TRANSCRIPTION START SITE MAPPING AND SMALL REGULATORY RNA PROFILING OF *MYCOBACTERIUM TUBERCULOSIS*

CHEAH HONG LEONG

UNIVERSITI SAINS MALAYSIA

2022

TRANSCRIPTION START SITE MAPPING AND SMALL REGULATORY RNA PROFILING OF *MYCOBACTERIUM TUBERCULOSIS*

by

CHEAH HONG LEONG

Thesis submitted in fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2022

ACKNOWLEDGEMENT

I would like to express my sincere and deepest gratitude to my supervisor, Prof. Tang Thean Hock for his motivation, patience, and support throughout my study. His wide academic interests and enthusiasm has motivated me to take on challenges and it was thrilling to learn new knowledge and lessons from situations. Throughout my study, I also received immense supports from Dr. Carsten Raabe and Dr. Timofey Rozhdestvensky, and wish to take this opportunity to show my appreciation for their guidance and insightful comments. There are several members of RNA-Bio Research Group that deserve appreciation: Dr. Citartan Marimuthu, Dr. Ng Siew Kit, Dr. Ch'ng Ewe Seng, and Dr. Hazrina Yusof Hamdani for their helpful advises and guidance; Dr. Siti Aminah Ahmed, for assisting me in the administrative works; as well as other laboratory fellows, including Mr. Yeoh Tzi Shien, Dr. Alvin Paul Dickson, Ms. Anna Andrew, Dr. Lee Li Pin, Dr. Ang Kai Cheen, Dr. Prabu Siva Sankar, Dr. Presela Ravinderan, etc. My gratitude also goes to all laboratory and administrative staffs of the Advanced Medical and Dental Institute (AMDI) for providing assistance throughout my study. Last but not least, my most sincere appreciation dedicated to two very precious persons in my life - my father, Mr. Cheah Chin Chai and my mother, Ms. Ang Chew Hong - for their unconditional supports, which gives me courage to take initiative in facing challenges.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS

%	percentage
^{o}C	degree Celcius
s	second
U	unit
µg/µL	microgram per microliter
µg/mL	microgram per milliliter
µg	microgram
µL	microliter
M	molar
mM	millimolar
µM	micromolar
g	gram
x *g*	relative centrifugal force
L	liter
mL	milliliter
mg/mL	milligram per milliliter
OD_{600}	optical density at 600 nm wavelength
pH	potential of Hydrogen
rpm	rotation per minute
V	volt

LIST OF ABBREVIATIONS

AG	Arabinogalactan
APS	Ammonium persulfate
Araf	D-arabinofuranosyl
asTSS	Antisense TSS
ATP	Adenosine triphosphate
BCG	Bacille Calmette-Guérin
bp	Base pair
cDNA	complementary DNA
CLI	Command-line interface
CPM	Count per million mapped reads
CTP	Cytidine triphosphate
DAT	Diacyltrehalose
DEG	Differentially expressed gene
DIP	Phthiocerol diphthioceranate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxynucleoside triphosphate
dRNA-seq	Differential RNA sequencing
e.g.	exempli gratia (for example)
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium bromide
EXO	Exonuclease
Galf	D-galactofuranosyl
GlcNAc	N-acetylglucosamine
GO	Gene ontology
GUI	Graphical user interface
HTS	High-throughput sequencing
i.e.	id est (that is)
iTSS	Internal TSS
LAM	Lipoarabinomannan
LM	Lipomannan

MA	Mycolic acid
mAGP	Mycoloylarabinogalactan-peptidoglycan
MDR-TB	Multidrug-resistant TB
miRNA	MicroRNA
mRNA	Messenger RNA
MurNAc	N-acetylmuramic acid
MurNGlyc	N-glycolylmuramic acid
nt	Nucleotide
OD	Optical density
ORF	Open reading frame
oTSS	Orphan TSS
PAGE	Polyacrylamide gel electrophoresis
PAT	Polyacyltrehalose
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDIM/DIM	Phthiocerol dimycocerosate
PE	Paired-end
PG	Peptidoglycan
PGL	Phenolic glycolipid
PI	Phosphatidylinositol
PIM	Phosphatidylinositol mannoside
pTSS	Primary TSS
RACE	Rapid amplification of cDNA ends
RBS	Ribosome-binding site
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RNase	Ribonuclease
RPKM	Reads per kilobase of transcript per million mapped reads
rRNA	Ribosomal RNA
SD	Shine-Dalgarno
SDS	Sodium dodecyl sulfate
SL	Sulpholipid
sRNA	Small regulatory RNA/small RNA
sRNA-seq	Small RNA sequencing

sTSS	Secondary TSS
TAE	Tris-acetate-EDTA
TAP	Tobacco acid pyrophosphatase
TB	Tuberculosis
TBE	Tris-borate-EDTA
TEMED	Tetramethylethylenediamine
TEX	Terminator™ 5'-phosphate-dependent exonuclease
TI	Transcriptional Interference
tRNA	Transfer RNA
TSS	Transcription start site
TTS	Transcription termination sites
UTR	Untranslated region
v/v	Volume per volume
w/v	Weight per volume
XDR-TB	Extremely drug-resistant TB

LIST OF APPENDICES

PEMETAAN TAPAK PERMULAAN TRANSKRIPSI DAN PEMPROFILAN RNA PENGAWALATUR KECIL DALAM *MYCOBACTERIUM TUBERCULOSIS*

ABSTRAK

Mycobacterium tuberculosis merupakan aktinobakteria yang menyebabkan jangkitan kronik. Patogenesisnya bergantung kepada dinding sel yang tebal dan kaya dengan lipid yang memudahkan kelangsungan hidupnya. Pengetahuan tentang rangkaian pengawalseliaan dalam *M. tuberculosis* penting untuk memahami patogenesisnya bagi mereka-bentuk strategi terapeutik. Selain gen protein, RNA pengawalseliaan kecil (sRNA) merupakan komponen modulasi utama dalam semua organisma. Kajian sejak sedekad yang lepas telah menemui pelbagai sRNA bakteria yang menunjukkan bahawa transkriptom bakteria jauh lebih rumit daripada jangkaan sebelum ini. Kajian ini telah menubuhkan profil sRNA *M. tuberculosis* yang terdedah kepada pelbagai tekanan melalui analisis data RNA-seq untuk menghuraikan pelbagai aspek sRNAome dalam *M. tuberculosis*. Kajian ini teleh mengenal pasti sejumlah 1,376 calon sRNA dan ujian 5'/3'RACE seterusnya telah berjaya mengesahkan lima sRNA, antaranya, predictedRNA_0578 mungkin mengekod peptida kecil yang baharu, manakala predictedRNA_0020 berpotensi mempengaruhi pembentukan dinding sel. Selain itu, kajian ini juga telah membina tapak TSS yang membekalkan ciri-ciri transkrip yang lain, seperti penganjur dan kawasan yang tidak diterjemahkan (UTR). Kesimpulannya, kajian ini telah mendedahkan kerumitan sRNAome mikobakteria dan berpotensi menyumbang kepada pembangunan antibiotik bersifat antisense. Namun demikian, mekanisme dan fungsi mereka masih memerlukan pencirian dan pengesahan selanjutnya.

TRANSCRIPTION START SITE MAPPING AND SMALL REGULATORY RNA PROFILING OF *MYCOBACTERIUM TUBERCULOSIS*

ABSTRACT

Mycobacterium tuberculosis is an actinobacterium that causes chronic infections. Its pathogenesis lies in the thick, lipid-rich cell envelope that facilitate persistence in hostile environments. A refined knowledge of regulatory networks within *M. tuberculosis* is important to understand its pathogenesis for therapeutic design. Small regulatory RNAs (sRNAs) are key modulatory components within all organisms. Extensive studies over the past decade have uncovered a plethora of bacterial sRNAs, implying that bacterial transcriptomes are far more sophisticated than previously anticipated. Here, sRNA profiles of *M. tuberculosis* exposed to different stresses were established by analyzing RNA-seq data to address different facets of sRNAome in *M. tuberculosis*. This study has identified a total of 1,376 sRNA candidates and the subsequent 5'/3'RACE assay has successfully validated a total of five sRNAs, among which predictedRNA_0578 may encode for a novel small peptide and predictedRNA_0020 potentially affects the formation of mycobacterial cell wall. Besides, this study also included transcription start site (TSS) landscapes that provide additional transcriptional features, e.g., promoters and untranslated regions (UTRs). Taken together, this study reveals the complexities of the mycobacterial sRNAome and may facilitate the development of antisense antibiotics. However, their mechanisms and functions still require further characterization and validation.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Tuberculosis (TB) is the most widespread bacterial infectious disease and remains prevalent particularly among those living in poverty in South-East Asia and Western Pacific regions (WHO, 2021). Despite recent medical advances, there are still more than 10 million TB patients worldwide with over 1.5 million deaths annually (WHO, 2021). TB is primarily attributed to *Mycobacterium tuberculosis*, which is a slow-growing bacillus belonging to the family Mycobacteriaceae (Cole *et al.*, 1998). It has a thick, hydrophobic cell envelop that controls nutrient and drug entries, representing the major contributor of antibiotic resistance and virulence (Cole *et al.*, 1998; Kieser & Rubin, 2014). Additionally, the cell envelop also contains various surface proteins (e.g., adhesins) that facilitate cell-cell contact during infections (Brennan & Nikaido, 1995; Kieser & Rubin, 2014). As a result, *M. tuberculosis* is capable of striving in various hostile environments in hosts.

Bacterial transcriptomes were once believed to be less complex with a much simpler regulatory fashion as compared to those in eukaryotes (Sorek & Cossart, 2010). The emergence of high-throughput sequencing (HTS) technology has allowed the discovery of numerous RNA regulatory elements (e.g., non-coding RNAs) that shed light on the complexity of transcriptome networks in prokaryotes. For instance, the widespread transcription of *cis*-antisense small RNAs (sRNAs) has previously been considered a hallmark of eukaryotic transcriptomes, but it has recently been revealed that prokaryotes also widely express RNA transcripts with similar features and mechanisms, implying that prokaryotic transcriptomes are more complex and

sophisticated than previously thought, and may act in similar ways to their eukaryotic counterparts (Yoder-Himes *et al.*, 2009; Thomason *et al.*, 2015; Sharma *et al.*, 2010a; Toledo-Arana *et al.*, 2009; Sittka *et al.*, 2008; Perkins *et al.*, 2009; Raabe *et al.*, 2011; Cheah *et al.*, 2018).

Most of our current understanding of bacterial regulatory sRNAs were primarily contributed by studies on model bacteria, e.g., *Escherichia coli*, *Salmonella typhimurium*, *Listeria monocytogenes*, *Bacillus subtilis*, and *Staphylococcus aureus*. Thousands of potential sRNAs have been identified in *M. tuberculosis* following the advent of HTS technologies over the past decade (Arnvig *et al.*, 2011; Miotto *et al.*, 2012; Namouchi *et al.*, 2016; Wang *et al.*, 2016), but current knowledge on non-coding elements (e.g., sRNAs) and their functions in *M. tuberculosis* remains elusive with only 21 annotated sRNAs and less than a dozen functionally characterized sRNAs (see Section 2.4.4). The current state of the poor annotations of the sRNAs is a hindrance to understanding of the regulatory roles of the sRNA associated with the pathogenesis of *M. tuberculosis*. To make matters worse, the genome of *M. tuberculosis* differs radically from that of the above-mentioned model bacteria, hampering the possibility of inferring sRNAs and their functions from homology. A much deeper insights into the identification of sRNAs can improve the understanding of the pathophysiology of *M. tuberculosis*. Therefore, further high-throughput and in-depth studies are required to address this issue and fill this knowledge gap.

1.2 Objectives of Study

This study aimed to establish sRNA profiles of *M. tuberculosis* in response to various stress conditions, so as to reveal different facets of sRNAome and sRNA-

mediated survival mechanisms in *M. tuberculosis*. This objective was achieved via the following sub-objectives:

- To identify novel sRNA candidates from RNA-seq data of *M. tuberculosis*;
- To identify and extract other transcriptome features, e.g., transcription start sites (TSS), promoters, and untranslated regions (UTR), to facilitate the subsequent characterization of sRNA candidates;
- To experimentally validate and characterize selected sRNA candidates via the 5'/3'RACE assay.

CHAPTER 2

LITERATURE REVIEW

2.1 History of Tuberculosis

The genus *Mycobacterium* is postulated to have existed for more than 150 million years (Barberis *et al.*, 2017). It has been demonstrated that the progenitor of *M. tuberculosis* might have infected early hominids in East Africa three million years ago (Gutierrez *et al.*, 2005), while the common ancestor of modern *M. tuberculosis* strains might have emerged between 15,000 and 20,000 years ago during the last glacial period (Kapur, Whittam, & Musser, 1994; Brosch *et al.*, 2002; Barberis *et al.*, 2017). Archaeological excavations have found TB in bone samples of a woman and her infant, who were buried together about 9,000 years ago in Atlit Yam which is now under the Mediterranean Sea (Hershkovitz *et al.*, 2008).

"Tuberculosis" is a modern medical term coined by Johann Lukas Schönlein in 1839 (Schönlein, 1839; Riccardi *et al.*, 2020), ever since then the disease has been described and given various names (e.g., yaksma, balasa, feilao, phthisis, white plague, scrofula, King's Evil, lupus vulgaris, and Consumption) in early civilizations (Daniel, 2006; Donoghue, 2011). The earliest written records of TB date back to 3,300 years ago in India and 2,300 years ago in China (Barberis *et al.*, 2017). Excellent descriptions of TB (called "phthisis") were made in Ancient Greece, where Ancient Greek physicians or scientists, e.g., Isocrates (436-338 BC), Aristotle (384-322 BC), and Galen (129-216 AD), have recognized its contagious nature (Riccardi *et al.*, 2020). Following the fall of the Roman Empire, there were no remarkable discoveries made regarding TB during the Middle Ages (476-1453 AD) and Renaissance (1,300-1,600 AD). One of the indicators for such backwardness is the healing ceremony called "royal touch" or "the king's touch", which refers to the

touch of the English and French monarchs who were believed to have curative powers due to the divine right of sovereigns (Murray *et al.*, 2016). The practice continued until 1712-1714 in England and 1825 in France (Barberis *et al.*, 2017). However, there were still some advancements made in European countries beyond England and France during the Late Renaissance. In 1546, Girolamo Fracastoro proposed that phthisis is a transmittable disease attributable to an invisible virus, which can survive for years on the clothes and belongings of patients and can be transmitted via direct contact or body fluids of patients (Sabbatani, 2004).

The Age of Enlightenment and the First Industrial Revolution from the 17th to 19th centuries, during which the incidence of TB peaked, witnessed greater breakthroughs in understanding TB. In 1720, the etiologic agent of TB was proposed by the English physician Benjamin Marten to be invisible microscopic living things that are similar to animalcule described by Anton van Leeuwenhoek in 1695 (Doetsch, 1978). In 1761, the percussion method for diagnosing various diseases, including TB, has been developed by the Austrian physician Josef Leopold Auenbrugger, who is considered one of the founders of modern medicine (Smith, 1962). A few years after Johann Lukas Schönlein created the term "tuberculosis", the German physician Philipp Friedrich Hermann Klencke has successfully inoculated rabbits with material from a miliary tubercle in 1843 (Agarwal *et al.*, 2017). However, Philipp Friedrich Hermann Klencke believed that TB is a cancer and it was not until 1865 when Jean Antoine Villemin experimentally showed for the first time that the disease could be transmitted to rabbits from humans, confirming its infectious nature (Daniel, 2015). In 1882, the German doctor Robert Koch reported the discovery, isolation, and cultivation of the microorganism responsible for pulmonary TB, i.e., "*Mycobacterium tuberculosis*" (see Figure 2.1) (Cambau &

Drancourt, 2014). The discovery made by Robert Koch marked an important milestone in TB eradication as it demonstrated that TB is attributable to a bacterium *M. tuberculosis* that can be acquired through transmission from infected patients or animals (Cambau & Drancourt, 2014).

BERLINER

KLINISCHE WOCHENSCHRIFT.

Organ für practische Aerzte.

Mit Berücksichtigung der preussischen Medicinalverwaltung und Medicinalgesetzgebung nach amtlichen Mittheilungen.

Redacteur: Professor Dr. C. A. Ewald. Verlag von August Hirschwald in Berlin.

Montag, den 10. April 1882. № 15. Neunzehnter Jahrgang.

I. Die Aetiologie der Tuberculose.

(Nach einem in der physiologischen Gesellschaft zu Berlin am 24. März cr. gehaltenen Vortrage.)

Von

Dr. Robert Koch,

Regierungsrath im Kaiserl. Gesundheitsamt.

Figure 2.1 The first-ever report of *M. tuberculosis* by Robert Koch. Adapted from "Steps towards the discovery of *Mycobacterium tuberculosis* by Robert Koch, 1882," by Cambau & Drancourt, 2014, Clinical Microbiology and Infection, 3(20), p. 996-201.

Robert Koch's discovery was followed by various major diagnostic and therapeutic breakthroughs (e.g., Mantoux tuberculin skin test, Bacillus Calmette–Guérin vaccine, streptomycin, and isoniazid), as well as declines in TB incidence due to sanatorium movement in the next decades. The first antituberculous drugs, i.e., streptomycin and para-Aminosalicylic acid, were administered to patients with TB in 1944 (Schatz *et al.*, 1944). In 1952, isoniazid, also known as isonicotinic acid hydrazide, opened the modern era of TB treatment as it was non-toxic and markedly

more potent than any previously tested antibiotic (Murray *et al.*, 2015). In addition, isoniazid is available as a relatively inexpensive generic medication (Murray *et al.*, 2015). The subsequent incorporation of rifampicin (also known as rifampin) and pyrazinamide into the treatment regimen for TB further reduced the treatment duration (Sensi, 1983; Whitfield *et al.*, 2015).

Declines in the global incidence and mortality of TB in the 20th century provided a glimmer of hope that the disease could be completely eliminated, especially in industrialized countries. However, the resurgence of TB due to the AIDS pandemic and antibiotic resistance in the late 1980s has proved this optimism to be unfounded (Keshavjee & Farmer, 2012; Borgdorff & van Soolingen, 2013). The World Health Organization (WHO) estimated that one-third of the global population in 1993 has latent TB infections, leading to 7-8 million active TB cases and 1.3-1.6 million deaths annually (WHO, 1994). The completion of the genome sequence of *M. tuberculosis* H37Rv in 1998 marked a significant milestone in TB research as it has provided abundant information about *M. tuberculosis* that will facilitate the development of new diagnostic and therapeutic techniques (Cole *et al.*, 1998).

2.2 Current Epidemiology

The 20th century has witnessed declines in TB incidence until the late 1980s when it re-surged due to the AIDS pandemic and emergence of multidrug-resistant TB (MDR-TB) strains (Keshavjee & Farmer, 2012; Borgdorff & van Soolingen, 2013). MDR-TB is defined as TB infections with *M. tuberculosis* strains that are at least resistant to the two major first-line antituberculous drugs, i.e., isoniazid and rifampicin (WHO, 2021). Antibiotic-resistant *M. tuberculosis* strains have emerged

following the widespread use of antituberculous drugs due to inappropriate drug prescription and administration, poor-quality generic drugs, low patient compliance, premature termination of treatment regimens, as well as intrinsic drug resistance and mutations of *M. tuberculosis* (Iacobino *et al.*, 2020). The development of resistance to streptomycin (one of the first reported effective antituberculous drugs) was reported shortly after its first administration for the treatment of TB (Schatz *et al.*, 1944). To date, antibiotic resistance in TB continues to be a major public health concern with about half a million MDR-TB cases worldwide (Iacobino *et al.*, 2020).

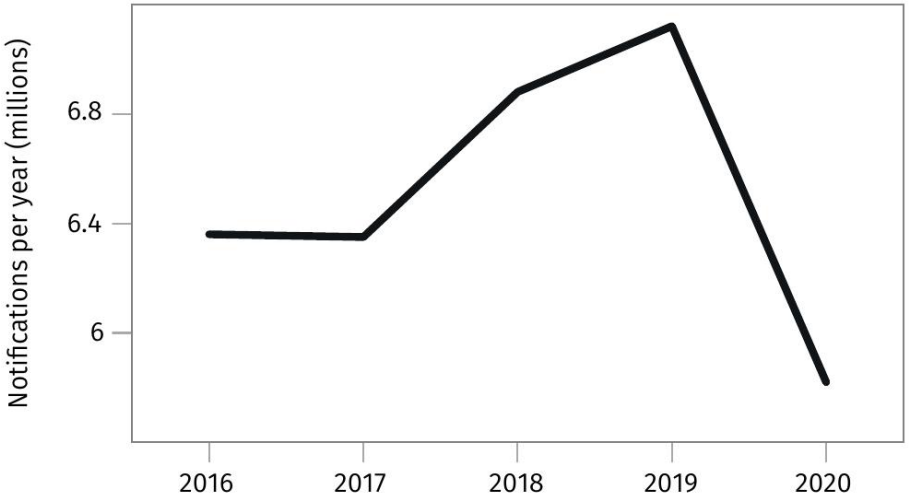

Figure 2.2 Global notified TB incidence (2016–2020). Adapted from "Global Tuberculosis Report 2021," by WHO.

Despite recent medical advances, there were more than 7 million notified cases of TB globally with an estimated actual incidence of 10.0 million cases, as well as over 1.2 million deaths from TB among HIV-negative patients in 2018 (WHO, 2019). The number of newly diagnosed cases of TB declined significantly from 7.1 million in 2019 to 5.8 million in 2020 following the COVID-19 pandemic (see Figure 2.2) (WHO, 2021). However, there was an increase in the number of TB deaths from 1.2 million in 2019 to 1.3 million among HIV-negative patients (WHO,

2021), suggesting that medical resources for the detection and treatment of TB patients are greatly affected by the COVID-19 pandemic.

About one-quarter of the global population (about 2 billion individuals) has latent (noncontagious) TB with 5-10% lifetime risk of developing active (contagious) TB (WHO, 2021). Individuals with active TB may develop mild chronic symptoms (cough, fever, weight loss, etc.) that lead to delays in seeking medical attention and transmissions of the pathogen to others. About 5-15 individuals may contract the disease from a single index patient with active TB in a year and there is one new TB case per second globally (WHO, 2019). In 2018, the South-East Asian region recorded the greatest number of new TB incidences, followed by the African region and Western Pacific region (WHO, 2019). The ranking did not change in 2020, during which the South-East Asian region accounted for 43% of global notified TB cases (WHO, 2021).

2.3 *Mycobacterium tuberculosis* - The Etiologic Agent

In humans, TB is caused by a non-motile, slow-growing bacillus, i.e., *M. tuberculosis*, which belongs to the phylum Actinobacteria, Class Actinobacteria, Order Actinomycetales, and family Mycobacteriaceae (Gordon & Parish, 2018). The tubercle bacilli are transmitted via respiratory droplets, and upon entering the lower respiratory tract and lungs, the tubercle bacilli will encounter the alveolar macrophages, which are non-specific phagocytes that engulf them into phagosomes (Chai *et al.*, 2018). Mature phagosomes will fuse with lysosomes into phagolysosomes that contain acid and nutrient-deprived environment, as well as antimicrobial peptides and lysosomal enzymes for degradation, but the tubercle

bacilli are often capable of blocking phagosomal maturation and acidification (Chai *et al.*, 2018).

The outcome of infection relies on the interplay between the host's immune system and *M. tuberculosis*. In most infected, immunocompetent individuals, the tubercle bacilli will be tightly sequestered within a granuloma formed by the aggregation of immune cells, leading to latent infections without obvious symptoms during which the bacilli switch to dormant, non-replicating state and can persist for decades until the host develops immunosuppression (Chai *et al.*, 2018; Gordon & Parish, 2018). In addition to the ability to enter dormancy, *M. tuberculosis* has several notable features that distinguish it from most other pathogenic bacteria, including metabolic versatility, thick hydrophobic cell wall, and intrinsic resistance to antibiotics. Collectively, these interlinked features have contributed to the resilience of *M. tuberculosis* in hosts and its success as a human pathogen with a long history that can be traced back to early human civilizations.

2.3.1 Dormancy, Persistence, and Metabolic Versatility

The metabolism in *M. tuberculosis*, like any other organism, is a complicated entity comprising vast networks of enzymes and metabolites, but the advent of omics approaches in recent years has provided a more comprehensive overview of the metabolic networks in *M. tuberculosis*, which previously remained understudied (Chang & Guan, 2021). Apart from its sophisticated immune evasion strategies (e.g., expression of surface lipids that mask the pathogen-associated molecular patterns, PAMPs) (Peddireddy *et al.*, 2017), the metabolic flexibility of *M. tuberculosis* has also contributed significantly to its success as a pathogen and persistence during adaptation to various hostile environments. For instance, tubercle bacilli that enter the dormant or persistent state by remodeling its metabolic networks display intrinsic

phenotypic resistance to antituberculous drugs that target actively growing bacteria (Stokes *et al.*, 2019). Such dormant or persistent subpopulations are hardly eradicable and may later undergo genetic mutations that confer acquired drug resistance, posing a major issue in antituberculous therapies (Stokes *et al.*, 2019).

The metabolic versatility of *M. tuberculosis* in hosts lies in its preference for simultaneous coutilization of multiple carbon sources, including fatty acids, cholesterol, lactate, pyruvate, etc., which depend on the site of replication and the type of host cells (Chang & Guan, 2021). Upon exposure to stresses (e.g., nutrient-deprived and hypoxic conditions), *M. tuberculosis* is capable of rewiring its metabolism, including switching its metabolism to triglyceride synthesis and cell wall thickening, to maintain the level of adenosine triphosphate (ATP) for long term survival despite entering dormancy (Baek *et al.*, 2011). Triglycerides serve not only as the energy source during starvation and resuscitation from dormancy, but also as an electron sink that maintains metabolic homeostasis based on environmental conditions (Chang & Guan, 2021).

Mycolic acid (MA), which is a cell wall lipid unique to mycobacteria and related species, also plays a crucial role in dormancy and resuscitation from dormancy, as it serves as a precursor for trehalose dimycolate (TDM) and trehalose monomycolate (TMM), which modulate the cell wall permeability (Chiaradia *et al.*, 2017; Sharma *et al.*, 2019). It has been previously reported that most MA-associated genes are downregulated to reduce MA biosynthesis during exposure to dormancy-inducing stresses, e.g., hypoxic condition (McGillivray *et al.*, 2015; Peterson *et al.*, 2019; Raghunandanan *et al.*, 2019). Besides, the expressions of genes associated with other mycobacterial lipids, e.g., sulfolipids (SL), polyacyltrehaloses (PAT), diacyltrehalose (DAT), and phthiocerol dimycocerosates (PDIMs), have also been

shown to be downregulated upon entering dormancy and upregulated following resuscitation from dormancy (Raghunandanan *et al.*, 2019). The tubercle bacilli may need to reduce the biosynthesis of lipids and MA upon entering dormancy, as they are energy-demanding processes (Peterson *et al.*, 2019). Taken together, metabolic remodeling plays significant roles in the dormancy and persistence of *M. tuberculosis*.

2.3.2 Mycobacterial Cell Wall

One of the most notable features of *M. tuberculosis* is its thick, hydrophobic cell wall that renders it neither Gram-positive nor Gram-negative. Being the protective layer encapsulating the tubercle bacilli, the cell wall provides the first line of defense against external stresses and the impermeability to antibiotics, as well as contains various immunomodulatory molecules, e.g., lipoarabinomannan (LAM), sulpholipids (SLs), and PDIMs, thereby playing a crucial role in its survival and adaptation to hostile environments, and contributing to many biological properties unique to mycobacteria, e.g., intrinsic resistance to antibiotics and impermeability to nutrients (Chiaradia *et al.*, 2017).

Besides, *M. tuberculosis* can alter its cell wall integrity and permeability in response to stresses by modifying the thickness and compositions of the cell wall (Sharma *et al.*, 2019). Given the importance of cell wall to *M. tuberculosis*, the biosynthesis of cell wall has become the target of numerous antituberculous drugs, e.g., isoniazid, ethambutol, and ethionamide (Gordon & Parish, 2018). As reviewed by Batt *et al.* (2020), the mycobacterial cell wall mainly comprises an outer capsule, a unique outer membrane (containing MAs, glycolipids, and PDIMs), the mycoloylarabinogalactan-peptidoglycan complex (mAGP), and a periplasmic space (see Figure 2.3).

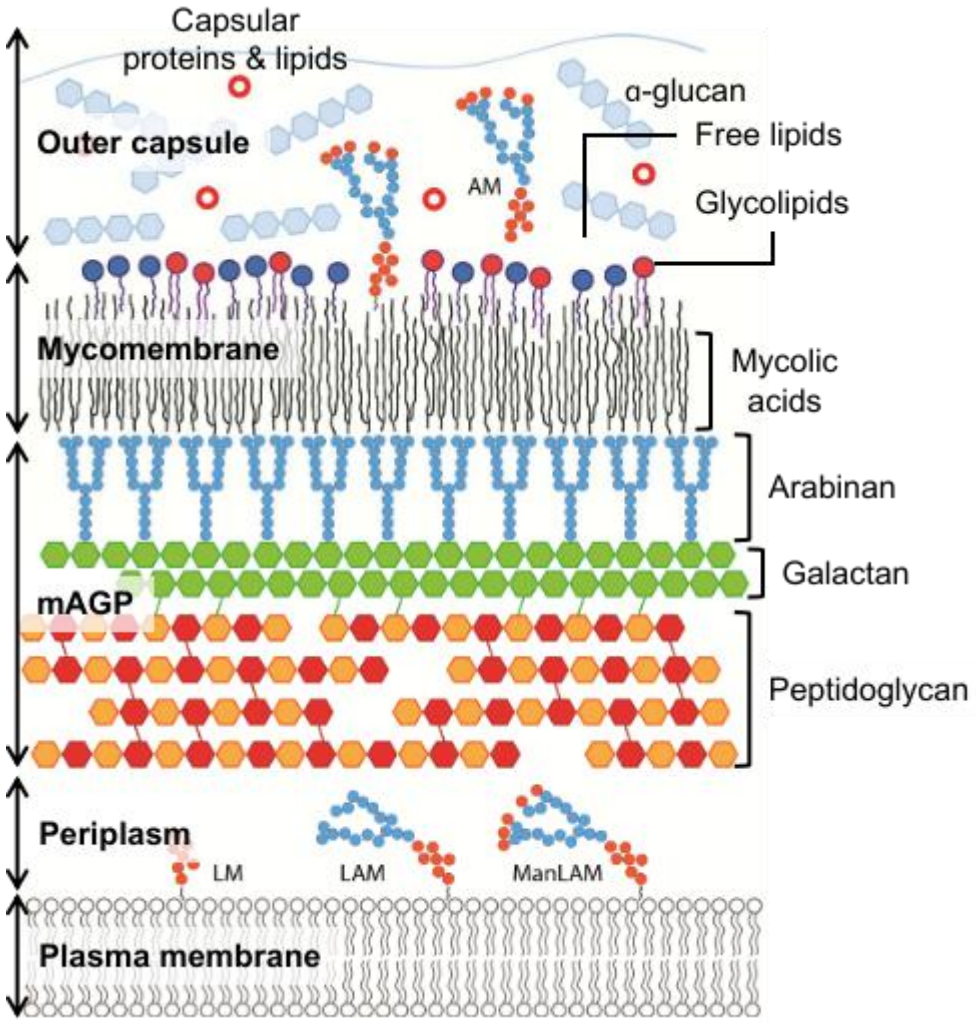

Figure 2.3 Structure of mycobacterial cell envelope. Adapted from "Interplay of human macrophages and *Mycobacterium tuberculosis* phenotypes," by Raffetseder, 2016, [Ph.D., Linköping University]. https://doi.org/10.3384/diss.diva-132321

2.3.2(a) Outer Capsule

Once controversial, the mycobacterial outer capsule is defined as the outer-most layer weakly attached to the outer membrane and can only be observed in cultures under certain conditions (Sani *et al.*, 2010). The outer capsule consists primarily of exopolysaccharides, i.e., α-D-glucan, D-arabino-D-mannan, D-mannan, as well as small amounts of capsular proteins and lipids (Ortalo-Magné *et al.*, 1995; Sambou *et al.*, 2008). α-D-glucan can be found in other bacteria and eukaryotes as cytoplasmic particles that serve as carbon reserves, but mycobacterial α-D-glucan,

apart from intracellular localization, also constitutes the outer capsule as the main exopolysaccharide (Sambou *et al.*, 2008). Other exopolysaccharides, i.e., AM and D-mannan, are structurally identical to and share biosynthetic genes with the major cell wall lipopolysaccharides, i.e., LAM and lipomannan (LM), respectively (Jankute *et al.*, 2015; Kalscheuer *et al.*, 2019).

As the outermost component of the surface of *M. tuberculosis*, the mycobacterial outer capsule plays a vital role in the initial interactions with the host immune system to favor its own survival. Previous studies have demonstrated that mycobacterial capsular polysaccharides prevent phagocytosis to facilitate the nonopsonic CR3-mediated uptake, which favors its intracellular survival by inhibiting IL-12 production and oxidative burst in macrophages and dendritic cells (Stokes *et al.*, 2004; Gagliardi *et al.*, 2007; Geurtsen *et al.*, 2009). Besides, it has been previously reported that vaccination of mice with encapsulated bacillus Calmette-Guerin (BCG) induced a greater protection efficacy than that with unencapsulated BCG (Prados-Rosales *et al.*, 2016).

2.3.2(b) Mycobacterial Outer Membrane

The mycobacterial outer membrane, also called mycomembrane, can be further divided into the inner leaflet and the outer leaflet (see Figure 2.3). As reviewed by Batt *et al.* (2020), the inner leaflet of mycomembrane is predominantly composed of MAs, which are abundant cell wall lipids that account for approximately 30% of the dry weight of cells. There are three classes of MAs: α-MAs, methoxy-MAs, and keto-MAs (see Figure 2.4), among which α-MAs are the most abundant. Besides, these MAs exist in two forms in the mycomembrane, i.e., esterified MAs and free MAs. The former is esterified either to trehalose, forming TDM and TMM, or to the arabinogalactan (AG) of the mAGP, while the latter exists

in smaller quantities and may be involved in the formation of mycobacterial biofilms (Bacon *et al.*, 2014; Batt *et al.*, 2020). Together, MAs are tightly packed into the hydrophobic and rigid yet flexible inner leaflet of mycomembrane, which contributes to impermeability to antibiotics, survival, and pathogenesis of *M. tuberculosis* (Groenewald *et al.*, 2014).

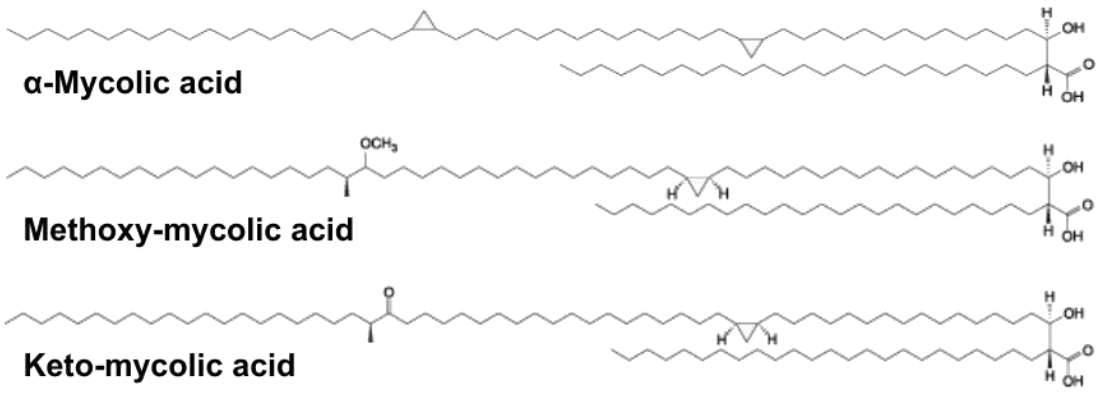

Figure 2.4 Major conformations of mycolic acids

The outer leaflet of mycomembrane, on the other hand, mainly comprises various free lipids, e.g., PDIMs and phenolic glycolipids (PGLs). Both PDIMs and PGLs share a similar lipid core, whereby the former is formed from a phthiocerol esterified with two mycocerosic acids, while the latter has an extra phenolic residue (phenolphthiocerol) and oligosaccharide chain (see Figure 2.5) (Daffé & Laneelle, 1988). Both the outer leaflet components exhibit important pathogenic roles by conferring impermeability and tolerance to antibiotics and facilitating immune evasion (Mohandas *et al.*, 2016; Quigley *et al.*, 2017). It is also noteworthy that despite favoring survival and virulence in hosts, the hydrophobic PDIMs restrict an efficient nutrient uptake and put the tubercle bacilli at a disadvantage when grown *in vitro* (Domenech & Reed, 2009).

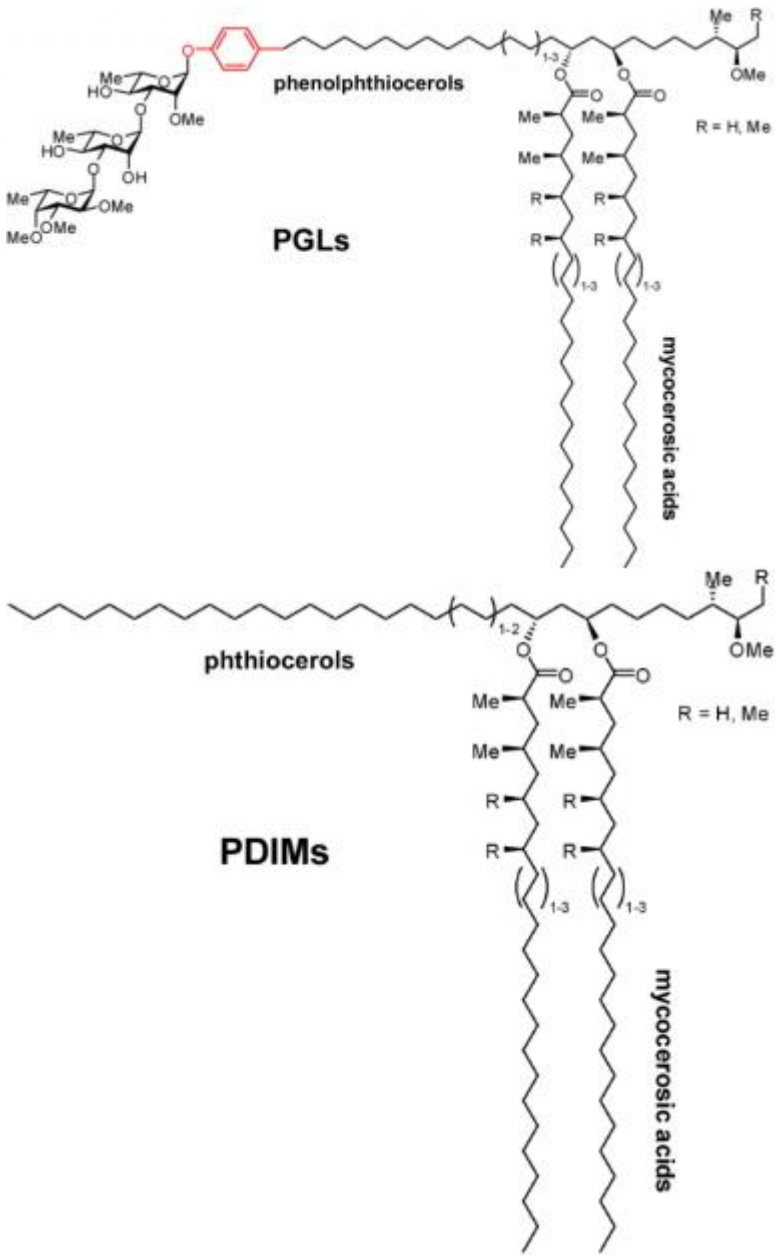

Figure 2.5 Structural differences between PDIM and PGL

2.3.2(c) Mycoloylarabinogalactan-Peptidoglycan Complex

The mAGP complex of *M. tuberculosis*, which constitutes the cell wall core structure, is essential for maintaining a rigid basal structure that supports the mycomembrane (Lederer *et al.*, 1975; Alderwick *et al.*, 2015). The mAGP complex comprises two main layers: peptidoglycan (PG) and AG.

PG is an important cell wall component with a complex, mesh-like structure in almost all bacteria, including *M. tuberculosis*, conferring rigidity to maintain the integrity and shape of bacterial cells (Brennan & Nikaido, 1995; Alderwick *et al.*, 2015; Maitra *et al.*, 2019). In general, bacterial PG comprises glycan strands of alternating N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) residues cross-linked by short peptides (Heijenoort, 2001; Maitra *et al.*, 2019). However, the mycobacterial PG has some modifications that make it structurally and functionally differs from that of other bacteria such as *E. coli*.

In addition to MurNAc, the PG of *M. tuberculosis* also contains N-glycolylmuramic acid (MurNGlyc), which improves its resistance to lysozyme (Raymond *et al.*, 2005). The mycobacterial PG has a peptide cross-linking proportion of 70-80%, which is almost twice that of *E. coli* (Wietzerbin *et al.*, 1974; Glauner *et al.*, 1988). Besides, the muramic acid residues of mycobacterial PG are covalently attached via a phosphodiester bond to the galactan domain of AG, which is also a cell wall feature absent in most other bacteria (McNeil *et al.*, 1990). Apart from structural and protective roles, mycobacterial PG, or more precisely PG fragments released from the cell wall, might serve as messengers for resuscitating dormant tubercle bacilli (Nikitushkin *et al.*, 2013; Machowski *et al.*, 2014).

AG is a heteropolysaccharide consisting of galactan, arabinan, and linker unit. Both the galactan and arabinan exist solely in the furanoid ring form, i.e., D-galactofuranosyl (Galf) and D-arabinofuranosyl (Araf), respectively (McNeil *et al.*, 1987). The galactan is a linear structure comprising 30 Galf residues with alternating β-(1-5) and β-(1-6) glycosidic linkages (Mikušová *et al.*, 2006). The arabinan, on the other hand, has a highly branched structure with about 30 Araf residues attached at three branch points of galactan (Daffe *et al.*, 1990). While the galactan is covalently

bound to the muramic acid residues of mycobacterial PG, the non-reducing end of arabinan is covalently attached to MAs of mycomembrane (Alderwick *et al.*, 2015).

2.3.2(d) Periplasm and Lipoglycans

Lipoglycans, i.e., phosphatidylinositol mannosides (PIMs), LM, and lipoarabinomanan (LAM), represent another group of mycobacterial cell wall components with an indefinite localization. These lipoglycans share a common phosphatidylinositol (PI) core decorated with mannose residues (see Figure 2.6A and B). However, PIMs only contain up to six mannose residues, while LM and LAM usually have more than 20 mannose residues (Figure 2.6C and D) (Hunter & Brennan, 1990; Bansal-Mutalik & Nikaido, 2014). Besides, LAM has additional arabinose branches whose non-reducing termini can be linked with mannose residues, forming ManLAM (see Figure 2.6D) (Chatterjee *et al.*, 1992).

PIMs have been demonstrated to be structural and functional components of the plasma membrane (Bansal-Mutalik & Nikaido, 2014), while its derivatives, LM and LAM, are believed to localize in both the plasma membrane (periplasmic space) and the mycomembrane, maintaining the cell envelope integrity and modulating interactions with the immune system (Pitarque *et al.*, 2008; Rahlwes *et al.*, 2019). Notably, LAM not only plays an important role in establishing infection by interacting with host receptors (e.g., Dectin-2 and lactosylceramide in lipid rafts) but also facilitates immune evasion by blocking the maturation of phagosomes containing the tubercle bacilli (Yonekawa *et al.*, 2014; Nakayama *et al.*, 2016; Batt *et al.*, 2020).

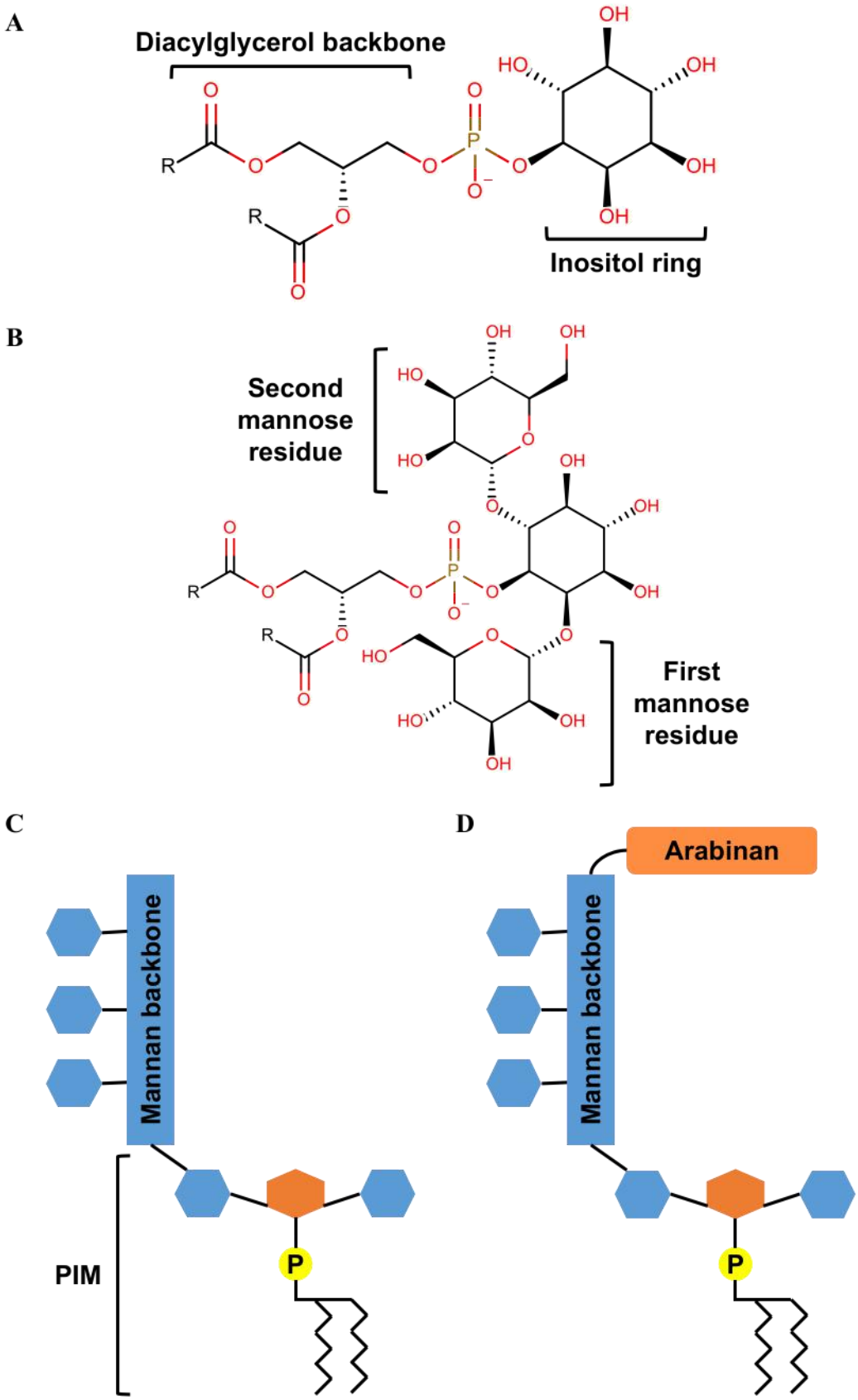

Figure 2.6 Structure of PI and lipoglycans. A. PI; B. PIM; C. LM; D. LAM.

2.3.3 Intrinsic Resistance to Antibiotics

Apart from the emergence of MDR-TB, the efforts at TB control and treatment have also been hampered by the intrinsic resistance of *M. tuberculosis* to many antituberculous drugs due to various mechanisms, including the above-mentioned unique features, i.e., thick cell wall and metabolic remodeling for dormancy or persistence (see Section 2.3.1 and 2.3.2). It is also worth noting that acquired antibiotic resistance in *M. tuberculosis* is predominantly attributed to chromosomal mutations instead of the extrinsic acquisition of antibiotic resistance genes via horizontal transfer of resistance plasmids or mobile genetic elements (Rabello *et al.*, 2012; Gygli *et al.*, 2017). In addition to the aforementioned mechanisms, drug/target modifications and drug efflux systems in *M. tuberculosis* also contribute to its intrinsic resistance to a broad range of antituberculous drugs (see Figure 2.7).

2.3.3(a) Drug and Target Modifications

Drugs capable of penetrating the mycobacterial cell wall may be inactivated enzymatically via either cleavage or modifications. For instance, the genome of *M. tuberculosis* harbors the Rv2068c gene that encodes for a class A (Ambler) β-lactamase gene (blaC), which hydrolyzes the β-lactam ring of β-lactam antibiotics, to which the tubercle bacilli are intrinsically resistant, including carbapenems (Gygli *et al.*, 2017). *M. tuberculosis* can also inactivate antibiotics via modifications (e.g. methylation or acetylation), for example, N-acetyltransferase (encoded by Rv2416c), also termed enhanced intracellular survival protein (Eis), can inactivate many aminoglycoside (kanamycin and amikacin) and cyclic peptide (capreomycin) antibiotics via acetylation (Houghton *et al.*, 2013a; Houghton *et al.*, 2013b).

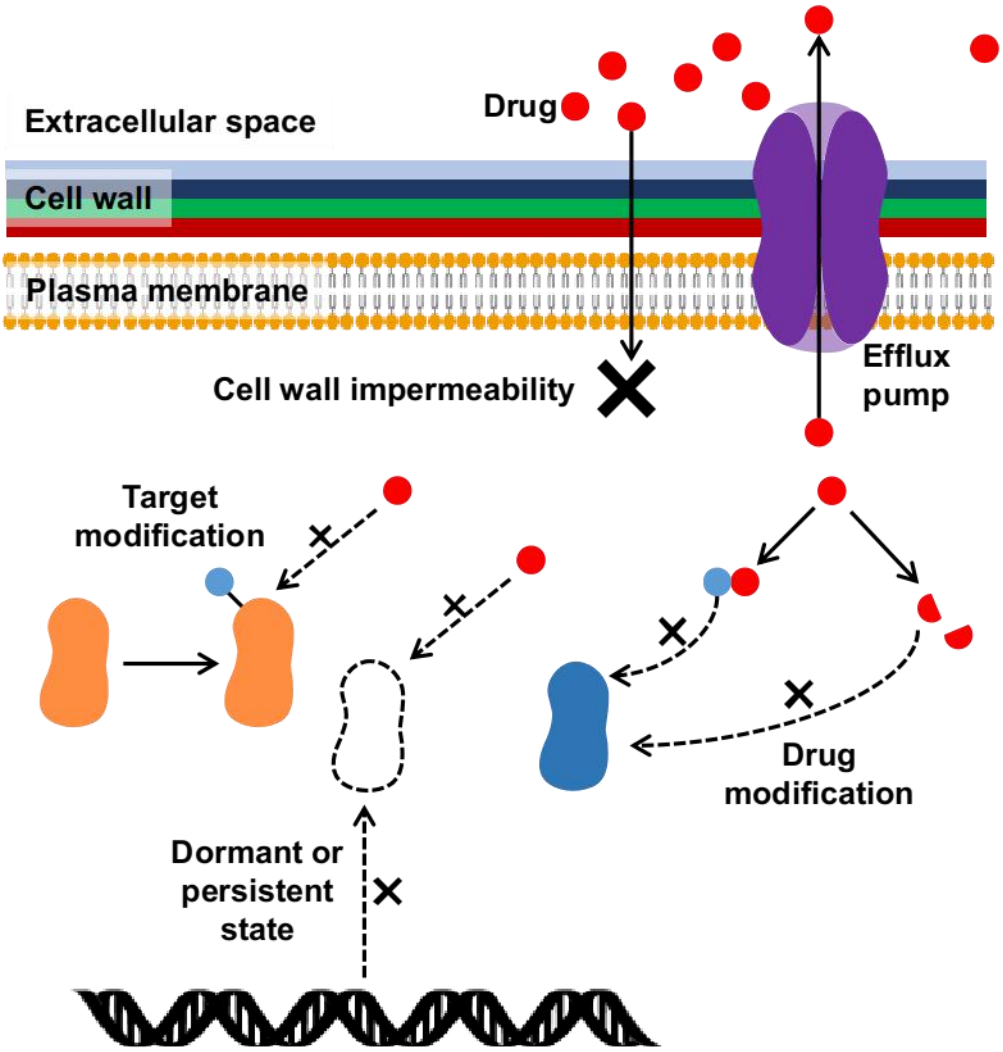

Figure 2.7 Mechanisms of antibiotic resistance in *M. tuberculosis*

A more recent example of drug modification by *M. tuberculosis* is the N-methylation of compound 14 (3-methyl-1-oxo-2-[3-oxo-3-(pyrrolidin-1-yl)propyl]-1,5-dihydrobenzo[4,5]imidazo[1,2-a]pyridine-4-carbonitrile) and TPSA (2-[5-(2-{[4-(2-thienyl)-2-pyrimidinyl]sulfanyl}acetyl)-2-thienyl]acetic acid) by 2-heptyl-1-hydroxyquinolin-4(1H)-one methyltransferase (htm, encoded by Rv0560c) (Warrier *et al.*, 2016; Chen *et al.*, 2019). The inhibitory effects of methylated compound 14 and TPSA against their targets, i.e., GlmU (Rv1018c) and decaprenylphosphoryl-β-D-ribose oxidase (dprE1, Rv3790), respectively, are affected by Rv0560c (Warrier *et al.*, 2016; Chen *et al.*, 2019). Besides, modifications can also be done on drug targets

instead of the drugs to prevent drug-target interactions in *M. tuberculosis*. For instance, the monomethylation of residues 2057-2059 of the 23S rRNA by 23S rRNA methyltransferase encoded by Erm(37)/Rv1988 confers resistance to various macrolide antibiotics (Warrier *et al.*, 2016).

2.3.3(b) Efflux Pumps

Mycobacterial efflux pumps are transmembrane transporters that export various small molecules, including various antibiotics, whose cytoplasmic concentrations can be reduced to sub-inhibitory levels, conferring resistance to nearly all known antituberculous drugs (Anthony Malinga & Stoltz, 2016; Chauhan *et al.*, 2021). The genome of *M. tuberculosis* encodes for six types of efflux pumps: ATP-binding cassette transporters (ABC transporters), Small Multidrug Resistance (SMR), Resistance Nodulation Division (RND), Major Facilitator Superfamily (MFR), Drug Metabolite Transporter superfamily (DTM), and Multidrug And Toxic Compound Extrusion (MATE), among which the first four types are unique to *M. tuberculosis* (Liu & Xie, 2014).

The induced expression of genes encoding efflux pumps upon exposure to stresses (e.g., antibiotics) confers resistance to antibiotics even in the absence of resistance-associated mutations, for examples, the upregulation of efflux pumps encoded by Rv2936-Rv2937/drrA-drrB, Rv3065/mmr, and Rv1258c/tap results in resistance to doxorubicin, erythromycin, and rifampicin respectively (Choudhuri *et al.*, 2002; Sharma *et al.*, 2010b). Besides, the ABC efflux pump complex encoded by Rv1217c-Rv1218c confers resistance to multiple structurally unrelated antibiotics, including rifampicin and isoniazid (Wang *et al.*, 2013).

2.4 Bacterial Small Regulatory RNAs

The resilience and adaptation of bacteria to various hostile environments, e.g., extreme pH, temperature, nutrient, and oxygen levels, are achieved via their complex yet fine-tuned regulation of gene expression (Hoe *et al.*, 2013; Citartan *et al.*, 2016). The exploration of regulatory mechanisms facilitating bacterial adaptation to different stresses is of great significance for future antibiotic development (Walsh & Wencewicz, 2014; Cheah *et al.*, 2018). Bacterial sRNAs are short transcripts that lack open reading frames (ORFs) with sizes ranging between 50 and 500 nts (Huang *et al.*, 2009; Cheah *et al.*, 2018). Over the past few decades, extensive transcriptome studies on bacteria, especially after the advent of high-throughput sequencing technology, concluded that sRNAs are important post-transcriptional regulators aside from transcription factors in various biological functions, including but not limited to stress response, virulence, and host-pathogen interactions (Kroger *et al.*, 2012; Baddal *et al.*, 2015; Barquist & Vogel, 2015; Westermann *et al.*, 2016; Pisu *et al.*, 2020). These transcriptome analyses revealed that prokaryotic transcription is not limited to coding regions and might be far more sophisticated than previously expected.

2.4.1 Classification of Bacterial sRNAs

Bacterial sRNAs can be further differentiated according to size, mechanism of action, and genomic localization, for instance, sRNAs can be broadly classified based on their genomic location into intergenic sRNAs and *cis*-antisense sRNAs. Antisense RNAs transcribed opposite to their target genes are *cis*-antisense sRNAs that often negatively regulate the expression of their target mRNAs via base-pairing interactions with perfect complementary or mechanisms associated with promoters (Georg & Hess, 2011). Intergenic sRNAs, on the other hand, are located between

annotated genes and often represent *trans*-antisense sRNAs that interact with distant mRNA targets via base-pairing interactions with imperfect complementary (Papenfort & Vogel, 2010).

However, the classification of sRNAs based on genomic locations is far-fetched as protein-coding genes in bacterial genomes are often annotated via computational ORF prediction, which often overlooks the UTRs of mRNAs, including 5'UTRs, 3'UTRs, and intercistronic regions of operons (Cheah *et al.*, 2018). As a result, it is possible that some of the presumed intergenic, *trans*-antisense sRNAs are indeed *cis*-antisense RNAs that only overlap in an opposite orientation with the UTRs of mRNAs or UTR-derived sRNAs that originates from 5'UTRs, 3'UTRs, or even intercistronic regions (see Figure 2.8).

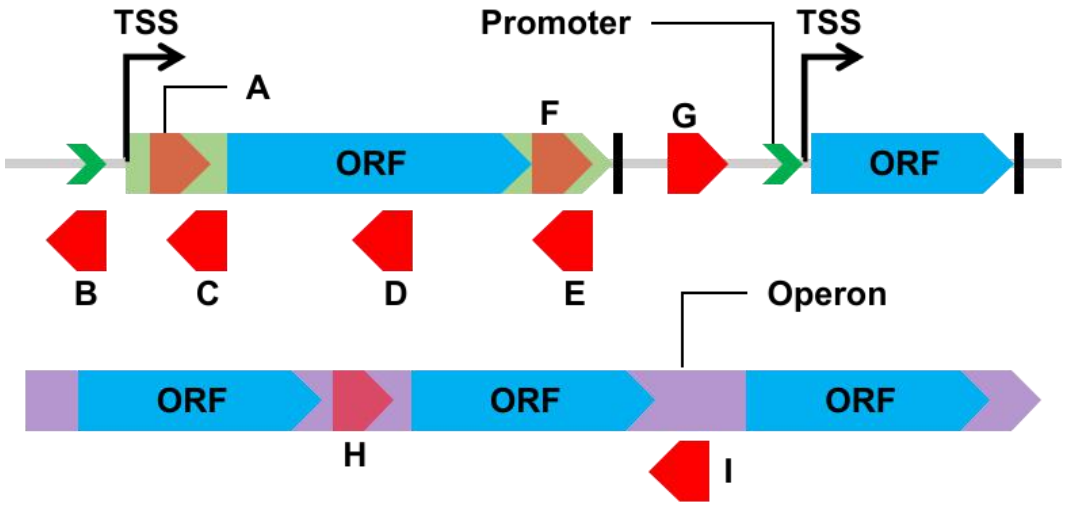

Figure 2.8 Classification of sRNAs based on their genomic locations. A. 5'UTR-derived sRNAs; B. Promoter-associated sRNAs; C. 5'UTR-antisense sRNAs; D. ORF-antisense sRNAs; E. 3'UTR-antisense sRNAs; F. 3'UTR-derived sRNAs; G. Individual intergenic sRNAs; H. Operon-derived sRNAs; I. Intercistronic-antisense sRNAs.