5.1. Summary

Antibiotic resistance in the human pathogens is now on the rise. Diseases that were once deemed treatable are no longer easy to treat. The pathogens are evolving at a higher rate than we are able to develop new drugs. Conventional antibiotic groups are now seemingly futile for treatment of many resistant varieties of infections. One such disease is tuberculosis, caused by the bacteria *Mycobacterium tuberculosis*. Tuberculosis has affected almost one-third of the world population. As resistant varieties of this pathogen are emerging, it is high time to look for novel drugs as well as drug targets. Hypothetical essential genes have been a grey area in *M. tuberculosis*. Such genes may provide important insights into the survival mechanisms of the pathogen. The focus of my study is on establishing novel drug targets by characterizing a hypothetical essential gene in this pathogen.

The gene selected here is *Rv2229c*. Preliminary bioinformatics analysis was performed using various sequence based tools (like BLAST, pfam, CDD, TMHMM, PSORTb, IntAct, STRING, 2ZIP) to determine possible function of the protein. Sequence based studies suggested the protein to be primarily bacterial with exceptional occurrences in few eukaryotic lower species. It contains a zinc ribbon domain at the C-terminal, which is classified in PFAM database as zf-RING_7, a domain of unknown function. Sequence based analysis didn't yield any significant hint towards the function of Rv2229c. Therefore, we explored the option of structure based function determination.

Rv2229c was successfully cloned in pET28b vector with N, C and N+C terminal His tag and TEV protease site for N-terminal His tag. Rv2229c protein was overexpressed in E. Coli BL21(DE3) expression system and the protein was obtained in soluble fraction. It was purified using Ni-NTA affinity chromatography followed by gel filtration chromatography. The presence of zinc ion was confirmed using PAR assay. Expression of domains individually also helped us to confirm that zinc is present only in the C-terminal domain. Circular dichroism studies suggested that protein is fairly stable with a melting point of about 50°C. Crystallization trials with Rv2229c didn't yield any crystals. Several other approaches, e.g., changing tag position, fusion protein and removing tags, also failed to yield the crystal. Finally we tried to obtain crystals using the homologs of Rv2229c. The homolog from *M. smegmatis* (msmeg_4306) yielded weakly diffracting crystals initially, which were improved in subsequent steps. The structure for msmeg_4306 has been solved by SAD phasing method using intrinsically bound zinc ion as anomalous scatterer. The X-ray diffraction data was collected at BM14 beamline, ESRF, Grenoble, France and was solved upto a resolutoin of 2.8

Å. Later on, a higher resolution data (2.6 Å) was collected at home source and was used for further analysis.

The msmeg_4306 structure consists of two domains – the N-terminal coiled coil helical domain and the C-terminal zinc ribbon domain. In order to predict a function for msmeg_4306, we performed a structure similarity search using DALI server. Two hits were obtained – HP0958 from *Helicobacter pylori* and CT398 from *Chlamydia trachomatis*. Although the sequence similarity or msmeg_4306 is very less as compared to HP0958 and CT398, they share a remarkably similar structure and domain organization. In fact, the structure of C-terminal zinc ribbon domain is unique to these three proteins (out of all the proteins with solved structures till the writing of this thesis). Both CT398 and HP0958 are involved in Type III secretion system where they act as a chaperon for sigma factor σ^{54} (RpoN). In *H. pylori*, HP0958 is also involved in co-translational export of FlaA protein. based on such close similarity in structure, we predict that msmeg_4306 may be involved in secretion system events and also may be interacting with multiple macromolecules. One more reason that strengthens our belief of msmeg_4306 (and thereby Rv2229c) involvement in secretion systems is that it has been reported that Rv2229c is localised in the membrane and cell wall fraction of *M. tuberculosis* H37Rv (Mawuenyega *et al.*, 2005; Målen *et al.*, 2010; de Souza *et al.*, 2011), despite being found to be quite soluble proteins in our experiments.

5.2. Future directions

The structure determination of msmeg_4306 will help in modelling of Rv2229c with good confidence given its close sequence identity. Thus, this study lays a foundation for future work wherein this structure can be useful at multiple frontiers. Given the essentiality of Rv2229c in growth of *M. tuberculosis*, it can help in drug discovery process. On the other hand, most of the mycobacterial secretion systems are still poorly understood. This structure might help in understanding the same and subsequently help in interpreting their roles in survival, pathogenesis and infection.