Chapter 5: Summary and Conclusions
Approximately eight million people develop active tuberculosis (TB) every year, with two million dying from the disease. In addition to this already huge burden of disease, it is estimated that up to two billion people have been infected with the causative agent, *Mycobacterium tuberculosis* (*Mtb*) (Dye *et al*., 1999). Most people control the initial infection by mounting a cell-mediated immune response that prevents disease but can leave a residual population of viable mycobacteria. Between 5–10% of individuals who become infected subsequently develop clinical disease (Bloom and Murray, 1992). Primary TB develops within 1 or 2 years after an initial infection and, particularly in children, is often associated with disseminated disease. Post-primary TB develops later in life, and can be caused either by reactivation of bacteria remaining from the initial infection or by failure to control a subsequent reinfection. Post-primary TB is predominantly a pulmonary disease, involving extensive damage to the lungs and efficient aerosol transmission of bacteria. The risk of disease is highly dependent on the immune status of the host; co-infection with HIV markedly increases the incidence of both forms of disease. The latency attribute of the pathogen represents a significant obstacle to the worldwide control and eradication of tuberculosis because the non-replicating bacilli may be in a state of ‘drug indifference’ wherein they are not killed by the drugs. In contrast to individuals with active tuberculosis, individuals with latent tuberculosis do not transmit the disease. The harsh environment faced by the bacillus
inside the human macrophages and dendritic cells include depletion of nutrients, shift in pH, production of growth limiting products and/or depletion of oxygen. On the other hand these extreme conditions give some how signal to the *Mtb* cells to slow-down this metabolism and go into dormancy/latency. Maintenance of dormancy within the host cells and granuloma for many years or decades is the outcome of a balanced war between the host immune response and the pathogen’s tactics to overcome it (Wayne and Sohaskey, 2001).

Transcriptional regulation in response to environmental changes encountered during infection is a common theme in bacterial pathogenesis, and similar concepts can be applied to dormant/latent bacteria. *Mtb* cells have to reprogram their transcriptional expression profiles to survive within the harsh environment presented by the macrophages. cAMP receptor protein/fumarate nitrate reductase (CRP/FNR) regulators is a family of eubacterial transcriptional factors associated with a variety of stress responses like hypoxia, nutrition depletion and redox regulation within the bacteria (Korner *et al.*, 2003). *Mtb* ORF *Rv3676* encodes cAMP receptor protein and knock out mutants have been reported to be defective for growth in animal models and macrophages (Rickmann *et al.*, 2005).

The present thesis is an attempt to study the properties of cAMP receptor protein from *Mtb* and also regulation by CRPs of other
mycobacterial species with an eventual goal to use this as drug candidate for intervention against latent and active TB.

This work started with the computer based bioinformatics approach aimed to identify the operons which could be regulated by CRP in mycobacteria. In earlier computational predictions of CRP-regulon, 73 binding elements in \textit{Mtb} genome were observed (Bai \textit{et al.}, 2005). Of the top 44 binding sites identified in the present study, 25 have been reported previously (Bai \textit{et al.}, 2005). These analyses thus highlighted 19 new \textit{Mtb}-CRP binding sites upstream of various operons in \textit{Mtb} genome. The CRP-binding element from \textit{fumarate reductase} (Rv1552) received the highest score. While this DNA element was also reported as a potential CRP-binding site in earlier reports (Bai \textit{et al.}, 2005; Rickman \textit{et al.}, 2005, Spreadbury \textit{et al.}, 2005), in the work presented here, experimental evidence for specific DNA: protein interaction between this DNA element and recombinant purified \textit{Mtb}-CRP was provided. Also, Spreadbury \textit{et al.} (2005) proposed some potential genes as members of the CRP-regulon. There are several overlaps between the results from this work and others. While previous studies (Bai \textit{et al.}, 2005, Spreadbury \textit{et al.}, 2005) utilized information from \textit{E. coli} CRP-regulon, in the present study only the available information from \textit{Mtb} CRP-regulon was used. Recently, a comprehensive comparative account of all \textit{Mtb}-CRP regulons has also been published (Krawczyk \textit{et al.}, 2009). Out of these novel identified
members predicted in the present analysis of CRP regulon in Mtb genome many are critical for pathogenesis and general life cycle of the bacterium. It includes genes related to cell wall biogenesis, 5’-3’ Cyclic Adenosine Monophosphate (cAMP) signaling, aminoacid biosynthesis pathways and recycling machinery of the cell.

CRP regulators from different species of mycobacteria have very similar DNA binding domains when compared with Mtb-CRP in terms of amino acid sequences (for M. avium, M. leprae and M. smegmatis are 96%, 96% and 97% identical respectively). Given the observation that the CRP proteins from all mycobacteria have identical DNA binding domains, the same profile matrix constructed for Mtb was extended to predict CRP binding sites in the genomes of M. leprae, M. avium subsp. paratuberculosis and M. smegmatis. This represented the first such attempt to interrogate other mycobacterial genomes. The operon context of these regulons across the genomes was also described.

A comparative analysis of CRP target genes in various species enabled the identification of the common CRP regulated genes across mycobacteria and at least 18 genes were found to be common. Conservation of these genes in the predicted CRP regulons suggests an important role of their cognate gene products in the mycobacterial life cycle. Mtb-CRP was earlier reported to be essential for the survival of mycobacteria inside macrophages and in animal models (Rickman et al., 2005). Further, in these analyses, a high conservation of these CRP
regulated genes among pathogenic mycobacteria than in non-pathogenic mycobacteria was found. This strengthened the notion that \textit{Mtb}-CRP and its regulated genes are important for pathogenesis of mycobacteria and that these might have co-evolved with the pathogenic branch as a result of genome optimization.

In the next part of study \textit{Mtb}-CRP was cloned, expressed and the recombinant protein was purified from \textit{E.coli}. It was characterized in terms biophysical and biochemical properties. Analytical size exclusion chromatography was carried out to determine the apparent oligomeric nature, if any, of the purified \textit{Mtb}-CRP protein. \textit{Mtb}-CRP was found to exist as a dimer of ~53 kDa (apparent molecular weight).

In most of oxygen tension-sensing proteins belonging to the CRP/FNR family of proteins, transition metals like Fe or Ni are associated with the protein to sense the fluctuations of oxygen availability \textit{via} redox mechanisms (Korner \textit{et al.}, 2003). The absorption spectrum of purified \textit{Mtb}-CRP was scanned to check for the presence of metal ion cofactor. Resulting spectra revealed only two characteristic peaks of proteins, one at 295 nm and the other at 280 nm. No bound associated metal cofactor was found. This suggested that \textit{Mtb}-CRP apparently uses some other mechanism(s) to sense effector signals.

Further, the cAMP binding properties of purified recombinant \textit{Mtb}-CRP were determined. Results of protein family search (pfam) revealed the presence of a putative cAMP-binding domain at the N-terminal end of
*Mtb*-CRP protein, thereby raising a strong probability that cAMP may be acting as an effector of *Mtb*-CRP. Purified *Mtb*-CRP was subjected to CD analysis in the presence and absence of cAMP as ligand. A comparison of CD spectra of these two forms provided evidence of binding as readout of change in secondary structure. The two tryptophan residues (Trp112 and Trp203) present in *Mtb*-CRP were used as probe to study the effect of cAMP concentration on conformational changes. This change in secondary structure clearly appeared to be a function of increasing concentration of cAMP. That cAMP indeed caused concentration dependent conformational alterations within *Mtb*-CRP was evident from tryptophan fluorescence spectrometry data. Physiological cAMP levels are in the range of 0–10 mM. At lower concentrations (6–10 mM), the binding showed positive cooperativity, and at 10 mM cAMP the protein existed in the most open conformation. With further increase of cAMP (12–16 mM), the protein was compacted which could be a reflection of a feedback regulation. To investigate the ability of *Mtb*-CRP to bind to its cognate DNA motif, EMSA was carried out using purified *Mtb*-CRP and a radiolabeled oligonucleotide carrying the CRP/FNR-binding site present upstream of the *frd* (*Rv1552*) gene encoding the fumarate reductase enzyme. This binding site was identified as a putative binding site in previous reports (Bai *et al.*, 2005; Spreadbury *et al.*, 2005). In the present work, during *in silico* regulon prediction studies, this motif elicited the highest score, and therefore it was selected for EMSA. It has been
reported that *Mtb*-CRP senses oxygen (Bai *et al*., 2005; Spreadbury *et al*., 2005) indirectly by controlling the expression of genes such as *frd*. Fumarate serves as an alternative electron acceptor in the absence of oxygen, and this is mediated by a membrane-linked fumarate reductase enzyme complex (Lambden and Guest, 1976). The putative CRP/FNR binding site, present upstream of the *frd* operon, was recognized by purified *Mtb*-CRP protein and this was evident from EMSA. The predicted cAMP-binding site in *Mtb*-CRP, indeed showed binding to cAMP leading to conformational changes in the protein as evident from spectral analyses. The extent of change in secondary structure was maximal in the presence of 10 mM cAMP.

To investigate the possible mechanistic role of *Mtb*-CRP in transcription, X-ray crystal structure of *Mtb*-CRP-cAMP-DNA ternary complex was determined. The structure of *Mtb*-CRP-DNA-cAMP was solved in the P2₁ spacegroup at a 2.9Å resolution. Using the CRP/FNR family protein structure (pdb entry: 2GAU) from *Porphyromonas gingivalis* as MR search model, it was possible to build complete protein structure as well as 23 basepairs of DNA and three cAMP molecules in the complex. In both of the subunits of the *Mtb*-CRP-DNA structure, it was observed that the very C-terminus residues forms an alpha-helix. It was investigated if this terminal could play a functional role in the ability of CRP to form complex. Results of deletion by side directed mutagenesis indicated that the helix-G was playing an active role in the folding of the
protein. Residues from the helix-G which could function as a flap to cover highly hydrophobic core formed at the interface of these three helices (E, F and G). This hydrophobic core consisted of Trp203, Ile204, Phe161, Val171, and Phe198 and Val150. These all hydrophobic side chains were facing the centre of the core. There were three residues (namely Leu219, Ala220 and Ala223) from helix-G which were facing this hydrophobic core, also hydrophobic in nature, and interacting with the core side chains via non-bonding interactions. Interestingly, neither the residues corresponding to helix-G were present in homologous proteins nor the hydrophobic core. These results indicated that this helix could be a unique feature of mycobacterial CRPs and related proteins.

In the structure there was one novel non-canonical cAMP binding site. The finding of a third cAMP molecule was unexpected, although secondary cAMP binding sites in EC-CRP, reported earlier (Passner 1997) was different from the one presented in this work. In EC-CRP the site is surface exposed and located near the DNA protein interaction area and interacting with Arg180 (located on DNA interacting helix-F), Glu58 (in nearby beta-sheet) and Gly173 and Gly177 (both on helix-E) (Passner 1997; pdb entry: 2CGP). In Mtb-CRP the secondary pocket was buried and located at the interface of three helices: helix-D (dimer interface helix), helix-E and helix-F (constituent of DNA interacting helix-turn-helix). This cAMP could directly interact with six of the side chains surrounding the pocket (Asn67, Leu 68, Asn137, Asp140, Glu156 and
Leu156) via several H-bonds and salt bridges. Least square superimposition analysis with *Mtb*-CRP-cAMP structure (PDB code: 3i54) did not reveal much difference with *Mtb*-CRP-DNA structure (RMSD=1.16 Å). This pocket environment was found to be conserved including the side-chains interacting with cAMP. For unknown reasons in *Mtb*-CRP-cAMP complex structure third cAMP was not bound. In case of EC-CRP it was known that interaction with DNA could give rise to secondary cAMP binding (Lin *et al.*, 2002, Scott and Jarjous, 2005). This might be similar for *Mtb*-CRP. This position seems interesting in terms of regulatory function for DNA interactions.

For this non-canonical pocket cAMP molecule was predicted to be present in only subunit B of *Mtb*-CRP with 50% occupancy. This indicated that this is seemingly a secondary binding site and not occupied fully in all molecules. On the other hand, some residual density on the similar NCS related position in sub-unit A was also seen although it was not sufficient to clearly position the cAMP. Taken together, this gave indication of concentration dependent sequential multi-step binding events for each sub-unit. Second non-canonical pocket has lower affinity than canonical one as it only had 50% occupancies. This observation was consistent with biophysical and biochemical studies for EC-CRP (Heyduk and Lee, 1989; Lee *et al.*, 1999; Takahashi *et al.*, 1980). In a surface-potential analysis a cleft for entry of cAMP to the pocket could be seen. It was also observed that the surface is highly positively charged
near this cleft. This positive charge could facilitate the entry of cAMP to the non canonical pocket.

Upon comparison analysis with the EC-CRP-DNA structure (pdb: 1O3S) different position of the helix-D (DBD) and helix-F (DNA interacting) of more than 3 Å was observed. This could be due to the presence of secondary binding pocket. This movement in helix-D translated as movement in helix-F which finally gives rise to more interaction with DNA. On looking in minute details it was observed that this helix rearrangement could be mediated by H-bonding between phosphate of cAMP and side-chain amino groups from Asn137 and Gln156. When the structure of *Mtb*-CRP-DNA was superimposed onto EC-CRP (PDB entry: 1O3S) it was found that the side chains of Met 59 and Lys129 occupied the binding pocket of CRP in *E. coli*, whereas the corresponding residues, Asn67 and Asn137 have lighter side chains in *Mtb*-CRP and were positioned to allow cAMP binding or specifically interact with the bound cAMP molecule. This indicated that the third binding pocket was specific to cAMP, but not conserved among all bacteria. Similar comparative analyses for the position of DNA interacting helix-F in *Mtb*-CRP-cAMP (PDB code: 3i54) were also performed but not any big movement in position could be seen. This might suggest that *Mtb*-CRP-cAMP form reported earlier in which only one cAMP is occupied in each sub-unit could be also the active form of *Mtb*-CRP ready to bind DNA and could start activation of transcription (Reddy *et al.*, 2009). In
Chapter 5 Summary and Conclusions

This case second cAMP binding could further regulate the transcription in a feed forward type of mechanism.

To investigate further, the importance of this secondary pocket, \textit{Mtb}-CRP mutants which could mimic the non-pocket environment found in EC-CRP were designed. The generated mutants were expected not to bind cAMP as the larger residues were expected to sterically hinder the binding of cAMP as seen in EC-CRP. The mutant allowed investigating the effect of the secondary binding pocket on the ability of \textit{Mtb}-CRP to bind the DNA. Four sets of experiments were performed: 1) Wild type \textit{Mtb}-CRP and the DNA motif; 2) Wild type \textit{Mtb}-CRP saturated with 500uM cAMP prior to addition of DNA; 3) \textit{Mtb}-CRPMut with DNA; and, finally 4) \textit{Mtb}-CRPMut saturated with cAMP (500uM). The Kd values for the DNA binding events for \textit{Mtb}-CRP with and without cAMP saturated states showed positive effect on DNA binding efficiency. The dissociation constant (Kd) dropped by almost 100 folds upon addition of cAMP (from 16nM to 0.16nM). In a parallel experiment with \textit{Mtb}-CRPMut which was devoid of novel non-canonical pocket no huge drop in Kd was evident rather it was only ~3 times (from 30nM to 10nM). Interestingly, EC-CRP showed very poor DNA binding when it was treated with higher concentrations (0.5mM) of cAMP (Harman, 2001; Garges and Adhya, 1995). This clearly indicated that the effects of cAMP binding on DNA interactions were different in EC-CRP and \textit{Mtb}-CRP and novel non-
canonical cAMP pocket have a critical role in regulation of DNA binding by *Mtb*-CRP.

Further results showed that energetics of DNA binding were also very distinct in *Mtb*-CRP. The change in enthalpy for cAMP liganded-*Mtb*-CRP (15.5 kcal/mol) and DNA interaction was rather higher than the energy released in unliganded-*Mtb*-CRP (9.15 kcal/mol). This high energy release was also compensated by high amount of change in entropy (-15cal/mol/K) in liganded interaction. Entropy change difference between liganded and unliganded interactions was almost 20cal/mol/K. While in similar experiment with *Mtb*-CRPMut big difference in release of energy was not seen. High amount of enthalpy changes could be a reflection of specificity of binding reaction. According to present accepted model for EC-CRP, binding of first cAMP favors the binding of cAMP molecules at secondary cAMP sites. Binding of extra cAMPs to the secondary binding site decreases its specificity for DNA (Lin, 2002; Scott, 2005). It should be noted that the secondary cAMP binding site in *Mtb*-CRP is at a different location than that in EC-CRP. It is therefore likely that the two binding site have different mechanism. ITC DNA binding data on *Mtb*-CRP and *Mtb*-CRPMut clearly showed that secondary cAMP increases the specificity of CRP for DNA instead of decreasing its specificity.

Currently, the only CRP-DNA complex structures available are from *E.coli*. Therefore a comparative analysis of the DNA-protein interactions between EC-CRP and *Mtb*-CRP was carried out. The minimal
motif in DNA as well as in protein which is necessary for interaction was defined. A predicted helix-turn-helix (176-LTQEEIAQLVGASRETVNKALA-196) motif which is involved in DNA binding for *Mtb*-CRP was identified. With the help of DNA-protein complex crystal structure it was further narrowed down to the exact minimal signature motif responsible for interaction. In EC-CRP this is Arg-Glu-thr-Xaa-Xaa-Arg (residues180-185) while for *Mtb*-CRP it was Arg-Glu-thr-Xaa-Xaa-Lys (residues188-193). At the last position in case of *Mtb*-CRP the lysine residue was substituted by arginine. While for DNA motif in EC-CRP the motif is “GCGA”, in case of *Mtb*-CRP it was “GTGA”. Except these base specific interactions some non-specific DNA backbone specific H-bonds were also observed. Val146, Asn192 and Arg65 could also interact with DNA backbone via H-bonds and salt bridges. Surprisingly, Ser187 and Glut178 in subunit B were able to form H-bond with the DNA backbone but not with subunit A. These obvious differences were observed between two subunits because two subunits were not identical in all respect as a least square superposition between two subunits yielded RMSD of 0.44Å. This could be an effect of binding of cAMP to secondary pocket in more molecules in subunit B than subunit A as seen in the structure. More DNA-protein interaction in subunit B comparison to subunit A was observed. Overall pattern of interactions was found to be similar except for two interactions, the lysine in *Mtb*-CRP could interact to only one base specifically while equivalent arginine in EC-CRP could make one
more interaction with thymine. On the other hand glutamate in \textit{Mtb}-CRP signature motif could interact with thymine. So, in conclusion the overall interactions were conserved.

Although, \textit{Mtb}-CRP is 32\% identical to EC-CRP in terms of aminoacid sequence, the minimal signature motifs for DNA-protein interaction were also very similar. The fashion of cAMP binding to the protein was unique in \textit{Mtb}-CRP. Further implications of binding which lead to fate and extent of DNA binding were completely different than EC-CRP. In \textit{Mtb}-CRP secondary cAMP pocket was completely unrelated.

The \textit{Mtb} genome encodes as many as 15 adenylate cyclases, suggesting that cAMP may have an important role in mycobacteria. It has indeed been reported that cAMP can alter the gene expression profile of \textit{Mtb} during anaerobic conditions (Gazdik and McDonough, 2005).

While these in vitro findings point to the importance of cAMP, it remains to be experimentally demonstrated whether cAMP is actually involved in regulating gene expression by recruiting \textit{Mtb}-CRP. While the biophysical features of purified \textit{Mtb}-CRP described here are physiologically relevant, experimental validation in vivo will be required to dissect the complete network of \textit{Mtb} genes regulated by \textit{Mtb}-CRP and cAMP. cAMP, acting as effector, is known to modulate the regulation of a large number of target genes, and it is likely that \textit{Mtb}-CRP will be involved in this process.
Mtb-CRP is essential for survival of TB bacilli inside host cells (Rickman et al., 2005). Small unique but important structural features of Mtb-CRP described in this thesis could be targeted for drug discovery approaches and could lead to design of effective therapeutic interventions against Mtb in future.