6 Summary

Tuberculosis (TB) is a public scourge, which has been declared as a global emergency by WHO in 1993. According to WHO reports in 2014, it accounts for 1.5 million deaths and 9.0 million new cases per year worldwide. The emergence of drug-resistant bacilli and its associations with HIV are the major factors to worsen the situation (Corbett et al., 2006).

The disease is caused by *Mycobacterium tuberculosis* (*Mtb*) and disseminates from one patient to other individual in an airborne route through the bacilli present in the cough (Russell et al., 2009). According to an estimation, by Russell and colleagues, the single bacillus is sufficient to spread the infection. The notorious bug is smart enough to evade our immune system and survive within the macrophage for a long time in quiescent stage (Raja, 2004).

A milestone in tuberculosis treatment was achieved with the discovery of anti-TB drugs, which saved million of lives (Tomioka and Namba, 2006). Till date, the best treatment for tuberculosis is a combination of four drugs that includes isoniazid, rifampicin, pyrazinamide and ethambutol (Gatey and Bouvet, 2012). But there are some issues like patient compliance, longer duration, and drug resistance posing a threat to treatment (Griffiths et al., 2010). Therefore, new possibilities for tuberculosis treatment are the need of the hour.

Calmette and Guerin had put their efforts to develop a vaccine from the bovine strain of mycobacteria. Their efforts brought fruits and after 13 years of continuous passaging, they were able to attenuate the bacilli. It was tested on the human in 1921 and found to be safe as well as protective. WHO have licensed the vaccine for human use in 1974 (WHO, 2014). Till date, BCG is the only approved vaccine against tuberculosis.

BCG has been found to be very protective against the pulmonary form of tuberculosis, especially among children and in developed countries (Mangtani et al., 2014). But, it failed to be protective in the high epidemic countries where it was most needed (Pönnighaus et al., 1992). The observed range of protection varies from 0 to 80%. BCG is also not recommended in immune-compromised individuals (WHO, 2007). Helminths' co-infection and environment mycobacteria are also attributing to the weaker protection mediated with BCG (Brandt et al., 2002; Elias et al., 2008). Therefore, better vaccine candidates are required to tackle this deadly bug.

6.1 Desired immune response

The generation of required immune response is a prerequisite for a better vaccine. *Mtb* has evolved an evasion mechanism to divert our Th population from Th1 to Th2 (Rook, 2007; Yahagi et al., 2010). Peptides activating the Th1 cells constitute a protective environment in tuberculosis (Radosevic et al., 2007; Sable et al., 2007). Interestingly, single amino acid substitution in peptide sequence may switch the immune response from Th2 to Th1 (Pfeiffer, 1995).

In the last decade, a number of algorithms from our group have been developed for the prediction of B-cell epitopes (Ansari and Raghava, 2010; Saha and Raghava, 2006; Singh et al., 2013), MHC binders (Lata et al., 2007; Singh and Raghava, 2001, 2003) and T-cell epitopes (Bhasin and Raghava, 2004a, 2007). These tools are very accurate in prediction of epitopes, but none of the tools can predict cytokine milieu generated after the immunization. Therefore, to extend our tools toward greater applicability for designing better vaccine candidates, we paid our focus on the development of tool that can predict the peptides inducing Th1 or Th2 cytokine.

To achieve our goal, we have selected the related dataset from IEDB, which stores the experimentally validated epitopes (Vita et al., 2010). We have filtered MHC class II binders which have been tested for the induction of IFN-gamma, which is signatory cytokine for Th1 population (Abbas et al., 1996; Romagnani, 2000). We observed that the amino acid composition, positional conservation, and peptide-length were contributing significantly to predict the IFN-gamma inducing peptides. We have also discovered important motifs responsible in this regard. Further, we employed a machine-learning approach to discriminate the interferon-gamma inducing peptide from non-IFN-gamma inducing, peptides inducing other cytokines, and random peptides. We achieved decent performance in terms of accuracy of the algorithm. To serve the scientific community, we have developed a freely available web server named as 'IFNepitope', available at http://crdd.osdd.net/raghava/ifnepitope.

Next, we have also made an attempt to develop an algorithm for the designing of peptides which activate Th2 population. IL4 cytokine has been recognized as a major cytokine in Th2 population (Romagnani, 2000). We have extracted the dataset from IEDB, where MHC class II binders have been tested for IL4 secretion. We observed that AAPs and dipeptide composition are able to distinguish the IL4 inducing epitopes from the peptides which do not induce IL4 and randomly generated peptides from Swiss-Prot (UniProt Consortium, 2012). In our analysis, we have observed that

certain MHC alleles were found to be associated with the release of an IL4 cytokine. Additionally, we have also discovered residues and their certain positional dominance for the release of the IL4 cytokine. We have extracted important motifs that favored the induction of IL4. We have employed motifs and AAP information to predict the peptides using the machine-learning algorithm. The performance of our algorithm was also evaluated using independent dataset. We have implemented our models in the form of an open source server 'IL4pred' available at http://crdd.osdd.net/raghava/il4pred. We hope the servers will help the scientific community working for the designing of vaccine candidates.

6.2 Selection criteria for antigen

Besides desired response, selection of antigens is also very critical for an improved vaccine against tuberculosis (Elhay and Andersen, 1997). The antigen selection in tuberculosis has broadly been described in two schools of thought:

Shared antigens: Protein or peptide sequences that are conserved across the mycobacterial species have been classified in the shared antigen category. The conserved mycobacterial sequences are diverse in origin from human, therefore, found to be immunogenic (Bashir et al., 2010; Ohara et al., 1997). But the recent reports from clinical trials of a vaccine candidate with mycobacterial conserved antigen 'Ag85' were not found to be fruitful (Tameris et al., 2013).

Exclusive tuberculous antigens: Mycobacteria are residing in a diverse habitat from the soil, water and air to our internal organs. The environmental mycobacteria interfere with the immune system generated with BCG (Young et al., 2007). Therefore, ideal vaccine candidate should avoid the antigenic sequence shared with NTM and select only the antigens found exclusively in tuberculous strains of bacilli.

In the era of next generation sequencing (NGS), whole genome sequencing became very affordable (Zhang et al., 2011b). The affordable cost and rapid sequencing facilitated the discovery of new mycobacterial strains (loerger et al., 2009; Madhavilatha et al., 2012). These sequencing projects submit their results to a publicly available repository of NCBI. There were 59 strains with complete chromosome assembly available; at that time when we conceived the current project. Fortunately, the strains were sequenced to touch each important point in the study that includes environmental strain, clinical isolates, parental BCG with its vaccine variants, drug-susceptible, MDR, and XDR. We have classified the strains into three groups; tuberculous (causing tuberculosis), vaccine (BCG variants and parental strain)

and NTM (bacilli present in environment or our body and not causing TB to a healthy individual). NTM may be an opportunistic pathogen.

6.3 Mapping of epitope

We have mapped the already known epitopes from IEDB to the proteome of each strain. The user can browse all of these sequences and mapped epitopes from our module http://crdd.osdd.net/raghava/mtbveb/browse_strain.php. We have also developed a module to map the sequence of user's interest. We segregated the mapped epitopes in tuberculous, vaccine and NTM group. From here, we extracted the epitopic sequences found exclusively in tuberculous strains. There were different assays, which have been performed with the sequences and we selected only those epitopes that were proven to be B-cell epitope, T-cell epitope or MHC binders. We ended up with the single peptide **P1** "DQVHFQPLPPAVVKLSDALI" from Phosphate-binding protein, pstS 1, with above filters. The results from multiple cytokine assays were very confusing, as the peptide was shown to induce a mixed (Th1 and Th2) kind of response from different studies. After that, we have moved to the selection of antigenic sequence, based on the prediction.

6.4 Vaccine candidate proteins

In the context of tuberculosis vaccine, three classes of proteins are very important that includes 125 virulent factors, 20 components of secretion system and 33 RD antigens. In reference to above 178 antigens from *Mtb* H37Rv, we have compared the protein sequences from all the 59 strains. At 95% sequence identity cut-off, we found that 23 proteins were conserved in tuberculous strains, but they have undergone substantial changes in NTM and vaccine strains. We have also developed a module on our server to compare the proteins of user's interest, which is available at http://crdd.osdd.net/raghava/mtbveb/compare.php. Our further analysis revealed 13 potential candidates for rBCG and/or subunit vaccine, which were found to be containing proven B- and T-cell epitopes. The above-suggested 13 candidates were conserved in tuberculous strains and in NTM and vaccine strains, the identical sequences were not present, so these candidates may also serve as diagnostic purpose for TB. Based on a proven T-cell epitopic sequence "YLLADTFTV", we are suggesting Rv2350c antigen as the most promiscuous candidate in our study.

6.5 Strain-specific vaccine

Parent bovine strain and four variants of BCG have been sequenced, annotated at chromosome level and available for public at NCBI. But there is no method to

compare the antigenic repertoire of BCG variants with respect to each tuberculous strain (Behr, 2002; Brewer and Colditz, 1995). The strain-specific vaccine section of the study was carried out to search suitable vaccine variant of BCG for each tuberculous strain. We have compared 178 antigenic sequences from tuberculous strains and extracted the 95% identical sequences from vaccine strains. We have observed that Tokyo strain of BCG shared a maximum number of antigens with each of the tuberculous strains. Additionally, we have also aligned the closest sequence from each strain taking reference sequences to *Mtb* H37Rv using ClustalW (Larkin et al., 2007). All the alignment files are available at our MtbVeb server for public use. Additionally, in a segment of our platform, "MtbVeb", user can submit sequence from SRS, contig or proteome format, to extract the antigenic sequences from our study and predict the closest variants of BCG, based on the sequence comparison.

6.6 Subunit vaccine

In our last section of the thesis, we have generated the peptides of 9 residues (9mers) from each antigenic sequence extracted from all the 59 strains. We have compared the 9mer among three categories of strains and peptides exclusive to tuberculous strain were selected. We have also created a pipeline using epitope prediction programs touching every angle of immune response desired for an ideal tuberculosis vaccine. The pipeline included the linear B-cell tool (LBtope), MHC-peptide promiscuity prediction algorithm (ProPred, ProPred1, nHLApred), identification of T-cell epitope server (CTLPred) and predicting cytokine inducing approach (IFNepitope and IL4pred) (Bhasin and Raghava, 2004a, 2007; Dhanda et al., 2013a, 2013b; Singh and Raghava, 2001, 2003; Singh et al., 2013). All the peptides with predicted immune response were made public at our platform "MtbVeb". We have applied different logical gates to select the peptides with desired immune response. In conclusion, we strongly recommend **seven subunit vaccine candidates** for experimental validation.