View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by ScholarBank@NUS

A CHEMICAL GENETICS APPROACH TO IDENTIFY TARGETS ESSENTIAL FOR THE VIABILITY OF MYCOBACTERIA

STEPHEN HSUEH-JENG LU

NATIONAL UNIVERSITY OF SINGAPORE

AND

THE UNIVERSITY OF BASEL

2007

A CHEMICAL GENETICS APPROACH TO IDENTIFY TARGETS ESSENTIAL FOR THE VIABILITY OF MYCOBACTERIA

STEPHEN HSUEH-JENG LU

(B.Sc. (Hons.), University of Auckland)

A THESIS SUBMITTED FOR THE DEGREE OF MASTER OF SCIENCE

IN

INFECTION BIOLOGY AND EPIDEMIOLOGY:

INFECTIOUS DISEASES, VACCINOLOGY AND DRUG DISCOVERY

DEPARTMENT OF MICROBIOLOGY

NATIONAL UNIVERSITY OF SINGAPORE

AND

THE UNIVERSITY OF BASEL

2007

Acknowledgements

I want to take this opportunity to acknowledge the generous financial support of the William Georgetti Scholarship from the New Zealand Vice-Chancellors Committee (NZVCC, Wellington, New Zealand).

I would like to thank my supervisors, Dr. Vasan Sambandamurthy and Dr. Thomas Dick, for the opportunity to work in the Tuberculosis Unit of the Novartis Institute for Tropical Diseases (NITD). I would also like to thank them for their full support throughout the duration of the course.

I thank all the members of the Tuberculosis Unit: particularly Srini, Mekonnen and in alphabetical order, Amelia, Angelyn, Bee Huat, Boon Heng Cui Feng, Florence, Karen, Kevin, Lay Har, Luis, Mahesh, Martin, Kai Leng, Kevin, Pamela, Penny, Sabai, Siew Siew, Sindhu, Wei Fun (plus members that have left recently Sabine and Kakoli) for their guidance, assistance, friendship and for being great lab members. In addition, I appreciate the support/encouragement of Dr. Thomas Keller and for his friendliness.

I also appreciate the support and kindness of Mark, David, Viral, Xinyi, Sarah, Joanne, Kim and Cindy of NITD. I also acknowledge the encouragement and assistance of the lecturers/professors of Universtät Basel and the National University of Singapore, particularly Dr. Markus Wenk and Professor Marcel Tanner.

I would like to thank my two flat-mates, Lukas and Tommy, for making this stay in Singapore memorable. Importantly, I would also like to thank Yunshan, Cheryl, Adrian

and Pei-Ying for being great classmates and hanging together in Switzerland and Singapore. I have many more people that I would like to thank but you know who you are and I appreciate everything that you guys have done for and with me in the last 18 months.

Finally yet most importantly, I would greatly appreciate the love of my family and friends (especially those that took the time to visit me from Germany, New Zealand, Hong Kong and Malaysia) – i.e. Kenny, Matthias, Claudia, Shannon, Joanna, Janice, Tania, Jo-Ann and Katie. My friends in Singapore, particularly Susie, Yunshan, Audrey, Joanne, Crystal, Ingrid and several others – you guys have been great as well. Thanks Mum and Dad – you guys are the best! And thanks Huimin for just being who you are.

Intellectual Property (IP) Statement

In compliance with the IP policies of Novartis, we are unable to display the chemical structure of compounds as well as their compound names used in this study. Instead, we have replaced the names of the two compounds used in this study as Compound X (the compound isolated from the initial screen – CpdX) and Compound Y (the structure derivative of CpdX taken from the Novartis compound library - CpdY).

Table of Contents

Chapter 1: Introduction

Chapter 2: Materials and Methods

Chapter 3: Results

Summary

The key goals for the development of new anti-mycobacterial drugs are to shorten the treatment time and to have efficacy against latent as well as multi-drug resistant tuberculosis. The best drug targets should be essential in both active and dormant phases of the *Mycobacterium tuberculosis* infection, so that a single drug would eradicate both populations. The only feasible way to elucidate such a novel target is to use a forward chemical genetics approach. Forward chemical genetics involves screening a library of compounds against the entire proteome for novel targets whose inhibition by one of the compounds results in bacterial death or growth inhibition. The candidate drug/target pair can be identified by microarray fingerprinting (Boshoff *et al*., 2004), proteomic profile comparison as well as whole genome sequencing of spontaneous resistant mutants (Andries *et al*., 2005).

We isolated *M. bovis* BCG mutants resistant to two structurally-related compounds, named compound X and compound Y. The magnesium and cobalt transport transmembrane protein, CorA, was identified as a putative target of these two compounds. This was based on the mapping of genetic mutations to the *corA* gene from the compound-resistant mutant strains. Moreover, the exogenous expression of the mutant copy of *corA* gene in wild-type mycobacteria conferred high levels of resistance to these two compounds. However, due to the non-essentiality of the *corA* gene and the bactericidal effect of the compounds, we suggest that CorA is not the actual target and that it mediates an indirect mechanism of resistance. More experiments are needed to identify and validate the biological target of compound X and compound Y.

List of Tables

List of Figures

List of Abbreviations

2D	Two-dimensional
AIDS	Acquired Immunodeficiency Syndrome
BCG	bacille Calmette-Guérin
CDC	Centers for Disease Control and Prevention
CFUs	Colony Forming Units
CpdX	Compound X
CpdY	Compound Y
DOTS	Directly Observed Treatment Short Course
HIV	Human Immunodeficiency Virus
HTS	High-Throughput Screening
LC-MS	Liquid Chromatography-Mass Spectrometry
MBC	Minimum Bactericidal Concentration
MDR-TB	Multi-Drug Resistant Tuberculosis
$[Mg^{2+}]$	Concentration of Magnesium
MIC	Minimum Inhibitory Concentration
OD_{600}	Optical Density at a wavelength of 600nm
PCR	Polymerase Chain Reaction
PPD	Purified Protein Derivative
SNPs	Single Nucleotide Polymorphisms
WHO	World Health Organization
XDR-TB	eXtensively-Drug Resistant Tuberculosis

Chapter One: Introduction

1. Introduction

The first part (Section 1.1 to Section 1.7) of this chapter will cover the clinical, epidemiological and scientific information about the disease tuberculosis and its causative agent. In the later parts, the use of chemical genetics to identify a compound that targets a novel biological target and the tools available for identifying such a target is presented (Section 1.8 to Section 1.11).

1.1. Tuberculosis

Tuberculosis is a contagious disease caused mainly by *Mycobacterium tuberculosis* and sometimes *M. bovis*. *M. tuberculosis* is a relatively large, non-motile, rod-shaped, acid-fast bacillus belonging to the family of actinomycetes (Parish and Stroker, 1998). *M. tuberculosis* is an obligate intracellular parasite. In the laboratory, *M. tuberculosis* can be grown on the agar-based Middlebrook medium or the egg-based Lowenstein-Jensen medium (Parish and Stroker, 1998). Since it has a slow generation time of around 18-20 hours, it takes 3-6 weeks for the bacteria to form visible colonies on these solid media (Parish and Stroker, 1998).

1.2. Epidemiology

In the 19th century in Europe, tuberculosis, then known as the *White Plague,* was responsible for 30% of total mortality (Merck, 2003). The World Health Organization (WHO) estimates that globally two billion people are latently infected with tuberculosis and that nine million people develop active disease annually, of which two million die each year (Dye *et al*., 2006; Dye *et al*., 1999). *M. tuberculosis* infections have reemerged as a major public health problem around the globe because

of poverty, neglect of the disease in the developed world, poor health services during a crisis, migration of people from endemic countries, multi-drug resistant tuberculosis (MDR-TB), and, lastly, due to HIV co-infection (Dye *et al*., 2006; Grange and Zumla, 1999; Manganelli *et al*., 2004; Espinal *et al*., 2001; Raviglione *et al*., 1997).

The WHO estimates that 90% of the tuberculosis cases occur in the developing world (50% of those in the sub-Saharan desert region (Zumla *et al*., 2000; WHO, 2004)) where the disease predominantly affects the 15-54 years age group. This has a severe economic impact on the patient's family and his/her community. Tuberculosis disease results in the loss of, on average, 20-30% of annual income or 15 years of income if the disease results in death (WHO, 2004; Ahlburg, 2000). In contrast, in the developed world the disease often occurs in the elderly or immunodeficient individuals.

Tuberculosis is classified as MDR-TB when the bacilli are resistant to at least the two front-line drugs rifampicin and isoniazid. MDR-TB is on the rise in many parts of the world, especially in the former Soviet Union (Espinal *et al*., 2001). In 2006, the outbreak of a virulent XDR-tuberculosis (eXtensively Drug-Resistant; i.e. MDR-TB that is additionally resistant to three or more second-line drugs) strain in South Africa where 52 out of 53 patients died within one month of diagnosis further highlights the importance of this disease (Associated Press, 2006; Centers for Disease Control and Prevention, 2006).

1.3. Biology of tuberculosis

1.3.1. Immunology of tuberculosis

Infection of humans *via* the aerosol route with *M. tuberculosis* will result in latent or, sometimes, active tuberculosis. Clinically, the primary infection could be controlled entirely by the innate immune system and the infected individuals remain asymptomatic (Grosset, 2003). However, the infection can also progress to a latent state with a 10% lifetime chance of reactivation to active disease (Gedde-Dahl, 1952; Grosset, 2003) (see Section 1.3.5. Persistence and Latent Tuberculosis).

1.3.2. Clearing of primary infection by the immune system

Most commonly, *M. tuberculosis* spreads when an infected patient expels small droplets that contain the bacilli during sneezing, coughing or talking (Merck, 2005). The most infective droplet is around 1-3μm in diameter (large droplets do not remain airborne for long and do not reach the alveoli to establish an infection) and contains up to three bacilli (Riley, 1974; Riley *et al*., 1962; Grosset, 2003). One to three weeks following infection, the bacilli multiply exponentially in the macrophages, which at this stage are not activated hence cannot effectively kill the mycobacteria (McDonough *et al*., 1993; it should be noted that dendritic cells can also phagocytose mycobacteria (Bodnar *et al*., 2001)). Subsequently, cellular immunity becomes activated when infiltrating $CD4^+$ T-lymphocytes recognize the *M. tuberculosis* antigens presented on MHC (Major Histocompatibility Complex) molecules of antigen presenting cells and release cytokines (such as γ-interferon) that activate the macrophages (Grosset, 2003). In addition, $CD8^+$ T-lymphocytes can also eliminate the infected macrophages (Houben *et al*., 2006). Humoral immunity is ineffective against mycobacteria because

the bacilli are intracellular and, even when the bacilli are in extracellular spaces, the thick cell wall prevents complement-mediated antibody killing.

At this stage, granulomas start to form at the foci of infection (see Section 1.3.2. Granuloma Formation and Caseous Necrosis; Grosset, 2003). Some bacilli may survive in this region of caseous necrosis for years (see Section 1.3.5. Persistence and Latent Tuberculosis). Occasionally, the bacilli may spread to other parts of the lung or to any part of the body via the bloodstream (see Section 1.3.5. Extrapulmonary Tuberculosis; Merck, 2005).

It should be noted that in an immunocompetent host, the infection does not always result in active disease (only 10% develop tuberculosis) (Enarson and Rouillon, 1994). The lesions heal to form either the fibrous, calcified Ghon complex (in the primary foci), nodular Simon foci (in other smaller foci) or calcified lymph nodes (Merck, 2005).

1.3.3. Granuloma formation and caseous necrosis

Tuberculous granulomas are a special type of lesion associated with tuberculosis disease. It consists of caseous necrosis in the center of the lesion surrounded by giant multinucleated Langhan's cells, epitheloid cells (activated macrophages), lymphocytes and fibroblasts (Canetti, 1955; Opie and Aronson, 1927; Adams, 1976; Bouley *et al*., 2001; Grosset, 2003; Cosma *et al*., 2003). Caseation (derived from the word *caseum* which means cheese) is a typical type of amorphous necrotic lesion that is associated with tuberculosis (Canetti, 1955; Opie and Aronson, 1927; Grosset, 2003). As these lesions develop, there is a huge reduction in the bacillary load within these lesions. In

the old caseous foci, there are very little, if any viable bacilli (Canetti, 1955; Opie and Aronson, 1927; Grosset, 2003).

Caseous necrosis results from the infiltration of activated cytotoxic T-lymphocytes that kill macrophages (or its derivatives the Langhan's giant cells and epitheloid cells) infected with *M. tuberculosis* as part of a necessary process to control the unimpeded bacillary replication (Grosset, 2003). This immunological activity damages the host tissue, but at the same time destroys a majority of bacteria. However, the bacilli do survive extracellularly but cannot replicate because of low oxygen tension, acidic environment within the caseous foci (Grosset, 2003). These physiological conditions may prompt the tubercle bacilli to enter a state of non-replicating persistence (or dormancy). This population of non-replicators is believed to be responsible for the long treatment period of over six months (see Section 1.6. Prevention, current treatment, DOTS and drug resistance).

In up to 90% of infected individuals, the T-cell activated macrophages will form the granulomas and eventually eliminate most of the bacteria. Sometimes, the caseation softens, spreads into the bronchial tree, and forms a lung cavity (see Section 1.3.4. Lung cavity) where the bacilli multiply exponentially following exposure to high oxygen levels (Canetti, 1955; Enarson and Rouillon, 1994). The softening of the caseum into the large airways of the lungs progresses asymptomatic *M. tuberculosis* infection into active tuberculosis disease.

1.3.4. Lung cavity

Lung cavities are the result of the softening of the caseum that is released into the bronchial tree, which in turn allows the bacilli to grow extracellularly due to the oxygen-rich environment (Long, 1935). A patient becomes infectious at this stage of the disease when thousands of bacilli in the lung cavity are released as small droplets during coughing or sneezing. Before the advent of antibiotics, around a quarter of all immunocompetent patients with lung cavities control the disease *via* cell-mediated immunity (see Section 1.3.3. Granuloma formation and caseous necrosis; Enarson and Rouillon, 1994). Half of the untreated patients that develop cavitary tuberculosis die within the first two years because bacilli released from the lung cavity will form new granulomas and in due course destroy the entire lung (Enarson and Rouillon, 1994). The remaining 25% of patients that do not receive any treatment will develop a chronic tuberculosis infection.

1.3.5. Extrapulmonary tuberculosis

Infection of tuberculosis outside the lung can also occur, due to the spread of bacilli either by the bloodstream or uncontrolled infection in the lung that spreads to nearby organs (Merck, 2005). Miliary tuberculosis is a severe form of tuberculosis where there is a widespread dissemination of bacteria throughout the body, presumably when the infection destroys the blood vessel walls thus releasing bacilli into the bloodstream (Merck, 2005). This form of tuberculosis is often fatal if left untreated. Tubercle bacilli can also infect the peritoneum, genitourinary system, pericardium, lymph nodes, bones, joints, gastrointestinal system, liver and meninges with varying degrees of severity and clinical outcomes (Merck, 2005). For example, tuberculous meningitis is associated with high morbidity and mortality in young children (Merck, 2005).

1.3.6. Persistence, dormancy and latent tuberculosis

Latent tuberculosis is a clinical condition where an individual is not sick with active tuberculosis but is purified protein derivative (PPD)-positive (see Section 1.5. Diagnosis of active and latent tuberculosis). As early as 1952, Gedde-Dahl described the latency phenomenon and this observation was confirmed by fact that a patient developed tuberculosis after 33 years of latent infection (Lillebaek *et al*., 2002). The World Health Organization currently estimates that two billion people are latently infected with *M. tuberculosis*, with the vast majority showing no symptoms or disease (Dye *et al*., 1999). Despite this, there is still an ongoing discussion as to the true nature of latency and how low bacillary numbers are maintained for many years. Specifically it is uncertain if the *M. tuberculosis* enters a dormant state, or that a fine balance between replication of the bacilli and its elimination by the host immune system is achieved (Parrish *et al*., 1998 Cosma *et al*., 2003).

1.3.6.1. The nature of persistence (dormancy) of mycobacteria

There is a theory among researchers that latent tuberculosis is the result of the bacteria lowering their metabolism and entering into a non-replicative state (also known as dormancy). This theory supported by several pieces of compelling evidence. One, no bacteria could be cultured when infected tissues containing acid-fast bacilli were used as the inoculum (Manabe and Bishai, 2000, McKinney, 2000, Parrish *et al*., 1998 and Cosma *et al*., 2003). It should be noted that the tissues were isolated from tuberculosis patients undergoing treatment and it is not unconceivable that drug-treated *M. tuberculosis* do not grow well under *in vitro* conditions. Two, despite sufficient penetration of antibiotics in the diseased tissue, it is well known that under *in vitro* conditions the bacilli are killed quickly, yet in human beings, a long course of therapy

is required (McKinney, 2000; Mitchison, 1979; Cosma *et al*., 2003). Nevertheless, the bacilli may have divide occasionally during the latent stage because isoniazid (a known cell-wall synthesis inhibitor; see Table 1.1) can decrease the risk of tuberculosis reactivation in PPD-positive patients (Comstock *et al*., 1979; Cosma *et al*., 2003).

1.3.6.2. Hypoxia-induced non-replicating persistence in *M. tuberculosis* – the Wayne Model

Many *in vitro* models have been developed to induce non-replicating persistence in *M. tuberculosis*. These include models based on altering the pH, nutrient starvation, hypoxia and nitric oxide levels (Dickinson and Mitchison, 1981, Heifets and Lindholm-Levy, 1992, Betts *et al*., 2002, Nathan and Shiloh, 2000; Wayne and Sohaskey, 2001). In particular, the hypoxia-induced Wayne model has been studied and established in great detail (Wayne and Sohaskey, 2001). Although oxygen levels within granulomas have not been determined, there is ample evidence to suggest that latent tuberculosis is the result of the survival of *M. tuberculosis* in an oxygen-deficient environment. Infection and reactivation is most commonly associated with the superior lobes of the lung where the oxygen tension is higher (Adler and Rose, 1996). Moreover, non-dividing bacilli can remain viable without oxygen in a presumably dormant state for several years (Canetti, 1955; Corper and Cohn, 1933).

In the Wayne model, mycobacteria are grown in sealed glass tubes with magnetic stirrers under a defined head space ratio of 0.5. The constant stirring ensures a uniform distribution of cells throughout the Dubos medium in such a way that the oxygen in the head space is slowly depleted to create a three-stage growth curve. The bacterial growth can be monitored by measuring the absorbance of the cultures periodically.

The bacteria grow exponentially for the first 5 days due to the presence of dissolved oxygen in the medium. The shift into the first stage of non-replicating persistence phase 1 (NRP1) occurs as the dissolved oxygen levels reach around 1%. NRP1 is characterized by a slight increase in turbidity without a concomitant increase in the bacterial colony forming units (CFUs). At day 10, the methylene blue indicator starts to fade and the decolorization is complete by day 12. This stage of non-replicating persistence phase 2 (NRP2) occurs when the oxygen level reaches 0.06% of normal saturation (anaerobic) and no further increase in optical density is seen (Wayne and Hayes, 1996; Wayne and Sohaskey, 2001). The bacteria from NRP2 show phenotypic isoniazid resistance.

1.3.6.3. Evidence of bacilli in the tissues of healthy PPD-positive individuals

These latent bacilli have been found in humans albeit in very low numbers. A few CFUs were isolated from pathologically normal lungs of people who had died from unrelated causes (Canetti, 1955; Feldman and Baggenstoss, 1938; Opie and Aronson, 1927; Grosset, 2003). Using *in-situ* PCR, *M. tuberculosis* DNA was found in the lung samples of individuals from Ethiopia and Mexico; both countries have a high prevalence of tuberculosis (Hernandez-Pando *et al*., 2000). Moreover, as noted earlier, isoniazid significantly reduced the risk of disease in PPD-positive people suggesting that there must be viable bacilli within the lung tissue (Comstock *et al*., 1979; IUAT, 1982). Moreover, hypoxic bacilli within tuberculosis lesions were imaged *in vivo* using pimonidazole probing (Barry *et al*., Tanzania, 2005).

1.3.6.4. Use of *in vitro* models of dormancy in drug discovery

The current drugs affect processes essential during replication and active metabolism, such as protein and cell wall synthesis, but show little or no activity against quiescent bacilli (see Section 1.8. Essential drug targets in *M. tuberculosis*; Dick, 2001; Boshoff and Barry, 2005). This resistance phenomenon observed with dormant bacteria is known as phenotypic, as opposed to genetic, drug resistance. Hence, one of the central goals of modern anti-mycobacterial drug discovery is the identification of compounds that can target both actively replicating as well as the non-replicating mycobacteria. By employing *in vitro* screens, it is possible to isolate potent compounds by determining their inhibitory/bactericidal concentrations on mycobacteria growing under aerobic and hypoxic conditions.

1.4. Symptoms of pulmonary tuberculosis

Symptoms of active pulmonary tuberculosis include constant coughing, tachycardia, chest pain and swollen lymph nodes in the neck (Merck, 2006). It should be noted that, contrary to common belief, bloody sputum is rarely seen in patients. Fatigue, fever, loss of appetite, chills and night sweats can also occur in many patients (Merck, 2006). However, symptoms of pulmonary tuberculosis are mild and usually develop gradually, thus may easily go unnoticed by patients. In extrapulmonary tuberculosis, the symptoms are more variable and dependent on the type of infected tissue, for example back pain and neurological defects could be symptoms of spinal tuberculosis (Merck, 2006).

1.5. Diagnosis of active and latent tuberculosis

The current gold standard for diagnosis of active tuberculosis is to microscopically determine the presence of acid-fact bacilli in patient's sputum samples following isolation of *M. tuberculosis* in culture (Merck, 2005; Nahid *et al*., 2006). Other useful tuberculosis diagnostic tools include chest X-ray, tuberculin skin test and IFNγ-based assays (Merck, 2005; Nahid *et al*., 2006).

Until recently, the tuberculin or Mantoux test, where purified protein derivative (PPD) of *M. tuberculosis* is injected intradermally, was the only method available for diagnosis of latent tuberculosis (Merck, 2005; Nahid *et al*., 2006). However, these results can be confounded by prior BCG vaccination and exposure to non-tuberculous environmental mycobacteria (Nahid *et al*., 2006; Pai, 2005). These problems led to the development of γ-interferon based assays (e.g. QuantiFERON-TB Gold test (Cellestis Ltd., Australia)) using two region of difference (RD-1) antigens, namely 6-kDa early secreted antigenic target (ESAT-6) and 10-kDa culture filtrate protein (CFP-10) (Nahid *et al*., 2006; Pai *et al*., 2004; Dheda *et al*., 2005; Pai, 2005; Lalvani, 2003). Studies have shown that these tests are more specific, especially in patients who have received BCG vaccination, and are proven to be more sensitive than the Mantoux test (Nahid *et al*., 2006; Pai, 2005). Furthermore, the tuberculin skin test is subjective (i.e. the reading is dependent on the attending physician). Polymerase chain reaction (PCR) based diagnostics tests have also been developed, such as Amplicor MTB tests (Roche Diagnostic Systems, United States) (Nahid *et al*., 2006).

Until recently, it is necessary to culture the *M. tuberculosis* strains isolated from patients in order to determine the antibiotic susceptibility patterns to design an

appropriate treatment regimen (Nahid *et al*., 2006). However, due to the long generation time of *M. tuberculosis*, treatment is often given prior the release of test results. Consequently, numerous kits that rapidly detect drug resistance based on molecular beacons and line probes have been designed to accelerate this screening process (Nahid *et al*., 2006; Lin *et al*., 2004; Morgan *et al*., 2006).

1.6. Prevention, current treatment, DOTS and drug resistance

A vaccine based on an attenuated *M. bovis* strain has been developed and used widely for the prevention of tuberculosis. Drs. Albert Calmette and Camille Guérin passaged pathogenic *M. bovis* several hundred times from 1908-1921 that led to several gene deletions and the eventual creation of the bacille Calmette-Guérin (BCG) vaccine. The BCG vaccine was first tested in man in 1921 (Bonah, 2005) and has since been shown to confer protection against tuberculosis for as long as 50-60 years after a single dose in childhood (Aronson *et al*., 2004). However, this protection is variable and is possibly due to exposure to *M. tuberculosis*-like antigens derived from environmental mycobacteria (Fine, 1995; Brandt *et al*., 2002). For example, in India the vaccine offers little or no protection against pulmonary tuberculosis while in the United Kingdom it is 70% effective (Fine, 1995; Brandt *et al*., 2002; Martin, 2006). Nevertheless, the BCG vaccine can protect against severe forms of tuberculosis in children, such as meningitis and miliary tuberculosis (see Section 1.3.5. Extrapulmonary tuberculosis; Zodpey *et al*., 1996; Powell and Hunt, 2006).

The current WHO-approved treatment for tuberculosis requires a multi-drug therapy comprising of: two months treatment with rifampicin, isoniazid, pyrazinamide and ethambutol (intensive phase) followed by four months treatment with rifampicin and

isoniazid (continuation phase) (WHO, 2003). In all countries, the WHO recommends that the medicines be taken under the supervision of medical professionals to ensure that the correct combination of drugs is taken regularly (this is the so-called DOTS program: Directly Observed Treatment Short Course). DOTS was developed to ensure patients compliance, thus reducing the emergence of drug resistance. In addition to direct observation of treatment, the success of the DOTS program also depends on political/financial support, correct diagnosis and the availability of high quality drugs (WHO, 2002).

Since its introduction, the DOTS campaign has reduced non-compliance and disease burden of tuberculosis. In fact, when effectively managed, DOTS results in a high cure rate for patients infected with tuberculosis (75% in a recent study in the Russian Federation; an additional 14% of patients lost contact; Balabanova *et al.*, 2006). Nonetheless, drug resistant tuberculosis is becoming increasingly prevalent. A recent Centers for Disease Control and Prevention (CDC)/WHO tuberculosis survey showed that MDR-TB is found in nearly 33% of new cases from the industrialized nations between 2000 and 2004 (Centers for Disease Control and Prevention, 2006). In Asia and South America, the percentage of clinical isolates that were MDR-TB was in excess of 50%. WHO estimates that there are around one million cases of MDR-TB cases worldwide annually. Out of the 109 countries surveyed in this study, all have cases of MDR-TB (see WHO website: http://www.who.int/mediacentre/news/releases/2006/pr24/en/index.html). Therefore, it is necessary to develop new second-line drugs that are more effective, less expensive and less toxic than the currently available ones. The international community is also making efforts to reduce the cost of anti-tuberculosis drugs. For instance, the Working Group on DOTS-Plus

for MDR-TB (the so-called Green Light Committee) that has dramatically reduced tuberculosis health care costs (sometimes by as much as 99%) through negotiations with the pharmaceutical industry to provide second-line tuberculosis drugs at reduced prices (see WHO website: http://www.who.int/tb/dots/dotsplus/management/en/; Gupta *et al*., 2001; Farmer and Kim, 1998; WHO, 2000; Onyebujoh *et al*., 2005)

1.7. HIV Infection, AIDS and tuberculosis

As noted previously, one third of the world's population has latent tuberculosis and reactivation of the disease could occur when the patient is old, immunodeficient or immunosuppressed (Dye *et al*., 1999). Studies have shown that around a third of HIV patients are also co-infected with tuberculosis (Onyebujoh *et al*., 2005; Zumla *et al*., 2000). Moreover, through the DOTS campaign, tuberculosis incidences have been decreasing in all countries, except in areas where HIV infection is prevalent (Elzinga *et al*., 2004). The loss of $CD4^+$ T-lymphocytes in HIV-infected individuals results in reactivation of tuberculosis (Chan and Kaufmann, 1994). Often AIDS patients also develop *M. avium* infections due to the same reason (Inderlied *et al*., 1993; Ellner *et al*., 1991; Biava *et al*., 2006). This presents a major reason for the need to develop new tuberculosis drugs, especially since HIV patients have reduced immune capability to control mycobacterial infections.

1.8. Essential drug targets in *M. tuberculosis*

It is generally believed that an ideal drug target should be an essential gene. In other words, essential genes are those that are required for maintaining the viability of the microorganism. Extensive transposon mutagenesis studies have indicated that one third of the *M. tuberculosis* genes are essential. These essential genes can be divided

into seven gene families, namely proteins/enzymes involved in aminoacyl tRNA synthase activities, purine ribonucleotide biosynthesis, polyketide and nonribosomal peptide synthesis, fatty acid and mycolic acid biosynthesis, Ser/Thr protein kinase and phosphatases, molybdopterin biosynthesis and PE-PGRS repeats (Lamichhane *et al*., 2003; Zhang, 2005). Transposon mutagenesis studies have also identified conditionally essential genes (Sassetti *et al*., 2001) and genes necessary for optimal growth under *in vitro* conditions (Sassetti *et al*., 2003). Besides this, gene deletion studies have illustrated the essentiality of several mycobacterial genes by demonstrating the feasibility of deleting a gene only when an additional copy of the gene is introduced into the genome using a plasmid (Zhang, 2005).

However, these results of 'essentiality' have to be viewed with caution. It is conceivable that genes that are determined to be non-essential under *in vitro* conditions could well be essential for survival in humans or important for establishing a successful infection (Sassetti and Rubin, 2003). Nevertheless, the data provides us with a guide for selecting biological targets for high-throughput screening and drug discovery.

The current mycobacterial drugs and their potential targets are discussed in Table 1.1.

Current Targets	Rationale	Anti-Tuberculosis Drug Example
Cell Wall Synthesis	The thick, waxy mycobacterial cell is an integral part of the bacteria's defenses against the hostile environment in the human macrophage. Moreover, to replicate the bacilli are required to produce more components of the cell wall, such as mycolic acids. Hence, drugs that target enzymes involved in cell wall synthesis would disrupt the ability of mycobacteria to divide or survive effectively in the macrophage.	Ethambutol, Isoniazid, Cycloserine, Ethionamide
DNA replication	The replication of DNA is an important part of cellular division. Hence, preventing proper DNA replication would halt the growth or kill the bacilli. DNA gyrase is also important in the assembly of proper folding of double-stranded DNA.	Moxifloxacin, Ciprofloxacin
Transcription	The transcription of the DNA genetic code into mRNA is a necessary part in the production of proteins. Inhibition of this process would disrupt the intracellular homeostasis and lead to bacterial death.	Rifampicin, Rifabutin
Protein Synthesis	The purpose of proteins in bacteria, like in other cells, is to maintain the metabolic homeostasis within the cells. Any disruption of cellular production of proteins would result in the cellular death or at least inhibition of division. Some of the best tuberculosis drugs target the essential machinery to produce proteins, i.e. the ribosomes.	Kanamycin, Streptomycin, Amikacin
Disruption of Membrane Energy	The survival of any bacteria (or cells in general) relies on the production of ATP, the so-called *energy currency* of any living entity. Thus, specific disruption of the membrane potential or electron transport mechanisms in the bacterial cell can inhibit the regeneration of ATP molecules from ADP. This will essentially shut down the bacteria's metabolism.	R207910, Pyrazinamide

Table 1.1: Current anti-tuberculosis drug targets with specific examples

1.9. Chemical genetics as an approach for drug discovery

In biology, it is possible to study biological systems by manipulating the genetic code (genetics) or by modifying the function of a protein using a chemical compound (chemical genetics) (Spring, 2005). 'Forward' method of genetics or chemical genetics is where the phenotype is selected and the gene/protein responsible for that phenotype is elucidated. Conversely, a 'reverse' method involves the manipulation of a known gene/protein followed by the analysis of the resulting phenotype (Spring, 2005). In antibacterial drug discovery, the phenotype desired is either cellular death or inhibition of bacterial replication.

Modern drug discovery using high-throughput screening (HTS) is a reverse chemical genetics approach where a library of chemically diverse compounds (usually about

1 million) is screened using an enzymatic assay. Subsequently, compounds capable of modulating the activity of the proteins are selected ('*hits*'). Critically, it is necessary that a drug target be validated genetically or pharmacologically prior to the HTS. Despite the speed and cost-effectiveness of HTS, there is no guarantee that the compounds identified can actually penetrate through the cell wall to reach its biological target *in vivo* (Hung and Rubin, 2006). Moreover, enzymes/proteins used in the HTS could have overlapping functions with other enzyme/protein. Thus, although the enzymes/proteins used for HTS are essential, the compound activity exhibited under *in vitro* conditions may not be reflected *in vivo* (Hung and Rubin, 2006).

Drugs that inhibit novel targets would eliminate the problem of cross-resistance with current therapies. This is especially a problem in infectious disease, such as tuberculosis, where multi-drug resistance is fast emerging. As a drug-discovery institute, our goal is to develop new drugs to treat tuberculosis. To accomplish this, we have chosen to use a forward chemical genetics approach because it enables us to develop new drugs with a novel mechanism of action. In other words, we are screening a library of compounds against the entire proteome for novel targets, that when inhibited results in bacterial death or inhibition of growth.

1.10. Technologies available for identification of drug targets

The identification of the biological target of a compound has several advantages for drug development. Critically, a known drug target would allow scientists to develop target-based assays and to facilitate lead optimization by establishing a structure-activity relationship (SAR). Moreover, with a known target it would be possible to predict potential side effects or toxicity issues.

One method to identify the target of a drug is to generate spontaneous mutants that are resistant to the compound of interest and characterize the genomic, proteomic and transcriptomics profiles of the mutants in comparison to the parental wild-type strain. From this information, the target can be elucidated. Other more direct methods include affinity-chromatography to pull down the protein target or the more traditional method of library screening. It should be noted that sometimes what we identify is not the 'target' but simply a resistance mechanism, e.g. mutating an activating enzyme of a pro-drug, or just a protein that binds with the compound (especially in affinity chromatography).

1.10.1. Confirmation of drug targets

The above methodologies and technologies, either used alone or in combination, can lead to candidate targets that need to be confirmed with further experiments. This is a crucial part of drug target identification because, *positive* results from the affinity chromatography or protein microarray, could just be a promiscuous binder of the compound, rather than a true pharmacological target. In addition, indirect methods such as microarray profiling only gives an indication of the pathway affected by the

compound. It does not pinpoint the gene product that is the actual site of action of the compound.

Confirmation of the drug target can be accomplished by over-expressing mutant gene in wild-type cells, thereby demonstrating the transferability of observed phenotype. This process is also known as complementation. In addition, gene knock-downs *via* antisense-mediate silencing or gene knockouts *via* homologous recombination can also be used for target confirmation. However, only the complementation approach is practical in mycobacterial drug discovery because gene-silencing techniques are not well established in mycobacteria and essential genes cannot be inactivated using genetic techniques.

	Approach	Used	Description	Example	References
Genetic code-based approaches	**Plasmid or Cosmid Library Screening**	**Yes**	A library of randomly cut genomic fragments are expressed in a bacterial population and selected for a desired phenotype. Plasmid library constructs can accommodate genomic DNA efficiently between 2kb and 6kb in size, while the cosmid library constructs can contain up to 45kb of DNA. One would transform either the mutant library into parental strains or the wild-type library into mutant cells and selecting for constructs that confer resistant or sensitivity, respectively.	The putative target of isoniazid, *inhA*	Banerjee *et al*., 1994; Vilcheze *et al.*, 2006
	Transposon Mutagenesis	**No**	A library of randomly inserted transposon clones, by definition, targeting non-essential genes can be screened for a phenotype, such as resistance against a given compound.	The inactivation of *rdxA* gene in the *Helicobacter pylori* results in resistance against the pro-drug metronidazole.	Goodwin *et al*., 1998
	Whole Genome Sequencing (WGS)	**No**	This method is based on the central dogma of biology (gene→mRNA→protein→phenotype). Using drug-resistant bacteria one can identify the location of the mutation by sequencing the complete genome. Just a few years ago, it was necessary to randomly sequence short segments of DNA that are aligned together ('shot-gun' approach). It is now possible to sequence the complete genome *of M. tuberculosis* in a few days.	A successful example of using this technology to identify the biological target of a compound is the diarylquinoline drug that targets the ATP synthase	Andries *et al*., 2005
	Comparative Genome Sequencing (CGS)	**Yes**	This approach is based on the differential hybridization patterns of fragmented DNA where the profiles between the parental and mutant are compared and differences identified. The single nucleotide polymorphisms (SNPs) are 'sequenced' with a DNA chip using a series of overlapping oligo-nucleotides with single base pair differences.		NimbleGen Systems Inc (see NimbleGen website: www.nimblegen.com, United States).
Direct Methods	**Affinity Purification**	**No**	Briefly, total cellular protein extracts are run through an affinity chromatography column (e.g. latex-based resins) onto which the compound is immobilized. Following elution, the bound proteins are identified by mass spectrometry.	Cyclosporine, Colchicines, Fumagillin, Acetylcholine, Rapamycin	Raftery *et al*., 1980; Noda *et al*., 1982; Mitchison, 1994; Brown *et al*., 1994; Harding *et al*., 1989; Borisy and Taylor, 1967; Borisy and Taylor
	Protein Microarray	**No**	One protein microarray approach is to directly spot purified proteins onto chemically derivatized glass or with immobilizing antibodies. In eukaryotes, another protein microarray approach is to spot a collection of plasmid-based vectors expressing different cDNAs and cover it with a layer of mammalian cells and transfection reagent. This would create an array of cells that are over-expressing different proteins and could be probed with a radioactively labeled compound.		MacBeath and Schreiber, 2000; Zhu *et al*., 2001
Systems Biology Approach	**Transcriptomics**	**Yes**	DNA microarrays can identify effects of drug exposure at the mRNA level. By comparing the mRNA gene expression profile of the drug with a database, it is possible to determine if drug acts in the same pathway as any of the reference drugs. Alternatively, it can also be determined that a gene-deletion has a similar profile to compound exposure hence suggesting that gene to be the target.	A microarray database of around 400 tuberculosis drugs has been compiled at the National Institutes of Health (Barry, C., USA) with which we can compare the profiles of our compounds; Another example is dyclonine.	Hughes *et al*., 2000; Boshoff *et al*., 2004
	Proteomics	**Yes**	Total protein extracts separated on a 2D gel electrophoresis format will be resolved in both isoelectric point (pI) and molecular weight (MW) directions; the identity of spots determined by tandem mass spectrometry. Cells grown under different conditions, such as different drug exposure or genetic background, would have a differential expression of proteins or a change in protein mobility.	Bengamides were found to be methionine amino peptidase inhibitors because certain proteins retained its initiation methionine based on the identification of 1500 different spots.	Towbin *et al*., 2003
	Lipidomics	**Yes**	Similar to microarray profiling, this technique needs a database of metabolite profiles treated with a spectrum of drugs that can be used for comparison with the profile of the compound of interest.		Watkins, 2000

Figure 1.2: Technologies that can be used to identify the biological target of compounds. Note that, by definition, the proteins/genes identified are only a resistance mechanism or binder of compound but, in some cases, it could be the biological target of the compounds. Confirmation and validation of these results are necessary (for more see in text).

1.10.2. Validation of drug targets

In antibacterial drug discovery, validation of a drug target involves the understanding of the protein's function in the microbe and how by the inhibition of the protein or alteration of its function could lead to the death or inhibition of the pathogenic organism. In other words, we have to prove that the target is essential for the survival/infectivity of the microorganism or at least is important in some stage of disease progression (Stockwell, 2000). This is often the bottleneck in the drug discovery process, especially for new drug targets where very little is known about its function (Stockwell, 2000).

1.11. Goals of Tuberculosis Drug Research and Discovery

In summary, there is a great need to develop drugs for tuberculosis and ideally these new drugs have to inhibit novel biological targets to avoid cross-resistance with current therapies. The best drug targets should be essential in both active and dormant phases of *M. tuberculosis* infection so that a single drug would eradicate both bacterial populations.

Ultimately, according to the *Scientific Blueprint for Tuberculosis Drug Development* (Global Alliance for TB Drug Development) any new drug would have to: (1) improve the current tuberculosis treatment by shortening the length of therapy, (2) be more effective against MDR-TB and (3) in light of the fact that two billion people are latently infected with *M. tuberculosis*, the drugs need to have efficacy against latent tuberculosis (O'Brien and Spigelman, 2005; Duncan, 2004; O'Brien and Nunn, 2001).

Objectives

In this study, we aim to isolate and characterize *M. bovis* BCG and *M. smegmatis* drug-resistant mutants to Compound X and Compound Y. Using these drug-resistant mutants, we intend to identify the drug target of Compound X and Compound Y using proteomic and genetic approaches.

Chapter Two: Materials and Methods

2. Materials and Methods

2.1. Bacterial Strains, Growth Media, Compounds and Drugs

2.1.1. Bacterial Strains

The slow-growing *Mycobacterium bovis* BCG (ATCC 35734; *M. bovis* BCG will be known as BCG) and fast-growing *M. smegmatis* (mc^2155) were used in the experiments. *Escherichia coli* TOP10 cells (Invitrogen, USA) were used in cloning experiments.

2.1.2. Bacterial Culture Media

Luria-Bertani (LB) broth and LB agar plates: Both media were used to culture *E. coli* cells. LB (Becton Dickinson, USA) broth was filter sterilized and stored at 37°C until use. LB Agar (Becton Dickinson, USA) was autoclaved and an appropriate amount of antibiotic was added immediately before the plates were prepared. After the plates set, they were stored at 4°C until required. When required, LB broth and LB agar plates was supplemented with 50μg/ml kanamycin (Sigma, USA), 150 μg/ml hygromycin (Roche, Switzerland) or 100 μg/ml of ampicillin (Sigma, USA).

Middlebrook 7H9 broth: Mycobacterial cultures were grown using this liquid media. Middlebrook 7H9 was prepared according to manufacturer's protocols (Becton Dickinson, USA) and supplemented with 0.2% glycerol, 0.05% Tween-80 and 10% (v/v) Albumin-Dextrose-Saline (ADS: 950ml dH_20, 8.1g NaCl, 50g Bovine Serum Albumin Fraction V, 20g D-dextrose). The complete medium was filter sterilized and stored at 37°C until use. When required, Middlebrook 7H9 broth were supplemented

with 25μg/ml kanamycin or 50 μg/ml hygromycin for antibiotic selection. Tween-80 was added in the media in order to decrease cellular clumping.

Middlebrook 7H10 agar plates: This solid media was used to isolated colonies of mycobacteria. Middlebrook 7H10 was prepared according to manufacturer's protocols and autoclaved. After sterilization, the media was supplemented with 0.5% glycerol, 0.05% Tween-80 and 10% (v/v) oleic acid-dextrose-albumin-catalase enrichment (OADC; Becton Dickinson, USA). When required, Middlebrook 7H10 agar plates were supplemented with either 25μg/ml kanamycin or 50 μg/ml hygromycin for antibiotic selection.

2.1.3. Glycerol stock of bacteria

E. coli: Glycerol stocks of overnight cultures were prepared by resuspending in LB broth containing 15% glycerol (v/v) and stored as 1ml aliquots at -80ºC.

BCG and *M. smegmatis*: Glycerol stocks of cultures were prepared by resuspending in Middlebrook 7H9 broth containing 15% glycerol (v/v) and stored as 1ml aliquots at -80ºC.

2.1.4. Compounds

All compounds were prepared as 5mM stocks in 90% DMSO and stored at 4°C until further use.

2.1.5. Drugs

Phosphomycin, imipenem, gatifloxacin, ethambutol, amikacin, ciprofloxacin, kanamycin, ofloxacin, vancomycin, ceftriaxone, azithromycin, gentamicin, ampicillin,

tetracycline, trimethoprim, moxifloxacin, cycloserine, sulfamethoxazole, rifampicin, streptomycin, isoniazid, rifabutin and linezolid stocks were prepared as 5mM stocks in 90% dimethyl sulfoxide (DMSO) and stored at 4°C until further use. These drugs were obtained from Sigma (USA) and Merck (USA).

2.2. Isolation and characterization of compound resistant mutants

2.2.1. MIC_{50} and MBC_{90} determination

The MIC_{50} (Minimum Inhibitory $Concentration_{50}$) is defined as the minimum concentration of compound at which 50% of mycobacterial growth is inhibited. 100µl of Middlebrook 7H9 was aliquoted into each well of a flat-bottom 96-well plate. A compound solution (in Middlebrook 7H9) at four-times the highest concentration required was prepared and aliquoted into the top wells of a 96-well plate followed by a two-fold serial dilution (for concentrations used see Table 2.1). Each of the wells were seeded with mycobacterial cells at a final optical density of 0.02 for BCG and 0.01 for *M. smegmatis*. The optical density was determined at 600nm (OD_{600}; SPECTRAmaxM2, Molecular Devices, USA) after 5 days of incubation for BCG and 2 days for *M. smegmatis* at 37ºC. The MIC_{50} values were calculated using the statistical software Prism (GraphPad, USA), where 100% growth is defined according to the average of the wells containing no compound. All compounds were tested in duplication in three independent trials.

The minimum bactericidal concentration 90 (MBC_{90}) is the lowest concentration of the compound that results in the killing of 90% of bacilli from the starting inoculum. This was calculated by determining the colony-forming units (CFUs) at day 0 and day 5 following compound treatment on Middlebrook 7H10 agar plates. Since the MBC_{90} is

higher than the MIC_{50}, we plated out the culture from the wells with concentrations higher than the MIC_{50} (usually up to 4 wells above the MIC_{50}). The CFUs of the no-drug control (using both initial inoculum at day 0 and the cells after 5 days of incubation) were also determined.

2.2.2. Isolation of spontaneous compound-resistant mutants (mutation frequency determination)

Based on the MIC_{50} assay results, mutants resistant to CpdX were isolated. 400ml of wild-type BCG were grown to an optical density of 0.4. Following centrifugation at 3,000rpm for 10 minutes, the cell pellet was resuspended in 1ml of fresh Middlebrook 7H9 broth. 100µl of concentrated culture was plated onto Middlebrook 7H10 agar plates containing five and ten times MIC_{90} concentration of the compound under investigation (in duplicates). 100-fold diluted cultures were also plated out on five and ten times MIC_{90} concentration on Middlebrook 7H10 agar plates in duplicates. In order to estimate the bacterial CFUs in the original culture, 100µl of 10^6, 10^7 and 10^8-fold diluted culture were plated onto Middlebrook 7H10 agar plates without drugs. The agar plates were incubated at 37°C for between 3 to 5 weeks prior to determining the CFUs counts and isolation of spontaneous resistant mutants. Note that the protocol used for *M. smegmatis* was similar with the only difference being that the mutants were isolated following 3-5 days of incubation at 37°C. In addition, the mutation frequency of CpdX and CpdY for BCG, and CpdY for *M. smegmatis* were also determined in the same experiment. We determined the mutation frequency of rifampicin as a positive control.

Compound-resistant mutants were colony purified by plating out on fresh Middlebrook 7H10 agar plates. From this, single colonies were isolated and grown in 2-3ml of drug-free Middlebrook 7H9 broth. Following growth, 0.5ml of the culture was transferred into 5ml of fresh Middlebrook 7H9 broth containing the appropriate concentration of drug with which each mutant was originally isolated. Following growth to an OD_{600} of 0.4 to 0.8, the bacterial cultures were spun down (3,000 rpm for 10 minutes), resuspended in fresh compound-free Middlebrook 7H9 broth with 15% glycerol and stored in aliquots at -80°C.

2.2.3. Selection of drugs to use in cross-resistance studies

The MIC_{50} of phosphomycin, imipenem, gatifloxacin, ethambutol, amikacin, ciprofloxacin, kanamycin, ofloxacin, vancomycin, ceftriaxone, azithromycin, gentamicin, ampicillin, tetracycline, trimethoprim, moxifloxacin, cycloserine, sulfamethoxazole, rifampicin, streptomycin, isoniazid, rifabutin and linezolid were determined in BCG and *M. smegmatis* according to protocols outlined in Section 2.2.1. (MIC_{50} and MBC_{90} determination) starting at 40µM for all drugs.

2.3. Molecular Biology

2.3.1. Polymerase Chain Reaction (PCR)

PCR was used to amplify the mutant and wild-type *corA* gene to facilitate cloning into the TOPO vector (pCR2.1-TOPO; Invitrogen) as outlined in the next section. Briefly, a PCR mixture (total volume 30µl), comprising of ~200ng of template DNA was mixed with Pfu Ultra HotStart mix (15µl; Stratagene; USA), 10mM primers (0.5µl) and made up to 30µl with dH_2O. Amplification conditions were as follows: hot start of 5min at 95°C; 25 cycles of denaturing (95°C for 30 s), annealing (58°C for 15 s), and

extension (72°C for 30s); and a final extension of 10min at 72°C was used for sequencing, expression and knock-out studies (PCR machine: Biometra T3000, Germany). The primers used in this work are listed in Table 2.2.

Primer Name	Primer Sequence
Sequencing	
H37Rv CorA Seq.fw1	5'-ATCCACCGCGATCGGAGGTT-3'
H37Rv CorA Seq.rv1	5'-TGTTCGCCGACGCGAAGCTA-3'
H37Rv CorA Seq.fw2	5'-CGGCGAACACGGCGGGTTAT-3'
H37Rv CorA Seq.rv2	5'-CCATGAGATTGGTCACCTCG-3'
Msmeg CorA Seq.fw1	5'-AACATGGCGAAGTTCCGGGC-3'
Msmeg CorA Seq.rv1	5'-TCCCTCCTGAGGCTCTAGAG-3'
Msmeg CorA Seq.fw2	5'-GATTCCTCGCCGTCGATCCT-3'
Msmeg CorA Seq.rv2	5'-GATCGGTGACATCGAGGTAG-3'
Msmeg CorA-like Seq.fw1	5'-TTGACTCACGGCATGGCGCAGG-3'
Msmeg CorA-like Seq.rv1	5'-CTACAGCCAATCGCGGCGCTTG-3'
Msmeg CorA-like Seq.fw2	5'-TCGGCGTTCGTGTTGCCGCACG-3'
Msmeg CorA-like Seq.rv2	5'-AGTTCGACGAGGTCCTTGCGC-3'
pMV261.fw	5'-CCGGTTTCATCCCCGATCCGGA-3'
pMV261.rv	5'-CGTACGCTAGTTAACTACGTCG-3'
pYUB415.fw	5'-GATAAGCGGTCAAACATGAG-3'
pYUB415.rv	5'-TGCCACCTGACGTCTAAGAA-3'
Expression Construct	
H37Rv CorA Exp.fw	5'-TTGGCCTCCAGCAGGTCACT-3'
H37Rv CorA Exp.rv	5'-ACCGCCAGACTTGCCGTCTT-3'
Msmeg CorA Exp.fw	5'-GCCGAACGTCTTGAACAAC-3'
Msmeg CorA Exp.rv	5'-AACTTCTGACCGCAACAC-3'

Table 2.1: List of primers used for sequencing and expression studies

2.3.2. TOPO cloning

The TOPO-cloning solution was prepared as follows: 4µl of fresh PCR product, 1µl of sterile water, 1µl of the salt solution (1.2M NaCl, 0.06M $MgCl_2$) and 1µl of the TOPO was incubated at room temperature for 5minutes. The entire solution was transformed into chemically competent *E. coli* Top10 cells (Invitrogen, USA).

2.3.3. Transformation of *E. coli* and mycobacteria

DNA was mixed gently with a thawed aliquot of chemically competent *E. coli* Top10 cells and incubated on ice for 30 minutes. The competent cells were heat shocked at 42ºC for 1 minute and chilled on ice. 500µl of LB was added to the cells followed by incubation in a shaker for 60-90 minutes at 37ºC to allow for expression of antibiotic resistance. The cells were centrifuged briefly and the LB decanted off leaving approximately 100µl. The cells were resuspended and spread onto LB Agar plates containing kanamycin and incubated overnight at 37°C. For the TOPO clones, 40µL of Xgal reagent was added onto the plates prior to plating cells to facilitate the selection of recombinant colonies, which will appear white due to the insertion of the DNA into the *lacZ* gene (*lacZ* gene codes for β-galactosidase that can break down Xgal to a blue product). A single white colony was used to inoculate 3ml of LB broth and plasmids were purified following protocols in Section 2.3.9. followed by restriction endonuclease digestion (Section 2.3.4.) and sequencing (Section 2.3.12.).

For BCG, electrotransformation was used to introduce foreign DNA. 100ml of mid-log phase cells were made electrocompetent by washing the cells with 10% glycerol three times and resuspending in 1ml of 10% glycerol. 200µl of cells was mixed with the DNA in a 0.2cm cuvette and incubated for 5-10minutes at room temperature.

Following electroporation (BioRad Gene Pulser, settings: 2500mV, 1000Ω, 25μF), 1ml of Middlebrook 7H9 media was added to the cells and incubated overnight at 37ºC in sterile tubes to allow expression of antibiotic resistance gene.

2.3.4. Restriction Enzyme Digestion

Analytical restriction enzyme digestion was used to screen for recombinant plasmids containing the DNA of interest. The digestion reaction mixture consist of ~200ng of plasmid DNA, 1 unit restriction enzyme, appropriate buffer and made up to a final volume of 10μl with dH_2O. The whole mixture was incubated at the appropriate temperature for 30-120 minutes and separated by agarose gel electrophoresis (Section 2.3.5. Agarose Gel Electrophoresis). The restriction enzymes used were supplied by Invitrogen, New England BioLabs or Roche.

2.3.5. Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate the DNA fragments of different sizes. A 1% agarose gel was used unless otherwise indicated. The appropriate amount of agarose (GibcoBRL) was added to the 1xTris-Acetate EDTA buffer (1xTAE Buffer; 40mM Tris, 1.4% (v/v) glacial acetic acid, 1mM EDTA) and microwave-heated until the agarose had completely dissolved. Ethidium bromide (around 1μl of 10mg/ml stock; GibcoBRL) was added and the gel poured into a casting tray and left to set. The gel was placed in an electrophoresis tank (BioRad) and submerged in 1xTAE buffer. Each sample was mixed with 6xDNA loading dye (0.03% xylene cyanol FF, 0.03% bromophenol blue, 60% glycerol, 1% sodium dodecyl sulfate (SDS), 100mM EDTA (pH 7.6)) and 3-5 μl of molecular mass standard 1kb plus ladder (GibcoBRL; 100μL H_2O, 25μl loading dye, 375μl 20mM NaCl) and loaded in adjacent wells.

Electrophoresis was performed at 120V for approximately 30minutes. The gel was exposed to UV light (Gene Genius, Syngene, United Kingdom) to visualize the DNA fragments.

2.3.6. Purification of digested plasmid DNA from agarose gels

1-2μg of plasmid DNA was digested with restriction enzymes and separated on a 1% (w/v) agarose gel for approximately 30minutes at 180V. With the aid of UV light, the gel bands that contained the vector and insert DNA were excised and purified from the agarose gel using the QIAquick gel extraction kit (Qiagen, USA). The DNA was eluted with TE buffer.

2.3.7. Dephosphorylation of DNA

Digested vector DNA was treated with 1U of Shrimp alkaline phosphatase (SAP) for 15 minutes to remove 5' phosphates. This would prevent self-ligation of the vector. The vector DNA was purified using the QIAquick gel extraction kit (Qiagen, USA) according to the manufacturer's instructions.

2.3.8. Ligation of DNA fragments

Ligation was carried according to manufacturer's instructions (Roche Rapid Ligation Kit) with 100ng of vector DNA and insert DNA at 3-5 molar excess. The ligation mixture was incubated at room temperature for 5 minutes prior to transformation into competent *E. coli* TOP10 cells (Section 2.3.3. Transformation of Plasmid DNA).

2.3.9. Small scale preparation of plasmid DNA

Chemically competent *E. coli* transformed with plasmids were inoculated in 3ml of LB containing 50µg/ml of ampicillin or 150µg/ml of hygromycin and cultured at 37^{o}C overnight. The plasmid DNA was purified according to the manufacturer's protocol (Roche High Pure Plasmid Isolation Kit Handbook). The DNA was eluted with TE (Tris-EDTA buffer: 10mM Tris HCl pH8.0, 1mM EDTA, 50µg/ml RNAseA (Roche)) buffer and stored at -20ºC until further use.

2.3.10. Large scale preparation of plasmid DNA

100ml of LB supplemented with 50µg/ml of ampicillin or 150µg/ml of hygromycin was inoculated with either a bacterial glycerol stock or a small starter culture of 2-5ml LB (with 50ug/ml of ampicillin). The bacteria were grown for 14 hours overnight on a shaker-incubator at 37ºC. Plasmid DNA was purified according to the manufacturer's instructions (Qiagen, USA). The plasmid DNA was resuspended in TE and stored at -20ºC until use.

2.3.11. Sequencing

Plasmids and gel-purified PCR products were sequenced using am ABI PRISM 3100 capillary sequencer (Applied Biosystems, USA) at Research BioLabs Technologies Pte Ltd (Singapore). For sequencing inserts in the TOPO vector the T3 forward and T7 reverse primers were used (for sequences of the primers see Invitrogen website; available at www.invitrogen.com). For sequencing PCR products, the primers used in amplification were used in addition to internal primers. Note that since the sequencing reaction can only accurately read out approximately 600bp, hence to obtain the

sequence of the full-length *corA* gene internal primers were required (H37Rv *corA* seq.fw2 and rv2).

2.4. Identification of drug target

2.4.1. Comparative Genome Sequencing

Comparative genome sequencing is a hybridization-based sequencing method (NimbleGen, USA). For more information please visit: www.nimblegen.com. We provided NimbleGen with phenol:chloroform purified genomic DNA comparative hybridization studies.

2.4.1.1. Genomic DNA preparation

100ml of culture (OD_{600} 0.4-0.8) was split into five tubes and spun down in a tabletop centrifuge and resuspended in 1ml of GTE solution (25mM Tris-HCl pH8.0, 10mM EDTA, 50mM glucose). In a 2ml microfuge tube, the cells were resuspended in 450µl of GTE solution with 50µl of 10mg/ml lysozyme solution (Sigma, USA). Following an overnight incubation at 37ºC, 100µl of 10% (w/v) SDS and 50µl of 10mg/ml proteinase K was added and incubated at 55ºC for 20-40 minutes. Following addition of 200ul of 5M NaCl, 160µl of preheated cetyltrimethylammonium bromide solution (CTAB solution: 4.1g NaCl, 10g CTAB in 90ml dH_2O) was added to the mixture and incubated at 65ºC for 10 minutes. Phase separation was achieved with equal volume of chloroform:isoamyl alcohol (24:1) coupled to centrifugation for 5 minutes (repeated twice using the previous aqueous phase). Equal volume of isopropanol was used to precipitate the genomic DNA and the pellet following centrifugation was washed with 70% ethanol. The air-dried, ethanol-washed DNA pellet was resuspended in 1xTE and stored at 4ºC until required.

2.4.1.2. Determination of DNA concentration and purity

A NanoDrop ND-1000 Sprectrophotometer (NanoDrop Technologies, USA) was used to determine the genomic DNA concentration. The A260/280 ratio was used to determine the purity of the DNA preparation, where a ratio of 1.8 indicates a high purity DNA preparation devoid of proteins, aromatic compounds and RNA contamination.

2.4.2. 2D gel electrophoresis

2.4.2.1. Sample preparation

Whole-cell extracts were prepared from BCG by harvesting 100ml of culture or wild-type and mutant strains grown for 3 days with or without CpdX drug pressure (at either 0.6µM or 1.2µM). Cultures were pelleted by centrifugation at 3000rpm for 15minutes. Cell pellets were washed three times with water containing 0.05% (v/v) Tween 80. Mycobacteria were resuspended in lysis buffer, containing 25mM Tris pH8, 0.1mM EDTA, 5% glycerol, 1% Triton-X 100, 1mM PMSF and 50mM dithiothreitol (DTT), and lysed by bead disruption in a ribolyser (Eppendorf, Germany) The debris in the lysate was spun down and the supernatant removed. The protein concentration of each sample was estimated using the BioRad Protein Assay (USA).

2.4.2.2. Electrophoresis

Isoelectric focusing sample loading buffer (8M urea, 2% (w/v) CHAPS, trace amount of bromophenol blue) was added to the samples and each separated using pH 4-7 18cm IPG strips (Amersham Bioscience) in the first dimension. Total protein (100µg made up to 350µl with sample loading buffer and 1.4µl 0.5M Tris-(2-carboxyethyl)phosphine hydrochloride (TCEP), 1.75µl IPG buffer) was loaded at the

cathode end and proteins focused as follows using Etthan™ IPGphor (Amersham Bioscience): 30V for 12h, 500V for 1h, 1000V for 1h, 8000V for 8h (total of 60,000Vh). Second-dimension electrophoresis was carried out using 18x18cm format, 1·5mm thick, 12% polyacrylamide gels (50ml: 16.5ml dH_2O, 20ml 30% acrylamide mix, 12.5ml 1.5M Tris pH8.8, 0.5ml 10% SDS, 0.5ml 10% ammonium persulfate, 0.02ml TEMED). IPG strips were equilibrated with prior to running on the second dimension using equilibration buffer (200ml: 6.7ml 1.5M Tris-HCl pH8.8, 6M urea, 69ml glycerol, 4g SDS and bromophenol blue) with 10mg/ml DTT for 15 minutes and equilibration buffer with 25mg/ml iodoacetamide for 20 minutes. Strips were then overlaid onto the second-dimension gels and sealed with 0.5% (w/v) agarose in cathode buffer (TGS Buffer, BioRad, USA), containing a trace amount of bromophenol blue.

2.4.2.3. Silver staining

For protein spot comparisons, gels were silver stained. Firstly, gels were immersed in fixing solution, comprising of 50% (v/v) methanol and 5% (v/v) acetic acid, for at least 1 hour. This was followed by successive incubations in 0.02% (w/v) sodium thiosulfate for 2min, 0.1% chilled silver nitrate for 40min and 0.014% formaldehyde plus 2% sodium carbonate. The reaction was stopped with 5% (v/v) acetic acid. Gel images were digitized by scanning on a flat-bed scanner (ScanJet2400, Hewlett Packard, USA) at a resolution of 300dpi.

2.4.2.4. Spot identification

Protein spots of interest were excised and send for identification at the Biopolis Shared Facility (Singapore) using Q-STAR (LC-MS) in order to concentrate the protein prior to mass spectrometry.

2.4.2.5. Database searching

Tandem mass spectra were extracted, charge state deconvoluted and deisotoped by ABI Analyst version 1.4. All MS/MS samples were analyzed using Mascot (Matrix Science, London, UK; version 2.1.03) and X! Tandem (www.thegpm.org; version 2006.04.01.2). X! Tandem was set up to search a subset of the uniprot_20060725 database also assuming trypsin. Mascot was set up to search the uniprot_20060725 database (selected for *M. tuberculosis* complex, unknown version, 9837 entries) assuming the digestion enzyme trypsin. Mascot and X! Tandem were searched with a fragment ion mass tolerance of 0.50 Da and a parent ion tolerance of 200 PPM. Oxidation of methionine and iodoacetamide derivative of cysteine were specified in Mascot and X! Tandem as variable modifications.

2.4.2.6. Criteria for protein identification

Scaffold (version Scaffold-01_06_03, Proteome Software Inc., USA) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 0.0% probability as specified by the Peptide Proph*et al*gorithm (Keller *et al*., 2002). Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii, 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

2.4.3. Bioinformatics

Sequences of *corA* in *M. tuberculosis* H37Rv, *M. tuberculosis* CDC1551, *M. smegmatis* mc^2155, *M. bovis*, *M. leprae*, *M. paratuberculosis*, *M. flavescens* and *M tuberculosis* strain C were aligned using VectorNTI to determine their homology to each other (AlignX, Invitrogen, USA).

Chapter Three: Results

3. **Results**

3.1. Anti-mycobacterial activity of Compound X and Compound Y

CpdX was selected from a library of 8,800 synthetic compounds following a growth inhibition screen on actively replicating BCG. CpdY was selected based on structural similarities to CpdX (personal communication, Vasan Sambandamurthy, NITD, Singapore).

The inhibitory and bactericidal activity of the two structurally-related chemical compounds, CpdX and CpdY, were determined. They both have good inhibitory activity on slow-growing mycobacterial species grown under aerobic conditions. CpdX has excellent bactericidal activity against BCG grown under aerobic conditions. However, CpdX has weak activity against *M. smegmatis* (over 25μM), while CpdY has good activity (MIC_{50} of 1-2μM).

Compound	MIC_{50}		MBC_{90} *M. bovis* BCG
	M. bovis BCG	*M. smegmatis*	
Compound X	0.62-1.73	>25	<2.5
Compound Y	0.36	1.47	ND

Table 3.1: Anti-mycobacterial properties of CpdX and CpdY against wild-type *M. bovis* BCG and *M. smegmatis*. This table summarized the basic biological properties of the two compounds against *M. bovis* BCG and *M. smegmatis*. (ND denotes not done and all MIC_{50} and MBC_{90} values are expressed in μM)

3.2. Isolation of spontaneous resistant mutants to Compound X and Compound Y in *M. bovis* BCG and *M. smegmatis*

Mutation frequency studies 1) can hint at whether the bacteria can quickly acquire genetic resistance, if used as a monotherapy and 2) allow for the isolation of resistant mutants that can be genetically analyzed so as to ascertain the mutation that confers resistance and possibly establish the biological target of these compounds. The frequency of appearance of spontaneous resistant mutants for both compounds in BCG was ~10^{-8} at 10 times MIC_{90}, and for CpdY in *M. smegmatis* it was around 10^{-8} (see Table 3.2) This is comparable with rifampicin, where spontaneous resistant mutants arise at a frequency of 10^{-8}, and significantly lower than that of isoniazid (10^{-6}) (data not shown). The mutation frequency experiment was not carried out using CpdX in *M. smegmatis* because the MIC_{50} value for the CpdX was more than 25μM and hence would require the use of very large amounts of the compound on *M. smegmatis*. The level of resistance acquired by the spontaneous resistant mutants are determined using the turbidimetry-based growth assay (Section 3.3).

Species	Compounds		Mutation Frequency		
	Name	Pressure	1	2	Average
M. bovis BCG	Compound X	$10 \times MIC_{50}$	1.5×10^{-8}	1.1×10^{-8}	1.3×10^{-8}
M. bovis BCG	Compound Y	$10 \times MIC_{50}$	4.9×10^{-9}	ND	ND
M. smegmatis	Compound Y	$10 \times MIC_{50}$	4.9×10^{-8}	ND	ND

Table 3.2: Mutation frequency experiment for CpdX and cpdY in wild-type *M. bovis* BCG and *M. smegmatis*. This table summarizes the results of the mutation frequency experiments. (ND denotes not done)

3.3. Characterization of *M. bovis* BCG and *M. smegmatis* compound-resistant mutants

CpdX-resistant BCG mutants were isolated and the level of resistance of each mutant strain were determined via turbidimetry-based growth assay. Moreover, to eliminate the possibility that the bacteria are resistant because of a mutation in one of the major transporter protein resulting in impermeability to drugs or mutation in a drug efflux system, we tested the mutants against a selection of frontline anti-tuberculosis drugs. Hypothetically, mutations in a transporter could either decrease the entry of drug or increase the efflux of drug out of the cell (De Rossi *et al*., 2006; specifically for *M. tuberculosis*: a transporter of the major facilitator superfamily - De Rossi *et al*., 2002; a transporter of the small multi-drug resistance family - De Rossi *et al*., 1998). In both cases, this leads to a decrease in the accumulation of the drug within the bacterial cell, thereby contributing to the increase in MIC_{50} value of mutants.

In order to decide which drugs are to be included in this panel, we tested 23 anti-bacterial drugs against wild-type BCG (see Table 3.3). From these, we chose streptomycin, moxifloxacin, rifampicin and ethambutol, which targets the protein synthesis, DNA coil relaxation, transcription and cell wall synthesis, respectively.

Colony purified CpdX-resistant BCG strains showed 6 to 14 times increase in the MIC_{50} values as compared to wild-type BCG (see Table 3.4). Similarly, the colony purified CpdY-resistant BCG mutants showed between 2 to 3 times increase in MIC_{50} value as compared to wild-type BCG. The MIC_{50} values only increased with the compounds and not with the standard drugs. For nearly all strains,

Drug	*M. bovis* BCG MIC_{50}	*M. smeg* MIC_{50}	Bacterial Target
		Aminoglycosides	
Amikacin	<0.156	0.79	30S ribosomal subunit
Gentamicin	1.26	0.70	30S ribosomal subunit
Imipenem	>20	>40	30S ribosomal subunit
Kanamycin	0.51	0.87	30S ribosomal subunit
Streptomycin	<0.156	<0.3	30S ribosomal subunit
		Quinolones	
Ciprofloxacin	0.21	0.51	DNA gyrase
Gatifloxacin	<0.156	<0.3	DNA gyrase
Moxifloxacin	<0.156	<0.3	DNA gyrase
Ofloxacin	>20	>40	DNA gyrase
		Rifamycins	
Rifabutin	<0.156	0.34	DNA-dependent RNA polymerase
Rifampicin	<0.156	6.11	DNA-dependent RNA polymerase
		Sulfonamides	
Sulfamethoxazole	1.64	>40	Dihydrofolate reductase
Trimethoprim	>20	19.61	Dihydrofolate reductase
		Others	
Ampicillin	>20	>40	Cell wall synthesis (Penicillins)
Azithromycin	0.47	2.43	Protein synthesis (Macrolides)
Ceftriaxone	6.56	>40	Cell wall synthesis (Cephalosporins)
Cycloserine	>20	>40	Cell wall synthesis
Ethambutol	2.51	1.07	Cell wall synthesis
Isoniazid	>20	>40	Cell wall synthesis
Linezolid	>20	0.84	Protein synthesis (Oxazolidinones)
Phosphomycin	>20	>40	Enolpyruvate transferase
Tetracycline	7.52	2.31	Translation (Tetracyclines)
Vancomycin	0.17	1.14	Cell wall synthesis

Table 3.3: The MIC_{50} values for 23 standard drugs against wild-type *M. bovis* BCG and *M. smegmatis*. (ND denotes not done and all MIC_{50} values are expressed in μM)

the MIC_{50} value for ethambutol, moxifloxacin, streptomycin and rifampicin remained comparable with the wild-type strain. Similarly, although colony-purified *M. smegmatis* CpdY-resistant strains showed MIC_{50} values of over 40μM for CpdY, the mutant strains remained sensitive to the other four drugs.

The data from this set of experiments suggest that these mutants have acquired specific resistance to the test compound through random mutations, presumably in its biological target. Hence, after successfully isolating and characterizing drug-resistant

mutants, we proceeded to characterize these mutants using genetic and proteomic approaches (Section 3.4-Section 3.10).

M. bovis BCG Strains	Minimum Inhibitory Concentration$_{50}$ (MIC$_{50}$ in µM)				
	Compound X	**Ethambutol**	**Moxifloxacin**	**Streptomycin**	**Rifampicin**
wild-type	1.73	2.28	0.070	0.075	0.0062
1	23.24	S	S	S	R (0.031)
2	19.69	S	S	S	S
6	16.51	S	S	S	S
7	10.44	S	S	S	S
10	15.60	S	S	S	S
13	20.79	S	S	S	S

M. bovis BCG Strains	Minimum Inhibitory Concentration$_{50}$ (MIC$_{50}$ in µM)				
	Compound Y	**Ethambutol**	**Moxifloxacin**	**Streptomycin**	**Rifampicin**
wild-type	0.36	2.28	0.070	0.075	0.0062
1	0.625-1.25	S	S	S	S
2	1.07	S	S	S	S

M. smegmatis Strains	Minimum Inhibitory Concentration$_{50}$ (MIC$_{50}$ in µM)				
	Compound Y	**Ethambutol**	**Moxifloxacin**	**Streptomycin**	**Rifabutin**
wild-type	1.47	1.00	0.16	0.19	0.65
1	>40	S	S	S	S
2	>40	S	S	S	S
4	>40	S	S	S	S
9	>40	S	S	S	S

Table 3.4: Cross-resistance study of CpdX and CpdY-resistant *M. bovis* BCG and *M. smegmatis* mutants. S denotes susceptibility to drug and R means resistance to drug; Susceptibility is defined as having an MIC$_{50}$ value very similar to wild-type with a very slight variation.

3.4. Sequencing of *M. bovis* BCG and *M. smegmatis* mutant

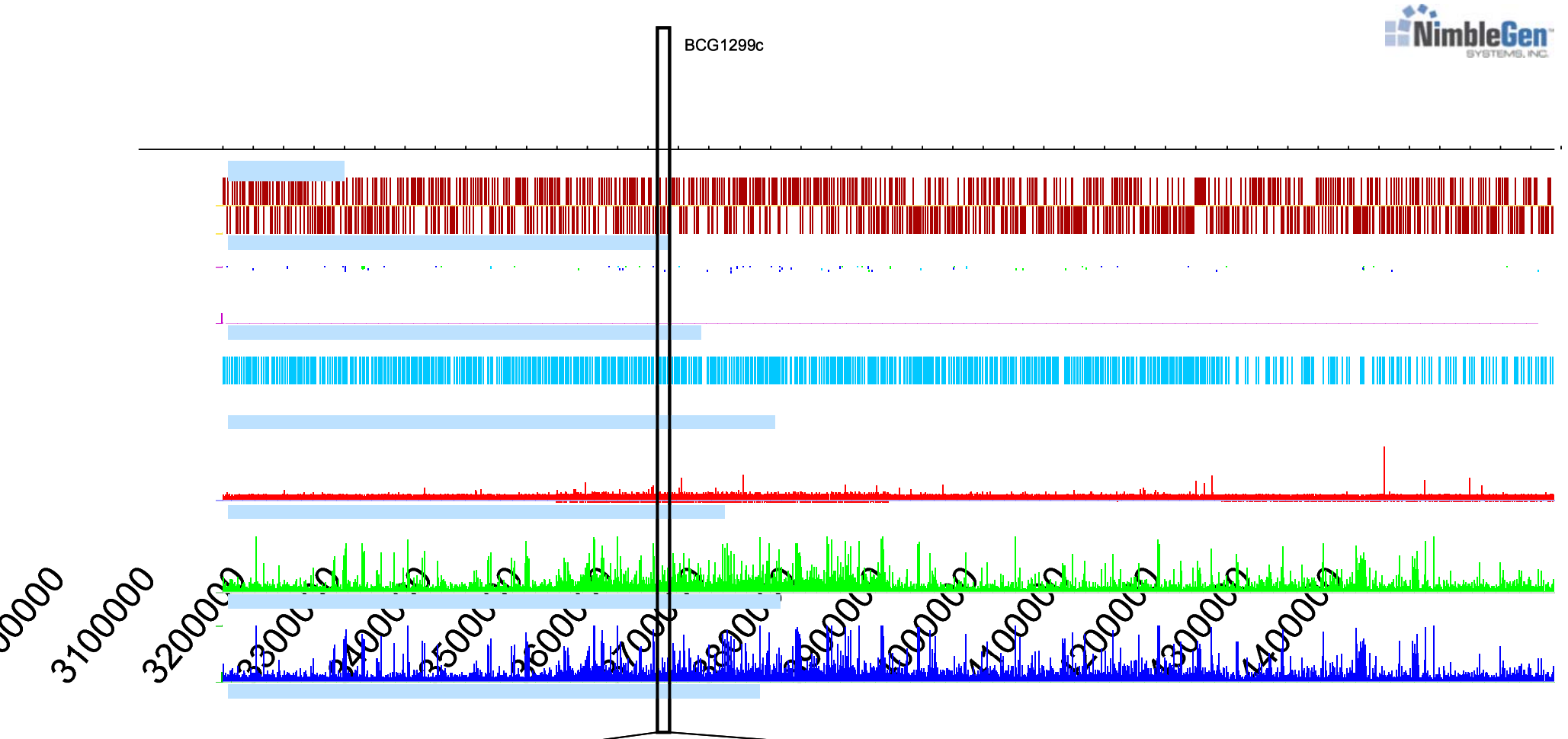
Without any prior indication or the knowledge of the biological target, the most promising option for target identification is to sequence the whole genome of several mutant strains to determine the SNP(s) that result in resistance. The open reading frame (ORF) in which the SNP(s) are located would very likely be the biological target of the compound or at the very least predicts the mechanism of resistance. The feasibility of this novel whole genome sequencing approach was elegantly demonstrated by the identification of ATP synthase to be the target for a diarylquinoline drug – designated R207910 (Andries *et al*., 2005).

As an alternative approach to whole genome sequencing, we chose to utilize a chip-based sequencing technology marketed by NimbleGen Inc (United States). One CpdX-resistant mutant of BCG (designated strain 10) was sent for comparative genome sequencing (CGS). A single SNP (Single Nucleotide Polymorphism) was present in *BCG1299c* (*Rv1239c*) open reading frame that codes for a putative magnesium and cobalt transport protein (CorA). The genetic mutation resulted in a non-synonymous substitution of alanine with serine at amino-acid position 317. CpdY-resistant mutants were not sent for comparative genome sequencing, however given the structural similarity between CpdX and CpdY, we believed that the *corA* gene in CpdY resistant mutants should have mutations in the *corA* gene as well (this was true for BCG mutants)

With this data, we designed primers to amplify and sequence the *corA* gene from CpdX and CpdY-resistant mutants of BCG. In all strains, except one BCG mutant, we found SNPs in the *corA* gene. Four individual types of genetic mutations were

mapped to the *corA* gene that resulted in a G299S, M300V, A313S and A317S substitution at the protein level, in both CpdX and CpdY-resistant mutants. The location of these amino acids in the folded protein is unknown and cannot be compared with the recently crystallized CorA protein from *Thermotoga maritima*, because the *corA* gene of *T. maritima* shares less than 30% homology with the *M. tuberculosis corA* (Eshaghi *et al*., 2006; Maguire, 2006; Lunin *et al*., 2006). No SNPs in the *corA* gene were isolated in strain 7 of CpdX-resistant BCG mutant. This result was confirmed with two independent PCR and sequencing reactions. Thus, it would be interesting to sequence this particular strain *via* comparative genome sequencing to isolate another SNP that is also capable of conferring resistance to CpdX. Unexpectedly, the *M. smegmatis* CpdY-resistant mutants had no mutations in *corA* as well as the *corA*-like gene (as annotated by The Institute for Genomic Research, TIGR), thus suggesting another resistance determinant.

In summary, a mutation within CorA at specific amino acid residues in BCG is able to confer resistance to CpdX and CpdY. However, in *M. smegmatis*, no mutation was observed in the *corA* or the *corA*-like gene.



Seq ID	Nucleotide Change			Confidence	Probability	Type	Highest Rank	Amino Acid Change			Locus Information					
	Genome Position	Ref	SNP					Type	Orginal	SNP	Gene Position	Designation	Start	End	Strand	Product
M. bovis BCG	1411816	C	A	HIGH	0.995	coding	10003.2	NON	A	S	949	BCG1299c	1411664	1412764	Non-coding	CorA

Figure 3.1: Results of Comparative Genome Sequencing. In summary, *corA* (*BCG1299c*) had a SNP at base pair 949 (which translates to A317S missense mutation at the protein level. For a complete explanation of the labels and graphs please see Appendix I: Comparative Genome Sequencing).

M. bovis BCG Strains	Compound X MIC_{50}	CorA Mutations Mapped
wild-type	1.73	No Mutation
1	23.24	Met_{300} --> Val_{300}
2	19.69	Met_{300} --> Val_{300}
6	16.51	Gly_{299} --> Ser_{299}
7	10.44	No Mutation
10	15.60	Ala_{317} --> Ser_{317}
13	20.79	Ala_{317} --> Ser_{317}

M. bovis BCG Strains	Compound Y MIC_{50}	CorA Mutations Mapped
wild-type	0.36	No Mutation
1	0.625-1.25	Ala_{313} --> Gly_{313}
2	1.07	Ala_{317} --> Ser_{317}

M. smegmatis Strains	Compound Y MIC_{50}	CorA and CorA-like Mutations Mapped
wild-type	1.47	No Mutation
1	>40	No Mutation
2	>40	No Mutation
4	>40	No Mutation
9	>40	No Mutation

Table 3.5: CorA mutations found in all the sequenced CpdX and CpdY-resistant *M. bovis* BCG and *M. smegmatis* mutants. (all MIC_{50} values are expressed in μM)

3.5. Expression of *corA* in *M. bovis* BCG

As noted in the introduction, one of the methods for determining if a gene is the biological target of a compound is to express the mutant copy of that gene in wild-type bacteria and determine if the resistance is transferred into the wild-type bacteria. In addition, the expression of the wild-type gene in mutant bacteria will allow us to determine if the mutant phenotype is dominant.

To this goal, we cloned separately the mutant and wild-type copy of *corA* gene along with a 500bp upstream region predicted to contain the promoter into the pMV306

vector (see Figure 3.6A). Recombinant BCG were selected out on Middlebrook 7H10 plates with 25μg/ml kanamycin alone or 25μg/ml kanamycin with 5μM or 10μM of CpdX. We isolated recombinants at all concentration of CpdX for BCG transformed with the mutant *corA* construct. Despite a similar transformation efficiency, no wild-type *corA* expressing recombinants that displayed resistance to CpdX were isolated.

Upon introduction of a compound-resistant copy of *corA* gene into wild-type BCG, the bacilli showed an increase in MIC_{50} values for CpdX to nearly over 14μM (this is comparable with the compound-resistant mutant itself). However, expression of neither wild-type *M. smegmatis corA* nor wild-type BCG *corA* resulted in such an increase in resistance to CpdX (MIC_{50} values of less than 4μM). As a control, the empty vector control (pMV306 only) did not cause a change in MIC_{50} when introduced into wild-type BCG. Conversely, we found that expression of wild-type *corA* carried on a plasmid in a CpdX-resistant background did not cause the mutant strains to become sensitive to CpdX.

Taken together, this data suggests that CorA could be the target of CpdX or at least mediates a non-target specific mechanism of resistance. The wild-type BCG has a compound-susceptible CorA that is blocked by CpdX thereby resulting in a low MIC_{50} value. However, upon expression of the compound-resistant CorA, the wild-type BCG develops a mutant phenotype, despite CpdX still inhibiting the wild-type CorA. The only logical explanation for this is that the mutant CorA gene product is functional and has a dominant phenotypic effect. In other words, the mutant CorA channel provides another avenue through which magnesium and other essential divalent cations can still enter the bacterium even in the presence of CpdX. For CorA to be the target of CpdX,

it would be required to establish the essentiality of magnesium transport through CorA channel in BCG.

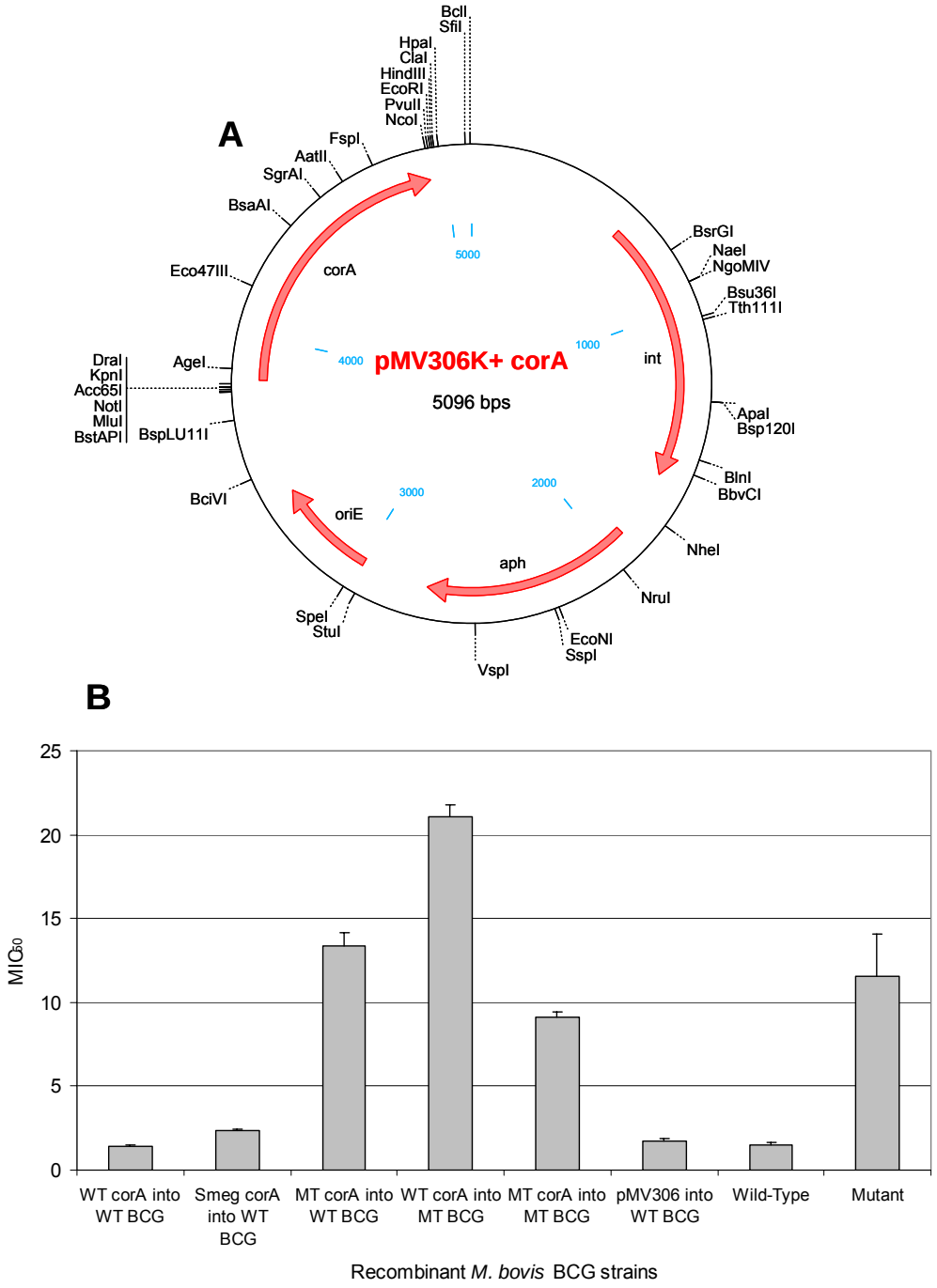


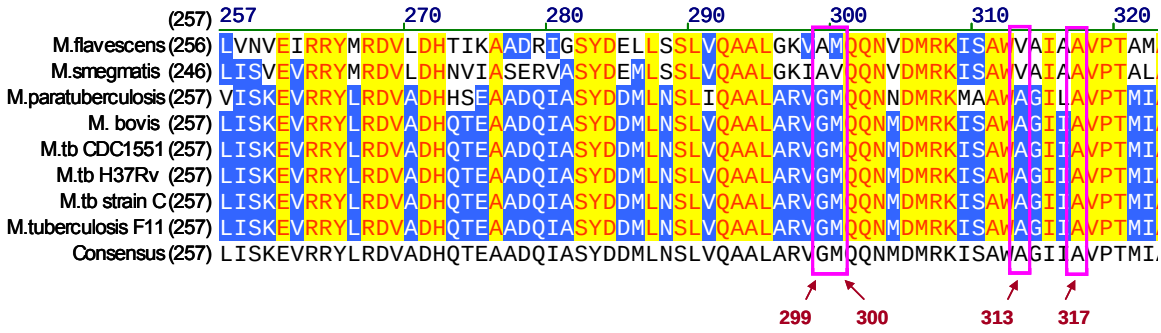
Figure 3.2: Expression of *corA* gene in wild-type *M. bovis* BCG. **A** shows a plasmid map of wild-type *corA* cloned into pMV306-Kan; all other constructs are similar except for the '*corA*' region. **B** shows the result of expression of the variants of *corA* gene into wild-type *M. bovis* BCG and mutant *M. bovis* BCG. (MIC_{50} values are expressed in µM; error bars indicate standard error; WT denotes wild-type; MT denotes mutant; Smeg denotes *M. smegmatis corA*)

3.6. Sequence alignment of the *corA* gene from various mycobacteria

The rationale behind the alignment of the corA gene of various species/strains belonging to the Mycobacterium complex is that we wanted to determine if there was any genetic explanation for the high MIC_{50} of CpdX in *M. smegmatis*, but low MIC_{50} in BCG. In other words, we wanted to determine if *M. smegmatis* had a 'natural mutation' that conferred resistance to CpdX but not CpdY.

We aligned the publicly available sequences of a selected number of species in the mycobacterial complex to determine the homology of *corA* within this group. We found that *M. smegmatis* had different residues at 299, 300 and 313 as compared to *M. bovis* and *M. tuberculosis* sequence of CorA. These 'natural mutations' are G299A, M300V and A313V, respectively. We observed that the intrinsic resistance of *M. smegmatis* to CpdX could be due to a natural mutation at position 300 from a polar amino acid methionine to a non-polar valine. It is of particular interest to note that M300V is one of the substitutions that BCG employs in order to gain resistance to CpdX.

A



B

Amino Acid Position	*M. bovis* BCG		*M. smegmatis*		*M. bovis BCG* mutant	
299	G	—H	A	—CH_3	S	—$CH(NH_2)$—COOH
300	M	—CH_2—CH_2—S—CH_3	V	—CH_2—OH	V	—CH_2—OH
313	A	—CH_3	V	—CH_2—OH	G	—H
317	A	—CH_3	A	—CH_3	S	—$CH(NH_2)$—COOH

Figure 3.3: Comparison of the *corA* sequences from various mycobacteria. **A** shows the multiple alignment of the annotated *corA* sequences of these species. Of particular interest, note amino acids 299, 300, 313 and 317 (more discussion in text; M.tb denotes *M. tuberculosis*). **B** shows the amino acid substitutions in the *M. bovis* BCG mutants and the 'natural mutation' in the *M. smegmatis* sequence from the wild-type *M. bovis* BCG. The side-chain of these amino acids are also shown.

3.7. Proteomics of Compound X-resistant *M. bovis* BCG mutant and wild-type *M. bovis* BCG with and without Compound X treatment

A proteomics approach was taken to identify the global differences in protein expression profile of wild-type BCG versus mutant BCG in the presence or absence of CpdX. This involved the separation of proteins according to their pI (isoelectric point) and MW (molecular weight). Following silver staining, differences in staining patterns were determined and the protein spots identified by LC-MS. The trypsin-digested proteins were compared to the known mycobacterial complex database using MASCOT and X! Tandem. The comparison of compound exposure versus non-exposure in wild-type bacilli could give us an indication as to which pathway is affected by the compound. For example, a series of enzymes are involved in the production of an essential chemical/nutrient. If the last enzyme was blocked by the compound then there could be an up-regulation of the protein expression of some (if not all) up-stream enzymes in the biosynthetic pathway. Conversely, these proteins would not be up-regulated in the mutant following drug exposure because the mutant is resistant to compound.

We identified nine out of the ten spots that were sent for LC-MS analysis (see Table 3.6). These were superoxide dismutase (possibly increased with treatment in wild-type and mutant), a probable ABC-type transporter ATP-binding protein (increased with treatment in wild-type and absent in mutant), gene product of *Rv1109c* (possibly increased with treatment in wild-type), gene product of *Rv2699c* (increased with treatment in wild-type), thioredoxin peroxidase (possibly increased with treatment in wild-type), phosphoribosyl isomerase A (slight increased with treatment in wild-type) and a probable peptidyl-isomerase cis-trans isomerase A (possibly decreased with

treatment in wild-type). Note that these results are preliminary, however, the results do give proteome profiles of wild-type BCG and mutant BCG that can be used for comparison with future studies.

Spot	Molecular Weight	Wild-type *M. bovis* BCG			Mutant *M. bovis* BCG		LC-MS Results		Essentiality	Heat-shock	Starvation	Essentiality during Infection
		0μM	0.6μM	1.2μM	0μM	1.2μM	Accession Number	Function/Identity				
1	19kDa	++/+++	-	+++	+++	+++	Rv3846	Superoxide dismutase	-	Up-regulated	Up-regulated	-
2	19kDa	+++/+	+++	+	+++	+++	Rv0009	Probable peptidyl-isomerase cis-trans isomerase A	Non-essential	Up-regulated	Up-regulated	-
3	29kDa	++/-	+++	+++	-	-	Rv1463	ATP binding protein ABC transporter	Essential	-	Up-regulated	-
4	23kDa	++/-	+++	+++	-	-	Rv1109c	Hypothetical protein	Non-essential	-	-	Essential
5	23kDa	+++/-	+++	+++	+++	+++	Rv1109c	Hypothetical protein	Non-essential	-	-	Essential
6	25kDa	+/-	++	+++	+++	+++	Rv2699c	Hypothetical protein	Non-essential	Up-regulated	-	-
7	-	+/-	++	+++	+++	+++	-	-	-			
8	26kDa	+/-	++	+++	+++	+++	Rv1603	Phosphoribosyl isomerase A	Essential	-	-	-
9	17kDa	+++/-	+++	+++	+++	+++	Rv1932	Probable thioredoxin peroxidase	Non-essential	-	-	-
10	19kDa	-/-	-	-	-	++	Rv3846	Superoxide dismutase	-	Up-regulated	Up-regulated	-

Table 3.6: Summary of 2D gel electrophoresis experiments. Differences in intensity level (relatively to surrounding spots) and the approximate size are indicated in this table for each 2D gel spot. The identity of spot according to LC-MS and the accession number, function and essentiality of each protein can also be found in this table. 2D gel electrophoresis was done after three days exposure to CpdX (+++ high intensity, ++ moderate, + low and – no detectable expression; compound X concentrations used are also given). The last four columns on the right shows the current published knowledge about the essentiality of each open reading frame during infection and under *in vitro* conditions; in addition, the involvement of each open reading frame in starvation and heat shock responses are also noted (Sassetti *et al.*, 2003; Sassetti and Rubin, 2003; Stewart *et al.*, 2002; Betts *et al.*, 2002)

A

Sequence Coverage	Protein	MS/MS Sample	Prob
	Hypothetical protein	DataNovartis 4.wiff (F006708)	100%
	Hypothetical protein	DataNovartis 5.wiff (F006709)	100%
	Hypothetical protein	DataNovartis 6.wiff (F006710)	100%
	PROBABLE CONSERVED ATP-BINDING PROTEIN ABC TRANSPORTER (ABC transporter, ATP-binding protein)	DataNovartis 3.wiff (F006707)	100%
	Phosphoribosyl isomerase A (1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino] imidazole-4-carboxamide isomerase)	DataNovartis 8.wiff (F006712)	100%
	Probable peptidyl-prolyl cis-trans isomerase A (EC 5.2.1.8) (PPIase A) (Rotamase A)	DataNovartis 2.wiff (F006706)	100%
	Superoxide dismutase	DataNovartis 1.wiff (F006705)	100%
	Superoxide dismutase	DataNovartis 10.wiff (F006704)	100%
	Tpx protein	DataNovartis 9.wiff (F006713)	100%

B

O53154_MYCTU (100%), 28823.1 Da
PROBABLE CONSERVED ATP-BINDING PROTEIN ABC TRANSPORTER (ABC transport
9 unique peptides, 13 unique spectra, 13 total spectra, 126/266 amino acids (47% cover

MTILEIKDLH VSVENPAEAD HEIPILRGVD LTVKSGETHA
LMGPNGSGKS TLSYAIAGHP KYHVTSGTIT LDGADVLAMS
IDERARAGLF LAMQYPVEVP GVSMSNFLRS AATAIRGEPP
KLRHWVKEVK AAMAALDIDP AFAERSVNEG FSGGEKKRHE
ILQLELLKPK IAILDETDSG LDVDALRVVS EGVNRYAESQ
HGGILLITHY TRILRYIHPE YVHVFVGGRI VESGGSELAD
ELDQNGYVRF SPASGRYPHQ PAPTGA

C

Sequence	Prob	Mascot Ion score	Mascot Identity score	X! Tandem
(K)IAILDETDSGLDVDALR(V)	95%	118.2	22.6	7.89
(M)TILEIK(D)	95%	37.9	25.6	
(R)IVESGGSELADELDQNGYVR(F)	52%			1.66
(R)IVESGGSELADELDQNGYVR(F)	95%			3.28
(R)SVNEGFSGGEK(K)	95%	62.3	16.7	3.17
(K)DLHVSVENPAEADHEIPILR(G)	95%	67.7	22.9	8.54
(R)VVSEGVNR(Y)	19%	20.6	25.1	1.22
(K)SGETHALMGPNGSGK(S)	95%	41.5	19.2	6.57
(K)SGETHALMGPNGSGK(S)	95%	38.9	20.4	4.96
(K)AAMAALDIDPAFAER(S)	95%	76.3	23.1	5.77
(K)AAMAALDIDPAFAER(S)	95%	76.3	22.5	7.24
(K)STLSYAIAGHPK(Y)	95%	45.7	23.8	1.00
(K)STLSYAIAGHPK(Y)	95%	45.5	23.2	3.77
(R)YPHQPAPTGA(-)	95%	35.3	22.8	1.41

D

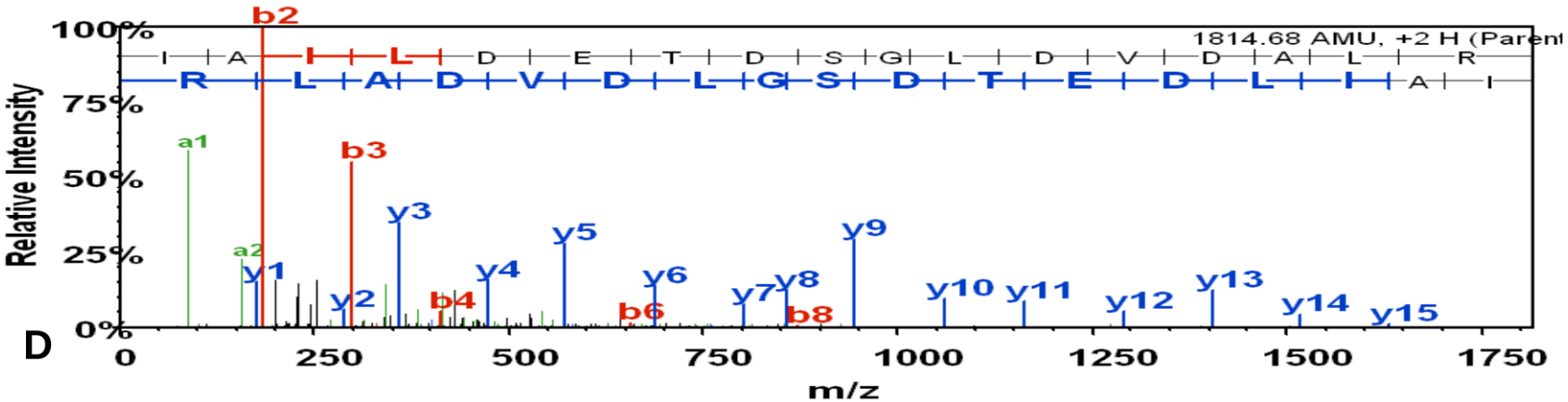


Figure 3.4: Example of protein identification by LC-MS; with analysis using MASCOT and X! Tandem. **A** shows the output file for all the protein spots that were identified. **B** shows the protein sequence of the ATP-binding ABC transporter (Rv1463) and the regions that positively identified by the LC-MS. **C** shows all the peptides that were identified by LC-MS from the ATP-binding ABC transporter (Rv1463). **D** shows the raw fragmentations patterns from the LC-MS.

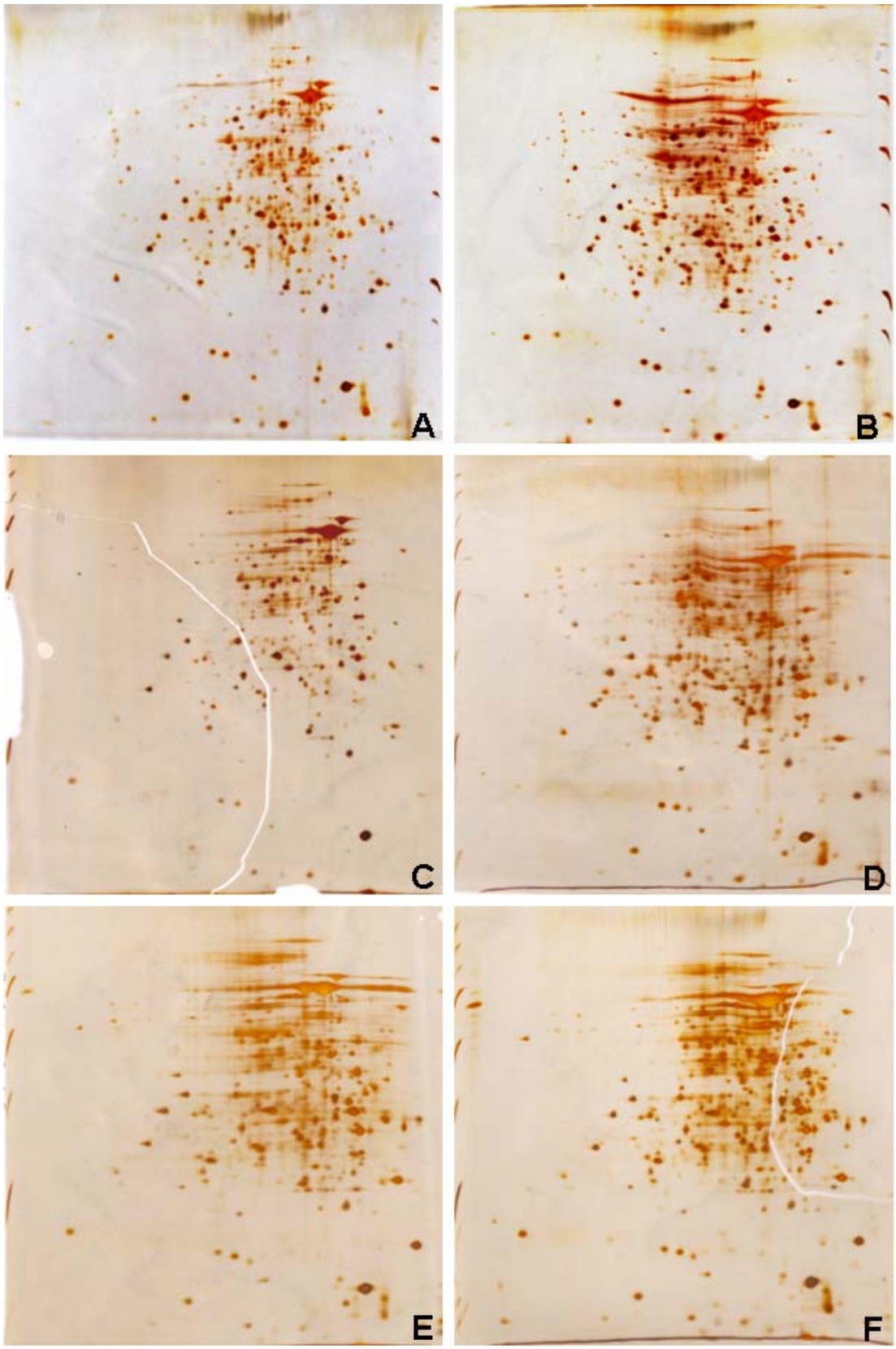


Figure 3.5: 2D gel electrophoresis (18cm pH 4-7 strips). **A, B, C** and **D** are 2D gels of total wild-type *M. bovis* BCG proteins, while **E** and **F** are from CpdX-resistant *M. bovis* BCG mutant. **B, D** and **F** are without CpdX treatment. **A** was treated with 0.6µM of CpdX for 3 days. **C** and **E** were treated with 1.2µM of CpdX for 3 days.

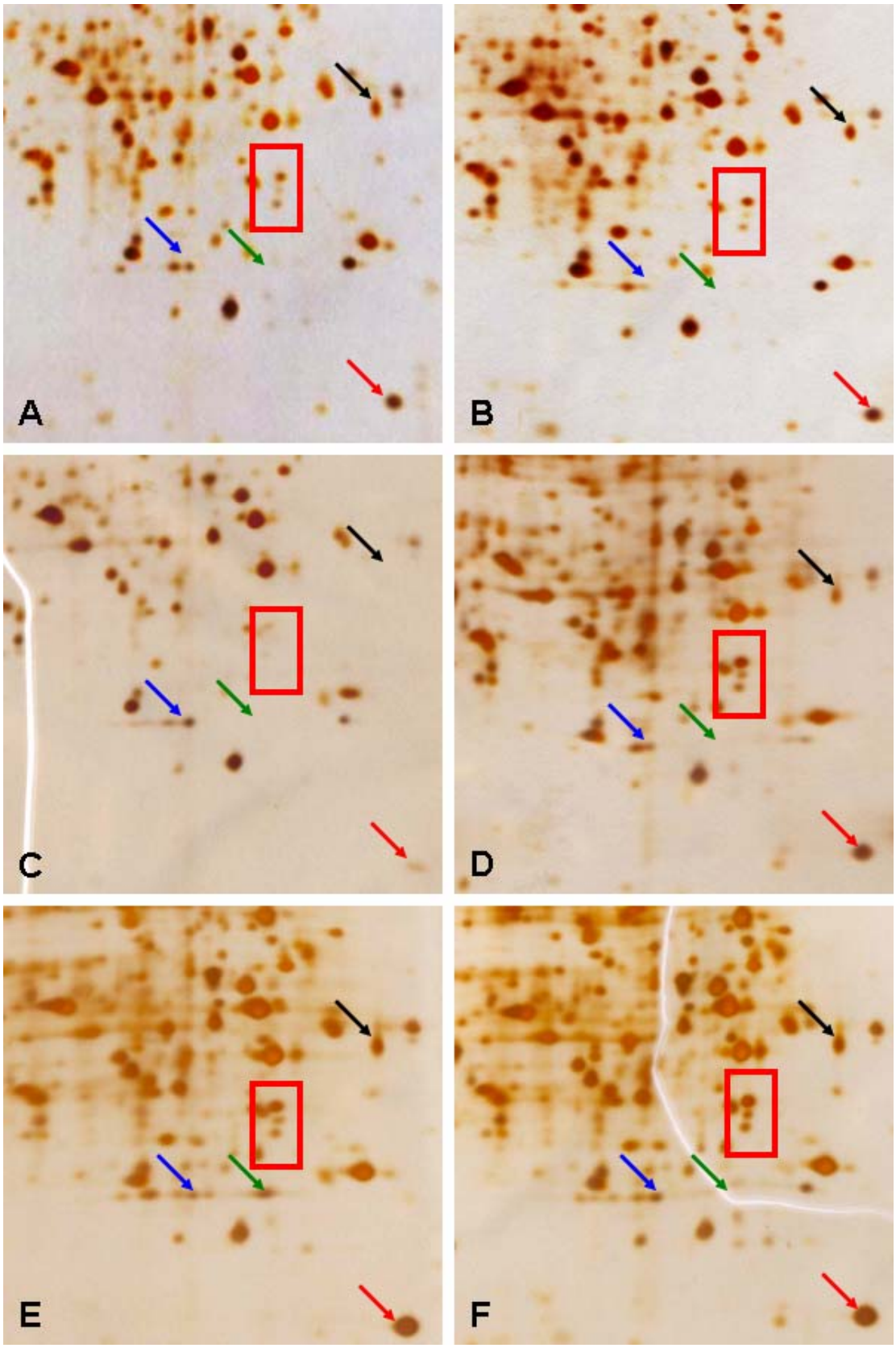


Figure 3.6: 2D gel electrophoresis (18cm pH 4-7 strips) – close up 1. A-F labels are as in Figure 3.14. Red arrow: spot number 9; Blue arrow: spot number 1; Green arrow: spot number 10; Black arrow: spot number 5; Red box (top to bottom): spot numbers 6-8.

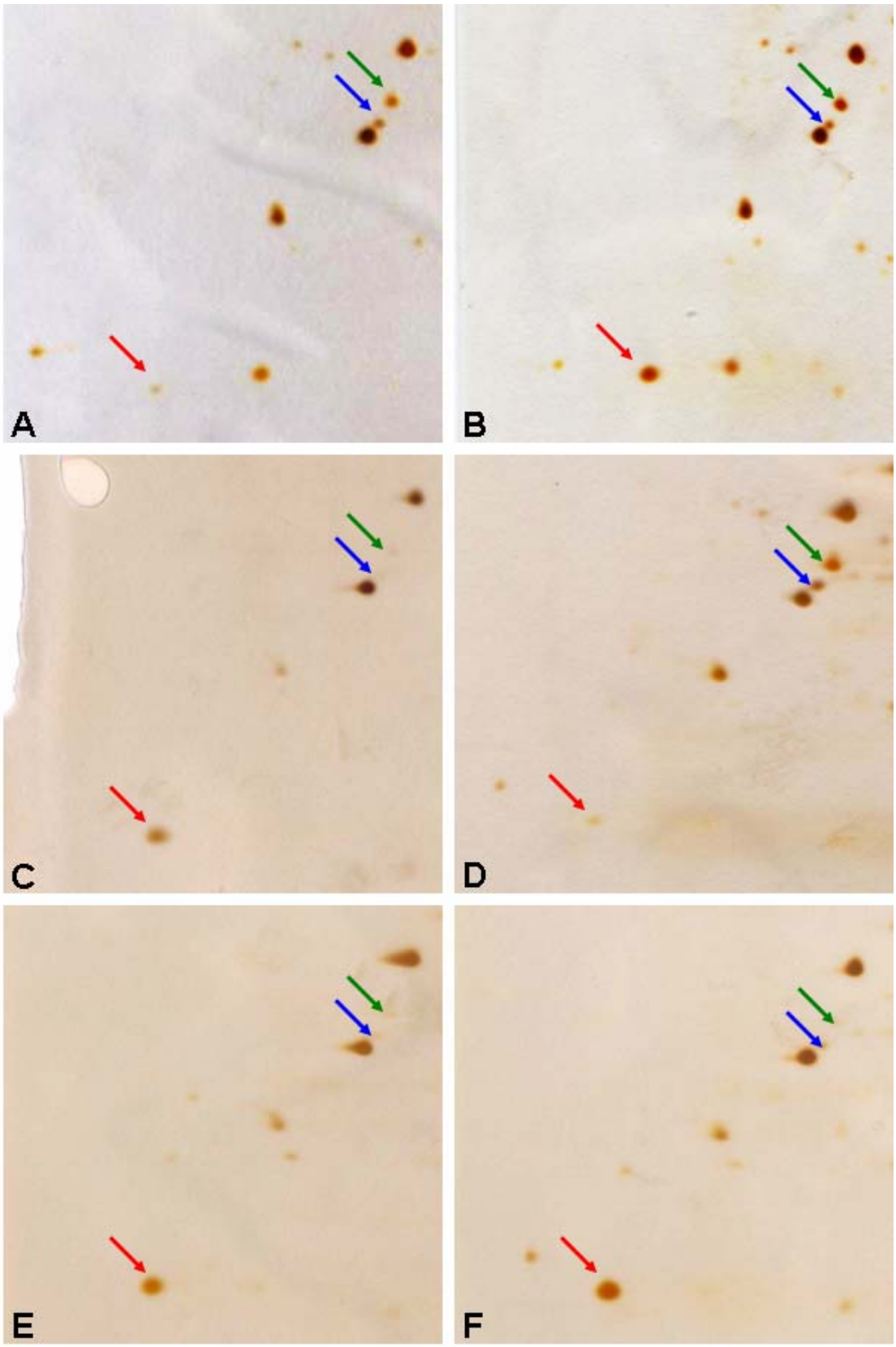


Figure 3.7: 2D gel electrophoresis (18cm pH 4-7 strips) – close up 2. **A-F** labels are as in Figure 3.14. Red arrow: spot number 2; Blue arrow: spot number 4; Green arrow: spot number 3.

Chapter Four: Discussion

4. Discussion

4.1. Physiological role of CorA

Magnesium is an essential cofactor of many enzymes involved in central metabolic pathways that are required to maintain bacterial viability. This includes reactions involving ATP as well as tRNA synthesis in both prokaryotic and eukaryotic systems (Cowan, 1998; Quigley *et al*., 1978; Szegedy and Maquire, 1999). Moreover, it has also been shown that magnesium contributes to the stability of biological membranes (Szegedy and Maquire, 1999). Hence, bacteria have developed multiple high-affinity magnesium transport systems to import this important divalent cation. In mycobacteria, these transport systems include the magnesium and cobalt transport transmembrane protein (CorA encoded by *Rv1239c*), magnesium transport P-type ATPase C (MgtC encoded by *Rv1811*) and magnesium transport transmembrane protein (MgtE encoded by *Rv0362*). Of all the known magnesium transport systems, CorA is thought to be the universal and most widely used system in both domains of bacteria as well as archaea.

In *Salmonella typhimurium*, it has been shown that ion transport through CorA is bidirectional (Smith and Maguire, 1993). The membrane potential of over -120mV creates an electrochemical gradient that drives magnesium inwards through CorA channel. CorA only mediates the outward flow of magnesium ions only when extracellular [Mg^{2+}] (magnesium concentration) exceeds 1mM, though the significance of this efflux is uncertain (Smith and Maguire, 1998). However, the P-type ATPase magnesium transporters (such as *mgtC*) are known to mediate influx only and are expressed under conditions of low extracellular [Mg^{2+}] (Smith and Maguire, 1998). In

M. tuberculosis, *mgtC* is not necessary for the invasion of macrophages but is required for survival inside these phagocytic cells (Buchmeier *et al*., 2000).

The *corA* gene from *M. tuberculosis* Erdman has been shown to be non-essential as it could be deleted in an earlier study (Sareen *et al.* 2003). This data is further supported by Himar1-based transposon mutagenesis studies in *M. tuberculosis* H37Rv in which *corA* was found to be non-essential (Sassetti *et al*., 2003). Given the presence of other cation transporters, mycobacteria can still survive by relying on other magnesium transporters under *in vit*ro conditions where levels of this cation are high (although *mgtC* is expressed only under low extracellular [Mg^{2+}] conditions).

In 2006, two groups expressed and crystallized the homo-pentameric closed-conformation CorA channel of *T. maritima* (Eshaghi *et al*., 2006; Lunin *et al*., 2006). This transporter has a narrow entrance (5Å) that opens up into the cytoplasm to approximately 20Å. Magnesium ions would have to travel at least 40Å to enter the cell and it was also shown that the channel pore comprises of two cavities created by three constricted regions. These three constricted regions are formed by amino acids residues B_{314}, M_{302}/T_{305} and L_{294}/M_{291}, respectively. The CorA transporter can transport around 10 million ions per minute. This is considerably slower than potassium channels, which can allow around 100 millions ions to pass through per second. This could be due to the fact that the CorA channel comprises of five subunits whereas magnesium needs to be hexacoordinated for transport. The fact that hydrated magnesium has the largest diameter of all biological ions (~5Å) and that the similarly sized but 3^{+}-charged cobalt(III)hexaammine is an inhibitor of CorA, it is clear that CorA does not rely on the charge nor hydrated size in its selectivity to transport ions

(Kucharshi *et al*., 2000; Eshaghi *et al*., 2006; Lunin *et al*., 2006). The authors suggest that, since magnesium has a hydrated radius about 350 times larger than its atomic radius, a selective filter based on this size difference would be very effective. Moreover, CorA is also known to transport cobalt and nickel ions, but not manganese, presumably, because manganese ion is significantly larger than cobalt, nickel and magnesium. In addition, the cobalt(III)hexaammine is about the same size as a fully-hydrated magnesium and yet cannot pass through the CorA channel. It should be noted that the three-dimensional structure could still be relevant for other prokaryotic and eukaryotic species because, despite little sequence homology, the secondary structure of many species are quite similar to the *T. maritima* CorA (Figure 4.2). The CorA of *M. tuberculosis* shows 28% identity at the protein level with *T. maritima* CorA (accession number NP_228371).

The key question though is whether magnesium transport via CorA is essential under physiological conditions (i.e. inside the human host) and this remains to be determined. In *Helicobacter pylori*, a *corA* gene deletion mutant was isolated and was unable to grow in media without magnesium (it requires supplementation with at least 20mM of magnesium). Given the low magnesium concentration within the human host, the authors concluded that CorA is essential for the survival of this gastric pathogen under *in vivo* conditions (Pfeiffer *et al*., 2002). We demonstrate here that CpdX-resistant mutants of BCG showed slightly slow growth (as measured via turbidimetry) than the parental strain (data not shown). In addition, it would be interesting to see if the requirement of magnesium for growth increases in BCG and *M. tuberculosis corA* knock-out mutants.

4.2. Is CorA the mode of entry or the target for CpdX and CpdY?

Since CorA is a channel to transport various ions, including cobalt and magnesium, two hypotheses were developed in regards to, (1) CorA being the actual drug target of CpdX or (2) CorA merely serves as a channel through which the compound enters the bacteria. In addition, it is also possible that the CorA protein is neither the target nor the entry channel and that the mutation in CorA mediates a non-target specific mechanism of resistance.

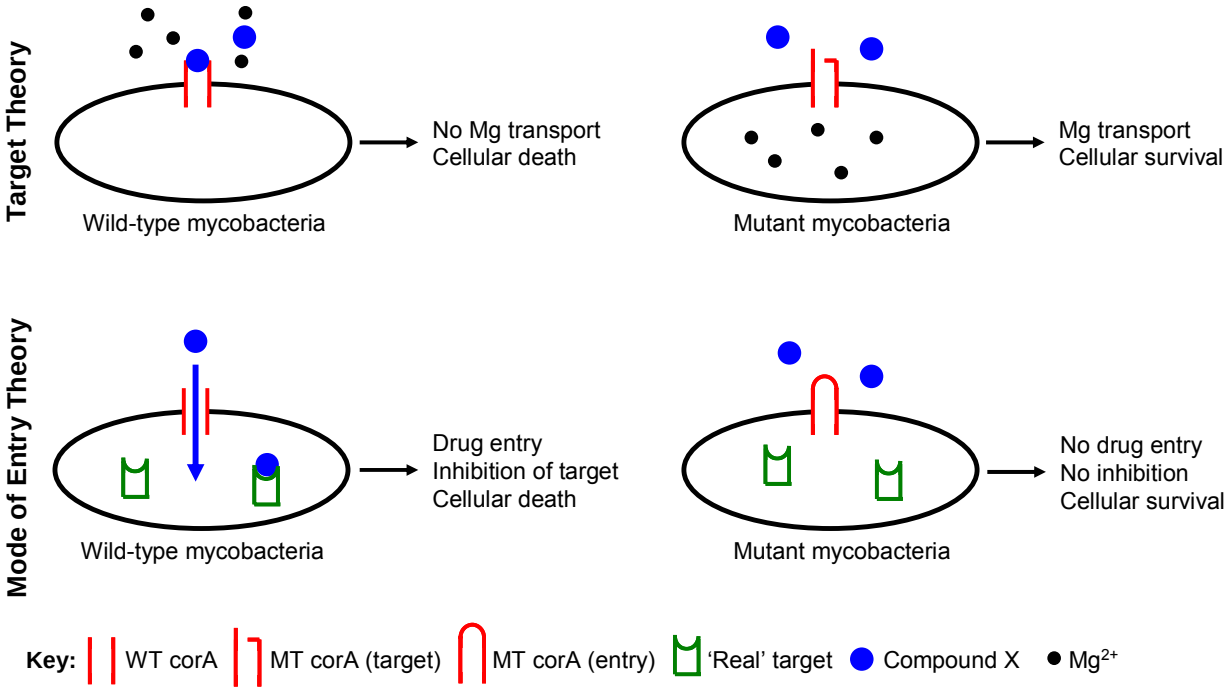


Figure 4.1: Target theory versus mode of entry theory. This figure highlights the way in which the wild-type strains are killed by compound X and how they could escape the bactericidal effect of the compound by mutating the **CorA** channel according to the target theory and the transporter theory.

4.2.1. Target theory

In this theory, we hypothesize that CpdX blocks the CorA channel and hence inhibits the entry of various divalent cations, including primarily magnesium but also cobalt and nickel. For this theory to be true, we have to prove that magnesium transport via *corA* is an essential process for mycobacterial growth *in vitro* and survival in the

human host. In addition, we also have to show that the exposure of mycobacteria to CpdX results in reduced magnesium or cobalt transport.

As shown earlier, the expression of mutant *corA* gene driven by its own promoter on a plasmid confers resistance to CpdX in wild-type BCG. In contrast, expression of the wild-type BCG *corA* or the wild-type *M. smegmatis corA* gene had no effect on this phenotype. Similarly, expression of wild-type *corA* in the BCG mutant strain did not result in the bacteria reverting to sensitivity. Therefore, it can be concluded that the mutant phenotype is dominant in the presence of CpdX because, as long as there is a functional *corA* gene the bacilli will grow unimpeded. This evidence is supportive of the theory that CorA is the biological drug target of CpdX.

To establish the essentiality of magnesium transport, we will have to develop a functional assay to measure the intracellular [Mg^{2+}] in BCG to determine if CpdX affects the movement of magnesium into and out of the cells. We could utilize the magnesium-specific fluorescent dye mag-fura 2 (acetoxymethyl ester form) to measure free magnesium levels with and without the treatment of BCG cells with CpdX (Froshauer *et al*., 2004). This protocol for the measurement of magnesium levels has traditionally been applied to measure concentrations of free magnesium within cells or organelles of eukaryotic cells; and more recently in *Salmonella enterica* (Froshauer *et al*., 2004). We would perform this experiment with five strains of bacteria, wild-type BCG, CpdX-resistant BCG, *corA* deletion mutant of BCG and *corA* knock-out BCG complemented with either wild-type *corA* or mutant *corA* gene on a plasmid. With each strain, we would measure the effect of addition of extracellular [Mg^{2+}] with or without the addition of CpdX.

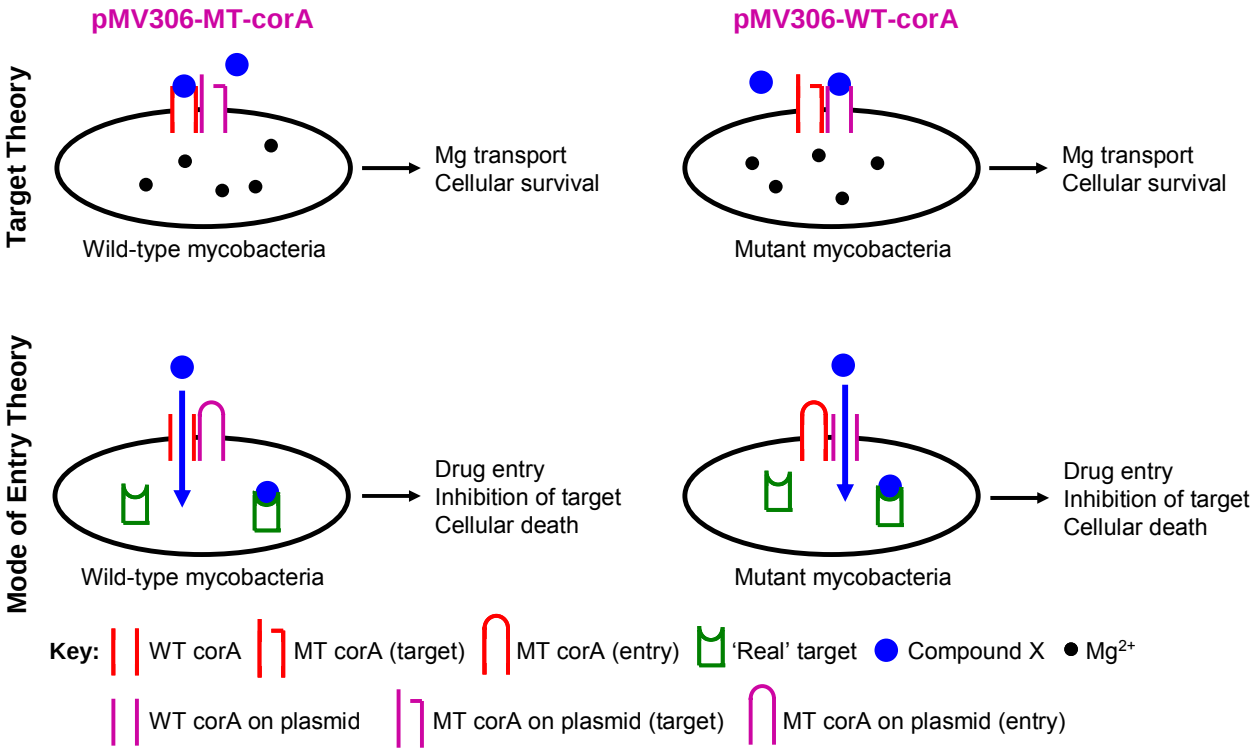


Figure 4.2: Target versus Transporter theory when *M. bovis* BCG is complemented with mutant corA gene or the wild-type *corA* gene. (WT denotes wild-type; MT denotes CpdX-resistant mutant)

However, we can delete the *corA* gene under laboratory conditions in *M. tuberculosis* and BCG (suggesting the gene is non-essential under the *in vitro* conditions) (data not shown) (Sareen *et al*., 2003). Nevertheless, the mere growth of this mutant in laboratory media may not reflect the importance of this gene product for intracellular survival in the host macrophages. To this end, we will be creating a BCG *corA* deletion mutant and test its ability to grow in macrophages and mice. The growth phenotype of the *corA* mutant in these models will be compared with that of wild-type, to determine if *corA* gene is important for *in vivo* survival. Nevertheless, due to the high bactericidal activity of the compounds in vitro, we believe that the biological target should be essential under *in vitro* conditions (i.e. the gene cannot be deleted). This suggest that CorA, although is a mediates resistance, could not be the biological target of CpdX and CpdY.

4.2.2. Mode of entry theory

We also hypothesized whether CorA could be a channel or pore through which the compound enters the bacterium and inhibits the 'real' target. The obvious question one would ask is – why would the bacterium mutate a channel protein and not the real target? Since this hypothetical 'real' target must be essential (based on the bactericidal activity of CpdX and CpdY), any mutation could be detrimental to its function and cause cessation of growth. However, according to this theory, the mutation of *corA* gene does not lead to bacterial death and, at the same time, block the entry of compound and hence it is a much more favorable option for the bacteria.

The biological target of any bactericidal compound/drug, by definition, has to be essential because if it is not essential then inhibition of the target should not have a phenotypic effect, i.e. death of bacteria. Genetic knock-out studies (Sareen *et al*., 2003) and Himar1-based transposon mutagenesis studies in *M. tuberculosis* have already shown that *corA* is non-essential for growth under *in vitro* conditions. This would suggest that the target for CpdX is another protein/enzyme and not the CorA gene product. It could be argued that since Middlebrook 7H9 broth has around 160µM of magnesium, the *corA* deletion mutant is able to grow by utilizing other magnesium transporters such as *mgtC* or *mgtE* (Piddington *et al*., 2000). This, however, does not explain the potent bactericidal activity of CpdX on BCG.

One could also propose strong arguments against this transporter hypothesis. If CorA is the transporter of CpdX and CpdY, the expression of wild-type corA in mutant BCG should provide a means for the compound to enter and should result in reversion back to compound sensitivity. This was not what was observed, in fact the bacteria

remained resistant. The transporter hypothesis can be disproved using a *corA* knock-out because, if this hypothesis is true, with the deletion of CorA, the compound cannot enter and will hence appear to have lost its activity. The research groups that crystallized CorA in *T. maritima*, suggest that the magnesium ion enters the bacteria in a partially dehydrated state (Eshaghi *et al*., 2006; Lunin *et al*., 2006). The Leu_{294} constriction in *T. maritima* CorA results in a mere 2.5Å making it unlikely for a 500g/mol compound (CpdX) to enter into such a small pore that has been shown to transport dehydrated magnesium under normal physiological conditions (Payandeh and Pai, 2006). It should be noted though that the protein was crystallized in a closed conformation and the pore could open up significantly in the open conformation.

4.2.3. Other hypotheses – Only a Mechanism of Resistance?

The non-essentiality of the *corA* gene and the results of the *corA* over-expression experiments suggest that there could be an alternative genetic explanation for a *corA* gene mutation mediating resistance to CpdX. Since, our mutant isolation studies were based on one single phenotype, resistance to the compound, we cannot rule out the possibility that CorA is not the target but is just an indirect resistance mechanism. It is possible that, *via* mechanisms that are still unknown, the mutation in *corA* gene leads to, for example, changes in the cell wall or alters a signaling pathway or restricts magnesium availability leading to the development of resistance to CpdX. The essential ATP binding ABC transporter (*Rv1463*) that was shown to be up-regulated upon drug exposure using 2D gels is a candidate gene that can be further evaluated. We have also plated out the A317S CpdX-resistant BCG mutant on Middlebrook 7H10 plates containing 50μM and 80μM of CpdX in order to isolate mutants with high level of resistance. Moreover, drug-resistant mutants can also be isolated using the *corA*

deletion BCG mutant. These strains will be sent for CGS and, if we are successful in identifying another gene that can confer compound resistance, this could be the real biological target. Similarly, given that the *corA* gene was not mutated in the *M. smegmatis* CpdY-resistant mutants, this suggests that there could be another putative drug target that could be identified by sequencing. However, comparative genome sequencing may not be possible at this stage for *M. smegmatis* because the annotation of the whole genome is yet to be completed. To design effective CGS chips, it is required for the whole genome sequence of the target organism to be completed. Hence, at this stage, we can only resort to the classical molecular biology technique of library screening (both wild-type and mutant libraries) to elucidate another candidate biological target in *M. smegmatis* that would subsequently be validated using biochemical and genetic techniques.

4.3. Differences in the proteome derived from wild-type and mutant *M. bovis* BCG

The proteomic profiling of wild-type BCG versus mutant BCG in the presence or absence of CpdX revealed several protein spot intensity differences. These include, superoxide dismutase (encoded by *Rv2946*) which has previously been shown to be up-regulated during heat-shock response and in a starvation model of *M. tuberculosis* (Stewart *et al*., 2002; Betts *et al*., 2002). An hypothetical protein (*Rv2699c*) and an essential ATP binding protein ABC transporter (*Rv1463*) were also found to be up-regulated during heat shock response and during nutrient starvation, respectively (Betts, 2002; Sassetti *et al*., 2003). Our results suggest that all three of these proteins were up-regulated with CpdX treatment. These results were expected since CpdX has

potent bactericidal activity on BCG and therefore we anticipated that the bacilli would express non-specific stress-related proteins/enzymes.

Microarray profiling of *M. tuberculosis* to CpdX exposure was done in collaboration with Dr. Clifton Barry (National Institutes of Health, USA), showed *sigE*, *sigD* and *sigH* genes to be up-regulated. *SigE* has previously been shown to be up-regulated at the mRNA level under stress conditions (specifically exposure to 0.05% of SDS) and takes part in the heat-shock response as shown by real-time quantitative PCR (Manganelli *et al*., 1999; Stewart *et al*., 2002). The product of oxidative stress-regulating gene *sigH* that regulates *sigE* was also up-regulated in an earlier study (Raman *et al*., 2001). Manganelli *et al*. (1999) also discovered that *sigD* is a sigma factor that is up-regulated at the mRNA level following heat-shock. Moreover, the superoxide dismutase (*Rv2946*) and thioredoxin peroxidase (*Rv1932*) proteins identified in our studies are known to destroy free radicals generated during oxidative stress.

Phosphoribosyl isomerase A (*Rv1603*), which catalyzes the fourth step of the essential histidine biosynthesis pathway, was also slightly up-regulated upon exposure to CpdX at the protein level in wild-type BCG, but not at the mRNA level. Based on the bactericidal activity of CpdX or CpdY, it could be anticipated that the target for these compounds should be an essential gene. For example, the inhibition of an essential amino acid biosynthetic pathway would inhibit the ability of the bacteria to multiply by removing one of its key building blocks.

Interestingly, all the genes in Table 3.6 excluding thioredoxin peroxidase, are found in the degenerate genome of the pathogen, *M. leprae*. This extremely slow-growing mycobacterium cannot be cultured *in vitro* and is believed to have undergone reductive evolution where genes not required for survival have been removed (Garnier *et al*., 2003 and Cole *et al*., 2001; Eiglmeier *et al*., 2001). In fact, the *M. leprae* genome is approximately 3.27Mb in size compared to the 4.41Mb genome of *M. tuberculosis*. Furthermore, in the *M. leprae* genome, over 1,100 genes are considered pseudogenes and are not functional. This suggests that, despite the fact that many of the genes in Table 3.6 can be disrupted by transposon mutagenesis in *in vitro* conditions, these genes themselves could still play some functional role in mycobacterial survival or pathogenicity under *in vivo* conditions.

Overall, the proteomics approach has led to the identification of primarily heat-shock or starvation factors being up-regulated following exposure to CpdX. The essential enzyme phosphoribosyl isomerase A, involved in histidine biosynthesis, was shown to be up-regulated at the protein level with CpdX exposure. The biological reason for this up-regulation requires further experiments and analysis. This data, along with future data, could help in elucidating the mechanism of action of CpdX and CpdY.

Chapter Five: Conclusion

5. Conclusion

In the present study, we have successfully isolated spontaneous resistant mutants to Compound X (CpdX) and its derivative Compound Y (CpdY) in both the slow-growing *M. bovis* BCG and fast-growing *M. smegmatis*. Using comparative genome sequencing and subsequent confirmation by sequencing several resistant clones, we have shown that resistance to CpdX and CpdY arises due to a single amino acid substitution in the universal magnesium transporter CorA. The ability of parental *M. bovis* BCG strain to acquire resistance to CpdX following the introduction of a mutant copy of the *corA* gene suggests that the *corA* gene could be the putative target for this compound class. Conversely, the introduction of the wild-type copy of the *corA* gene into the mutant *M. bovis* BCG did not reverse the phenotype to sensitivity, possibly because of the dominance of the mutant phenotype. Nevertheless, this has to be confirmed with live measurement of magnesium fluxes using a *corA* gene deletion mutant. On the other hand, *corA* has been shown to be non-essential under *in vitro* conditions because it can be deleted without losing viability. It is possible that another gene product is the actual gene target and that corA confers resistance *via* an yet unknown mechanism. In order to gain more knowledge about the biological target, we also employed transcriptomic and proteomic-based profiling of wild-type and mutant *M. bovis* BCG with and without CpdX treatment. From these systems biology approaches, we identified mainly genes involved in the heat shock response, and genes known to be up-regulated during starvation or oxidative stress. A forward chemical genetics approach is an attractive method for the identification of novel biological targets in mycobacteria and these novel targets could be utilized in future high-throughput screening (HTS) of compound libraries.

Bibliography

Bibliography

Global Alliance for TB Drug Development website, please see: www.tballiance.org.

Adler, J.J. and Rose, D.N. (1996) Transmission and pathogenesis of tuberculosis, 129-140 pp. Tuberculosis, edited by S.M. Garay. Little, Brown & Co, Boston.

Ahlburg, D. (2000) The economic impacts of tuberculosis. World Health Organization, Geneva.

Andries, K., Verhasselt, P., Guillemont, J., Gohlmann, H.W., Neefs, J.M., Winkler, H., Van Gestel, J., Timmerman, P., Zhu, M., Lee, E., Williams, P., de Chaffoy, D., Huitric, E., Hoffner, S., Cambau, E., Truffot-Pernot, C., Lounis, N. and Jarlier, V. (2005) A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. Science 307(5707), 223-7.

Aronson, N.E., Santosham, M., Comstock, G.W., Howard, R.S., Moulton, L.H., Rhoades, E.R. and Harrison, L.H. (2004) Long-term efficacy of BCG vaccine in American Indians and Alaska Natives: A 60-year follow-up study. Jama 291(17), 2086-91.

Balabanova, Y., Drobniewski, F., Fedorin, I., Zakharova, S., Nikolayevskyy, V., Atun, R. and Coker, R. (2006) The Directly Observed Therapy Short-Course (DOTS) strategy in Samara Oblast, Russian Federation. Respir Res 7, 44.

Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K.S., Wilson, T., Collins, D., de Lisle, G. and Jacobs, W.R., Jr. (1994) *inhA*, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. Science 263(5144), 227-30.

Baulard, A., Jourdan, C., Mercenier, A. and Locht, C. (1992) Rapid mycobacterial plasmid analysis by electroduction between *Mycobacterium* spp. and *Escherichia coli*. Nucleic Acids Res 20(15), 4105.

Betts, J.C., Lukey, P.T., Robb, L.C., McAdam, R.A. and Duncan, K. (2002) Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. Mol Microbiol 43(3), 717-31.

Biava, M., Porretta, G.C., Deidda, D. and Pompei, R. (2006) New trends in development of antimycobacterial compounds. Infect Disord Drug Targets 6(2), 159-72.

Bodnar, K.A., Serbina, N.V. and Flynn, J.L. (2001) Fate of *Mycobacterium tuberculosis* within murine dendritic cells. Infect Immun 69(2), 800-9.

Bonah, C. (2005) The 'experimental stable' of the BCG vaccine: safety, efficacy, proof, and standards, 1921-1933. Stud Hist Philos Biol Biomed Sci 36(4), 696-721.

Borisy, G.G. and Taylor, E.W. (1967) The mechanism of action of colchicine. Binding of colchincine-3H to cellular protein. J Cell Biol 34(2), 525-33.

Borisy, G.G. and Taylor, E.W. (1967) The mechanism of action of colchicine. Colchicine binding to sea urchin eggs and the mitotic apparatus. J Cell Biol 34(2), 535-48.

Boshoff, H.I. and Barry, C.E., 3rd. (2005) Tuberculosis - metabolism and respiration in the absence of growth. Nat Rev Microbiol 3(1), 70-80.

Boshoff, H.I., Myers, T.G., Copp, B.R., McNeil, M.R., Wilson, M.A. and Barry, C.E., 3rd. (2004) The transcriptional responses of *Mycobacterium tuberculosis* to inhibitors of metabolism: novel insights into drug mechanisms of action. J Biol Chem 279(38), 40174-84.

Brandt, L., Feino Cunha, J., Weinreich Olsen, A., Chilima, B., Hirsch, P., Appelberg, R. and Andersen, P. (2002) Failure of the *Mycobacterium bovis* BCG vaccine: some species of environmental mycobacteria block multiplication of BCG and induction of protective immunity to tuberculosis. Infect Immun 70(2), 672-8.

Brown, E.J., Albers, M.W., Shin, T.B., Ichikawa, K., Keith, C.T., Lane, W.S. and Schreiber, S.L. (1994) A mammalian protein targeted by G1-arresting rapamycin-receptor complex. Nature 369(6483), 756-8.

Buchmeier, N., Blanc-Potard, A., Ehrt, S., Piddington, D., Riley, L. and Groisman, E.A. (2000) A parallel intraphagosomal survival strategy shared by mycobacterium tuberculosis and Salmonella enterica. Mol Microbiol 35(6), 1375-82.

Canetti, G. (1955) The tubercle bacillus in the pulmonary lesion of man. Springer Publishing Co., Inc., New York, USA.

Centers for Disease Control and Prevention (2006) Emergence of *Mycobacterium tuberculosis* with Extensive Resistance to Second-Line Drugs - Worldwide, 2000-2004. MMRW Weekly 55(11), 301-305.

Chan, J. and Kaufmann, S.H.E. (1994) Immune mechanisms of protection. Tuberculosis: pathogenesis, protection, and control, edited by B.R. Bloom. American Society for Microbiology, Washington, D.C., USA.

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M.A., Rajandream, M.A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S.,

Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J.R. and Barrell, B.G. (2001) Massive gene decay in the leprosy bacillus. Nature 409(6823), 1007-11.

Comstock, G.W., Baum, C. and Snider, D.E., Jr. (1979) Isoniazid prophylaxis among Alaskan Eskimos: a final report of the bethel isoniazid studies. Am Rev Respir Dis 119(5), 827-30.

Corper, H.J. and Cohn, M.L. (1933) The viability and virulence of old cultures of tubercle bacilli. Ann Rev Tuberc 28, 856-874.

Cosma, C.L., Sherman, D.R. and Ramakrishnan, L. (2003) The secret lives of the pathogenic mycobacteria. Annu Rev Microbiol 57, 641-76.

Cowan, J.A. (1998) Metal Activation of Enzymes in Nucleic Acid Biochemistry. Chem Rev 98(3), 1067-1088.

De Rossi, E., Ainsa, J.A. and Riccardi, G. (2006) Role of mycobacterial efflux transporters in drug resistance: an unresolved question. FEMS Microbiol Rev 30(1), 36-52.

De Rossi, E., Arrigo, P., Bellinzoni, M., Silva, P.A., Martin, C., Ainsa, J.A., Guglierame, P. and Riccardi, G. (2002) The multidrug transporters belonging to major facilitator superfamily in Mycobacterium tuberculosis. Mol Med 8(11), 714-24.

De Rossi, E., Branzoni, M., Cantoni, R., Milano, A., Riccardi, G. and Ciferri, O. (1998) *mmr*, a *Mycobacterium tuberculosis* gene conferring resistance to small cationic dyes and inhibitors. J Bacteriol 180(22), 6068-71.

Dheda, K., Udwadia, Z.F., Huggett, J.F., Johnson, M.A. and Rook, G.A. (2005) Utility of the antigen-specific interferon-gamma assay for the management of tuberculosis. Curr Opin Pulm Med 11(3), 195-202.

Dick, T. (2001) Dormant tubercle bacilli: the key to more effective TB chemotherapy? J Antimicrob Chemother 47(1), 117-8.

Dickinson, J.M. and Mitchison, D.A. (1981) Experimental models to explain the high sterilizing activity of rifampin in the chemotherapy of tuberculosis. Am Rev Respir Dis 123(4 Pt 1), 367-71.

Dye, C., Maher, D., Weil, D., Espinal, M. and Raviglione, M. (2006) Targets for global tuberculosis control. Int J Tuberc Lung Dis 10(4), 460-462.

Dye, C., Scheele, S., Dolin, P., Pathania, V. and Raviglione, M.C. (1999) Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. Jama 282(7), 677-86.

Eiglmeier, K., Parkhill, J., Honore, N., Garnier, T., Tekaia, F., Telenti, A., Klatser, P., James, K.D., Thomson, N.R., Wheeler, P.R., Churcher, C., Harris, D., Mungall, K., Barrell, B.G. and Cole, S.T. (2001) The decaying genome of *Mycobacterium leprae*. Lepr Rev 72(4), 387-98.

Ellner, J.J., Goldberger, M.J. and Parenti, D.M. (1991) *Mycobacterium avium* infection and AIDS: a therapeutic dilemma in rapid evolution. J Infect Dis 163(6), 1326-35.

Elzinga, G., Raviglione, M.C. and Maher, D. (2004) Scale up: meeting targets in global tuberculosis control. Lancet 363(9411), 814-9.

Enarson, D.A. and Rouillon, A. (1994) The epidemiological basis of tuberculosis control, 19-32 pp. Clinical tuberculosis, edited by P.D.O. Davis. Chapman & Hall, London, United Kingdom.

Eshaghi, S., Niegowski, D., Kohl, A., Martinez Molina, D., Lesley, S.A. and Nordlund, P. (2006) Crystal structure of a divalent metal ion transporter CorA at 2.9 angstrom resolution. Science 313(5785), 354-7.

Espinal, M.A., Laszlo, A., Simonsen, L., Boulahbal, F., Kim, S.J., Reniero, A., Hoffner, S., Rieder, H.L., Binkin, N., Dye, C., Williams, R. and Raviglione, M.C. (2001) Global trends in resistance to antituberculosis drugs. World Health Organization-International Union against Tuberculosis and Lung Disease Working Group on Anti-Tuberculosis Drug Resistance Surveillance. N Engl J Med 344(17), 1294-303.

Farmer, P. and Kim, J.Y. (1998) Community based approaches to the control of multidrug resistant tuberculosis: introducing "DOTS-plus". Bmj 317(7159), 671-4.

Faul, M. (2006) AP Interview: Fatal TB likely widespread, Associated Press, Tugela Ferry, South Africa

Feldman, W.H. and Baggenstoss, A.H. (1939) The occurrence of virulent tubercle bacilli in presumably non-tuberculous lung tissue. Am J Pathol 15, 501-515.

Fine, P.E. (1995) Variation in protection by BCG: implications of and for heterologous immunity. Lancet 346(8986), 1339-45.

Fortun, J., Martin-Davila, P., Navas, E., Perez-Elias, M.J., Cobo, J., Tato, M., De la Pedrosa, E.G., Gomez-Mampaso, E. and Moreno, S. (2005) Linezolid for the treatment of multidrug-resistant tuberculosis. J Antimicrob Chemother 56(1), 180-5.

French, G. (2003) Safety and tolerability of linezolid. J Antimicrob Chemother 51 Suppl 2, ii45-53.

Froschauer, E.M., Kolisek, M., Dieterich, F., Schweigel, M. and Schweyen, R.J. (2004) Fluorescence measurements of free [Mg2+] by use of mag-fura 2 in *Salmonella enterica*. FEMS Microbiol Lett 237(1), 49-55.

Garnier, T., Eiglmeier, K., Camus, J.C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempe, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, P.R., Parkhill, J., Barrell, B.G., Cole, S.T., Gordon, S.V. and Hewinson, R.G. (2003) The complete genome sequence of *Mycobacterium bovis*. Proc Natl Acad Sci U S A 100(13), 7877-82.

Gedde-Dahl, T. (1952) Tuberculous infection in the light of tuberculin matriculation. Am J Hyg 56(2), 139-214.

Goodwin, A., Kersulyte, D., Sisson, G., Veldhuyzen van Zanten, S.J., Berg, D.E. and Hoffman, P.S. (1998) Metronidazole resistance in *Helicobacter pylori* is due to null mutations in a gene (*rdxA*) that encodes an oxygen-insensitive NADPH nitroreductase. Mol Microbiol 28(2), 383-93.

Grange, J.M. and Zumla, A. (1999) Paradox of the global emergency of tuberculosis. Lancet 353(9157), 996.

Grosset, J. (2003) *Mycobacterium tuberculosis* in the extracellular compartment: an underestimated adversary. Antimicrob Agents Chemother 47(3), 833-6.

Gupta, R., Kim, J.Y., Espinal, M.A., Caudron, J.M., Pecoul, B., Farmer, P.E. and Raviglione, M.C. (2001) Public health. Responding to market failures in tuberculosis control. Science 293(5532), 1049-51.

Harding, M.W., Galat, A., Uehling, D.E. and Schreiber, S.L. (1989) A receptor for the immunosuppressant FK506 is a cis-trans peptidyl-prolyl isomerase. Nature 341(6244), 758-60.

Heifets, L. and Lindholm-Levy, P. (1992) Pyrazinamide sterilizing activity in vitro against semidormant *Mycobacterium tuberculosis* bacterial populations. Am Rev Respir Dis 145(5), 1223-5.

Hernandez-Pando, R., Jeyanathan, M., Mengistu, G., Aguilar, D., Orozco, H., Harboe, M., Rook, G.A. and Bjune, G. (2000) Persistence of DNA from *Mycobacterium tuberculosis* in superficially normal lung tissue during latent infection. Lancet 356(9248), 2133-8.

Houben, E.N., Nguyen, L. and Pieters, J. (2006) Interaction of pathogenic mycobacteria with the host immune system. Curr Opin Microbiol 9(1), 76-85.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., Kidd, M.J., King, A.M., Meyer, M.R., Slade, D., Lum, P.Y., Stepaniants, S.B., Shoemaker, D.D., Gachotte, D., Chakraburtty, K., Simon, J., Bard, M. and Friend, S.H. (2000) Functional discovery via a compendium of expression profiles. Cell 102(1), 109-26.

Hung, D.T. and Rubin, E.J. (2006) Chemical biology and bacteria: not simply a matter of life or death. Curr Opin Chem Biol 10(4), 321-6.

Inderlied, C.B., Kemper, C.A. and Bermudez, L.E. (1993) The *Mycobacterium avium* complex. Clin Microbiol Rev 6(3), 266-310.

International Union against Tuberculosis Committee on Prophylaxis. (1982) Efficacy of various durations of isoniazid preventative therapy for tuberculosis: five years of follow-up in the IUAT trial. Bull World Health Organ 60, 555-564.

Keller, A., Nesvizhskii, A.I., Kolker, E. and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 74(20), 5383-92.

Kucharski, L.M., Lubbe, W.J. and Maguire, M.E. (2000) Cation hexaammines are selective and potent inhibitors of the CorA magnesium transport system. J Biol Chem 275(22), 16767-73.

Lalvani, A. (2003) Spotting latent infection: the path to better tuberculosis control. Thorax 58(11), 916-8.

Lamichhane, G., Zignol, M., Blades, N.J., Geiman, D.E., Dougherty, A., Grosset, J., Broman, K.W. and Bishai, W.R. (2003) A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 100(12), 7213-8.

Lillebaek, T., Dirksen, A., Baess, I., Strunge, B., Thomsen, V.O. and Andersen, A.B. (2002) Molecular evidence of endogenous reactivation of *Mycobacterium tuberculosis* after 33 years of latent infection. J Infect Dis 185(3), 401-4.

Lin, S.Y., Probert, W., Lo, M. and Desmond, E. (2004) Rapid detection of isoniazid and rifampin resistance mutations in *Mycobacterium tuberculosis* complex from cultures or smear-positive sputa by use of molecular beacons. J Clin Microbiol 42(9), 4204-8.

Long, E.R. (1935) From pathology to epidemiology in tuberculosis: softening of the caseous tubercle and its results. JAMA 104, 1883-1888.

Lunin, V.V., Dobrovetsky, E., Khutoreskaya, G., Zhang, R., Joachimiak, A., Doyle, D.A., Bochkarev, A., Maguire, M.E., Edwards, A.M. and Koth, C.M. (2006) Crystal structure of the CorA Mg2+ transporter. Nature 440(7085), 833-7.

MacBeath, G. and Schreiber, S.L. (2000) Printing proteins as microarrays for high-throughput function determination. Science 289(5485), 1760-3.

Maguire, M.E. (2006) Magnesium Transporters: Properties, Regulation and Structure. Frontiers in Bioscience 11, 3149-3163.

Maguire, M.E. (2006) The structure of CorA: a Mg(2+)-selective channel. Curr Opin Struct Biol 16(4), 432-8.

Manabe, Y.C. and Bishai, W.R. (2000) Latent *Mycobacterium tuberculosis*-persistence, patience, and winning by waiting. Nat Med 6(12), 1327-9.

Manganelli, R., Dubnau, E., Tyagi, S., Kramer, F.R. and Smith, I. (1999) Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. Mol Microbiol 31(2), 715-24.

Manganelli, R., Provvedi, R., Rodrigue, S., Beaucher, J., Gaudreau, L. and Smith, I. (2004) Sigma factors and global gene regulation in *Mycobacterium tuberculosis*. J Bacteriol 186(4), 895-902.

Martin, C. (2006) Tuberculosis vaccines: past, present and future. Curr Opin Pulm Med 12(3), 186-91.

McDonough, K.A., Kress, Y. and Bloom, B.R. (1993) Pathogenesis of tuberculosis: interaction of *Mycobacterium tuberculosis* with macrophages. Infect Immun 61(7), 2763-73.

McKinney, J.D. (2000) In vivo veritas: the search for TB drug targets goes live. Nat Med 6(12), 1330-3.

Merck. (last updated: 2005) Tuberculosis (TB), Merck Manual Professional. Merck, USA.

Mitchison, D.A. (1979) Basic mechanisms of chemotherapy. Chest 76(6 Suppl), 771-81.

Mitchison, T.J. (1994) Towards a pharmacological genetics. Chem Biol 1(1), 3-6.

Morgan, M., Kalantri, S., Flores, L. and Pai, M. (2005) A commercial line probe assay for the rapid detection of rifampicin resistance in *Mycobacterium tuberculosis*: a systematic review and meta-analysis. BMC Infect Dis 5, 62.

Nahid, P., Pai, M. and Hopewell, P.C. (2006) Advances in the diagnosis and treatment of tuberculosis. Proc Am Thorac Soc 3(1), 103-10.

Nathan, C. and Shiloh, M.U. (2000) Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens. Proc Natl Acad Sci U S A 97(16), 8841-8.

Nayyar, A. and Jain, R. (2005) Recent advances in new structural classes of anti-tuberculosis agents. Curr Med Chem 12(16), 1873-86.

Nesvizhskii, A.I., Keller, A., Kolker, E. and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 75(17), 4646-58.

Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Furutani, Y., Hirose, T., Asai, M., Inayama, S., Miyata, T. and Numa, S. (1982) Primary structure of alpha-subunit precursor of *Torpedo californica* acetylcholine receptor deduced from cDNA sequence. Nature 299(5886), 793-7.

O'Brien, R.J. and Nunn, P.P. (2001) The need for new drugs against tuberculosis. Obstacles, opportunities, and next steps. Am J Respir Crit Care Med 163(5), 1055-8.

O'Brien, R.J. and Spigelman, M. (2005) New drugs for tuberculosis: current status and future prospects. Clin Chest Med 26(2), 327-40, vii.

Onyebujoh, P., Zumla, A., Ribeiro, I., Rustomjee, R., Mwaba, P., Gomes, M. and Grange, J.M. (2005) Treatment of tuberculosis: present status and future prospects. Bull World Health Organ 83(11), 857-65.

Opie, E.L. and Aronson, J.D. (1927) Tubercle bacilli in latent tuberculous lesions and in lung tissue without tuberculosis lesions. Arch. Pathol. Lab. Med. 4, 1-21.

Pai, M. (2005) Alternatives to the tuberculin skin test: interferon-gamma assays in the diagnosis of *Mycobacterium tuberculosis* infection. Indian J Med Microbiol 23(3), 151-8.

Pai, M., Riley, L.W. and Colford, J.M., Jr. (2004) Interferon-gamma assays in the immunodiagnosis of tuberculosis: a systematic review. Lancet Infect Dis 4(12), 761-76.

Parish, T. and Stoker, N. (1998) Mycobacteria protocols, edited by T. Parish and N. Stoker. Humana Press, Totowa, New Jersey, USA.

Parrish, N.M., Dick, J.D. and Bishai, W.R. (1998) Mechanisms of latency in *Mycobacterium tuberculosis*. Trends Microbiol 6(3), 107-12.

Payandeh, J. and Pai, E.F. (2006) A structural basis for Mg2+ homeostasis and the CorA translocation cycle. Embo J 25(16), 3762-73.

Pfeiffer, J., Guhl, J., Waidner, B., Kist, M. and Bereswill, S. (2002) Magnesium uptake by CorA is essential for viability of the gastric pathogen *Helicobacter pylori*. Infect Immun 70(7), 3930-4.

Piddington, D.L., Kashkouli, A. and Buchmeier, N.A. (2000) Growth of *Mycobacterium tuberculosis* in a defined medium is very restricted by acid pH and Mg(2+) levels. Infect Immun 68(8), 4518-22.

Powell, D.A. and Hunt, W.G. (2006) Tuberculosis in children: an update. Adv Pediatr 53, 279-322.

Quigley, G.J., Teeter, M.M. and Rich, A. (1978) Structural analysis of spermine and magnesium ion binding to yeast phenylalanine transfer RNA. Proc Natl Acad Sci U S A 75(1), 64-8.

Raftery, M.A., Hunkapiller, M.W., Strader, C.D. and Hood, L.E. (1980) Acetylcholine receptor: complex of homologous subunits. Science 208(4451), 1454-6.

Raman, S., Song, T., Puyang, X., Bardarov, S., Jacobs, W.R., Jr. and Husson, R.N. (2001) The alternative sigma factor *SigH* regulates major components of oxidative and heat stress responses in Mycobacterium tuberculosis. J Bacteriol 183(20), 6119-25.

Raviglione, M.C., Harries, A.D., Msiska, R., Wilkinson, D. and Nunn, P. (1997) Tuberculosis and HIV: current status in Africa. Aids 11 Suppl B, S115-23.

Riley, R.L. (1974) Airborne infection. Am J Med 57(3), 466-75.

Riley, R.L., Mills, C.C., O'Grady, F., Sultan, L.U., Wittstadt, F. and Shivpuri, D.N. (1962) Infectiousness of air from a tuberculosis ward. Ultraviolet irradiation of infected air: comparative infectiousness of different patients. Am Rev Respir Dis 85, 511-25.

Sareen, D., Newton, G.L., Fahey, R.C. and Buchmeier, N.A. (2003) Mycothiol is essential for growth of *Mycobacterium tuberculosis* Erdman. J Bacteriol 185(22), 6736-40.

Sassetti, C.M., Boyd, D.H. and Rubin, E.J. (2001) Comprehensive identification of conditionally essential genes in mycobacteria. Proc Natl Acad Sci U S A 98(22), 12712-7.

Sassetti, C.M., Boyd, D.H. and Rubin, E.J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol 48(1), 77-84.

Sassetti, C.M. and Rubin, E.J. (2003) Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci U S A 100(22), 12989-94.

Shim, J.S. and Kwon, H.J. (2004) Chemical genetics for therapeutic target mining. Expert Opin Ther Targets 8(6), 653-61.

Smith, D.L. and Maguire, M.E. (1993) Molecular aspects of Mg2+ transport systems. Miner Electrolyte Metab 19(4-5), 266-76.

Smith, R.L. and Maguire, M.E. (1998) Microbial magnesium transport: unusual transporters searching for identity. Mol Microbiol 28(2), 217-26.

Spring, D.R. (2005) Chemical genetics to chemical genomics: small molecules offer big insights. Chem Soc Rev 34(6), 472-82.

Stewart, G.R., Wernisch, L., Stabler, R., Mangan, J.A., Hinds, J., Laing, K.G., Young, D.B. and Butcher, P.D. (2002) Dissection of the heat-shock response in *Mycobacterium tuberculosis* using mutants and microarrays. Microbiology 148(Pt 10), 3129-38.

Stockwell, B.R. (2000) Chemical genetics: ligand-based discovery of gene function. Nat Rev Genet 1(2), 116-25.

Szegedy, M.A. and Maguire, M.E. (1999) The CorA Mg(2+) transport protein of *Salmonella typhimurium*. Mutagenesis of conserved residues in the second membrane domain. J Biol Chem 274(52), 36973-9.

Tortoli, E., Dionisio, D. and Fabbri, C. (2004) Evaluation of moxifloxacin activity in vitro against *Mycobacterium tuberculosis*, including resistant and multidrug-resistant strains. J Chemother 16(4), 334-6.

Towbin, H., Bair, K.W., DeCaprio, J.A., Eck, M.J., Kim, S., Kinder, F.R., Morollo, A., Mueller, D.R., Schindler, P., Song, H.K., van Oostrum, J., Versace, R.W., Voshol, H., Wood, J., Zabludoff, S. and Phillips, P.E. (2003) Proteomics-based target identification: bengamides as a new class of methionine aminopeptidase inhibitors. J Biol Chem 278(52), 52964-71.

Vilcheze, C., Wang, F., Arai, M., Hazbon, M.H., Colangeli, R., Kremer, L., Weisbrod, T.R., Alland, D., Sacchettini, J.C. and Jacobs, W.R., Jr. (2006) Transfer of a point mutation in *Mycobacterium tuberculosis inhA* resolves the target of isoniazid. Nat Med 12(9), 1027-9.

Villarino, M.E. and Bliven, E.E. (2006) Back to the future: where now for antituberculosis drugs? Enferm Infecc Microbiol Clin 24(2), 69-70.

Watkins, S.M. (2000) Comprehensive lipid analysis: a powerful metanomic tool for predictive and diagnostic medicine. Isr Med Assoc J 2(9), 722-4.

Wayne, L.G. and Hayes, L.G. (1996) An in vitro model for sequential study of shiftdown of *Mycobacterium tuberculosis* through two stages of nonreplicating persistence. Infect Immun 64(6), 2062-2069.

Wayne, L.G. and Sohaskey, C.D. (2001) Nonreplicating persistence of mycobacterium tuberculosis. Annu Rev Microbiol 55, 139-63.

Westwood, N.J. (2004) Chemical genetics: how does it function? Philos Transact A Math Phys Eng Sci 362(1825), 2761-74.

World Health Organization (WHO). (2000) Guidelines for establishing DOTS-Plus pilot projects for the management of multidrug resistant tuberculosis (MDR-TB). World Health Organization, Geneva.

World Health Organization (WHO). (2002) An expanded DOTS framework for effective tuberculosis control. Int J Tuberc Lung Dis 6(5), 378-88.

World Health Organization (WHO). (2003) Treatment of Tuberculosis: Guidelines for National Programmes (Third Edition). World Health Organization, Geneva.

World Health Organization (WHO). (2004) Global tuberculosis control - surveillance, planning, financing, WHO report 2004. WHO report 2004. World Health Organization, Geneva.

Zhang, Y. and Amzel, L.M. (2002) Tuberculosis drug targets. Curr Drug Targets 3(2), 131-54.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R.A.,

Gerstein, M. and Snyder, M. (2001) Global analysis of protein activities using proteome chips. Science 293(5537), 2101-5.

Zodpey, S.P., Maldhure, B.R., Dehankar, A.G. and Shrikhande, S.N. (1996) Effectiveness of bacillus Calmette Guerin (BCG) vaccination against extra-pulmonary tuberculosis: a case-control study. J Commun Dis 28(2), 77-84.

Zodpey, S.P., Maldhure, B.R., Shrikhande, S.N. and Tiwari, R.R. (1996) Effectiveness of bacillus of Calmette-Guerin (BCG) vaccination against tuberculous meningitis: a case-control study. J Indian Med Assoc 94(9), 338-40.

Zumla, A., Malon, P., Henderson, J. and Grange, J.M. (2000) Impact of HIV infection on tuberculosis. Postgrad Med J 76(895), 259-68.

Appendix

Appendix I: Comparative Genome Sequencing

This file explains the data output from the CGS process and modified from the file provided by NimbleGen Inc. (USA).

NOTE: For organisms with genomic G/C content of >60%, such as Mycobacterium or Pseudomonas species, please note that the resequencing step often has a high false positive rate, calling more than 100 false positive SNPs per genome analysis. To eliminate the majority of these false positive, please look at the "HIGHEST__RANK" column in the <Samplename>_snp_info.txt, SNPs with "HIGHEST__RANK" scores above 10,000 are most likely to be real, and any score below 1, are likely to be false positives.

Genome Mapping: <Samplename>_ratio.gff (see Figure 3.2: Top Panel)

Genome Resequencing: <Samplename>_snp_info.txt (see Figure 3.2: Bottom Panel)

<Samplename>_ratio.gff

The <Samplename>_ratio.gff file is designed to be viewed using NimbleGen's SignalMap visualization software. This viewer allows the data from the entire genome to be easily browsed. It contains data "tracks" that indicate the positions of SNPs and other potential mutations, as well as raw data. A users manual for SignalMap (in PDF format) is available from NimbleGen if one was not delivered with the Software. The following data tracks may or may not be present, depending on the availability of genome annotation and resequencing information:

Genome Annotation (generally named with the genome identifier, typically the RefSeq number if available): Displays colored blocks representing annotated genes. Blocks above the line indicate genes encoded on the forward strand of the genome, while bars below the line represent genes encoded on the reverse strand. The CDS data field from the Genbank annotation file is used to generate this track. If the Genbank file was annotated with hyperlinks, double-clicking on a gene block will open a web browser and target the NCBI database for know information about the selected gene. If no Genbank annotation file was available at the time of analysis, this track will be absent.

DIFFERENCES_<samplename>: The track will only be displayed if resequencing was not performed. Indicates positions in the genome that have a significant ratio shift between reference and test probes for both strands. These positions may represent SNPs, deletions, sites of insertion, the end points of inversions or translocations. It is sometimes possible to distinguish mutation signatures. Deletions tend to have large ratio shifts, while SNP ratio shifts tend to be subtler. The number of consecutive probes shifted by any mutation will allow for an estimation of its size, based on the mapping probe spacing. Most of the mutations represented by SNPs can be resolved by performing follow up resequencing array analysis, as long as the genomes do not appear to be too divergent (>0.5%).

SNP_<samplename>: This track is only displayed if resequencing was performed. Displays colored coded points or vertical bars that indicate the position of SNPs (base substitutions) identified in the genome by resequencing. If the data is displayed as

points, it is highly recommended that the track style is changed to bars, by right clicking on the left side of the y-axis for this track, selecting style, and clicking on bars. This will make the SNP positions much easier to visualize. SNP positions are colored light blue if they fall between annotated genes, green if they fall inside a gene but does not change the amino acid sequence of the encoded protein (synonymous mutation), and dark blue if the SNP changes the amino acid sequence of the encoded protein (non-synonymous mutation). Holding the cursor over the SNP position will display the nucleic acid and amino acid changes at the top of the screen.

NON_CALLED_RIO_<samplename>: This track is only displayed if resequencing was performed. Displays positions in the genome that have a significant ratio shift between reference and test probes for both strands, but were not called SNPs by resequencing. These positions may represent non-SNP mutations, such as deletions, sites of insertion, the end points of inversions or translocations, tightly clustered SNPs, or individual SNPs that were not called reliably by resequencing. The number of consecutive probes shifted by any mutation will allow for an estimation of its size, based on the mapping probe spacing. The exact nature of these positions can be obtained by other techniques, such as PCR amplification followed by capillary sequencing.

RESEQUENCED_<samplename>: This track is only displayed if resequencing was performed. Displays positions in the genome that were interrogated by resequencing arrays.

RATIO_<samplename>: Plot of the probe intensity ratio (Reference/test). Reference probes and test probes that do not span a mutation should represent full-length perfect match hybridization, and thus should have similar intensities, with a reference/test ratio near 1. Probes that span mutations in the test genome typically display lower hybridization intensities than the corresponding reference probes, and thus the reference/test ratio will shift above 1. If the test genome contains an amplification event (increased copy number when compared to the reference), then the reference/test ratio will shift below 1.
Note: If the reference genome used to design the arrays is different from the actual reference used in the experiment, the reference probe intensity values may not represent full-length perfect match hybridization signal. These regions may not provide enough signal to allow robust data generation.

AVE_TEST_<samplename>: Plot of the average of the forward and reverse probe intensities from the test genome hybridization.

AVE_REFERENCE_<samplename>: Plot of the average of the forward and reverse probe intensities from the reference genome hybridization.

<Samplename>_snp_info.txt

The snp_info file provides information on SNPs that were identified in the genome. The data columns produced by the SNP analysis are provided below.

SEQ_ID: The genome identifier, typically the RefSeq number if available.

GENOME_POSITION: The genomic position of each SNP.

REF: The reference nucleotide of each SNP.

SNP: The mutant nucleotide of each SNP.

CONFIDENCE: A composite of several parameters to determine the likelihood that any SNP is real (not a false positive). This score is based on the quality of the SNP call and the similarity of the sequencing flanking the SNP to other sequences in the genome that may cross hybridize. SNPs with HIGH confidence scores have a high probability of being true SNPs. LOW confidence scores generally result from failed uniqueness testing, and have a very high likelihood of being false positive base calls. These sites should generally be ignored. SNPs with MEDIUM confidence should be validated by capillary sequencing before biological conclusions are drawn.

PROBABILITY: A relative measure of the quality of a SNP call. The score typically ranges between 1.0 and 0.9. The lower the probability score, the less likely a SNP is to be real. This criterion alone should not be used to assign SNP calls. SNP loci that are not unique in the genome often have excellent probability scores and are still often false positive.

TYPE: Lists whether or not a SNP is found within an annotated gene, or between annotated genes. SNPs in genes are designated as coding, and are listed first. SNPs between genes are designated as intergenic, and are listed second.

NMER_SCORE: A weighted measure of how often perfect matches for increasing numbers of central bases in each probe are found in the genome is calculated: the "Nmer frequency". Each such perfect match contributes to the Nmer frequency score, with shorter oligos contributing less than longer oligos according to the following equation:

Nmer frequency score = □ [N-mer frequency]*[0.75^((29mer-Nmer length)/2)]

The sum of the frequency scores for central Nmers from 19, 21, 23, 25, 27, and 29mers is calculated for each Probe calling a SNP. The higher the Nmer frequency score, the more likely it is that the probe will also hybridize elsewhere in the genome. Any SNP being called by a probe with an Nmer frequency above zero is considered a potential false positive.

UNIQUE: The second uniqueness test calculates a "sequence dissimilarity" score that is calculated by comparing the 29 bases spanning the SNP to every possible 29 base window in the reference genome. Mismatches rear the end of a probe tend to disrupt hybridization less than centrally located mismatches, and hence are weighted differently. The weight value vs. position used in this score are listed below:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1	1	1	2	2	2	2	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	2	2	2	2	1	1	1

The dissimilarity score for each possible 29mer in the genome is calculated by adding up the weight vectors for each base not matching its corresponding base in the 29 bases surrounding a SNP. If the lowest dissimilarity score for every 29mer in the genome is

greater than 10, the SNP position is considered unique. For example, if the 29mer that most closely matches the 29 bases surrounding a SNP contained 3 mismatches between positions 8 and 22, that SNP position would have a dissimilarity score of 12 (4+4+4), and be sufficiently dissimilar to be considered unique, whereas six mismatches in positions 1 to 6 would not meet the uniqueness threshold of 10 (1+1+1+2+2+2 = 9). SNP sites that pass this criterion (threshold value >10) are given a passing score of 1, while SNP sites that fail (threshold value <10) are given a failing score of 0. SNP sites that fail this criterion are considered potential false positives.

HIGHEST_RANK: A measure of how strongly the SNP position was indicated by the mapping data. Ranks above 10,000 are strongly indicated by both strands, whereas numbers between 10,000 and 1 were only strongly indicated a probe from one strand. Rank scores below 0 were not strongly mapped by either strand. Occasionally true SNPs are identified by resequencing that were not strongly indicated by mapping. NOTE: For organisms with genomic G/C content of >60%, such as Mycobacterium or Pseudomonas species, please note that the resequencing step often has a high false positive rate, calling more than 100 false positive SNPs per genome analysis. SNPs with "HIGHEST__RANK" scores above 10,000 are most likely to be real, and any score below 1, are likely to be false positives.

FORWARD: The average probability score of the five base calls in either direction from the SNP on the forward strand. This is a measure of the SNP "neighborhood" quality.

REVERSE: The average probability score of the five base calls in either direction from the SNP on the reverse strand. This is a measure of the SNP "neighborhood" quality.

FORWARD_INTENSITY: The raw intensity of the probe that called the SNP on the forward strand.

REVERSE_INTENSITY: The raw intensity of the probe that called the SNP on the reverse strand.

AA CHANGE: Categorizes coding SNPs base on whether or not they change the amino acid sequence of a protein. SYN indicates synonymous SNPs (no amino acid change). NON indicates nonsynonymous SNPs (altered amino acid).

ORIG_AA: The amino acid associated with the reference sequence for the corresponding SNP position.

SNP_AA: The amino acid associated with the test sequence, for the corresponding SNP position.

GENE_POSITION: The nucleic acid position mutated within a gene.

LOCUS_TAG: The locus tag field from the Genbank annotation file.

START: The genomic start position of the mutated gene.

END: The genomic end position of the mutated gene.

STRAND: The genome strand on which the mutated gene is located.

PRODUCT: The annotation field from the Genbank file.

NOTE: The note field from the Genbank file.