

## ***In silico* Screening of Protein Rv3228 to have a Vision towards Survival and Pathogenesis of *Mycobacterium tuberculosis* H37Rv**

**Hanisha Sharma, Swati Meena, Laxman S Meena\***

CSIR-Institute of Genomics and Integrative Biology, Mall Road, Delhi-110007, India

### **Research Article**

Received date: 24/09/2019

Accepted date: 04/10/2019

Published date: 11/10/2019

#### **\*For Correspondence**

Laxman S Meena, CSIR-Institute of Genomics and Integrative Biology Mall Road, Delhi-110007, India,  
Tel: 0091-11-27002200; Ext:#234;  
Fax: 0091-11-27667471.

**E-mail:** meena@igib.res.in  
laxmansm72@yahoo.com

**Keywords:** *Mycobacterium tuberculosis* H37Rv, ATP/GTP binding protein, Mutation analysis, Ligand binding prediction

**Abbreviations:** TB: Tuberculosis; *M. tuberculosis*: *Mycobacterium tuberculosis* H37Rv; *M. smegmatis*: *Mycobacterium smegmatis*; *M. leprae*: *Mycobacterium leprae*; HIV: Human Immunodeficiency Virus; AIDS: Acquired Immunodeficiency Syndrome; WHO: World Health Organisation; GTP: Guanosine triphosphate; MDR-TB: Multidrug-Resistant TB; XDR-TB: Extremely Drug-Resistant TB ; MSA: Multiple Sequence Alignment; VICM: Prediction of Virulence factors, Information molecule; Cellular process and Metabolism molecule; ProBiS: Protein Binding Sites; BCPREDS: B Cell Epitope Prediction Server; MUSCLE: Multiple Sequence Comparison by Log-Expectation

#### **ABSTRACT**

Tuberculosis has appeared as a main world health problem with almost 1/3rd of the world society today infected with causative pathogen *Mycobacterium tuberculosis* (*M. tuberculosis*). *M. tuberculosis* is a gram-positive bacterium makes so many difficulties in its abolition totally. Rv3228 is a conserved hypothetical gene of *M. tuberculosis*. Rv3228 is an expected GTP/ATP binding protein. It also shows metal ion binding or GTPase activity. GTP/ATPs are energy-rich molecules that facilitate binding of respective protein factor either to ribosomes or to the tRNA. This manuscript considered some of the valuable aspects of Rv3228 protein being the function as unknown. The main visions of this study include retrieval of protein sequence database, multiple sequence alignment, string interaction study, sub-cellular localization, ligand binding prediction, B-cell and T-cell epitopes prediction, structure-based function prediction by COFACTOR and VICMpred. VICMpred predicts that this gene is a virulence factor. Ab-Initio modelling by RAPTOR X and validates by RAMPAGE, ERRAT, and VERIFY3D, mutation analysis by MAESTRO WEB SERVER. This study will be helpful in the development of new drugs and the treatment of tuberculosis disease.

## **INTRODUCTION**

Tuberculosis (TB) is potentially extensively dispersed perilous and fatal disease that spreads quickly and affect grievous bodily harm and it caused by *Mycobacterium tuberculosis* H37Rv (*M. tuberculosis*)<sup>[1]</sup>. Tuberculosis (TB) widens periodically and growing the numbers of persons have been suffering from this illness. *M. tuberculosis* is a significant human pathogen claiming more lives every twelve months than any further infectious disease<sup>[2]</sup>. Use of H37Rv as one and only reference genome in analysing clinical isolates present some limitations to completely investigating mycobacterium infectious disease virulence<sup>[3]</sup>. *M. tuberculosis* strain is mostly pathogenic strains in among *Mycobacterium* species like *Mycobacterium bovis* (*M. bovis*), *Mycobacterium leprae* (*M. leprae*), are pathogenic and *Mycobacterium smegmatis* (*M. smegmatis*) is a non-pathogenic strain. Mainly tuberculosis is a respiratory disease. It is also infecting other body parts including bone, brain, urinary tracts, etc.<sup>[4,5]</sup>. *M. tuberculosis* is aerobic

and gram-positive bacteria and vastly successful GC and lipid-rich bacteria. It is a facultative intracellular macrophage pathogen<sup>[6]</sup>. Crossing the respiratory tract, Mycobacterium lastingly resides within the alveolar macrophages wherever they reside for while with none hindrance by host immune system<sup>[7]</sup>. *M. tuberculosis* can enter in host movement by (M-cells) Microfold cell. M-cells originate within the (GALT) gut-associated lymphoid tissue in small intestine's Peyer's patches, and in the (MALT) mucosa-associated lymphoid tissue of further division of the human gastrointestinal tract<sup>[8]</sup>. Resistance to aerophilic killing, inhibition of phagosome-lysosome fusion and formation of the supposed electron-transparent zone (ETZ) impairs diffusion of lysosomal enzymes. Some number of the mechanisms which shall make a case for the survival of *M. tuberculosis* inside macrophages<sup>[9]</sup>. *M. tuberculosis* is often transmitted by aerosols and reaches the lungs, wherever macrophages and different immune cells are recruited during the first innate response to infection. Tuberculosis has become a major threat to public health<sup>[10]</sup>. Tuberculosis is the most dangerous communicable disease. Particularly, tuberculosis, being the predominant one, is extremely contagious. In past, the two hundred years, one billion infectious diseases (TB) deaths had occurred, and it's anticipated that within the next twenty-five years, over forty million folks could also be killed by TB unless control measures are enforced. There are varied causes that increase the susceptibility to *M. tuberculosis* infection; these embraces weakened immune system that happens through varied diseases and medications like human immunodeficiency virus (HIV)/immunodeficiency syndrome (AIDS), type II diabetes, end-stage renal disorder, alcoholism and endovenous drug use, certain cancers, cancer treatment like therapy, malnutrition and very young or advanced age. Other factors embrace tobacco use that will increase the chance of obtaining TB and dying from it<sup>[11]</sup>. The facade of multi-drug-resistant strains of this microorganism is additionally making a worldwide crisis. The confrontation of the microorganism within the cellular immunologic response and their capability to endure also as reactivation on an afterward part is barely implicit<sup>[12]</sup>. Pulmonary TB is common whereas *extra pulmonary tuberculosis* (EPTB) is additionally a current clinical drawback. In spite of having Bacille-Calmette-Guérin (BCG), vaccine and also the efficient short-course chemotherapy DOTS: Directly observed treatment, short-course, the exposure of drug-resistant strains with the deadly combination of the HIV is the only reason of increased patients of TB among developing countries<sup>[13]</sup>. Tuberculosis can be treated effectively by using 1st line drugs (FLD), isoniazid (INH), rifampicin (RIF), pyrazinamide (PZA), ethambutol (EMB) and streptomycin (SM). These are terribly effective for the treatment of TB but these drugs are currently unproductive to rehabilitate TB because of drug resistance in the bacterium. The emergence of multidrug-resistant TB (MDR-TB), i.e. resistance for INH and rifampicin, requires the utilization of second-line drugs that aren't simple to acquire and are expensive or poisonous than 1st line drugs<sup>[14]</sup>. TB is one among the highest ten causes of death and therefore the leading cause from a single agent (above HIV/AIDS). The World Health Organization (WHO) Global Tuberculosis Report indicated that in 2017 the best estimate is that 10.0 million individuals (range, 9.0–11.1 million) developed Tuberculosis disease: 5.8 million men, 3.2 million ladies, and 1.0 million kids. Drug-resistant Tuberculosis continues to be a public health crisis. The most effective estimate is that, worldwide in 2017, 558 000 people (range, 483 000–639 000) developed Tuberculosis that was resistant to rifampicin (RR-TB), the foremost effective 1st line drugs, and of these, 82 had multidrug-resistant Tuberculosis (MDR-TB). 3 countries accounted for pretty much half of the world's cases of MDR/RR-TB: India (24%), China (13%) and therefore the Russian federation (10%)<sup>[15]</sup>. In this manuscript, we performed a bioinformatics analysis to identify gene pairs associated with *M. tuberculosis* drug resistance. We predicted functions for hypothetical proteins and analyzed the detailed properties of the Conserved hypothetical protein Rv3228 of *M. tuberculosis* which showed ATP/GTP binding motif A. It is uncharacterized protein and one of the several proteins that assist in the late maturation steps of the functional core of the 30S ribosomal subunit. It helps in release of RbfA from mature subunits. It may play a role in the assembly of ribosomal proteins into the subunit. This protein circularly permuted GTPase that catalyzes slow GTP hydrolysis, GTPase activity is stimulated by the 30S ribosomal subunit that belongs to the TRAFAC class YlqF/YawG GTPase family. G-proteins (GTP-binding proteins) are extremely preserved signaling molecules that participate in cellular signalling and microorganism pathologic process by regulation the activity of cognate GTPases<sup>[16]</sup>. GTPases are stated as molecular modification proteins. These proteins specifically bind and hydrolyse GTP that successively activates or inactivates the GTPase in a very cyclic way. GTPases are highly conserved and function through ribonucleic acid (RNA) or ribosome binding<sup>[17]</sup>.

## MATERIALS AND METHODS

### Retrieval of Rv3228 protein sequence databases

Mycobrowser database has been used for retrieving sequences (gene and protein). *M. tuberculosis* data is available easily like nucleotide sequence or protein sequence else physiochemical properties are also easily available here. We have selected the protein sequence of the Rv3228 hypothetical gene for the IN-SILICO study. It has contained ATP/GTP site motif<sup>[18]</sup>.

### Multiple sequence alignment

Multiple sequence alignment of hypothetical conserved gene Rv3228 of *M. tuberculosis* had been done with various mycobacterium genes like *Mycobacterium marinum* (*M. marinum*), *M. bovis*, *M. leprae*, and *M. smegmatis*. For detection of

MSA, we use MUSCLE server for conserved hypothetical protein analysis. M-U-S-C-L-E stands for “Multiple Sequence Comparison by Log-Expectation” it is professed to accomplish both better normal precision and preferred speed over ClustalW2 or T-Coffee, contingent upon the picked selections <sup>[19,20]</sup>.

### Interaction study

We used the STRING-10 server to predict the cooperating or interacting partners of protein-protein interactions. This database uses a mixture of prediction methods and incorporation of other data. Maximum of the biological procedures, STRING-10 is an attractive target for discovering new mechanisms of the drug resistance. STRING-10 for protein-protein interaction has now arisen as major analytical tools for identification and characterization of expression of proteins and its native species <sup>[21]</sup>. The complete number of interacted proteins dataset had been saved in STRING and evaluate in the range of 0 and 1 the score shows <0.4 means low interaction between proteins, between 0.4 to 0.7 scores shows medium interaction and above 0.7 shows higher interaction between proteins <sup>[22]</sup>.

### Subcellular localization

TBpred is a prediction database that predicts confined subcellular localization like an integral membrane, secretory and membrane-attached by lipid anchor and cytoplasmic mycobacterial proteins. It is an SVM based process that shows different properties of the protein (dipeptide composition, position-specific scoring matrix and amino acid composition). It depends on provision vector appliance learning to predict the local subcellular places. With the help of this database different parameters showed their different values. The nth SVM model belongs to nth class samples with positive labels and rests other samples with negative labels. An unknown sample prediction is based upon the highest score out of 4 scores, generated by 4 models specific to 4 different subcellular compartments <sup>[23]</sup>.

### Prediction Of B-cell and T cell epitopes

The prediction of B cell and T-cell epitopes found in Rv3228 protein was completed by several online In-Silico approaches. BCPREDS server predicts B-cells epitopes, ABCpred server use for T-cells epitopes and ProPred tool predicts MHC-Class II Binding Peptide <sup>[24]</sup>.

### Structure-based function prediction

Function prediction through structure-based is predicted by the online server COFACTOR. COFACTOR is a protein-protein based collaboration, structure and arrangement-based procedure for an organic clarification of protein atoms <sup>[25]</sup>. Starting from the 3-dimensional basic model, cofactor will track the problem through the BioLiP protein function database by nearby and universal structure matches to distinguish functional conditions and homologies. Practical bits of information like Enzyme Commission (EC), ligand restricting locales and including gene ontology (GO) will get through the finest practical homology designs. The COFACTOR structure-based function prediction calculation was positioned as the best strategy for protein work forecast. We applied this approach for the prediction of molecular function and biological process. In the COFACTOR server, C score GO is the confidence score of predicted GO terms. C score GO values range within [0-1], where a higher value indicates better confidence in predicting the function using the template <sup>[26,27]</sup>. A VICMpred tool is an insilico approach; it is a direct method for prediction and classification of major functions like virulence factors, information molecules, cellular process and metabolism of bacterial and cellular proteins. Most of the proteins involved in virulence factors containing toxins, adhesions and haemolytic molecules <sup>[28]</sup>. KEGG Pathway Dataset- Kyoto encyclopaedia of genes and genomes (KEGG) pathways are a directly graphical representation of intersection between several types of proteins and genes <sup>[29]</sup>.

### Modeling

I-TASSER is a hierarchical protein structure modeling method which is based on the 2° structure raised Profile-Profile threading Alignment (PPA). The hypothetical protein Rv3228 protein modelling is done by I-TASSER database that shows the prediction of homology modeling of the protein. Through Ab Inito display approach I-TASSER creates a 3D model of protein of our targeted protein using FASTA sequence. It predicts also dynamic binding sites of our desired protein and it follows three ways to predict the 3-Dimensional model of the respective protein. For a further advance illustration of the 20 structure of the protein, I-TASSER introduced a Local Meta Threading Server (LOMETS) that uses H, E and C articulate for alpha-helix, beta-sheet and curl respectively <sup>[30]</sup>. In I-TASSER each protein models are quantitatively measured by C- score value which is given in Monte Carlo theory. The confidence score is a C-score for calculating and estimating the superiority of the predicted models. It is typically in the range of -5 to 2, where a higher value signifies a model with high confidence and vice versa <sup>[31]</sup>. A local threading meta server (LOMETS) is very quick and automated freely available server which predicts the tertiary structure and spatial constraints of proteins <sup>[32]</sup>. For targeted protein, we also performed homology modelling or cross evolution for the Phyre2 and RaptorX <sup>[33]</sup>. uGDT and GDT are the p-value of RaptorX to evaluating the quality of the model structure. GDT is figured as uGDT separated by the area length and duplicated by a 100. uGDT is figured between the model built from the best-positioned layout and indigenus. The cut-off value

of the RaptorX showed respective protein is p-value>4, 95% of the models have uGDT more prominent than 50. Then again, if for estimations of p-value<4, 98% of the models have uGDT under 50. This type of values validates if a model has uGDT larger than 50, which is acceptable<sup>[34,35]</sup>. PSIPRED server predicts the secondary structure of a protein with highly accurate results. This is a user-friendly server in which users can get result according to their choice. It gives results on the bases of PSI-BLAST and gives output in publication quality graphical representation<sup>[36]</sup>.

### Validation of the modeled protein RV3228

SAVES meta-server used to validate of the targeted modeled protein Rv3228. SAVES meta-serve has ERRAT, RAMPAGE, and Verify3D for the modification of modeled protein Rv3228. In SAVES meta-server the RAMPAGE, evaluated for its spine compliance utilizing a Ramachandran plot. RAMPAGE examine and analyses the stereochemical value of a protein assembly by analyzing residue-by-residue geometry and overall structure geometry demonstrated the protein validate score for the validated result in the most favoured region, additionally allowed region, generously allowed region and disallowed region<sup>[37]</sup>. Verify3D was utilized to approve the refined structure. The 3D structure of the protein was contrasted with its own amino acid sequence thinking about a 3D profile ascertained from the nuclear directions of the structures of correct proteins<sup>[38]</sup> and the general characteristics of the demonstrated structures were assessed by the ERRAT server<sup>[39]</sup>. A user can use all these programs in an input PDB structure or it can also do it individually run on the program one after one by the SAVES meta-server. ProQ online tool is used to find accurate models in contrast to other procedures which are used to discover the connate structures. With the help of ProQ server, we find two values measures which are predicted LG score and MaxSub<sup>[40]</sup>.

### Mutation analysis

A point mutation can have a solid effect on protein stability; Different in-silico methods to predict the change in stability upon point mutations have been developed. A recently introduced server MAESTRO proposed further benefits which are (i) operation on multimeric proteins, (ii) report of a prediction confidence value (iii) estimation of potential disulfide bonds and (iv) a scan mode for the most de-stabilizing n-point mutations. MAESTRO web provides four different experimentation-First one is the study of user definite mutations. The mutations sites and types can be specified via menus or by flexible mutation arrangement. Second is the type of experiment which scans for the most de-stabilizing n-point mutation ( $n \leq 5$ ), which is valuable for protein engineering tasks. Third is MAESTRO web offers the calculation of mutation sensitivity outlines, where the effect of the particular mutation at the particular possible location is visualized, similar to the PoPmusic server. This allows the identification of cold or hot spot sites individually, which are resistant or sensitive to mutations. The fourth experiment type is the evaluation of potential disulfide bonds. In adding to  $\Delta\Delta G$  values, geometric constraints are applied. Both geometry and  $\Delta\Delta G$  are combined with a disulfide bond score. Mutations can be restricted to certain amino acid classes, exposed or buried residues and user-specified regions on the basis of different experiments<sup>[41]</sup>. I-Mutant is a web server that predicts the stability of protein after point mutation. It provides stability of protein after point mutation on the basis of structural information and sequence information<sup>[42]</sup>.

### Ligand binding prediction

ProBiS plugin, freely accessible on our server at <http://insilab.org/probis-plugin>, is an associate extension of our earlier ProBiS-ligands approach accessible at <http://probis.cmm.ki.si> that predicts protein ligands by finding out structurally similar binding sites and transposition of ligands between these sites. Major enhancements within the ProBiS plugin over the sooner approach include: ProBiS plugin allows prediction of tiny molecule ligands and binding sites for all ~290,000 macromolecule chains within the PDB, whereas the ProBiS-ligands approach solely enabled prediction of ligands for the 42,000 protein chains within the ninety-fifth non-redundant PDB. The plugin calculates three-dimensional grid models of binding sites that outline the scale and also the form of the predicted binding sites. The new database of binding website comparisons is concentrated on tiny matter binding sites solely, whereas the info employed in ProBiS-ligands considered whole protein surfaces. With tiny ligand, mean synthetic or naturally occurring chemical compound, not peptides, proteins or other biological macromolecules. Consequently, a database explore for similar binding sites is much quicker<sup>[43]</sup>.

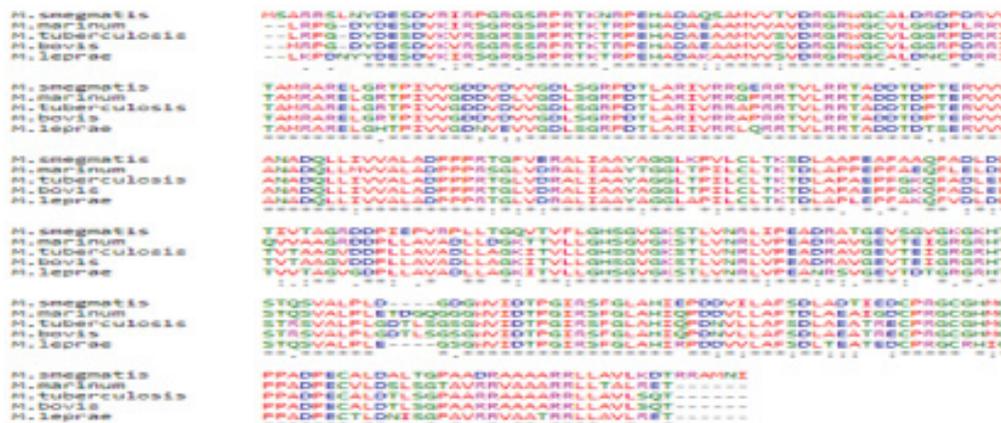
## RESULTS AND DISCUSSION

### Retrieval of RV3228 protein database

FASTA format or protein sequence of targeted protein Rv3228 has been retrieved from mycobrowser server. This protein 34857.8 Da and 993 bp long protein. Rv3228 is CDS type and a conserved hypothetical protein. This protein has not studied before and its functions are also unknown.

### Multiple sequence alignment

In MSA tool MUSCLE, we studied our query protein Rv3228 of *M. tuberculosis* which shows homology with their orthologues (*M. smegmatis*, *M. bovis*, *M. leprae*, *M. marinum*). Most of the sequences are conserved shown by an asterisk (\*), some are conserved between groups with strong similarity as shown by a colon (:), and some few are conserved between groups with less similarity as shown by period (.) (as shown in **supplementary Figure 1**).

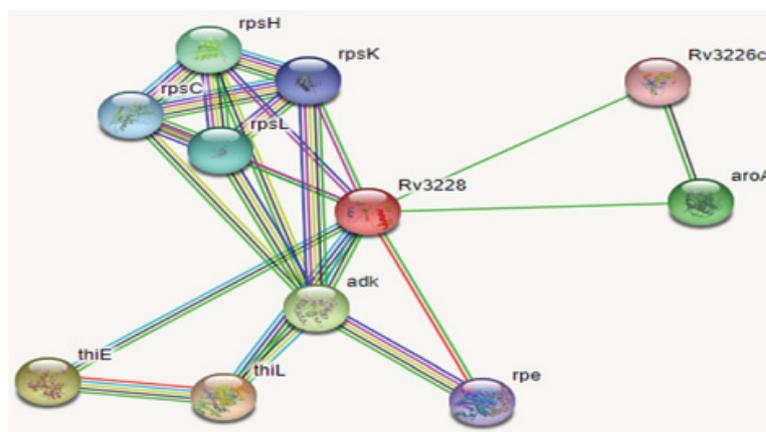


**Supplementary Figure 1:** Multiple sequence alignment.

This figure shows Systematic demonstration of the multiple sequence alignment by using MUSCLE Tool. The multiple sequence alignment of Rv3228 protein of *M. tuberculosis* result out the homology of this protein sequence with other proteins of *M. tuberculosis*.

**Interaction study**

String discovers the functional association or basic interaction between proteins that contribute jointly to a specified function. String database result predicts that Rv3228 interact with thiL, thiE, adk, aroA, rpsH, rpsL, rpsC, rpsK, rpe, Rv3226. Rv3228 shows highest interaction score with thiL (0.920) and with other proteins, it shows minimum interaction range up to 0.6-0.9 (as shown in **Supplementary Figure 2 and Table 1**).



**Supplementary Figure 2:** STRING Server: Protein-Protein Interaction.

In this figure shows Protein-Protein Interaction of Rv3228 by STRING tool. String server predicts the interacting partner of the respective protein with their efficiency of interaction. The figure shows that Rv3228 interacts with thiL with 0.920 score the cut-off value of this server is (0-1).

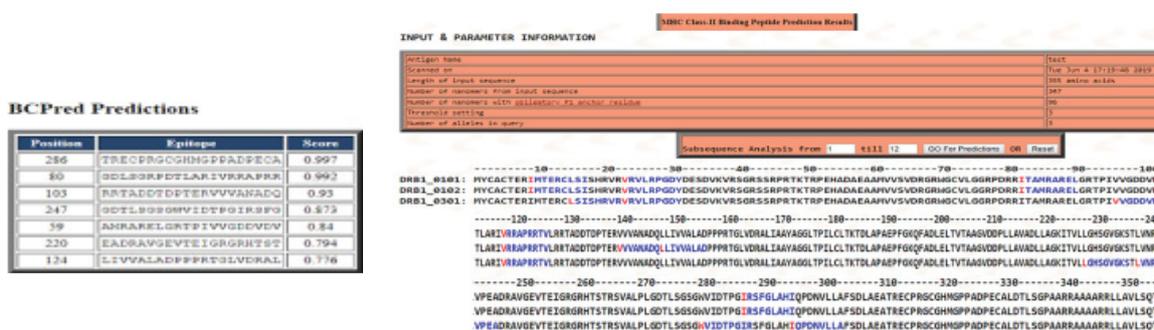
**Table 1:** Protein-Protein interaction study.

S. No.	Predicted functional partner	Predicted function	Score
1	thiL	Thiamine-monophosphate kinase protein	0.92
2	thiE	Thiamine-phosphate synthase	0.905
3	adk	Adenylate kinase	0.903
4	aroA	3-phosphoshikimate 1-carboxyvinyltransferase	0.876
5	rpsH	30S ribosomal protein S8	0.803

STRING server enlist protein-protein interactions with Rv3228 protein where score shown separately for each data confirms cut-off value within [0-1], low confidence: scores<0.4; medium: 0.4 to 0.7; high: >0.7. This table is showing interaction of protein Rv3228 with other proteins.

**Prediction of B-Cell and T-Cell Epitope**

B-cell epitopes predict by BCPred, it takes overlapping by the window of 16 amino acids that consequences in the preminent possibility of the score 0.96 from the residue region “RECPGCGHMGPPADP” that starts at 312 positions. Multiple DRB (DR β-1) alleles were used like HLA-DRB1\*0101, HLA-DRB1\*0102, HLA-DRB1\*0301 for the prediction of the T-cell epitopes with the MHC class-II binding region in the antigenic protein sequence of Rv3228. Five consensus epitopes were - IMTERCLSI, VRVLRPGDY, ITAMRAREL, VRRAPRRTV and IRSFGLAHI in (DRB1\_0101, HLA-DRB1\*0102) at the 09th-17th, 23rd-31st, 82nd-90th, 118th-126th, and 287th-290th residue positions. Two consensus epitope sequences were different in DRB1\_0102 VVANADQ, LLIVALAD at position 140th-147th and 148th-156th respectively, and nine consensus epitopes - LSISHRVR, VRVLRPGDY, VGDDVD, VVDLSGRP, VRRAPRRTV, LGHSGVGKS, LVNRLVPEA, WVIDTPGIR were observed in DRB1\*0301 at 15th-22nd, 23rd-32nd, 96th-102nd, 103rd-111th, 118th -126th, 228th-236th, 238th-246th, 280th-288th, 295th-302nd residue positions respectively in sequence as their respective alleles at 3% threshold as shown in T-cell epitope prediction (as shown in **Supplementary Figure 3A, Supplementary Figure 3B**).

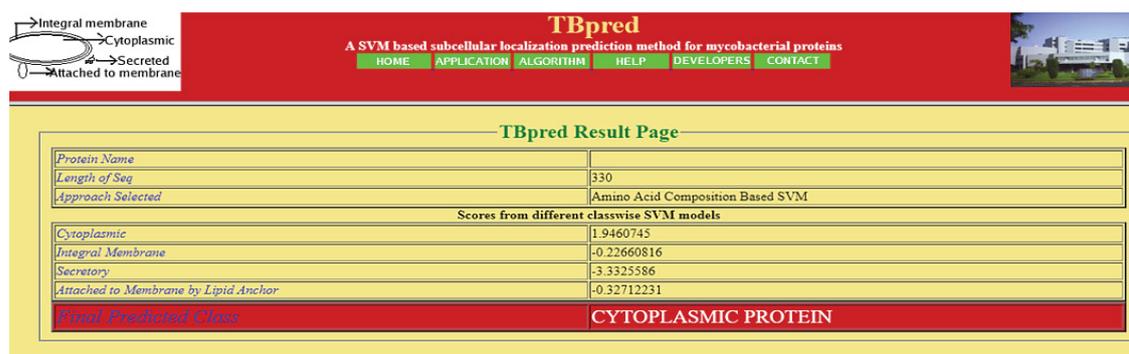


**Supplementary Figure 3: BCPRED and PROPREL Prediction.**

BCPRED and PROPREL server predicts B-cell epitopes and MHC Class II in targeted proteins where (A) Figure shows presence of B-cell epitopes in targeted protein by BCPRED server and (B) MHC class II binding peptides in targeted protein by using PROPREL server.

**Sub-cellular localization**

According to TBpred database, we predicted that Rv3228 is a cytoplasmic protein. It shows 1.9460 scores with the cytoplasmic region which is higher than the scores of different predicted classes such as integral membrane, secretory protein and attached to the membrane by lipid anchor (as shown in **supplementary Figure 4**).



**Supplementary Figure 4: TBpred: Localization Prediction.**

This figure shows Rv3228 localization where it shows that it is a cytoplasmic protein.

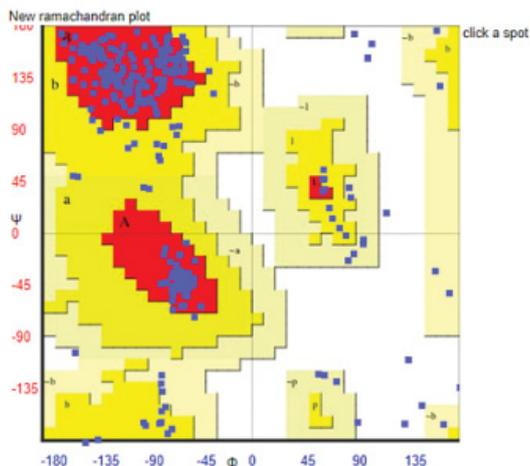
**Modelling**

Molecular modelling is the collection of procedures and methods to represent the biomolecules. The structure of Rv3228 was modelled by I-TASSER, RaptorX, Phyre2, and Lomets. In I-TASSER, quality of modelled protein determined by the percentage of the favourable region which lies above 90% of the value of C-score. Through I-TASSER, RaptorX, Phyre2, and Lomets, we create 12 models of the targeted protein and we finalized the top 5 models according to their C-score value (as shown in **Figure 1**).



**Model validation**

After modelling of structure, the protein structure was validated through and SAVES server (RAMPAGE, ERRAT, and Verify3D). The targeted protein was validated by RAMPAGE –Ramachandran plot investigation which is an online server. After examination of the Ramachandran plot of our proteins, the structure demonstrated that have been present in a favoured region. Although, other residues were laid in the allowed region and number of residues were laid in the outlier region. According to this 83.9% residues are in favoured region (A, B, L), 11.7% are in additional allowed region (a, b, l, p), 2.2% are in generously allowed region (~a, ~b, ~l, ~p) and 2.2% in disallowed region. These parameters of protein structure demonstrating that our displayed protein is of nearly good quality stable and adequate (as shown in **Figure 3**).



**Figure 3:** Model validation.

This figure shows Ramachandran plot for the protein Rv3228 which shows that 83.9% amino acids in targeted protein are present in favoured region and 11.7% are in additional allowed region.

ERRAT- an online server approves the protein structure on the principle of the nuclear connection between various sorts of atoms. The ERRAT analysis shows an overall quality factor of our model protein is good and satisfactory. The Verify3D approach calculates protein structure by utilizing 3D profiles. This program examines the similarity of a nuclear model (3D) with its own amino acid sequence which is 1 dimensional. According to this, 84.55% residues have average 3D-1D score  $\geq 0.2$  while cut off value to pass a model is 80% amino acids should have average score  $\geq 0.2$ . So this model is passed according to Varify3D (as shown in **Table 2**).

**Table 2:** Model validation.

Model no.	Saves model			Pro Q	
	Verifying 3D	ERRAT	RAMPAGE	LG Score	Max. Sub.
L1	89.39%	42.8105	83.90%	4.338	0.411
L5	84.55%	45.4861	89.70%	3.68	0.353
R	97.27%	63.0915	84.60%	4.152	0.426
P	84.55%	53.0945	83.90%	3.836	0.385
T1	97.27%	95.6386	65.90%	5.584	0.251
T2	99.39%	88.8199	64.10%	5.235	0.265
T3	99.70%	94.0994	67.40%	5.3682	0.212
T4	99.70%	94.0994	62.30%	5.504	0.261
T5	97.88%	93.1677	62.50%	5.034	0.247

This table shows model validation and protein quality prediction by saves (statistical analysis and verification server) and pro Q server Model evaluation by SAVES (Statistical Analysis and Verification Server).

**Structure-based function prediction**

We used COFACTOR online server for structure-based function prediction. COFACTOR is a protein-protein, structure and arrangements-based approach for natural analysing of protein particles. COFACTOR results predicted structural analogue in PDB, molecular capacity, biological process, cellular segment, enzyme homolog in PDB, and layout protein with comparative binding sites (as shown in **Figures 4A- 4C and Tables 3A and 3B**).

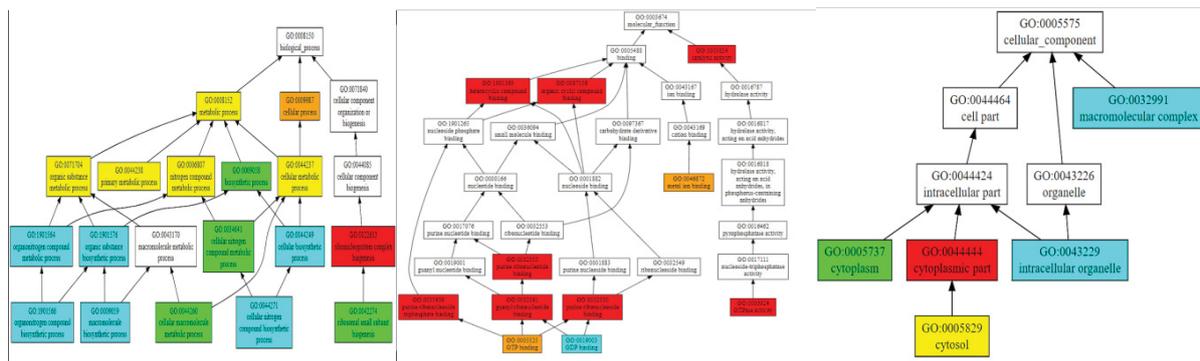


Figure 4: Function Prediction by COFACTOR server.

This figure shows function prediction by COFACTOR server where (A) shows involvement of Rv3228 in biological process, (B) shows molecular function of Rv3228, and (C) shows involvement of Rv3228 with cellular components.

Table 3A: Molecular function prediction by COFACTOR.

GO Term	C Score <sup>GO</sup>	Name
GO:0003824	1	Catalytic activity
GO:1901363	0.98	Heterocyclic compound binding
GO:0097159	0.98	Organic cyclic compound binding
GO:0003924	0.97	GTPase activity
GO:0032555	0.96	Purine ribonucleotide binding
GO:0032550	0.96	Purine ribonucleoside binding
GO:0032561	0.94	Guanyl ribonucleotide binding
GO:0035639	0.91	Purine ribonucleoside triphosphate binding
GO:0005525	0.88	GTP binding
GO:0046872	0.87	Metal ion binding

This table shows molecular function prediction by COFACTOR through C-SCOREGO values in range between [0-1], where higher value shows a better confidence in predicting the function.

Table 3B: Biological process prediction study.

GO term	C Score <sup>GO</sup>	Name
GO:0022613	1	Ribonucleoprotein complex biogenesis
GO:0009987	0.82	Cellular process
GO:0008152	0.79	Metabolic process
GO:0071704	0.77	Organic substance metabolic process
GO:0044237	0.76	Cellular metabolic process
GO:0006807	0.73	Nitrogen compound metabolic process
GO:0044238	0.72	Primary metabolic process

This table shows the predicted terms within the Gene Ontology for Biological Process on the basis of C score GO. Cofactor server enlist biological process with C-score GO, it is the confidence score of predicted GO terms. C-score GO values range in between [0-1]; where a higher value indicates a better confidence in predicting the function using the template.

C-score GO is the confidence score of predicted GO terms. COFACTOR results in the estimated quality metaphysics GO (gene ontology) expressions which are organized by atomic capacity, organic process, and cell part with a definite C-score. VICMpred server predicts the class of protein whether it is involved in cellular process, metabolism and signalling or virulence factor. According to this, score value 2.2138826 for Rv3228 which indicate that this gene is involved in virulence (as shown in **Supplementary Figure 6**).

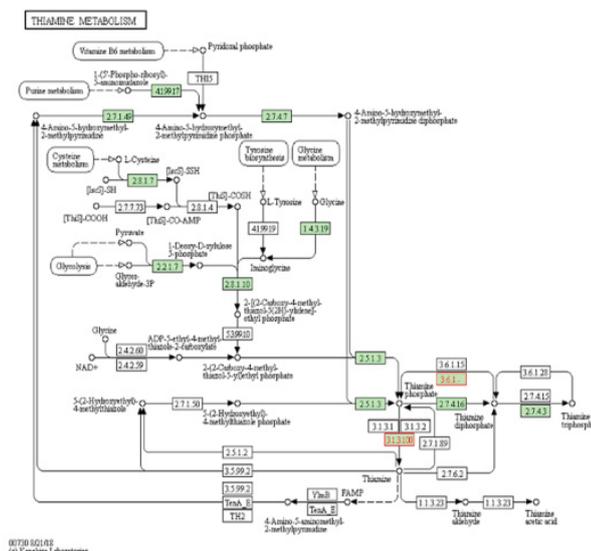
Score of Different Functional Class	
Function	Score
cellular Process	-3.483447
Information Molecule	0.55373422
Metabolism	-2.3297038
Virulence factors	2.2138826

**Predicted Functional Class**  
**Virulence factors**

**Supplementary Figure 6:** VICMpred prediction for targeted protein.

This figure shows that Rv3228 is a virulence factor with 2.213886 score value.

KEGG pathway set a map view of the interaction of molecules with different molecules of involvement in a cellular process. According to KEGG pathway, our protein Rv3228 is involved in thiamine metabolism (as shown in **Figure 5**).



**Figure 5:** KEGG PATHWAY for protein Rv3228.

KEGG Pathway: This figure shows Rv3228 protein involved in Thymine metabolism pathway.

**Mutation Analysis**

We purpose to contribute a fully unique methodology for predicting variations in stability upon point mutation in proteins referred to as maestro. Maestro is structure-based and differentiates itself from similar methods in the following points: (i) maestro implements a multi-tasker machine learning system. (ii) It also provides predictable free energy alteration ( $\Delta\Delta G$ ) values and conforming prediction confidence estimation. (iii) It delivers high throughput scanning for multi-point mutations where sites and types of mutation can be comprehensively controlled (iv) Finally, the software provides a particular mode for the prediction of stabilizing disulfide bonds. According to maestro server, mutation at D81M, D81L, D81V, G84M, G84H, G204A, H205M, H205L, S206V, G207A, G209H, T259M, T259V, and G261H is decreasing the stability of protein as the ddG value of these protein is zero and Cpred value is between 0-1 as zero is not reliable and 1 is highly reliable value. These mutations also compared by I-Mutant server and similar results were found that these mutations were decreasing the stability of the protein (as shown in **Table 4**).

**Table 4:** Mutation analysis.

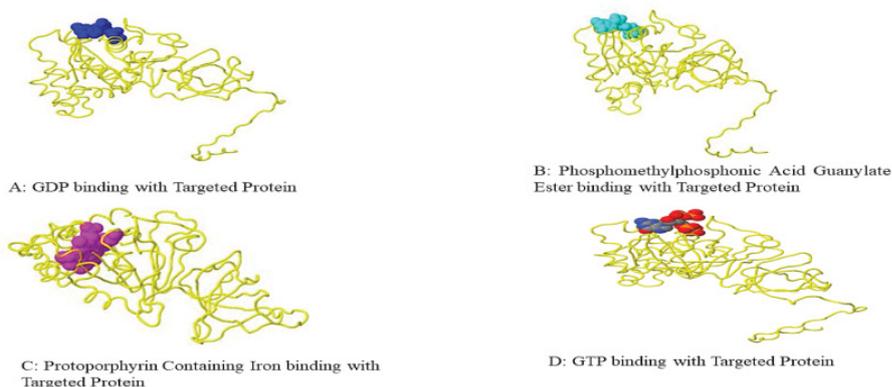
S. No.	Substitution	ddG-pred	C-pred	I-Mutant (ddG Kcal/mol)	Stability after mutation
		MAESTROweb server	MAESTROweb server		
1	D81{M}	0.365330513	0.803181791	-1.42	Decrease

2	D81{L}	0.337906721	0.76547246	-1.44	Decrease
3	G84{M}	0.167923675	0.899949101	-1.28	Decrease
4	G84{H}	0.04949247	0.910594466	-1.91	Decrease
5	G204{A}	0.334202773	0.882327346	-2.33	Decrease
6	H205{M}	0.127230538	0.91968627	-0.3	Decrease
7	H205{L}	0.217314855	0.910067747	-0.76	Decrease
8	G207{A}	0.013449798	0.829182347	-2.48	Decrease
10	G209{H}	0.355621578	0.953610612	-1.21	Decrease
11	T259{M}	0.278813446	0.947244749	-0.44	Decrease
12	T259{V}	0.349927504	0.876326772	-0.35	Decrease
13	G261{H}	0.285719592	0.961099119	-2.02	Decrease

This table is showing results of two servers, one is MAESTROweb server and second is I-MUTANT. Both the servers are predicting decrease in stability of protein after mutation in wild type protein.

**Ligand Binding Prediction**

ProBiS is a web server which measures the frequency of occurrence of particular residues. According to this server, our targeted protein has a specific and nonspecific binding site for Guanosinen-5', 3' tetraphosphate, Guanosine 5' Diphosphate with higher score. Sulfate ions are also having binding property but they are showing non-specific binding. There are some other molecules that are predicted for specific binding (as shown in **Figure 6** and **Table 5**).



**Figure 6:** ProBiS: Ligand binding prediction.

This figure is showing ligand binding with the protein Rv3228. Here it predicts that GDP, Phosphomethylphosphonic acid Guanylate Ester, Protoporphyrin containing iron and GTP are having specific binding with targeted protein.

**Table 5:** Probis ligand binding prediction.

S.No.	Name	Source	Confinder	Binder
1	Guanosine-5',3'-Tetraphosphate	1LNZ	2.09	Specific
2	Guanosine-5'-DiPhosphate	2GF9	2.03	Specific
3	Protoporphyrin IX Containing FE	1GCW	1.91	Specific
4	Protoporphyrin IX Containing FE	1GCW	1.91	Specific
5	Protoporphyrin IX Containing FE	1GCV	1.91	Specific
6	Guanosine-5'-DiPhosphate	3W5J	1.91	Specific
7	Sulfate Ion	3W5J	1.91	Non-Specific
8	Sulfate Ion	3W5I	1.91	Non-Specific
9	Guanosine-5'-Di Phosphate	3GEE	1.83	Specific
10	PhosphomethylPhosphonic Acid Guanylate Ester	3GEI	1.83	Specific
11	PhosphomethylPhosphonic Acid Guanylate Ester	3GEI	1.83	Specific
12	Adenosine-5'-DiPhosphate	3THO	1.75	Specific
13	PhosphoaminoPhosphonic Acid-Adenylate Ester	4W9M	1.75	Specific

This table is predicting ligand binding with the targeted protein Rv3228 by PROBIS server. Here it predicts that GDP, Phosphomethylphosphonic acid Guanylate Ester, Protoporphyrin containing iron and GTP are having specific binding with targeted protein.

## CONCLUSION

In the present circumstance, we are able to see that there no defensive and curative treatment to destroy TB completely apart from BCG. Past a few years of analysis as of currently demonstrates that BCG provides forced insurance against tuberculosis nonetheless fails in securing MDR, TDR and XDR instances of Tuberculosis<sup>[44,45]</sup>. There is a persistent effort has been placed by researchers with the end goal to make the adequacy of the antibody and in searching for new medication targets. GTP binding genes initiate to be as novel targets of treatment of this disease. In this manuscript, we highlighted the Rv3228 gene of *M. tuberculosis*. Afterwards, computational examination of this gene is 34 kDa proteins by Mycobrowser database<sup>[46]</sup>. Mutation analysis has shown a decrease in stability of protein and ligand binding prediction has suggested some molecules that could be used for the drug development against tuberculosis.

## ACKNOWLEDGEMENTS

The author acknowledges financial support from the Department of Science and Technology-SERB, Council of Scientific and Industrial Research-Institute of Genomics and Integrative Biology under the research project GAP0145.

## REFERENCES

1. Meena S, et al. Interaction of Mycobacterium tuberculosis H37Rv with Microfold Cell leads to a New Era of Infection in Host. *Ann Clin Lab Res.* 2018;6:246.
2. Dolin PJ, et al. Global tuberculosis incidence and mortality during 1990-2000. *Bull World Health Organ.* 1994;72:213-220.
3. O'Toole RF, et al. Limitations of the Mycobacterium tuberculosis reference genome H37Rv in the detection of virulence-related loci. *Genomics.* 2017;109:471-474.
4. Meena LS, et al. Survival mechanisms of pathogenic Mycobacterium tuberculosis H37Rv. *FEBS J.* 2010;277:2416-2427.
5. Barry CE, et al. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat Rev Mol Cell Biol.* 2009;7:845-855.
6. Russell DG. Mycobacterium tuberculosis: here today, and here tomorrow. *Nat Rev Mol Cell Biol.* 2001;2:569-586.
7. Gagneux S. Ecology and evolution of Mycobacterium tuberculosis. *Nat Rev Microbiol.* 2018;16:202-213.
8. McDonough KA, et al. Pathogenesis of tuberculosis: Interaction of Mycobacterium tuberculosis with macrophages. *Infect Immun.* 1993;61:2763-2773.
9. Arruda S, et al. Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. *Science.* 1993;261:1454-1457.
10. Cui ZJ, et al. Bioinformatics Identification of Drug Resistance-Associated Gene Pairs in Mycobacterium tuberculosis. *Int J Mol Sci.* 2016;17:1417.
11. Kumari P, et al. Factors Affecting Susceptibility to Mycobacterium tuberculosis: A Close View of Immunological Defence Mechanism. *Appl Biochem Biotechnol.* 2014;174:2663-2673.
12. Akif M, et al. Conformational flexibility of Mycobacterium tuberculosis thioredoxin reductase: crystal structure and normal-mode analysis. *Acta Crystallogr D Biol Crystallogr.* 2005;61:1603-1611.
13. Espinal MA, et al. Global trends in resistance to antituberculosis drugs. *N Engl J Med.* 2001; 344:1294-1303.
14. Masjedi MR, et al. Extensively drug-resistant tuberculosis: 2 years of surveillance in Iran. *Clin Infect Dis.* 2006;43:840-847.
15. Global Report Tuberculosis. World Health Organisation. 2017.
16. STRING (2019) Protein-Protein Interaction Networks Functional Enrichment Analysis.
17. Rajni, et al. Guanosine triphosphatases as novel therapeutic targets in tuberculosis. *Inter J Infect Dis.* 2010;14:682-687.
18. Kapopoulou A, et al. The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis (Edinb).* 2011;91:8-13.
19. Madeira F, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 2019;47: 636-641.

20. Li W, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res* 2015;43:580-584.
21. Sharma D, et al. Secretary Proteome Analysis of Streptomycin-Resistant Mycobacterium tuberculosis Clinical Isolates. *SLAS DISCOVERY: Advancing Life Sciences R&D*. 2017;22:1229-1238.
22. Lewis AC, et al. Predicting protein-protein interactions in the context of protein evolution. *Molecular Biosystem*. 2010;6:55-64.
23. Rashid M, et al. Support Vector Machine-based method for predicting subcellular localization of mycobacterial proteins using evolutionary information and motifs. *BMC Bioinformatics*. 2007; 8:337.
24. Monu M, et al. Biochemical characterization of PE\_PGRS61 family protein of Mycobacterium tuberculosis H37Rv reveals the binding ability to fibronectin. *Iran J Basic Med Sci*. 2016;19:1105-1113.
25. Zhang C, et al. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucleic Acids Res*. 2017;45:291-299.
26. Roy A, et al. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res*. 2012;40:471-477.
27. Roy A, et al. Recognizing protein-ligand binding sites by global structural alignment and local geometry refinement. *Structure*. 2012;20:987-997.
28. Fabris F, et al. New KEGG pathway-based interpretable features for classifying ageing-related mouse proteins. *Bioinformatics*. 2016;32:2988-2995.
29. Stular T, et al. Discovery of Mycobacterium tuberculosis InhA Inhibitors by Binding Sites Comparison and Ligands Prediction. *J Med Chem*. 2016;59:11069-11078.
30. Wu ST, et al. LOMETS: A local meta-threading-server for protein structure prediction. *Nucl Acids Res*. 2007;35:3375-3382.
31. Yang J, et al. The I-TASSER Suite: protein structure and function prediction. *Nat Methods*. 2015;12:7-8.
32. Roy A, et al. I-TASSER: a unified platform for automated protein and function prediction. *Nat Protoc*. 2010;5:725-738.
33. Ma J, et al. Protein threading using context-specific alignment potential. *Bioinformatics*. 2013;29:257-265.
34. Peng J, et al. RaptorX: exploiting structure information for protein alignment by statistical inference. *Proteins*. 2011;79:161-171.
35. Dakal TC, et al. Structural modelling of human organic cation transporters. *Comput Biol Chem*. 2017;68:153-163.
36. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*. 1999;292:195-202.
37. Lovell SC, et al. Structure validation by C $\alpha$  geometry: phi, psi and C beta deviation. *Proteins*. 2003;50:437-450.
38. Ho BK, et al. The Ramachandran plots of glycine and pre-proline. *BMC Structural Biology*. 2005;5:1-11.
39. Lundstrom J, et al. Pcons: a neural-network-based consensus predictor that improves fold recognition. *Protein Sci*. 2001;10:2354-2362.
40. Wallner B, et al. Can correct protein models be identified? *Protein Sci*. 2003;12:1073-1086.
41. Laimer J, et al. MAESTROweb: a web server for structure-based protein stability prediction. *Bioinformatics*. 2016;32:1414-1416.
42. Capriotti E, et al. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res*. 2005;33:306-310.
43. Konc J, et al. ProBiS-CHARMMing: Web Interface for Prediction and Optimization of Ligands in Protein Binding Sites. *J Chem Inf Model*. 2015;55:2308-2314.
44. Shaw S, et al. GTPases: Prerequisite Molecular Target in Virulence and Survival of Mycobacterium tuberculosis. *Int J Mol Biol*. 2016;1:26-27.
45. Meena LS. GTPases: a significant signalling molecule in TB infection. *Int J Biotech Bioeng*. 2018;4:124-129.
46. Kapopoulou A, et al. The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis*. 2011;91:8-13.