**ABSTRACT**

*Mycobacterium tuberculosis* (MTB) H₃₇Ra is an avirulent strain of MTB which is genotypically related to the virulent MTB H₃₇Rv strain. In spite of large scale study, variation in virulence between MTB H₃₇Rv and MTB H₃₇Ra strain is still to be understood. Difference in protein expression or structure due to mutation may probably be an important factor for virulence property of MTB H₃₇Rv strain. In this study, we have designed an algorithm epcALGO using NCBI standalone BLAST program, for entire proteome comparison and carried out whole proteome comparative analysis between these two strains using Bioinformatics approaches to elucidate differences in their protein sequences. On comparison of whole proteome between these two strains, revealed 3759 identical proteins in both the strains out of 4003 proteins in MTB H₃₇Rv and 4034 proteins in MTB H₃₇Ra. 244 proteins of MTB H₃₇Rv and 260 proteins of MTB H₃₇Ra were found to be non-identical. A total of 172 proteins were identified with mutations (Insertions / Deletions / Substitutions) in MTB H₃₇Ra while 53 proteins of MTB H₃₇Rv and 85 proteins of MTB H₃₇Ra were found to be distinct. Among those 244 non identical proteins, 40 proteins were identified with single amino acid variation in MTB H₃₇Ra. As 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, we further studied the effect of single amino acid mutations on *Mycobacterium tuberculosis* H₃₇Rv and H₃₇Ra using different mutation analysis systems such as SIFT, PolyPhen, PROVEAN and HOPE. Subsequently, structure comparison of proposed modeled structure of mutant proteins of MTB H₃₇Ra with native proteins of MTB H₃₇Rv was performed. We observed that amongst 40 single amino acid mutated proteins of MTB H₃₇Ra, 5 were found to be damaged by all the mutation analysis systems. All the five proteins showed important biological function in MTB.
Among those five proteins, one protein namely Dihydroxy acid dehydratase (DHAD) encoded by Rv0189c gene was reported as essential enzyme for the survival of *M. tuberculosis*, non-homologous to human and it was proposed as potential drug target. Therefore, structure based virtual screening was implemented for identification of potential novel inhibitors targeting dihydroxy acid dehydratase. A 3D model of DHAD was built using Phyre 2 server followed by structural refinement and energy minimization by YASARA server. Procheck, ProSA and ProQ were employed to access the reliability of refined model. A dataset of 135 Drug like compounds were retrieved from ZINC database based on the properties similar to the natural substrate (2,3-dihydroxy-3-methylbutanoate) of DHAD. Molecular docking program Auto Dock Vina in PyRx 0.8 was used as a tool for virtual screening and two drug-like compounds, ZINC40397312 (1-(1H-1,2,3,4-tetrazol-5-yl)cyclobutan-1-amine) and ZINC00330490 (6-hydroxy-2-(methylamino)-3,4-dihydropyrimidin-4-one) were identified as potential inhibitors against DHAD of MTB. The docking result was further established by molecular dynamics simulations. These inhibitors might be useful for design of drugs targeting BCAA biosynthesis pathway.

Proteome Comparison Database on MTB H$_{37}$Rv versus H$_{37}$Ra has also been designed in this study to provide access to proteome sequence comparison data along with detailed analysis on these two strains of MTB. This database presently hosts 8031 protein sequence comparison data obtained through whole proteome comparison of MTB H$_{37}$Rv against H$_{37}$Ra and *vice versa*. The database is uploaded and hosted on online web server to make it freely available over the internet.

This study reports the protein differences with mutation between MTB H$_{37}$Rv and H$_{37}$Ra, which may help in better understanding of MTB virulence and pathogenesis.