Summary

The information released by a recent report of World Health Organization (WHO) is a matter of great concern, as it states that Tuberculosis (TB) is one of the top ten leading cause of deaths worldwide. This increase in the number of deaths is largely contributed by drug resistant strains such as in MDR and XDR cases of TB and therefore it necessitates the requirement for development of novel drugs against these strains. In order to develop novel drugs, new targets need to be identified and to do so, we need to understand the complete pathobiology of *Mycobacterium tuberculosis* (*Mtb*), which is a causative agent of TB. A significant step towards understanding the biology of this pathogen was provided by whole genome sequencing of *Mtb* which resulted in 3984 genes (Cole et al., 1998). The genome sequencing of *Mtb* has also helped us to identify the essential genes which are required for survival of pathogen (Griffin et al., 2011; Sassetti et al., 2003; Xu et al., 2013). Some of the essential genes have been characterised but about one quarter of essential genes still corresponds to proteins of unknown function and are annotated as 'hypothetical protein' in the database. Structural and functional characterization of these essential hypothetical proteins will help in understanding the pathogenic and survival mechanism of bacteria which will further assist in designing new drug molecules.

The Rv1828, is an essential and uncharacterised gene from *Mtb*. It is predicted to be a HTH MerR transcriptional regulator and therefore anticipated that it may play a role in regulating some of the essential processes of the pathogen. Owing to the essentiality and its predicted function, the structural and functional characterization of Rv1828 may help in understanding the biology of the pathogen and may be explored as new drug target against multidrug resistant *Mtb*. Thus, to understand structural and functional properties of Rv1828 and to develop it as a potential drug target we initiated the characterization of Rv1828. Accordingly, the *Rv1828* gene was cloned into pET28b vector, expressed in BL21 (DE3) cells and purified using Ni-NTA and gel filtration chromatography. The size exclusion chromatography studies revealed that Rv1828 exist as dimer in solution. The far UV CD scan of Rv1828 confirmed that it is properly folded and contain both helices and sheets.

It is well known that the MerR family transcription regulator consists of two domains, a DNA binding domain at its N-terminal and an effector binding domain at its C-terminal region. In addition, it is known that the MerR family transcription regulators tend to bind to

its own promoter region thereby regulate its own expression. Since Rv1828 is predicted to be a MerR transcription regulator we checked its DNA binding ability by EMSA method. Using EMSA, we show that Rv1828 can bind to a 239 bp fragment of DNA from its own putative promoter region and subsequently we show that even 80 bp (that lies close to the start site of operon Rv1828) from the 239 bp fragment of DNA can bind to Rv1828. Apart from own putative promoter, the Rv1828 is reported to bind to the putative promoter region of seven genes (Rv0042c, Rv1141c, Rv0307c, Rv0308, Rv2150c, Rv2742c, Rv2890c). These putative promoter regions were identified by genome wide binding analysis of all predicted transcriptional regulators of *Mtb* (Minch et al., 2015). To confirm that Rv1828 indeed binds to these putative promoter regions we carried out the EMSA study of Rv1828 protein with all seven putative promoter regions. These binding experiments revealed that the promoter region of Rv2150 (*FtsZ*) showed maximum binding with Rv1828, in fact, even better than its own putative promoter binding. It is important to note that the FtsZ protein, encoded by Rv2150c in *Mtb* is crucial for bacterial life cycle as it is required for its cell division (de Boer et al., 1992). Thus, we speculate that the binding of Rv1828 to the promoter region of Rv2150 gene may play a role in regulation of this gene at transcriptional level.

As N-terminal domain of Rv1828 is predicted to have DNA binding domain, we cloned, expressed and purified the N-terminal domain (Rv1828-NTD) and the C-terminal domain (Rv1828-CTD) of Rv1828 to further understand and validate their function. In size-exclusion chromatography studies the Rv1828-NTD exist as monomer in solution while Rv1828-CTD exist as dimer in solution. The EMSA study was carried out with Rv1828-NTD and Rv1828-CTD proteins along with the DNA from its own putative promoter as well as with the promoter region of Rv2150c. These results revealed that the Rv1828-NTD show binding with both the promoter region of DNA but with little lesser affinity than full length Rv1828 protein. However, the Rv1828-CTD does not show any binding with any of the DNA indicating that the Rv1828-NTD domain is responsible and sufficient for its DNA binding. In addition, we also checked the binding of Rv1828 and Rv1828-NTD proteins with promoter region of Rv2150c in presence of metal ions like $HgCl_2$, $CuCl_2$ and $ZnCl_2$. We observed a loss of binding in the presence of $HgCl_2$, $CuCl_2$ in both Rv1828 and Rv1828-NTD proteins while its binding was intact in presence of $ZnCl_2$. The role played by metal ions which leads to loss in binding of Rv1828 with DNA is still not clear and needs further study.

In an attempt to structurally characterise the Rv1828 protein, we initiated its crystal

determination. However, in spite of different strategies and extensive attempts we could not get a diffraction quality crystal. However, we are fortunate enough to get a crystal structure of C-terminal domain, which putatively encompass a novel fold. We have determined the crystal structure of Rv1828-CTD upto 1.5 Å resolution using SAD method. To solve the structure by SAD method, we did mercury atom soaking of native crystals to obtain required phases. The crystal structure of Rv1828-CTD showed a stable dimer with two monomers in the asymmetric unit. In fact, the structural data validates our biophysical observation where the C-terminal domain shown to be responsible for the Rv1828 dimerization. The search for structural homologs of Rv1828-CTD through DALI serer revealed that it shares a remarkable structural similarity to SCO5550 protein from *Streptomyces coelicolor* with r.m.s.d of 2.8 Å despite they don't share any sequence similarity at the C-terminal domain. Importantly, both Rv1828 and SCO5550 form dimer through their C-terminal domains, which is a unique characteristic of MerR family of transcription factors and thus they are considered to belong to unique subfamily of MerR transcriptional regulators. Since, SCO5550 is also an uncharacterised protein and holds an unknown function for its C-terminal domain, we could not identify any ligand that would bind to Rv1828-CTD on similarity basis.

During our attempts for crystallisation of Rv1828, we also tried crystallization of homologous proteins of Rv1828 from four *Mycobacterial* species, which are *Mycobacterium parascrofulaceum*, *Mycobacterium avium*, *Mycobacterium smegmatis* and *Mycobacterium xenopi*. We cloned, expressed and purified the Rv1828 homolog proteins from these four species. Although the purified protein from these four species did not yield any crystals, they showed similar DNA binding affinity with promoter region of Rv2150c as that of Rv1828 protein indicating that the function of these proteins will be similar to that of Rv1828.

We also predicted the monomeric structure of Rv1828 full length protein using I-TASSER (Zhang, 2008) server which used SCO5550 as its template. Based on the organization of SCO5550 as dimer we generated a dimer for Rv1828 too. In order to analyse interactions between Rv1828 and promoter region of Rv2150c, we predicted the Rv1828-DNA complex structure and identified the amino acids as well as nucleotide residues important for binding with DNA. To validate the modelled structure, we mutated the interacting amino acid residues and did EMSA with mutant proteins and DNA fragments of promoter. As a result it was observed that the residues from 31-35 of Nα2-helix and Tyr-51 and Arg-52 residues of Nβ2 strand are important for interacting with DNA. The prediction

also indicated that the nucleotides of inverted repeat sequence present in the promoter region were also found to interact with amino acid residues of α2-helix and Nβ2 strand in the modelled structure. Strikingly, when these repeat nucleotides were mutated to some random sequence, a loss of binding was observed with Rv1828 as shown by EMSA experiments. This confirms our prediction that the Rv1828 binds to the inverted-like repeat sequences of promoter region.

In summary, we elucidated the structure-function relationship of Rv1828 from *Mtb*. Since Rv1828 happens to be a potential drug target, the information provided by this work can be helpful in developing drugs against *Mtb*.