COMPARISON OF PROTEOMES OF

AVIRULENT *Mycobacterium tuberculosis* H37Ra AND

VIRULENT *Mycobacterium tuberculosis* H37Rv

ANG KAI CHEEN

UNIVERSITI SAINS MALAYSIA

2011

COMPARISON OF PROTEOMES OF AVIRULENT *Mycobacterium tuberculosis* H37Ra AND VIRULENT *Mycobacterium tuberculosis* H37Rv

by

ANG KAI CHEEN

Thesis submitted in fulfillment of the requirements for the Degree of Master of Science

September 2011

ACKNOWLEDGEMENT

First of all, I would like to thank my supervisor, Assoc. Prof. Dr Gam Lay Harn for her patience, motivation, enthusiasm, and immense knowledge throughout this study whether in designing experimental works or in writing this thesis. I could not have a better advisor and supervisor for my M.Sc study.

Special thanks to my co-supervisor, Assoc. Prof. Dr Pazilah binti Ibrahim who has a *Mycobacterium tuberculosis* laboratory at my disposal for the culture of *Mycobacterium tuberculosis*.

My sincere thanks to Assoc. Prof. Dr. Tang Thean Hock who has given me excellent opinions, sound advice and encouragement in my study.

I am grateful to the Institute of Postgraduate Studies, Universiti Sains Malaysia for the Graduate Assistance and Fellowship Scholarship Scheme which helped to finance my studies. Furthermore, I would like to extend my gratitude to the staff of the School of Pharmaceutical Sciences and the National Poison Centre for providing facilities, equipments and assistance in my work.

My gratitude and thanks to my fellow lab mates, Seng Liang, Lay Chin, Chu Ai, Hui Hua, Siew Yong, Siang Yian, Chaw Sen, Thaigarajan and Shi Rou for providing moral support, assistance and sharing their knowledge with me.

Last but not the least, my deepest gratitude to my parents in Johor Bahru for everything they have done for me. You have being a source of inspiration and confidence throughout my life, and I am proud to be your daughter.

TABLE OF CONTENTS **Page**

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

2-DE	: Two dimensional electrophoresis
3D	: Three dimensional
4CN	: 4-Chloro naphthol
CAN	: acetonitrile
Acr	: α-crystallin
ADC	: Albumin-Dextrose-Catalase
AFB	: Acid Fast Bacillus
APS	: ammonium persulphate
ATCC	: American Type Culture Collection
BSA	: bovine serum albumin
BCG	: Bacillus Calmette-Guérin
C	: cytosine
$CaCl_2$	: calcium dichloride
CBB	: Coomassie Brilliant Blue
CID	: collision induced dissociation
CNRs	: classical nitroreductases
CO_2	: carbon dioxide
CFP-10	: culture filtarte protein-10
DI	: deionised water
DIMs	: phthiocerol dimycocerosates
DNA	: deoxyribonucleic acid
DTT	: dithiothreitol
ESAT-6	: early secretory antigenic target-
ESI	: electrospray ionization
ETF	: electron transfer flavoprotein
G	: guanine
HIV	: human immunodeficiency virus
Hrs	: hours
ICC	: ion charge control
IEF	: isoelectric focusing
IFN	: interferon
INH	: isoniazid
IPG	: immobilized pH gradient
kDa	: kilo dalton
LAM	: lipoarabinomanan
LC	: liquid chromatography
LCMS/MS	: liquid chromatography and tandem mass spectrometry
LM	: lipomannan
m/z	: mass to charge ratio
MHC	: major histocompatibility complex
MDR resistance	: multi-drug resistance
Mgf	: mascot generic format
Min	: minute
MS	: mass spectrometry
MSDB	: Mascot Protein Database
MSMS	: tandem mass spectrometry
NBDs	: nucleotide binding domains
NH_4HCO_3	: ammonium bicarbonate
NTP	: National TB Control Program

OFN	: oxygen free nitrogen
ORF(s)	: open reading frame(s)
PAS	: paraaminosalicylic acid
PDIM	: phthiocerol dimycoserosates
PGL	: phenolic glycolipids
PMNs	: polymorphonuclear granulocytes
p.p.m.	: part per million
PCR	: Polymerase Chain Reaction
PE	: proline-glutamic acid
PGLs	: phenolic glycolipids
p*I*	: isoelectric point
PPD	: purified protein derivatives
PPE	: proline-proline-glutamic acid
rBCG	: recombinant BCG
r.p.m.	: revolutions per minute
R^2	: correlation coefficient
RF	: radio frequency
SD	: standard deviations
SDS-PAGE	: sodium dodecyl sulphate polyacrylamide gel electrophoresis
Sec	: second
SIB	: Swiss Institute of Bioinformatics
SL	: sulfolipids
TB	: tuberculosis
TCA	: trichloracetic acid
TEMED	: N,N,N',N'-Tetramethylethylenediamine
TIFF	: tagged image file format
TM	: transmembrane
TMA	: transcription mediated amplification
TLB	: thiourea lysis buffer
WHO	: World Health Organization
XDR	: extremely drug resistant
ZN	: Ziehl-Neelsan

PERBANDINGAN PROTEOM ANTARA *Mycobacterium tuberculosis* AVIRULEN H37Ra DENGAN *Mycobacterium tuberculosis* VIRULEN H37Rv

ABSTRAK

M. tuberculosis H37Ra adalah strain terbitan daripada strain virulen H37 yang dijumpai pada tahun 1935. Bagi mengenalpasti faktor-faktor virulen untuk *M. tuberculosis* H37Rv, proteom- proteom strain virulen yang dihasilkan melalui elektroforesis gel dua-dimensi pada tiga jangka masa kultur yang berbeza (14, 28 dan 50 hari) telah dibandingkan dengan proteom bagi strain avirulen *M. tuberculosis* H37Ra. Sejumlah 250 tompok protein telah dikenalpasti selepas pewarnaan Coomassie Blue dan proteom *M. tuberculosis* H37Rv dan *M. tuberculosis* H37Ra adalah hampir sama (R^2~0.70), tiga puluh satu tompok protein yang menunjukkan perbezaan ekspreksi pada tempoh 14, 28 dan 50 hari pertumbuhan telah dikenalpasti, iaitu 14 tompok protein daripada ekstraksi Tris dan 17 tompok protein daripada ekstraksi TLB. Daripada 26 tompok protein yang dianalisis dengan mengunakan LC-MS/MS, 32 protein telah dikenalpasti. *L*-guluno-1,4-laktone dehidrogenase (Rv1771) menunjukkan peningkatan-regulasi pada hari ke-14 untuk *M. tuberculosis* H37Rv sahaja. Tiga protein menunjukkan perbezaan ekspresi pada hari ke-28 untuk *M. tuberculosis* H37Rv, iaitu protein hipotetikal terabadi Acg (*acg*, Rv2032), pengangkut makrolida, protein pengangkut kaset pengikat ATP (pengangkut ABC, Rv2477c) dan asid-lemak-CoA ligase (*fadD28*, Rv2941). Acg yang terlibat dalam dormasi regulon menunjukkan penurunan-regulasi dalam *M. tuberculosis* H37Rv. Rv2477c dan fadD28 pula menunjukkan peningkatan-regulasi dalam *M. tuberculosis* H37Rv. *M. tuberculosis* menunjukkan kerintangan terhadap antibiotik umum, hal ini mungkin disebabkan oleh kehadiran beberapa pengangkut ABC pada membran sel *M. tuberculosis* yang boleh membawa antibiotik keluar daripada sel. Penurunan-regulasi protein fadD28 yang terlibat dalam biosintesis dinding sel membawa implikasi

kelemahan pertumbuhan *M. tuberculosis* H37Ra secara intraselular. Kajian ini memberi maklumat proteom untuk *M. tuberculosis* H37Rv dan *M. tuberculosis* H37Ra dan perbezaan ekspresi protein antara kedua-dua strain ini boleh memberi pemahaman yang lebih baik tentang fenotip dan mekanisme pathogenesis antara *M. tuberculosis* H37Rv dengan strain tak virulen *M. tuberculosis* H37Ra. Penyiasatan lanjutan ke atas protein-protein yang dilaporkan mungkin dapat membantu dalam perkembangan biopenanda yang berpontensi untuk tujuan diagnostik dan pengubatan penyakit tuberkulosis.

COMPARISON OF PROTEOMES OF AVIRULENT *Mycobacterium tuberculosis* H37Ra AND VIRULENT *Mycobacterium tuberculosis* H37Rv

ABSTRACT

M. tuberculosis H37Ra, which was discovered in 1935, was derived from the virulent strain H37. In order to identify the factors that are correlated to virulence in *M. tuberculosis* H37Rv, 2D-gel profile of three virulent strains at culture grown for three time points (14, 28 and 50 days) were compared with its avirulent counterpart *M. tuberculosis* H37Ra. A total of 250 protein spots were visualized by Coomassie blue staining. The protein profiles of the *M. tuberculosis* H37Rv and *M. tuberculosis* H37Ra were highly similar (R^2~0.70), thirty one spots were found to be differentially expressed at 14, 28 and 50 days cultures, whereby 14 spots were identified from Tris extract and 17 spots were identified from TLB extract. Twenty-six spots could be identified by LCMS/MS and showed to consist of 32 proteins. *L*-guluno-1,4-lactone dehydrogenase (Rv1771) was significantly up-regulated in 14-day culture of *M. tuberculosis* H37Rv only. Three proteins showed differential expression in 28-day culture only. These proteins include the conserved hypothetical protein Acg (*acg*, Rv2032), macrolide-transport ATP-binding cassette protein (ABC transporter, Rv2477c) and fatty-acid-CoA ligase (*fadD28*, Rv2941). Acg which is involved in dormancy regulon was down-regulated in *M. tuberculosis* H37Rv. Rv2477c and fadD28 were up-regulated in *M. tuberculosis* H37Rv. *M. tuberculosis* is usually resistant to common antibiotics; this may be due to the presence of several ABC transporters on the cell membrane that could transport the antibiotics out from cells. Down-regulation of fadD28, which is involved in cell wall biosynthesis, may explain the low survival of *M. tuberculosis* H37Ra intracellularly. This work has provided the information on the proteome of *M. tuberculosis* H37Rv and *M. tuberculosis* H37Ra, the differential protein expression between the two strains may give a better understanding on phenotypic and pathogenesis

between *M. tuberculosis* H37Rv and its avirulent counterpart *M. tuberculosis* H37Ra. Further investigation on the reported proteins may also lead to identification of candidate biomarkers for diagnosis and treatment of the disease.

CHAPTER 1

INTRODUCTION

1.1 The historical background and the present burden of the tuberculosis disease

Tuberculosis (TB) is an ancient disease (Daniel, 2006). It is known to be one of the most persistent diseases of mankind. TB causes illness and death in a wide geographical area (Figure 1.1). The first evidence of TB in human was detected in Egyptian mummies dated from 2400 B.C. whereby the identified microbial genetic material recovered from the bone tissues was shown to be similar to that of Mycobacterium genetic material (Morse *et al.*, 1964; Zimmerman, 1979; Crubéy *et al.*, 1998).

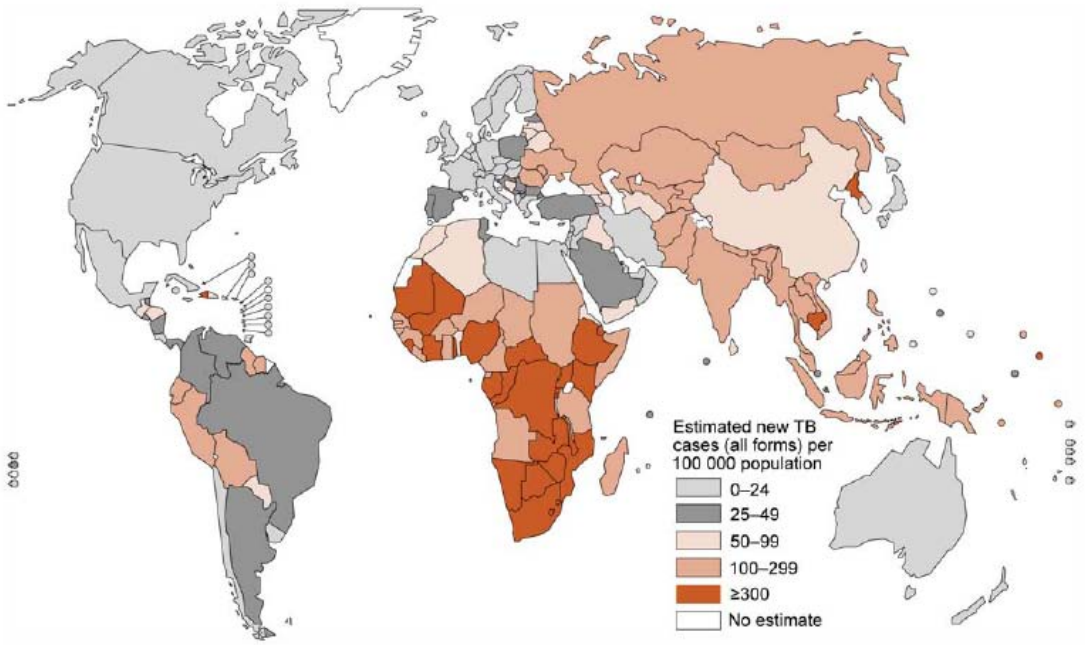

Figure 1.1 The global incidence of TB. The number of new TB cases per 100,000 populations for the year 2008 according to WHO report (Adapted from WHO, 2009).

Around 460 B.C., a term ‘phthsis’ was used to describe TB. Hippocrates stated that ‘phthsis’ is the most widespread disease of that time. Later, Aristotle (382 to 322 B.C) and Galen (130 to 200 A.D) reported that TB can be transmitted from man to man. Hence, TB epidemic which is known as ‘white plague’ was a dreadful threat during the 17^{th} and 18^{th} centuries in Europe.

In the period from 1483 to 1553, Frascatrius suggested that the mode of TB transmission from human to human is via aerosol. In 1865, a French military doctor, Jean-Antoine Villemin, performed an experiment by infecting rabbits with liquids from tuberculosis cavities. Indeed, the first breakthrough of understanding TB was reported by a German bacteriologist, Robert Koch in 1882. He has isolated the infectious agent of tuberculosis and then infected the disease to animals. Later, he had also developed a staining method for the infectious agent and examined a rod shaped agent under microscope (Koch, 1882; Kaufmann, 2003; Daniel, 2005). Later, he further discovered a compound, named tuberculin which could be used as an agent to diagnose TB. The tuberculin skin test is still used today with some modifications (Kaufmann and Schaible, 2005). In 1896, Lehmann and Newman classified this organism as Mycobacterium genus that belonged to Mycobacteriaceae family (Ridley, 1981).

Between 1908 and 1920, an attenuated strain derived from *Mycobacterium bovis* (*M. bovis*) was developed by two Frenchmen (Calmette and Guérin). *M. bovis* is a species closely related to *M. tuberculosis* that was found in cattle. Nearly 200 times of subculturing of virulent *M. bovis*, resulting in the discovery of attenuated strain called Bacillus Calmette-Guérin (BCG) and this has served as a vaccine against tuberculosis (Barreto *et al.*, 2006).

With the advent of BCG vaccination, the disease was substantially under control, although it was not totally eliminated. In 1985, the rapid increase of new TB cases globally urged WHO to declare tuberculosis as a 'global emergency' and subsequently WHO implemented controlled TB plan in 1993 (WHO, 1993). The resurgence of TB cases was partly due to the increased urbanization, civil wars, mass migration of population, homeless individuals and the HIV epidemic (Frieden *et al.*, 2003). WHO estimated that eight million people will be infected with TB yearly, of which 95 % of these cases would be from developing countries. WHO also estimated that two millions people will die of TB yearly. In order to actively prevent the spread of TB, WHO and Stop TB Partnership endorsed a global plan to reduce TB prevalence and mortality by 2015 and to eliminate TB by 2050 (WHO, 2006).

In spite of the global preventive TB plan, the disease remains as an alarming public health problem. This is also partly due to the emergence of multi-drug (MDR) and extremely drug (XDR) resistance TB cases. Multi drug resistance TB is defined as TB caused by strains resistant to the first line anti-TB drugs, such as isoniazid and rifampicin (Dye *et al.*, 2002), whereas XDR-TB is defined as MDR-TB that is also further resistant to several main classes of the second line anti-TB drugs, such as fluoroquinolones (Raviglione and Smith, 2007). All these problems reflect the importance of an intensive research to investigate:

a) an effective TB vaccine for prevention.
b) an effective anti-TB medication for treatment.
c) an improvement of the current diagnosis of TB.

1.2 Tuberculosis in Malaysia

In the early 1940s and 1950s, TB was the main contagious disease in Malaysia. Although there was effective chemotherapy for TB in late 1950s, yet TB remained as a major cause of mortality at that time. The Malaysian government realized the seriousness of TB and launched the National TB Control Program (NTP) in 1961. At that time, a combination of three drugs regime was recommended and the duration of treatment was 1-2 years. Streptomycin, isoniazid and paraaminosalicylic acid (PAS) were administered for the first two months, followed by isoniazid and PAS for the next 12 months. Furthermore, school children were vaccinated by BCG vaccination (Figure 1.2).

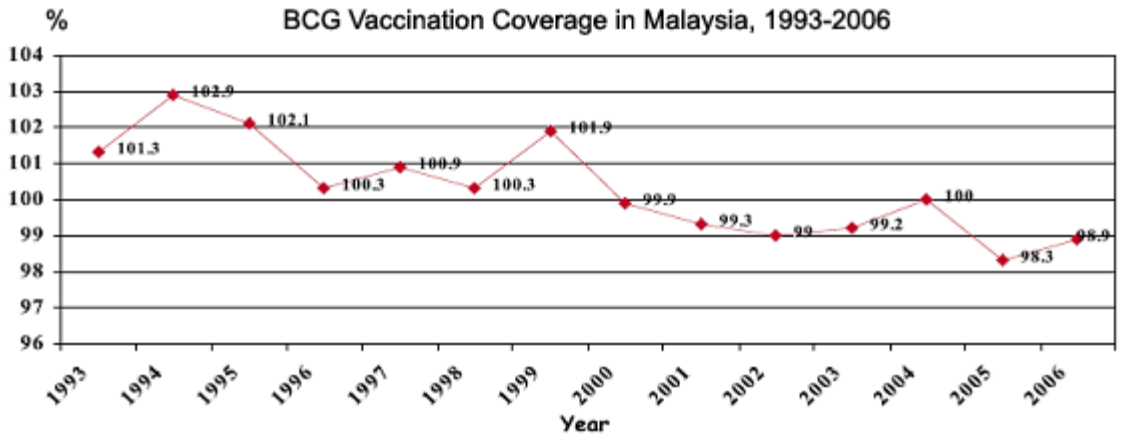

Figure 1.2 The coverage rate (%) of BCG vaccination in Malaysia from 1993 to 2006 (Adapted from Venugopalan, 2007).

The headquarters of National TB Centre is located in Kuala Lumpur while the state general hospitals with chest clinics functioned as the state directorates. The state directorate is known as the State TB Managerial Team; this team is mainly responsible for the implementation of the activities of the NTP at the state and district levels. Owing to the proper control prevention strategy, TB cases were dramatically reduced recently.

The resurgence of TB is mainly due to human immunodeficiency virus (HIV) in Malaysia (Figure 1.3). In 1990, the number of TB/HIV co-infection cases has escalated and the number of deaths due to TB/HIV co-infections has increased. In 1996, 329 TB/HIV co-infection cases were reported compared to 6 cases reported in 1991. The National TB Centre has now been renamed as The Institute of Respiratory Medicine. In 2007, TB was one of the top 5 prominent infectious diseases, therefore WHO classified Malaysia as an intermediate-TB-burden country.

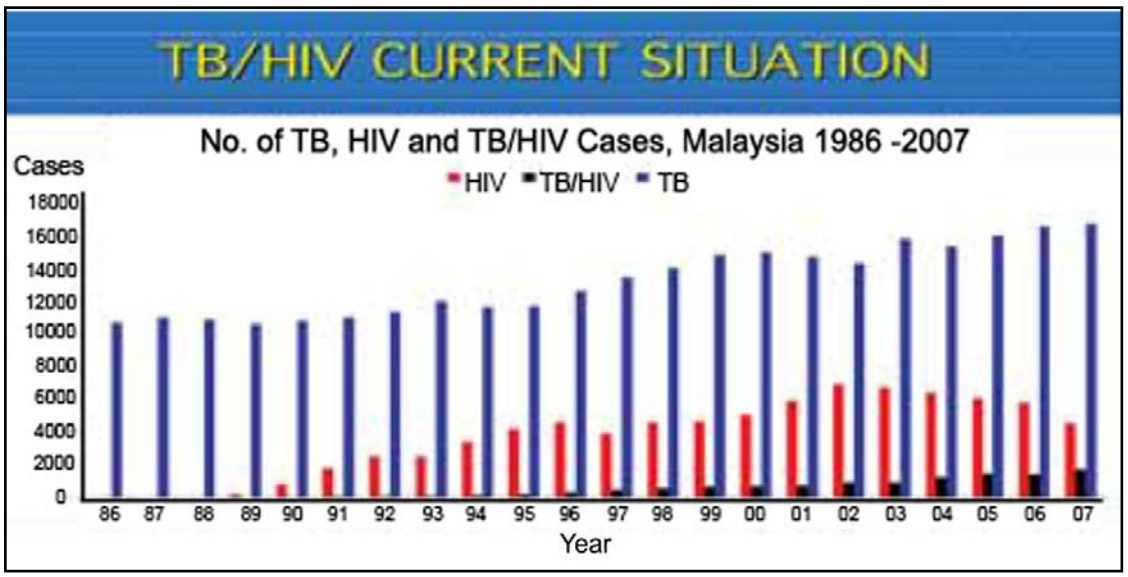

Figure 1.3 The cases of TB, HIV, TB/HIV in Malaysia from 1986 to 2007 (Adapted from Venugopalan, 2009).

1.3 The structure of Mycobacteria cell envelope

Mycobacterium is a non-motile and non-sporulated rod-shaped bacterium. The bacteria are closely related to Actinomycetes with high lipid content in the cell wall (Bloom, 1994; Kolattukudy *et al.*, 1997). The cell envelope of Mycobacterium comprises of plasma membrane, cell wall and outer like capsule layer (Figure 1.4).

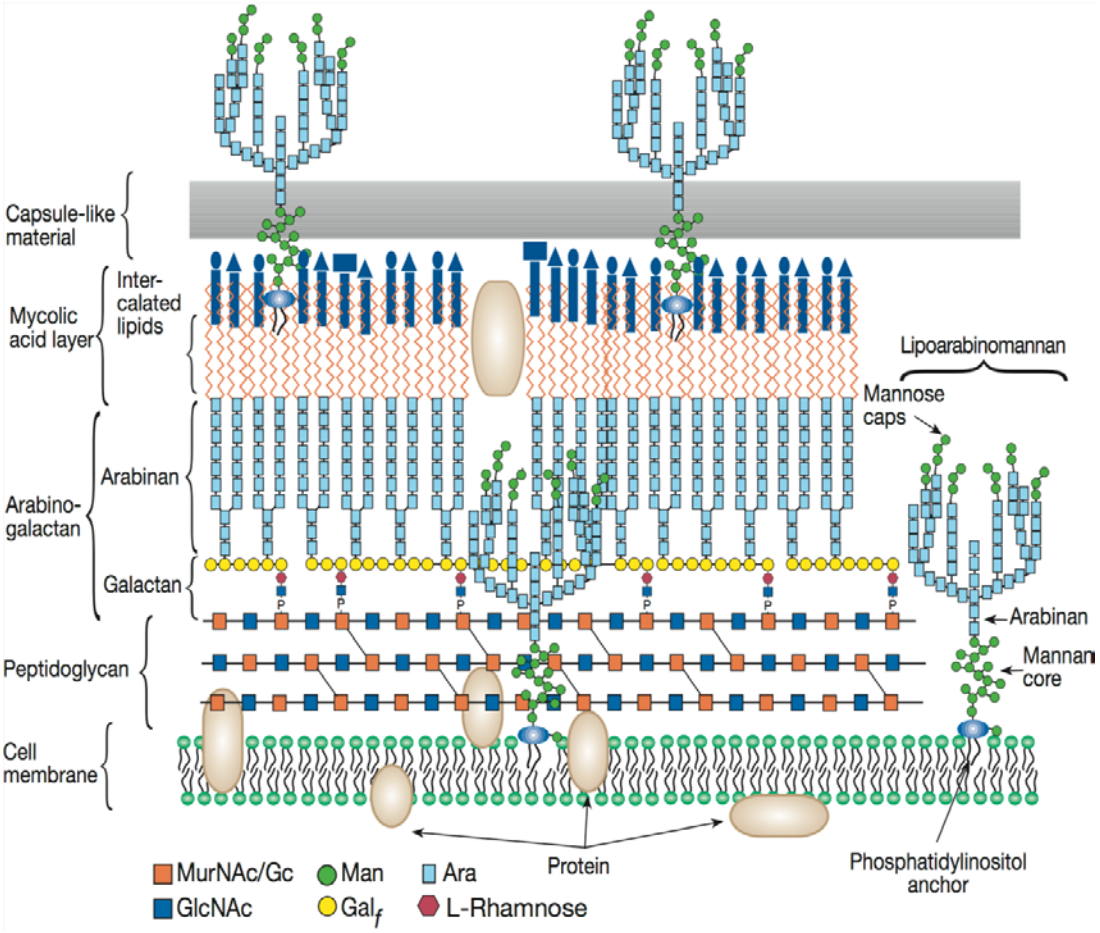

Figure 1.4 Structure of the cell envelope of Mycobacteria (Adapted from Brennan and Crick, 2007).

1.3.1 Plasma membrane

The plasma membrane of Mycobacteria contains some lipopolysaccharides that provides osmotic protection, regulates the traffic of specific solutes between the cytoplasm and the environment, and also organizes the cell house-keeping tasks (Mahapatra *et al.*, 2005).

1.3.2 Cell wall

The cell wall of Mycobacteria is very unique and is associated with pathogenicity of *M. tuberculosis*. The innermost cell wall layer is peptidoglycan, which is responsible for shape-forming property and the structural integrity of the bacterium. The building block of peptidoglycan layer in Mycobacteria consists of *N*-glycolyl muramic acid units

(Crick *et al.*, 2001). The arabinogalactan layer is covalently bound to peptidoglycan. The surface of arabinogalactan layer has mycolic acids that are typically long and branches carbon chains containing 60- to 90- carbon atoms. The outer layer of the cell wall consists of an array of free lipids such as phthiocerol dimycoserosates (PDIM), phenolic glycolipids (PGL), trehalose-containing glycolipids and sulfolipids (SL). Some glycolipids are anchored to the plasma membrane and extended to the exterior of the cell wall such as the phosphatidyl-myoinositol mannosides, lipomannan (LM) and lipoarabinomanan (LAM). These fatty acids and lipids pose Mycobacteria virulence. Furthermore, they protect the bacteria against some antibiotics and from the immune system of the host (Chatterjee and Khoo, 1998; Takayama *et al.*, 2005).

Interspersed proteins are also found in Mycobacteria cell wall, some of these proteins are responsible for the construction of Mycobacteria cell wall. For example, porins form hydrophilic channels which allow the aqueous solutes to enter the mycolic acid layer (Brennan and Nikaido, 1995).

1.3.3 Outer capsule like layer

Polysaccharides, proteins and certain extractable lipids are found in the outer-most capsule (Brennan and Nikaido, 1995). Hence, the lipid and fatty acid biosynthetic pathways in Mycobacteria hold the potential study of cell membrane and wall proteins because they may be involved in virulence pathogenesis.

1.4 Mycobacteria growth and metabolism

1.4.1 Growth

Mycobacteria are divided into two groups namely rapid-growing and slow-growing *Mycobacterium* species. Rapid-growing Mycobacteria consists of *M. smegmatis* and *M. fortuitum* (doubling time is ~ 2 to 3 hrs) whereas slow-growing Mycobacteria are *M. avium-intracellulare* complex (doubling time ~ 10 to 12 hrs), *M. tuberculosis* complex (doubling time ~ 22 to 24 hrs) and *M. leprae* (doubling time ~ 14 days *in vivo*).

Mycobacterium is prototrophic (*i.e.* it can synthesize all its components from basic carbon and nitrogen sources) and heterotrophic (*i.e.* it uses synthetic organic compounds as a source of carbon and energy). Most of the *Mycobacterium* requires carbon source, nitrogen source and essential metal ions such as iron and magnesium for propagation. In general, a simple growth medium contains all of these sources.

M. tuberculosis multiplies slowly in every 12 to 24 hours even under favorable laboratory conditions. This is partially due to the cell wall impermeability that limits nutrient uptake (Chauhan *et al.*, 2006). The life cycle of *M. tuberculosis* depends on the nutritional quality of the environment, either in the natural habitat or in the culture media. A high concentration of fatty acids slows down the growth of *Mycobacterium*.

1.4.2 Metabolism

M. tuberculosis has the ability to synthesize the essential amino acids, vitamins and enzyme co-factors (Bloom, 1994). The enzymes for essential pathways such as glycolysis, pentose phosphate pathway, tricarboxylic acid and glyoxylate cycles are found in *M. tuberculosis* (Koníček, 1984).

Under aerobic respiration, *M. tuberculosis* produces ATP from oxidative phosphorylation, which involves a ubiquinone cytochrome *b* reductase complex and cytochrome *c* oxidase. Nevertheless, *M. tuberculosis* can survive in anaerobiosis due to the presence of several anaerobic phosphorylative electron transport chains in mitochondrial membrane. *M. tuberculosis* conducts different physiological pathways during the change in environment. A study reported that *M. tuberculosis* is capable of alternating the metabolism from aerobic and carbohydrate-metabolizing mode to microaerophilic and lipids-utilizing mode (Segal and Bloch, 1956).

1.5 Pathogenesis of tuberculosis

TB is a transmissible disease, affecting various parts of body's tissues especially the lung. The lung is the main entrance of *M. tuberculosis*. Generally the infection can be subdivided into 5 stages as below:

1.5.1 Stage I of *M. tuberculosis* infection

When a person inhales aerosol droplets, the larger droplets (> 0.5 μm) are lodged in the upper respiratory tract (the nose and throat) but the smaller droplets (< 0.5 μm) may reach the alveoli. In the alveolar space, the bacteria encounter polymorphonuclear granulocytes (PMNs) and inactive macrophages (Persson *et al.*, 2009). Hence, the initial infection begins.

1.5.2 Stage II of *M. tuberculosis* infection

Between 1-3 weeks after initial infection, *M. tuberculosis* stays dormant and replicates freely within the inactive macrophages. Due to this, the other macrophages from peripheral blood will gather at the infection site, these inactive macrophages will try to phagocytose *M. tuberculosis*. In order to survive within the inactive macrophage, *M.*

tuberculosis has to overcome intra-phagosomal degradation from the fusion of phagolysosome (Armstrong and Hart, 1975), thus delaying apoptosis (Zhang *et al.*, 2005; Chen *et al.*, 2006). In other words, these inactive macrophages cannot diminish *M. tuberculosis*.

1.5.3 Stage III of *M. tuberculosis* infection

In order to activate macrophages to destroy *M. tuberculosis*, the cell-mediated immune system has to be triggered. The lymphocytes such as T-cells will infiltrate the infection site. T-cells will recognize the *M. tuberculosis* antigens which have been processed and presented by major histocompatibility complex (MHC) molecules. As a result, T- cell is activated and produce cytokines such as gamma interferon (IFN). IFN will activate the inactive macrophage and promote *M. tuberculosis* destruction (Smith, 2003).

At this stage, patients infected by TB will show positive tuberculin skin test due to the vigorous cell-mediated immune response (Puri and John, 1997). *M. tuberculosis*, an intracellular pathogen, can only be controlled by cell-mediated immune system. In addition, activation of macrophages results in the synthesis of lytic enzymes and reactive intermediates. These compounds are the major impact on the immune pathology. Tubercle formation begins at this stage, and the center of tubercle contains cheesy-semi-solid materials known as caseation necrosis. *M. tuberculosis* cannot duplicate within these tubercles because of the low pH and anoxic environment. However, some *M. tuberculosis* can persist within these tubercles for a long period of time (Grosset, 2003).

1.5.4 Stage IV of *M. tuberculosis* infection

After the activation of macrophages, the tubercles will be surrounded by many activated macrophages and also inactivate macrophages. The persistence *M. tuberculosis* in tubercles will start to duplicate in the inactivate macrophage and hence the tubercle is enlarged. *M. tuberculosis* in the enlarged tubercle can invade the bronchus through hematogenic route. Eventually, *M. tuberculosis* disseminate to various parts of the body through hematogenic and lymphatic route. This infection is known as extrapulmonary TB. The secondary lesions caused by extrapulmonary TB can be categorized into:

a) Exudative lesions involve the accumulation of polymorphonuclear granulocytes (PMNs) around the tissue infected with *M. tuberculosis* which can propagate freely. The result of such situation is the formation of 'soft tubercle'.

b) Granulomatous or productive lesions take place when the host becomes hypersensitive to tuberculoproteins. The consequence of this is the formation of 'hard tubercle'.

1.5.5 Stage V of *M. tuberculosis* infection

The caseous center of the tubercles is enlarged and liquefied through an unknown process (Bentrup and Russell, 2001). The liquid from the surrounding tissues provides nutrient for the growth of *M. tuberculosis*. The antigens from *M. tuberculosis* bind to the walls of bronchi and cause necrotic and rupture. A cavity is formed at this stage, resulting in local or systemic TB disease.

In fact, only a very small percentage of *M. tuberculosis* infections cause the disease, and even a smaller percentage advance to a stage of being uncontrollable or

serious. For a person who has effective cell-mediated immunity, the granulomas heal, become fiborous and calcify from the lesions. This substance is called Ghon complex. Thus, recovery from the TB disease is probable.

A small number of *M. tuberculosis* may also calcify from metastatic foci. However, the escaped *M. tuberculosis* can survive in the foci and subsequently reactivate the *M. tuberculosis* over a certain period of time, which is known as latent infection (Pandey and Sassetti, 2008) (Figure 1.5).

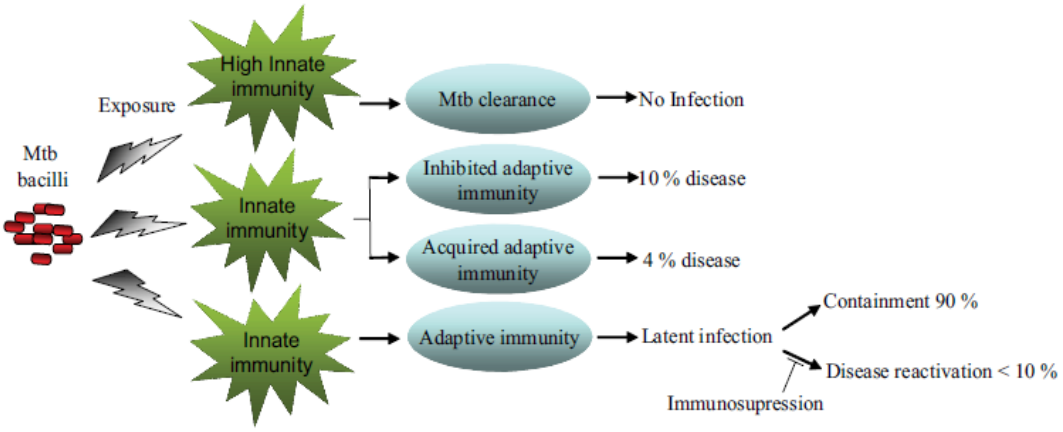

Figure 1.5 The possible outcomes of *M. tuberculosis* exposure (Adapted from Målen, 2007).

1.6 Diagnosis and Anti-TB therapy of the disease

1.6.1 Active TB

Active TB is the detection of *M. tuberculosis* complex bacilli from respiratory tract (pulmonary TB) specimens or specimens from extrapulmomary TB. The common and inexpensive diagnosis methods of active TB are Zielh-Neelsen (ZN) staining and culture on solid or liquid medium. Zielh-Neelsen (ZN) staining is simple, fast, inexpensive and very usable diagnosis method for the contagious patients. Due to the low sensitivity and specificity of ZN staining in low Mycobacteria loads specimen, the culture diagnosis methods are used to examine the specimen for the drug susceptibility

test and cross contamination of the Mycobacterium genus. Cultivation of *M. tuberculosis* required 6 weeks or longer on solid culture media (*e.g* the egg-based Lowenstein-Jensen medium or the agar-based Middlebrook 7H10 or 7H11) whereas observed growth of *M. tuberculosis* in liquid culture media takes 7-21 days only (*e.g* Middlebrook 7H9) (Morgan *et al.*, 1983). Both solid and liquid media are required, the former is to examine the colony morphology and the latter is for advanced molecular diagnostic assays. In addition, ZN staining and the culture diagnosis methods are also used in monitoring the effectiveness of the direct observed therapy strategy (DOTS).

In addition to active TB, a variety of new molecular diagnostic assays are developed for rapid diagnosis of TB and other Mycobacteria diseases:

a) Immunological tests

Immunological tests based on the interaction of antigen with the antibodies from the serum active TB patient. ELISA-based serological tests and immunochromatographic assays can easily be administered to TB patients in low resource countries as they are rapid and inexpensive (Chan *et al.*, 2000) ELISA–based serological test is used to detect specific TB antibodies which are bound to various purified or complex antigens of *M. tuberculosis* (Chan *et al.*, 2000; Gennaro, 2000; Silva *et al.*, 2003; Kanaujia *et al.*, 2005; Weldingh *et al.*, 2005). However, due to large variations observed in the interaction between antibodies from tuberculosis patients and Mycobacteria antigens, an improvement in sensitivity and specificity of ELISA- based serogical test is required (Lyashchenko *et al.*, 1998; Abebe *et al.*, 2007). Another immunochromatographic assay is known as lateral-flow tests or simple strip tests. The benefits of immunochromatographic tests include: 1) user-friendly; 2) fast result; 3) long-term stability over different climates; 4) relatively inexpensive.

b) Nucleic acid amplification assays

Nucleic acid amplification assays are polymerase chain reaction (PCR), transcription-mediated amplification (TMA) and strand displacement amplification (SDA). These techniques are used to detect nucleic acid of microorganisms and have high sensitivity (95 %) and specificity (98 %) result (Boman *et al.*, 1999). However, these assays may give false positive result in people who had TB previously; this is due to the high sensitivity of the assays that could detect the residual of genetic material from the previous infection. Nevertheless, these assays are not very practical in field study and expensive.

1.6.2 Latent tuberculosis

Tuberculin skin test is a delayed-type hypersensitivity test which comprises a cocktail of purified *M. tuberculosis* antigens, known as purified protein derivatives (PPD). It can detect active TB disease or TB infection. A person, who is infected with *M. tuberculosis* for a duration of 6-8 weeks may be presented with positive result of tuberculin skin test. An intradermal injection of PPD is administered and the induration period will be recorded after 48 to 72 hours. The effects of the local injection site within 24 to 72 hours are induration, swelling and monocytic infiltration. An induration is the diameter measurement of hard swelling site of tuberculin skin reaction. Enarson (2004) reported that a person who has induration tuberculin skin test greater than 10 mm may develop TB easily in the future. An induration result of 20 mm or more indicates that a person has active TB disease. However, a false positive result is shown when the purified protein derivatives interact with BCG vaccine and other common Mycobacteria in the environment (Iseman, 2007).

In the last decade, interferon-gamma (IFN- γ) assays suitable for detecting TB infection have been developed. *M. tuberculosis* antigens [(early secretory antigenic target-6 (ESAT-6) and culture filtarte protein-10 (CFP-10)] will respond to IFN-γ production by T-lymphocytes (Doherty *et al.*, 2002) and produce positive result. These tests are more specific compared to tuberculin skin test for latent TB (Brook *et al.*, 2001) as they are not affected by BCG vaccine (Lalvani *et al.*, 2001).

Currently, IFN- γ assays are commercially available. They are the enzyme-linked immunospot (ELISPOT) T SPOT-TB assay (Oxford Immunotec, Oxford, United Kingdom) and the QuantiFERON-TB Gold (Cellestis International, Carnegie, Australia). Both of these tests are approved by Food and Drug Administration (FDA) for detecting latent TB infection in Europe and USA (Mazurek *et al.*, 2003; Meier *et al.*, 2005). QuantiFERON-TB Gold test is a quantification test of IFN- γ released from T lymphocytes in blood. T SPOT-TB assay is a test that detects a number of T lymphocytes that produce IFN- γ in the blood samples upon TB infection (Mazurek *et al.*, 2003; Meier *et al.*, 2005). However, these tests are very costly and cannot be easily performed in resource-limited settings and developing countries such as Malaysia.

1.6.3 BCG vaccine and new vaccine candidates

A good vaccine is a prerequisite for the control, prevention and eventual elimination of tuberculosis. Bacillus of Calmette and Guérin (BCG, named after the two Frenchmen that developed it), is a live attenuated vaccine that is currently available.

The efficacy of BCG vaccination has shown to be highly variable across different populations (Haile and Källenius, 2005). BCG vaccination is widely used in children for prevention of primary TB infection such as military TB, but its effectiveness in

preventing adult pulmonary TB varies from 0 % to nearly 80 % (Aronson *et al.*, 2004; Young and Dye, 2006). Even though the effectiveness of BCG vaccination is varied, it has been continually used to protect severe forms of TB in children. Therefore, an improvement of the current BCG vaccine is urgently needed. Several factors need to be considered when generating a new version of BCG vaccine, *i.e.* it can be used for:

a) protection of pre-exposure and post-exposure to *M. tuberculosis* individuals;
b) latently infected and healthy individuals;
c) combination of immunotherapeutic agent with antimicrobials to booster the eradication *M. tuberculosis*;
d) minimizing the side effect of the individuals who had vaccinated.

In general, the new generation of vaccines should be compatible to BCG and HIV patients (Vuola *et al.*, 2003).

Recently, subunit and live vaccine approaches have been widely used to improve the TB vaccine. Subunit vaccines can deliver immunodominant Mycobacteria antigens such as ESAT-6 (6 kDa early secretory antigenic target). Live vaccines which focus mainly in the recombinant engineering of existing BCG strains are hope to improve expression of the immunodominant antigens and inactivate the virulence genes of *M. tuberculosis* (Britton and Palendira, 2003). Currently, several new generations of vaccines are being developed at the laboratory level.

Studies on multi-variant antigens in recombinant BCG (rBCG) against numerous pathogens have shown protective immune response in animal models (Winter *et al.*, 1995; Ohara and Yamada, 2001; Santangelo *et al.*, 2007). rBCG is an insertion of targeted DNA from one species into the DNA of another. rBCG in the cytokine genes (IL-2, IFN-

γ) is to improve the immune-stimulatory properties of BCG (Murray *et al.*, 1996) whereas the insertion of listeriolysin gene from *Listeria monocytogenes* to the rBCG is to present the microbial antigen to the host cytoplasm in order to enhance the production of $CD8^+$ (Hess *et al.*, 1998; Eddine and Kaufmann, 2005). A vaccine developed from heterologous prime-boost immunization strategies is being studied as an alternative way to boost the immunity. Furthermore, acellular vaccines (virus vaccine) have shown to give a better protection than BCG in preclinical testing. For example, a recombinant modified vaccinia virus that expresses *M. tuberculosis* antigen 85A had been investigated to be a booster of BCG vaccine. This vaccine was named Ankara vaccine (McShane and Hill, 2005; McShane *et al.*, 2004).

In view of the fact that the majority of the targeted vaccine candidates are made up of virulence proteins, therefore proteomic investigation in *M. tuberculosis* for identification of virulence proteins in TB is the upmost important.

1.7 *M. tuberculosis* H37Rv and H37Ra

M. tuberculosis H37 is a clinical strain that was first isolated by Edward R. Baldwin from a 19 year-old male patient with chronic pulmonary TB in 1905 (Steenken and Gardner, 1946). Both *M. tuberculosis* H37Ra and *M. tuberculosis* H37Rv were derived from the same virulent parent strain *M. tuberculosis* H37 through a process of aging and dissociation *in vitro* culture as demonstrated in Figure 1.6 (Steenken, 1935). The parental virulent *M. tuberculosis* H37 was inoculated onto solid egg media and incubated for 3-4 months at 37 °C until the culture dried, discrete colonies lysed and transformed into a confluent viscous mass. Then a secondary growth of *M. tuberculosis* emerged with different colony morphology from the midst of the viscous mass. The new growth was isolated into fresh media and infected into guinea pigs, but the guinea pigs

showed no symptom of TB infection (Steenken, 1935; Steenken, 1938). *M. tuberculosis* H37Ra ('a' for avirulent) also has a decrease ability to survive in macrophages or anaerobic conditions as compared to its parent strain. Therefore, the laboratory strain has lost its virulence, and has some different phenotypes from its virulent sibling *M. tuberculosis* H37Rv [('v' for virulent) (Steenken and Gardner, 1946)]. Consequently, *M. tuberculosis* H37Ra and *M. tuberculosis* H37Rv have been widely used as laboratory reference strains for studying and understanding the virulence and pathogenesis of *M. tuberculosis* worldwide (Zheng *et al.*, 2008). Therefore, a comparison of proteomes between MTB H37Ra and H37Rv would shed light in the understanding of the pathophysiology of both strains and also in the understanding of protein that could be involved in the avirulent pathway.

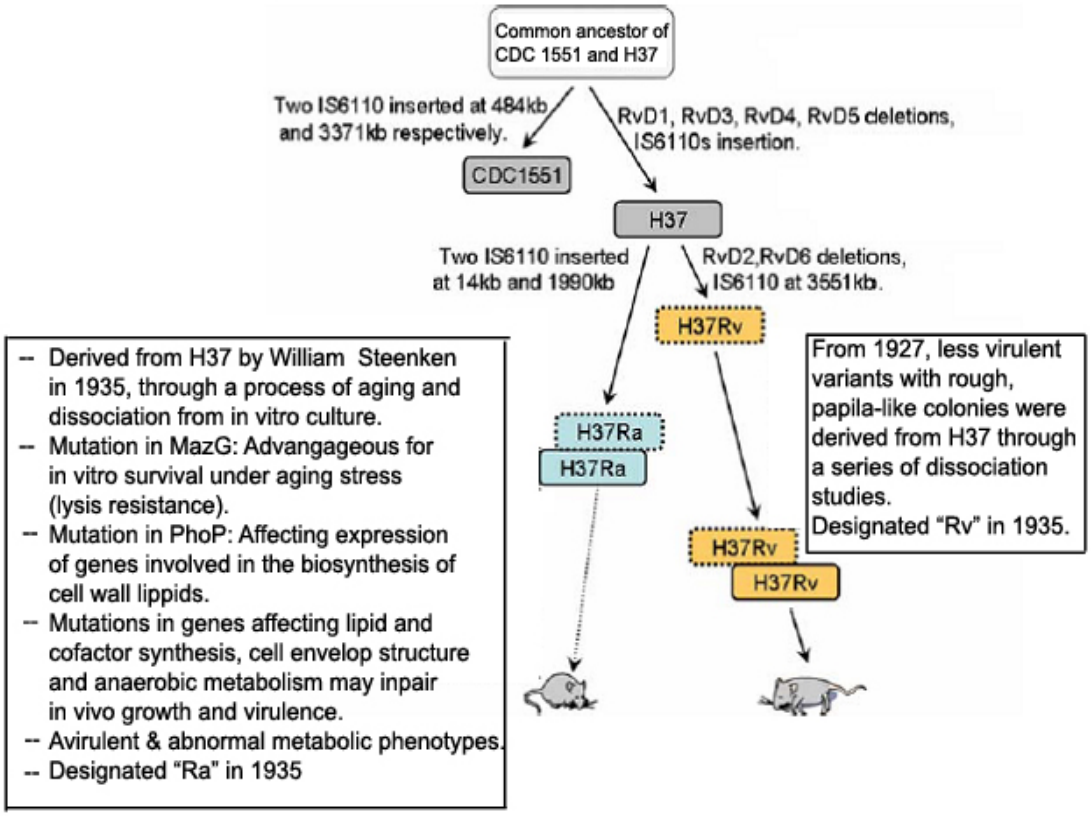

Figure 1.6 The process of *M. tuberculosis* strain evolution (Adapted from Zheng *et al.*, 2008).

1.8 The complete genome of Mycobacteria

The availability of complete genome sequence of *M. tuberculosis* H37Rv in 1998 was a big step forward for TB research (Cole *et al.*, 1998). Subsequently, complete genome sequences of the clinical isolates, *M. tuberculosis* CDC1551 (Fleischmann *et al.*, 2002) and *M. bovis* AF2122/97 (Garnier *et al.*, 2003) were also being established. Recently, the genome sequence of *M. tuberculosis* H37Ra has also been reported (Zheng *et al.*, 2008) and the database of *M. tuberculosis* H37Ra could be obtained from http://cmr.jcvi.org/cgi-bin/CMR/GenomePage. The complete genome of *M. tuberculosis* H37Rv comprises of 4,411,529 base pairs with approximately 4,000 genes, accounting to 91 % of the potential coding capacity (Brosch *et al.*, 1998; Cole *et al.*, 1998). The genomic deoxyribonucleic acid (DNA) consists of 65.5 % of guanine (G) and cytosine (C) (Cole *et al.*, 1998).

More than 20 % of *M. tuberculosis* chromosome consists of fatty acid metabolism; acidic, glycine-rich polypeptides; proline-glutamic acid (PE) and proline-proline-glutamic acid (PPE) proteins family. Forty percent of the *M. tuberculosis* proteome consisting of proteins with known functions were mainly responsible for its metabolism pathway. The repetitive DNA (~ 51 %) of *M. tuberculosis* is due to recent evolution descent or extensive functional redundancy of the organism (Cole *et al.*, 1998).

In addition, some genomic regions of TB have been detected as deletions in both virulent Mycobacteria and attenuated Mycobacteria. By using *Dra*I digestion, it was found that a 480 kb fragment in *M. tuberculosis* H37Rv is replaced by two fragments of 220 and 260 kb in *M. tuberculosis* H37Ra. On the other hand, the 7.9 kb fragment was detected in *M. tuberculosis* H37Ra but not in *M. tuberculosis* H37Rv (Cole *et al*., 1994; Brosch *et al.*, 1999). Several studies were focused in different genomic regions of

virulent *M. tuberculosis*, *M. bovis* and avirulent BCG (Mahairas *et al.*, 1996; Brosch *et al.*, 1998; Behr *et al.*, 1999), some deletion regions have been identified but the role of deleted genes remains unclear.

1.9 Proteomic study

Proteomics is a study of all proteins in a cell that are translated in a given physiological condition. The proteome of an organism refers not only to characterization of proteins but also the translation rate and the posttranscriptional events that occur in the organism (Betts, 2002). A proteomic study in pathogenic bacteria may result in the discovery of potential biomarkers for the identification and monitoring of the disease progression.

1.9.1 Current strategy in proteomic study

The combination of two dimensional electrophoresis (2-DE) and liquid chromatography in tandem with mass spectrometry (LC-MS/MS) is widely used in studying proteomic. Electrophoresis is the movement of charged particles or molecules under the influence of an electric field. Two dimensional electrophoresis (2-DE) comprises of two protein separation modes, namely isoelectric focusing for separation of proteins according to isoelectric points (p*I*) and then SDS-PAGE for separation of proteins by molecular weights (Nelson and Cox, 2005; Weiss and Görg, 2009). Hence, the information obtained from the 2-DE is not only on molecular size but also the charge of protein. Two dimensional electrophoresis is very useful in resolving the complex mixtures of protein samples. One has to be aware that 2-DE separated proteins with differential electrophoretic mobilities may not solely represent the proteins encoded by different open reading frames (ORFs); they may also be identified as proteins with similar ORF but with a diversification of a primary translation product by co- and

posttranslational protein modifications. The early 2-DE on Mycobacteria has led to detection of 50-300 protein spots (Britton *et al.*, 1987; Daugelat *et al.*, 1992; Wallis *et al.*, 1993; Lee and Horwitz, 1995; Mahairas *et al.*, 1996; Rosenkrands *et al.*, 2000b). Currently, the use of chemical modification and immobilized pH gradients in the first dimension of 2-DE has increased gel resolution that lead to the detection of 2000 protein spots (Urquhart *et al.*, 1997; Sonnenberg and Belisle, 1997; Celis *et al.*, 2006; Wittmann-Liebold *et al.*, 2006).

The targeted protein spot can be subjected to in-gel enzyme digestion and the resulting peptides can be identified using LC-MS/MS. Reverse-phase liquid chromatography (LC) is commonly applied in the analysis of proteins and peptides (Hancock *et al.*, 1978; Davis and Lee, 1992; Ausubel *et al.*, 2002; Simpson, 2003). The principle of reverse phase LC is based on the use of stationary phase that is non-polar material [silica linked with carbon-18 (C18)] and the mobile phase that is polar solvent (water, methanol, acetonitrile and IPA). Reverse-phase LC aids in the removal of contaminants that might interfere with MS/MS analysis (Shaw, 1994). Basically, the separation of peptides in reverse-phase LC is based on the strength of hydrophobic interaction between the amino acid side chains of the peptides and the stationary phase (Stone and Williams, 2002). The elution of binding peptides is based on the gradually increasing strength of the organic solvent over a period of time thus resulting in concentrated and purified peptides. In the reverse phase LC, the non-polar peptides elute slower than the polar compounds in the column (Simpson, 2003).

Electrospray ionization mass spectrometry (ESI-MS) is one of the methods to accomplish mass determination of biomolecules (Figure 1.7). The principle of electrospray ionization mass spectrometry is the use of soft ionization techniques in order to determine the mass of biomolecules. ESI-MS consists of 4 main components:

1) ionization source;

2) ion trap mass analyzer;

3) ion detector and

4) data processor.

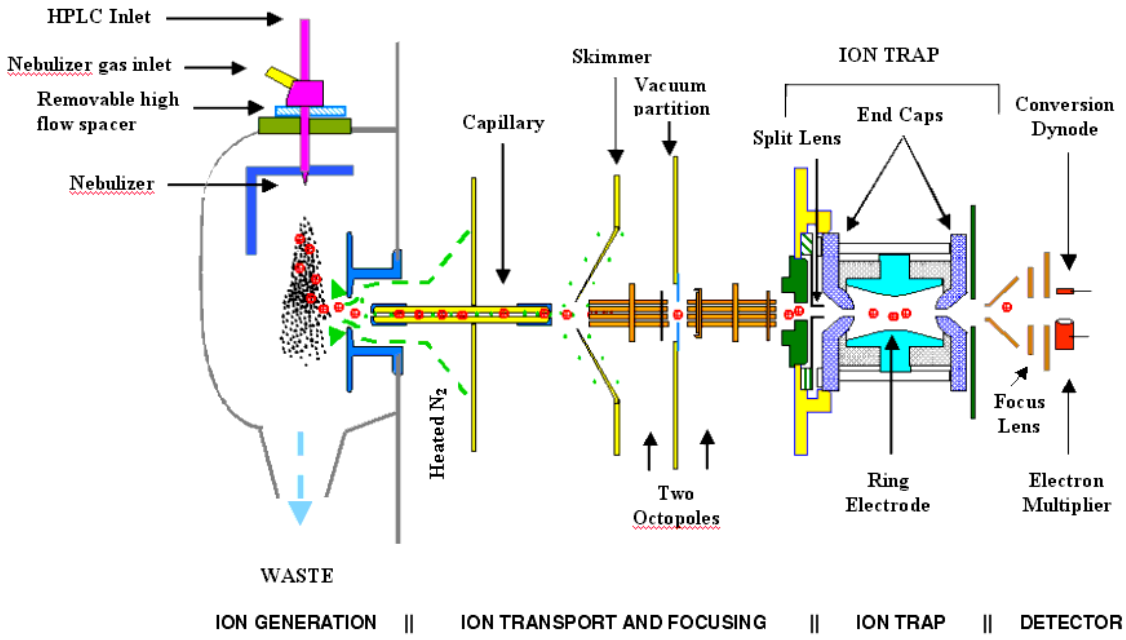

Figure 1.7 The structure of ESI-MS (Reproduced with permission from Agilent Technologies).

A solution of sample and solvent is sprayed through a capillary with a low flow rate (0.1-10 μL/min) by applying a high voltage (2-6 kV). This high-voltage could increase the charge density at the surface of the droplets and the solvent is evaporated by heated drying gas. Therefore, the ions are formed as a result of coulombic repulsion in a high-voltage electrostatic gradient (Iribarne and Thomson, 1976). The difference in

pressure between the ionization source and the ion transport and focusing region pushes the ions through glass capillary. Subsequently, the ions are focused as through ion trap mass analyzer to ion detector.

Within the ion trap mass analyzer, the ions are collected and stored for further MS/MS analysis by controlling the radio frequency (RF) devices. Helium gas is used for efficient trapping and cooling of the ions generated by the electrospray interface. The helium pressure in the trap is set by mechanical pressure regulator in order to control the helium gas amount. In full scan mode, the RF voltage on the electrodes is changed to sequentially eject the ions based on their *m/z* values. This produces a spectrum that represents all peptide ions in the trap. In MS/MS mode, the targeted precursor ions are trapped in the ion trap and ejected out all the unwanted ions by manipulating the radio frequency. During isolation of target precursor ion, the electronics system on the ion trap generates a broadband frequency spectrum with target mass resonating frequency. The target precursor ion then undergoes induction fragmentation when the frequency increases resulting in collision induction dissociation of the target ions to forming product ions, which are scanned according to their *m/z* values (Liebler, 2002).

1.9.2 Proteomic study of pathogenic and non-pathogenic bacteria

Comparison proteomic studies of pathogenic and non-pathogenic bacteria are widely used in order to detect vaccine candidates or biomarkers for diagnosis and treatment of the disease. In general, proteomics studies are involved separation of proteins in one- or two- dimensional polyacrylamide gel electrophoresis (Nagai *et al.*, 1991; Garbe *et al.*, 1996) followed by analysis of proteins using tandem mass spectrometry. It is a widely used method to allow fast overview of changes in cell processes by analysis of cell's proteome. To date, the development of high-resolution 2-

DE systems and highly sensitive mass spectrometric techniques has led to high throughput screening of pathogen's functional genomics in different cell conditions.

The early comparison studies of virulent and avirulent *Mycoplasma pneumoniae* isolates were carried out by (Hansen *et al.,* 1979) and (Hansen *et al.,* 1981). The former study identified three novel proteins expressed by the *Mycoplasma pneumoniae* M129 virulent isolates. The latter was the comparison study of a virulent parental strain of *M. pneumonia* M129 with two derived avirulent chemical mutant strains. The study reported the differential expressed protein at quantitative and qualitative levels when analyzed by 2-DE. However, no data was provided in the identities of these proteins.

A similar approach was used to compare *Brucella. abortus* virulent strain 2308, avirulent vaccine strain S19 and LPS deficient strains by using 2-DE (Sowa *et al.,* 1992). Approximately 935 proteins were resolved by 2-DE although 2129 proteins were expected based on the size of the 3.13×10^{6}-bp *B. abortus* genome (Allardet-Servent *et al.,* 1991). The comparison of the 2-DE protein profiles of the virulent and avirulent strains has resulted in the discovery of 86 differentially expressed proteins and 6 unique proteins.

Vibrio vulnificus is a food borne pathogen. The virulence-related proteins of this bacterium were ToxR and ToxS. In order to further understanding the pathophysilogy pathway, Lee and his co-workers compared the proteome of wild-type *Vibrio vulnificus* and *toxR* mutant of *V. vulnificus* (Lee *et al.,* 2006). They found that 12 proteins were either up- or down-regulated in the *toxR* mutant and these proteins may regulate the expression of toxR.

In another study, the proteomes of *Francisella tularensis* virulent subspecies *tularensis* strain SCHU S4 and subspecies *holarctica* strain were compared. Pavkova *et*