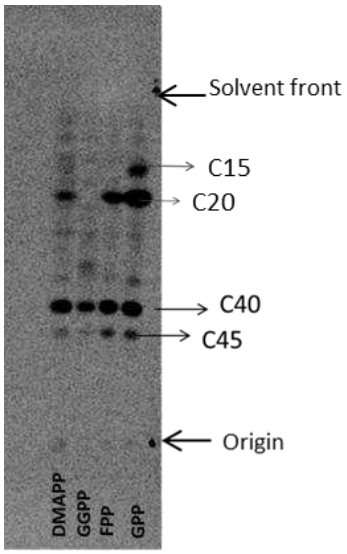


Figure 40: Reverse-phase TLC of reaction products of GrcC2. Equal amounts of radioactivity were spotted on reverse-phase TLC plates and the plates were developed in methanol:acetone (8:2, v/v). Products labeled with [^{14}C]IPP were visualized using Amersham Typhoon trio phosphoimager. Standard polyprenols were located with an anisaldehyde spray reagent.

The enzyme final products were also analyzed following the assay that was described in the study by Peters's group "Mann, et al 2011 and 2012" (as described in the methods section). The assay components used in Mann's studies were similar to the assay used in this study, but the major difference was in the way that the assay was stopped and the final product dephosphorylation, in our assay we used enzymatic dephosphorylation for final product analysis and Mann, *et al* used acidic dephosphorylation; the reaction was stopped and dephosphorylated by the addition of 9:1 methanol and 3N hydrochloric acid. The two assays were tested at the same time. TLC analysis was used to detect the final products of both assays (Figure 41). The results showed that using the assay conditions and enzymatic dephosphorylation optimized in this study; multiple products with different chain lengths were observed (C15, C20, C30, and C45) for both GrcC2 and control enzymes (Figure 41, a). The TLC analysis of the final products of the enzyme reaction for GrcC2 that was assayed using

Peters's group assay and using acidic dephosphorylation, resulted in no observed long-chain polyprenyl products with any of the substrates (Figure 41, b).

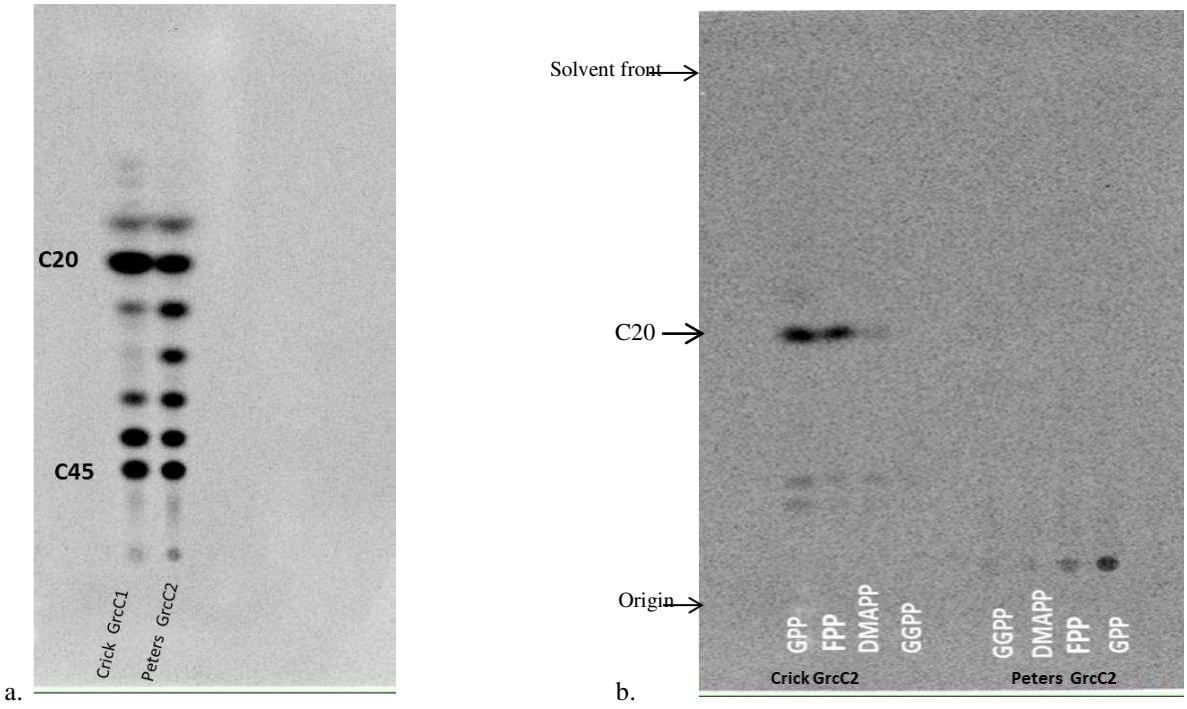


Figure 41: Comparing the dephosphorylation protocols of the two assays. Results of final products of the reaction assays with allylic diphosphate substrate were characterized using the thin layer chromatography (TLC) after the dephosphorylation of the synthesized radiolabeled compounds. In this assay, the purified protein expressed from the plasmid was provided by the Peters' group. a. Purified proteins (GrcC1 as a control and Peters' GrcC2) assayed in the reaction mix contained 50mM MOPS, 2 mM $MgCl_2$, 2mM DTT, 0.2% Triton X-100, 50 µM GPP and 35 µM [^{14}C]IPP and enzymatically dephosphorylated. b. Purified protein assayed in the reaction mix contained 50 mM sodium phosphate buffer (pH7), 10 % glycerol, 5 mM $MgCl_2$, 2 mM DTT, IPP*, Substrates, Final volume = 1 ml. Assay stopped and dephosphorylated by adding 9:1 methanol: 3N hydrochloric acid.

Chapter 5

Characterization of a long chain polyprenyl diphosphate synthase of *Mycobacterium smegmatis*[1] MS1133 homologous to GrcC1 and GrcC2 in *Mycobacterium tuberculosis*

In the laboratory, *Msmeg* is an important model system for *Mtb*. It has lower risk (non-pathogenic),is fast growing (doubling time ~ 3 hours) and is easy to manage and manipulate, which makes it an extremely useful model to study mycobacteria (Cordone, 2011; Lim, 2001). Studying and searching the *Msmeg* enzymes' functions and roles helps to understand the potential roles of these enzyme's homologues *Mtb* and other mycobacterial species (Tran & Cook 2005).

The BLAST homology search for long-chain polyprenyl diphosphate synthases showed that the protein expressed from *ms1133* in *Msmeg* is homologous with *Mtb* long chain prenyltransferases *grcC1* & *grcC2*. The alignment results show identity of MS1133 with GrcC1 (74%) and GrcC2 (60%). There is no reported function yet for *ms1133* in the *Msmeg* database (SmegmaList). *Msmeg* database annotated this gene as an orthologue of *Mtb grcC1* (Rv0562). In *Msmeg*, *ms1133* is located up stream of the gene *ms1132* (Figure 42), which encodes the protein (MenJ) a known MK reductase involved in the MK biosynthesis pathway (Upadhyay, 2015).

[1] Considered old nomenclature; based on Gupta, 2018, NCBI has adopted new amended divisions of mycobacterial species (5 distinct groups), which changed the names of the studied species including *M.smeg*.

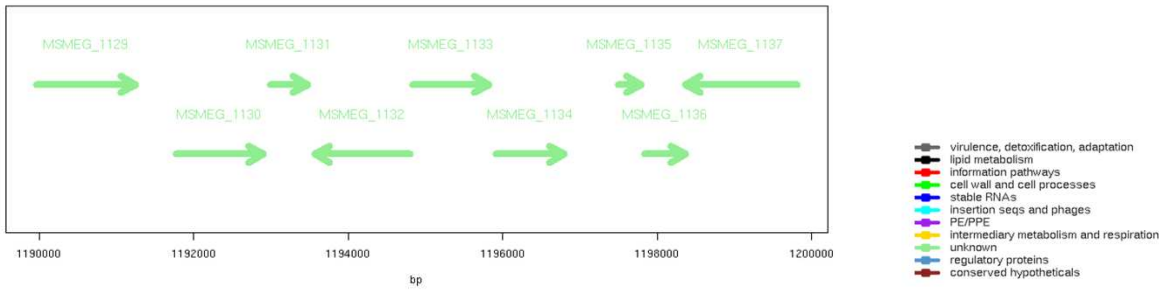


Figure 42: Organization of genomic region in *Msmeg* showing that the gene *ms1133* is located upstream region from the previously identified gene *ms1132* which are identified and involved in MK biosynthesis pathway.

Bacterial strains, plasmids and materials, used in this study:

The bacterial strain *E. coli* DH5α (from Invitrogen) was used as the cloning host; and the BL21 strain was used as the expression host. PCR kit and the plasmid vector pGEM-T easy vector system were purchased from Promega). The expression vectors were pET28a 5369 bp, and pET29b 5371 bp purchased from Novagen. pBlueScript (pBSK) 3000 bp was from Stratagene, and pUC4K 3914bp purchased from Amersham. The Expand High Fidelity PCR system and dNTPs were purchased from Roche. Diagnostics and restriction endonucleases were purchased from New England Biolabs. FastDigest enzyme was purchased from Fermentas Life Science and the DNA ligase from BioLabs Inc. Potato acid phosphatase was purchased from Sigma-Aldrich. The media for growing the bacterial strains was Lysogeny broth (LB), purchased from Invitrogen. B-PER Bacterial protein extraction reagent was purchased from Thermo Science. Monoclonal anti-polyHistidine antibody produced in mouse was purchased from Sigma-Aldrich and anti-mouse IgG (whole molecule) alkaline phosphatase antibody produced in goat was purchased from Sigma-Aldrich. BCA assay kit was purchased from Sigma. QIAprep

Spin Miniprep Kit and QIAquick Gel Extraction microcentrifuge Kit were purchased from QIAGEN. PD-10 Desalting columns were purchased from GE Healthcare Bio-Sciences. Amicon Ultra-15 30K centrifugal filter devices were purchased from Millipore. The EDTA-free protease inhibitor cocktail tablets were purchased from Sigma.

Amplification of *Msmeg* (MS_1133):

To facilitate gene amplification, we used the PCR approach. Two primers were designed based on a homology search of the *M. smegmatis* genome sequence, conducted on the webserver (SmegmaList) database. The designed primers were: atata **catatg** gtg gtg gca ggc gtt gac ttc (forward) and atata **aagctt** tcagccttcccggctgatcgt (reverse). The restriction sites for *NdeI* and *HindIII* (in bold) were introduced to the forward and reverse primers and genomic DNA of *M. smegmatis* was used as the template. The primers were synthesized and confirmed by Integrated DNA Technologies "XXIDT". Reaction mixes contained the enzyme *Taq* polymerase (1 unit), dNTPs (nucleotide mix 25%), DMSO (10%), and $MgCl_2$ PCR buffer (1.5 mM) as well as the primers. The PCR reaction was performed on a Perkin Elmer PCR system 2400.

The amplified 990 bp DNA fragment (PCR product) was extracted and gel purified using QIAquick Gel Extraction Kit following the manufacturer's protocol and digested with the restriction enzymes *NdeI* and *HindIII*. Both the PCR purified fragment and the ligation vector were digested with the restriction enzymes.

The purified fragment was ligated into the pGEM-T Easy vector (Promega). The ligation standard reaction in the kit protocol was followed. The ligation reaction was used to transform Subcloning Efficiency $DH5_{\alpha}$ Competent Cells (Invitrogen) following the manufacturer's guidelines. The cells containing plasmid were selected on Lennox agar containing Ampicillin

(100mg / ml), X-Gal (50 mg/ml) and 1 M IPTG (isopropyl-β-D-thiogalactopyranoside). Single colonies were picked and used to inoculate 5ml of Lennox/Ampicillin broth and the plasmid DNA was purified using QIAprep Spin Miniprep Kit following the manufacturer's procedure. The purified plasmidDNA (pGEMT- *ms1133* construct) was digested using the restriction enzyme *NdeI* and *HindIII* to confirm the correct size of the product (~1000 bp) using electrophoresis in 1%TBE agarose gel. The resulting 1000 bp fragment was purified and ligated into the expression vector pET28a+ and used to transform DH5α competent cells. The plasmids were purified using Qiagen mini-prep kit and subjected to restriction enzyme *Hind III* and *NdeI* and gel purified. The constructed pET28-*ms1133*, was used to transform into *E.coli* BL21(DE3)pLsS cells. The empty vector (pET28) was used to transform cells as a negative control. The transformed cells were cultivated in 100 ml of LB broth supplemented with kanamycin (50µg/ml) overnight. For over-expression, the transformed cells were grown overnight, transfered into 1 liter of LB broth and incubated with shaking at 37˚C to an optical density of 600 = 0.5. Protein expression was then induced by the addition of 0.5 mM of isopropyl β-D-thiogalactopyranoside (IPTG) and incubated for an additional 8 hours at 20˚C. The induced bacterial cells were harvested by centrifuging at 3000 rpm for 10 minutes, the pellets were lysed using B-PER bacterial protein extract solution and used for BCA protein estimation. Western blot analysis and (SDS-PAGE) were used to confirm protein expression. Primary (anti-His antibodies) and secondary (polyclonal goat anti-mouse) monoclonal anti-polyhistidine antibodies were used for Western blot analysis.

Purification of the recombinant *ms1133*:

The harvested cells were disrupted by sonication on ice (12 cycles of 30 seconds) in 50 mM MOPS (pH 7.9) lysis buffer containing: 200 mM $NaCl_2$, 10 mM $MgCl_2$, protease inhibitor

cocktail tablet, 5 mM β-mercaptoethanol (BME), and 10% glycerol. The homogenized cell fractions were centrifuged at 15,000 xg for 15 minutes at 4˚C. The precipitate (resulted from sonication supernatant) was suspended with buffer containing 50mM MOPS, 200mM NaCl, 1mM $MgCl_2$, 1mM BME, and 10% glycerol. Affinity purification of the recombinant protein was done using Bio-RAD poly-prep chromatography nickel columns and washed with different concentrations of imidazole buffer. The protein was eluted in 1 ml volumes, with increasing imidazole concentrations (25, 50, 100, 150 and 200 mM). The level of protein purification was determined by Western blot and 10 % sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) on a series of different concentrations of imidazole. Western blot is a sensitive method that can detect purified proteins, probed and analyzed with monoclonal anti-polyhistidine taq produced in mouse. The purified protein was desalted using PD-10 desalting columns (from GE Healthcare Bio-sciences). The desalted protein (3.5 ml) was concentrated using centrifugal filter units for concentration and purification of biological solutions (30,000 MWCO) Amicon Ultra-15. BCA protein estimation kit was used to determine protein concentration. The protein quantified using a calibration curve created with BSA (bovine serum albumin), and the measurements where done at 562 nm.

Protein activity assay for MS1133 (enzyme assay):

The activity of the resulting recombinant protein was determined by estimating the amount of radiolabeled [^{14}C]IPP incorporated into butanol-extracted polyprenyl diphosphate, which is an accurate and sensitive detection method. The 50 µl standard reaction assay mix contained 50 mM MOPS buffer pH 7.9, 2 mM $MgCl_2$, 2 % Triton X-100, 2 mM DTT, 35 µM [^{14}C]IPP, 50 µM allylic substrate (GPP) and 35 ng purified protein. Starting with this reaction mix the optimal

reaction conditions for the assays were identified.The active *M. tuberculosis* protein GrcC1 was used as a control.

Enzymatic assays and reaction requirements:

Assay conditions required for optimal activity of the enzyme were standardized for protein concentration, time, divalent cations and pH, and were linear for protein concentration and time.

The effect of protein concentration in the reaction:

To determine the protein concentration, the assay was conducted to adjust the optimal protein concentration in the previous reaction mix using 5 different protein concentrations in the mixes (50, 100, 200, 300 and 400 ng). The reaction mixes were incubated in a water bath at 37˚C for 30 minutes. The reaction was stopped by adding 1 ml of water saturated with NaCl. The [^{14}C]radiolabeled products were extracted with butanol saturated with water. Aliquots of the extracted product (100 µl) were used for liquid scintillation spectrometry.

Optimization of reaction time:

To optimize the reaction time, the previous reaction mix was incubated in water bath at different time ranges (0, 5, 10, 15, 20, 25, 30 and 35 minutes). The reaction was stopped by adding 1 ml of water saturated with NaCl. The [^{14}C]radiolabeled products were extracted with butanol saturated with water. Aliquots of the extracted product (100 µl) were used for liquid scintillation spectrometry.

Effect of metal divalent cation on the activity of MS1133:

The rate of radioactivity incorporated into product was tested in the absence and presence of divalent cations. The recombinant protein was incubated for 15 minutes with Bio-Rex 70 cation exchange resin to remove endogenous cations. The incubated protein was tested in the previous reaction mix with and without divalent cation ($MgCl_2$) at 2 mM concentration, with a control that

containing recombinant protein not incubated with Bio-Rex resin. To determine which divalent cation increases enzyme activity, the effect of different divalent cations was studied. The endogenous divalent cations were removed by using Bio-Rex mesh and different cations, namely $CaCl_2$, $MgCl_2$, $MnCl_2$, $CoCl_2$ and $ZnCl_2$, were added to the reaction mix and the radioactivity was determined by scintillation counter.

pH optimization:

The pH of the assay was adjusted using 3 different buffers (MES, MOPS and Tris). The reaction mixes were incubated in a water bath at 37˚C for 30 minutes. The reaction was stopped by adding 1 ml of water saturated with NaCl. The [^{14}C]radiolabeled products were extracted with butanol saturated with water. Aliquots of the extracted product (100 µl) were used for liquid scintillation spectrometry.

Detergent effect:

The effect of detergents on MS1133 activity was tested as well; the rate of radioactivity incorporated into product was analyzed in the absence and presence of detergent in the reaction mix. The enzymatic reaction was conducted as described previously with the addition of different detergents to test if they enhance the enzyme activity in the assay. The effect of the detergents Triton x-100, CHAPS and Tween-20 were tested at 0.2 % concentration.

Using the optimized reaction mix, the enzymatic activity of the recombinant protein was tested with different allylic substrates which are dimethyl allyl diphosphate (DMAPP), farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP), and different products of each allelic substrate were identified.

The characterization of kinetic parameters:

The kinetic parameters for MS1133 were determined with different concentrations of the four substrates (GPP, DMAPP, FPP or GGPP) with constant concentration of [^{14}C]IPP and using Michaelis-Menten assumptions. The values of V_{max} , K_{cat} and K_m for each substrate were calculated. The data were analyzed by nonlinear regression using the SigmaPlot 11.0 statistical program.

The reaction product analysis:

The final products of the reaction assays with different allylic diphosphate substrates were characterized using the thin layer chromatography (TLC) after the dephosphorylation of the synthesized radiolabeled compounds.

The radiolabeled products (after the removal of butanol under nitrogen) were subjected to the dephosphorylation process to analyze the product chain length. The enzyme potato acid phosphatase was used for the dephosphorylation of the final products. Starting with 100 µl aliquot of butanol extracted products; the butanol was removed under nitrogen stream and the radiolabeled products were dissolved in 5 ml of buffer (100 mM sodium acetate pH 4.8, 0.1% triton-X, 60% methanol) and bath sonicated for 2 minutes. Then, 20 units of the potato acid phosphatase were added to the mix and incubated overnight at 25˚C. The dephosphorylated products were extracted with 1 ml of n-hexane (3 times). The n-hexane pooled extracts were washed with water (1 ml) and the solvent evaporated under stream of nitrogen. The extracts were dissolved in 200µl of (2:1, v/v) chloroform:methanol and aliquots were used for scintillation spectrometry (20 µl) and TLC analysis. The radiolabeled product aliquots were spotted on reversed-phase C18 TLC plates which were developed in methanol:acetone (8:2, v/v) and visualized by the phosphoimager (Bioscan system imaging scanner). Geraniol (GOH), farnesol

(FOH), geranylgeraniol (GGOH) and polyprenols were used as standard polyprenols for the TLC plate and these standards were located and visualized using an anisaldehyde spray reagent. All data presented is representatives of multiple experiments conducted in triplicate.

Results

Amplification of *M. smegmatis* (MS1133):

Msmeg ms1133 was successfully cloned, PCR reaction was performed as previously described, and the amplified 990bp sized band of *ms1133* was extracted and gel purified and ligated into the expression vectors (Figure 43) as previously described. Western blot analysis and 10% SDS-PAGE were used to confirm protein expression and size. Primary (anti-His antibodies) and secondary (polyclonal goat anti-mouse) monoclonal anti-polyhistidine antibodies were used for Western blot analysis.

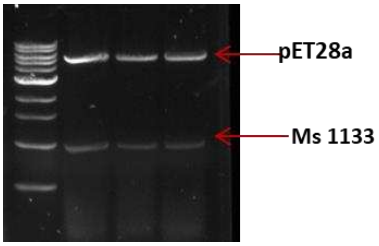


Figure 43: PCR amplified *ms1133* DNA fragment after double digestion using *NdeI* and *HindIII*. Reaction mixes contained the enzyme *Taq* polymerase (1 unit), 200 µM dNTPs, DMSO (10%), and 1.5 mM $MgCl_2$ buffer as well as the primers. The PCR reaction was performed on a Perkin Elmer PCR system 2400. The 3 lanes are 3 PCR reactions with the confirmed clones.

Purification of the recombinant *ms1133*:

Affinity purification of the recombinant protein was done as described in methods section. The extracted protein was washed and eluted with different concentrations of imidazole buffer (25,

50, 100, and 150 mM). The level of protein purification was determined by Western blot and 10 % SDS-PAGE on a series of different concentrations of imidazole (Figure 44). BCA protein estimation kit was used to determine protein concentration.

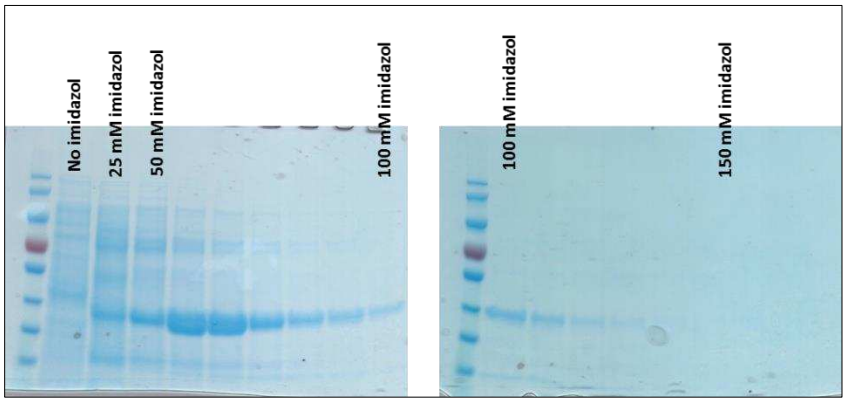


Figure 44: Purification of MS1133: 10% SDS-PAGE analysis of extracted protein showing the purification of MS1133 at the imidazole concentrations 25, 50, 100 and 150 mM. A lane contained 25 mM immidazol, 6 lanes contained 50 mM, 6 lanes contained 100 mM and 3 lanes contained 150 mM.The 10% SDS-PAGE were visualized using Coomassie Brilliant Blue G-250 stain.

Protein activity assay for MS1133 (Enzymatic assay):

The activity of the resulting recombinant protein was determined by estimating the amount of [^{14}C]IPP incorporated into butanol-extracted polyprenyl diphosphate in 50 µl standard reaction assay mix contained 50 mM MOPS buffer pH 7.9, 2 mM $MgCl_2$, 2 % Triton X-100, 2 mM DTT, 35 µM [^{14}C]IPP, 50 µM allylic substrate (GPP) and 35 ng purified protein. The amplified protein was active under the assay conditions, GrcC1 active protein was used as a control.

Enzymatic assays and reaction requirements:

The purified enzyme was active and the assay conditions required for optimal activity of the enzyme were standardized for protein concentration, time, divalent cations and pH. The previously mentioned reaction mix content was used to standardize the enzymatic assay requirements.

Protein concentration:

The assay was conducted to adjust the optimal protein concentration in the reaction mix using 5 different protein concentrations. The optimal protein concentration for the reaction was 300 ng which was used for further assays (Figure 45).

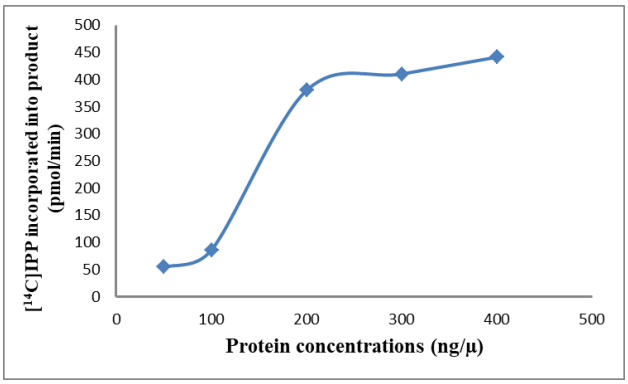


Figure 45: Protein concentration curve shows the effect of protein concentration on the activity of MS1133. Protein concentrations in the reaction mixes were (50, 100, 200, 300 and 400 ng). The reaction was incubated at 37˚C for 30 minutes. The [^{14}C]radiolabeled products were extracted and analyzed using liquid scintillation spectrometry.

Optimization of reaction time:

To optimize the reaction time, the previous reaction mix was incubated in water bath at different time ranges (0, 2, 4, 8, 16, 24, 28 and 32 minutes). The reaction was linear for approximately 7 minutes (Figure 46). Further experiments were conducted with 7 minutes incubation time.

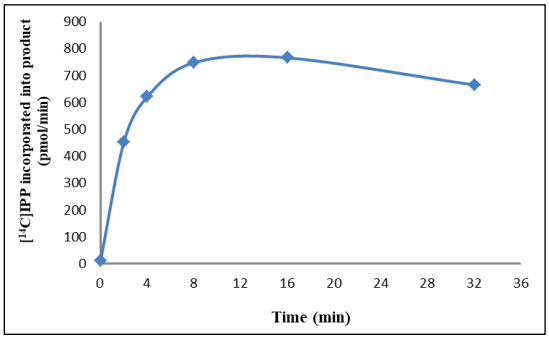


Figure 46: Reaction time optimization for MS1133. The time range started at 0 minutes to 35 minutes were tested using the previously described assay mix. The enzyme activity was linear for approximately 7 minutes.

Effect of metal divalent cation on the activity of MS1133:

The incubated protein was tested in the previous reaction mix with and without divalent cation ($MgCl_2$) at 2 mM concentration, with a control that containing recombinant protein not incubated with Bio-Rex 70 resin. The results showed that the enzyme requires the presence of divalent cations for the activity (Figure 47). To determine which divalent cation increases the enzyme activity, the effect of different divalent cations was studied. The endogenous divalent cations were removed by using Bio-Rex 70 cation exchange resin mesh and the cations, $CaCl_2$, $MgCl_2$, $MnCl_2$ and $ZnCl_2$, were added to the reaction mix. The enzyme optimal activity was shown in the present of $MgCl_2$ at 2mM of concentration (Figures 48). MS1133 activity requires the presence of divalent cations and $MgCl_2$ was the preferred cation. The effect of different concentrations of $MgCl_2$ on the protein activity was analyzed. To optimize the Mg^{++} concentration in the reaction, concentration range of 0.25 mM to 4 mM were tested and 1 mM was optimal and used in subsequent assays (Figure 49).

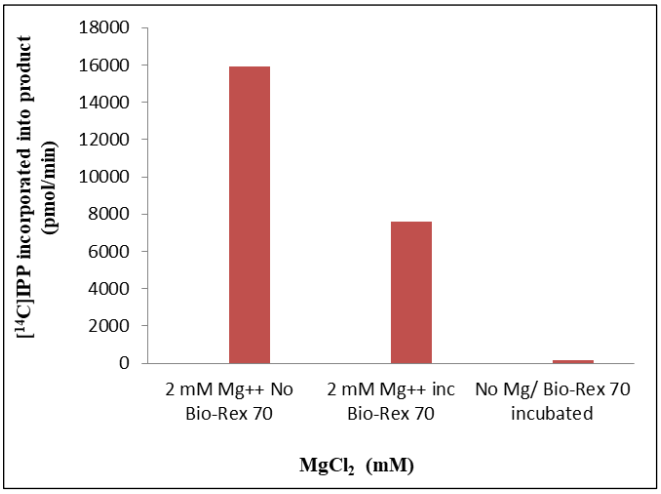


Figure 47: Divalent cation requirements for the activity of MS1133. The rate of incorporated [^{14}C]IPP into product was tested in the absence and presence of divalent cations. The recombinant protein was incubated for 7 minutes with Bio-Rex 70 cation exchange resin. The incubated protein was tested in the reaction mix with and without divalent cation ($MgCl_2$) and, with control containing recombinant protein not incubated with Bio-Rex resin.

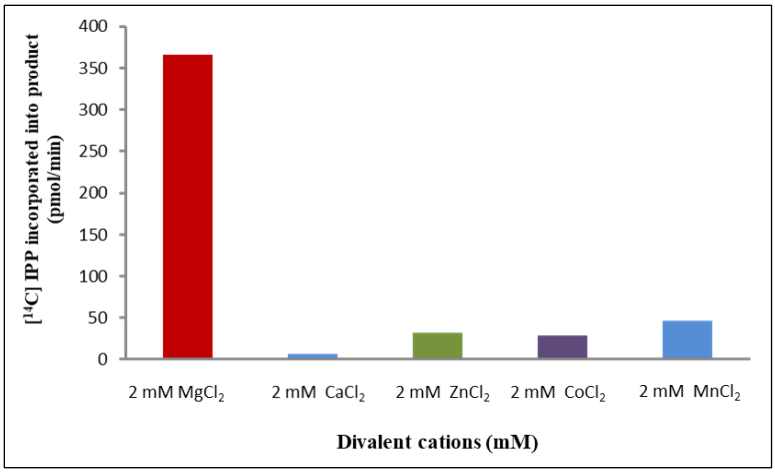


Figure 48: The enzyme MS1133 requires the divalent cations to support activity and $MgCl_2$ was the preferred cation. The effect of different divalent cations was tested and the endogenous cations were removed by using Bio-Rex 70 cation exchange resin. The cations, $CaCl_2$, $MgCl_2$, $MnCl_2$ and $ZnCl_2$, were added to the reaction mix at 2 mM concentration. The enzyme optimal activity was shown in the present of $MgCl_2$.

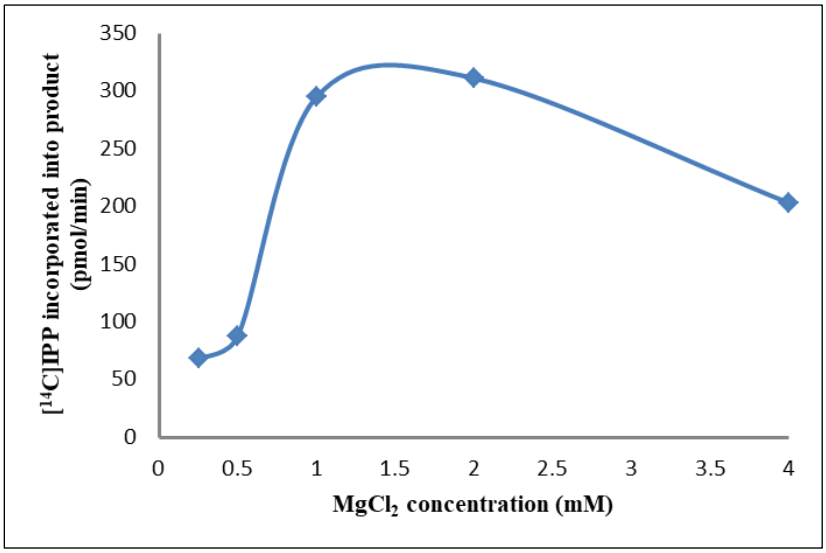


Figure 49: Optimization of Mg^{++} concentration for MS1133 activity. 1 mM Mg^{++} concentration gives the highest reaction activity. Different concentrations of Mg^{++} range of 0.25 mM to 4 mM were tested and 1 mM was optimal and used in further assays.

pH optimization:

The enzymatic assay was conducted to adjust the reaction pH in the reaction mixes using 3 different buffers (MES, MOPS and Tris) with pH ranges start from 5.0 to 9.0. The optimal pH for ms1133 activity was determined to be between 7 and 7.5 in MOPS buffer (Figure 50).

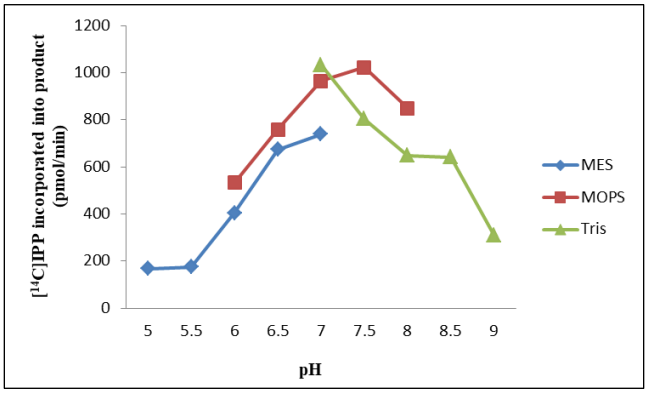


Figure 50: Effect of pH on MS1133 activity. The enzymatic activity was tested in three different buffering systems with pH ranged from 5.0 to 9.0. The optimum pH for the reaction was 7.5 (MOPS buffer) which was used for the further assays.

Detergent effect:

The effect of different detergents including Triton X-100, CHAPS and Tween-20 was tested at 0.2% concentration. MS1133 showed a decrease in its enzymatic activity in the presence of detergents (Figure 51). This enzyme did not require the presence of detergents in the reaction. No detergents were used in the further assays.

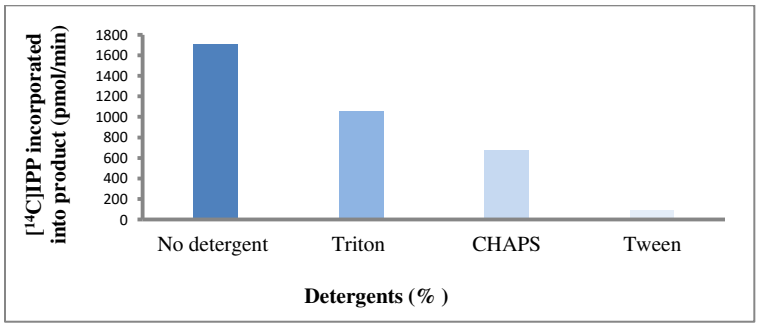


Figure 51: The effect of detergents on MS1133 protein activity. The rate of incorporated [^{14}C]IPP into product was tested in the presence of the detergents: Triton X-100, CHAPS and Tween-20 at 0.2% concentration. The enzyme activity was decreased in the presence of detergents.

Specificity of the allylic substrates

By using the standardized reaction mix, the optimized assay conditions were used to specify the enzyme substrates, characterize the enzyme kinetic parameters and final product. The enzymatic activity of the recombinant protein was tested with the allylic substrates: DMAPP, GPP, FPP, and GGPP. The concentration for each substrate in the reaction was 50 µM, and 35 µM of IPP. The products of each allelic substrate were analyzed. The enzyme demonstrated high activity with the substrates GPP and FPP and lower activity was with using GGPP as a substrate, and almost no activity with DMAPP. FPP was likely to be the preferred substrate of this enzyme (Figure 52). The effect of substrates different concentrations on the enzyme activity was determined for the four substrates DMAPP, GPP, FPP, and GGPP in varying concentrations ranging from 0.25 µM to 40 µM (Figure 53). In general, ms1133 enzymatic reaction was linear at the very low ranges of substrate concentrations (0.25 to 7 µM) for the substrates GPP and FPP, and at higher concentrations (from 5-40 µM) for the substrates GGPP and DMAPP.

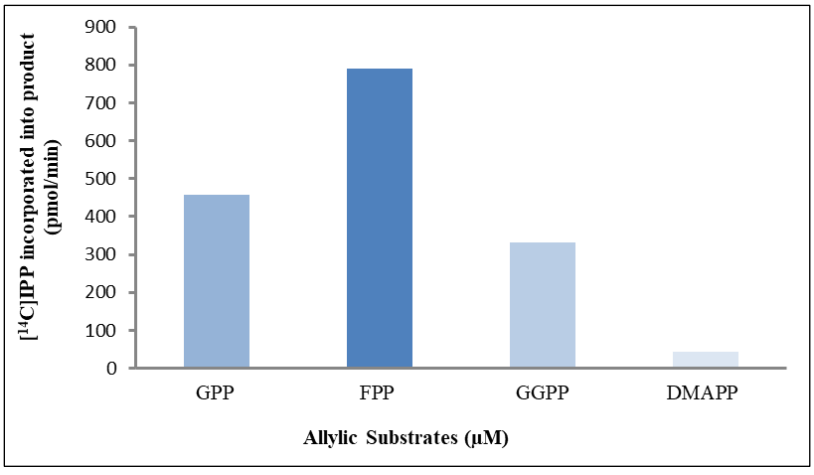


Figure 52: The activity of MS1133 was measured in the presence of the indicated allylic substrates as described on methods section. The optimized reaction mix contained: 50 mM MOPS buffer (pH 7.5), 2 mM $MgCl_2$, 300 ng purified protein, 2 mM DTT, 35 µM [^{14}C]IPP, 50 µM allylic substrate was used. MS1133 showed maximum activity in the presence of the substrates FPP and GPP, and FPP is likely to be the preferred substrate of this enzyme.

The characterization of kinetic parameters:

The standardized assays were used to determine the enzyme kinetic parameters. The values of K_m, V_{max} and K_{cat} were calculated using non-linear regression (SigmaPlot 11), Figure 53. To reflect the enzyme's affinity for the used substrates, we calculated the K_m for MS1133 with different substrates. The low K_m value indicates that the enzyme has greater affinity (0.4 µM) for the longer chain substrates FPP and GGPP (see Table 3).

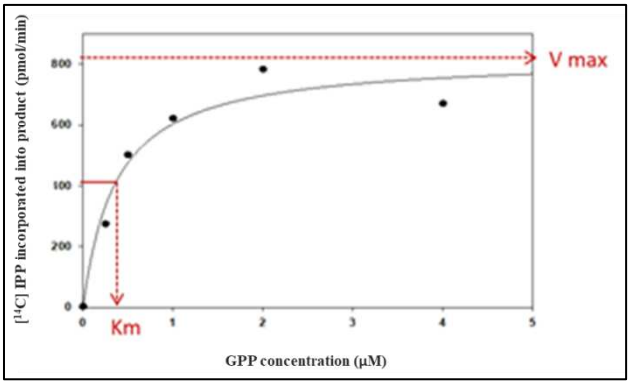


Figure 53: A representative activity curve of the effect of allylic substrate (GPP) varying concentrations on MS1133 activity. Similar curves were obtained for the substrates FPP, DMAPP and GGPP. These data were used to calculate the kinetic parameters of MS1133; V_{max} , K_m and K_{cat} values shown in Table (3) using Michaellis-Menten assumptions.

Table 3: Kinetic parameters of MS1133. The kinetic parameters were determined using the assay conditions as described for figure (52). The data was analyzed using nonlinear regression (Sigma Plot 11.0) (figure 53); V_{max} represents the maximum rate achieved by the system, at maximum (saturating) substrate concentrations. K_m indicates the enzyme affinity for substrate. K_{cat} value gives direct measure of the catalytic production of product. K_{cat}/K_m measures enzyme efficiency.

	Substrates			
	DMAPP (C5)	GPP (C10)	FPP (C15)	GGPP (C20)
V_{max} (pmol/min)	307	824	1944	1408
K_m (µM)	70	0.4	2.8	22
K_{cat} (1/min)	0.0008	0.0002	0.0005	0.0004
K_{cat}/K_m	0.00001	0.0005	0.0002	0.000002

The reaction product analysis:

The final products of the reaction assays with different allylic diphosphate substrates were characterized and the chain length was determined by detecting the radiolabeled compounds on reverse phase TLC plate after the dephosphorylation of the synthesized radiolabeled compounds. Multiple products were observed under the standardized assay conditions (C20, C35, C40 and C45). The results showed that MS1133 catalyzes the enzyme reaction that produces multiple products with different chain lengths, the main products were C35 and C40 in the presence of the substrate GPP, FPP and GGPP. Also with the substrates GPP and GGPP the final products included C45, but with the presence of DMAPP, no visible products were shown (Figure 54). With substrates specificity, the enzyme synthesizes a long polyprenyl chain (length C40- C45) which may need for MK synthesis in mycobacteria.

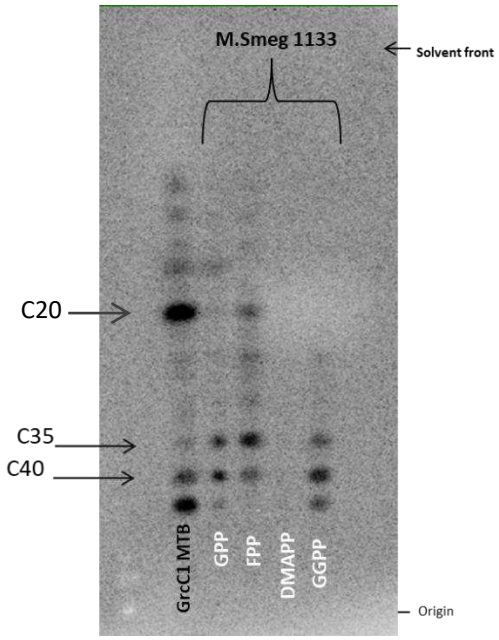


Figure (54): Reverse-phase TLC of reaction products of MS1133. Equal amounts of radioactivity were spotted on reverse-phase TLC plates and the plates were developed in methanol:acetone (8:2, v/v). Products labeled with [^{14}C]IPP were visualized by Amersham Typhoon trio phosphoimager. Standard polyprenols were located with an anisaldehyde spray reagent, also a sample of GrcC1 was used as a radioactive standard.(left lane). Compared to non-radioactive standards (GGOH, geranylgeraniol: FOH, farnesol) on the same plate.

Chapter 6

Essentiality of *grcC1*, *grcC2* and *ms1133*

In this study we investigated the essentiality or non-essentiality of the enzymes GrcC1, GrcC2 and MS1133. According to mycobacteria database, *grcC1* is predicted to be essential but *grcC2* is not. *Msmeg* MS1133 a homologue of GrcC1 and GrcC2 and is annotated in the bioinformatics database as of unknown function. An *in vitro* gene knock-out technique was used for *grcC1* and *grcC2* from *Mtb* H37Rv and *Msmeg ms1133* compared with the wild type cells.

The methodology that was used in this part of the study depends on the use of counter selection with *sacB,* which allows the positive selection of insertional mutations. This method was used to allow efficient gene exchange mutagenesis.

Materials and Methods

Construction of *grcC1*-Kan, *grcC2*-Kan and *ms1133*-Kan on the plasmid pBKS:

The pBSK vector from Stratagene, was used as a vector for the in vitro transcription process. To generate knockouts of *grcC1*, *grcC2* and *ms1133*; forward and reverse primers were designed for each gene: KOGrcC1F (atatctagaccgttgcccagcggaaagatc) KOGrcC1R (atatctagatgagcgctggtcgccagctct), KOGrcC2F(atatctagaggatatcggcgtttttgcgcc) KOGrcC2R (atatctagaaaccgggaaacgcgtccaact), and KOMS1133F(atatctagacggcggcgaccgagcggat) KOMS1133R (atatctaga gtgcatggcccgcagcgcc). All the forward and reverse primers were flanked with restriction sites (*XbaI*). The primers were synthesized and confirmed by Integrated DNA Technologies "XXIDT". The PCR amplification reaction for each gene was performed on a Perkin Elmer PCR system 2400. The amplified DNA fragment (PCR product) was extracted and gel purified (1% agarose gel) using QIAquick Gel Extraction Kit following the manufacturer's protocol. All the PCR purified fragments and the ligation vector were digested with the restriction enzyme *XbaI*. Each of the purified fragments was ligated into the pGEM-T Easy vector (Promega), following the kit protocol for the standard reaction. Each of the ligation reactions was transformed into Subcloning Efficiency DH5$_\alpha$ Competent Cells (Invitrogen) following the manufacturer's guidelines. The cells containing plasmid were selected on Lennox agar containing Ampicillin (100 mg/ml). Single colonies were picked and used to inoculate 5 ml of Lennox/Ampicillin broth and each of the plasmid DNA's was purified using QIAprep Spin Miniprep Kit following the manufacturer's procedure. Each of the purified plasmidDNA was digested using the restriction enzyme *XbaI* to confirm the correct size of the products using electrophoresis 1% TBE agarose gel. The *XbaI* purified fragments of *grcC1*, *grcC2* and *ms1133*

were ligated into pBKS shuttle vector (*XbaI* digested and dephosphorylated), then used to transforme in DH5$_\alpha$ competent cells. The pBKS vector dephosphorylation protocol was: 1/10 volume of 10x antarctic phosphatase reaction buffer added to 5 µg of pBKS (previously digested with *XbaI*), then 5 units of anthractic phosphatase were added and mixed. The reaction was incubated for 15 minutes at 37 C°. The enzyme was heat inactivated for 7 minutes at 65 C°.The constructed plasmids were purified using QIAprep Spin Miniprep Kit following the manufacturer's protocol and digested with *XbaI* to confirm the clone. The purified plasmids (pBKS-*grcC1*, pBKS-*grcC2* and pBKS-*ms1133*) were double digested with restriction enzymes *Asc1*, *NruI*, *BglII* and *NarI* for deletion, to be ligated with kanamycin cassette from pUC4K.

Construction of knock out mutations tagged with kanamycin resistance gene:

We started with the construction of deletion alleles tagged with an antibiotic resistance gene (kanamycin) flanked by specific recombination sites followed by recombination of the plasmid borne deletion into the chromosome. The constructed plasmid pBKS-*grcC1* was digested with the enzymes *Asc1* and *NruI*, pBKs-*grcC2* with *NruI* and *BglII*, and pBKs-*ms*1133 with *NarI*. The digestions for pBKS-*grcC1*and pBKs-*grcC2* were performed in two steps. The sizes of digested plasmid DNA were confirmed using 1% TBE agarose gel electrophoresis. The DNA bands for knock out constructs *grcC1*, *grcC2* and *ms1133* were gel purified following the manufacturer's protocol. All the purified DNA's were blunt-ended using the Quick Blunting Kit from BioLabs. To disrupt these genes with antibiotic cassette; the vector pUC4K (from Amersham) was used as the source of kanamycin cassette. The pUC4K vector was transformed and cloned in *E.coli* DH5α then column purified. The purified pUC4K vector was digested with restriction enzyme *HincII*, and gel purified the kanamycin cassette. The *HincII* blunt-ended *grcC1*, *grcC2* and *ms1133* were ligated with the kanamycin cassette following Quick Ligation

Protocol (BioLabs) and transformed into MAX Efficiency DH5α competent cells (Invitrogen). The transformed cells were incubated overnight in LB media with the addition of kanamycin, X-Gal and IPTG. The plasmid-DNA-kanamycin cassette (*grcC1*-Kan, *grcC2*-Kan and *ms1133*-Kan) were extracted (following the protocol of High Pure Plasmid kit from Roche), gel purified and digested with *XbaI* to confirm the clones' sizes and sequenced.

The thermosensitive plasmid vector pPR27 carrying the counter selectable marker *sacB* was gel extracted and digested with *XbaI*. A copy of each knockout gene disrupted with kanamycin and digested with *XbaI*, was cloned into the suicide vector pPR27 and used to transforme MAX Efficiency DH5α competent cells.

Electrocompetent cells of *Msmeg* and *Mtb* were prepared according to the Broussard procedure (Broussard, 2009). The construct *ms1133*-Kan cloned into the suicide vector pPR27 was electroporated in the prepared *Msmeg* electrocompetent cells following the procedure described by Jackson et al, 2002. The kanamycin and sucrose resistant clones in *Msmeg* colonies were tested using catechol solution, cultured and used for colony PCR analysis. Colonies were picked, dissolved in water, boiled and used as DNA template in the PCR reaction mix. The forward and reverse colony PCR primers were: FKOMS1133: ataGGATCGCCACGCCGTACACGGTC and RKOMS1133: ataGTTCCGTCCCGACGCGAACGCGT and PCR reaction using wild type DNA template was used as control. The constructs *grcC1*-Kan and *grcC2*-Kan cloned into the suicide vector pPR27 were electroporated in the prepared *Mtb* electrocompetent cells following same procedure. The wild type and the mutant cells were grown to test the ability of each cell to survive in the absence and the presence of the genes.

Growth curve:

The bacterial growth curve was obtained by measuring the OD 600 nm of cultures growing in 7h9 medium over 100 hours for *Msmeg* and 20 days for *Mtb*. The OD measurements may influence by the cell clumps, and we avoided that by shaking with magnetic beads.

Lipid extraction for menaquinone quantitation:

The bacterial cells of both wild type and knock-outs constructs were cultured into 50 ml 7h9 to an OD of 0.8. The cultures were centrifuged at room temperature (25,000 RPM for 15 min). Then 15 mls of chloroform:methanol (2:1) was added to the cells. After 10 minutes, the mix of cells and solvent was centrifuged for 20 minutes. The bottom layer for each mix was washed with 1 ml of water and evaporated under nitrogen. The dry samples were re-suspended in 100% chloroform for silica column extraction. The lipid samples in the chloroform were passed through the silicic acid column and each was washed with 5 mls of chloroform. The extracted lipid samples were dried under nitrogen. The lipid samples were dissolved in 100 ethanol and vitamin K2 solution was added to the samples and the control (2 µg total concentrations) as an internal recovery standard. The purified lipid molecular mass was analyzed using liquid chromatography–mass spectrometry technique (LC-MS) performed at CSU, CIF Mass Spectrometry.

Results

Construction of *grcC1*-Kan, *grcC2*-Kan and *ms1133*-Kan on the plasmid pBKS:

The PCR products were analyzed using 1% agarose gel. The sizes of the resulted DNA fragments matched the expected sizes, for KO*grcC1* the fragment size was 2 Kbp, KO*grcC2* was 2.1Kbp and for KO*ms1133* was 1.5 Kbp (Figure 55).

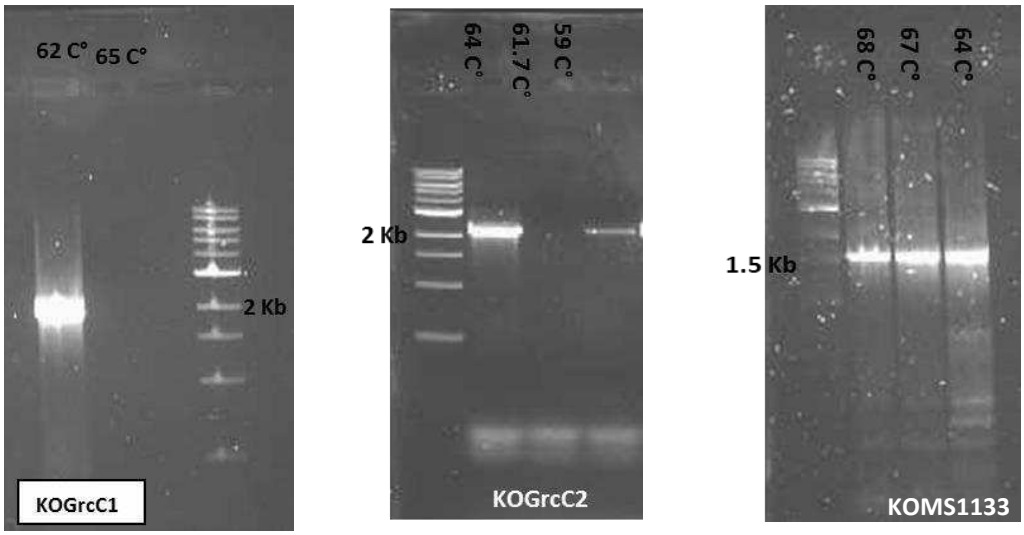


Figure 55: Agarose gel (1%) electrophoresis analysis showing amplified PCR products. The fragments of KO*grcC1*, KO*grcC2* and KO*ms1133* with expected sizes 2, 2.1 and 1.5 Kb respectively.

The DNA fragments that resulted from the PCR were ligated in the cloning vector pGEM-T easy and transformed into *E.coli* DH5α. Clones with the expected inserts of KO*grcC1*, KO*grcC2* and KO*ms1133* were successfully identified. The clones were then digested with *XbaI* restriction enzyme, and gel purified using QIAquick kit (Figure 56).

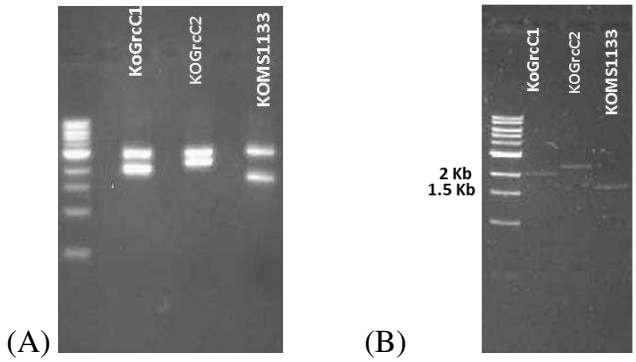


Figure 56: Ligation and purification of knock out PCR resulted fragments. (A) Showing the *XbaI* digested DNA fragments ligated into pGEM-T easy. (B) Showing the gel purified KO*grcC1*, KO*grcC2* and KO*ms1133* using QIAquick kit.

The gel purified KO*grcC1*, KO*grcC2* and KO*ms1133* fragments were successfully cloned into the shuttle vector pBSK (3K) and the clones were confirmed by *XbaI* digestion and sequencing (Figure 57).

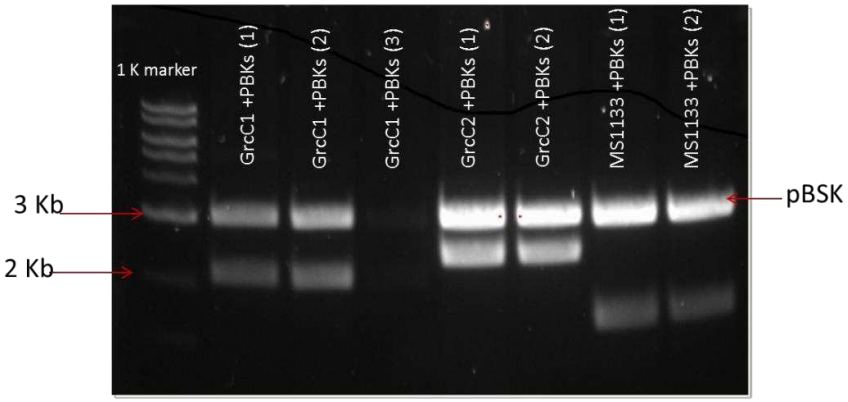


Figure 57: *XbaI* digestion to confirm the clone of *grcC1*, *grcC2* and *ms1133* in pBKS.

The *XbaI* digested constructs of *grcC1*, *grcC2* and *ms1133* with pBKs showed the correct fragments sizes on the gel electrophoresis, ~5 kb for each of pBKs-*grcC1* and pBKs-*grcC2* and 4.5 kb for pBKs-*ms1133*. The constructs were confirmed also with sequencing.

Construction of knock out mutations tagged with kanamycin resistance gene:

The constructed plasmids pBKs-*grcC1*, pBKs-*grcC2* and pBKs-*ms1133* were successfully disrupted with the kanamycin cassettes. The gel purified plasmids were digested with *XbaI* enzyme to confirm the size of the DNA fragments in the clones.

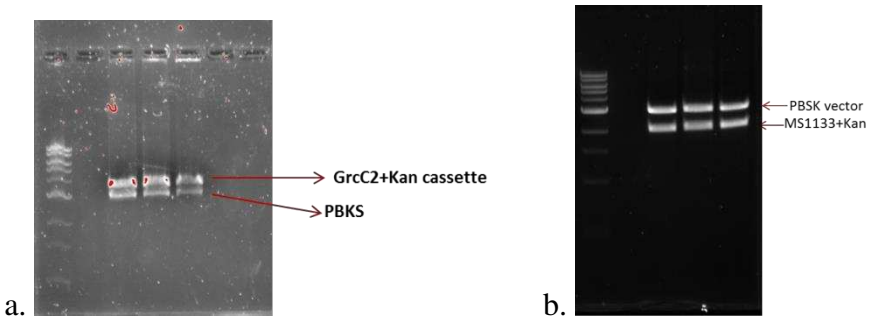


Figure 58: *XbaI* digestion for confirmation of the clones: a. *grcC2*+ Kan + pBKs. The size of the vector plasmid pBKs ~2.9Kb and *grcC2*+Kan ~3.2Kb comparing with the 10 K DNA ladder.

The constructed plasmids pBKs-*grcC1*-Kan, pBKs-*grcC2*-Kan and pBKs-*ms1133*-Kan were cloned into mycobacteria replicating vector pPR27 containing a *XylE* thermosensitive selectable marker. The clones were transformed into Max Efficiency DH5α competent cells. The clones in the resulted colonies grown on LB media supplemented with Kanamycin were confirmed by colony PCR confirmation and *XbaI* digestion (Figure 59). The plasmids were extracted from the resulting colonies and the constructed plasmids were ready to be electroporated into competent *Mtb* and *Msmeg*. The suicide vector pPR27containing *ms1133*-Kan was introduced into the electrocompetent *Msmeg* by electroporation and plasmids pBKs-*grcC1*-Kan and pBKs-*grcC2*-Kan in pPR27 electroporated into electrocompetent *Mtb* cells. After the incubation period on 7h10 plates containing kanamycin, there were multiple colonies of antibiotic resistant transformants on the *Msmeg* plates, few colonies on *Mtb* with pBKs-*grcC2*-Kan in pPR27 and no growth for plates inoculated with pBKs-*grcC1*-Kan in pPR27. To isolate the cells contain

clones that were allelic exchange mutants, the colonies were tested with a catechol test and 5 yellow colonies were picked and grown in liquid media and then plated onto solid media containing 2% sucrose.

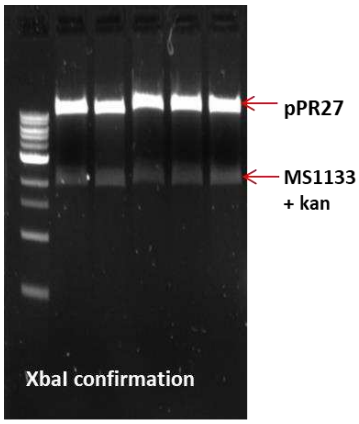


Figure 59: *XbaI* digestion for confirmation of the clones: a. *grcC2*+ Kan + pBKs. The size of the vector plasmid pBKs ~2.9Kb and *grcC2*+Kan ~3.2Kb comparing with the marker.

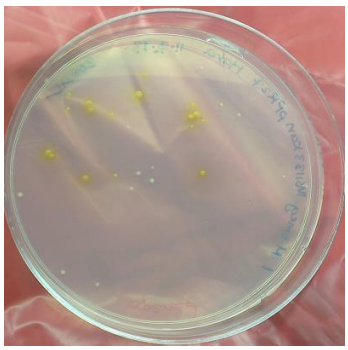


Figure 60: Catechol test for the culture of *Msmeg* transformants with the suicide vector pPR27containing *ms1133*-Kan.

The picked colonies were catechol tested again (Figure 60), and the results of this test allowed selection of the positive colonies that carry the disrupted copy of the targeted gene in their chromosomes (double crossover exchange mutants). From catechol test, four positive white colonies were picked with the phenotype of antibiotic and sucrose resistance. The colonies were growing into liquid media 7H9 and analyzed by PCR analysis. The PCR analysis determined the

correct size of the insert in the vector which confirmed the isolation of the colonies that were allelic exchange mutants, which integrated the disrupted gene into the chromosome. These observations demonstrate that the *Mtb* cells transformed with *grcC2* knock out were growing slowly with the interrupted gene, while the cells contained the construct of *grcC1* knock out did not grow, thus *grcC2* was not essential for bacterial survival while *grcC1* was essential. For *Msmeg* with the *ms1133* knock out, the cells grew which indicates that this gene is not essential for bacterial growth and survival.

Growth curves:

The growth curves for both wild type and *ms1133* knock out of *Msmeg* were obtained by growing the cell cultures over 100 hours. The results showed that in the *ms1133* gene knock out slowed the cells growth (increased lag time) compared to the wild type.

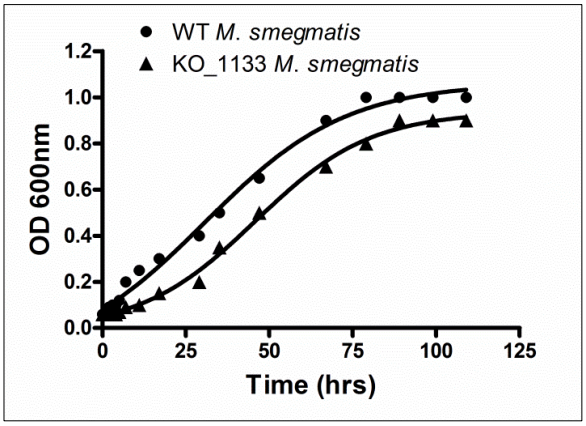


Figure 61: Wild type and KO *ms1133 Msmeg* growth curves. The growth curve were generated by OD_{600} nm measurement for both wild type and mutant cells cultured in 7h9 medium over 100 hours.

Also, the growth curve for wild type *Mtb* cells and the *grcC2* – knockout, to demonstrate the effect of the gene mutant on the growth of the *Mtb* cells since it did not affect the survival. Comparing with the wild type growth curve, the results showed that the gene knock out for *grcC2* slightly slowed down the bacterial growth under both aerobic and anaerobic growth conditions (Figure 62).

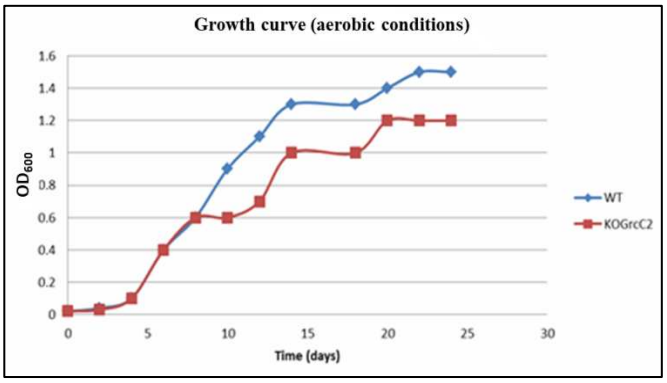


Figure 62: Wild type and KO*grcC2 M. tuberculosis* growth curves. Both wild type and mutant cells were cultured in 7h9 medium over 4 weeks. OD measured under aerobic conditions. The KO*grcC2* slowed down the bacterial growth.

Lipid extraction for menaquinone quantitation:

The wild type and mutated *Msmeg* cells were harvested, lipid extracted and the extracted ion chromatogram (EIC) were generated and analyzed using LC-MS (liquid chromatography-mass spectrometry). The lipid extract was analyzed to quantitate and compare the MK production in both wild type and mutant *Msmeg* cells (Figure 63). The presented data suggests the KO*ms1133* slowed down the bacterial growth and showed also slight reduction of MK production.

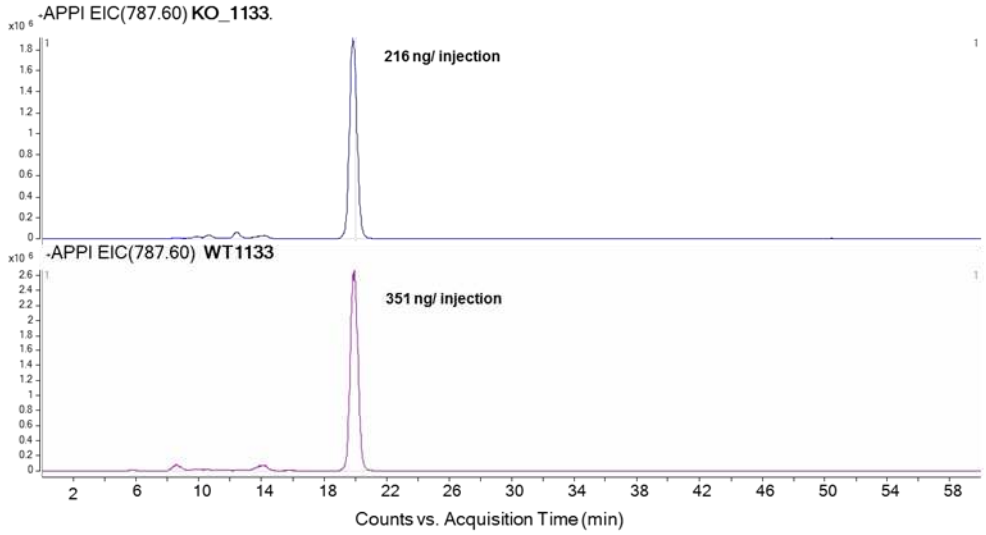


Figure 63: LC-MS chromatogram for MK (EIC) quantitation for both wild type and mutant *Msmeg*. Based on calculated [M+H]+ values of MK, KO*ms1133* synthesizes MK at a level near that of wild type *Msmeg*.

Chapter 7

Discussion and Conclusion

We successfully identified three genes encoding proteins constituting long-chain polyprenyl diphosphate synthases, and the products of GrcC1, GrcC2 and MS1133 were conclusively determined. GrcC1 and GrcC2 were predicted to belong to the gene family trans-hexaprenyltransferase, the annotated functions of which, are a medium chain polyprenyl diphosphate synthase for GrcC1 (*Mtb* database) and a short chain polyprenyl diphosphate synthase for GrcC2 (Mann, 2012). Our results showed that the three enzymes GrcC1, GrcC2 and MS1133 are long-chain prenyldiphosphate synthases that make products consistent with an intermediate in MK synthesis. The three synthases catalyze the chain elongation up to C45 starting from C5.

In addition, the assays for all three enzymes were optimized and the enzymes were characterized. The expression of the gene *grcC2* in *E. coli*, resulted in a good amount of protein which was purified but inactive. The gene expressed in competent *Msmeg* produced protein that was purified and active. This may be explained by the presence of regulatory genes that regulate the expression of *grcC2* in mycobacteria and those genes are not in *E. coli* (the lab expression host). The study of Fu, et al (2007) indicated that at the transcriptional level, a group of regulatory genes regulate other genes in the same gene cluster and this also is associated with the gene function (Zhang, 1998), so, *grcC2* the annotated *E*-isoprenyl diphosphate synthase may

possibly require another gene's product for catalysis during protein expression. Alternately, the expressed protein maybe misfolded.

The expressed and purified enzymes encoded by *grcC1*, *grcC2* in *Mtb* and *ms1133* in *Msmeg* were active and catalyzed the condensation reaction that added [^{14}C]IPP to allylic substrates of varying chain-lengths, including DMAPP, GPP, FPP, and GGPP. Enzyme assay conditions and activity requirements were determined. Our results showed that the divalent cation Mg^{2+} was required for enzyme activity for all three of the proteins and the optimal concentrations are shown in Table 4 below. The previous literature supports our results, prenyl diphosphate synthases including FPP, GGPP synthases and long-chain prenyl diphosphate synthases require divalent metal ions such as Mg^{2+} or Mn^{2+}, which are commonly required by all prenyltransferases (Ogura 1998; Liang 2002). Ogura's and Liang's studies showed that in *M. luteus* the products of solanesyl diphosphate synthase require specifically Mg^{2+} as a co-factor for enzyme activity and this supports our results for protein activity requirements. The effect of detergents on the enzyme activity was tested as well; both GrcC1 and GrcC2 require the presence of detergent in the assay, but MS1133 does not, which is unusual for long-chain prenyltransferases.

Table 4: Summarizes the optimal assay conditions and requirements for GrcC1, GrcC2 and MS1133 activity.

Assay condition	GrcC1	GrcC2	MS1133
Protein concentration / assay	100 ng	400 ng	300 ng
Time	10 min	15 min	7 min
Divalent cation (Mg^{++})	2 mM	2 mM	1 mM
Detergent	0.2% Triton X- 100	0.2 % CHAPS	----
Dithiothreitol (DTT)	-	-	2 mM
Buffering system and pH	MES (pH 6.5)	Tris (pH 8)	MOPS (pH 7.9)

The enzymes' natural substrates also were determined, and our results showed that GrcC1, GrcC2 and MS1133 can utilize GPP as a substrate, as well as other prenyl diphosphates (DMAPP, FPP & GGPP). The enzymes' affinity for these substrates vary; GrcC1 and GrcC2 showed greater affinity for GGPP and FPP as substrates (see Tables 1 and 2). GPP is likely the preferred substrate of GrcC1, FPP is likely the preferred substrate of GrcC2 and MS1133. Our results are supported by the previous studies in the literature; for polyprenyl diphosphate synthases that produce long-chain prenyl diphosphates, such as solanesyl diphosphate and hexaprenyl diphosphate, it is known that they do not prefer DMAPP as a substrate but require GPP, FPP, or GGPP as a priming substrate (Fujii, 1982; Sagami,1977; Baba ,1978; Ericsson, 1992; Ohnuma, 1998). Since the mid-1970s the substrate specificities of prenyl synthases have been investigated extensively and some studies indicated that FPP synthase utilizes a broad spectrum of artificial substrates to produce analogues of GGPP (Ogura 1998). The proteins GrcC1, GrcC2 and MS1133 generally preferred the longer chains allylic substrates of 10 or more carbon atoms. Multiple products were observed under assay conditions (C20, C35, C40, and C45).

The kinetic parameters for GrcC1 and GrcC2 were determined by calculating the values of V_{max}, K_m and K_{cat}, which identify them as individual enzymes. These indicate the characteristics of the enzymes and their substrate preferences. Since the role of the enzyme in the MK biosynthesis pathway has been suggested, elucidating the properties of these enzymes and obtaining more information on their biological role was important in demonstrating the involvement of the enzymes in MK biosynthesis. Kellogg & Poulter in 1997 demonstrated that SPS synthases are C45 *E*-prenyltransferase that catalyzes the addition of IPP to the allylic substrate FPP.

To determine the enzyme's product chain length, the dephosphorylated [C^{14}]radiolabeled products were analyzed on reverse-phase C18 TLC plates. This method helps to detect most of the radiolabeled products in the extract. By looking at the chain length of the final product, our results showed that GrcC1 is a SPS, which catalyzes the chain elongation reaction from C15 to C45. Also, GrcC2 is a long-chain prenyl diphosphate synthase that produces C45 products but its major product is C40 (octaprenyl diphosphate). In the presence of the preferred substrate; GrcC1 mainly produces C45 and GrcC2 C40. MS1133 also is a long-chain prenyl diphosphate synthase which catalyzes the chain elongation reaction from C20 to C45. The results of a study by Takahashi, 1980 showed that the SPS from *Micrococcus luteus* catalyzes the trans-chain elongation starting from C10 to C40 and C45 and it is suggested that when the chain elongation exceeds a specific chain length, two prenyl synthases may be involved (Takahashi, 1980), which may explain our results of having GrcC1 and GrcC2 producing the same products including long chain solanesyl diphosphate.

Our results identified GrcC2 as a long chain prenyl diphosphate synthase (OPS); however, the results of a study by the Peters' group in 2012 reported that GrcC2 is a short chain prenyl diphosphate synthase (GPP synthase). The results of the Peters' study are inconsistent not only with our results but with the bioinformatics database (in which GrcC2 is annotated as a heptaprenyl synthase). In long-chain *E*-polyprenyl-diphosphate synthases enzymes, the first asparate-rich motif (FARM) and the upstream fifth amino acid are important conserved regions in the amino acid sequence. These enzymes have product specificity that results from the position difference of the amino acids. The extreme importance of the amino acid location or position in determining the chain length of the enzyme's final product has been demonstrated (Ohnuma, 1996, Ohnuma, 1996).

A possible explanation for these differing results is the final product extraction procedure that Peters' group used. They extracted the enzyme product under acidic conditions, which likely broke down the long chain to short chain products. The studies of Collins, 1981; Dunphy, 1971; and Links, 1960 support this explanation. Allylic prenyl diphosphates are extremely sensitive to acids and breakdown into many products. Those studies and many others, revealed the effect of acid on breaking down prenyl diphosphate long-chain products (Dunphy, 1971). Under acidic conditions, the C55-isoprenyl diphosphate products isolated from *Micrococcus lysodeikticus* were completely hydrolyzed and new compounds appeared (Stone, 1972). For enzymatic dephosphorylation, a study by Fujii et al, 1982, reported that using potato acid phosphatase; all *E*-solanesyl diphosphate products were completely dephosphorylated to their respective alcohols.

In our study, knock-out mutations were used to test the essentiality of the genes *grcC1*, *grcC2* of *Mtb* and *ms1133* of *Msmeg* for the survival and growth. Our results demonstrated that *grcC1* is an essential gene for *Mtb* survival as predicted by transposon mutagenesis (DeJesus, 2017; Sun, 2001) and might be a good drug target. Neither *grcC2* nor *ms1133* are essential for survival, suggesting that neither may be involved in MK synthesis; however, the absence of these genes delays bacterial growth in *Mtb* and *Msmeg* somewhat. Since the characterized, non-essential gene *grcC2* promotes bacterial growth, it may be involved in a different intermediate metabolic role or in *Mtb* persistence since genes affecting growth of *Mtb* occur more frequently in non-essential pathways.

Our results have implications for developing new agents against TB. Here we identified GrcC1 and GrcC2 as solanesyl and octaprenyl diphosphate synthases, and we determined GrcC2 was not a GPP synthase as previously reported. Our results indicate that GrcC1 and GrcC2 are

polyprenyl diphosphate synthases which produce products that include mainly C40 and C45 using different allylic substrates. C45 produced by GrcC1 is the isoprene chain length required for MK biosynthesis in mycobacteria. In addition, our knock out mutation results supports the idea that *grcC1* is a conserved, essential gene and any changes in this gene could be lethal.

With all the research conducted for more than a century, TB continues to be the leading cause of death worldwide. Combining genomic and bioinformatics tools to identify mycobacterial genes potentially generate the knowledge that will help the development and design therapies that are needed to treat the airborne disease caused by *Mtb* (Sassetti, 2003). Identifying essential and non-essential genes for optimal growth holds promise for our results to feed into the design of new *Mtb* therapies, and the bioinformatics databases for both *Mtb* and *Msmeg*. Despite the fact that MK (vitamin K2) is an essential and uniqe supplement that human and animals require but can't synthesize, prenyl synthases in prokaryotes have significant homologues in eukaryotic organisms. Compounds which inhibit prenyl diphosphate synthases in bacteria may well show toxicity due to inhibition of analogus prenyl diphosphate synthases in eukaryotes.

The knowledge gained by identifying these three enzymes will open further needed investigations to fullfill the upcoming need to find new drug targets in the mycobacteria synthetic pathways.

References

- American Lung Association (ALA) (2013). Extensively Drug-Resistant Tuberculosis (XDR TB) Fact Sheet. Retrieved from http://www.lung.org/lung-disease/tuberculosis/factsheets/extensively-drug-resistant.html

- Andries, K., et al. (2005) A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science*. 14;307(5707):223-227.

- Baba, T., Allen, C. (1978). Substrate specificity of undecaprenyl pyrophosphate synthetase from Lactobacillus plantarum. *Biochemistry*.26;17(26):5598-604.

- Bald, D., et al. (2017). Targeting Energy Metabolism in Mycobacterium tuberculosis, a New Paradigm in Antimycobacterial Drug Discovery. *mBio*. 8. e00272-17. 10.1128/mBio.00272-17.

- Berney, M. and G. Cook. (2010). Unique flexibility in energy metabolism allows Mycobacteria to combat starvation and hypoxia. PLoS ONE.

- Beste, D. et al. (2009). The Genetic Requirements for Fast and Slow Growth in Mycobacteria. PLoS ONE 4(4): e5349. https://doi.org/10.1371/journal.pone.0005349

- Black, P. *et al*. (2014). Energy metabolism and drug efflux in Mycobacterium tuberculosis. *Antimicrobial agents and chemotherapy*, 58(5), 2491-503.

- Boshoff, H. and C. Barry. (2005) Tuberculosis — metabolism and respiration in the absence of growth. Nature reviews.

- Boucher, Y. & Doolittle, W. (2000). The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Molecular Microbiology*, 37(4), 703-716. doi.org/10.1046/j.1365-2958.2000.02004.x

- Broussard, G., (2009). Electrocompetent Mycobacterium Cells. https://gregorybroussard.com.

- Brown, A., et al. (2010). The Nonmevalonate Pathway of Isoprenoid Biosynthesis in *Mycobacterium tuberculosis* Is Essential and Transcriptionally Regulated by Dxs. *Journal of Bacteriology*, 192 (9) 2424-2433. DOI: 10.1128/JB.01402-09.

- Campos, N., et al. (2001). Identification of gcpE as a novel gene of the 2-C-methyl-D-erythritol 4-phosphate pathway for isoprenoid biosynthesis. *FEBS* letters 488: 170-173

- Centers for Disease Control and Prevention. (2006). Emergence of *Mycobacterium tuberculosis* with extensive resistance to second-line drugs—worldwide, 2000–2004. MMWR Morb. Mortal. Wkly. Rep. 2006, 55, 301–305. *Chemical Reviews*, *98* (4): 1263-1276. DOI: 10.1021/cr9600464

- Chen, M., et al. (2013). Identification of a hotdog-fold thioesterase involved in the biosynthesis of menaquinone in *Escherichia coli*. *Journal of Bacteriology*. DOI: 10.1128/JB.00141-13

- Chen, A., Dale Poulter, C. and Kroon, P.A. (1994), Isoprenyl diphosphate synthases: Protein sequence comparisons, a phylogenetic tree, and predictions of secondary structure. Protein Science, 3: 600-607. https://doi.org/10.1002/pro.5560030408

- Cole, S., et al. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.*Nature*, volume 393, 537–544.

- Collins, M. D., & Jones, D. (1981). Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implication. *Microbiological Reviews*, 45(2), 316–354.

- Cook, G. *et al*. (2014). Energetics of Respiration and Oxidative Phosphorylation in Mycobacteria. *Microbiology spectrum*, 2(3), 10.1128/microbiolspec.MGM2-0015-2013.

- Cook, G. *et al*. (2017). Oxidative Phosphorylation as a Target Space for Tuberculosis: Success, Caution, and Future Directions. *Microbiology spectrum*. https://doi.org/10.1128/microbiolspec.TBTB2-0014-2016

- Cordone, A., et al. (2011). Characterization of a *Mycobacterium smegmatis* uvrAmutant impaired in dormancy induced by hypoxia and low carbon concentration. *BMC Microbiol*. https://doi.org/10.1186/1471-2180-11-231

- Crick, D. et al. (2000) Polyprenyl Phosphate Biosynthesis in *Mycobacterium Tuberculosis* and *Mycobacterium Smegmatis*." *Journal of Bacteriology* 182: 5771–5778

- Crofton,J. and D. Mitchison, (1948) Streptomycin Resistance in Pulmonary Tuberculosis. Br Med J.; 2(4588): 1009–1015. PMCID: PMC2092236

- Dam H, Schønheyder F. (1936). The occurrence and chemical nature of vitamin K. Biochem J. 30(5):897-901. doi: 10.1042/bj0300897.

- Daniel, T. (2006). The history of tuberculosis. *Respiratory Medicine*: 100, 1862–1870 DOI: https://doi.org/10.1016/j.rmed.2006.08.006

- Debnath, J. et al. (2012). Discovery of Selective Menaquinone Biosynthesis Inhibitors against Mycobacterium tuberculosis. J. Med. Chem., 55.

- Dewick, M. (1995). The biosynthesis of C5-C20 terpenoid compounds. *Nat. Prod. Rep*.12, 507-534

- Dhiman RK, et al. (2019). Characterization of MenA (isoprenyl diphosphate:1,4-dihydroxy-2-naphthoate isoprenyltransferase) from *Mycobacterium tuberculosis*. *PLoS ONE* 14(4): e0214958. https://doi.org/10.1371/journal.pone.0214958

- Dhiman, R, et al (2009). Menaquinone synthesis is critical for maintaining mycobacterial viability during exponential growth and recovery from non-replicating persistence. Mol Microbiol.;72(1):85-97.

- Dhiman, R. et al. (2005). 1-Deoxy-d-Xylulose 5-Phosphate Reductoisomerase (IspC) from Mycobacterium tuberculosis: towards Understanding Mycobacterial Resistance to Fosmidomycin. *Journal of Bacteriology* 187 (24) 8395-8402 DOI: **10.1111/j.1365-2958.2009.06625.x**

- Dormandy, T. (2000). The white death, a history of Tuberculosis. New York university press.

- Dunphy, P. , G. David, P. Philip & B. Arnold. (1968). A New Natural Naphthoquinone in *Mycobacterium phlei*. Journal of Biological Chemistry. 243. 398-407.

- Eoh, H., Brennan, P. J., & Crick, D. C. (2009). The Mycobacterium tuberculosis MEP (2C-methyl-D-erythritol 4-phosphate) pathway as a new drug target. Tuberculosis (Edinburgh, Scotland), 89(1), 1–11.

- Ericsson, J. et al. (1992). Substrate specificity of cis-prenyltransferase in rat liver microsomes. *The Journal of Biological Chemistry* 267, 19730-19735.

- Fu, L. and Fu-Liu, C. (2007). The gene expression data of *Mycobacterium tuberculosis* based on Affymetrix gene chips provide insight into regulatory and hypothetical genes. BMC Microbiology20077:37. https://doi.org/10.1186/1471-2180-7-37

- Fujii, H., Koyama, T., and Ogura, K. (1982) Hexaprenyl pyrophosphate synthetase from Micrococcus luteus B-P 26. Separation of two essential components. *The Journal of Biological Chemistry,* 257, 14610-14612.

- Fujii, H., Koyama T., and Ogura K. (1982). Efficient enzymatic hydrolysis of polyprenyl pyrophosphates. *Biochimica et Biophysica Acta* (BBA) - *Lipids and Lipid Metabolism*: 716-718. https://doi.org/10.1016/0005-2760(82)90304-6.

- Gandhi,R., et al. (2006). Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. Lancet, 368, 1575–1580.

- Gupta, S., *et al.* (2018). Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus *Mycobacterium* into an Emended Genus *Mycobacterium* and Four Novel Genera. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2018.00067

- Gygli, S., *et al*. (2017). Antimicrobial resistance in *Mycobacterium tuberculosis*: mechanistic and evolutionary perspectives, *FEMS Microbiology Reviews*, 354 373. https://doi.org/10.1093/femsre/fux011

- Haddock, B. A., & Jones, C. W. (1977). Bacterial respiration. *Bacteriological Reviews*, 41(1), 47–99.

- Halder M, *et al*. (2019). Vitamin K: Double Bonds beyond Coagulation Insights into Differences between Vitamin K1 and K2 in Health and Disease. Int J Mol Sci. 20(4):896. doi: 10.3390/ijms20040896.

- Heider, S. A., Wolf, N., Hofemeier, A., Peters-Wendisch, P., & Wendisch, V. F. (2014). Optimization of the IPP Precursor Supply for the Production of Lycopene, Decaprenoxanthin and Astaxanthin by *Corynebacterium glutamicum*. *Frontiers in bioengineering and biotechnology, 2, 28*. https://doi.org/10.3389/fbioe.2014.00028

- Hemmi, H., Ikejiri, S., Yamashita, S., & Nishino, T. (2002). Novel medium-chain prenyl diphosphate synthase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. *Journal of bacteriology*, 184(3), 615–620. https://doi.org/10.1128/jb.184.3.615-620.2002

- Hershkovitz ,I., *et al*. (2008) . Detection and Molecular Characterization of 9000-Year-Old *Mycobacterium tuberculosis* from a Neolithic Settlement in the Eastern Mediterranean. PLOSONE: http://dx.doi.org/10.1371/journal.pone.0003426

- Hirooka, K., et al. (2000). Mechanism of product chain length determination for heptaprenyl diphosphate synthase from Bacillus stearothermophilus. *European Journal of Biochemistry*, 267: 4520-4528. doi:10.1046/j.1432-1327.2000.01502.x

- Jackson, M. *et al*. (2002). Gene replacement and transposon delivery using the negative selection marker *sac*B. From: methods in molecular medicine, vol, 54: *Mycobacterium tuberculosis* protocols. Edited by: T. Parish and G. Stoker. Humana Press.

- Kellogg, B. & Poulter, C. (1997) Chain elongation in the isoprenoid biosynthetic pathway. Current opinion in chemical biology (1): 570-578.

- Keshavjee, S., and P. Farmer. (2012) Tuberculosis, drug resistance, and the history of modern medicine. *New England Journal of Medicine.*

- Kharel , Y. et al (2006) Manipulation of prenyl chain length determination mechanism of cis-prenyltransferases. FEBS Journal 273 : 647–657

- Kollas, A., et al. (2002). Functional characterization of GcpE, an essential enzyme of the non-mevalonate pathway of isoprenoid biosynthesis. *FEBS Letters* 532: 1873-3468. doi.org/10.1016/S0014-5793(02)03725-0

- Kurosu, M., & Crick, D. C. (2009). MenA Is a Promising Drug Target for Developing Novel Lead Molecules to Combat *Mycobacterium tuberculosis*. Medicinal Chemistry (Shariqah (United Arab Emirates), 5(2), 197–207.

- Kurosu, M., et al. (2007). Discovery of 1,4-Didydroxy-2-naphthoate Prenyltransferase Inhibitors: New Drug Leads for Multidrug-Resistant Gram-Positive Pathogens. *Journal of Medicinal Chemistry*, 50(17), 3973–3975. http://doi.org/10.1021/jm070638m

- Kurosu, M.; Begari, E. (2010). Vitamin K2 in Electron Transport System: Are Enzymes Involved in Vitamin K2 Biosynthesis Promising Drug Targets? *Molecules*, *15*, 1531-1553.

- Layre, E. et al. (2014). Molecular profiling of Mycobacterium tuberculosis identifies tuberculosinyl nucleoside products of the virulence-associated enzyme Rv3378c. PNAS (8): 2978-2983

- Li, K. et al. (2014). Multitarget Drug Discovery for Tuberculosis and Other Infectious Diseases. *Journal of Medicinal Chemistry 57* (7), 3126-3139. DOI: 10.1021/jm500131s

- Liang, P., et al (2002) Structure, mechanism and function of prenyltransferases. *Eur. J. Biochem*. 269, 3339-3354.

- Lim, A. and Dick, T. (2001). Plate-based dormancy culture system for *Mycobacterium smegmatis* and isolation of metronidazole-resistant mutants. *FEMS Microbiology Letters*. https://doi.org/10.1111/j.1574-6968.2001.tb10718.x

- Mann, F., et al. (2012). Functional characterization and evolution of the isotuberculosinol operon in *Mycobacterium tuberculosis* and related mycobacteria. *Front. Microbiol*. https://doi.org/10.3389/fmicb.2012.00368

- Mann, F., et al. (2011). Rv0989c encodes a novel (E)-geranyl diphosphate synthase facilitating decaprenyl diphosphate biosynthesis in *Mycobacterium tuberculosis*, FEBS Letters, 585. https://doi:10.1016/j.febslet.2011.01.007

- Matsoso, L. et al. (2005). Function of the cytochrome bc1-aa3 branch of the respiratory network in mycobacteria and network adaptation occurring in response to its disruption. *Journal of bacteriology*,187: 6300–6308.

- Mdluli, K., Kaneko, T. & Upton,A. (2015). The Tuberculosis Drug Discovery and Development Pipeline and Emerging Drug Targets. *Cold Spring Harb Perspect Med* .

- Meganathan, R., & Kwon, O. (2009). Biosynthesis of Menaquinone (Vitamin K2) and Ubiquinone (Coenzyme Q). *EcoSal Plus*, *3*(2), 10.1128/ecosalplus.3.6.3.3. https://doi.org/10.1128/ecosalplus.3.6.3.3

- Mitchell, P. (1979). Keilin's respiratory chain concept and its chemiosmotic consequences. *Science*, 206: 1148-1159

- Mitchell, P. (2011). Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. *Biochimica et Biophysica Acta*, 1807: 1507-1538. https://doi.org/10.1016/j.bbabio.2011.09.018

- Muramatsu, M., et al. (1997). Prenyl diphosphate synthetase genes. *European Patent Application*: EP0812914A2

- Nagel, R., et al. (2018). Arginine in the FARM and SARM: A Role in Chain-Length Determination for Arginine in the Aspartate-Rich Motifs of Isoprenyl Diphosphate Synthases from *Mycobacterium tuberculosis*. *Molecules*, 23(10), 2546; https://doi.org/10.3390/molecules23102546

- Narayanasamy, P. Eoh, H. Brennan, P. & Crick, D. (2010). Synthesis of 4-Diphosphocytidyl-2-C-Methyl-D-Erythritol 2-Phosphate and Kinetic Studies of *Mycobacterium tuberculosis* IspF.*Chemistry & Biology*. (17): 117-122. https://doi.org/10.1016/j.chembiol.2010.01.013

- Nowicka , B. and Kruk , J. (2010) Occurrence, biosynthesis and function of isoprenoid quinones. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1797: 1587-1605

- Ogura, K. & Koyama, T. (1998). Enzymatic Aspects of Isoprenoid Chain Elongation. *Chem. Rev*., 98, 4, 1263–1276. acs.org

- Ohara, K., et al. (2010). Two solanesyl diphosphate synthases with different subcellular localizations and their respective physiological roles in Oryza sativa. *Journal of experimental botany*, 61(10), 2683-92.

- Ohnuma, S. et al. (1996). A role of the amino acid residue located on the fifth position before the first aspartate-rich motif of farnesyl diphosphate synthase on determination of the final product. *J Biol Chem*. 271(48):30748-54.

- Ohnuma, S., et al. (1996). Conversion from farnesyl diphosphate synthase to geranylgeranyl diphosphate synthase by random chemical mutagenesis. *The Journal of Biological Chemistry*. 271(17):10087-95.

- Ohnuma, S., et al. (1998). A Pathway Where Polyprenyl Diphosphate Elongates in Prenyltransferase. The Journal of Biological Chemistry .273: 26705-26713. doi: 10.1074/jbc.273.41.26705

- Oldfiel, E. (2015). Tuberculosis Terpene Targets. *Chemistry and Biology* (22): 437-439.

- Oldfield, E. and Lin, F. (2012), Terpene Biosynthesis: Modularity Rules. *Angew. Chem*. Int. Ed., 51: 1124-1137. doi:10.1002/anie.201103110

- Palomino JC, and Martin A. (2014). Drug Resistance Mechanisms in *Mycobacterium tuberculosis*. *Antibiotics* (Basel), (3):317-40. DOI: 10.3390/antibiotics3030317

- Pandey, A. & C. Sassetti. (2008). Mycobacterial persistence requires the utilization of host cholesterol. *Proceedings of the National Academy of Sciences* 105 (11) 4376-4380; https://doi.org/10.1073/pnas.0711159105

- Pedro, E. Andrea, G. Martin, A. and JuanCarlos, P. (2011) Efflux as a mechanism for drug resistance in Mycobacterium tuberculosis. FEMS Immunology & Medical Microbiology, Volume 63: (1–9).

- Poulter, C. & Rilling, H. (1976) Prenyltransferase: the mechanism of the reaction. *Biochemistry.* 15(5):1079-83.

- Ratledge, C. and Stanford, J. (1982). The biology of the Mycobacteria, volume 1: Physiology, Identification and Classification. Academic Press.

- Richardson, D. and G. Sawers. (2002). PMF through the redox loop. Science.

- Rip, J. et al. (1985) Distribution, metabolism and function of dolichol and polyprenols. *Prog. Lipid Res*. Vol.24: 269-309.

- Rodríguez-Concepción, Manuel & Boronat, Albert. (2002). Elucidation of the Methylerythritol Phosphate Pathway for Isoprenoid Biosynthesis in Bacteria and Plastids. A Metabolic Milestone Achieved through Genomics. *Plant physiology*. 130. 1079-89. 10.1104/pp.007138.

- Sacchettini, J. & Poulter, D. (1997). Creating Isoprenoid Diversity. Science: 277, 1788-1789. DOI: 10.1126/science.277.5333.1788

- Sagami,H., Ogura, K. and Seto, S. (1977). Solanesyl pyrophosphate synthetase from Micrococcus lysodeikticus . *Biochemistry*, 16 (21): 4616-4622; DOI: 10.1021/bi00640a014

- Sassetti, M., Boyd, .H. and Rubin, .J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology*, 48: 77-84. doi:10.1046/j.1365-2958.2003.03425.x

- Scherman , M. et al. (2012). Screening a library of 1,600 adamantyl ureas for anti-Mycobacterium tuberculosis activity in vitro and for better physical chemical properties for bioavailability. *Bioorg Med Chem*.; 20(10): 3255–3262.

- Schulbach, M. (2000). Isoprenoid chain elongation in *Mycobacterium tuberculosis*. PhD thesis, Colorado state university.

- Shamaei M, *et al*. (2009). First-line anti-tuberculosis drug resistance patterns and trends at the national TB referral center in Iran--eight years of surveillance. *Int J Infect Dis*.; 13(5):e236-40. doi: 10.1016/j.ijid.2008.11.027.

- Shineberg , B. and Young, I. (1976). Biosynthesis of bacterial menaquinones: the membrane-associated 1,4-dihydroxy-2-naphthoate octaprenyltransferase of *Escherichia coli.* Biochemistry 15 (13), 2754-2758 .DOI: 10.1021/bi00658a007.

- Smith, I. (2003). *Mycobacterium tuberculosis* Pathogenesis and Molecular Determinants of Virulence. Clinical Microbiology Reviews, 16 (3) 463-496; DOI: 10.1128/CMR.16.3.463-496.2003

- SIGMA-ALDRICH, Anisaldehyde soluton (SRA1) - Datasheet (sigmaaldrich.com).

- Soballe, B., Poole, R. (1999) Microbial ubiquinones: multiple roles in respiration, gene regulation and oxidative stress management. *Microbiology*, 145: 1817-1830.

- Stone, J., et al. (1972). Isolation of C55-Isoprenylpyrophosphate from *Micrococcus lysodeikticu*. *Journal of Biological Chemistry*, Volume 247, Issue 16, 5107 – 5112. doi: https://doi.org/10.1016/S0021-9258(19)44945-4

- Sun, Z., et al (2001;). Salicylate uniquely induces a 27-kDa protein in tubercle bacillus. *FEMS Microbiol Lett* 203, 211–216.

- Takahashi, I., *et al*. (1980). Heptaprenyl pyrophosphate synthetase from *Bacillus subtilis*. J Biol Chem.: 255(10):4539-43.

- Takahashi, I. and Ogura, K. (1982). Prenyltransferases of *Bacillus subtilis*: Undecaprenyl Pyrophosphate Synthetase and Geranylgeranyl Pyrophosphate Synthetase. *The Journal of Biochemistry*, (92): 1527-1537

- Takahashi,S. , et al . (1998).A 1-deoxy-d-xylulose 5-phosphate reductoisomerase catalyzing the formation of 2-C-methyl-d-erythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis. *Proceedings of the National Academy of Sciences*, 95 (17): 9879-9884; DOI: 10.1073/pnas.95.17.9879

- Tran, S., & Cook, G. (2005). The F1Fo-ATP synthase of Mycobacterium smegmatis is essential for growth. Journal of bacteriology, 187(14), 5023–5028. doi:10.1128/JB.187.14.50235028.2005.

- TubercuList web server, http://genolist.pasteur.fr/TubercuList/

- Upadhyay, A. et al. (2015). Partial Saturation of Menaquinone in *Mycobacterium tuberculosis*: Function and Essentiality of a Novel Reductase, MenJ. *ACS Central Science*. DOI: 10.1021/acscentsci.5b00212

- Vandermoten, S. *et al*. (2009). Structural features conferring dual Geranyl/Farnesyl diphosphate synthase activity to an aphid prenyltransferase. *Insect Biochemistry and Molecular Biology*, 39: 707-716. https://doi.org/10.1016/j.ibmb.2009.08.007

- Wang, K. and Ohnuma, S. (2000). Isoprenyl diphosphate synthases. *Biochimica et Biophysica Acta* (BBA) - Molecular and Cell Biology of Lipids,Volume 1529, 33-48. doi.org/10.1016/S1388-1981(00)00136-0.

- Wang, X., & Dowd, C. S. (2018). The Methylerythritol Phosphate Pathway: Promising Drug Targets in the Fight against Tuberculosis. ACS infectious diseases, 4(3), 278–290. doi.org/10.1021/acsinfecdis.7b00176

- Watanabe, S., et al. (2011). Fumarate reductase activity maintains an energized membrane in anaerobic *Mycobacterium tuberculosis*. *PLoS pathogens*, 7(10), e1002287.

- Weinstein, E. A., *et al*. (2005). Inhibitors of type II NADH:menaquinone oxidoreductase represent **a** class of antitubercular drugs. Proceedings of the National Academy of Sciences of the United States of America, 102(12), 4548–4553. http://doi.org/10.1073/pnas.0500469102

- www. STRING-db.org

- Yasuhiro, a. (1988). Bacterial electron transport chains. *Ann. Rev. Biochem*. (57): 101-312.

- Ye, Y., et al. (2007). Geranylgeranyl diphosphate synthase in fission yeast is a heteromer of farnesyl diphosphate synthase (FPS), Fps1, and an FPS-like protein, Spo9, essential for sporulation. *Molecular biology of the cell*, 18(9), 3568-81.

- Zhang, Y. et al. (1998). Two Subunits of Heptaprenyl Diphosphate Synthase of *Bacillus subtilis* Form a Catalytically Active Complex. *Biochemistry*, 37 (38), 13411-1342 DOI: 10.1021/bi972926y

LIST OF ABBREVIATIONS

CDP-ME, 4-diphosphocytidyl-2-C-methyl-D-erythritol

DMAP, dimethylallyl diphosphate

DXP, 1-deoxy-D-xylulose 5-phosphohate

DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase

DXS, 1-deoxy-D-xylulose 5-phosphate synthase

EIC, extracted ion chromatogram

ETC, electron transport chain

FARM, the first aspartate rich motif

FPP, farnesyl diphosphate

GGPP, geranylgeranyl diphosphate

GGPS, geranylgeranyl diphosphate synthase

GPP, geranyl diphosphate

HepPS, heptaprenyl diphosphate synthase

HexPP, hexaprenyl diphosphate

HexPS, hexaprenyl diphosphate synthase

IP, isopentenyl monophosphate

IPP, isopentenyl diphosphate

IPTG, isopropyl 1-thio-D-galactoside

MDR-TB, multi-drug resistant tuberculosis

MEP, 2C-methyl-erythritol 4-phosphate pathway

MK, menaquinone

Msmeg, *Mycobacterium smegmatis*

Mtb, *Mycobacterium tuberculosis*

MVA, mevalonate pathway

OPS, octaprenyl diphosphate synthase

SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

SPS, solanesyl diphosphate synthase

TB, Tuberculosis

TLC, thin-layer chromatography

XDR-TB, extensively drug-resistant tuberculosis