I. Review of Literature
1.1. Tuberculosis (TB): a challenge of unicellular to multicellular

Tuberculosis, a disease of great antiquity holds lineage to saprophytic soil organisms whose later introduction as a human pathogen likely coincided with the domestication of cattle approximately 10,000 years ago (Smith et al., 2003). Throughout history, tuberculosis has been classified by many names, from Phthisis ("to waste") by the ancient Greeks, to consumption in the 1800's (Hippocrates, 2006). In 1882 Robert Koch isolated the causative agent, *M. tuberculosis*, commonly known as tubercle bacilli, from crushed lung tubercles. It was only Edward Trudeau who showed that TB could be induced in rabbits with a purified culture of virulent *M. tuberculosis*, and environmental conditions play very important role in the development and progression of disease (Angelichio et al., 2002). Currently, *M. tuberculosis* is known as the world's leading cause of death from a single infectious agent, with a global prevalence of greater than 1.7 billion persons which is equal to one-third of the entire world population (Dolin et al., 2000 and Dye et al., 1999). As of recent WHO report, there were 9 million new TB cases and approximately 2 million TB deaths in 2004. This alarming situation in addition to the increasing prevalence of multidrug resistant strains due to poor compliance of the drug regime to the long term treatment course and reactivations in the later stages of life from latent infection in immunocompromised hosts such as HIV patients has provoked the need for finding new effective therapeutics.

1.2. Mycobacterium tuberculosis: the causal organism

*M. tuberculosis* is a fairly large, nonmotile, aerobic, rod-shaped pathogenic bacterium. The rods are 2-4 μm in length and 0.2-0.5 μm in width. *M. tuberculosis* is a slow grower with a generation time of 18-20 h and is classified as acid-fast as once stained with certain dyes, it is resistant to destain when treated with acidified organic compounds. Bacteria grows well under aerobic conditions as in well aerated spaces of lungs but can also survive in the low oxygen conditions and enters into nonreplicating persistent state as in macrophages infection (facultative intracellular parasite). Generally mycobacteria can be divided into two groups depending upon their growth rate. Slow growers such as *M. tuberculosis, M. bovis, M. leprae* etc. have generation time of around 18-20 h and are pathogenic whereas fast growers including *M. smegmatis, M. aurum, M. fortuitum* etc. have generation time of 3-5 hrs (Rogall et al., 1990).
1.3. Members of *M. tuberculosis* complex

The complex comprises *M. tuberculosis*, *M. africanum*, *M. microti*, *M. bovis*, *M. bovis* BCG and *M. canetti* (Rogall et al., 1990).

*M. tuberculosis* is the most common human pathogen and is responsible for the vast majority of human tuberculosis. It only exceptionally causes disease in other mammals.

*M. africanum* is very closely related to *M. tuberculosis* and was isolated from patients in Africa. Phenotypically they appear to be intermediate between *M. bovis* and *M. tuberculosis*.

*M. bovis* has the widest range of hosts. In addition to its primary host cattle, it may affect several domestic and wild animals including dears, lions, seals and laboratory animals like rabbits. It can infect humans but with pasteurization of milk and other safe practices, its effect is limited.

*M. bovis* BCG is an attenuated strain of *M. bovis* generated by several passages (230 passages) of *M. bovis* in a potato medium by Calmette and Guérin at the Pasteur Institute, Lille in 1882. As such, in 1921 the “Bacille de Calmette et Guérin” (BCG) was obtained and was then used as attenuated live vaccine against tuberculosis. BCG is the most widespread used vaccine in the world.

*M. microti* causes rodent tuberculosis and was first isolated in 1930 in UK. It has also been reported from severely immunocompromised patients.

*M. canetti*, a smooth variant that is very rarely encountered but causes human disease.

1.4. Mycobacterial cell wall

The important features shared by other members of *M. tuberculosis* complex include a cell wall of unique composition composed by a complex outer cell wall consisting of large amount of cell wall lipid. It consists of several unique components such as lipoarabinomannan (LAM), lipomannan (LM), phthiocerol dimycocerosate (PDIM), mycocerostate, mycolic acid, trehalose dimycolate (TDM) and sulpholipids (Beman et al., 1990; Bersa and Chatterjee, 1994).

![Diagram of Mycobacterial cell wall structure](image)

**Fig. 1.1.** Schematic representation of the Mycobacterial cell wall structure: contains 2 layer peptidoglycan layer adjacent to cell membrane and mycolyl-arabinogalactan layer on the outer side.
These components are suggested to be responsible for mycobacterial hydrophobicity, ability to form clumps or cords, ability to survive intracellularly and it is the cell wall that gives *M. tuberculosis* its acid-fastness, enabling it to retain basic dyes in the presence of acid alcohol (Madison *et al.*, 2001). The metabolic activity of mycobacteria, including assimilation of nutrients, energy production, metabolism and biosynthesis of macromolecules, are similar to those of other bacteria (Ratledge, 1982; Wheelar and Ratledge, 1994).

1.5. Pathogenesis of *M. tuberculosis*: stages of infection

Development of TB disease can be studied into three separate inter-related stages that occur with in the host infected by *M. tuberculosis*. In the first stage, individuals become infected by the inhalation of air-borne droplet nuclei containing *M. tuberculosis* (Fig. 1.2). Within the lungs, *M. tuberculosis* primarily infects resident phagocytic cells such as alveolar macrophages and dendritic cells, or alternatively, in monocytes recruited from peripheral blood. The bacteria are phagocytosed in a process that is initiated by bacterial contact with macrophage mannose and/or complement receptors (Schlensinger, 1993) or a macrophage cell entry protein expressed by virulent mycobacteria (Riley, 1995). On entry into the host macrophages, *M. tuberculosis* initially resides in phagosome which on its way to normal phagosomal maturation cycle forms phagolysosome, thereby exposing the bacilli to hostile environment that includes acidic pH, reactive oxygen intermediates (ROIs), lysosomal enzyme, and toxic peptides. Reactive nitrogen intermediates (RNIs) produced by activated mouse macrophages are major elements in antimicrobial activity (Nathan and Hibbs, 1991). The presence of RNIs in human macrophages and their potential role in disease has been the subject of controversy, but the alveolar macrophages of a majority of TB-infected patients exhibit iNOS (inducible nitric oxide synthase) activity (Nicholson *et al.*, 1996). The mycobacterial phagosome is not acidified (Crowle *et al.*, 1991). This is presumably due to the extrusion of proton ATPase from the mycobacterial phagosome (Sturgill-Koszycki *et al.*, 1994), but it is not clear that the blocking of endosomal maturation is essential for *M. tuberculosis* survival in macrophages. Infected macrophages in the lung, through their production of chemokines, attract inactivated monocytes, lymphocytes, and neutrophils (van Crevel *et al.*, 2002); none of which kill the bacteria very efficiently (Fenton *et al.*, 1996).
Adapted from: Zahrt et al., 2003,

**Fig. 1.2 Stages of M. tuberculosis infection:** M. tuberculosis infection is characterized by three interrelated stages of infection. In the first stage, M. tuberculosis acquired by inhalation is ingested by resident macrophages of the lung. Since macrophages are unable to eliminate bacilli, the bacterial burden increases to moderate levels during the initial stage of infection. The emergence of a cell-mediated immune response during the second stage of infection results in the formation of granulomas and recruitment of CD4+ T cells, CD8+ T cells, and activated macrophages to the site(s) of infection. Mature granulomas control bacillary growth and limit dissemination to extrapulmonary sites of infection. The third and final stage of infection is characterized by bacillary reactivation and subsequent secondary acute infection. M. tuberculosis in this stage exhibits uncontrolled bacterial growth, hematogenous spread to other infection sites, and secondary transmission to susceptible hosts. Reactivation typically results in death in those individuals not receiving antibiotic therapy.

Then, granulomatous focal lesions composed of macrophage derived giant cells and lymphocytes begin to form (Dannenberg and Rook, 1994). It has been demonstrated that TNF-α (Chensue et al., 1994) and IFN-γ are involved in granuloma formation (Enelow et al., 1992). In contrast, the course of infection during the second stage is primarily dictated by the host’s immune response. For example, infection of an immunocompromised host by M. tuberculosis will typically result in the establishment of an acute infection characterized by uncontrolled bacillary proliferation and dissemination of the organism to distal sites. Symptomatically, individuals suffering from acute disease exhibit persistent fatigue, anorexia, progressive weight loss, low-grade fever, and production of a chronic and contagious cough.
Alternatively, if the infected host is immunocompetent, the host’s immune system will typically resolve the initial infection, or alternatively, hold the infection in check using mechanisms that prevent further bacillary proliferation, limit the dissemination of the organism, and concentrate the immune response directly to sites of infection. Although these individuals continue to be persistently infected by *M. tuberculosis* and remain latent carriers, they do not exhibit overt signs of disease symptoms and are not infectious; however, they do test positive for a delayed-type hypersensitivity response. The third and final stage of infection is characterized by reactivation of the tubercle bacillus from latent stage, and subsequent initiation of secondary acute infection in the host. The mechanisms responsible for the transition from persistence to resumed growth following reactivation is presently unclear; however, it is likely to be influenced by factors associated with the host’s immune status, as conditions known to suppress the immune system including HIV infection, steroid therapy, age, and malnutrition, increase the likelihood of *M. tuberculosis* reactivation from latency (Zahrt et al., 2003).

### 1.6 Granuloma: primary residence of mycobacteria in the host

It is the primary mechanism to control the *M. tuberculosis* growth during different stages of infection and limits its dissemination. Granulomas are organized aggregates of immune cells such as macrophages, T cells that surround foci of infected tissues (Fig.1.3). Productive granulomas contain highly activated macrophages, which form multinucleate giant cells, and larger epithelioid-like cells that contain tightly interdigitated cell membranes surrounded by lymphocytic effector cells including CD4+ and CD8+ T cells. This structure walls off the organism from the rest of the environment. The stratified structure of the granuloma is such that it benefits both host and pathogen. From the host’s perspective, the granuloma limits dissemination of infection by ‘walling’ it off, yet this benefits the pathogen as well because the T cells capable of releasing macrophage-activating cytokines are restricted to the periphery of the granuloma, away from the infected macrophages. The formation of the granuloma involves both innate and acquired immune responses, and bacterial cell wall components such as trehalose dimycolate (cord factor) which is potent initiator of the response (Lima et al., 2001). Interior of the granuloma contains either calcified or necrotic tissue, is an area of low oxygen and high CO₂ concentration, contains hydrolytic enzymes and increased levels of aliphatic organic acids (Wayne et al., 2001). Finally, the activation of macrophages and other immune effector cells
surrounding the granuloma results in the release of numerous anti-microbial agents such as IFN-γ, TNF-α as well as reactive nitrogen intermediates and/or reactive oxygen intermediates.

![Diagram of granuloma composition](adapted_from: Bentrup et al., 2001)

**Fig.1.3.** Composition of mature granulomas: Granulomas are organized aggregates of immune cells surrounding *M. tuberculosis*-infected tissues(s). The granuloma center is composed of primarily necrotic tissue and macrophages that have fused to form multinucleated giant cells or epithelioid-like cells that contain tightly interdigitated membranes. The granuloma periphery is composed of activated macrophages and CD4⁺ and CD8⁺ T cells. The barricade-like structure of the granuloma acts to wall off *M. tuberculosis* and prevent further bacillary multiplication and spread to other infection sites.

While the significance of this environment on *M. tuberculosis* growth and survival during persistent infection remains relatively unclear, tubercle bacilli are frequently observed extracellularly in infected human tissues (miliary or extrapulmonary TB). When this happens, the person becomes infectious and requires antibiotic therapy to survive (Dannenberg and Rook, 1994).

### 1.7 Mycobacterial persistence

Three terms, latency, persistence, and dormancy are frequently used in literature describing *M. tuberculosis* and TB pathogenesis. Latency was defined by Amberson, (1938) as "the presence of any tuberculosis lesion which fails to produce symptoms of its presence". It's a state in which person has proven *M. tuberculosis* infection (either by skin test or by immunological test) without clinical signs or symptoms of tuberculosis. This represents a state in which the host immune response is able to contain the infection. Latency is achieved either by early restriction of *M. tuberculosis* growth in the lungs prior to the onset of TB disease, or through the spontaneous resolution of primary TB. Most people exposed to *M. tuberculosis* mount a vigorous cell-mediated immune response that arrests the progress of the infection, largely limiting it to the
initial site of invasion in the lung parenchyma and the local draining lymph nodes (called Ghon's complex) (Ghon, 1923). The complete elimination of the pathogen, however, is slow and difficult to achieve. Without antibiotic treatment, chronic or latent infection is thought to be the typical outcome of TB infection. Latent TB can reactivate after years or even decades of subclinical persistence, leading to progressive disease and active transmission of the pathogen.

Dormancy has been used to describe both TB disease as well as the metabolic state of the tubercle bacillus (Wayne, 1994; Gangadharam, 1995; Cunningham and Spreadbury, 1998). In terms of bacterial physiology, dormancy is defined as "a reversible state of low metabolic activity, in which cells can persist for extended period without cell division" (Kell et al., 1995). TB lesions are described as active or dormant, based on whether the associated pathology is progressing or healing, respectively. The word persistence literally means "continuing steadfastly or obstinately, especially in the face of opposition or adversity". Persistence is likely to be the combined effect of both the immune system and bacterial physiology (Bloom and Mckinney, 1999). As a pathogen, *M. tuberculosis* manifests its unusual capacity to persist in many ways. On the cellular level, mycobacterium resides within macrophages, cells that typically function to eliminate pathogens and other foreign materials from the body. At a more systemic level, *M. tuberculosis* is able to avoid elimination from the human host despite the development of vigorous cell-mediated immunity. Another less obvious but profoundly important manifestation of *M. tuberculosis* persistence is the slow rate at which the bacterium is cleared by anti-TB drugs.

1.7.1. Characteristic features of persistent infection

Different approaches have been used to explain the metabolic state of persistent bacteria. The controversy over whether the persistent bacteria are metabolically active and constantly replicating or in a spore-like form has been addressed using several different approaches. Chemoprophylaxis trials for latent tuberculosis with isoniazid showed efficacy in decreasing the risk of reactivation, thus, some bacterial metabolic activity is clearly present as activity of INH would function only in metabolically active cells (Comstock et al., 1979). Also, detection of Ag 85 transcripts by RT-PCR in multiple tissues that were negative for cultivable bacteria, further confirmed that the bacilli are viable and metabolically active (Pai et al., 2000). Results of these
CJ{eview of Literature

studies indicate an active, but restricted metabolism, present below the threshold level of immune detection.

The most peculiar characteristic of persistent state is related to its replication ability. Evidences in humans that the persisting bacteria are in a dormant, spore-like state comes from the results of culturing and staining diseased tissues from patients who have undergone chemotherapy, which might have resulted in false negative culturing results (Manabe and Bishai, 2000; McKinney, 2000; Parrish et al., 1998). Other hypothesis entails that persistent bacteria are either in a non-replicative state or have low levels of replication within the granuloma (Parrish et al., 1998), and a constant bacterial load is maintained by a balance between bacterial replication, and killing by the immune system. Mouse plateau model representing replication status of persistent bacteria has also shown that persistent bacteria either replicates very slowly or not at all (Rees and D'Arcy Hart, 1960). Another fact associated with persistent bacteria is their altered colony morphology and thickening of cell wall which make these persistent bacteria resistant to different drugs. Dormant form of bacilli show uncharacteristic staining properties and are no longer acid-fast (Seiler et al., 2003). Persistent forms of mycobacterium are resistant to anti-tubercular drugs, to which they are susceptible in the proliferating state of infection (Stewart et al., 2003). Drug that act on cell division processes and cell wall synthesis have a reduced effect on non-replicating cultures. The induction of stress-response defenses and a decrease in metabolic rate also tend to increase tolerance to drug action. Reduction in bacterial replication is also associated with a reduction in drug susceptibility.

1.8. Model systems of persistent infection

There are a number of important questions that remain to be answered with respect to latency and reactivation in tuberculosis. Since processes regulating latency and M. tuberculosis persistence can take years to manifest in the human host, several in vitro and in vivo systems have been developed to mimic aspects of latent infection. While these model systems are limited in their ability to fully recapitulate host and bacterial characteristics and have limitations, they provide a useful platform from which to initiate studies addressing specific aspects of the infection, dormancy and reactivation process. Furthermore, analyses using infection of surrogate mycobacterium species in their respective host(s) have provided additional insights into processes regulating latent tuberculosis (Ramakrishnan et al., 2000; Chan et al., 2002).
1.8.1. In vitro model of M. tuberculosis persistence

All in vitro models are based on the assumption that latent bacilli reside inside granuloma, and all in vitro parameters are defined in a way to mimic the harsh environmental conditions that bacteria encounter inside granuloma.

1.8.1.1. The Wayne non-replicating persistence model

Wayne's non-replicating model (Wayne and Hayes, 1996; 2001) is based on the assumption that M. tuberculosis encounters anaerobic conditions within granulomatous lesions inside the host. In this model, cultures of in vitro grown bacteria are subjected to gradual oxygen depletion by incubation in sealed culture tubes. Oxygen depletion triggers a dormancy response in the bacilli that is termed non-replicating persistence (NRP), a physiological state thought to mimic the one exhibited by M. tuberculosis during various stages of persistent infection. The transition of M. tuberculosis to a state of NRP is characterized by three distinct growth stages. In the first stage, the dissolved oxygen concentration in the medium is high and the bacilli exhibit normal logarithmic growth. This stage may recapitulate the physiological growth state of M. tuberculosis in vivo prior to the emergence of cell-mediated immunity. The second stage, called NRP-1, occurs when dissolved oxygen in medium reaches 1%. This stage mimics the physiological growth state of M. tuberculosis in vivo following the emergence of acquired immunity. During this stage, bacteria curtail protein synthesis, show discontinuation in DNA, and to some extent RNA synthesis, become resistant to several anti-mycobacterial drugs including isoniazid and rifampicin, and also show the upregulation of genes involved in glyoxylate shunt pathway which allow the utilization of alternative energy sources. During the third and final stage, NRP-2, dissolved oxygen concentrations drop below 0.06% in the culture medium, and bacilli switch from aerobic respiration to anaerobiosis. Bacilli in this state arrest growth at a uniform stage of the cell cycle. Furthermore, they exhibit sensitivity to metronidazole, a drug active against anaerobically growing organisms (Wayne and Saramek, 1994) and remain competent for growth reactivation following transfer into an oxygen-rich medium. The NRP-2 growth stage recapitulates the environment that M. tuberculosis encounters in vivo during periods of extended persistence within a mature granuloma. While the NRP cannot recapitulate the influence of the host's immune system, it may explain some characteristics observed during persistent infection of M. tuberculosis infection in humans (Yuan et al., 1998).
1.8.1.2. The nutrient starvation model

The second in vitro growth model is the nutrient starvation model (Betts et al., 2002). This model is based on the assumption that M. tuberculosis resides in tissues where nutrients and other essential cofactors are likely to be limiting. This model is initiated by the growth of M. tuberculosis in nutrient-rich medium, transfer of the culture in nutrient-limiting medium such as phosphate-buffered saline, and prolonged incubation under these conditions. M. tuberculosis cultures grown in this approach exhibited no loss in viability over an incubation periods of 6 weeks, decreased respiration rate and showed increased resistance to anti-tubercular drugs. At genetic level, nutrient starved cultures exhibit a global downshift in gene expression as determined by microarray analysis. 2-D gel analysis of cell extracts has further confirmed reduced protein level in the starvation model.

1.8.2. In vivo models

1.8.2.1. Cornell mouse model

The Cornell mouse model (also know as drug induced model) was the first animal model system for latent tuberculosis. This model was developed at Cornell University in 1950s (McCune et al., 1956). In this model M. tuberculosis infected mice are treated with antimycobacterial drugs, such as isoniazid and pyrazinamide, for a defined period (usually 12 weeks) which reduces the bacterial burdens to undetectable level in mouse lungs (McCune et al., 1966). Following cessation of antibiotic treatment, reactivation of infection can either occur spontaneously or in response to immunosuppressive agents like glucocorticoids. This model mimics the low bacterial burden in mice, as observed during human latent infection. But use of antibiotics for reducing bacterial burden is not similar to protective immune response developed in humans. Variants of Cornell model has been reported differing in inoculating dose, antibiotic regimen and rest period before immune suppression (Scanga, et al., 1999).

1.8.2.2. Low-dose murine model

The low-dose murine model (also known as the chronic or plateau model) was developed by O’rme (1988). In this model, low-dose bacterial inocula are administered to “genetically resistant” strains of mice such as C57BL/6 or BALB/c intravenously or via aerosol. This results in acute phase of bacterial replication followed by a steady state of infection directed by the
onset of adaptive immune response, followed by a stable maintenance of CFU at high levels in the lung over many months. This chronic model resembles latency in humans in that it depends on host immune response to contain the infection. However, due to large bacterial numbers and progressive pathology, it is more similar to chronic infection rather than human latent TB infection (Rhoades et al., 1997).

Other animal models also have been developed. One such model involves the use of \textit{M. marinum}, which causes TB in ectothermic hosts such as frogs and fish. This model is used to study mycobacterial pathogenesis (Cosma et al., 2003).

1.9 Proteins implicated in mycobacterial persistence

The ability of \textit{M. tuberculosis} to undergo persistence requires the coordinated expression of several bacterial virulence determinants at specific times during infection. The identification of genes expressed by \textit{M. tuberculosis} during period of persistence is of great importance because they may define the host conditions encountered by \textit{M. tuberculosis} during latent infection, as well as they allow the development of novel antibiotics and more efficacious vaccine that could target bacilli during latent/persistent stages and provide alternative strategies for the development of rationally attenuated \textit{M. tuberculosis} vaccine that are safer and more efficacious than the current \textit{M. bovis} BCG vaccine (Zahrt et al., 2003). A range of genes that are expressed during mycobacterial persistence have been identified. The products of these genes fall into five basic categories: respiratory enzymes, stress-related products and metabolic enzymes, proteins involved in fatty acid metabolism, PE/PE-PGRS proteins and regulatory proteins.

1.9.1. Respiratory Proteins

During persistence organisms undergo metabolic downshift, eventually allowing survival with minimal metabolic activity in the host (Wayne and Hayes, 1996; Smeulders, 1999). The bacteria adapt their metabolism to anaerobiosis by switching to nitrate respiration and reductive amination of glyoxylate (Wayne and Hayes, 1996). The most commonly expressed respiratory protein during persistent infection is NarX, a putative “fused nitrate reductase”. NarX might function as a respiratory nitrate reductase when oxygen is no longer available as the terminal electron acceptor. This gene is present in the form of cluster as \textit{narGHJI} in both \textit{M. tuberculosis} and BCG (Weber et al., 2000). It has been reported that \textit{narG} deficient mutant was unable to exploit nitrate as a source of oxygen and because of the scarcity of oxygen in the mature
granuloma, the bacteria showed reduced virulence as the infection progressed. Wayne and Hayes (1996) showed a marked increase in nitrite production in *M. tuberculosis* cultures grown with nitrate during hypoxic shiftdown from aerobic growth to anaerobic persistence. Another gene whose expression is up-regulated during hypoxic response is NarK2, a putative nitrite-extrusion protein has been found to be upregulated in response to hypoxic conditions (Sherman et al., 2001). Nitrate respiration helps the bacteria to survive in oxygen depleted areas of inflammatory or necrotic tissues where nitrate occur as an NO degradation product. Macrophages expressing inducible nitric oxide synthase (iNOS) release RNIs and GSNO, which are bactericidal *in vitro* at a pH characteristic of the phagosome of activated macrophages (Chan et al., 1991). Therefore, if the level of RNIs is elevated in the phagosomes of activated macrophages, a cascade of proteins could convert these compounds to NO₃⁻. Nitrate in turn could be used as a terminal electron acceptor under anaerobic condition, which would provide *M. tuberculosis* with a distinct survival advantage under nitrite conditions (Virtanen, 1960; Hutter et al., 1999).

### 1.9.2. Stress-response and general metabolic proteins

The nature of metabolic adaptations to the anaerobic non replicating state are largely unknown (Wayne et al., 1982). However, the *M. tuberculosis* genome revealed the presence of enzymes that are involved in anaerobic respiration and fermentation. In *M. tuberculosis*, transcription of *hmp*, increases under microaerophilic conditions when the culture enters stationary phase and under nitrosative stress. This enzyme is homologous to genes encoding flavohemoglobin in *E. coli*. Hmp protein plays a role in protecting *M. tuberculosis* from nitrosative stress, which is the main mechanism utilized by host to clear mycobacterial infections. This protein might act as an NO di-oxygenase, which converts NO to NO₃⁻ (aerobically) or N₂O (anaerobically), thereby functioning as an NO trap when induced by NO or nitrosative stress (Poole et al., 2000).

Another protein, whose expression increases manifold, is 16 kDa *α*-crystalline like protein (Acr protein). This protein is associated with stationary phase (Yuan et al., 1996), is induced under RNI stress (Garbe et al., 1999) and is an immunodominant antigen that is necessary for survival in macrophages (Yuan et al., 1998). When subjected to anaerobic stress, *M. tuberculosis* and *M. bovis* BCG induce a massive up-regulation of the production of this protein, which is associated with a thickened cell envelope (Cunningham et al., 1998; Lim et al., 1999; Boon et al., 2001; Desjardin et al., 2001). This protein can stabilize cell structure during long-term survival and
permits the bacilli to survive within the low-oxygen environment of the granuloma (Wilkinson et al., 1998). Its expression is controlled by the sigma factor SigF, an alternate RNA polymerase sigma factor is expressed in stationary phase and under stress conditions in vitro (Michele et al., 1999). It is nonessential both in axenic culture and for survival in macrophages in vitro. However, although loss of SigF does not prevent the mutant strain from producing a lethal infection, death of BALB/c mice infected by the mutant strains was significantly delayed (Chen et al., 2000).

1.9.3. Proteins involved in fatty acid metabolism
The most important proteins, activated during adaptation to persistence and microaerophilic conditions (Wayne et al., 1982) involve enzymes of glyoxylate bypass. There is a significant increase in a putative enzyme glyoxylate dehydrogenase and isocitrate lyase (ICL) activity. The up regulated expression of these enzymes allows *M. tuberculosis* to utilize acetyl Co-A as a carbon source produced by the metabolism of fatty acid via the β oxidation cycle and produces succinate which is the main precursor for the synthesis of sugars. Consequently, enabling *M. tuberculosis* to synthesize carbohydrate from fatty acids, as well as supply intermediates to support the TCA cycle. This is obviously of major importance when fatty acids are the main source of carbon and energy, as has been suggested to be the case for *M. tuberculosis* and *M. leprae* in chronically infected tissues (Bloch, 1956; Wheeler and Ratledge, 1988). Recently studies have revealed that ICL expression is up regulated during infection of macrophages by mycobacterium species (Höner zu Bentrup et al., 1999). It has been shown that *M. tuberculosis* in which the gene encoding ICL was deleted, impaired survival of bacilli occur in activated murine macrophages as well as in the late stage infection in the lungs of mice. The requirement for the ICL gene for bacterial survival during late-stage infection indicates a change of environment that require the bacteria to alter their diet from carbohydrate to lipid and also confirm the link between the immune status of the host and the metabolism of *M. tuberculosis* because the dependence on ICL seems to coincide with the onset of acquired immunity against the bacterium. The abundance of genes encoding enzymes involved in fatty acid degradation supports the suggestion that *M. tuberculosis* uses host lipids while growing in vivo (Cooper, 1999). This was further confirm by and observed that in vitro-grown bacteria had a preference
for carbohydrates, whereas in vivo-grown bacteria preferred fatty acids (Nathan and Shiloh, 2000).

1.9.4. PE/PE-PGRS proteins

The PE/PE-PGRS proteins represent a novel class of proteins found in the genomes of several pathogenic Mycobacterium spp. including M. tuberculosis and M. marinum. The PE proteins have conserved Pro-Glu motif near their N-terminal, and in case of PE-PGRS proteins, there is C-terminal extension of tandem repetition of Gly–Gly–Ala or Gly–Gly–Asn (Cole et al., 1981). In mycobacteria, genes from this family accounts for nearly 5% of the coding sequences, and suggested that these genes are required for virulence and persistent infection. First, several proteins from this family localize to the cell surface and influence cell surface interactions between mycobacteria and the macrophage (Brennan et al., 2001). Second, significant humoral and cellular immune responses are generated in vivo following the expression of several PE-PGRS and PE family members, respectively (Delogu, 2001). Finally, a PE-PGRS protein is a dominant antigen recognized by sera of asymptomatic latent carriers (Singh et al., 2001). The role of PE and PE-PGRS in mycobacterial persistence is assured by the expression of several genes from this family in persistent infection by M. marinum (Ramakrishnan et al., 2000). For example, several PE and PE-PGRS genes are expressed following M. marinum infection in macrophages. Some genes are involved during the growth of M. marinum in granulomatous lesions in vivo, while others are required for long-term persistence in the poikilothermic animal model system of infection. For example, an M. marinum strain carrying a mutation in a PE-PGRS gene, mag-24 is defective in persistent infection and exhibited reduced bacterial burdens in the spleen and liver of infected frogs. Frogs infected with this mutant exhibited attenuated granulomatous response. Taken together, these results suggest a role for PE-PGRS genes in aspects of persistent infection that include modulation of the host's immune response.

1.9.5. Transcription factors (Regulatory proteins implicated in mycobacterial persistence)

Apart from the requirement of secondary metabolism systems and lipids, genes encoding transcription factors are also required by M. tuberculosis for long-term persistence. Transcription factors play an important role in modulating bacterial response during infection, because they provide a direct mechanism to quickly initiate adaptive responses. At a molecular level, transcription factors allow the tubercle bacilli to rapidly increase or decrease effectors gene
expression in response to changes in the local environment. The large number of putative transcription regulators identified in the *M. tuberculosis* genome sequence indicates that much of the regulation required for these adaptations by *M. tuberculosis* occurs at the transcriptional level.

1.9.5.1. Two-component signal transduction systems

Two-component signal transduction system mediates adaptive processes in response to physical or chemical environmental stimuli. *M. tuberculosis* has eleven complete two-component systems (Cole *et al.*, 1998), and one of these systems, *mprA-mprB*, participate in persistent processes (Zahrt *et al.*, 2001). The expression of this system in bacteria is tissue- and time course-specific. Mice infected with an *mprA* mutant strain of *M. tuberculosis* exhibit reduced bacterial burdens in the lungs during persistent, but not acute stages of infection, and the growth of the *mprA* mutant is reduced in the spleen during both acute and persistent stages of infection. Several other *M. tuberculosis* two-component systems have also been implicated but their role in persistent infection remains unclear.

*phoP* (Rv0757), which codes for a putative transcription regulatory factor of the two component system PhoP/PhoR, is a response regulator required for intracellular growth of *M. tuberculosis* during acute stages of infection (Perez *et al.*, 2001; Walter *et al.*, 2006). The *prrA-prrB* system is expressed in *M. tuberculosis* during growth in human macrophages *in vitro* (Graham and Clark-Curtiss, 1999), and is required during the initial stages of acute infection *in vivo* (Ewann *et al.*, 2002).

A novel two-component system *devS/devR* was reported which is expressed at higher levels in *M. tuberculosis* H37Rv as compared to H37Ra (Das Gupta *et al.*, 2000) and may have role in bacillary persistence.

The study of two-component systems offer a multifaceted approach for understanding latency related processes because identification of the signals recognized by sensor kinase proteins, such as MprB, will ultimately define host conditions experienced by *M. tuberculosis* *in vivo* and determination of the downstream effector genes regulated by response regulator proteins, such as MprA, will help define virulence determinants required for the establishment and maintenance of persistent infection.
1.9.5.2. Sigma factors

A second group of transcription factors, required by *M. tuberculosis* during persistent stages of infection are sigma factors. These proteins are subunits of RNA polymerase and are responsible for directing the transcription of genes. In *M. tuberculosis*, 13 sigma factor genes have been annotated in the genome (Cole et al., 1998), nine of which belong to a special subfamily and direct extra cytoplasmic functions and various other stress responses. For example several sigma factors are activated following exposure of *M. tuberculosis* to various environmental stresses *in vitro*, including temperature, oxidative stress, pH, and infection of macrophages (Manganelli et al., 1999).

One such sigma factors, *sigH* is required by *M. tuberculosis in vivo* for progressive pulmonary disease during latent infection (Kaushal et al., 2002). Although *sigH* expression is not required for the growth and survival of *M. tuberculosis* in a mouse model of tuberculosis, infection with an *M. tuberculosis* *sigH* mutant results in a significant reduction in overall lung histopathology during persistent stages of infection, as well as reduction in the recruitment of CD4+ and CD8+ T cells to infection sites. Microarray analyses of the *sigH* mutant suggest the up-regulation of genes that are involved in resistance to oxidative and other denaturing stresses (Kaushal et al., 2002). Thus, *sigH* is likely to play an important role in modulating gene expression in the tubercle bacilli in response to *in vivo* conditions, including those involved in host immunity.

1.9.5.3. WhiB3

A third putative transcription factor expressed in *M. tuberculosis* during persistent infection is *whiB* (Steyn et al., 2002). This determinant, *whiB*3 is homologous to the *whiB* gene of *Streptomyces coelicolor*, a transcription factor involved in sporulation processes. In *M. tuberculosis*, WhiB3 interacts with the C-terminal region of RpoV, the principal sigma factor of strain H37Rv, to activate expression of yet-to-be-identified virulence determinants. The role of *rpoV* in *Mycobacterium* virulence is already known, as an *rpoV* mutation in *M. bovis* reduces virulence of this strain in a guinea pig model of tuberculosis (Collins, 1995). In *M. tuberculosis*, mutations in *whiB3*, although affect bacterial persistence but overall bacterial burden remains same. Similarly, *M. tuberculosis whiB3* mutant grows normally in the organs of infected mice during either acute or persistent stages of infection. These mice exhibit significantly prolonged survival times compared to mice infected with wild-type *M. tuberculosis*. This increase in
survival time is due to the result of changes in bacterial gene expression in the whiB3 mutant that reduce the cell-mediated immune response.

1.9.5.4 Rv3291c (LrpA)

Rv3291c is annotated as putative transcription factor of Lrp/AsnC family (leucine-responsive regulatory protein/ regulator of asparagine synthase C gene product) in the *M. tuberculosis* H37Rv genome data bank (www.sanger.ac.uk). The Lrp/AsnC family of transcriptional regulators includes proteins that act as global or specific regulators of transcription (Brinkman *et al.*, 2000). These proteins are known to regulate amino-acid metabolism and related processes. Members of this family have been identified in 45% of bacterial genomes and 95% of archaea. To date there are no confirmed homologues available in eukaryal genomes, indicating that this family is probably restricted to prokaryotes (Brinkman *et al.*, 2003). LrpA is a global regulator that controls the expression of a large number of operons in *E. coli*, including those involved in the synthesis and degradation of amino acids (Brinkman *et al.*, 2003). As the name implies, many Lrp responsive operons are co-regulated by L-leucine that can either antagonize or potentiate the effect of Lrp (Calvo *et al.*, 1994). Members of the Lrp/AsnC family typically have a molecular mass of nearly 15 KDa but populate a range of multimeric species in solution that include dimers, tetramers, octamers and hexa decamers (Willins *et al.*, 1991; Madhusudhan *et al.*, 1995; Brinkman *et al.*, 2000; Jafri *et al.*, 1999; Chen *et al.*, 2001).

Homologues of Lrp and AsnC are referred as the Feast/ Famine Regulatory Proteins (FFRPs). Calvo & Matthews coined the term Feast/ Famine Regulatory Proteins (Calvo *et al.*, 1994) for Lrp/AsnC proteins to summarize the general function of these proteins. The term Feast/ Famine is used for a class of proteins that are involved in sensing the concentration of amino acids/ effector molecules from nutrient media and effecting appropriate changes to gene expression (Suzuki, 2003). These proteins are also involved in DNA bending, condensation of DNA into globular nucleo-protein structures, chromosome structure and organization among other roles (Brinkman *et al.*, 2000; Leonard *et al.*, 2001). *E.coli* contain three members of this family: LrpA, to which Mtb LrpA has 40% amino acid identity, YbaO and AsnC, which regulate asparagine biosynthesis (Kolling and Lother 1985).

Crystal structure for six proteins in the Lrp/AsnC family have been solved to date: LrpA in *Pyrococcus furiosus* (Leonard *et al.* 2001), FL11 in *Pyrococcus OT3* (Koike *et al.* 2004), LrpA
These structures reveal that proteins in this family typically have two domains: (i) a DNA binding domain which contain the helix-turn-helix (HTH) motif, which has been proposed to be involved in DNA binding (Platko and Calvo 1993). The $\alpha_1$ and $\alpha_2$ helices are joined by a short turn that contains a critically conserved glycine at position 32 (fig.1.4). All HTH-containing proteins characterized to date have a glycine at the analogous position in the HTH motif (Mondragon et al., 1989) and (ii) a regulatory domain for binding the effector. This second domain is termed as RAM domain (Regulator of amino acid metabolism) for binding the effector (Ettema et al., 2002). Lrp proteins naturally form dimers and likely interact with helical DNA in this form. However, they have been reported to form a number of higher order oligomers (tetramers, hexamers, octamers), which is believed to be related to formation of histone-like particles with the DNA helix wrapped around it and making interactions at multiple evenly spaced sites (Koike et al., 2004). Rv3291c form stable protein-DNA complexes. Recently it was

**Figure 1.4 Overview of M. tuberculosis LrpA structure.** Ribbon diagram of the *M. tuberculosis* LrpA subunit secondary structure. The helix-turn-helix (HTH) motif is formed by residues Leu 5–Ser 47. The C-terminal domain consists of residues Leu 67–Ile 137. The HTH domain and the C-terminal domain are connected by a two-hinge region, which encompasses residues Arg 48–Leu 66 and Ile 138–Pro 150. The $\alpha$-helices are colored in rust, the $\beta$-strand is colored in purple, and the connecting loops are colored in yellow.
shown that a specific oligomeric transition between hexadecameric and octameric/lower order oligomers in presence of Phe that supports an effector-mediated model for the disassembly of a nucleosome like particle (Shrivastava et al., 2009). In the recent study, two mutants were generated Gly102Thr and Glu104Ala, which are part of the essential 100-106 effector binding loop (Shrivastava et al., 2009). The Gly102Thr mutant adopts an unusual ‘open’ quaternary structure and offers interesting functional insights co-related to the binding of an effector.

LrpA is strongly up-regulated during starvation (Betts et al., 2002), implicating it in response to nutrient limitation in M. tuberculosis, and therefore suggesting it might play a role in persistence (Gomez and McKinney 2004). There are several in-vitro models of persistence, including stationary-phase growth (DeMaio et al. 1996; Voskuil et al. 2004) and hypoxia (Wayne and Hayes 1996; Wayne and Sohaskey 2001). Nutrient starvation is another important model for M. tuberculosis dormancy, as a nutrient-limiting environment is suspected to develop in the granulomatous lesions that encase the sites of infection (Loebel et al., 1933a, b). Betts et al (2002) developed an in-vitro system to characterize the effects of starvation on M. tuberculosis gene expression. This culture system generates non-replicating bacilli that exhibit a low level of respiration and exhibit no loss of viability in this nutrient deprived condition for up to 6 weeks. The bacilli develop resistance to isoniazid and rifampicin similar to persistence in-vivo (Gomez and McKinney 2004), but don’t acquire susceptibility to metronidazole as in hypoxic conditions (Wayne and Sohaskey 2001), suggesting differences in the physiological responses. The Rv3291c has also been observed upon inhibition of septum formation and is inversely proportional to bacterial growth rate consistent with a feast/famine regulatory role (Slayden et al., 2006; Landgraf et al., 1996). It also binds to the promoter region of lysine-ε-aminotransferase (lat), 150 bp upstream of the translational site of lat. A 23bp region located at -55 to -33 resembles the AT-rich E.coli LrpA binding site pattern having sequence CATAATTT TTCCAGCATAAATAT (Reddy et al., 2008).

Recently it was shown that in M. fortuitum the lrpA homologous gene has been found to be highly conserved and in-vivo screening involving a murine infection model of persistence using this pathogen led to the identification of a mutant by transposon insertional mutagenesis where this gene was disrupted. It was found that the mutation affected the ability of the pathogen to persist in the kidney in the murine infection model (Parti et al., 2008). It was also shown that a
mutant of *M. fortuitum* deficient in virulence and persistence in a murine infection model could be complemented by the *M. tuberculosis* rv3291c (Srivastava et al., 2009). The above data suggest that the protein is important for maintenance/adaptation to long term persistence although it is not essential for growth under normal conditions (Sassetti et al., 2003). Rv3291c therefore has the potential to be an important therapeutic target for developing strategies to counter persistence of the disease.

**1.10 Techniques used for study of mycobacteria**

Study of mycobacteria presents several problems. The bacterium has a very long generation time of 15-20 hr. Normal transformation and other genetic manipulation techniques are very difficult to apply in mycobacteria (McFadden, 1996). Earlier approaches dealt largely with developing methods for creating mutations in specific genes of fast-growing nonpathogenic species because of the relative ease of working. Several techniques were developed to inactivate genes, with the first reports of *M. smegmatis* and *M. bovis* BCG transformation being published in 1988 (Snapper et al., 1988) and the subsequent development of a highly transformable *M. smegmatis* strain, mc2 155, being published in 1990 (Snapper et al., 1990). The choice of genes to mutate and inactivate to study virulence was frequently based on (i) existence of naturally occurring mutations in virulent strains that affected pathogenicity, or (ii) on predictions of genes important in some aspects of virulence and/or physiology of *M. tuberculosis*, or (iii) by inference from studies of other bacterial pathogens, for example, genes encoding sigma factors and iron acquisition regulators. Current genetic methods are far more robust and have higher rate of success. With the combination of various reporter systems, it has been possible to unravel several aspects of mycobacterial virulence, some of which are discussed below.

**1.10.1 Directed Gene Disruption (Gene knockout)**

The method involves the insertion of an antibiotic resistance cassette in the gene of interest using allelic replacement.

The directed gene replacement is mostly done using shuttle plasmid containing the cassette for gene replacement. It usually entails a two-step process in which the plasmid containing the desired gene disrupted with an antibiotic resistance cassette integrates into the genome by a single crossover event (Campbell-type integration) at the region of homology, selecting for the
antibiotic resistance cassette. The event forms a direct repeat at one of two positions relative to the antibiotic cassette. In a second crossover event, the plasmid backbone is excised by means of recombination at the other direct repeat than the one initially used, when the antibiotic selection is maintained, resulting in the desired gene disruption. The technique can be relatively efficient when coupled with counterselection to facilitate plasmid elimination in the second step. One type of counterselection is sucrose resistance when *sacB* encoding levansucrase is in the plasmid backbone; the enzyme converts sucrose to levans. Levans are toxic to mycobacterial cells which lack functional levanase that can covert levans to fructose and glucose. The presence of the vector in the bacterial chromosome prevents bacterial growth on selective media containing sucrose. The drawbacks of the two-step procedure is the time required for the process, i.e., approximately 3 to 5 months from the initial transformation to the verification, by DNA analysis, of the gene disruption in survivors after counterselection. In addition, mutations that inactivate the plasmid-borne *sacB* during the selection and counterselection procedure will be erroneously scored as desired events in which the plasmid has been excised from the chromosome.

A new method of directed gene inactivation, termed specialized transduction, using a *ts* bacteriophage delivery system has been described recently that bypasses these problems and is the current method of choice for directed mutation (Bardarov et al., 2002). The technique is very efficient because essentially all of the recipient cells can be transduced and the selection is robust.

### 1.10.2 Transposon Based Global Gene Inactivation

The principle of transposon mutagenesis involves insertion of foreign DNA, a transposable element, into many sites in the bacterial genome, ideally on a completely random basis. These events require a selectable phenotype, generally an antibiotic resistance marker carried within the transposon. Transposons are DNA elements that can move from one genetic location to another. They contain defined terminal inverted repeats and encode a transposase that excises the transposable element from a donor site and rejoins it to DNA at a second location. Transposition intermediates called "transposomes" are synaptic complexes between two transposase molecules and any DNA with inverted repeat ends (Goryshin et al., 2000). During the excision of the transposable element from the donor DNA a transposome is formed in which the transposon ends are brought together by the dimerization of the transposase. Magnesium ions are essential for the
catalytic activity of transposomes, these ions activate the transposase and allow insertion of transposon into cellular DNA by a cut and paste mechanism (Goryshin et al., 2000; Hoffman et al., 2000). Genomic mutagenesis is thus possible without conjugations between bacterial strains.

The transposome method mentioned above can also be used to make random mutations in mycobacterial genomes (Detweiler et al., 2001). The Tn5 based transposon inserts randomly (Meis., 2000) and can create knock-out in genes. Tn5 normally has a very low rate of transposition in bacteria. Efficient transposition systems have been developed which produce integration events in the genomes of mycobacterial species. Transposons can be introduced into genomic DNA by use of 'suicide vectors', or bacterial phages, carrying transposon.

The transposon mutagenesis system has been utilized to identify M. tuberculosis virulence genes by using the signature-tagged mutagenesis (STM) (Arruda et al., 1993). STM is a powerful method that enables the identification of transposon mutants that are defective for growth in an animal host, by allowing simultaneous screening of large number of mutants. The hallmark of STM is that the transposons in individual mutants contain distinct DNA tags. Attenuated mutants are identified as distinctly marked bacteria present among a pool of mutants, used in an initial inoculum to infect the host but absent or under-represented in the host after a defined period of infection. Uniquely tagged transposon M. tuberculosis mutants were made in broth cultures and were negatively screened by hybridization for those that did not survive during animal infections (Camacho et al., 1999). A transposition system has been developed more recently, its recognition sequence is simply an A-T base pair and is be expected to be truly random in its integration. The Himar-1 system has been used to introduce transposition events in M. smegmatis (Rubin et al., 1999), M. bovis BCG (Sassetti et al., 2001)

1.10.3 Antisense Methods

Antisense technology involves use of complementary RNA, which binds to the mRNA of a specific gene thereby preventing its translation. Antisense technology is particularly useful for silencing of essential genes as their knockout mutants are unable to survive, and in systems where gene inactivation is difficult. Regulatable acetamide/acetamidase system is generally used for conditionally controlling the production of antisense RNA in mycobacteria (Parish et al., 1997). Applications of the antisense method were in M. bovis to lower the levels of ahpC (Wilson et al., 1998) and in M. tuberculosis H37Rv to reduce the amounts of sodA (Edwards et
Recently, the antisense approach was used to decrease the level of sigA (Wu et al., 2004) and rv3303c in *M. tuberculosis* (Akhtar et al., 2006).

### 1.10.4 Subtractive Hybridization and SCOTS

Initially developed in eukaryotic cells, subtractive hybridization is now being used to study prokaryotic systems, such as the infection of cultivated macrophages by *M. avium* (Plum and Clark-Curtiss 1994). In this study, mycobacterial mRNA was converted to cDNA by reverse transcription. Biotin-labelled cDNA prepared from *M. avium* grown in broth was used to subtract constitutively expressed housekeeping genes from the cDNA of macrophage-derived *M. avium* using streptavidin-coated paramagnetic beads. Unfortunately, the construction of representative cDNA libraries for subtractive hybridization is hampered by the instability of bacterial mRNA and the problems of isolating sufficient high quality mRNA from small populations of bacteria growing *in vivo*. In addition, transiently expressed genes may not be well represented in cDNA library.

### 1.10.5 *In vivo* Induced Antigen Technology (IVIAT)

IVIAT is designed to identify antigens expressed specifically *in vivo* during human infection (Handfield et al., 2000; Handfield et al., 2003). Convalescent sera from infected individuals are pooled and adsorbed against a pathogen of interest grown under typical laboratory conditions, thus retaining antibodies that recognize antigens specifically expressed during *in vivo* growth. Adsorbed serum is then used to probe an inducible genomic expression protein library prepared from the cognate organism and identified clones are subject to further characterization to confirm antigenicity, and pathogen DNA contained within reactive clones is analysed to identify the reactive gene product. IVIAT has been applied for identification of *in vivo* induced genes in *M. tuberculosis* (Deb et al., 2002), *Vibrio cholerae* (Hang et al., 2003), *Candida albicans* (Cheng et al., 2003).

### 1.10.6 Microarray

The availability of genome sequences has facilitated an important development of nucleic acid hybridization, the DNA microarray or chip (Ness, 2006). Microarrays are generally manufactured on glass microscope slides, and consist of an ordered grid of thousands of DNA spots. These can either be oligonucleotides or PCR products corresponding to every predicted
ORF on the genome. Hybridization of labelled mRNA to the microarray allows the relative level of expression of each gene to be measured. Microarray has made it possible to get information for the expression of all bacterial genes from a single experiment.

The first bacterial application was described by De Saizeu et al. (1998) using an oligonucleotide-based microarray to look at expression of a subset of 100 genes from *Streptococcus pneumoniae*. DNA microarrays have the potential to revolutionize the study of *in vivo* induced (ivi) gene expression since they can yield global information concerning temporal and spatial aspects of gene expression.

**1.10.7 Reporter Fusions and Promoter Traps**

Reporter technology has become the method of choice to study the bacterial gene expression under various host conditions. Use of various reporters has helped in analysis and characterization of various promoter elements. Specific promoter sequences from known genes are cloned upstream of the promoterless reporter gene and electroporated into the mycobacterial cell so that the activity of the gene in question can be simply measured under different conditions e.g. during infection of macrophages (Dellagostin et al., 1995) and iron starvation (Rodriguez et al., 1999).

Among the reporter genes used for this purpose in mycobacteria and the proteins they encode are *lacZ* (β-galactosidase), *xylE* (catechol-2,3-dioxygenase), *phoA* (alkaline phosphatase), *lux* (luciferase), *gfp* (green fluorescence protein, GFP), and *cat* (chloramphenicol acetyltransferase).

**β-galactosidase (lacZ)**

*E.coli lacZ* gene encodes for the β-galactosidase enzyme (Fowler and Zabin, 1983), which catalyzes the β-1,4 cleavage of lactose to liberate monomeric sugars glucose and galactose. β-galactosidase reporter system has been extensively used to monitor gene expression in response to various environmental conditions, including the activity of promoters during intracellular growth of mycobacteria within macrophages (Dellagostin et al., 1995). The β-galactosidase activity can be easily assayed in cell extracts or intact mycobacterial cells (Srivastava et al., 1997; Jain et al., 1997; Das Gupta et al., 1998). However, a major limitation of the β-galactosidase reporter system is that it is invasive, requiring permeabilization of target bacterial cells. Because efficiency of permeabilization varies from cell to cell, observed differences in β-
galactosidase activity can reflect altered substrate uptake rather than changing levels of gene expression, thus preventing accurate analysis.

**Luciferases**

In recent years, bioluminescence assays have become a popular method to quantify reporter activity (Bronstein *et al.*, 1994). Vectors containing luciferase gene provide highly sensitive and rapid quantitation of gene expression. Of these, the gene encoding firefly luciferase has been used widely. The most well-characterized firefly luciferase is that isolated from the common North American firefly *Photinus pyralis*. Firefly luciferase catalyzes the ATP-dependent oxidative decarboxylation of luciferin (LH₂) to oxyluciferin, resulting in the production of light as shown in the reaction, where ‘P’ denotes the product oxyluciferin.

\[
\text{Mg}^{2+} \quad \text{LH}_2 + \text{ATP} + \text{O}_2 \quad \rightarrow \quad \text{P} + \text{AMP} + \text{PPi} + \text{CO}_2 + \text{hv}
\]

The assay is highly sensitive and can be performed with intact cells (Bronstein *et al.*, 1994). Firefly luciferase is useful in a variety of applications. As there is no endogenous level of luciferase in mycobacteria, it is an attractive reporter for monitoring *in vivo* gene expression in pathogenic mycobacteria such as *M. tuberculosis* (Jacobs *et al.*, 1993). Because of its specificity for ATP, it can be used to measure the amount of ATP present in biological samples with no interference from other nucleotide triphosphates.

**Green Fluorescent protein**

Green Fluorescent protein (GFP) was obtained from jellyfish *Aequorea victoria*, which emits green fluorescence naturally after excitation by blue light. It has gained rapid and widespread popularity as a reporter of gene expression on account of its special properties (Tsein *et al.*, 1998). It does not require the addition of any substrate or co-factor. GFP is a small protein producing hybrid proteins with fluorescence when fused to gene of interest, making it ideal for protein localization studies. It can be used for detection and quantification of bacterial gene expression in *in vitro* and *ex vivo* conditions and real-time visualization of gene expression in different pathogens. GFP has been successfully used for *in vivo* analysis of mycobacterial gene expression to study mycobacterium–macrophage interactions (Dhandayuthapani *et al.*, 1995; Kremer *et al.*, 1995).
1.11. Promoters

1.11.1 Prokaryotic Promoters

Prokaryotic RNA polymerase comprising of the four types of subunits (α, β1, β2, and σ) form a tight complex with the promoter region. The sigma (σ) unit of the enzyme is believed to be responsible for directing the RNA polymerase to the promoter for binding and transcription initiation. Most bacteria produce a primary sigma factor and several alternative sigma factors. *E. coli* have five sigma factors:

Primary sigma factor, such as *E. coli* Sigma 70 directs the synthesis of most mRNA in the actively multiplying cells and is essential for growth. The ability to induce or activate alternative sigma factors that can replace the primary sigma and change the RNA Polymerase promoter specificity is a mechanism that allows the expression of different set of genes in response to varying environmental conditions. Sigma 54 is required for nitrogen fixation, flagellum and pilus biosynthesis. There are other sigma factors such as sigma 32, sigma 28, sigma 38 in *E. coli*.

Prokaryotic promoters are less complex (size and number of elements recognizable by sigma factors) than their eukaryotic counterparts though there are some similarities. Prokaryotic promoters vary in their affinities for RNA polymerase, a factor very important with regard to controlling the frequency of transcription and therefore the extent of gene expression. Unregulated transcription initiations at many prokaryotic promoters have been found to require only an RNA polymerase holoenzyme, which consists of four core subunits with a dissociable sigma factor. Multiple sigma factors have been identified and each programs the core enzyme to transcribe from different class of promoters. Prokaryotic promoters direct not only the site of transcription initiation but also the rate of transcription. Earlier studies (Chamberlin, 1974) have established that promoter strength (as defined by degree with which transcripts are produced) is primarily determined by two factors: the binding affinity to RNA polymerase and the rate of isomerization from ‘closed promoter complexes’ (DNA remains duplex) to ‘open promoter complexes’ (DNA opened by ‘melting’).

1.11.1.1 *Escherichia coli* Promoters

More than 300 promoters have been experimentally characterized by various researchers. A striking observation is lack of any extensive conservation of sequence over the 60 bp commonly associated with interaction with RNA polymerase. There are four notable features in most *E. coli*
promoters; the transcriptional start site, the -10 hexamer, the -35 hexamer and the distance between the -10 and -35 sequences.

The transcriptional start site has been found to be purine in more than 90% of characterized promoters. Just upstream of the start site, a six base pair (bp) region is recognizable in most promoters called -10 sequences. The distance from transcription start site varies in known promoters from 18 to 9 bp. Its consensus is TATAAT and can be summarized in the form T_{80} A_{95}T_{45}A_{60}A_{50}T_{96} where the subscripts denote the per cent occurrence of the most frequently found base.

The other conserved hexamer is around ~35 bp upstream of the start site. The consensus for -35 has been universally accepted as TTGACA. The distance separating the -35 and -10 sites has been found to be between 16 and 18 bp in 90% of the promoters (Hawley and McClure 1983). The distance may be critical in holding the two sites at the appropriate distance for the geometry of RNA polymerase. The -35 region is said to provide the signal for recognition by RNA polymerase, while the -10 sequence allows the complex to convert from 'closed' to 'open' form.

Other researchers have established another important sequence element in addition to the four mentioned in some E. coli promoters (Ross et al., 1993). The seven E.coli rrn genes, which encode ribosomal RNA, are unusually strong, accounting for more than 60% of total RNA system in rapidly growing cells. The exceptional strength of the rrn promoter has been attributed to an AT-rich sequence of ~20 bp located immediately upstream of the -35 region. This region with the AT-rich motif has been termed upstream element or UP element (Ross et al., 1993).

1.11.1.2 Mycobacterial Promoters

The M. tuberculosis H37Rv contains 13 potential sigma factors which determine the differential regulation of various genes (Hatfull and Jacobs, Jr. 2000). In addition to transcription of genes essential for growth, these sigma factors regulate mechanisms for stress response, sporulation, heat shock and extracytoplasmic factors (ECF).

SigA and SigB are the two principal/ principal-like sigma factors found in both fast and slow growing mycobacteria. SigA is involved in the transcription of several house keeping genes and mutation leads to attenuation of virulence (Collins et al., 1995). SigB is located downstream of SigA and its expression increase under oxidative stresses such as exposure to SDS and H_2O_2 and heat shock (Manganelli et al., 1999).
SigE and SigH have a role in stress response to pH, heat shock, SDS and H$_2$O$_2$ stress. SigF is a stress response factor. It is upregulated in stationary phase and in response to cold shock, oxygen depletion and exposure to alkyl peroxides (Michele et al., 1999).

GTG is often used as start codon in mycobacteria instead of ATG. The Shine Dalgarno sequences of mycobacteria are located at varying distances from translation start site and have minimal complementarity to the 3'end of mycobacterial 16S rRNA, which suggests that ribosome binding site recognition may be less stringent in the mycobacteria, as in Streptomyces species. Some mycobacterial promoters like \textit{M. tuberculosis katG} gene promoter require for optimal expression, a 155 bp region which is located about 300 bp upstream of the translation start codon and approximately 200 bp upstream of -35 region. Similar region of 41 bp is present 269 bp of the \textit{M. tuberculosis recA} gene and is essential for expression (Mulder et al., 1997).

Multiple promoters are also a common feature in mycobacteria. These include \textit{ahpC} gene in \textit{M. smegmatis}, \textit{glnAI}, \textit{recA} and \textit{katG} of \textit{M. tuberculosis}, and \textit{hsp60} of \textit{M. bovis} BCG (Mulder et al., 1997).

Mycobacterial promoters show a conserved -10 region with consensus sequence T$_{80}$A$_{90}$Y$_{60}$G$_{40}$A$_{60}$T$_{100}$ for \textit{M. tuberculosis} where Y is pyrimidine base and the numbers indicate the percent conservation. However there is poor consensus at -35 regions. It is thought to be the reason for lower specificity of RNAP in mycobacterium. In several cases, absence of -10 is compensated by the presence of an extended -10 motif with the consensus TNTGN immediately upstream of -10 hexamer. The initiating nucleotide at the transcription start point (+1) is G (52%) and A (27%).

Mycobacterial promoters are divided into 4 groups on basis of -10 and -35 regions.

**Group A:** Have putative -10 and -35 regions similar to \textit{E. coli} sigma70 consensus.

**Group B:** Have only putative -10 regions similar to \textit{E. coli} sigma70 consensus.

**Group C:** Have no consensus.

**Group D:** Include \textit{M. paratuberculosis} promoters with -10 and -35 consensus clearly different from \textit{E. coli} sigma70 promoters.

For Group A promoters, the distance between transcription start point and -10 ranges from 4-8 bp and between -10 and -35 is 16-21 bp.
1.11.2. Functionality in *E. coli*

Mycobacterial genomes have a high G+C content, for example, *M. tuberculosis* contains 65.9% G+C. Since the G+C content of a genome affects codon usage and promoter recognition, mycobacterial promoters function poorly in *E. coli* (Mulder *et al.*, 1997). However there are some exceptions like the mycobacterial heat shock promoters (Stover *et al.*, 1991). Among other mycobacterial promoters shown to be active in *E. coli* is the 16S rRNA promoter of *M. bovis*, 65 kDa antigens of *M. tuberculosis*, *M. bovis* BCG (Timm *et al.*, 1994), *M. leprae*. More examples include *M. tuberculosis* 38 kDa antigen, *M. paratuberculosis* pAN promoter, *M. fortuitum* blaF (Timm *et al.*, 1994), and the *M. tuberculosis* katG (Thole *et al.*, 1985). In all cases, expression in *E. coli* was less efficient than in the natural hosts.

1.11.3 Promoters in Fast and Slow Growers

The expression of genes in fast growers such as *M. smegmatis* using promoters from slow growers, e.g. *M. tuberculosis* have provided evidence that, transcriptional signals are generally conserved among mycobacteria. Bashyam *et al.* (1996), demonstrated that the efficiency and specificity of transcriptional recognition is conserved in *M. tuberculosis*, *M. smegmatis* and *M. bovis* BCG. The promoter clones examined in these three hosts exhibited similar activities and utilized the same transcription start sites. Although certain promoter sequences appear to be conserved among mycobacteria, there are likely to be differences in other aspects of the transcription machinery between the slow growers and the fast growers. Timm *et al.*, (1994) reported differences in the relative strengths of three mycobacterial promoters in *M. smegmatis* and *M. bovis* BCG. Thus the viability of studying *M. tuberculosis* promoters in other mycobacterial hosts may depend on the particular promoter to be examined.

1.11.4 Promoter Prediction

The driving force behind the recognition of any promoter region is the adjacent gene(s) to be transcribed since similar or 'stronger' promoter-like sequences may function like promoter in one case but not in the other cases. In any case, certain sequences have features that mask them as promoter sequences. Computational methods, as compared to traditional laboratory methods previously used in finding genes and protein binding sites, are being used frequently. The methods of gene
Prokaryotic promoter detection/prediction used by various programs cover statistical, analytical and training/learning methods such as artificial neural network, Markov models, hidden Markov models and Bayesian networks.

Prokaryotic promoter detection/prediction is less researched and few softwares are available for analysis. Due to the variability of the transcriptional machinery in prokaryotes as reflected on the availability of several known sigma factors, promoter prediction in prokaryotes has to be at least species specific for it to be very effective. Species specific programs would be more accurate if enough training datasets are available.

Prokaryotic promoter prediction methods used to date include Statistical Analysis, Hidden Markov Models (Pedersen et al., 1996), Artificial Neural Network (Pedersen & Engelbrecht 1995; Mahadevan & Ghosh 1994; Kalate et al. 2003), and Cluster Analysis (Ozoline et al., 1997). Most of the above approaches have had some degree of success with the task of promoter prediction, but they also predicted many sequences that were not known to have promoter activity according to existing data on *E. coli* promoters.

### 1.12. Variable Number Tandem Repeat (VNTRs) in Mycobacteria

The genomes of eukaryotes and prokaryotes are laiden with tandemly repeated DNA sequences occurring in several to thousands of copies which are dispersed throughout the genome. Some of these repeats show inter-individual length variability and are designated as variable number of tandem repeat (VNTR) loci. VNTR loci have been of interest because of hypervariability in repeat numbers for each locus in animal and human population and have emerged as useful markers to allow individual identification, in tracking genetic drift and speciation in population both in eukaryotes and prokaryotes (Jeffreys et al. 1985; Van Belkum et al. 1998). Contiguous repeats have been characterized in various prokaryotic genomes (Keim et al., 2000; Frothingham and Meeker O’conell, 1998; Supply et al., 2000; Van balkum et al., 1997). Microsatellites with short sequence repeats are more common and have been characterized in various prokaryotic genomes in genes encoding products as diverse as microbial surface components recognizing adhesive matrix molecules and specific virulence factors such as lipopolysaccharides modifying enzymes or adhesions (Van Belkum et al., 1998). Certain types of bacteria are known to use microsatellites containing genes called contingency genes which can confer survival advantages upon a subset of bacterial population under changing environmental conditions for e.g. in
Minisatellite structures composed of 40-100 bp repetitive units were first reported in *M. tuberculosis*. Also known as mycobacterial interspersed repetitive units (MIRUs), they were originally reported in 41 locations throughout the chromosome (Supply et al. 1997). Twelve of these displayed polymorphism in MIRU copy numbers among non related *M. tuberculosis* isolates (Supply et al. 2000). The number of polymorphic loci has since then increased and variations of the repeat units in the majority of the repeats (39 loci) have been reported in *M. tuberculosis* isolates (Smittipat et al., 2005). These VNTRs have been located both in the coding and non coding regions. The size polymorphism resulting from their insertion within coding regions has been suggested as source of antigenic variation for evading immune response to provide bacterial populations with evolutionary flexibility to adapt to unpredictable environmental changes in situations in which classical regulation of gene expression is not sufficient (Supply et al. 2000; Skuce et al. 2002).

### 1.12.1 Variable-number tandem repeat 3690 polymorphism in Indian clinical isolates of *M. tuberculosis*

The VNTR 3690 locus is located in the intergenic region between *rv3303c* and *rv3304*, in the 5'-flanking region of the *lpdA* (*rv3303c*) gene. LpdA is a flavoprotein disulfide reductase originally annotated as a probable lipoamide dehydrogenase in *M. tuberculosis* (Argyrou et al., 2004). Whilst the laboratory *M. tuberculosis* H37Rv strain TMC102 contain four tandem repeats, *M. bovis* BCG has been found to have only one repeat unit. The copy number of VNTR 3690 has been reported to vary extensively, ranging from one to eight copies in one study (Supply et al., 2006) and from one to 22 in another (Le Fleche et al., 2002). As such, the VNTR 3690 locus has been included as one of the polymorphic VNTRs utilized in the 15 and 24 subsets of VNTRs utilized in molecular epidemiological investigations of tuberculosis (Supply et al., 2006). The presence of one imperfect repeat (with two nucleotide difference) present in all isolates and included as one of the copies in molecular typing is noteworthy. Overall, VNTR 3690, also referred to as Mtub39 (Fabre et al., 2004) or allele no. 7, is 58 bp long and has been found to have an allelic diversity of 0.64 (Le Fleche et al., 2002). Among 65 clonally and epidemiologically related isolates and single-colony cultures, only one variant was identified.
Presence of VNTR in the intergenic regions has been widely reported in eukaryotes and shown to influence the gene expression however this role has not been addressed in *M. tuberculosis*. The present study was the first attempt to study the effect of VNTR polymorphism on downstream gene expression in *M. tuberculosis*. 