NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF CHEMISTRY

BÁRBARA FERREIRA RODRIGUES

BSc in Biochemistry

# STUDIES ON MYCOBACTERIA GLYCOSYLTRANSFERASES, PROMISING TARGETS TO DEVELOP ANTITUBERCULOSIS DRUGS

MASTER IN BIOTECHNOLOGY

NOVA University Lisbon
September, 2023

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF CHEMISTRY

# STUDIES ON MYCOBACTERIA GLYCOSYLTRANSFERASES, PROMISING TARGETS TO DEVELOP ANTITUBERCULOSIS DRUGS

**BÁRBARA FERREIRA RODRIGUES**

BSc in Biochemistry

**Adviser:** Margarida Archer
*ITQB NOVA*

**Co-adviser:** Maria João Catalão
*FFUL, iMEDULisboa*

**Examination Committee**

**Chair:** Sérgio Joaquim Raposo Filipe
*Associate Professor, NOVA University Lisbon*

**Rapporteur:** João Filipe Bogalho Vicente
*Researcher, ITQB NOVA*

MASTER IN BIOTECHNOLOGY

NOVA University Lisbon
September, 2023

**Studies on mycobacteria glycosyltransferases, promising targets to develop antituberculosis drugs**

*"*

*Segue o teu destino,*
*Rega as tuas plantas,*
*Ama as tuas rosas.*
*O resto é a sombra*
*De árvores alheias.*

*"*

**-Fernando Pessoa**

# Agradecimentos

Este projeto é o reultado de várias horas de trabalho e não seria possível sem a ajuda de várias pessoas, que direta ou indiretamente me apoiaram, e a quem eu gostaria de agradecer:

À Dra. Margarida Archer e à Dra. Maria João Catalão pela oportunidade e confiança em mim para abraçar este projeto ao longo deste ano.

À Cátia, por todo o apoio e ajuda incansável desde o primeiro dia, e à Vanessa por mesmo estando longe ter dado o seu melhor para me ajudar. Um obrigada às duas por nunca se cansarem de me ajudar e motivar, especialmente nos dias em que as experiências corriam inevitavelmente mal.

Ao Federico e ao Cristiano, por toda a disponibilidade e paciência que tiveram comigo.

A todas as pessoas do ITQB e da FFUL que de alguma forma se cruzaram comigo e me ajudaram no meu trabalho.

Aos meus pais e à Tilde por serem o meu maior apoio faça eu o que fizer.

À Matilde, Joana e Helena, por todos os cafés na Mini Nova e por terem tornado estes 2 anos um bocadinho mais suportáveis.

Aos meus amigos, Inês, Cance, Barros, Brás, Tixa, Chico, João e Manel, por estarem presentes não só neste último ano, mas em todos desde que vos chamo de amigos, em especial pela capacidade infinita de me ouvir sem se cansarem (pelo menos disfarçam bem).

À Mafalda, que nunca me deixa desistir e tornou tudo melhor e mais bonito, o meu percurso na Nova nunca seria o que foi sem ti e ainda bem que me permitiu conhecer-te.

À Chica, por ser o meu braço direito nestes últimos 5 anos e acreditar sempre em mim, principalmente nos dias em que eu não acredito.

Por fim, ao Manel, por ser o meu melhor amigo e o meu maior apoio todos os dias. A tese, assim como todos os desafios que me aparecem, tornam-se mais fáceis contigo ao meu lado.

# Abstract

Tuberculosis (TB), caused by the pathogen *Mycobacterium tuberculosis* (*Mtb*) continues to be a major infectious disease responsible for global mortality. Although treatment exists, resistance to first- and second-line drugs is a well-known global public health problem. The emergence of multi- and extensive-drug resistant strains of *Mtb* reflects the urgent need to develop new drugs and effective treatment approaches. The success of *Mtb* is mostly due to the complexity of its cell wall (CW) structure, well-known as the mycolyl-arabino-galactan-peptidoglycan (mAGP) complex. Arabinofuranosyltransferases (EmbA-C and AftA-D) play a crucial role in the biosynthesis of arabinogalactan, being vital for mycobacterial growth and survival. The biosynthesis of arabinogalactan presents some promising targets for anti-TB drugs. The arabinosyl transferases Emb are inhibited by ethambutol (EMB), a first-line antibiotic used to treat TB, therefore compromising CW integrity. Mutations in the *embB* gene have been linked to EMB resistance, while Afts appear as new distinct targets for the creation of innovative TB treatments. Therefore, the main goal of this thesis was the study of EmbB, AftA and AftD, to further understand their essentiality and their role in mycobacterial fitness and drug resistance. First, knockdown mutants of *embB*, *aftA* and *aftD* were generated in the bacterial study model, *Mycobacterium smegmatis* (*M. smegmatis*), using the clustered regularly interspaced short palindromic repeat (CRISPR) interference (CRISPRi) system. CRISPRi is based on the anhydrotetracycline (ATc)-inducible expression and relies on the small single guide RNA (sgRNA), that is specific to a target gene sequence, together with the enzymatically inactive Cas9 (dCas9) protein. Using this system, *embB* and *aftA* were confirmed to be essential in *M. smegmatis*, and the level of *embB* and *aftA* knockdown seemed to be directly related with the strength of the associated proto-spacer adjacent motif (PAM). The *aftD* knockdown mutant did not show any viability defect phenotype and, for that reason, no further studies were carried out with this mutant. The basal gene expression, the distance of the target gene region from the transcription start site and the role of the target genes in the biosynthesis of arabinogalactan may also influence CRISPRi-mediated gene silencing and, consequently, the associated phenotypes. The level of knockdown was also assessed by quantitative real-time polymerase chain reaction (qRT-PCR), and a significant decrease in *embB* and *aftA* mRNA expression level was obtained, in

agreement with previous reports, there was evidence of a downstream polar effect when the *aftA* gene was silenced by CRISPRi. In this work, *M. smegmatis* competent cells were also transformed with the pVVAP-AftA expression vector. A prediction of the structure of AftA in *Mtb* was obtained using the artificial intelligence program AlphaFold and, docking was performed to predict the complementarity of the protein binding sites with several compounds, which were selected based on a drug repurposing strategy. A total of four compounds were selected. The effectiveness of each compound against AftA from *M. smegmatis* was assessed through minimum inhibitory concentration (MIC), minimum bactericidal concentration (MBC) assays and growth curves. All the compounds tested showed very high MIC values, except for compound A. This compound had a reasonable MIC value and proved to be bactericidal. In parallel with the work performed with AftA, this project at ITQB also aimed to purify and performed activity assays with the AftD protein. Although, the protein purification attempts were not successful, promising results were obtained in the activity assays. The results of the activity assays were analyzed by NMR spectroscopy, where the arabinofuranose transfer from $^{13}$C-labeled donor to acceptor can be followed by changes of the chemical shifts on NMR $^{13}$C-spectrum.

**Keywords:** Tuberculosis, Mycobacterial cell wall, Arabinofuranosyltransferases, Antibiotic resistance, Drug repurposing

# Resumo

A tuberculose (TB), causada pelo agente patogénico *Mycobacterium tuberculosis* (*Mtb*), continua a ser uma das principais doenças infeciosas responsável por milhões de mortes em todo o mundo. Embora a tuberculose seja uma doença infeciosa para a qual já existe um tratamento, a resistência aos medicamentos de primeira e segunda linha é um conhecido problema de saúde pública mundial. O aparecimento de estirpes de *Mtb* multi ou extensivamente resistentes, reflete a urgente necessidade de desenvolver novos medicamentos e abordagens de tratamento eficazes. O sucesso de Mtb deve-se sobretudo à complexidade da estrutura da sua parede celular (PC), comumente conhecida como o complexo micolil-arabino-galactano-peptidoglicano (mAGP). As arabinofuranosiltransferases (EmbA-C e AftA-D) desempenham um papel crucial na biossíntese do arabinogalactano, sendo vital para o crescimento e sobrevivência das micobactérias. A biossíntese do arabinogalactano apresenta alguns alvos promissores para fármacos contra a TB. As arabinosiltransferases Emb são inibidas pelo etambutol (EMB), um antibiótico de primeira linha utilizado no tratamento da TB, comprometendo assim a integridade da PC. A resistência ao EMB tem sido associada a mutações no gene *embB*, enquanto as Afts parecem ser um novo alvo distinto para o desenvolvimento de novas terapêuticas para a TB. Desta forma, o principal objetivo desta tese de mestrado foi o estudo das arabinosiltransferases EmbB, AftA e AftD, para melhor compreender a sua essencialidade e função nas micobactérias. Primeiro, foram construídos mutantes knockdown dos genes *embB*, *aftA* e *aftD* no organismo de estudo *Mycobacterium smegmatis* (*M. smegmatis*), através de um sistema de interferência de repetições palindrómicas curtas regularmente espaçadas (CRISPRi). CRISPRi é um sistema que se baseia na expressão induzida pela anidrotetraciclina (ATc) e depende de um pequeno RNA guia (sgRNA), que será específico para uma sequência genética alvo, juntamente com a proteína Cas9 enzimaticamente inativa (dCas9). Através deste sistema confirmou-se que os genes *embB* e *aftA* são essenciais para a sobrevivência de *M. smegmatis*, e ainda que a eficiência do knockdown parece estar diretamente relacionada com a força do motivo adjacente ao proto-espaçador (PAM) associado. Os mutantes knockdown do gene *aftD* apresentaram um fenótipo de baixa viabilidade, por isso, não foram realizados mais estudos com estes mutantes. A expressão génica basal, a distância da região alvo

do gene ao local de iniciação da transcrição e o papel dos genes alvo na biossíntese do arabinogalactano parecem também influenciar o silenciamento de genes mediado por CRISPRi e, consequentemente, os fenótipos associados. Os knockdowns foram ainda avaliados por uma reação em cadeia da polimerase quantitativa em tempo real (qRT-PCR), tendo-se obtido uma diminuição significativa do nível de expressão do mRNA de *embB* e *aftA*. Em concordância com estudos anteriores, houve evidência de um efeito polar a jusante quando o gene *aftA* foi silenciado através do sistema CRISPRi. Neste projeto, células competentes de *M. smegmatis* foram ainda transformadas com o vetor de expressão pVVAP-AftA. Para além disto, através de um programa de inteligência artificial designado AlphaFold, foi possível obter uma previsão da estrutura da AftA de *Mtb*. Assim sendo, um estudo de docking foi realizado para prever a complementaridade dos locais de ligação da proteína a possíves compostos. Quatro compostos foram, então, selecionados com base numa estratégia de reaproveitamento de fármacos. A eficiência de cada composto na ação da proteína AftA em *M. smegmatis* foi determinada através dos ensaios da concentração mínima inibitória (CMI), concentração bactericida mínima (CBM) e ainda através de curvas de crescimento. Todos os compostos testados demonstraram valores de CMI elevados, com a exceção do composto A. Além de apresentar um valor de CMI razoável, este composto mostrou ainda possuir atividade bactericida. Em paralelo com o trabalho realizado com a proteína AftA, este projeto no ITQB também tinha como objetivo purificar e realizar ensaios de atividade com a proteína AftD. Embora as tentativas de purificação da proteína não tenham sido bem sucedidas, foram obtidos resultados promissores nos ensaios de atividade. Os resultados dos ensaios de atividade foram analisados através de espectroscopia de $^{13}$C-RMN, onde as alterações nos espetros são observadas através de desvios químicos, uma vez que se trata de um espetro de carbono marcado, pois dador tem grupo arabinofuranosil marcado com $^{13}$C.

**Palavras-chave:** Tuberculose, Parede celular micobacteriana, Arabinofuranosiltransferases, Resistência a antibióticos, Reaproveitamento de fármacos

# Contents

# List of Figures

# List of Tables

# Acronyms

| | |
|---|---|
| ***B. subtilis*** | *Bacillus subtilis (p. 17)* |
| ***M. smegmatis*** | *Mycobacterium smegmatis (p. 4)* |
| **Acp** | acyl carrier protein *(p. 10)* |
| **AG** | arabinogalactan *(p. 4)* |
| **Ala** | alanine *(p. 39)* |
| **apoB** | apolipoprotein B *(p. 14)* |
| **AraT** | arabinofuranosyltransferase *(p. 8)* |
| **AsnB** | glutamine amidotransferase *(p. 6)* |
| **ATc** | anhydrotetracycline *(pp. 17, 20)* |
| **BCG** | Bacille Calmette-Guérin *(p. 2)* |
| **CCC** | chemically competent cells *(p. 22)* |
| **CRISPR** | clustered regularly interspaced short palindromic repeat *(p. 16)* |
| **CRISPRi** | CRISPR interference system *(p. 16)* |
| **crRNA** | CRISPR RNA *(p. 16)* |
| **cryo-EM** | single-particle cryogenic electron microscopy *(p. 10)* |
| **CW** | cell wall *(p. 4)* |
| **dCas9** | endonuclease-deficient Cas9 *(p. 16)* |
| **dCas9**$_{Spy}$ | nuclease-dead *S. pyogenes* Cas9 protein *(p. 16)* |
| **dCas9**$_{th1}$ | dCas9 from *S. thermophilus (p. 17)* |
| **DNA** | deoxyribonucleic acid *(p. 2)* |
| **DOTS** | Directly Observed Therapy Short course *(p. 3)* |
| **DPA** | decaprenylphosphoryl-D-arabinose *(p. 8)* |
| **EMB** | ethambutol *(pp. 3, 10)* |
| **ETH** | ethionamide *(p. 8)* |

**FT** flow through *(p. 27)*
**Fw** forward primer *(p. 22)*

**GL** glycolipids *(p. 6)*
**Glu** glucose *(p. 20)*

**HIV** human immunodeficiency virus *(p. 1)*
**HoFH** homozygous familial hypercholesterolemia *(p. 14)*

**IGRA** interferon gamma release assay *(p. 2)*
**INH** isoniazid *(p. 3)*

**Kan** kanamycin *(pp. 20, 22)*

**LAM** lipoarabinomannan *(p. 4)*
**LB** Luria-Bertani *(p. 20)*
**LDL-C** low-density lipoprotein cholesterol *(p. 14)*
**LM** lipomannan *(p. 4)*

**MA** mycolic acid *(p. 4)*
**MDR** multidrug resistant *(p. 3)*
**Mtb** *Mycobacterium tuberculosis* *(p. 4)*
**MurNAc** N-acetylmuramic acid *(p. 4)*
**MurNGlyc** N-glycolylmuramic acid *(p. 4)*

**NamH** N-acetyl muramic acid hydroxylase *(p. 6)*
**NT** non-template strand *(p. 17)*

**OD** optical density *(p. 20)*

**PAMs** proto-spacer adjacent motif *(p. 16)*
**PCR** polymerase chain reaction *(p. 22)*
**PG** peptidoglycan *(p. 4)*
**PIMs** phosphatidyl-*myo*-inositol mannosides *(pp. 4, 6)*
**pre-crRNA** primary CRISPR RNA *(p. 16)*
**pRpp** phospho-$\alpha$-D-ribosyl-1-pyrophosphate *(p. 8)*

**RF** rifampicin *(p. 3)*
**RR** rifampicin-resistant *(p. 3)*

**Rv** reverse primer *(p. 22)*

**SEM** standard error of the mean *(p. 23)*
**sgRNA** single artificial guide RNA *(p. 16)*

**T** template strand *(p. 17)*
**TC** total cholesterol *(p. 14)*
**TDM** trehalose dimycolate *(p. 6)*
**TM** transmembrane *(p. 10)*
**Tm** triple mutant *(p. 39)*
**TMM** trehalose monomycolate *(p. 6)*
**tracrRNA** trans activator crRNA *(p. 16)*
**trans-crRNA** trans-acting RNA *(p. 16)*
**TST** tuberculin skin test *(p. 2)*
**Tyl** tyloxapol *(p. 20)*

**WHO** world health organization *(p. 3)*
**WT** wild type *(p. 20)*

**XDR** extensive drug resistant *(p. 3)*

# Chemical Symbols

**$MgCl_2$** Magnesium chloride *(p. 25)*
**$MgSO_4$** Magnesium sulfate *(p. 27)*

**NaCl** Sodium chloride *(p. 27)*

# 1

# Introduction

## 1.1 Tuberculosis

Tuberculosis (TB), the second leading infectious disease, right after COVID-19, causing death worldwide, is caused by infection with the pathogen *Mycobacterium tuberculosis*. An estimated 10 million new cases of TB are diagnosed each year, of which 1.5 million were reported as fatal in 2022 [1]. The TB epidemic is highly influenced by socioeconomic conditions since its incidence is unequal throughout the world, prevailing in developing countries. TB has a higher incidence in India (26%), Indonesia (9%), China (8%), Pakistan (6%), South Africa (4%) and Nigeria (4%) [2]. Co-infection with human immunodeficiency virus (HIV) is also very common, with TB being the lead cause of death of HIV-positive patients.

TB is transmitted through the respiratory tract and most commonly affects the lungs, although it can damage any tissue. A simplified transmission cascade has been proposed (**Figure** 1.1), where an individual infected with active TB can generate aerosol particles that survive in the air and will be inhaled by a susceptible individual, that will potentially become infected and develop the disease [3]. In this case, once they contract the disease, the pathogen can spread to the lungs or other organs (pulmonary and extrapulmonary TB, respectively). The most common symptoms of the disease are coughing, fatigue, weight loss, or even more severe such as bone pain or spitting blood [4].

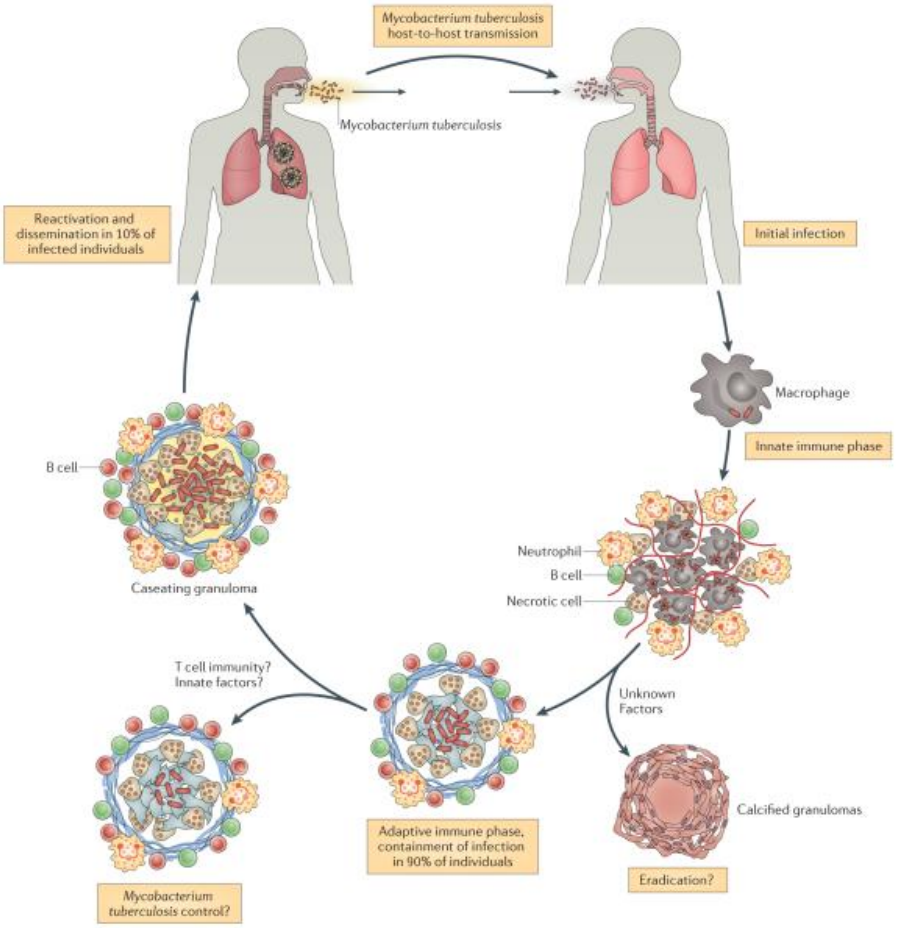


Figure 1.1: **Cascade of tuberculosis transmission** [5].

The majority of individuals who become infected develop a latent infection, which is defined by an absence of clinical signs, since the bacteria cannot replicate itself, and so the disease cannot be transmitted to other hosts [3]. About 25% of the global population harbors latent TB infection, thus being the ecological reservoir for the transmission of TB. However, just 10% of those infected progress to active disease, which appears heterogeneous, suggesting that the host role is significant in the progression and susceptibility of the disease [1].

The most prevalent tuberculosis prevention strategy is the Bacille Calmette-Guérin (BCG) vaccine, which was created almost a century ago. It offers defense against severe forms of tuberculosis, such as TB meningitis and miliary TB, in children [1]. Available diagnostic assays comprise the tuberculin skin test (TST) and the interferon gamma release assay (IGRA), however these tests are unable to differentiate between individuals whose immune system has completely eliminated the infection and those who are still infected. Nowadays, the most advanced diagnostic tests are nucleic acid amplification tests, namely the Xpert MTB/RIF assay, which can identify *Mtb*'s DNA and detect resistance to rifampicin, or imaging tests, such as magnetic resonance imaging [6].

The current treatment used to combat TB consists in a combination of several drugs

over a 6-month regimen. There are four first-line drugs, isoniazid (INH), ethambutol (EMB), rifampicin (RF), and pyrazinamide. This treatment strategy is a Directly Observed Therapy Short course (DOTS), a therapy proposed by the WHO, where direct observation of treatment is performed. The main goal of DOTS is to guarantee that the individuals with TB complete their treatment, since the 6-month regimen sometimes leads to lower adherence of patients, and thereby also helps in the prevention of drug resistance [7, 8].

Although TB is an infectious disease for which treatment exists, resistance to first- and second-line drugs is a well-known global public health problem. There are mono-resistant infections where patients develop resistance to one of the first-line drugs, multidrug resistant (MDR) infections, where patients develop resistance to at least isoniazid and rifampicin, extensive drug resistant (XDR) infections, in which the causative strains are not only multidrug resistant, but also rifampicin-resistant (RR) and have resistance to any fluoroquinolones and other second-line injectable drugs [9]. The WHO have also defined pre-XDR-TB strains which are MDR/RR and also resistant to any fluoroquinolones [1].

MDR-TB infections are nearly incurable by first-line drug treatment and the spread of MDR-TB (**Figure** 1.2) is a major problem in global TB control. In 2020, only 50% of TB cases and 32% of MDR/RR-TB cases were successfully treated [1].


Figure 1.2: **Occurrence of MDR/RR-TB in 2021, focusing on countries reporting a minimum of 1000 new incident cases.** (Reproduced from World Health Organization) [1].

The WHO "End TB strategy" has set several targets towards the elimination of the TB epidemic by 2035. However, the advent of COVID-19 has inverted the previous decline in TB cases and more recently, there has been a 50% drop in the diagnosis of TB cases in many regions across the globe, estimated to have resulted in 400 000 additional TB deaths by 2020 [2]. The decline in diagnosis coupled with the continuing high mortality rates due to MDR-TB prompt the development of new therapeutic agents, not affected by the known mechanisms of resistance, and that allow shortening the treatment time to circumvent patient compliance problems and thus decrease subpopulations that develop resistance [10].

## 1.2 Mycobacterium

The direct study of *Mycobacterium tuberculosis* (Mtb) is crucial for the understanding of its pathogenicity. However, studying it in the laboratory is very difficult and laborious because *Mtb* is a category 3 human pathogen. Therefore, this pathogen requires a biosafety level 3 laboratory, and inevitably poses a risk of accidental exposure. In addition, it has a very slow duplication time, which makes experiments very time-consuming [10].

There are three predominant species of mycobacteria that serve as models for TB studies: *Mycobacterium bovis* (BCG strain), *Mycobacterium marinum* and *Mycobacterium smegmatis*. Comparing the characteristics of all of them, *Mycobacterium smegmatis* (*M. smegmatis*), isolated in 1885, *M. smegmatis* $mc^2$ 155 strain in particular, has been used in genetic analysis, and is the most suitable for the study of *Mtb*, since in addition to being nonpathogenic, it grows rapidly and can easily be genetically manipulated [11].

A characteristic of mycobacteria, which is very similar between *M. smegmatis* and *Mtb* is that both have N-acetylmuramic acid (MurNAc) and N-glycolylmuramic acid (MurNGlyc) as peptidoglycan constituents. The simultaneous presence of these two derivatives of muramic acids, similarities in CW structure and metabolism, and PG amidation improves lysozyme resistance and influence in antibiotic resistance and in host immune recognition [12].

## 1.3 The Mycobacterial Cell Wall

The success of *Mtb* as an obligate intracellular pathogen is mostly due to the complexity of its cell wall (CW) (**Figure** 1.3). The main constituents are mycolic acid (MA), arabinogalactan (AG) and peptidoglycan (PG) mentioned as the MA-AG-PG (MAPc) complex. The PG is covalently linked to the arabinogalactan, which are coated with mycolates [13, 14].

In addition to PG, AG and MA, important lipids, namely lipomannan (LM), lipoarabinomannan (LAM) and phosphatidyl-*myo*-inositol mannosides (PIMs), are also present in the CW of mycobacteria and play an important role in the virulence and survival of *Mtb*. Also on the outer membrane are glycolipids, which contain trehalose and phthiocerol dimers, intertwined with mycolic acids and crucial for the host immune system as well as interactions host-pathogen [15].

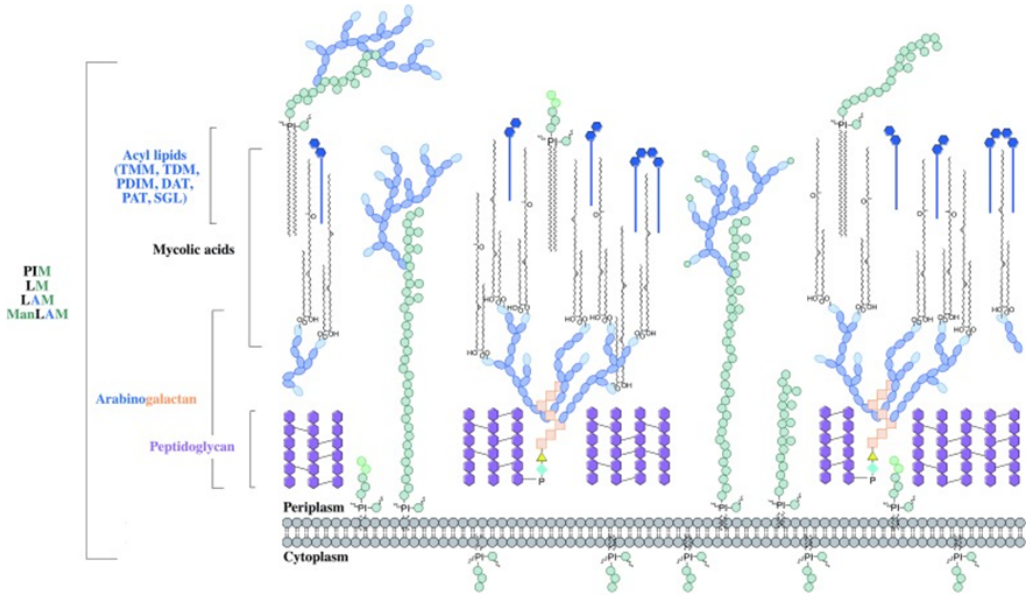


Figure 1.3: **Mycobacterial cell wall.** An illustrative representation of the mycobacterial cell wall, highlighting key components, such as glycolipids (PIMs, LM, LAM, mannosylated lipoarabinomannan), PG, AG and MA. Embedded within the mycolate layer are the acyl lipids, including trehalose monomycolate, trehalose dimycolate, diacyltrehalose (DAT), polyacyltrehalose (PAT), phthiocerol dimycocerosate (PDIM) and sulfoglycolipid (SGL). The capsular material is not depicted [13].

Their CW resembles both Gram-negative and Gram-positive bacteria, being impermeable and hydrophobic, but mycobacteria have a classification that is neither gram positive or negative. They are classified as acid fast when it comes to staining, and usually said gram-positive, in genetic terms, because it was assumed that they had no outer membrane (**Annex** Fig. I.4).

Since there are many specific components in the CW crucial for the survival of *Mtb*, its unsurprising that two of the four primary TB drugs, target the CW biosynthesis, and therefore it is the CW constituents or the proteins in its biosynthesis that have been the focus and main target in the search for new TB drugs.

### 1.3.1 Peptidoglycan

PG is present in the cell wall of both gram-positive and negative bacteria, confers strength, rigidity and shape, and it is mainly responsible for neutralizing turgescence pressure, becoming essential for cell growth and viability. In most bacteria, it consists of glycan chains with alternating residues of N-acetylglucosamine and N-acetylmuramic acid [16]. The mycobacterial PG is highly cross-linked, contributing to the increased resistance and complexity of the CW [17].

Although having a conserved structure, the mycobacterial PG undergoes quite a few modifications, namely additional N-glycolyl and N-acetyl-muramic acid residues, linked in a $\beta$-(1→4) configuration. The N-glycolylation of these residues is an evolutionary

advantage characteristic of *Mycobacterium* and closely related genera (*Tsukamurella*, *Gordonia*, *Nocardia*, *Rhodococcus* and *Micromonospora*). This modification takes place when the N-acetyl group of the muramic acid residues is transformed into the N-glycolyl group, resulting in the formation of MurNGlyc. The hydroxylase responsible for this modification, N-acetyl muramic acid hydroxylase (NamH), is encoded by the nonessential gene *namH*, a promising drug target [18]. A *namH* deletion seems to increase the vulnerability to $\beta$-lactams and lysozyme in *M. smegmatis* [14, 18].

Other unique characteristics of the mycobacterial PG are the amidation of *m*-DAP carboxylic acids, the amidation of D-*iso*-glutamate and, the presence of complex side chains through the addition of glycine or serine residues [19–21]. Distinct from *namH*, the genes responsible for catalysing these reactions, *murT*, *gatD*, and *asnB*, are essential for *M. tuberculosis* growth [22, 23]. D-*iGlu* amidation, catalyzed by the MurT/GatD amidotransferase complex, appears to influence the susceptibility to $\beta$-lactams and lysozyme and the PG cross-linking [14, 24], while *m*DAP amidation, catalyzed by the glutamine amidotransferase (AsnB), seems to be involved in PG hydrolysis modulation [14, 22].

In additional to this, there is another important difference between the PG of mycobacteria compared to other rod-shaped bacteria, such as *E. coli*. These have a protein called MreB that controls the function of PG elongation along the cell, ensuring that the glycan filaments are inserted in a circumferential manner. However, mycobacteria lack this protein, and therefore contain PG differentially at both poles of the cell, leading to asymmetric cell division which results in heterogeneous cells [14, 25, 26].

Since PG s unique to bacterial cell walls, both PG and the enzymes responsible for its synthesis have become promising targets in the development of new antibiotics [13, 16].

Disrupting peptidoglycan synthesis can result in the formation of L-shaped bacilli in mycobacteria, wich can exhibit diverse appearances and cell morphologies.

### 1.3.2 Mycolic Acids

MA represent an important class of CW lipids of mycobacteria. These long-chain $\alpha$-alkyl-$\beta$-hydroxy lipids can show considerable diversity depending on the strain and species of mycobacteria. The outer leaflet of the asymmetric lipid bilayer is composed of a mixture of extractable glycolipids (GL) and for the MAs, which are connected to the AG. These GLs include trehalose monomycolate (TMM), trehalose dimycolate (TDM) and PIMs. As a result, the asymmetric lipid bilayer is crucial for cell viability and pathogenicity as it serves as a barrier with variable permeability to both nutrients and chemicals (such as antibiotics) [13, 19].

They can be categorized into three groups based on the presence of functional groups. The $\alpha$-mycolic acids constitute the most prevalent group and contain cyclopropane rings in *cis* conformation. The other classes are called methoxy acids (methoxy group) and ketomycolic acids (ketone group), in both of which the cyclopropane rings can have *cis* or *trans* conformation [20, 21].

MAs form a hydrophobic barrier that confer protection to the mycobacteria within macrophages as well as against antibiotics, and therefore are important for the viability and virulence of *Mtb*. They can be found bound to the arabinose component of the arabinogalactan, in free form or bound to other saccharides [21, 27].

The free MAs can cause innate immune responses in hosts, including the activating of alveolar macrophages. They also serve as antigens in serological tests and are recognized by T cells due to their carbohydrate esters [21, 27].

#### 1.3.2.1 Mycolic Acid Biosynthesis

MA biosynthesis begins with two synthetases, FAS I and II respectively, and occurs in the cytoplasm (**Figure** 1.4) [28–30]. FAS I is a polypeptide responsible for the synthesis of an C 14/16 fatty acyl-CoA, which will be extended by FAS II. The FAS II via a group of enzymes (FabH, heterodimer dehydratases HadAB/BC, reductase MabA, enoyl-reductase InhA and the condensates KasA/B) carries out cycles of elongation to the acyl chain attached to AcpM, adding two carbon units per cycle. During the process, *cis/trans*-cyclopropanation modifications are introduced to the meromycolate chains, as well as addition of methoxy and keto groups [31]. Then FabD32 activates the meromycolate chain and meromycolic-AMP binds to the $\alpha$-alkyl-CoA ester to generate $\alpha$-alkyl- $\beta$-keto-mycolic acid [32]. Finally, a reduction step occurs where a mature mycolic acid is produced by the CmrA reductase [33], which will finally be transferred to the cell envelope by Mmpl3, forming TMM.

Later, the mycolyltransferase Antigen 85 complex acts to catalyze the binding of the MA to the AG, and a molecule of trehalose is released in this process [34].

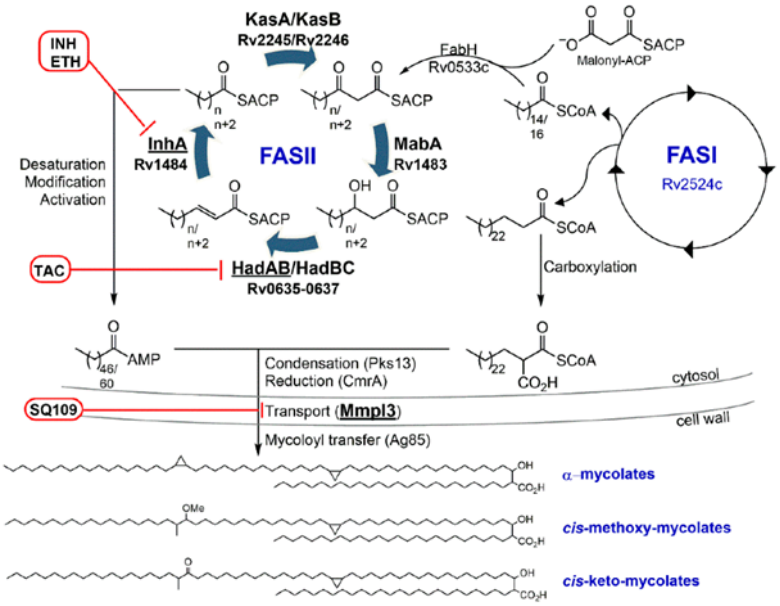


Figure 1.4: **Schematic representation of mycolic acid biosynthesis** [12].

There are several anti-TB agents that inhibit the biosynthesis of MA, namely INH and its analogue ethionamide (ETH). Isoniazid and ethionamide are first- and second-line TB drugs, respectively, that reach *Mtb* cells as a prodrug, meaning that it does not affect any eukaryotic cell, but is activated by a specific enzyme of the pathogen, leading to its death [35]. Both antibiotics are activated by KatG catalase and can bind to and inhibit InhA and enoil-ACP of the FAS II system [36–39], thus inhibiting the biosynthesis of MA and consequently leads to mycobacterial cell death [40–42].

### 1.3.3 Arabinogalactan

AG is a heteropolysaccharide that is covalently attached to the PG muramic acid in the cell wall, creating a macropolymer bound by a single bond. The AG consists mostly of arabinose and galactose, both in the furanose ring form, which are uncommon in nature. The galactose part is a linear chain of approximately 30 residues, while the arabinose component has 3 highly branched arabinan chains [32].

#### 1.3.3.1 Arabinogalactan Biosynthesis

The AG biosynthesis starts with the creation of the bond between PG and AG, initiated by the enzyme WecA (**Figure** 1.5). The galactofuranosyl-transferases GlfT1 and GlfT2 carry out the synthesis of the linear galactan chain [32, 34].

The remaining steps occur outside the cell. The Araf residues are transferred into the chain via decaprenylphosphoryl-D-arabinose (DPA). DPA is synthesized by several steps, beginning with the addition of a decaprenyl group to phospho-$\alpha$-D-ribosyl-1-pyrophosphate (pRpp) by UbiA, forming decaprenol-1-monophosphate 5-phosphorribose. Subsequently, this enzyme is subjected to epimerization and dephosphorylation by DprE1 and DprE2, finally forming DPA.

The transfer of the inicial arabinose unit from DPA to the galactan chain is catalyzed by the first arabinofuranosyltransferase (AraT), AftA. The AraTs, EmbA and EmbB, catalyze the addition of $\alpha$ $(1 \rightarrow 5)$ of Araf so that arabinan is formed. AftC and AftD introduce $\alpha$ $(1 \rightarrow 3)$ branching. The structure ends with a hexa-arabinofuranosyl (Ara f6) created by EmbA, EmbB, AftC, AftD and AftB [13, 28]. Despite these advances, the precise structural details with respect to the number and further conjugation remain unclear.

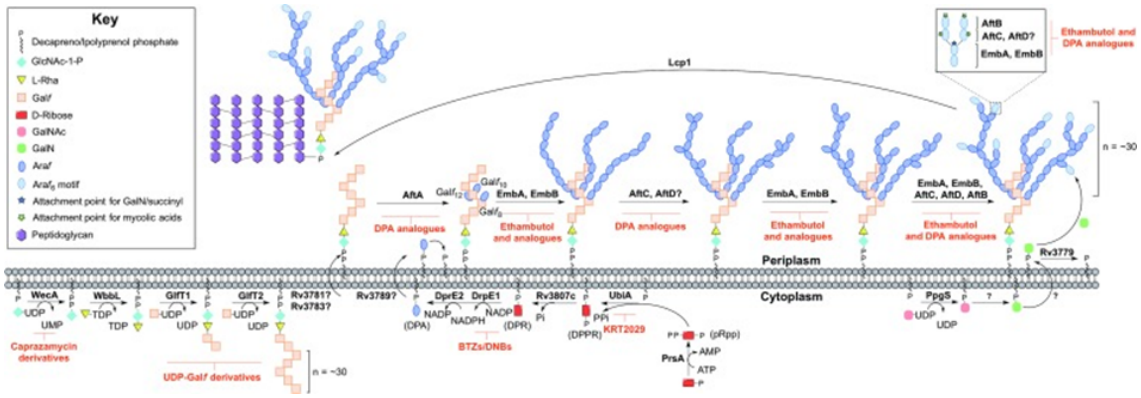


Figure 1.5: **Enzymes involved in the biosynthesis of arabinogalactan, with inhibitors highlighted in red** [13].

AG biosynthesis presents some promising targets for anti-TB drugs. The arabinosyltransferases Emb are inhibited by EMB, a first-line antibiotic used to treat TB, therefore compromising CW integrity [15]. Resistance to this antibiotic is related to mutations in a group of genes (*embCAB*, Rv3793-3795) responsible for encoding arabinosyltransferases, which increase the transcription levels of *embA* and *embB* [43]. In addition, this antibiotic can inhibit the synthesis of arabinan to arabinogalactan, since mutations have also been found in the enzyme UbiA, which participates in DPA synthesis, and mutated leads to an increase in intracellular DPA, which can thus bind to Emb and contribute to ethambutol resistance [28].

## 1.4 Structures of the arabinofuranosyltransferases AftD and EmbB

As previously mentioned, the CW of mycobacteria is fundamental to their growth and virulence, and is also a major contributor to their ability to resist against antibiotics. AraTs enzymes, as Afts and Embs, standing out in the biosynthesis of the mycobacterium CW, using DPA as substrate [44]. Therefore, AraTs become promising targets for new therapeutic targets against tuberculosis.

In the present study, the enzymes AftD, AftA and EmbB will be the main focus. The most widely recognized AraT are Embs, and they receive their name because are reported targets of EMB [45, 46], a first-line antibiotic against TB. Among the AraTs family, EmbB stands out as one of the most extensively investigated members, whith a molecular mass of 117 kDa, while AftD with a mass of 150 kDa, stands as the largest protein predicted within this family [47]. In order to have a deeper perception into the drug resistance and catalytic mechanisms of these proteins, it is important to know their 3D structure.

The *embA*, *embB* and *embC* genes are co-located in the *Mtb* genome and is suggested that the EmbB-EmbA proteins form a heterodimer within cells, functioning in a coordinated manner. Lu Zhang *et al.* reported the structure of *M. tuberculosis* and *M. smegmatis* EmbA-EmbB complexes expressed in *M. smegmatis* cells [48]. They were able to elucidate its structure at 2.81 to 3.10 Å resolution, using single-particle cryogenic electron microscopy (cryo-EM). The structures were obtained in complex with EMB and also with the acceptor and donor substrates, diarabinose and DPA, respectively. Interestingly, the structures revealed the presence of an acyl carrier protein (Acp) attached to the cytoplasmic surface of each EmB protomer. [48].

The structure (**Figure** 1.6) has a transmembrane (TM) domain with 15 helices, periplasmic and cytoplasmic loops that allow connections between the helices, and also periplasmic domains at the N- and C-terminals, PN and PC respectively. The periplasmic loop 5 (PL5) forms two helices that contribute to the dimerization of the complex and blockage of the DPA; PL2 also forms two cross helices that shelter the highly conserved DDX motif [48].

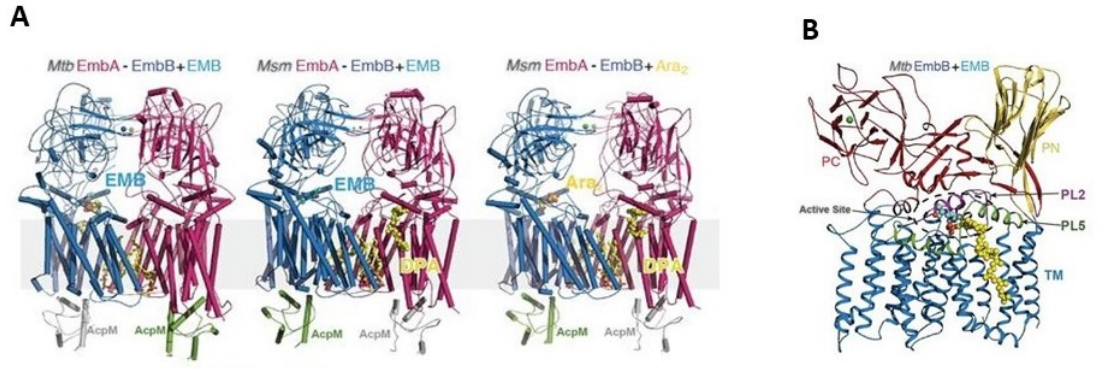


Figure 1.6: **Structure of EmbA-EmbB complexes from *Mtb* and *M. smegmatis* with ethambutol (EMB), donor (DPA) and acceptor (diarabinose) substrates and bound to AcpM.** (**A**) Overview of the cryo-EM structures of *Mtb* and *M. smegmatis* complexes EmbA-EmbB-AcpM2 in complex with diarabinose or ethambutol. The unmodeled AcpM protomer bound to Mtb EmbB or *M. smegmatis* EmbA is shaded in gray. The EMB drug, DPA and diarabinose substrates and the lipids are presented as spheres. In (**B**) the overall folding of the Emb proteins is shown, with *Mtb* EmbB represented. The PC, PN and TM domains have different colors. Functionally crucial PL2 and PL5 regions are also emphasized [48].

The periplasmic domain of the C-terminus can be divided into two subdomains, the first exhibiting a mixed $\alpha$ and $\beta$ structure, and the second exhibiting a gelatinous fold that is stabilized by a $Ca^{2+}$ ion that is bound to it. At the intersection of the TM and periplasmic domains, it is a space where the active site is found (**Figure** 1.6B), which is where EMB and substrates are observed, allowing clear conclusions about drug inhibition, resistance and catalysis [48].

To also better understand their function and activity, Yong Zi Tan *et al.* using as well cryo-EM, were able to determined the 2.9 Å structure of the AftD protein from *M. abscessus*, which was produced in *E. coli* through recombinant techniques [49].

As shown in **Figure** 1.7, both terminals of AftD are in the cytoplasm, and at the N-terminus are 11 TM helices, which have a common fold of the GT-C glycosyltransferases. In the periplasm is the polypeptide chain. AftD is composed of a soluble portion that lies on the periplasmic side of the inner membrane and a membrane-embedded area with 16 TM helices. AftD's extensive periplasmic segment, which accounts for 60% of the protein's mass, compromise five distinct domains. Following these domains, the polypeptide chain reverts to the membrane through a helix, extending up to the C-terminus [49].

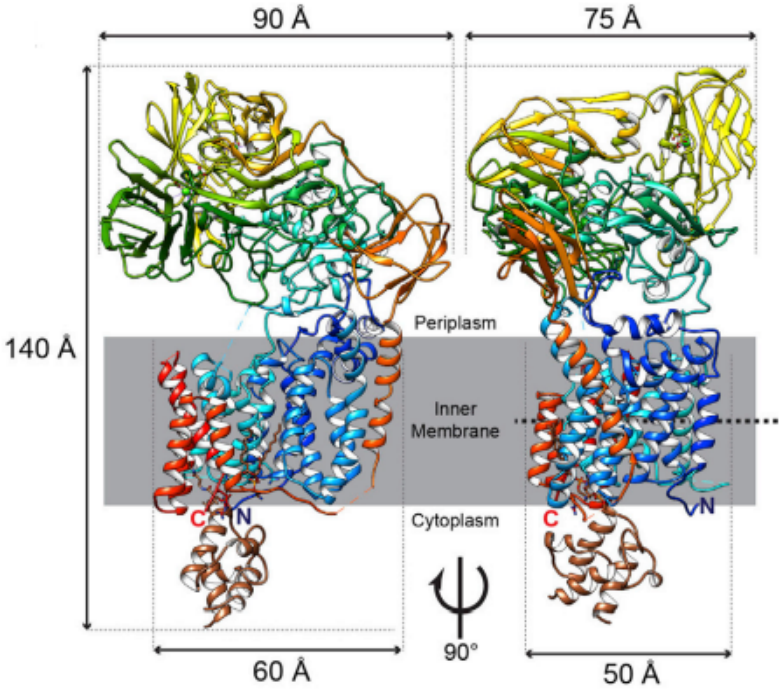


Figure 1.7: **Cryo-EM structure of AftD.** The single-particle cryo-EM structure of the AftD complex is represented in a rainbow-colored cartoon format, from the N- (blue) to the C- (red) terminus. The complexed includes *E. coli* ACP and ligands, depicted in brown, with both proteins shown as cartoons. Green spheres represent $Ca^{2+}$ ions. Two distinct perspectives, each perpendicular to the membrane plane are presented. The estimated monomer dimensions measure approximately 85 Å in width, 145 Å in height, and 75 Å in depth. The membrane boundaries were defined by identifying the interface between the nanodisc lipids and the solvent [49].

AftD has DPA as donor and a putative substrate polyarabinose, that is part of the growth of AG or LAM, as acceptor [50, 51]. Using a server named Dali [52], a search for atomic models resembling AftD, found matches with all GT-C structures available in PDB. The first 600 residues of AftD constitute the GT-C superdomain, where a large cavity is present between the ends facing the periplasm and the periplasmic domain. Various sequence alignment of AftD among members of the *Actinobacteria* class, showed a large number of highly conserved residues in this cavity, making them strong candidates for catalytic residues [49].

Structural superposition of AftD with other GT-C glycosyltransferases, like PgIB, was performed to compare their active sites. PgIB requires the presence of bivalent metal cations for proper functioning [53] and AftD also possesses a negatively charged pocket, which suggests its potencial for this purpose (**Figure** 1.8; red circles). Although, curiously, around the presumed AftD active site there was also observed some positive charge.

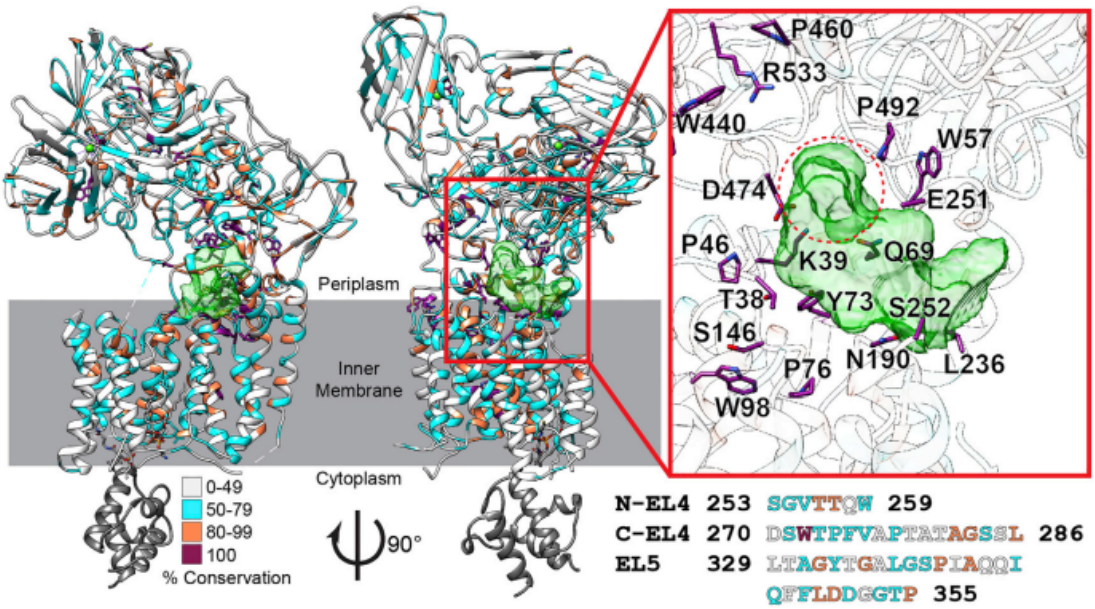


Figure 1.8: **Sequence Conservation and the Presumed Active Site of AftD.** The structure of AftD is despicted in a cartoon representation and colored according to sequence conservation. Distinct sequences obtained through a PSI-BLAST search were initially filtered to retain those that share 95% of the length of *M. abscessus* AftD [49].

Both structures, EmbB and AftD, were determined by cryo-EM. The structure of AftD was the first complete mycobacterial GT-C structure to be successfully acquired. 3D structures of EmbB were characterized from *M. smegmatis* and *Mtb* sources. On the other hand, in AftD a potential active site was identified and its structures has not yet been solved in *Mtb*, which may mean that some observations are not transferable to pathogenic bacteria, but are nevertheless an excellent starting point for further studies.

## 1.5 Drug Repurposing for AftA

A prediction of the structure of AftA in *Mtb* could be obtained using the artificial intelligence program AlphaFold. This program, based on the sequence of AftA and comparing with other Afts, is able to predict its structure and active center. Once this structure was predicted, docking of the protein and a virtual screen was performed. Molecular docking uses protein structure to predict the complementarity of the binding site between the ligand (a drug, for example) and the target (AftA, in this case) [54].

The list of compounds used (RepoDB library) was based on a drug repurposing strategy, and consists of drugs that are already commercially available for the treatment of other diseases, being reused for a different disease, as in this case for TB. Using this approach, instead of developing an entirely new medication for a specific application, offers several advantages, such as reduced drug development time, lower risk of failure and low cost.

In this study, after *in silico* calculations, 4 drugs were chosen from this library to be tested using the MIC technique, which determines the lowest drug concentration capable of inhibiting visible bacterial growth. The selected drugs, based on molecular docking of AftA structure, were the coumpounds A, B, C and D.

Compound A is employed as an inhibitor of microsomal triglyceride transfer protein in homozygous familial hypercholesterolemia (HoFH) patients to lower levels of low-density lipoprotein cholesterol (LDL-C), total cholesterol (TC) and apolipoprotein B (apoB). This drug is contra-indicated in pregnancy and severe gastrointestinal adverse reactions may occur.

Compound B function as a leukotriene receptor antagonist to alleviate symptoms of allergic rhinitis and asthma. It is a drug with some side effects including abdominal or stomach pain and others. Compound C, as compound B, is also used for the treatment of asthma and may cause some people to be agitated, irritable, or display other abnormal behaviors.

Compound D is an antiemetic agent used for the prevention and symptomatic treatment of nausea and vomiting. This medicine is contraindicated during lactation and for use by children up to 2 years of age.

For reasons of confidentiality regarding the use of these compounds against tuberculosis, their names are not referenced.

Table 1.1: **Compounds chosen based on drug repurposing strategy for testing in AftA**

| Compounds | Associated Conditions | Date of approval |
|---|---|---|
| **A** | Homozygous Familial Hypercholesterolaemia (HoFH) | 2012 |
| **B** | Allergic Rhinitis (AR) Asthma | 1998 |
| **C** | Asthma Chronic Asthma | 2007 |
| **D** | Dizziness Menière's Disease Vomiting | 1967 |

Recently, it was possible to determine the structure of AftA in *M. smegmatis* by cryo-EM at 3.1 Å resolution [55]. AftA also belongs to the GT-C superfamily but seems to present a dimer never seen before, a more open dimer compared to those of more compact Embs, with 13 TM helices and a soluble C-terminal domain in the periplasm. **Figure** 1.9 shows its structure.

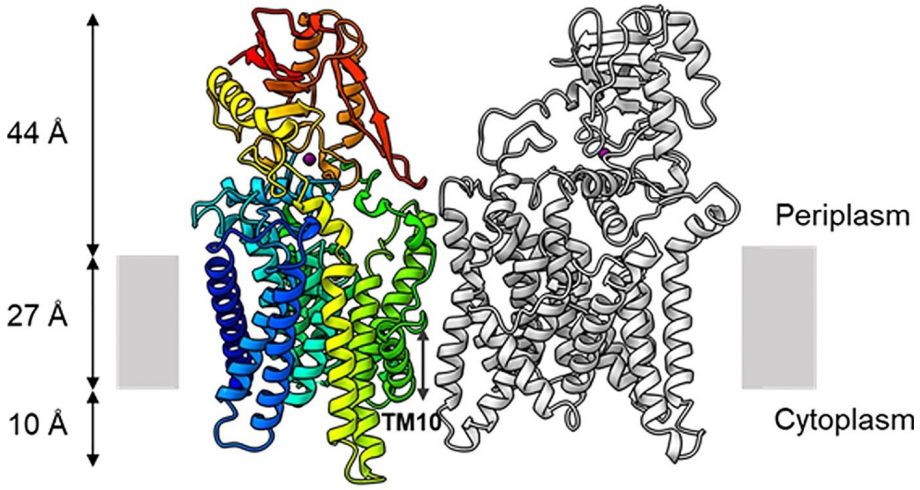


Figure 1.9: **Cryo-EM structure of AftA.** Represention of the dimeric AftA, the protein appears to have TM helices of different sizes, with TM10 highlighted as the shortest helix crossing only half of the membrane [55].

Gong *et al.* also used the Dali server to search models with similarity to AftA, and the top structural homologue results were members of the GT-C superfamily, confirming that the TM domain of AftA exhibits a conserved GT-C fold. In addition, they also used unique AftA sequences from *Corynebacterium* and *Mycobacterium* genera to identify a highly conserved region with Asp128 above the membrane. To deeper understanding the importance of this amino acid in the structure, they mutated it and realized it eliminated AftA activity, suggesting this residue is relevant for catalysis [55].

Interestingly, in this study, Acp was not observed attached to the cytoplasmic side in the AftA structure, as in the previously mentioned structures of Afts. When purifying AftA without Acp, the protein was active, suggesting that Acp may not be required for all arabinosyltransferases [55].

Since this article was recently published, the selection of compounds were based on computational studies and dockings that were already done using the AftA protein structure predicted by Alphafold.

## 1.6 CRISPR

There is a growing need for new and more effective genetic methods for drug-target characterization. For TB, with drug-resistance cases, the required development of new therapeutics is crucial. Recently, the use of the clustered regularly interspaced short palindromic repeat (CRISPR) system together with its associated proteins (CRISPR-Cas) appears as a very useful and alternative genetic tool for targeted gene regulation.

Approximately 40 % of bacteria own a CRISPR system for defense against foreign DNA elements [56]. The Cas heterogeneous family of proteins has typical domains of nucleases, polymerases, helicases and polynucleotide-binding proteins [57, 58]. Usually, different organisms can have distinct types of the CRISPR-Cas system, however all consist in 3 separated steps: adaptation, expression, and interference [57]. The most commonly employed CRISPR system is *Streptococcus pyogenes* (*S. pyogenes*) type II which is essentially composed of two components: Cas9 nuclease and a single artificial guide RNA (sgRNA), which is a fusion of a CRISPR RNA (crRNA) and a trans activator crRNA (tracrRNA) [59]. The first step is adaptation, characterized by Cas-mediated cleavage and insertion of the invading protospacer sequence between two CRISPR repeats. The recognition of proto-spacers from the invading nucleic acids is achieved by a prior identification of unique proto-spacer adjacent motif (PAMs) [60, 61]. The second step is expression and processing, in which occurs the transcription and processing of a long primary CRISPR RNA (pre-crRNA) into a mature crRNA by a trans-acting RNA (trans-crRNA), who has complementarity with precrRNA repeat fragments and guides the RNase III and the endonuclease Cas9 protein to catalyse precrRNA cleavage in a repeat or spacer fragment, respectively. Finally, the last step is interference, characterized by crRNA-guided Cas-mediated cleavage of the complementary target sequence [62–64].

### 1.6.1 CRISPRi

More recently, a new and altered CRISPR system, the CRISPR interference system (CRISPRi), has been developed (**Figure** 1.10). This new system features two changes from the original CRISPR-Cas9 [65, 66]:

1. the sgRNA comprises three regions: a 20 bp base-pairing sequence, a 42 bp dCas9-binding hairpin and a 40 bp species-specific transcriptional terminator;

2. the Cas9 protein contains 2 point mutations (D10A and H840A) in the nuclease domains, creating an endonuclease-deficient Cas9 (dCas9) with retained DNA-binding activity.

The nuclease-dead *S. pyogenes* Cas9 protein (dCas9$_{Spy}$) when co-expressed with a designed sgRNA, complementary to a target sequence located next to a PAM sequence, causes a specific gene silencing [65, 67].

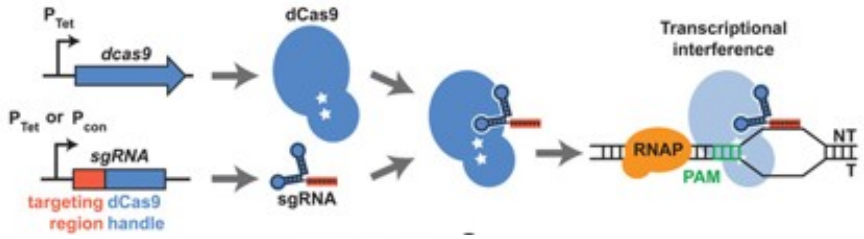


Figure 1.10: **Schematic representation of the mechanism of action of CRISPRi system to achieve transcriptional repression** [64].

The CRISPRi system has been thoroughly studied in the bacterial species *E. coli* and *Bacillus subtilis* (*B. subtilis*) [66, 68], in both, the dCas9$_{Spy}$-sgRNA complex causes a steric block to transcription elongation or represses transcription of the target gene by hiding the RNA polymerase from the target promoter [64]. Targeting the non-template strand (NT), strand of the target gene increases silencing efficiency (up to 300-fold repression) when compare to targeting the template strand (T) strand. This could indicate that the RNA polymerase is able to overcome the sgRNA-dCas9 obstruction and continue transcription in the T strand orientation [66].

Interestingly, initial endeavors to implemented the dCas9$_{Spy}$ from CRISPRi system in micobacteria were not very successful [68]. Nevertheless, these limitations seemed to be overcome by the use of a simpler and more efficient CRISPRi system, based on the employment of dCas9 from *S. thermophilus* (dCas9$_{th1}$). Rock *et al.* have found this system which was developed by constructing a single plasmid system that contains [64]:

1. a codon-optimized dCas9$_{th1}$ under the control of an anhydrotetracycline (ATc)-inducible promoter;
2. a sgRNA under the control of a constitutive or ATc-inducible promoter;
3. a Tet repressor;
4. a single-copy L5-integrative backbone that contains the integrase gene and an *attP* site;
5. an *E. coli* derived pBR3222 origin of replication;
6. a kanamycin resistance cassette .

Considerably, in order to increase the number of possible target sites and allowing different efficiencies of gene silencing, 24 permissive PAM position variants for dCas9$_{th1}$ were identified (**Annex** Fig. I.2), each one showed a different fold-repression ("PAM strength") [64, 67]. In line with previous data [69], no relation was found between fold-repression and the distance of the designed sgRNA from the transcription start site (TSS), the targeting of essential genes caused an inhibition of growth and a robust gene knockdown for all genes and PAM variants tested [64].

This CRISPRi system was found to exhibit exceptional specific and very robust in both *Mtb* and *M. smegmatis* [64].

## 1.7 Objectives

There is an urgent need of research for more effective anti-TB drugs and strategies, in which the complex mycobacterial CW comprises a promising target. The AG biosynthesis process includes multiple key enzymes for bacterial survival and for the success in host immune evasion.

This project aims to functionally characterize Afts, identify of promising candidates with high specificity and affinity to inhibit AftA and evaluate their potency using *in vitro* and *in vivo* assays. The work will be developed at the Membrane Protein Crystallography Lab, supervised by Margarida Archer, ITQB NOVA, in close collaboration with Maria João Catalão, Host-Pathogen Interactions Lab, at iMed-ULisboa.

The specific aims of this thesis are to:

1. Construct of *aftA*, *aftD* and *embB* knockdown mutants in *M. smegmatis* $mc^2$ 155 using CRISPRi. The knockdown mutants will then be functionally characterized, through growth curves and spotting dilutions assays, and confirmation of gene repression by qRT-PCR;

2. Evaluate the inhibition power of promising compounds, selected based on a drug repurposing strategy, in opposition to a predicted structure of AftA in *Mtb*, using the artificial intelligence program AlphaFold. The selected compounds will be characterized by determination of the minimum inhibitory and minimum bactericidal concentrations (MICs and MBCs, respectively);

3. Produce AftA wild type (WT) in *E. coli* and *M. smegmatis*;

4. Produce AftD (WT and mutants) in *M. smegmatis*, followed by protein purification using affinity-chromatography;

5. Performe a functional assay with AftD using NMR spectroscopy to assess the enzymatic activity of the WT and mutants forms of the protein of interests.

Tasks 1-2 will be performed at iMed-ULisboa, whereas tasks 3-5 will be completed at ITQB.

# 2

# Materials and Methods

## 2.1 Bacterial strains and grown conditions

The bacterial strains used in this study were *E. coli* JM109, *E.coli* XL-10 and *M. smegmatis* $mc^2$ 155.

*E. coli* was cultured in Luria-Bertani (LB) medium (Merck) supplemented with 25 µg/mL kanamycin (Kan) (Sigma-Aldrich) when necessary.

To amplify the CRISPRi backbones, *M. smegmatis* $mc^2$ 155 was grown in liquid DifcoTM Middlebrook 7H9 (BD) supplemented with 0.2% (v/v) glycerol (Thermo Fischer Scientific), 0.05% (v/v) tyloxapol (Tyl) (Sigma-Aldrich) and 0.5% (v/v) glucose (Glu) (Sigma-Aldrich), at 37 °C with agitation at 180 rpm. *M. smegmatis* culture plating occurred in 7H10 agar medium supplemented with 0.5 % Glycerol. For mycobacterial gene silencing, in *M. smegmatis*, the CRISPRi backbone plasmid used was PLJR962 (Addgene ID:115162). The media were supplemented with 100 ng/mL of ATc (Sigma-Aldrich) to induce CRISPRi.

To grow *M.smegmatis* wild type (WT) and overexpress a protein of interest, cells were first cultured on a pre-inoculum with 50 mL of 7H9 medium supplemented with 0.5% Glucose, 0.05% Tween 80 and 15 µg/mL Kan. The starter culture was grown overnight at 37 °C in an incubator shaking at 200 rpm. The next day, the inoculum was prepared with 900 mL 7H9 medium, 9 mL of 20% succinate, 10% Tween 80 and 50 µL/mL Kan, and left grown at 37 °C shaking at 240 rpm until cells reached an optical density $(OD)_{600}$ at 0.8. Once this OD was reached, the cells were induced by addition of 0.2% acetamide. *M. smegmatis* cultures plating occurred in 7H10 agar medium supplemented with 0.5 % Glu and 15 µg/mL Kan.

## 2.2 Construction of *aftA*, *aftD* and *embB* knockdown mutants in mycobacteria

To silence the expression of *aftA*, *aftD* and *embB* in *M. smegmatis* using CRISPRi, various sgRNAs were designed, according to the following strategy.

### 2.2.1 Design of the sgRNA

1. Finding a PAM sequence on the antisense strand of the gene of interest;
2. Select the sgRNA as the first ~20-25 nucleotides upstream the PAM sequence;
3. Verify that the sgRNA sequence start with an "A" or "G";
4. Perform a BLAST search to assert that the base-pairing region is unique in the genome so as to circumvent off-target inhibition.

In this work, the *in silico* design of the sgRNAs was performed as optimized by Rock *et al.* [64, 70].

Table 2.1: **sgRNAs used to target *aftA*, *aftD* and *embB* genes in *M. smegmatis***

| ***M. smegmatis mc²* 155** | | | | | |
|---|---|---|---|---|---|
| **Searched PAM** | **Actual PAM** | **Base-pairing region of the sgRNA** | **Size (bps)** | **BLAST (seed region + PAM)** | **Unique in Genome** |
| ***aftA (MSMEG_6386)*** | | | | | |
| 5′-NNGGAAC-3′ | 5′ – GCGGAAC – 3′ | 5′ – AGCCGGGTGAGGTACTCGGT – 3′ | 20 | GAGGTACTCGGTGCGGAAC | Yes |
| 5′-NNAGGAA-3′ | 5′ – CCAGGAA – 3′ | 5′ – GTAGAACAGCGCAGCTACAC – 3′ | 20 | GCGCAGCTACACCCAGGAA | Yes |
| ***embB (MSMEG_6389)*** | | | | | |
| 5′-NNAGCAG-3′ | 5′ – CCAGCAG – 3′ | 5′ – ACCTCACGGGACAGCAGCAG – 3′ | 20 | GGACAGCAGCAGCCAGCAG | Yes |
| 5′-NNAGCAG-3′ | 5′ – CCAGCAG – 3′ | 5′ – GATGACGGTCGACACGATCG – 3′ | 20 | TCGACACGATCGCCAGCAG | Yes |
| ***aftD (MSMEG_0359)*** | | | | | |
| 5′-NNAGCAG-3′ | 5′ – TCAGCAG – 3′ | 5′ – GCACACCCCAGAACCCGACCG – 3′ | 20 | AGAACCCGACCGTCAGCAG | Yes |
| 5′-NNAGGAA-3′ | 5′ – CGAGGAA – 3′ | 5′ – GCCCGACGACTCGATGAAGT – 3′ | 20 | ACTCGATGAAGTCGAGGAA | Yes |

### 2.2.2 Amplification of the CRISPRi backbones

To amplify the CRISPRi backbones in *E. coli* JM109, chemically competent cells (CCC) were transformed with PLJR962 plasmid (**Annex** Fig. I.1) using a double heat-shock approach. Shortly, the suspension mix was put in contact with ice for 30 min, carefully mixed and, heated at 42°C for 45 seconds. After, the cells were iced for 10 min and 950 µL of warm LB media were added. The cells were then incubated at 37°C for 1 h, with agitation. Finally, the cells were pelleted at 3000 xg for 5 min, the supernatants were discarded and, the pellets were resuspended in 100 µL of LB. Pre-heated LB supplemented with 25 µg/mL Kan plates were used to selected for the desired transformants. The recombinant DNA was extracted using a NZYMiniprep kit (NZYTech) and then, quantified using Nanodrop (Thermo Fischer Scientific).

Afterwards, the plasmid was digested at the *BsmBI* restriction site (40 U of restriction enzyme were used to digest 1 µg of plasmid DNA), the reaction occurred at 37°C for 4 h in a thermocycler block, in the presence of DTT and, purified using the QIAquick polymerase chain reaction (PCR) Purification Kit (Qiagen) [64]. The obtained DNA was afterwards quantified using Nanodrop and stored at - 20°C.

### 2.2.3 Cloning the sgRNA into the backbone vector

To clone the *aftA/aftD/embB* sgRNA targeting sequence into the sgRNA scaffold, two complementary primers were designed, as previously described. The sequence in the template strand (forward primer (Fw)) corresponds to the transcribed sequence of the sgRNA and the sequence in the non-template strand (reverse primer (Rv)) is the reverse complement. All pairs of sgRNAs oligos were synthesized (Eurofins Genomics), annealed, and then ligated into the digested CRISPRi backbone. For the annealing reaction, 4 µL of each oligo (100 µM) was added to 42 µL of annealing buffer (50 mM Tris, pH 7.5; 50 mM NaCl; 1 mM EDTA) and, the program (95°C for 2 min, 0.1°/sec to 25°C, END) was run on a thermal cycler. Afterwards, the annealed oligos were ligated into digested PLJR962 using T4 DNA ligase (400 U/µL; NEB), at 22°C for 30 min. The ligated plasmid was then transformed into *E. coli* CCC by double heat-shock transformation and the plasmid DNA from the recombinant clones was extracted using the Miniprep kit (NZYTech) and quantified with NanoDrop. The sequence of the recombinant plasmids was confirmed by Sanger sequencing (Eurofins Genomics).

Finally, *M. smegmatis* was transformed with 300 ng of the obtained clones by electroporation [71] in the presence of 10% (v/v) glycerol. The cells were recovered in supplemented 7H9 (Glu + Tyl) for 2 hours and, later plated in 7H10 supplemented with 25 µg/mL Kan for selection of Kan-resistant transformants. The mycobacteria harboring the CRISPRi systems were then grown in 7H9+Glu+Tyl at 37°C with shaking at 180 rpm.

## 2.3 Characterization of the mutants

The mutants were initially characterized by performing spotting dilutions and growth curves.

First, the mycobacterial cultures were grown to reach the log phase in 5 mL of supplemented 7H9 medium (Glu+ Tyl). The cultures then were centrifuged at 3000 x g for 5min. Afterwards, the pellets were resuspended in the initial volume of growth media, subjected to an ultrasound bath (Bardelin sonorex RK-52) for 5 minutes as a way to remove bacterial clumps and to separate the remaining cell aggregates and, again, centrifuged at 500 xg for 1 min to obtain a single-cell supernatant. The $OD_{600}$ of each suspension was then measured by a spectrophotometer (Biophotometer-Eppendorf), and the cultures were normalized to a $OD_{600}$ of 0.001.

To perform growth curves [68], mycobacterial cultures were grown, washed, and normalized as described before, in a final volume of 20 mL. Each culture was incubated with or without ATc and was then kept at 37 °C with shaking at 180 rpm. The culture growth evaluation was performed for 45 h and the $OD_{600}$ of each culture were measured at 18, 21, 24, 42 and 45 h respectively. The amount of ATc in culture was replenished every 24 h. This experiment was performed three times for each culture and the average and standard error of the mean (SEM) were calculated.

To perform spotting dilutions [64], the cultures of *M. smegmatis* were previously grown, washed, and normalized as described before. Each bacterial suspension was added to a well of a 96-well plate (Sarstedt) at a volume of 200 µL, serial two-fold dilutions were prepared, and 5 µL of each dilution were then plated into 7H10 or 7H10+ ATc plates. The plates then were incubated at 37 °C for 48 h. Three independent biological replicates were performed.

## 2.4 Evaluation of mRNA expression levels by qRT-PCR

Before extracting RNA, the cultures were grown, washed, and normalized to an $OD_{600}$ of 0.001. The cultures were then incubated at 37 °C with shaking at 180 rpm until an $OD_{600}$ of 0.05-0.1 was reached. After 18 h of incubation (t=0 h), 100 ng/mL of ATc were added to induce CRISPRi-mediated gene silencing and, the cultures were again incubated at 37 °C for 6 h. The cells were harvest by centrifugation at 3000 xg for 5min, 500 µL of RNA protect were added to resuspend the pellets and these were stored at -20 °C.

To perform RNA extraction a specific kit was used (Nzytech Total RNA Isolation Kit), according to the manufacturer's instructions. The cells were first thawed, centrifuged at 5000 xg for 5 min, and then the supernatant was completely removed. Mycobacterial cells were lysed by resuspension in 350 µL of buffer NR (Kit buffer) and 3.5 µL $\beta$-mercaptoethanol, after which the bacterial suspensions were transferred to a 2 mL screw-cap tubes and lysed using a bead-beater (BeadBug 6; Benchmark). Afterwards, the beads were spinned-down, and the lysate was transferred into an NZYSpin Homogenization column. To

avoided genomic DNA contamination, the extracted RNA was treated with TurboDNase (Invitrogen).

The purity and concentration of RNA was quantified using Nanodrop. The absence of DNA was confirmed by performing PCR, using the NZYTaq II 2x Green Master Mix kit (NZYTech), followed by a 1.5% agarose gel electrophoresis analysis.

To synthesize cDNA, the NZY First-Strand cDNA Synthesis kit (NZYTech) was used. Further, the cDNA was quantified by a qPCR using the NZY qPCR Green Master Mix (2x), ROX plus (NZYTech) kit, a QuantStudio™ 7 Flex Real-Time PCR System (Applied Biosystems) and, different sets of primers (the primers were verified to have a 90-110% amplification efficiency). The 40 cycles PCR reaction included a polymerase activation step (95°C for 10 min), a desnaturation step (95°C for 15 s) and an anneling/extension step (60°C for 30 s). The data were quantified using the ΔΔCT method and normalized against the housekeeping gene *sigA*. Two biological replicates, each comprising two technical replicates, were used for each culture of interest. The average and SEM were then calculated.

## 2.5 Construction of AftA expression vector

The construction of the AftA expression vector in *M. smegmatis* was carried out using the Gibson reaction. The previous performed PCR amplifications of the DNA from the PVVAP vector (*M. smegmatis* expression vector) and the DNA encoding for the AftA protein were used for the reaction, together with the TEV cut linker, the FLAG tag and the 10-histidine tag (amplified from an *E. coli* expression vector previously used in the laboratory to express AftA). The primers used (**Annex** Table I.1) for the amplifications were designed so that the final product of each amplification had complementary ends, according to the desired order of binding. After assembling everything in a single reaction with the Gibson cloning Master Mix (which consists of 3 different enzymes: T5 Exonuclease; Phusion DNA Polymerase and Taq DNA ligase) [72] the mix was incubated at 50 °C for 1 h. After that, *E. coli* XL10 competent cells were transformed with the expression vector by the heat shock method. For outgrowth, after the heat shock, the cells were incubated in 500 µL of LB at 37 °C for 1 h. Thereafter, the cells were centrifuged at 5000 rpm for 2 minutes, and then plated in LB medium containing Kan as selective antibiotic.

To screen for positive clones, two methods were used. Cell material from randomly chosen colonies were either submitted to the Colony PCR technique or transferred to falcons and grown overnight at 37 °C with LB supplemented with 15 µg/mL Kan.

Two different strategies were tested for Colony PCR. In both employed strategies, colonies were picked from the plate using a pipette tip. In the first strategy, when the colony was picked, it was rubbed against the walls of the eppendorf, followed by the addition of the PCR components. The used pipette tip was further discarded into a falcon used for the pre-inoculum. In the second strategy, the colony was picked and then diluted in an eppendorf with 20 µL of MilliQ water and, after this dilution, half of the volume

would be for the Colony PCR reaction and the other half for the respective pre-inoculum. The colony PCR results were analyzed through eletrophoresis using an 0.7% agarose gel.

The pre- inoculum tubes that were left with colonies growing overnight were centrifuged at 7000 rpm for 10 minutes and the resulting pellets were used for plasmid DNA purification with the NZYMiniprep kit(NZYTech). The DNA obtained from each clone was then analyzed using a UV-Vis spectrophotometer (NanoDrop), which enabled the measurement of plasmid DNA concentration and its degree of purity.

The resulting plasmid DNA samples had suitable concentration and purity parameters (with an absorbance ratio 260/280 between 1.8 and 2). These samples were then submitted to enzymatic digestion using Tango buffer and the Xba*I* enzyme. The digestion reaction occurred overnight at 37 °C, followed by 20 minutes at 65 °C. The digestion reaction results were analyzed after eletrophoresis using an 0.7% agarose gel.

DNA from allegedly positive clones were sent for Sanger sequencing (Eurofins Genomics) and those showing a correct alignment with the AftA protein sequence in the PVVAP plasmid, were selected to proceed. At this stage, competent cells of *M. smegmatis* WT were transformed with the positive clone by electroporation. In this transformation method, the cells are placed in cuvettes in an electroporator and subjected to high-voltage electrical pulses that increase the membranes transport potential and allow for the exchange of DNA. Afterwards, the cells were inoculated in LB medium, Tween 0.05% and glucose 0.5%, and incubated at 37 °C for 2 h. After this incubation step, the cells were centrifuged at 15000 rpm for 10 minutes at 4 °C, and then plated on solid 7H10 medium containing Kan as selective antibiotic and 0.5 % glucose. The plates containing the transformants were incubated at 37 °C for 48 h.

The *M. smegmatis* transformants with pVVAP to express AftA were grown in pre-inoculums, each consisted of 20 mL of 7H9 medium with 0.5% glucose, 0.05% Tween 80 and 15 µg/mL Kan. The starter culture was grown overnight at 37 °C with shaking at 200 rpm. The following day, the inoculum was prepared (20 mL 7H9 medium, 200 µL of 20% succinate, 100 µL of 10% Tween 80 and 50 µL/mL Kan) with a starter culture at an $OD_{600}$ of 0.02 growth. The cultures were then incubated at 37 °C with shaking at 240 rpm until the cells reached an OD of 0.8 (~48 h). Once this $OD_{600}$ was reached, the cells were induced by addition of 0.2% acetamide. The cultures were once again grown overnight at 37 °C with shaking at 200 rpm. After 2 overnights incubations, the cells were picked by centrifugation and the final pellet was collected.

To begin the expression assay, the resulting pellet was then ressuspended in a lysis solution (Bugbuster reagent; 0.1 mg/mL Lysozyme; 3 U/mL Benzonase; 2 mM $MgCl_2$ and 0.5 mM PMSF) containing 1:1000 of protease inhibitors and lysed by vigorous vortexing for 20 minutes. Afterwards, DDM was added to each sample and these were placed in a cold room to shake for 1 h. Subsequently, the cells were centrifuged at 12000 rpm for 20 minutes at 4 °C. A sample of the supernatant was then loaded in a 10% SDS-PAGE gel.

## 2.6 Minimum inhibitory and bactericidal concentration assays

To investigate the antimycobacterial potential of the compounds obtained through a molecular docking search that intended to repurpose currently employed drugs as AftA inhibitors, MIC assays were performed as described elsewhere [73]. First, the compounds were diluted in media to achieve the desired initial concentration of 1028 µg/mL. A volume of 200 µL of each compound was added to the first row of a 96-well plate (VWR). Six two-fold serial dilutions of each compound were performed, and the last two plate rows were used as positive (media with bacteria) and negative controls (only medium).

The bacteria were treated and added to the plates at $OD_{600}$ of 0.002 (2x $10^5$ CFU/mL) as previously described [74]. After 2 days of incubation, the plates were inspected for signs of mycobacterial growth and the MIC was determined by direct observation. Three independent assays were performed and then the median values were calculated.

For the growth curves in the presence of the compounds in study, *M. smegmatis* cultures were first grown to log phase, treated and, normalized to an $OD_{600}$ of 0.006 in a final volume of 20 mL. The bacterial cultures were allowed to grow overnight in erlenmeyer flasks. The next day, the compounds were added to the cultures at 1x, 2x or 4x the MIC, depending on the MIC value. The cultures were then diluted to achieve an initial concentration of 0.4 in a final volume of 5 mL. The $OD_{600}$ of each culture in study was measured at timepoint 0 h and then every 1.5 h for almost 8 h.

The MBC assay for compound A was performed as described previously with some alterations [48, 75]. After MIC determination, a 50 µL sample from the 1x and 4x MIC assays was transferred to a new 96-well plate and then, 5 µL of each well were platted into 7H10 agar plates. The plates were incubated at 37°C for three days.

The MBC is defined as the lowest concentration of antibiotic capable of killing the tested bacteria. The tested antibiotics were considered bactericidal if the MBC$\leq$ 4x MIC. In contrast, if MBC >4x MIC, the antibiotics were considered bacteriostatic. Two independent MBC assays were performed for compound A and, the MBC medians were then calculated.

## 2.7 AftD purification

The purification of AftD began with its overexpression in *M. smegmatis*. The pre-inoculum, for 8L of growth, consisted of 50 mL of 7H9 medium (with 0.5% glucose, 0.05% Tween 80 and 15 µg/mL Kan) and *M. smegmatis* colony with PVVAP to express AftD, from a 7H10 plate with 0.5% glucose and 15 µg/mL Kan. The starter culture was grown overnight at 37 °C with shaking at 200 rpm. The following day, the inoculum was prepared in 900 mL 7H9 medium, 9 mL of 20% succinate, 4.5 mL of 10% Tween 80 and 50 µL/mL Kan with a starter culture at an $OD_{600}$ of 0.02 growth in 2 L baffled flasks each. The cultures were then incubated at 37 °C with shaking at 240 rpm until the cells reached an $OD_{600}$ of 0.8 (~48 h). Once this OD was reached, the cells were induced by addition of 0.2% acetamide. The cultures were once again grown overnight at 37 °C with shaking at

200 rpm. After 2 overnights incubations, the cells were picked by centrifugation at 4500 rpm using a JA-10 rotor for 30 min at 4 °C. The supernatant was discarded and the pellet was resuspended in culture medium in a smaller volume and centrifuged again at 6000 ×g for 20 min at 4 °C, to ensure that all the cells were collected.

Once the final pellet was obtained, we started the purification of AftD. The purification process began with the lysis of the cells, in which the cells were passed through a cell disruptor at 22Kpsi 4 times. Lysis was performed using Buffer A (200 mM HEPES at pH 7, 500 mM NaCl, 20 mM $MgSO_4$, 1 mM PMSF, 1:1000 protease inhibitors, and 4 µL/100 mL Benzonase). Afterwards, the cells were placed in ultracentrifuge tubes to be centrifuged at 42000 rpm for 40 min at 4 °C (using an Ultracentrifuge Optima LE-80K). Thereafter, the supernatant was discarded and the membrane pellet was homogenized again using a glass homogenizer, this time with Buffer B (20 mM HEPES pH 7, 200 mM NaCl, 20 mM $MgSO_4$, 1 mM PMSF and 1:1000 protease inhibitors). After being homogenized, the cells were centrifuged at 42000 rpm for 30 min in an ultracentrifuge at 4 °C. The resulting pellet was resuspended in Buffer B and 1% DDM was added to the mix, which was then incubated for 2 hours at 4 °C with slow shaking. Afterwards, the mixture was centrifuged at 42000 rpm for 30 min at 4 °C in an ultracentrifuge.

The purification was performed using a flag resin (Sigma-Aldrich), which has high specificity for the flag tag fused with AftA and can be reused up to 7 times. In this work, 1.5 mL of flag resin was used, and it was its third time. After the first wash with a wash buffer (20 mM HEPES pH 7, 150 mM NaCl and 0.5% DDM), the resin is placed on a gravity column and washed again with the same buffer. Once prepared, the supernatant from the last centrifugation is added to the resin and left to incubate overnight at 4 °C with gentle agitation.

The following day, the resin was added to a gravity column and the flow through (FT) is collected. The resin was washed with 10 column volumes of wash buffer, followed by 4 column volumes of elution buffer (20 mM HEPES pH 7, 150 mM NaCl and 0.1% DDM). After 4 column volumes of elution buffer, 5 mg/mL of Flag peptide, was added to the column so that the protein of interest was released from the resin and collected. The samples were then loaded in a 10% SDS-PAGE gel.

## 2.8 *M. smegmatis* activity assays

*M. smegmatis* cells were grown for the activity assays using the same protocol as for protein purification. Along with the cells overexpressing AftD, *M. smegmatis* WT cells were also grown to be used as a negative control in the assays. The first step for the activity assays is the preparation of the membranes. The pellets were lysed in Buffer C (50 mM MOPs at pH8, 200 mM NaCl, 20 mM $MgCl_2$, 1 mM PMSF, protease inhibitors, and 4 µL/100 mL Benzonase), with the use of a cell disruptor at 23 Kpsi. After being lysed, the cells were ultracentrifuged at 27000 xg for 1 hour at 4 °C. Both the resulting supernatants and the pellet will be used: the pellet will be used to prepare the p60 fraction and the supernatants to prepare the bacterial membrane, both of which will be crucial in the activity tests.

1. The supernatants were once again ultracentrifuged at 1000000 xg for 2 hours at 4 °C, and the resulting pellets correspond to the bacterial membrane from *M. smegmatis* WT and overexpressing AftD to be used in the assay,

2. The pellet was resuspended in 10 mL of Buffer C and then percol was added, aiming for a 60% suspension, followed by ultracentrifugation at 27000 xg for 1 hour at 4 °C.

The ultracentrifugation with percol results in different phases, since a density gradient centrifugation is used and the p60 fraction has a known specific density with the appearance of a flocculent white layer. This layer was collected and washed 3 times with Buffer C. Between each wash, a centrifugation step at 15000 rpm, 4 °C for 4 minutes was performed. After the last centrifugation, the obtained pellet was resuspended with 1 mL of Buffer C, resulting in the p60 fraction that was used in the activity assays.

For the activity assays, 150 mg/mL of bacterial membranes (wild type or overexpressed AftD), 100 mg/mL of P60 fraction, 1 mM of acceptor compound, 2 mM donor compound (linear penta-arabinose, and $^{13}$C-labeled Z-FPA, respectively) and Buffer D (50mM MOPs at pH8, 200 mM NaCl, 5% glycerol, 10 mM $MgCl_2$ and 2 µL/mL of protease inhibitors) were mixed. These compounds were synthesized by Cristiano Conceição, a PhD student in Dr. Rita Ventura's laboratory at ITQB NOVA. In total, 6 assays of 2 mL each were prepared and then divided into equal parts. One of them was added directly to 1 mL of absolute ethanol (quench/stop reaction), serving as time 0 (T0), the start of the reaction. The other was first incubated at 37°C for two hours. After incubation, 1 mL of absolute ethanol was added to stop the reaction ( quench reaction).

After all assays were prepared, 1 mL was left at 37 °C for 2 hours (T=2) and another milliliter was added in eppendorfs with absolute EtOH (T=0). Finally, all reactions were centrifuged at 12000 rpm for 5 min at 4°C. The supernatant was collected, vacuum-dried, dissolved in deuterated water and analyzed by NMR.

# 3

# Results and Discussion

## 3.1 Construction and characterization of *aftA*, *aftD* and *embB* knockdown mutants in mycobacteria

### 3.1.1 Construction of *aftA*, *aftD* and *embB* knockdown mutants in *M. smegmatis* mc$^2$ 155 using CRISPRi

Given the demonstrated success of the CRISPRi-dCas9 system in repressing mycobacterial genes, including essential genes [60, 76], the first objective of this project was to construct *aftA*, *aftD* and *embB* knockdown mutants in *M. smegmatis* mc$^2$ 155.

The first step to construct the *aftA*, *aftD* and *embB* knockdown mutants in mycobacteria was to design various sgRNAs, with different PAM strengths and different targeting sites as detailed in the material and methods section.

### 3.1.2 Phenotypic characterization of *aftA*, *aftD* and *embB* knockdown mutants in *M. smegmatis*

#### 3.1.2.1 Spotting dilutions

In order to investigate the effects of CRISPRi-mediated gene silencing on mycobacterial growth, spotting dilutions were performed. All knockdown mutants and controls were investigated, with or without 100 ng/mL ATc, and the results of spotting dilutions are shown in **Figure** 3.1.

The first conclusion from the obtained results is that both negative controls, WT and PLJR962, show similar growth, with and without ATc induction, suggesting that ATc is not toxic to mycobacteria. These results align with earlier research [77]. In contrast, the knockdown mutants displayed major growth defects, consistent with the expected essential role of both genes in mycobacteria [53, 64].

The gene suppression produced by the ATc-inducible CRISPRi system used in this study is influenced by a number of variables, including the complementarity that exists between the sgRNA-DNA target and the strength of the PAM. It is expected taht a more robust PAM will lead to greater gene fold-repression and, consequently, increased growth

inhibition for essential genes [64]. Likewise, as can be seen in **Figure** 3.1, *aftA*.1 (PAM 11) shows greater repression when compared to *aftA*.2 (PAM 14).

The cell viability of two *aftD* knckdown mutants was also tested. As an essential gene, the knockdown of *aftD* should result in observable growth defects. However, no growth deficits were observed in both cases, suggesting an inefficient silencing of *aftD*. The employed PAMs (PAM 13 and 14) are likely not strong enough to produce an efficient binding of dCas9 to the target sites. In addition, the considerable size of the *aftD* gene seems to be a fundamental factor for the lack of repression efficiency.

Since *embB.1* and *embB.2* knockdown mutants have the same PAM (PAM 13), similar growth phenotypes were expected. Contrarily, *embB.2* exhibited more growth defects than *embB.1*. Because *embB.1* targets a site at the beginning of the gene (around 679 bp) while *embB.2* targets a site at the middle of the gene (around 1038 bp), greater viability defect phenotype was expected for this mutant. Surprisingly, that did not occur – likely due to other factors affecting the efficiency of CRISPRi. The observed silencing results indicate that CRISPRi can achieved substancial gene knockdown regardless of the distance from the target site, as previously revealed by Rock *et al.* [76].

Both the *aftD* and *embB.1* knockdown mutants did not show any viability defect phenotype, and for that reason no further studies were carried out with these mutants.

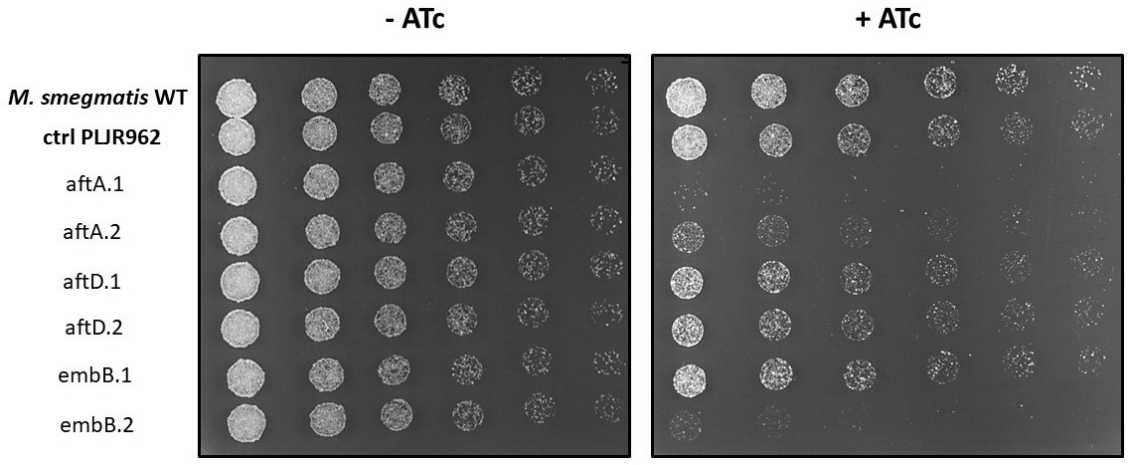


Figure 3.1: **Spotting dilutions of *M.smegmatis* mc$^2$ 155 WT, PLJR962, and *aftA*, *aftD* and *embB* knockdown mutants in the presence and absence of ATc (n=6).**

According to the literature, the AftA protein acts first in AG biosynthesis, followed by the Embs proteins [13]. It is also known that the EmbB protein acts more than once in AG biosynthesis, whereas AftA is predicted to act only once in the beginning. Beyond this, it is believed that the proteins cannot functionally replace each other, the EmbB protein needs AftA to act in order to act as well [78]. These suggestions could possibly justify the results of the spotting dilutions, where although with a weaker PAM, the *embB* gene was the most repressed, affecting the viability of the bacteria the most.

### 3.1.2.2 Growth curves

To further answer the effects of CRISPRi-mediated *aftA* and *embB* silencing on mycobacterial growth, *M. smegmatis* WT, PLJR962 and the mutants were incubated with and without inducer for 45 h, during which the $OD_{600}$ was registered. The growth curves, that is, the evolution of bacterial growth given by the $OD_{600}$ over time, are shown in **Figure** 3.2.

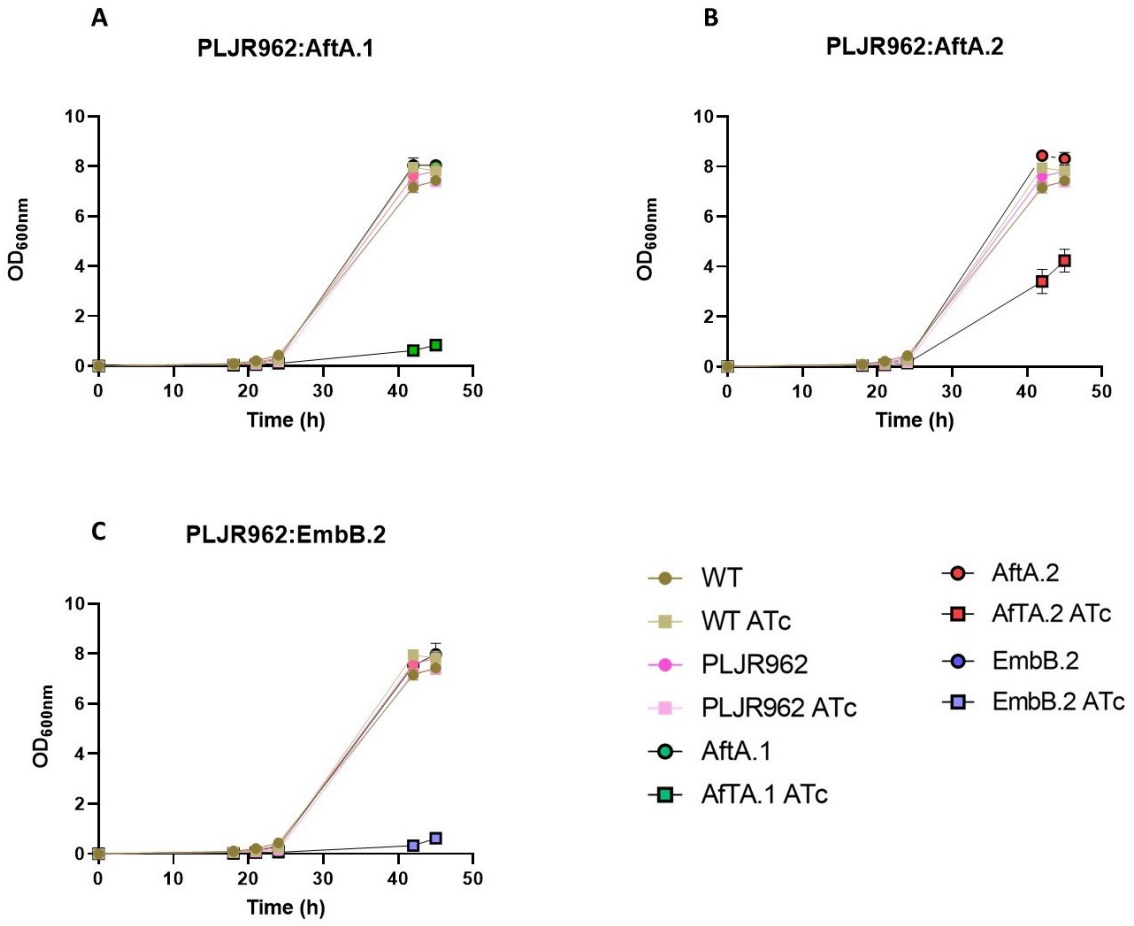


Figure 3.2: **(A-C) Growth curves of *M. smegmatis* mc$^2$ 155 WT, PLJR962, and *aftA* and *embB* knockdown mutants with and without ATc.** *M. smegmatis* mc$^2$ 155 WT and ctrl PLJR962 were used as controls. The cultures were grown from mid- to late-log phase, washed and diluted to a theoretical $OD_{600}$ of 0.001 (t = 0 h). The 100 ng/mL ATc addition was made at 0 and 24 h. The growth profile of cultures was followed by $OD_{600}$ measurement at 0, 18, 21, 24, 42, and 45 h. The growth of each culture in function of the incubation time was evaluated in at least three different experiments (n=3), after which the average was calculated. The error bars represent the SEM.

Similar to the spotting dilutions results, all the controls had a similar and normal growth in the presence and absence of the inducer, confirming once more, a non-toxic effect of ATc [77].

Likewise, all induced knockdown mutants exhibited a reduced growth rate when compared to the controls, supporting the predicted essentiality of *aftA* and *embB* and the fact that its inhibition can lead to effects in cellular division [64, 79].

Comparing the growth curves of the *aftA* knockdown mutants, *aftA*.1 (PAM11, +318 bp) is associated with the strongest PAM and showed the most notorious growth inhibition phenotype when compared to *aftA*.2 (PAM14, +721 bp). Interestingly, of the three mutants analyzed, the *embB* knockdown mutant (PAM13, +1038 bp) was the one with the most notable viability deficiency, although it has a stronger PAM when compared to the *aftA*.2 mutant.

The results of both the spotting dilutions and the growth curves support the essentiality of *aftA* and *embB* genes in *M. smegmatis* and show the success of the targeted gene knockdown using the CRISPRi method. The verified essentiality classification is in accordance with previous literature reports [80].

### 3.1.3 Assessment of CRISPRi-mediated gene repression by qRT-PCR

In order to confirm the efficiency of the CRISPRi system and to complement the results obtained through the growth curves and spotting dilutions assays, a qRT-PCR was performed with the control strains *M.smegmatis* WT, ctrl PLJR962 and the two knockdown mutants, once more, with and without ATc. Since polarity effects had been reported using this system [64], besides quantifying the *aftA* and *embB* mRNA expression levels, the expression levels of the genes immediately upstream and downstream were also measured. In the case of the *aftA* gene (**Figure** 3.3) *aftA* has the ms_6385 (upstream) and the ms_6387 (downstream) genes and, the *embB* gene has the ms_6388 (upstream) and the ms_6390 genes (downstream).

CRISPRi-mediated repression is achieves maximum inhibition between one and two cell division [64]. The mRNA levels were measured using the $\Delta\Delta$CT method and normalized against the *M. smegmatis* housekeeping gene *sigA*. This gene is vital and strikingly resembles the primary sigma factors from other bacteria, indicating that it is an essential sigma factor in mycobacteria [81]. *M. smegmatis* WT was used as calibrator sample.

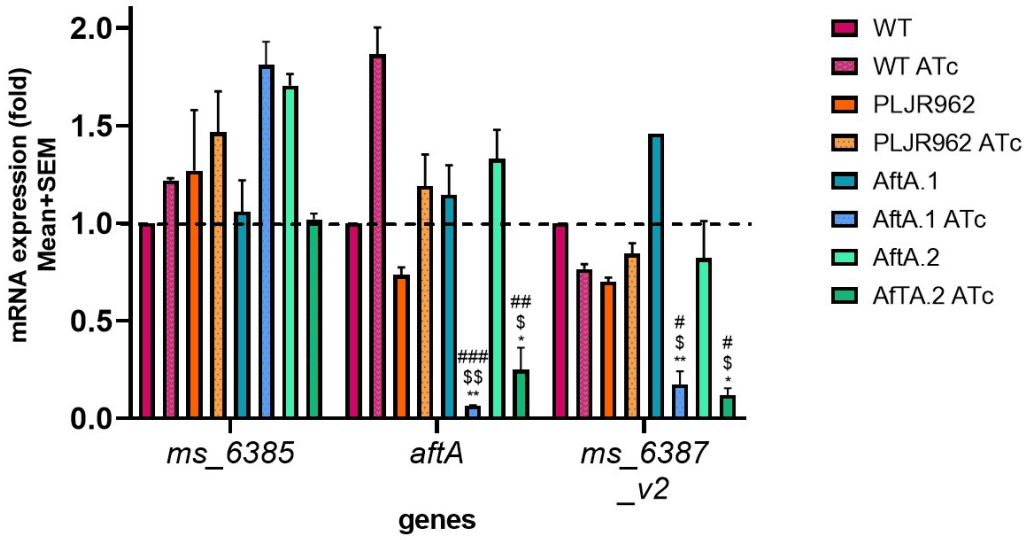


Figure 3.3: **Relative ms_6385, *aftA* and ms_6387 mRNA expression levels in *M. smegmatis* mc$^2$ 155 WT, PLJR962, and *aftA* knockdown mutants with and without 100 ng/mL ATc treatment for 6 h (n=2).** The graph shows the mean relative mRNA expression of genes of interest normalized to *sigA*. Error bars show the standard error of the mean. Multiple comparisons were made using one-way ANOVA, * P < 0.05; ** P < 0.01; *** P < 0.001. Significant differences are specified with symbols: # comparing to control *M. smegmatis* WT ATc, $ comparing to PLJR962 ATc, * comparing to the respective non-induced strain.

**Figure** 3.3 shows that there was a clear reduction in the mRNA expression levels of the *aftA* gene, proving that the CRISPRi system is functioning properly, and the gene is effectively repressed. The same can be observed for its downstream gene (ms_6387), indicating that there is a polar effect. In fact, the ms_6387 and *aftA* genes are under the same promoter and therefore suffer from the repression of *aftA*.

Previous studies have stated three determinants of CRISPRi efficacy, the PAM strength; targeting strand and PAM position along the targeted gene [60, 65]. In terms of PAM strength, stronger PAMs are expected to produce a higher level of knockdown, and in this study that was observed, since there is noticeably higher repression in *aftA*.1 which has a PAM 11, compared to *aftA*.2 which has a PAM 14.

In the case of the *embB* knockdown mutant, the repression of mRNA expression levels was observed (**Figure** 3.4). This confirms once again the effectiveness of the CRISPRi system and efficient target knockdown.

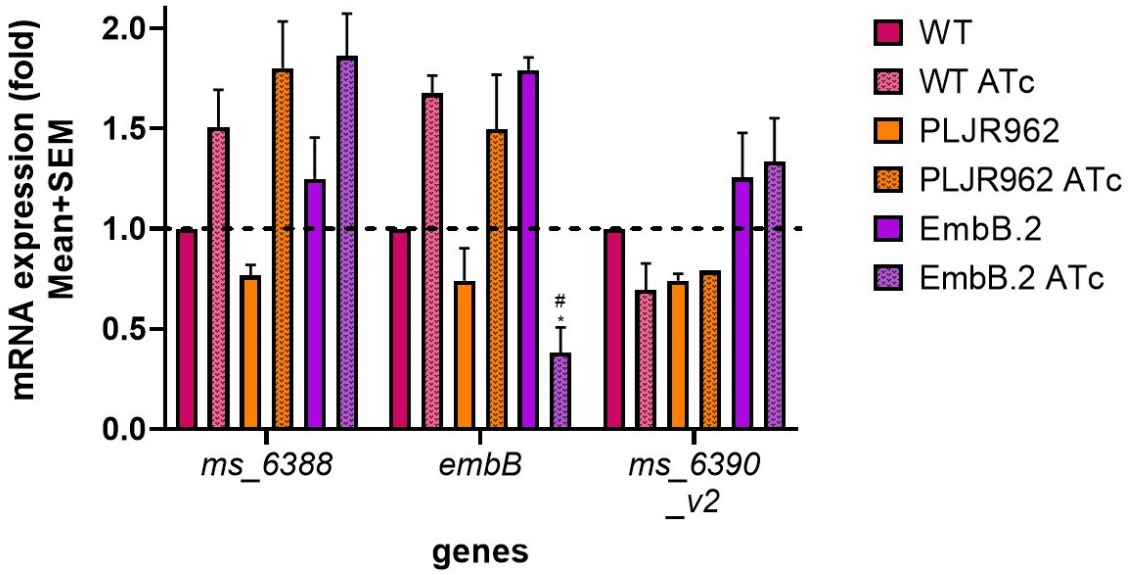


Figure 3.4: **Relative ms_6388, *embB* and ms_6390 mRNA expression levels in *M. smegmatis* mc$^2$ 155 WT, PLJR962, and *embB* knockdown mutant with and without 100 ng/mL ATc treatment for 6 h (n=2).** The graph shows the mean of the relative mRNA expression of genes of interest normalized to *sigA*. Error bars represent the standard error of the mean. Multiple comparisons were made using one-way ANOVA, * P < 0.05; ** P < 0.01; *** P < 0.001. Significant differences are indicated with symbols: # comparing to control *M. smegmatis* WT ATc, $ comparing to PLJR962 ATc, * comparing to the respective non-induced strain.

In this case, no polar effects were observed. The *embB* gene is the last in its operon and its repression does not affect the transcription of neighboring genes, whether upstream or downstream. Comparing the differences between the bars in the graph of the *aftA* ATc and *embB* ATc gene, a lower repression is observed in the *embB* gene, which can be justified because it has a weaker PAM (PAM 13) compared to *aftA*.1 (PAM 11).

Interestingly, the same was not observed with the *aftA*.2 gene, which although it had a weaker PAM (PAM 14) than the *embB* gene (PAM 13), there was greater repression of mRNA expression levels with the *aftA*.2 gene. Even though this is not expected, several factors can affect the CRISPRi system machinery, in addition to the PAMs, and the expression of genes can be different during the growth phases, so the *aftA*.2 gene could be less expressed during the growth phase in which the qRT-PCR was performed.

These results demonstrate the capability of CRISPRi to significantly inhibit the expression of a particular gene in *M. smegmatis*, with the previously discussed phenotype characterization experiences, demonstrate that this inhibition is sufficient to result in a deficiency in bacterial growth.

## 3.2 Construction of AftA expression vector

Given that the gene coding for AftA protein and the PVVAP expression vector had already been amplified during a previous work at ITQB, the first step was to construct the *aftA* expression vector through the Gibson assembly technique. After this reaction, *E. coli* XL-10 competent cells were transformed with the *aftA* expression vector using the heat shock method and, the first transformants were obtained. The first observation from these results was the appearance of many satellite colonies on the plates, which would be false positives as a result of the transformation. Satellite colonies are colonies smaller than the transformants, which can appear due to prolonged incubation when the antibiotic has lost its effect, or at the time of plating, the medium was added while it was still hot and the antibiotic may have denatured. In order to avoid these satellite colonies, the Gibson reaction as well as the transformation were repeated using twice the concentration of antibiotic, Kan (30 µg/mL).

Once kanamycin resistant colonies and the desired transformants were obtained, it was necessary to confirm that these were effectively positive clones. In this way, the next steps were crucial: colony PCR and enzymatic digestion, both with subsequent verification through agarose gel electrophoresis.

### 3.2.1 Colony PCR

In this work, two different strategies were tested for Colony PCR. In parallel with the Colony PCR, the colonies repicked for this technique were also placed in a pre-inoculum together with LB and Kan medium to observe if they would grow in liquid medium. From the results of this pre-inoculum, colonies that did not grow in liquid medium were immediately considered false positives. Those that grew in liquid medium supplemented with Kan were expected to result in a band after the Colony PCR, if a band was not observed, the transformants probably only have the sequence of the pVVAP plasmid, which recirculated during the Gibson reaction process.

In both employed strategies, colonies were picked from the plate using a pipette tip. In the first strategy, when the colony was picked, it was rubbed against the walls of the eppendorf, followed by the addition of the components for the Colony PCR. In the second strategy, the colony was picked and diluted in an eppendorf with 20 µL of MilliQ water and, after this dilution, half of the volume would be for the Colony PCR reaction.

The second strategy is believed to be more effective, because it has been found that the success of the reaction is increased by diluting the colony as several cells may be present, which will inhibit the reaction [55]. However, the results were negative for both strategies.

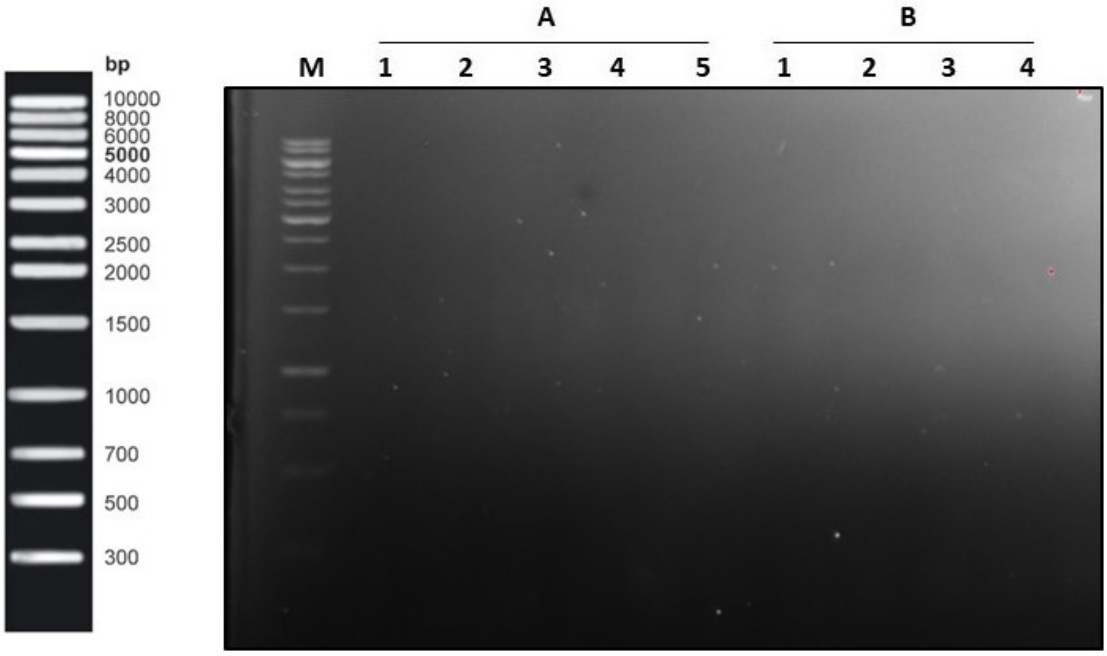


Figure 3.5: **0.7% Agarose gel of the Colony PCR.** This electrophoresis was meant to verify the colony PCR of AftA insert in the pVVAP plasmid. **A** and **B** correspond to different plates and each number to a different colony on the respective plate.No positive results were obtained. The employed molecular weight marker was Thermo Scientific Gene Ruler 1 Kb DNA Ladder.

### 3.2.2 Enzymatic digestion

Another technique, enzymatic digestion, was used since the Colony PCR results were inconclusive and some pre-inoculums showed bacterial growth.

To verify the presence of a positive clone, the growing inoculums were used for the purification of plasmid DNA with a commercial kit (Miniprep kit, NZYtech). The resulting DNA samples were then analyzed using a UV-Vis spectrophotometer (NanoDrop), and only samples with good concentration and purity parameters were selected for the enzymatic digestion step. Two restriction enzymes were tested for the reaction: the first was the Not*I*-HF enzyme which did not show good results in the agarose gel; for this reason the Xba*I* enzyme was used.

The enzyme Not*I*-HF would make a double cut on the pVVAP plasmid with the AftA protein sequence, where is expected to observed two bands in the gel, which should sum to approximately 9000 bps (a 7 kbps band for the plasmid + a 2kbps band for the AftA sequence). The enzyme Xba*I* would make a single cut, where is expected to observe one band the size of 9000 bps. One band the size of the plasmid (7 kbps) was considered a negative result for both enzymes tested.

In this work, using the Xba*I* enzyme, two positives clones were obtained as shown in **Figure** 3.6.

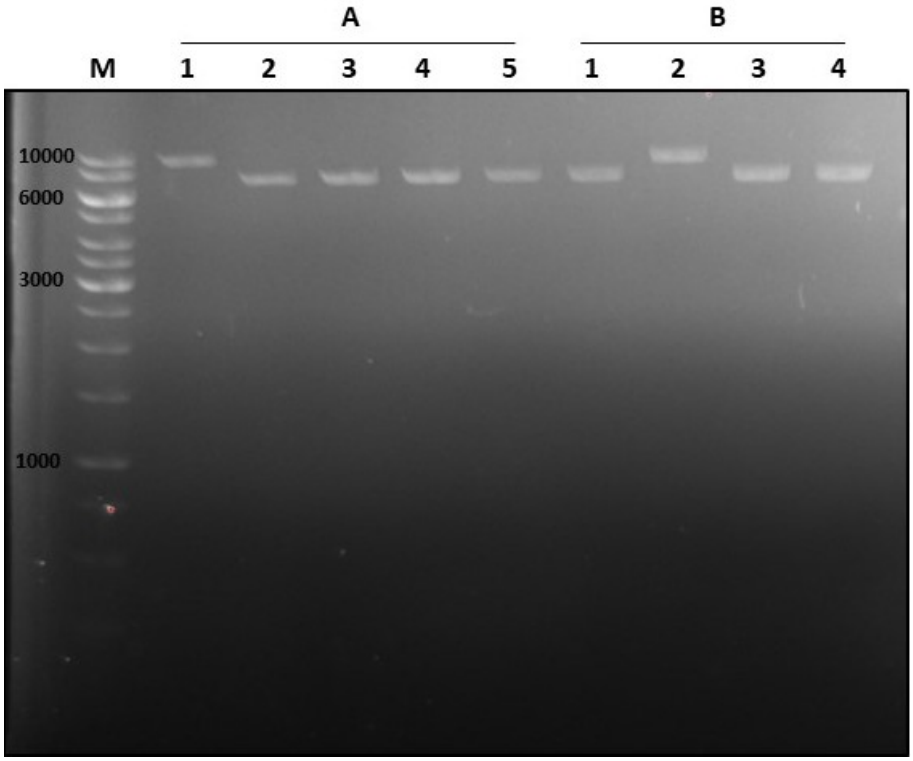


Figure 3.6: **0.7% Agarose gel of the enzymatic digestion with the cutting enzyme Xba*I*.** This electrophoresis was to verify the enzymatic digestion of the AftA insert in the PVVAP plasmid. **A** and **B** correspond to different plates and each number to a different colony on the respective plate. A1 and B2 were the positives obtained from the reaction. The molecular mass marker used was Thermo Scientific Gene Ruler 1 Kb DNA Ladder.

To verify whether the obtained clones represented the successful error-free cloning of AftA in the pVVAP vector, one DNA sample from each clone was sent for sequencing. The analysis of these results was performed with the ExPASy and Clustal Omega programs. From the obtained results, only the A1 sample showed a correct alignment with the AftA protein sequence and was, therefore, used to proceed to the next step, the transformation of *M. smegmatis* cells with the recombinant plasmid pVVAP.

### 3.2.3 AftA expression assay

Although *M. smegmatis* has already been transformed with pVVAP-AftA, it is necessary to understand if the obtained colonies are producing the desired protein efficiently.

Since the aim is to understand if the cells are expressing the AftA protein correctly, the assessment began with the lysis of the cells, so, BugBuster (reagent with the ability to break the cell wall) was added and the cells were lysed by vigorous vortexing for 20 minutes. Afterwards, DDM (detergent that ruptures the membrane and ensures that the membrane proteins remain in the soluble part) was added to each sample and placed in a cold room to shake for 1 h. Thereafter, the cells were centrifuged and, as micelles DDM-AftA expected to be soluble, our protein of interest should be in the supernatant,

each were loaded in a 10% SDS PAGE gel. AftA has a size of 67 KDa, so a band around that weight was expected to be observed.

The obtained results were not as expected since none of the bands observed had the size of the AftA protein. These results may have occurred because of an inefficient lysis, or because the sample may have been too diluted when the lysis buffer were added. Several factors can also affect the growth of colonies in the inoculum such as temperature, shaking or the induction OD. Although not the reason for the non-appearance of the AftA band, the gel in **Figure** 3.7 should also have been left running for longer. A possible alternative to this lysis assay would be to perform a western blot with anti-histidine antibody in order to confirm if there were positive bands on the gel, since the expression levels were low.

Due to time constraints, it was not possible to assess why no positive results were obtained in the expression assays.

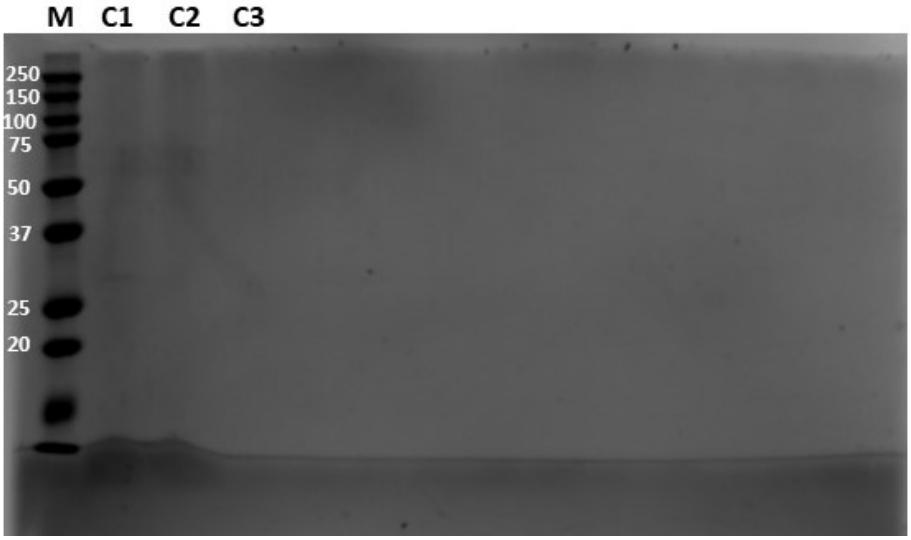


Figure 3.7: **10% SDS-PAGE gel to analyze AftA expression.** SDS-PAGE gel with the result of the expression test for AftA protein. **C** represents each colony selected for testing (3). Although there appear to be bands between 75 and 50 KDa in samples C1 and 2, further studies are nedeed to confirm the expression. The molecular mass marker used was Precision Plus Protein Unstained Standards (Bio-Rad).

## 3.3 AftD purification

In parallel with the previous work with AftA, this project at ITQB also aimed to purify the AftD protein. When I started this work at ITQB, *M. smegmatis* cells had already been transformed with the vector meant to express AftD and a positive clone had already been found. Following this work, it was possible to grow the bacteria expressing AftD in order to purify it and, later follow up with activity assays.

In order to have a deeper perception of the function of AftD at a structural level, we first started by purifying it in *M. smegmatis*. Although its structure has already been determined by cryo-EM, with the protein being expressed in *E. coli*, there may be conformational differences as AftD was purified through heterologous production. When arabinosyltransferase proteins are produced in *E. coli*, they are not part of the organism's system, while in mycobacterium they are already part of it and play an active role.

Furthermore, an ACP domain has been reported to bind to the structure of both AftD and Embs when purified [48, 49]. For this reason, an AftD mutant in which this interaction is disrupted was constructed. In this way, we can better understand if there is any conformational change or any effect on the proteins activity without the ACP domain.

In this work, the purification of AftD WT and of the triple mutant (Tm) was carried out. In the triple mutant, 3 residues of the protein sequence were exchanged for alanine (Ala) with the aim of disrupting the binding of the AftD protein to the ACP domain.

In this purification, all used buffers were prepared at a pH of 7. The cells were lysed using a French Press at a pressure of 16 Kpsi (cell lysis is usually done using a cell disruptor, however it was out of order) and incubated in the resin overnight to be collected the next day. After this incubation period, the FT was collected on a gravity column where all proteins that had no affinity for the resin were eluted. The resin was then washed with washing buffer and after, the desired samples were finally eluted with an elution buffer containing a flag peptide. The flag peptide is a very specific peptide with affinity to the flag resin where AftD with flag tag was bound. In this way, to the flag peptide facilitates the release and elution of the AftD protein from the resin by competing to its fusion flag tag N-terminal tail. The results were then analyzed on a 10% SDS-PAGE gel.

The results were not successful as the AftD band, which should appear at 150 KDa, was not observed in either the WT or Tm proteins. The elution samples were concentrated to a final volume of 20 µL and, a pool of all samples was prepared to improve the observation of results on the SDS-PAGE gel.

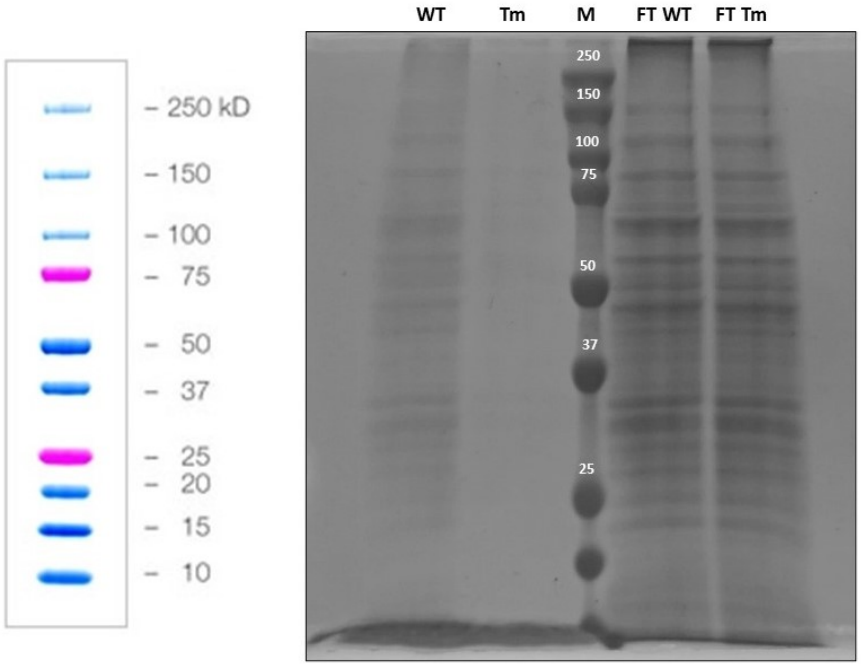


Figure 3.8: **10% SDS-Page gel to analyze the purification of the WT AftD protein and of the Tm AftD protein.** The first two samples in the gel are a pool of all elution samples of Wt and Tm proteins, respectively. M is the molecular mass marker used, Precision Plus Unstained Standards (Bio-Rad). The last two samples correspond to the FT (collected after the incubation period) of WT and Tm proteins, respectively.

Based on **Figure** 3.8, the pool pattern of the AftD WT samples is quite similar to their FT, which indicates that the washing of the resin on the column with the wash buffer may not have been long enough and when the elution of WT AftD started, contaminants were eluted.

Thus, with the results of this first purification, it cannot be concluded whether the protein was not sufficiently expressed during bulk growth or whether there might have been some error during the purification steps, such as an inefficient incubation with the DDM detergent, being the membrane protein insufficiently extracted, A possible alternative to this lysis assay would be to perform a western blot with anti-histidine antibody in order to confirm if there were positive bands on the gel, since the expression levels were low or a bad lysis of the cells with the French Press.

## 3.4 *M. smegmatis* activity assays

In addition to protein purification experiments, activity assays were also performed with the membrane of WT bacteria and with bacteria overexpressing the AftD protein. The aim of these assays is to discover whether AftD has a linear or branched acceptor and thus better understand its function and in what step part of the arabinogalactan biosynthesis the AftD protein acts. The donor compound, Z-FPA, has a $^{13}C$ marked atom and, therefore, the results of the activity assays were analyzed by $^{13}C$-NMR spectroscopy, where changes can be observed through chemical shifts, since it is a C-labeled spectrum and changes on the bonds of the labeled C can be followed.

First, the bacterial membranes and the p60 fraction were prepared for the tests. The P60 fraction is a highly conserved protein in mycobacteria that has been suggested to play an important role in the virulence and survival of *Mtb* [82]. The p60 fraction, is purified in the activity assays of this work in order to help stabilize the AftD protein. Through a density gradient centrifugation with the percol reagent, the cellular components were separated based on their densities, with the heavier ones at the bottom and the lighter ones floating. The p60 protein has a known specific density (1.08-1.10 g/mL) and can therefore be identified with the appearance of a flocculent white layer (about 60% of the suspension).

For the activity assays, bacterial membranes (WT or overexpressed with AftD), P60 fraction, the acceptor compound, the donor compound (Penta-arabinose, linear and Z-FPA with a $^{13}C$ tag, respectively) and Buffer (50mM MOPs at pH8, 200 nM NaCl, 5% glycerol, 10 mM $MgCl_2$ and 2 µL/mL of protease inhibitors) were mixed. In total, 6 assays were prepared and then divided into equal parts. One of them was added directly to 1 mL of absolute ethanol (quench/stop reaction), serving as time 0 (T0), the start of the reaction. The other was first incubated at 37°C for two hours. After incubation, 1 mL of absolute ethanol was added to stop the reaction ( quench reaction).

After all assays were prepared, 1 mL was left at 37 °C for 2 hours (T=2) and another milliliter was added in eppendorfs with absolute EtOH (T=0). Finally, all reactions were centrifuged.

The activity tests were performed in collaboration with Dra. Rita Ventura's laboratory at ITQB and the assay samples (with different types of acceptors and donor Z-FPA) were synthesized by Cristiano Conceição, a PhD student. The supernatant resulting from the last centrifugation of the assay was collected, dried in vacuum and diluted in deuterated water for posterior NMR analysis.

The NMR spectrophotometer was a 500 MHz with a cryo-probe, in which a $^{13}C$ spectrum marked with a scale of 80 to 120 ppm Hertz (frequency unit) was obtained. The presence of a cryo-probe in the spectrophotometer enables a higher specificity in these assays, since the cryo-probe is designed to operate at three fixed frequencies, one of them being $^{13}C$. This probe-head allows the capture of multi-dimensional double resonance spectra from protein samples that have been double-labeled.

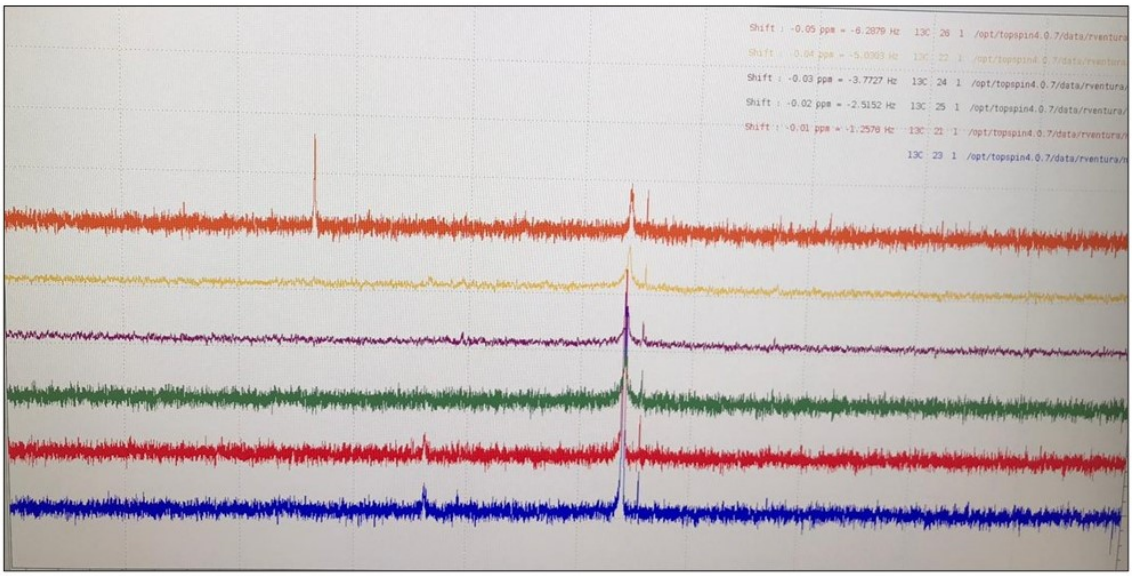


Figure 3.9: **$^{13}$C-NMR spectrum on a scale of 80 to 120 ppm Hertz.** Each line in the spectrum corresponds to 6 assays, all containing p60 fraction. The first 2 lines at the bottom (blue and red) correspond to the negative control assays and their replicate respectively, i.e. without bacterial membrane and only with the Z-FPA donor. The green line corresponds to the test with *M. smegmatis* WT membrane and donor. The purple line to the assay with *M. smegmatis* membrane with overexpressed AftD and donor. Yellow line corresponds to the assay with *M. smegmatis* WT membrane, donor and penta-arabinose acceptor and, finally, the top line (orange) corresponds to the assay with *M. smegmatis* membrane with overexpressed AftD, donor and penta-arabinose acceptor. The signal observed in all of them is the $^{13}$C signal (around 97 ppm). The top line (orange) shows a new chemical shift (around 107 ppm) that results in a larger signal, typical of a glycosidic bond.

As shown in **Figure** 3.9, in the assay with the membrane of *M. smegmatis* overexpressing AftD, donor and acceptor penta-arabinose, it was observed a signal that is typical of a glycosidic bond. This result suggests that the enzyme (membrane overexpressing AftD) transferred an arabinose moiety from sugar donor Z-FPA to linear penta-sugar acceptor causing a chemical shift. More studies are needed with different sugar acceptors to better characterize the enzymatic reaction of AftD. Moreover, this assay will allow the evaluation of the inhibition capacity of the aforementioned selected compounds, such as Lomitapide, directly on the activity of AftD.

Recently the result was repeated in the laboratory and the same results were obtained, confirming that it is indeed a positive result.

## 3.5 Evaluation of the potential of selected compounds as AftA *M. smegmatis* inhibitors through MIC determination assays

To understand whether molecular docking between the AftA from *Mtb* protein and the chosen compounds would indicate increased inhibitory activity against the AftA from *M. smegmatis* protein, thereby leading to a reduction in the viability of *M. smegmatis* WT bacteria, MIC assays were performed.

The MICs were assessed through direct observation and defined as the lowest concentrations at which no bacterial growth was visually observed in the liquid medium. This assay was conducted at least three times for each compound and the results are presented in the **Table** 3.1. Ethambutol and isoniazid MICs were used for comparison, since both are first-line anti-TB drugs [74].

Table 3.1: **MIC values obtained for the tested compounds and first-line anti-TB drugs, ethambutol and isoniazid**

| | MIC (µg/mL) | MIC (M, mol/L) |
|---|---|---|
| **Compound A** | 16-32 | (4.6-9.2) x$10^{-6}$ |
| **Compound B** | 256 | 1.06x$10^{-4}$ |
| **Compound C** | 256 | 1.08x$10^{-4}$ |
| **Compound D** | 512 | 1.65x$10^{-4}$ |
| **Ethambutol** | 0.5-1 | (2.4- 4.9) x$10^{-7}$ |
| **Isoniazid** | 8-16 | (5.8- 1.2) x$10^{-6}$ |

The obtained results were not as expected, since except for compound A, the remaining compounds showed very high MIC values. However, the MICs were performed using *M. smegmatis* and not *Mtb*, and it is known that the cell wall of *M. smegmatis* is much more lipid-rich, thus, being more drug-impermeable when compared to *Mtb* [83]. The prospective inhibitory activity of the compounds in study against the AftA from *Mtb* protein should thereafter be tested with *Mtb* cultures, but these experiments should be carried by BSL-3 trained personnel, and are thus, out of scope for master students.

Growth curves with the compounds were also performed to better understand their effect on the susceptibility of *M. smegmatis* WT. The $OD_{600}$ was measured every 1.5 h for approximately 8 h. Since the compound D presented a very high MIC, it was automatically excluded, and no growth curves were made with it. The first-line anti-TB agents, ethambutol and isoniazid were also tested in the growth curves to serve as reference.

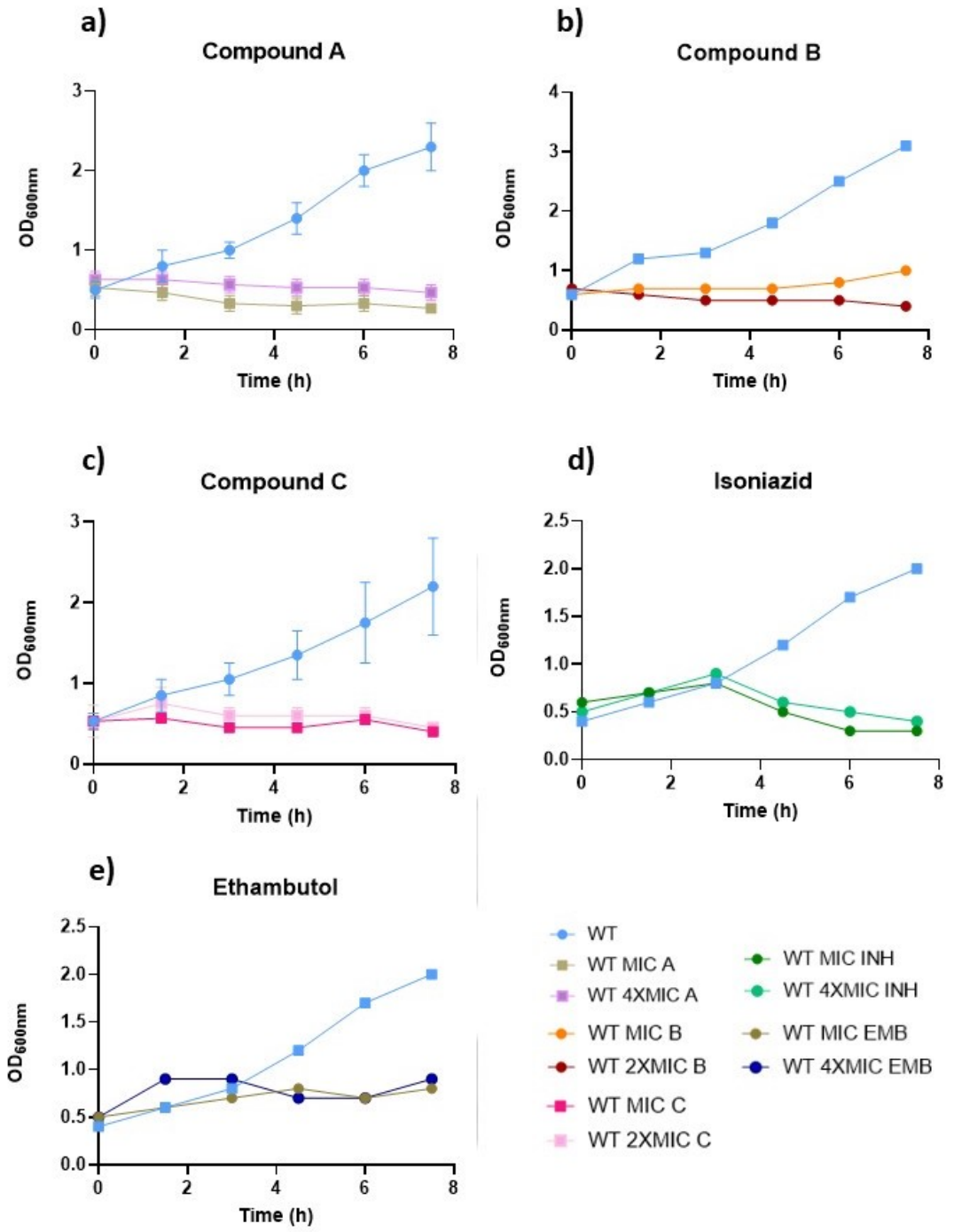


Figure 3.10: **(A-E) Growth curves of *M. smegmatis* mc$^2$ 155, the compounds A, B and C, and the first-line anti-TB drugs Ethambutol and Isoniazid.** *M. smegmatis* mc$^2$ 155 WT was used as a control. The cultures were grown from mid- to late-log phase, and diluted to a theoretical $OD_{600}$ of 0.3-0.4 (t = 0 h). The error bars represent the SEM for compounds A and C.

Interestingly, although the compounds had very different MICs, they showed very similar growth curves.

Based on **Figure 3.10**, of the 3 compounds tested, compound B is the only one that shows a greater decrease curve with the doubling of its MIC concentration. Both compounds A and C lead to a higher growth defect with the MIC concentration than with the quadruple or double of it, respectively. In addition, compound B **(graph b)** at the MIC concentration shows a recovery in the growth of the bacteria in the last $OD_{600}$ measurement, but a decrease with twice the MIC in the same measurement. In the case of compound C **(graph c)**, in the last measurement, a decrease was already observed for both concentrations, MIC and 2x MIC. Compound A **(graph a)** showed more constant growth curves when compared to the other two compounds.

Isoniazid showed a higher growth decrease profile when compared to the tested compounds, however the same was not observed with Ethambutol. Three hours after the addition of the anti-TB drugs, granules were observed in the liquid medium which suggested that the bacteria were clumping and that sedimentation was occurring. This sedimentation may have affected the measurement of ODs and could account for the inconsistent growth curve of Ethambutol.

With these results of growth curves and MICs, although compounds B and C presented very high MIC values, all compounds were shown to affect the growth of *M. smegmatis*. Since, compound A showed a much lower MIC value, it was possible to perform several independent growth curves assays. The remaining compounds had high MIC values, so a larger amount of each was needed to perform the growth curves assays and, for this reason, compound C was tested twice and compound B was tested only once. Overall, compound A showed the best MIC value and a more constant growth curve.

### 3.5.1 Minimum bactericidal concentration assay

To better understand the potential of the compound A, the MBC assay was performed. This assay aims to evaluate if the compound has the ability to only inhibit bacterial growth in vitro (bacteriostatic activity) or if it can kill the bacteria (bactericidal activity). In each assay, there was no bacterial growth on the 7H10 plates either at their MIC concentration or at 2x or 4x MIC, indicating MBC of compound A to be around 16-32 µg/mL. In this way, we were able to conclude that this compound is bactericidal.

The results of both the MIC and MBC assays indicate potential of compound A as an AftA inhibitor and possibly as a future new anti-TB drug.

# 4

# Conclusions and Future Perspectives

Over the past few decades, TB has consistently ranked among the leading causes of death, claiming the lives of millions of people annually. TB is the second leading infectious disease, from a single pathogen *Mycobacterium tuberculosis*, right after COVID-19.

The main objective of this thesis was to investigate the role of the mycobacterial glycosyltranferases AraTs and their potencial to be promising targets for the development of new antituberculosis drugs. For this purpose, various techniques were used not only to understand its essentiality in mycobacteria, but also its structure and function.

The CRISPRi platform optimized by Rock *et al.* is a very useful and effective tool for silencing essential mycobacterial genes. This system allowed for the inhibition of the *aftA*, *aftD* and *embB* genes.

Using this system, the mycobacterial essentiality of the *embB* and *aftA* genes was investigated, and the obtained results confirmed its essentiality for mycobacteria. Through spotting dilutions and growth curves it was found that the level of *aftA* and *embB* knockdown appeared to be directly related with the associated PAM strength. However, although the chosen PAMs are important to dictate silencing efficiency, there are other factors that can interfere. For instance, the distance of the target gene from the transcription start site, the length of the employed sgRNA, the amount of GC content of the targeted region but also, other factors, such as the basal level of expression of the target genes or the role that each one plays in the synthesis of the mycobacterial wall, affect the efficiency of knockdown. The *aftD* knockdown mutant did not show any viability defect phenotype and, for that reason, no further studies were carried out with this mutant.

The level of knockdown was also assessed by qRT-PCR, and a significant decrease in mRNA expression levels of *aftA* and *embB* was observed. This result, together with phenotype characterization experiments, supports the ability of CRISPRi to lead to considerable and efficient target inhibition in *M. smegmatis*. Likewise, it indicates that the produced target inhibition is enough to cause a defect in growth, that is, a proportional phenotypic effect.

After confirming that the *aftA* gene was essential for the viability of the bacteria, further studies were carried out using different techniques to better understand its role

in the synthesis of the mycobacterial cell wall and its structure. First, *E. coli* XL-10 cells were transformed with the pVVAP plasmid containing the sequence of the AftA protein. The success of this protocol was confirmed by 2 methods: Colony PCR and Enzymatic digestion. These two methods were used to confirm that the transformation resulted in a positive clone. Colony PCR did not yield any positive results which is why the screening strategy was changed to enzymatic digestion and the results were subsequently confirmed by agarose gel electrophoresis. Once a positive clone was obtained, and the sequence of the plasmid DNA sample confirmed, it was possible to proceed to transformation of *M. smegmatis* cells with pVVAP-AftA.

Expression activity assays were performed to find out if the obtained transformants were expressing the AftA protein correctly. However, no positive results were obtained and, due to time constraints, it was not possible to confirm whether it was a problem with the protocol, such as inefficient lysis, or if the colonies were in fact not expressing the protein properly. To note that the growth of mycobacteria is not very reproducible.

At ITQB, computational studies were being carried out in parallel using artificial intelligence programs (AlphaFold) to predict the structure of AftA and to search for possible compounds that could bind to the protein and inhibit it function, based on this structure prediction (*in silico* screens - molecular dynamics and molecular docking). During my thesis, I was able to test these compounds using MIC and MBC determination assays at Faculty of Pharmacy.

Although some of the selected compounds presented high MIC values, growth curves were performed at the MIC concentration and the results were very similar to all the tested compounds. There was one compound that particularly stood out, presenting the same MIC value as the first-line anti-TB drug isoniazid, the compound A. In this way, MBC assays were performed with compound A, confirming its bactericidal activity. With these results, all compounds were shown to affect the growth of *M. smegmatis* and should be subject to further investigation, particularly in *Mtb*.

In this work, I also had the opportunity to accompany some on-going studies with the AftD protein. *M. smegmatis* cells had already been transformed with an AftD expression vector and following this work, it was possible to grow mycobacteria expressing the AftD protein in order to purify it and, afterwards, perform activity assays.

Unfortunately, no suitable results were ever achieved with the purification of AftD. Several factors may have affected this, such as the pH of the buffers that were used, ineffective cell lysis, insufficient cell growth/protein expression. More time would be needed to find the cause of these results.

Activity assays were performed with the membrane of *M. smegmatis* WT or with *M. smegmatis* overexpressing the AftD protein. The activity assays were performed in collaboration with Dra. Rita Ventura's laboratory at ITQB. The samples were analyzed by NMR spectroscopy by Cristiano Conceição, a PhD student. A signal was obtained in the C-labeled spectrum that corresponds to a glycosidic bond in one of the assays carried out with the synthesized donor and acceptor. This test was performed with the membrane that

overexpresses the AftD protein and this result was very promising, and further studies will continue.

Given the results obtained in this thesis and all the questions that remain to be answered, future perspectives should focus on:

1. improve the protocol for AftA from *M. smegmatis* expression tests so that, once colonies that correctly express the protein are obtained, the protein purification and activity assays can be carried out as with the AftD protein;

2. carry out checkerboard assays to determine whether the compounds in study, can synergize with first-line anti-TB antibiotics, despite of their MIC values alone;

3. repeat the activity assays with the membrane of *M. smegmatis* overexpressing AftD and optimize the protocol;

4. in addition, new assays should be performed with different acceptors;

5. improve the protein purification protocol to better understand the interaction between AftD and the ACP domain.

To conclude, the work presented in this thesis enlightens the essentiality of the *embB* and *aftA* genes in *Mycobacterium*. In addition, *M. smegmatis* cells were transformed with pVVAP-AftA, so protein purification and activity assays can also be conducted for AftA, similar to the activity assay with AftD. This work also helped implement a drug repurposing strategy where the compounds chosen against the molecular docking of AftA from *Mtb* showed promising results in MIC assays with *M. smegmatis*, with compound A emerging as a possibly future anti-TB drug.

# Bibliography


[1] World Health Organization 2022. *Global Tuberculosis Report 2022*. Tech. rep. 2022 (cit. on pp. 1–3).

[2] Koegelenberg et al. "Tuberculosis: The Past, the Present and the Future". In: (2021). DOI: 10.1159/000516509 (cit. on pp. 1, 3).

[3] Gavin Churchyard et al. "What We Know about Tuberculosis Transmission: An Overview". In: (2017). DOI: 10.1093/INFDIS/JIX362 (cit. on pp. 1, 2).

[4] Marco Schito et al. "Perspectives on Advances in Tuberculosis Diagnostics, Drugs, and Vaccines". In: (2015). DOI: 10.1093/CID/CIV609 (cit. on p. 1).

[5] Cláudio Nunes-Alves et al. "In search of a new paradigm for protective immunity to TB". In: (2014). DOI: 10.1038/NRMICRO3230 (cit. on p. 2).

[6] Peter D.O. Davies. "The role of DOTS in tuberculosis treatment and control". In: (2003). DOI: 10.1007/BF03256649 (cit. on p. 2).

[7] Paolo Miotto et al. "Drug resistance mechanisms and drug susceptibility testing for tuberculosis". In: (2018). DOI: 10.1111/RESP.13393 (cit. on p. 3).

[8] Kwonjune J. Seung et al. "Multidrug-Resistant Tuberculosis and Extensively Drug-Resistant Tuberculosis". In: (2015). DOI: 10.1101/CSHPERSPECT.A017863 (cit. on p. 3).

[9] Christoph Lange et al. "Drug-resistant tuberculosis: An update on disease burden, diagnosis and treatment". In: (2018). DOI: 10.1111/RESP.13304 (cit. on p. 3).

[10] Michael U. Shiloh and Patricia A. DiGiuseppe Champion. "To catch a killer. What can mycobacterial models teach us about Mycobacterium tuberculosis pathogenesis?" In: (2010). DOI: 10.1016/J.MIB.2009.11.006 (cit. on pp. 3, 4).

[11] Mudassar Altaf et al. "Evaluation of the Mycobacterium smegmatis and BCG models for the discovery of Mycobacterium tuberculosis inhibitors". In: (2010). DOI: 10.1016/J.TUBE.2010.09.002 (cit. on p. 4).

[12] Joseph Antony Sundarsingh T et al. "Features of the biochemistry of Mycobacterium smegmatis, as a possible model for Mycobacterium tuberculosis". In: (2020). DOI: 10.1016/J.JIPH.2020.06.023 (cit. on pp. 4, 7).

[13] Katherine A. Abrahams and Gurdyal S. Besra. "Mycobacterial cell wall biosynthesis: a multifaceted antibiotic target". In: (2018). DOI: 10.1017/S0031182016002377 (cit. on pp. 4–6, 8, 9, 30).

[14] Arundhati Maitra et al. "Cell wall peptidoglycan in Mycobacterium tuberculosis: An Achilles' heel for the TB-causing pathogen". In: (2019). DOI: 10.1093/FEMSRE/FUZ016 (cit. on pp. 4, 6).

[15] Sarah A. Stanley and Jeffery S. Cox. "Host–Pathogen Interactions During Mycobacterium tuberculosis infections". In: (2013). DOI: 10.1007/82_2013_332 (cit. on pp. 4, 9).

[16] Luke J. Alderwick et al. "The Mycobacterial Cell Wall—Peptidoglycan and Arabinogalactan". In: (2015). DOI: 10.1101/CSHPERSPECT.A021113 (cit. on pp. 5, 6).

[17] Monika Jankute et al. "Assembly of the Mycobacterial Cell Wall". In: (2015). DOI: 10.1146/ANNUREV-MICRO-091014-104121 (cit. on p. 5).

[18] Jon B. Raymond et al. "Identification of the namH Gene, Encoding the Hydroxylase Responsible for the N-Glycolylation of the Mycobacterial Peptidoglycan". In: (2005). DOI: 10.1074/JBC.M411006200 (cit. on p. 6).

[19] Patrick J. Brennan and Hiroshi Nikaido. "The envelope of mycobacteria". In: (1995). DOI: 10.1146/ANNUREV.BI.64.070195.000333 (cit. on p. 6).

[20] Maria João Catalão et al. "Revisiting Anti-tuberculosis Therapeutic Strategies That Target the Peptidoglycan Structure and Synthesis". In: (2019). DOI: 10.3389/FMICB.2019.00190 (cit. on p. 6).

[21] Mira Holzheimer et al. "Chemical Synthesis of Cell Wall Constituents of Mycobacterium tuberculosis". In: (2021). DOI: 10.1021/ACS.CHEMREV.1C00043 (cit. on pp. 6, 7).

[22] Alex Dajkovic et al. "Hydrolysis of peptidoglycan is modulated by amidation of meso-diaminopimelic acid and Mg2+ in B acillus subtilis". In: (2017). DOI: 10.1111/MMI.13673 (cit. on p. 6).

[23] M Sassetti et al. "Genes required for mycobacterial growth defined by high density mutagenesis". In: (2003). DOI: 10.1046/J.1365-2958.2003.03425.X (cit. on p. 6).

[24] Teresa A. Figueiredo et al. "Identification of Genetic Determinants and Enzymes Involved with the Amidation of Glutamic Acid Residues in the Peptidoglycan of Staphylococcus aureus". In: (2012). DOI: 10.1371/JOURNAL.PPAT.1002508 (cit. on p. 6).

[25] Karen J. Kieser and Eric J. Rubin. "How sisters grow apart: mycobacterial growth and division". In: (2014). DOI: 10.1038/NRMICRO3299 (cit. on p. 6).

[26] Julia Domínguez-Escobar et al. "Processive movement of MreB-associated cell wall biosynthetic complexes in bacteria". In: (2011). DOI: 10.1126/SCIENCE.1203466 (cit. on p. 6).

[27] D. Branch Moody et al. "CD1b-mediated T cell recognition of a glycolipid antigen generated from mycobacterial lipid and host carbohydrate during infection". In: (2000). DOI: 10.1084/JEM.192.7.965 (cit. on p. 7).

[28] Catherine Vilchèze. "Mycobacterial Cell Wall: A Source of Successful Targets for Old and New Drugs". In: (2020). DOI: 10.3390/APP10072278 (cit. on pp. 7–9).

[29] S. T. Cole et al. "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence". In: (1998). DOI: 10.1038/31159 (cit. on p. 7).

[30] D. N. Brindley, S. Matsumura, and K. Bloch. "Mycobacterium phlei Fatty Acid Synthetase—A Bacterial Multienzyme Complex". In: (1969). DOI: 10.1038/224666a0 (cit. on p. 7).

[31] Sylvain Cantaloube et al. "The Mycobacterium Tuberculosis FAS-II Dehydratases and Methyltransferases Define the Specificity of the Mycolic Acid Elongation Complexes". In: (2011). DOI: 10.1371/JOURNAL.PONE.0029564 (cit. on p. 7).

[32] Omita A. Trivedi et al. "Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria". In: (2004). DOI: 10.1038/NATURE02384 (cit. on pp. 7, 8).

[33] Apoorva Bhatt et al. "Loss of a Mycobacterial Gene Encoding a Reductase Leads to an Altered Cell Wall Containing $\beta$-oxo- Mycolic Acid Analogs and Accumulation of Ketones". In: (2008). DOI: 10.1016/J.CHEMBIOL.2008.07.007 (cit. on p. 7).

[34] Mary Jackson et al. "Inactivation of the antigen 85C gene profoundly affects the mycolate content and alters the permeability of the Mycobacterium tuberculosis cell envelope". In: (1999). DOI: 10.1046/J.1365-2958.1999.01310.X (cit. on pp. 7, 8).

[35] Márió Gajdács. "The Concept of an Ideal Antibiotic: Implications for Drug Design". In: (2019). DOI: 10.3390/MOLECULES24050892 (cit. on p. 8).

[36] Benfang et al. Lei. "Action mechanism of antitubercular isoniazid. Activation by Mycobacterium tuberculosis KatG, isolation, and characterization of inha inhibitor". In: (2000). DOI: 10.1074/JBC.275.4.2520 (cit. on p. 8).

[37] Asesh Banerjee et al. "inhA, a Gene Encoding a Target for Isoniazid and Ethionamide in Mycobacterium tuberculosis". In: (1994). DOI: 10.1126/SCIENCE.8284673 (cit. on p. 8).

[38] Andréa Dessen et al. "Crystal Structure and Function of the Isoniazid Target of Mycobacterium tuberculosis". In: (1995). DOI: 10.1126/SCIENCE.7886450 (cit. on p. 8).

[39] Annaik Quémard et al. "Enzymatic characterization of the target for isoniazid in Mycobacterium tuberculosis". In: (1995). DOI: 10.1021/BI00026A004 (cit. on p. 8).

[40] K. Takayama, L. Wang, and H. L. David. "Effect of Isoniazid on the In Vivo Mycolic Acid Synthesis, Cell Growth, and Viability of Mycobacterium tuberculosis". In: (1972). DOI: 10.1128/AAC.2.1.29 (cit. on p. 8).

[41] Florence Brossier et al. "Performance of the genotype MTBDR line probe assay for detection of resistance to rifampin and isoniazid in strains of Mycobacterium tuberculosis with low- and high-level resistance". In: (2006). DOI: 10.1128/JCM.01054-06 (cit. on p. 8).

[42] Hiroki Ando et al. "Downregulation of katG expression is associated with isoniazid resistance in Mycobacterium tuberculosis". In: (2011). DOI: 10.1111/J.1365-2958.2011.07547.X (cit. on p. 8).

[43] Zhenling Cui et al. "Mutations in the embC-embA Intergenic Region Contribute to Mycobacterium tuberculosis Resistance to Ethambutol". In: (2014). DOI: 10.1128/AAC.03285-14 (cit. on p. 9).

[44] Sarah M. Batt et al. "Antibiotics and resistance: the two-sided coin of the mycobacterial cell wall". In: (2020). DOI: 10.1016/J.TCSW.2020.100044 (cit. on p. 10).

[45] Aimee E. Belanger et al. "The embAB genes of Mycobacterium avium encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol." In: (1996). DOI: 10.1073/PNAS.93.21.11919 (cit. on p. 10).

[46] R. Goude et al. "The arabinosyltransferase EmbC is inhibited by ethambutol in Mycobacterium tuberculosis". In: (2009). DOI: 10.1128/AAC.00162-09 (cit. on p. 10).

[47] Yong Zi Tan et al. "Cryo-EM structure of arabinosyltransferase EmbB from Mycobacterium smegmatis". In: (2020). DOI: 10.1038/s41467-020-17202-8 (cit. on p. 10).

[48] Lu Zhang et al. "Structures of cell wall arabinosyltransferases with the anti-tuberculosis drug ethambutol". In: (2020). DOI: 10.1126/SCIENCE.ABA9102 (cit. on pp. 10, 11, 26, 39).

[49] Yong Zi Tan et al. "Cryo-EM Structures and Regulation of Arabinofuranosyltransferase AftD from Mycobacteria". In: (2020). DOI: 10.1016/J.MOLCEL.2020.04.014 (cit. on pp. 11–13, 39).

[50] L. J. Alderwick et al. "Structure, function and biosynthesis of the Mycobacterium tuberculosis cell wall: arabinogalactan and lipoarabinomannan assembly with a view to discovering new drug targets". In: (2007). DOI: 10.1042/BST0351325 (cit. on p. 12).

[51] Henrieta Škovierová et al. "AftD, a novel essential arabinofuranosyltransferase from mycobacteria". In: (2009). DOI: 10.1093/GLYCOB/CWP116 (cit. on p. 12).

[52] Liisa Holm and Laura M. Laakso. "Dali server update". In: (2016). DOI: 10.1093/NAR/GKW357 (cit. on p. 12).

[53] Chandra B. SHARMA et al. "N-Glycosylation of Yeast Proteins". In: (1981). DOI: 10.1111/J.1432-1033.1981.TB05306.X (cit. on pp. 12, 29).

[54] Sudeep Pushpakom et al. "Drug repurposing: progress, challenges and recommendations". In: (2018). DOI: 10.1038/nrd.2018.168 (cit. on p. 13).

[55] Yicheng Gong et al. "Structure of the priming arabinosyltransferase AftA required for AG biosynthesis of Mycobacterium tuberculosis". In: (2023). DOI: 10.1073/PNAS.2302858120 (cit. on pp. 15, 35).

[56] Guanwen Liu et al. "The CRISPR-Cas toolbox and gene editing technologies". In: (2022). DOI: 10.1016/J.MOLCEL.2021.12.002 (cit. on p. 16).

[57] Kira S. Makarova et al. "Evolution and classification of the CRISPR–Cas systems". In: (2011). DOI: 10.1038/nrmicro2577 (cit. on p. 16).

[58] Ruud Jansen et al. "Identification of genes that are associated with DNA repeats in prokaryotes". In: (2002). DOI: 10.1046/J.1365-2958.2002.02839.X (cit. on p. 16).

[59] Martin Jinek et al. "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity". In: (2012). DOI: 10.1126/SCIENCE.1225829/SUPPL_FILE/JINEK.SM.PDF (cit. on p. 16).

[60] Eira Choudhary et al. "Gene silencing by CRISPR interference in mycobacteria". In: (2015). DOI: 10.1038/ncomms7267 (cit. on pp. 16, 29, 33).

[61] Rodolphe Barrangou et al. "CRISPR provides acquired resistance against viruses in prokaryotes". In: (2007). DOI: 10.1126/SCIENCE.1138140/SUPPL_FILE/BARRANGOU.SOM.PDF (cit. on p. 16).

[62] Jennifer A. Doudna and Emmanuelle Charpentier. "The new frontier of genome engineering with CRISPR-Cas9". In: (2014). DOI: 10.1126/SCIENCE.1258096 (cit. on p. 16).

[63] Robert Heler et al. "Cas9 specifies functional viral targets during CRISPR–Cas adaptation". In: (2015). DOI: 10.1038/nature14245 (cit. on p. 16).

[64] Jeremy M. Rock et al. "Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform". In: (2017). DOI: 10.1038/nmicrobiol.2016.274 (cit. on pp. 16–18, 21–23, 29, 30, 32).

[65] Matthew H. Larson et al. "CRISPR interference (CRISPRi) for sequence-specific control of gene expression". In: (2013). DOI: 10.1038/nprot.2013.132 (cit. on pp. 16, 33).

[66] Jason M. Peters et al. "A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria". In: (2016). DOI: 10.1016/J.CELL.2016.05.003 (cit. on pp. 16, 17).

[67] Lei S. Qi et al. "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression". In: (2013). DOI: 10.1016/J.CELL.2013.02.022 (cit. on pp. 16, 18).

[68] Atul K. Singh et al. "Investigating essential gene function in Mycobacterium tuberculosis using an efficient CRISPR interference system". In: (2016). DOI: 10.1093/NAR/GKW625 (cit. on pp. 17, 23).

[69] Jeremy Rock. "Tuberculosis drug discovery in the CRISPR era". In: (2019). DOI: 10.1371/JOURNAL.PPAT.1007975 (cit. on p. 18).

[70] Jeremy Rock. "CRISPRi gene silencing in mycobacteria". In: (2017) (cit. on p. 21).

[71] Víctor Campo-Pérez et al. "Easily applicable modifications to electroporation conditions improve the transformation efficiency rates for rough morphotypes of fast-growing mycobacteria". In: (2021). DOI: 10.1016/j.nbt.2021.02.003 (cit. on p. 22).

[72] *Addgene: Gibson Assembly Protocol*. URL: https://www.addgene.org/protocols/gibson-assembly/ (visited on 2023-06-07) (cit. on p. 24).

[73] Irith Wiegand et al. "Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances". In: (2008). DOI: 10.1038/NPROT.2007.521 (cit. on p. 26).

[74] Cátia Silveiro et al. "CRISPRi-mediated characterization of novel anti-tuberculosis targets: Mycobacterial peptidoglycan modifications promote beta-lactam resistance and intracellular survival". In: (2023). DOI: 10.3389/FCIMB.2023.1089911/BIBTEX (cit. on pp. 26, 43).

[75] Paulo Bettencourt et al. *Application of confocal microscopy for quantification of intracellular mycobacteria in macrophages*. 2010 (cit. on p. 26).

[76] Jing Liu and Arcady Mushegian. "Three monophyletic superfamilies account for the majority of the known glycosyltransferases". In: (2003). DOI: 10.1110/PS.0302103 (cit. on pp. 29, 30).

[77] Sabine Ehrt et al. "Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor". In: (2005). DOI: 10.1093/NAR/GNI013 (cit. on pp. 29, 32).

[78] Luke J. Alderwick et al. "Identification of a novel arabinofuranosyltransferase (AftA) involved in cell wall arabinan biosynthesis in Mycobacterium tuberculosis". In: (2006). DOI: 10.1074/JBC.M600045200 (cit. on p. 30).

[79] Yusuke Minato et al. "Genomewide Assessment of Mycobacterium tuberculosis Conditionally Essential Metabolic Pathways". In: (2019). DOI: 10.1128/MSYSTEMS.00070-19 (cit. on p. 32).

[80] Yusuke Minato et al. "Genomewide Assessment of Mycobacterium tuberculosis Conditionally Essential Metabolic Pathways". In: (2019). DOI: 10.1128/MSYSTEMS.00070-19 (cit. on p. 32).

[81] Kelley Hurst-Hess et al. "Mycobacterial SigA and SigB Cotranscribe Essential Housekeeping Genes during Exponential Growth". In: (2019). DOI: 10.1128/MBIO.00273-19 (cit. on p. 32).

[82] John P. Bannantine et al. "NlpC/P60 domain-containing proteins of Mycobacterium avium subspecies paratuberculosis that differentially bind and hydrolyze peptidoglycan". In: (2016). DOI: 10.1002/PRO.2884 (cit. on p. 41).

[83] Nada Lelovic et al. "Application of Mycobacterium smegmatis as a surrogate to evaluate drug leads against Mycobacterium tuberculosis". In: (2020). DOI: 10.1038/S41429-020-0320-7 (cit. on p. 43).

# I

# Annex

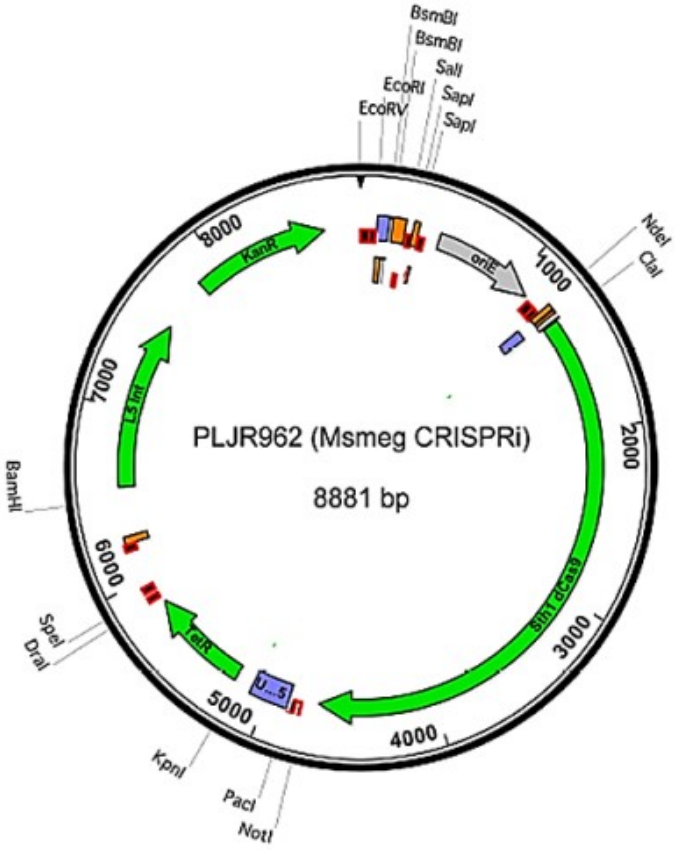


Figure I.1: **The CRISPRi backbone for *M. smegmatis* (PLJR962; 8881bp).** The vector contain: 1) dCas9Sth1 regulated by an ATc-inducible promoter; 2) a sgRNA scaffold regulated by an ATc-inducible promoter; 3) a Tet repressor; 4) a single-copy L5- integrative backbone; 5) a pBR3222 origin of replication; 6) a kanamycin resistance cassette.

| PAM | Fold repression | SD |
|---|---|---|
| 5'-NNAGAAG-3' | 216.7 | 10.0 |
| 5'-NNAGAAT-3' | 216.2 | 10.4 |
| 5'-NNAGAAA-3' | 158.1 | 22.8 |
| 5'-NNGGAAG-3' | 145.2 | 5.3 |
| 5'-NNAGAAC-3' | 120.5 | 7.9 |
| 5'-NNGGAAA-3' | 110.5 | 26.4 |
| 5'-NNAGCAT-3' | 84.6 | 5.2 |
| 5'-NNAGGAG-3' | 82.2 | 9.2 |
| 5'-NNAGGAT-3' | 64.7 | 8.7 |
| 5'-NNAGCAA-3' | 53.4 | 9.9 |
| 5'-NNGGAAC-3' | 51.5 | 6.2 |
| 5'-NNGGAAT-3' | 47.3 | 3.3 |
| 5'-NNAGCAG-3' | 42.2 | 7.0 |
| 5'-NNAGGAA-3' | 38.5 | 5.2 |
| 5'-NNAGGAC-3' | 25.5 | 0.8 |
| 5'-NNGGGAG-3' | 24.7 | 1.9 |
| 5'-NNGGGAT-3' | 24.2 | 3.4 |
| 5'-NNGGGAA-3' | 12.3 | 0.8 |
| 5'-NNAGCAC-3' | 11.9 | 1.2 |
| 5'-NNGGGAC-3' | 7.9 | 1.0 |
| 5'-NNGGCAT-3' | 6.7 | 0.9 |
| 5'-NNGGCAG-3' | 4.0 | 0.3 |
| 5'-NNGGCAA-3' | 3.3 | 0.3 |
| 5'-NNGGCAC-3' | 2.7 | 0.3 |
| ctrl sgRNA | 1.3 | 0.1 |

Figure I.2: **Possible PAM variants for dCas9$_{Sth1}$ mediated gene silencing.** Their correspondent silencing efficiency is indicated as "fold repression" and PAMs are organized in descending order. The subsequent reference of PAMX, where X is any number between 1 and 11, in the materials and methods and in the results and discussion, refers to the order by which the PAMs are organized in this table.

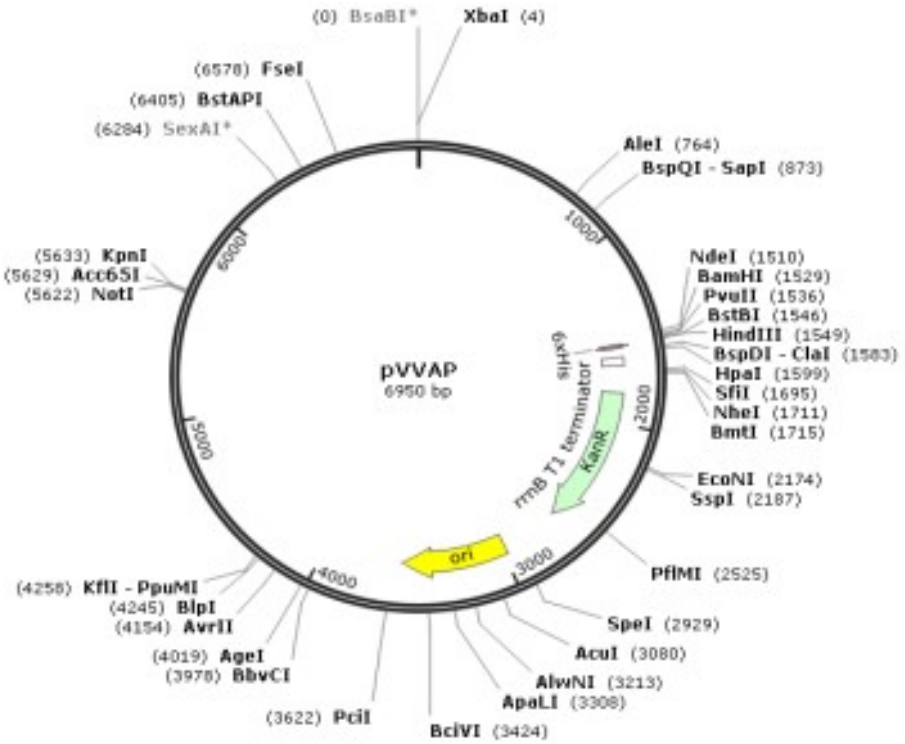


Figure I.3: **Restriction Map of shuttle vector pVVAP (6950bp).** KanR – kanamycin resistance cassette; ori$_{ColE1}$ – ColE1 origin of replication; ori$_{pAL5000}$ – pAL5000 origin of replication; *rrnB* T1 – terminator; His$_6$ - sequence encoding 6 histidines.

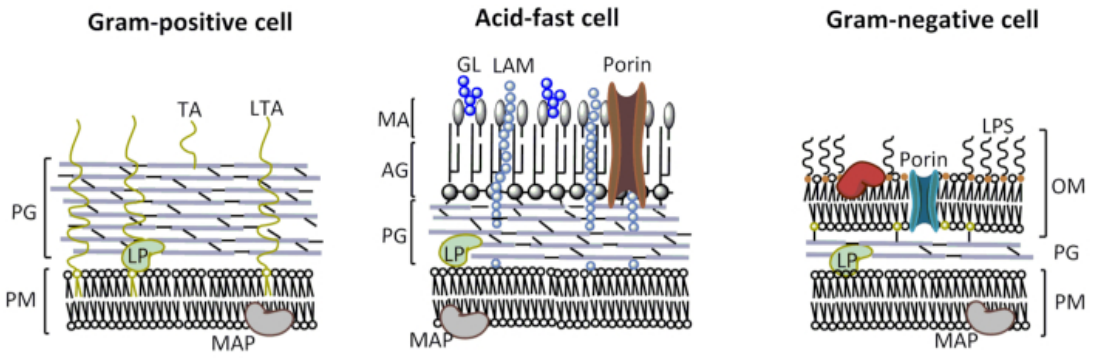


Figure I.4: **The cell envelope structures vary among Gram-positive, acid-fast and Gram-negative bacteria. AG, arabinogalactan; GL, glycolipid; LAM, lipoarabinomannan; LP, lipoprotein; LPS, lipopolysaccharide; LTA, lipoteichoic acid; MA, mycolic acid; MAP, membrane-associated protein; OM, outer membrane; PG, peptidoglycan; PM, plasma membrane; TA, teichoic acid .**

Table I.1: **PCR primers used in the Gibson reaction**

| PCRs | Forward and Reverse Primers (5′-3′) | Size (bp) | % GC | Tm ( ºC) |
|---|---|---|---|---|
| PCR1 | **Fwd- CAATTGCGGATCCAGCTGCAGAATTCG** | 27 | 52 | 64 |
| | **Rev- GGCCATATGTGGACTCCCTTTCTCTTATCGG** | 31 | 52 | 64 |
| PCR3 | **Fwd- CCGATAAGAGAAAGGGAGTCCACATATGGCC** | 31 | 39 | 60 |
| | **Rev- CGAATTCTGCAGCTGGATCCGCAATTG** | 27 | 52 | 61 |