Chapter 1

GENERAL INTRODUCTION
1.1 **Introduction:**

Tuberculosis (TB) remains a public health threat and still a major neglected cause of death and disability worldwide. Globally, 22 high-burden countries account for more than 80% of the active tuberculosis cases in the world showing inequitable distribution of the disease. There are more than 9 million new cases every year worldwide, and the incidence rate is falling at less than 1% per year.

The etiologic agent of tuberculosis is *Mycobacterium tuberculosis* (Mtb), which can evade and manipulate the host immune response. Mtb was first described in 1882 by Robert Koch. The bacterium is also known as Koch's bacillus. Infection with Mtb follows a pattern of events that have been established through animal models, as well as observations from human TB. The infectious bacilli are inhaled as droplet nuclei that have been exhaled into the atmosphere. These droplets are small enough to remain airborne for several hours and cause disease.

1.2 **Mtb Complex:**

TB is caused by several species of gram-positive bacteria known as tubercle bacilli or *M. tuberculosis* complex. Mtb complex includes obligate human pathogens such as Mtb and *M. africanum*, as well as organisms adapted to various other species of mammal like *M. canetti* and *M. microti* (Comas et al., 2009)

1.3 **Reason behind success of Mtb as a pathogen:**

Mtb has a penetrance of its host population that would be the envy of most human pathogens and causes more deaths in humans than any other bacterial pathogen. Intracellular parasites, like Mtb, must attach to and enter the host cells, avoiding potentially hazardous host responses before they are internalized. After becoming intracellular, these pathogens face a new hostile environment in which host cellular defense mechanisms are sophisticated and effective. In response, these pathogens must be able to sense when they have entered a host cell and adapt accordingly. The capacity of Mtb to establish an infection within an individual and efficiently disseminate within the human population is mediated in part by its ability to survive within phagocytic cells, remain dormant over long periods of latent infection, and resume growth upon disease reactivation (Parrish et al., 1998). The physiological and environmental signals during periods of active disease, dormancy or
disease reactivation are likely to contribute to Mtb adaptation during various stages of infection. This ability to adapt should also be important for Mtb pathogenicity, because when this pathogen infects its host, it can enter many different cells of the immune system during disease progression, e.g. dendritic cells, macrophages, alveolar epithelial cells, neutrophils, etc. In addition, Mtb is found in lung granulomas that can change markedly during different stages of an infection and can also reside in other organs (Smith, I. 2003).

Mtb does not possess the classic bacterial virulence factors such as toxins, capsules and fimbriae. However, a number of structural and physiological properties of the bacterium are known to contribute to bacterial virulence. These include:

i) **Mtb can grow intracellularly.** This is an effective means of evading the immune system. Once Mtb is phagocytosed, it can inhibit phagosome-lysosome fusion. Mtb interferes with the toxic effects of reactive oxygen intermediates produced in the process of phagocytosis by two mechanisms:

(a) macrophage uptake via complement receptors may bypass the activation of a respiratory burst,

(b) compounds like glycolipids, sulfatides and LAM down regulate the oxidative cytotoxic effect.

ii) **Mtb has special mechanisms for cell entry:** The tubercle bacillus can bind directly to mannose receptors on macrophages via the cell wall-associated mannosylated glycolipid, LAM, or indirectly via certain complement receptors or Fc receptors.

iii) **Antigen 85 complex:** This complex is composed of a group of proteins secreted by Mtb that are known to bind fibronectin. Although direct evidence has been lacking, these proteins are predicted to aid in wailing off the bacteria from the immune system.

iv) **Slow growth:** As Mtb has longer generation time, the immune system may not readily recognize the bacteria or may not be triggered sufficiently to eliminate them. Many other chronic diseases are caused by bacteria with relatively long generation times, for example, slow-growing *M. leprae* causes leprosy, *Treponema pallidum* causes syphilis, and *Borrelia burgdorferi* causes Lyme disease.

v) **High lipid concentration in cell wall:** High lipid content of the cell wall accounts for impermeability and resistance to antimicrobial agents, resistance to killing by acidic and
alkaline compounds in both the intracellular and extracellular environment, and resistance to osmotic lysis via complement deposition and attack by lysozyme.

vi) **Cord factor:** The cord factor is primarily associated with virulent strains of Mtb. It is known to be (a) toxic to mammalian cells and (b) an inhibitor of polymorphonuclear leukocyte (PMN) migration. However, its exact role in Mtb virulence remains unclear.

Most important factor towards success of mycobacterium is persistence. It is the hallmark of tuberculosis. In most common clinical pattern of tuberculosis, T lymphocytes get activated after infection and this causes infiltration of activated macrophages to the site of infection. These lymphocytes arrest the organism inside a granulomatous lesion called a tubercle. Such immune responses of the host are capable of containing the infection. Indeed, most people infected with the bacillus show no signs of disease and only develop active disease when their immune system is perturbed, for example, after onset of AIDS, with aging or with malnutrition. This asymptomatic infection is called the latent disease.

After infection, the incubation period of tuberculosis ranges from a few weeks to a lifetime. The persistence of Mtb is because of its ability to evade the immune system of the host. During latency, the infected person does not have clinically apparent tuberculosis, but harbors the dormant bacilli, which restore to virulence when the immune system is weakened because of old age, poor nutrition or HIV. All conventional anti-tuberculosis drugs target processes involved in its growth and division. These drugs are only effective against actively growing bacteria but they are not potent against the latent bacteria. Only recently a series of compounds containing nitroimidazopyran nucleus, possessing bactericidal activity against both replicating and static Mtb have been reported (Stover *et al.*, 2000). It appears, thus, that an understanding of the mechanisms by which this organism persists might allow the design of drugs that would specifically target latent bacilli (Manabe and Bishai, 2000).

Adaptation to the environment is particularly important for Mtb which is transmitted by aerosol route with the lung being the primary organ affected. Pulmonary tuberculosis, the most common form of tuberculosis, is a highly contagious and life-threatening infection. Once Mtb reaches the alveoli it is engulfed by professional phagocytes such as macrophages. Initially, Mtb is able to replicate within macrophages until a cell-mediated immunity is mounted by the host. Then, macrophages are activated by interferon-γ (IFN-γ) and, are able to control the intracellular growth of Mtb by triggering a hostile environment that includes acidification of the phagosome, lysosome maturation and production of NO and other
reactive oxygen/nitrogen intermediates. In those hosts that cannot control the bacterial infection, Mtb can overcome the hostile conditions and is usually able to replicate in the macrophages. To survive in this cell type, Mtb has developed strategies to arrest phagosome maturation at an early stage, maintaining a relative non-acidic pH and avoiding fusion with lysosomes (Russell, 2001). An immuno-competent individual infected with Mtb is usually able to develop a strong immune response whereby numerous peripheral blood monocytes and T lymphocytes migrate to the lung and contain the focus of infection by forming a granuloma. In these stages of containment the bacteria may remain for several months or years usually until the person becomes immuno-compromised for some reason (e.g. immunosuppressive drugs, AIDS, aging, etc.) which marks the reactivation of the disease. This state of dormancy and persistence is one of most intriguing phenomena of Mtb lung infection. In spite of global research efforts, mechanisms underlying pathogenesis, virulence and persistence of Mtb infection remain poorly understood.

1.4 Present cures for TB and their limitations:

Although tuberculosis has been the focus of medical research for more than 100 years, M. bovis Bacillus Calmette-Guérin (BCG) is the only approved vaccine against TB. It was developed though the serial in vitro passage of M. bovis until it became nonpathogenic. It is used in countries with endemic TB because it protects children against severe forms of disease, such as TB meningitis or disseminated infection. However, although effective against development of TB in some countries such as the United Kingdom, its efficacy has been questioned in several studies, most notably in India, where very limited (or no) protection has been reported. While it has clear beneficial effects against tuberculosis in childhood (Colditz et al., 1995), it only provides protection against the disease for a limited number of years (Andersen and Doherty, 2005) in highly tuberculosis endemic regions. Therefore, a new vaccine is urgently needed (Russell et al., 2010).

Tuberculosis can be cured in most cases by a cheap course of antibiotic treatment. However, the difficulty of a timely diagnosis, socioeconomic factors in tuberculosis endemic areas and the fact that bacterial clearance requires many months of treatment combines to prevent successful global tuberculosis control by antibiotics. Furthermore, the key front-line antituberculosis drugs are effective in treating individuals with acute primary or reactivated tuberculosis infections, these drugs are ineffective in eliminating Mtb during the persistent stages of latent infection (Wayne and Hayes, 1996). Rifampicin, discovered 40 years ago,
Chapter 1: General Introduction

represents the last novel class of antibiotics introduced for the first-line treatment of tuberculosis. Drugs in this class are part of a 6-month, regimen that is ineffective against multi drug-resistant (MDR) and extreme drug-resistant (XDR) tuberculosis, and are difficult to use with many antiretroviral drugs. Although, ten compounds are in the pipeline of clinical development for the treatment of tuberculosis, additional sustainable efforts are needed to tackle global burden of tuberculosis (Zhenkun et al., 2010).

The history of anti-tuberculosis drug resistance is fairly recent. With the discovery of rifampicin in 1966, and the expansion of its use between 1970 and 1990, patients who were already carriers of isoniazid resistant Mtb strains became resistant to rifampicin. This was the start of a progressively growing problem, multidrug-resistant tuberculosis (MDR-TB, defined as resistance to at least both isoniazid and rifampicin). Curing TB in these patients is difficult, as they carry strains resistant to the two most efficient anti-tuberculosis drugs. With the misuse of other drugs with anti-tuberculosis action, in particular the fluoroquinolones, the most effective among the second-line drugs, resistance has broadened to extensively drug-resistant TB (XDR-TB). XDR-TB can be defined as MDR-TB plus resistance to any fluoroquinolone and to at least one of the second-line injectables. XDR TB is still relatively rare (an estimated 5% of cases) and gradual progression of this form towards the current epidemic began only 15 years ago, in the second half of the 1990s. XDR-TB did not spread in a uniform fashion around the world. Patients with XDR-TB are left with treatment options that are much less effective and often have worse outcomes. Thus, it is not uncommon that people with XDR-TB die even after entering treatment (Jassal and Bishai, 2009).

1.5 HIV and TB:

HIV-infected individuals are highly susceptible for TB, i.e., 40–80 time more likely, to be infected with or reactive latent TB. The HIV virus replicates vigorously in activated CD4+ T cells and in dendritic cells (DCs), first in the genital and gut mucosa, respectively, and then subsequently throughout the lymphatic tissue leading to systemic and harmful immune activation. TB/HIV co-infection has become a major global health problem and about 25% of TB deaths occur among HIV patients, also making TB the leading killer in HIV-positive patients (Brighenti et al., 2010).
1.6 Two-component regulatory systems (TCS):

TCSs are ubiquitous bacterial regulatory elements commonly associated with signal recognition and adaptive responses. They form part of complex cellular networks which regulate metabolism, environmental adaptation, and phenotypic changes in biological systems. Among the elements forming regulatory networks in bacteria are regulatory proteins such as transcription factors, which respond to exogenous and endogenous conditions. To date, researchers have found hundreds of such systems in eubacteria, archaea, and a few eukaryotic organisms. TCS serve as a basic stimulus-response coupling mechanism to allow organisms to sense and respond to changes in many different environmental conditions. These are sophisticated signaling systems marked by a highly modular design that has been adapted and integrated into a wide variety of cellular signaling circuits. The prototypical system consists of a histidine protein kinase (HK), containing a conserved kinase core, and a response regulator protein (RR), containing a conserved regulatory domain. Extracellular stimuli are sensed by, and served to modulate the activities of, the HK. The HK transfers a phosphoryl group to the RR, in a reaction catalyzed by the RR. Phosphotransfer to the RR results in activation of a downstream effector domain that elicits the specific response. Biochemical activities and three dimensional structures are known for representatives of all conserved core domains of HKs and RRs, as well as for several variable input domains of HKs and for most of the subfamilies of output domains of RRs (for a review, see Gao and Stock, 2009).

Most HKs proteins are located in the cytoplasmic membrane with their transmitters projecting into the cell. Their periplasmic domains are structurally unrelated and have diverse receptor functions. For example, EnvZ senses osmolarity changes and UhpB detects periplasmic hexose phosphates. In contrast, response regulators can occur alone (R), in tandem (RR) or in combination with different output domains (RO), most of which have regulatory functions and are grouped into subfamilies on the basis of sequence similarity. Proteins in the RO₁ group have unrelated output domains with disparate functions. For example, Spo0A regulates gene expression leading to sporulation, whereas CheB is a methylesterase that covalently modifies chemo-receptors during sensory adaptation to chemotactic stimuli. Output domains of the RO₂ subfamily appear to have DNA-binding roles. Some of these RRs (OmpR, PhoB and VirG) bind to specific target sequences upstream of the promoters they regulate. The RO₃ group is still poorly understood but may involve a DNA-binding function as well, since the output domain of FixJ resembles the promoter
recognition region of sigma factors. The RO1 subfamily contains two linked output domains. The C-terminal domain binds at enhancer sequences in the vicinity of the regulated target genes, whereas the central domain interacts with the $\sigma^{54}$ from of RNA polymerase holoenzyme to promote formation of open-complex capable of transcription initiation.

1.7 **Phosphotransfer Chemistry:**

The chemistry of the basic two-component phosphoryl transfer signal transduction pathway involves three phosphotransfer reactions and two phosphoprotein intermediates: the $\gamma$-phosphoryl group of ATP is first transferred to a conserved His-side chain of the HK. The RR then catalyzes the transfer of this phosphoryl group from the phosphor-His residue to a conserved Asp side chain within its own regulatory domain. Finally, the phosphoryl group is transferred from the phosphor-Asp residue to water in a hydrolysis reaction. All three reactions require divalent metal ions, with $\text{Mg}^{2+}$ presumably being the most relevant cation in vivo (Stock *et al.*, 2000).

1.8 **Importance of the TCSs in Mtb:**

The in vivo environment encountered by the bacilli is the granulomatous lesion, which presents a nutritionally deprived and an anoxic environment. Mtb is one of the oldest human pathogens and has evolved strategies for survival. Despite the fact that it stimulates a strong immune response by the host, Mtb has evolved to resist the body's attempts to eradicate it. To be so well adapted to the numerous environmental conditions that the host offers, Mtb must be able to mount an array of functions, like inducing metabolic pathways to utilize the carbon source available inside the macrophage, scavenging oxygen radicals to avoid cellular damage, and acquiring iron, to name a few. Thus, to establish a successful infection, bacteria have to constantly sense the medium, and efficiently signal the changes that enable a quick adjustment to new conditions. This series of events occurs coordinately by means of numerous mechanisms of signal transduction and transcriptional regulation. Signal transduction in bacteria mainly involves TCSs. However, mycobacteria also have several serine/threonine kinases and tyrosine phosphatases that were originally thought to participate only in eukaryotic signal transduction (Fontan *et al.*, 2004). After the signals indicating the environmental changes are sensed and transduced, bacteria respond by synthesizing new proteins and down regulating others. This balanced regulation of gene expression in bacteria occurs primarily at the level of transcription.
The TCSs form a large family of proteins involved in signal transduction that allows bacteria to detect and respond to many different kinds of stimuli. In its most common form, TCS consists of two proteins, a sensor and a receiver, that are involved in a phosphotransfer reaction. After interaction with the appropriate stimulating ligand, the membrane bound sensor protein, called HK, catalyze autophosphorylation of a conserved histidine residue. The phosphate is then transferred to the cognate cytoplasmic receiver protein, known as RR. Once the phosphotransfer occurs, the response regulator is activated, allowing it then to carry out its specific function. In most cases, the response is the modulation of transcription through the response regulator interacting with DNA at specific binding sites located in target promoter regions. The total effect is change in global gene expression that aids the organism in responding to the initial signal sensed by the HK.

In agreement with this view, a number of TCSs have been shown to play critical role in latency and other stages of Mtb life cycle (Zahrt, 2003). The genome sequence of Mtb contains 11 complete two-component systems (Cole et al., 1998). The MtrA-MtrB system was the first TCS to be characterized in Mtb (Via et al., 1996). While mutation in a subset of RRs or HKs were shown to attenuate in vitro growth of Mtb in macrophages compared to the wild type bacillus, mutations in other systems were shown to enhance the in vivo growth of the pathogen in immunocompromised SCID mice (Haydel and Clark-Curtiss, 2004; Parish et al., 2003, Perez et al., 2001, Zahrt and Deretic, 2001) suggesting that the Mtb virulence is also regulated by a subset of HK and RR proteins that are normally repressed. For example, the transcripts from the co-transcribed devR-devS genes are expressed at lower levels in the avirulent H37Ra strain compared with the virulent H37Rv strain of Mtb (Dasgupta et al., 2000). The cDNAs corresponding to co-transcripts from the prrA-prrB locus have been recovered from Mtb cultured in human blood monocye-derived macrophages but not from bacteria cultured in standard laboratory medium, indicating that the genes are transcribed in response to host–cell interaction (Graham and Clark-Curtiss, 1999). In addition, TrcRS and KdpDE mutants are hypervirulent (Parish et al., 2003), suggesting that these systems play important role(s) during infection. In an attempt to identify which TCSs are induced during macrophage infection and thus possibly required for virulence, Deretic and coworkers characterized the expression of Mtb RRs in 7H9 medium (in vitro) and in macrophages (ex vivo). Using green fluorescent protein (GFP) expression vectors fused with the promoters of selected Mtb RRs, constructs placed in M. bovis BCG and Mtb were screened for GFP expression during infection of macrophages (Zahrt and Deretic, 2001). The results show that
phoP and Rv0818 constitutively express in vitro and ex vivo in M. bovis and Mtb. The RR, mtrA, was induced in both M. bovis and Mtb in macrophages, while the RRs Rv0981 and Rv3143 were induced in macrophages in M. bovis but not in Mtb. Several RRs did not show induction of expression in macrophage. As mtrA was the only RR induced during macrophage infection in this study, an attempt was made to delete the mtrA-mtrB system. The effort was unsuccessful suggesting that this system is essential for growth (Zahrt and Deretic, 2000). Although this was a preliminary analysis to understand RR’s role in Mtb pathogenesis, a lack of gene expression in macrophages does not rule out their involvement in virulence regulation since there are many aspects to Mtb infection and disease progression. Also, constitutively active RRs may still be very important for Mtb survival and growth in macrophages, as seems to be the case with the Mtb phoP/phoR system.

1.9 Role of PhoP in Mtb virulence:

Among eleven TCSs of the Mtb genome, the phoP/phoR locus is annotated as being involved in phosphate metabolism, because of the observed similarity with phoP/phoR system in B. subtilis (Cole et al., 1998). The Mtb phoP/phoR also shows a high degree of similarity to other TCSs including Salmonella phoP/phoQ, a well studied TCS that controls the transcription of several key virulence genes. Furthermore, phoP mutant of Salmonella is unable to synthesize many proteins normally expressed upon interaction with macrophages (Buchmeier and Heffron, 1990). The pleiotropic PhoPQ has also been shown to control, directly or indirectly many genes not involved in virulence (Groisman, 2001).

The first evidence in support of a role for PhoP in virulence regulation of Mtb was obtained when Mtb strain with a deletion in the phoP/phoR TCS was unable to grow in murine bone marrow macrophages or in mice (Perez et al., 2001). Thus inactivation of phoP results in high attenuation of Mtb; however, it is not completely eliminated and persists in in vitro cultured-macrophages and also in mouse organs. The mutant cells also have an altered, rounded shape and they show difference in levels of lipoarabinomannan derivative compared to the wild type (Ludwiczak et al., 2002). In an attempt to identify genes requiring PhoP for expression, global gene expression profiling indicates that 44 genes are up-regulated and another 70 genes are down-regulated by PhoP (Walters et al., 2006). Many of the up-regulated genes are involved in general or lipid metabolism and in substrate transport across the plasma membrane. However, mechanism of regulation of these target promoters have not been demonstrated yet and it is not known which, if any of them are direct targets of PhoP.
The observation that a *phoP* knockout mutant Mtb strain lacks sulfatides, diacyltrehaloses and polyacyltrehaloses in the cell envelope, suggest that the *phoP/phoR* system is involved in complex lipid biosynthesis (Gonzalo Asensio *et al.*, 2006; Walters *et al.*, 2006). This also provides an explanation for the altered morphology of the *phoP* mutant compared to wild-type Mtb H37Rv.

The importance of PhoP as a regulator of virulence was further underscored by the fact that live Mtb *phoP* mutant strain SO2 has been shown to be more attenuated than BCG and confers protective immunity against tuberculosis in mice and guinea pig model systems (Martin *et al.*, 2006). Complementation of Mtb SO2 with the wild-type *phoP* gene restored the virulence of the strain in the SCID mice, confirming that the attenuated phenotype is due to the *phoP* mutation (Martin *et al.*, 2006). Further supporting the role of PhoP in virulence regulation, more recent studies have demonstrated that a point mutation in PhoP contributes to avirulence of the H37Ra strain, since this mutation abrogates secretion of the ESAT-6 antigen and the synthesis of acyltrehalose-based lipids implicated to be involved in virulence (Frigui *et al.*, 2008; Lee *et al.*, 2008). Lee *et. al.* further showed that comparison of attenuated strain H37Ra of Mtb with its virulent counterpart H37Rv revealed three unique coding region polymorphisms. One of the point mutation was located in the DNA-binding domain of the transcriptional regulator PhoP, replacing a serine at position 219 by leucine causing the protein’s reduced DNA-binding capacity. The authors argued that the mutation in *phoP* gene as the most critical to account for the phenotypic differences. It is noteworthy that complementation of H37Ra with the *phoP* gene of H37Rv caused reversion of the colony morphology resembling that of H37Rv (Lee *et al.*, 2008).

In a recent study it was shown that members of the *phoP* regulon are differentially expressed between intracellular Mtb H37Ra and H37Rv (Li *et al.*, 2010) According to the study, twelve members of the *phoP* regulon (*pk53, pk54, mmpL8, mmpL10, papA3, fadD21, PPE18, PPE19, Rv1639c, Rv2376c, PE31, and PPE60*) were under-expressed in intracellular H37Ra compared to H37Rv. These observations are in agreement with previous studies by Issar Smith and coworkers in which transcriptomes of broth-grown H37Rv and a *phoP* knockout mutant of Mtb was compared (Walters *et al.*, 2006). Results of these studies also suggest a strong connection between the ESX-1 secretion system and members of the *phoP* regulon (Walters *et al.*, 2006).
1.10 **Scope and objectives of the present study:**

The PhoPR TCS has attracted attention in the past few years. A number of investigations show that inactivation of \textit{phoP} in Mtb H37Rv leads to significant growth attenuation (Perez \textit{et al.}, 2001; Gonzalo-Asensio \textit{et al.}, 2006; Walters \textit{et al.}, 2006). Also, biochemical studies reveal that PhoP regulates sulphatides, diacyltrehaloses and polyacyltrehaloses and absence of these lipid molecules in the \textit{phoP} mutant is the major reason for its attenuated growth in a mouse model (Gonzalo Asensio \textit{et al.}, 2006; Ludwiczak \textit{et al.}, 2002) (for review, see Ryndak \textit{et al.}, 2008). While two independent studies show that a point mutation in \textit{phoP} contributes to avirulence of Mtb H37Ra (Chesne-Seck \textit{et al.}, 2008, Lee \textit{et al.}, 2008), more recently PhoP has been implicated in the ESAT-6 secretion and specific T-cell recognition during virulence regulation of the bacilli (Frigui \textit{et al.}, 2008). Thus, accumulating evidences suggest that PhoP is a key regulator of Mtb. However, molecular mechanism of how it functions remains largely unknown.

Many members of the PhoP subfamily use different mechanisms to regulate their DNA binding and transcription regulation despite having high sequence similarity. In this study we set out to investigate mechanism of action of PhoP as a regulator of more than 100 genes of Mtb. Chapter 2 comprises of detailed study of DNA protein interaction of PhoP. Although, PhoP has been shown previously to bind to its own promoter, we identified a direct repeat sequence as the primary target site for sequence-specific DNA binding by PhoP (Gupta \textit{et al.}, 2006). Here, we showed that two PhoP protomers are recruited on it’s target DNA comprising a 9-bp direct repeat motif. We also show (i) that DNA binding stimulates the dimerization of PhoP, and (ii) the two molecules are structurally organized in a specific head-to-head orientation.

The crystal structure of PhoPC clearly shows that the primary DNA binding of the protein involves winged helix-turn-helix motif (PDB ID: 2PMU) and the surface around the PhoP residues comprising the recognition helix (residues Asn212-Tyr224 of \(\alpha 8\)) display strong positive electrostatic potential, indicating that these residues are likely to be critical in DNA binding and nucleotide sequence recognition (Wang \textit{et al.}, 2007). To this end, we used structure-guided mutagenesis to obtain single alanine substitutions of 10 solvent-exposed residues spanning \(\alpha 8\). Our results of rational mutagenesis coupled with DNA-binding affinity study of the \(\alpha 8\)-DNA interface in the complex formed by PhoP and its cognate DNA demonstrate that most PhoP mutants have significantly reduced DNA-binding affinity while
possessing near wild-type stability. However, alanine substitution of Glu215 of α8 shows major effect on the specificity of DNA recognition. Using structural insights coupled with biochemical analyses, we identify that Glu215 of PhoP appears to establish a base-specific interaction with (G/C)⁹ of the upstream repeat motif (DR1 of DR1,2) to contribute significantly to the recognition specificity of the regulator. Biochemical experiments corroborate these results showing that DNA recognition specificity can be altered by as little as a single residue change of the protein or a single base change of the DNA.

Another objective was to investigate domain structure of Mtb PhoP and how does it contribute to PhoP’s function. The several functions of PhoP are apportioned between a C-terminal effector domain (PhoPC) and an N-terminal receiver domain (PhoPN), phosphorylation of which regulates activation of the effector domain. In the 3rd chapter we show that PhoPN, on its own, demonstrates PhoR-dependent phosphorylation. PhoPC, the truncated variant bearing the DNA binding domain, binds in vitro to the target site with affinity similar to that of the full-length protein. To complement the finding that residues spanning Met1 to Arg138 of PhoP constitute the minimal functional PhoPN, we determined Arg150 as the first residue of the distal PhoPC domain capable of DNA binding on its own, thereby identifying an inter-domain linker. We further show that coupling of two functional domains together in a single polypeptide chain is essential for phosphorylation-coupled DNA binding by PhoP.

Chapter 4 originates from the interesting domain structure that was detailed in chapter 3. To better understand inter-domain interaction(s) in effector domain regulation, we sought to investigate domain structure of PhoP. To this end, we identify an 11-residue long inter-domain linker that tethers two functionally-independent domains of PhoP together and regulates inter-domain interactions. While the newly-identified linker region is not required for either domain functions of PhoP, most strikingly, it plays an essential role for phosphorylation-dependent DNA binding to msl3 promoter, previously suggested to be regulated by PhoP (Walters et al., 2006). Interestingly, biochemical studies reveal that one of the major differences between OmpR and PhoB reside in the inter-domain linker region that tethers together the N-terminal domain with the C-terminal domain (Walters et al., 2003). Consistent with this view, a previous study had shown that C-terminal DNA binding by OmpR could influence phosphorylation of the N terminus in which the linker region underwent a conformational change (Ames et al., 1999), thus suggesting a key role of the linker region in regulation of inter-domain interaction(s). Together, our results suggest that
although the DNA binding energy and specificity of regulator-promoter interactions is
contributed primarily (but not entirely) by the C-domain, linker region of the protein likely
allows the regulator to adopt a different phosphorylation-dependent conformation enabling it
to discriminate target promoters while it regulates a vast array of genes to either activate or
repress transcription.