Introduction
A curable and preventable disease, Tuberculosis (TB) continues to be a leading cause of mortality and morbidity worldwide. TB is a complex disease caused by *Mycobacterium tuberculosis* (*M. tuberculosis*) which has evolved with highly successful mechanisms to equivocate host defenses and existing classes of antibiotics. Decades after the discovery of *M. tuberculosis* test for *M. tuberculosis* infection, vaccine against TB and drugs to treat disease, TB remains a major cause of morbidity and mortality in many developing countries. One third of the world’s population is considered to be infected with *Mycobacterium tuberculosis*, which leads to nearly 9.4 million new patients and 1.7 million deaths in the year 2009 (WHO Report 2010). Multi-drug-resistant strains of this pathogen, emerging in association with HIV, have added a frightening dimension to the problem (Raviglione et al. 1995). Outbreaks of extensively drug-resistant (XDR) tuberculosis have also been an increasing threat in certain regions around the world (Shah et al. 2007). Despite many studies been done on TB, this disease poses a considerable risk in many developed countries. However, the global health costs in both humanistic and economic terms require an unrelenting pursuit of new and effective treatments - an important enterprise where computational biology can make significant contributions. *M. tuberculosis* is very virulent. But there is no simple answer found yet for, what makes *Mycobacterium tuberculosis* so virulent? Although, much has been learned about the structure of this organism, the epidemiology of the tuberculosis, the physiological and immunological responses of the host to infection and the physiology of *M. tuberculosis* in vitro conditions, much about the basic biology of *Mycobacterium tuberculosis* in the infected human remains to be elucidated (Clark-Curtiss et al. 2003). Due to the advancement of Bioinformatics and genome sequencing project, whole genome and proteome sequences of different organisms are available in the public domain. Currently complete genome sequences of four clinical strains of *Mycobacterium tuberculosis* (*H*37Rv, CDC 1551, F11 and KZN 1435) and one avirulent strain *H*37Ra is available (http://www.ncbi.nlm.nih.gov/genome/). These strains are phenotypically and genotypically different from each-other but they cause the same disease in humans, although the virulence power is different among these strains. Current evidences suggest that as a species *M. tuberculosis* exhibits very little genomic sequence diversity (Musser et al. 2000; Sreevatsan et al. 1997). *M. tuberculosis* should also exhibit very little phenotypic variation in immunologic and virulence factors. However, evidence of phenotypic diversity among clinical isolates conflicts with this hypothesis (Manca et al. 1999; Valway et al. 1998).
M. tuberculosis H₃⁷Rv is virulent and susceptible to most of the antitubercular drugs used so far, H₃⁷Ra which is an avirulent strain (Zheng et al. 2008) and M. tuberculosis KZN strain is resistant to different drugs like isoniazid, rifampicin, kanamycin, ofloxacin, ethambutol, pyrazinamide etc (Ioerger et al. 2009), there must be some genetic or proteomic mutations present in them. So, there is a need for genomic as well as proteomic analysis among different strains of M. tuberculosis to know the variation among them. The complete sequence and annotation of the M. tuberculosis genome has allowed many new genetic approaches to studies of the physiology and pathogenicity of this organism. Many tools have also been developed for the complete determination of the genome sequence of a huge number of bacterium, but still, their proteomes remain relatively poorly defined.

Historically Mycobacterium tuberculosis H₃⁷Ra is the avirulent counterpart of virulent strain H₃⁷Rv and both strains are derived from their virulent parent strain H₃⁷, which was originally isolated from a 19 year old male patient with chronic pulmonary tuberculosis by Edward R. Baldwin in 1905 (Steenken et al. 1946). Variation in virulence between M. tuberculosis H₃⁷Rv with respect to H₃⁷Ra strains is still to be understood. A single amino acid mutation in protein sequence may cause alteration in protein structure and function that may account for virulence and drug resistance properties of pathogenic organisms. Identification of the virulence factors of M. tuberculosis is a fundamental goal if new vaccines and antimycobacterial drugs against this pathogen are to be developed. Therefore, the development of an in silico technology to study the proteomic variations of different strains of genetically intractable pathogens such as M. tuberculosis will enhance the analysis of virulence and drug resistance properties and significantly advance the understanding of the mechanisms of disease. The proposed work aims to study proteins undergone mutations in M. tuberculosis i.e. H₃⁷Rv and H₃⁷Ra strains, which may explore a unique discovery platform for comparative proteomics among other strains of M. tuberculosis also and give insights into the discovery & development of TB drugs, vaccines, biomarkers etc. Proteome comparison is significant as proteins represent the actual functional molecules in the cell and helps to determine whether there are any major differences between different strains at the protein level. When mutations occur in the DNA, it is the proteins that are ultimately affected. The Proteome comparison analysis provides information on domain structure and function, gene duplication and protein families in different genomes. Different Bioinformatics tools to interrogate and compare entire proteomes of different organisms make it possible to identify systematically conserved proteins, conserved families that are missing in a given genome and proteome or proteins unique to a particular species.