SIGNIFICNCE ANALYSIS OF TARGET PROFILE IN TUBERCULOSIS USING GENE INTERACTIONS AND INSILICO DOCKING APPROACH TO FIND POTENTIAL LIGANDS

Dr.Manjunath, Shambhu M.G Apeksha,Chandana N,Lakshmi S,Mounika S.
Department of Biotechnology, The Oxford College of Engineering, Bengaluru-560068
*correspondence should be addressed to Mr Shambhu M G :*shambhumg13@gmail.com*

Abstract: Tuberculosis is a common and deadly infectious disease caused by Mycobacteria. WHO estimates that one third of global population is infected with *Mycobacterium tuberculosis.* Tuberculosis is a Multidrug resistant since their are a lot of mutations occur in genes .Our study focused on uncharacterized mutated tuberculosis target identification by using Systems Biology Approach, which is used to find the better drug targets. From the advance search by using UniProt we observed seven receptors are the significant targets 2CCA,1P44, 3VZ1,3IFZ, 1KOR,2EYQ .Tuberculosis target candidates are screened and validated by docking studies using Auto dock Software .Analysis has revealed that among 14 ligands it has been observed that Josamycine and Rifapentine showed a better interaction score (-10.3 kcal/mol and -12.7 kcal/mol) and could be potential ligands. These potential ligands have also shown better ADMET properties in *Insilico* studies by using ADMET SAR software.

1. INTRODUCTION

Tuberculosis (TB) is caused by *Mycobacterium tuberculosis.* TB is an infectious disease that usually affects the lungs .Some strains of the TB bacteria developed resistance to the standard drugs through genetic changes [2].TB affects 24% of world's total population .According to WHO its world's top infectious disease, about 5000 people deaths occurs everyday .*Mycobacterium tuberculosis* (MTB) is a rod shaped bacteria that can thrive only human beings[3] .TB is often called Multidrug resistance(MDR).The TB bacteria has natural defences against some drugs, and can acquire drug resistance through genetic mutations. The bacteria does not have the ability to transfer genes for resistance between organisms through plasmids some mechanisms of drug resistance include:**Cell wall**: The cell wall of *M. tuberculosis* (TB) contains complex lipid molecules which act as a barrier to stop drugs from entering the cell [9].**Drug modifying & inactivating enzymes :**The TB genome codes for enzymes (proteins) that inactivate drug molecules. These enzymes usually phosphorylate, acetylate, or adenylate drug compounds.**Drug efflux systems:** The TB cell contains molecular systems that actively pump drug molecules out of the cell.**Mutations**: Spontaneous mutations in the TB genome can alter proteins which are the target of drugs[6], making the bacteria drug resistant.[1]In the present study , Initially the receptors are obtained from the target pathogen database 30 receptors were obtained from the literature survey .Among the 30 receptors, 7 potential genes were obtained by string software .Validation was performed using Rampage software. Homology model was performed for all seven receptors .Among the seven receptors , only two receptors model was not found .These two receptors model was built using Swiss model and validated using Rampage and ERATT tool.Fourteen potential ligands were obtained from drug bank . These ligands satisfied and passed lipnsiki's rule which were performed to obtain better drug candidates.Interaction studies between the receptors and ligands were observed by docking studies by using Autodockvina software.It was observed that among the 14 ligands , two ligands showed good interaction studies they are Rifapentine and josamycin .Insilico ADMET properties predictions was performed using ADMET SAR SOFTWARE .

2. MATERIALS AND METHODS

2.1 IDENTIFICATION AND PREPARTION OF TARGET PROTEIN

Target identification is the process of identifying the direct molecular target for example protein or nucleic acid of a small molecule. In clinical pharmacology, target identification is aimed at finding the efficacy target of a drug/pharmaceutical .Target proteins are functional biomolecules that are addressed and controlled by biologically active compounds [1] .Target proteins control the action and the kinetic behaviour of drugs within the organism.Initially by using database Target pathogen and literature review 4000 genes was obtained which occurred in Tuberculosis[1].These 4000 genes was further analyzed and screened by using the string software . By using string software on the basis of protein-protein interaction 30 potential genes were screened based on their functions and characteristics using UniProt advance filters.

2.2 GENES NETWORK STUDIES

Using string database these 30 genes were screened based on their functions and characteristics using UniProt advance filters .Seven significant genes involved in tuberculosis were shortlisted among 30 targets. All the 7 targets were further studied based on structural information[15]. Among these seven targets we observed that 2 targets did not have the structure which was further modelled using Swiss model.

2.3 HOMOLOGY MODELLING

The Homology Modelling server template library ExPDB is extracted from the PDB. To select templates for a given protein, the sequences of the template structure library are searched. If these templates cover distinct regions of the target sequence, the modeling process will be split into separate independent batches.Homology modelling was used for the construction of atomic resolution model of the target protein. Swiss model was used to obtain 3d protein structure models for the genes which we have selected for finding the better drug candidate.The template protein of gabD1 is 3VZL and of mfd is 2EYQ. For two protein the model was built based on template as the structure was not available and this was done using Swiss model.

2.4 VALIDATION

The Ramachandran plot has been the mainstay of protein structure validation for many years.Its detailed structure has been continually analysed and refined as more and more experimentally determined models of protein 3D structures have become available, particularly at high and ultra-high resolution. These plots are typically split in forbidden and allowed regions. Around 40% of all the amino acids in a structure are contained in just the 2% of the Ramachandran plot the so called “allowed areas . Rampage revealed the information of the dihedral angles of residues with respect to protein structures. Ramachandran plot was analysed for the 2 protein models by giving the pdb format.Validation was done using the ERRAT tool and this tool aanalyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function versus position of a 9-residue sliding window, calculated by a comparison with statistics from highly refined structures.We uploaded the pdb file of modeled protein and we obtained a graph which specifies the error % and the warning %. These graphs were used for the validation process.

2.5 SCREENING OF LIGANDS

The ligands were collected from the DRUGBANK and advance search. The ligands were screened based on the Lipinski’s rule whichstates that poor adsorption is anticipated , if the molecular weight is greater than 500 LogP is greater than 5 and hydrogen bond accepters is greater the all the ligands should satisfies Lipinski’s rule and also indicates good drug candidates.In our study 14 potential ligands were screened[17].

2.6 DOCKING

AutoDockVina, a new program for molecular docking and virtual screening, is presented. AutoDockVin a significantly improves the accuracy of the binding mode predictions, than the Autodock 4.Six targets were docked with the selected 14 ligands.The best interaction is taken based on the score given by autodockvina. The general functional form of the conformation-dependent part of the scoring function AutoDockVina (referred to as Vina here) is designed to work with is[7] ,

$$c = \sum_{i<j} f_{t_i t_j}(r_{ij}),$$

Where the summation is over all of the pairs of atoms that can move relative to each other, normally excluding 1–4 interactions, *i.e.* atoms separated by 3 consecutive covalent bonds. Here, each atom i is assigned a type t_i, and a symmetric set of interaction functions f_{titj} of the interatomic distance r_{ij} should be defined.

c= c inter + c intra

This value can be seen as a sum of intermolecular and intermolecular contributions[7]. The optimization algorithm, described in the following section, attempts to find the global minimum of c and other low-scoring conformations, which it then ranks.

2.7 ADMET (Absorption, Distribution, Metabolism, Excretion, and toxicity) Test

ADMET stands for Adsorption, Distribution, Metabolism, Excretion,Toxicity. To select drug-like molecule, ADMET SAR software was used to screen the selected five molecules based on filters namely Lipinski's rule[16],Quantitative Estimate of Drug likeness. The selected compounds in SDF format was given to the ADMET software interface and proceeded to calculate the properties.

3. RESULTS AND DISCUSSIONS

3.1 IDENTIFICATION OF TARGET PROTEIN

After the characterization of genes we obtained 1373 unknown genes. Analysis of genes was carried out based on their functional characteristics. Among the 30 targets, 7 targets have to be further analyzed through network analysis. The mapping of genes was carried out using UniProt ID mapping. The several specific genes were obtained as such when provided with identifiers from UniProtKB AC/ID to PDB.

3.2 Gene interaction network analysis

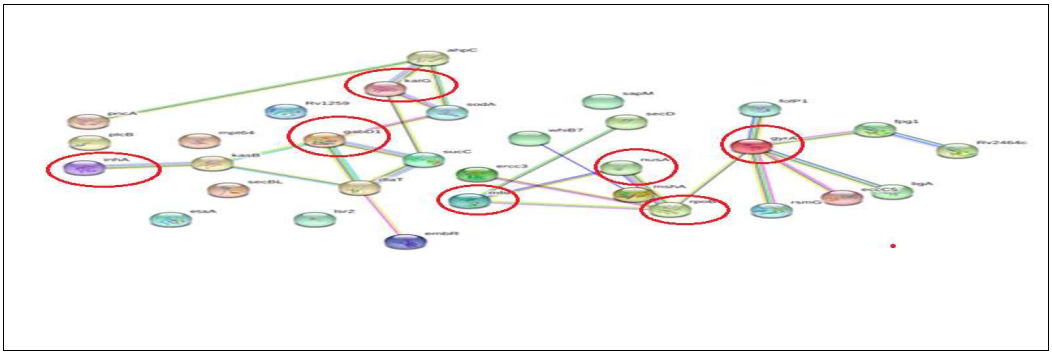
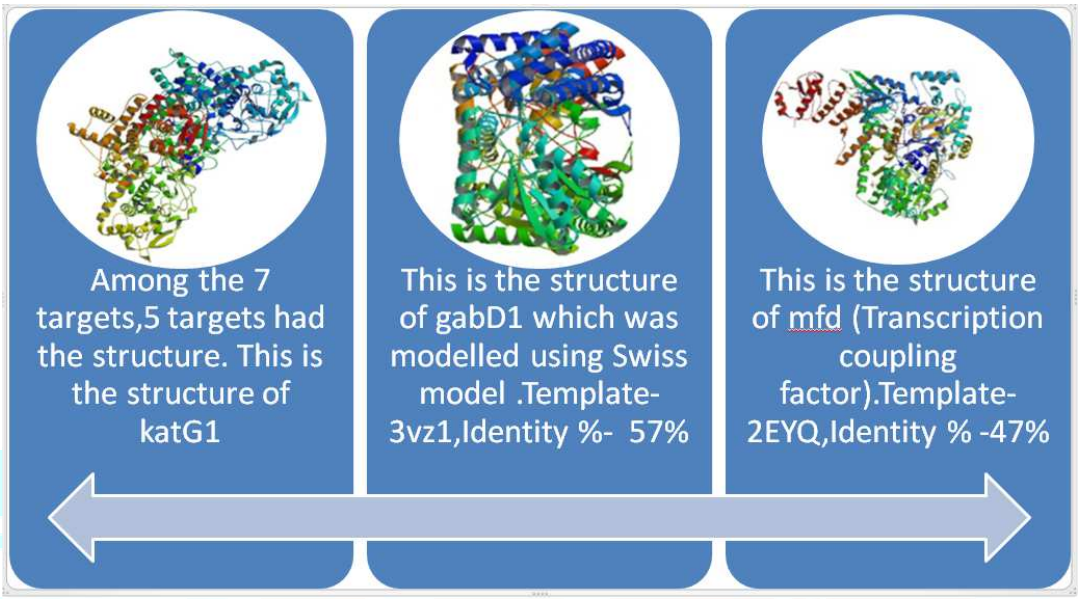


Fig 2: protein-protein interactions where done using the string database and 7 targets were found to have better interactions and those genes were inhA,katG,r

The 30 Target protein which were obtained from the advance search n the characterized genes, these were given to string software and using the string software we found out the gene-gene interactions and among the 30 genes we found that 7 genes had the better drug interactions and based on these interactions the 7 genes were shortlisted and those are the inhA,katG,rpoB,gabD1,mfd,gyrA,nusA. Among these 7 genes 5[22] had the structures and 2 did not have the structures which were modelled using the homology modeling.

3.2 HOMOLOGY MODELLING



In the homology modelling we used the swiss model software and the 2 genes which did not have the structure was built a model and the template was found for the these genes and we obtained the structure.

3.3 VALIDATION

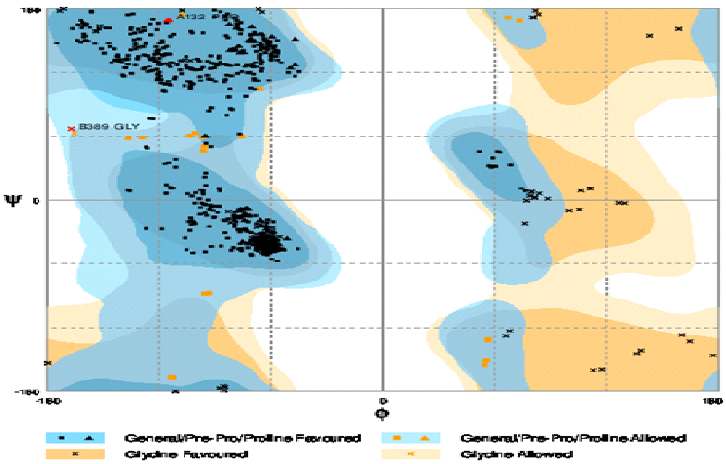


FIGURE 8:
Number of residues in favoured region (~98.0% expected) : 881 (97.5%)
Number of residues in allowed region (~2.0% expected) : 21 (2.3%)

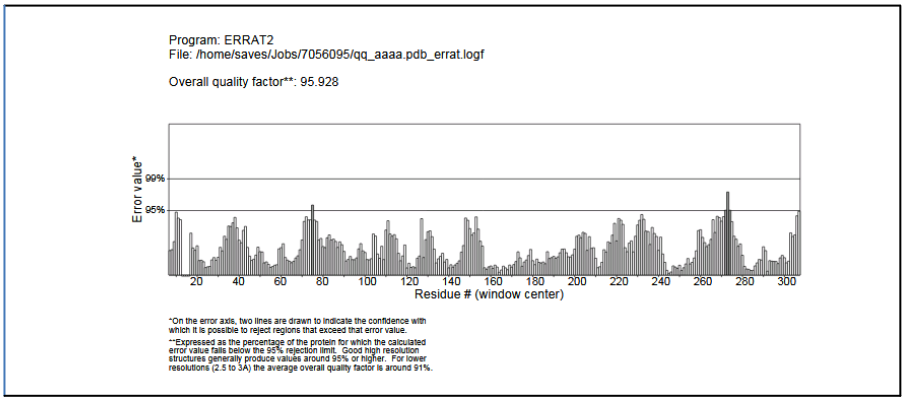


Fig 7 : Validation of molecules using ERRAT tool

3. 4 DOCKING

Autodockvina was performed to predict the bound conformation the binding affinity[23].The grid maps will be automatically formed by the software.The configuration values will be saved in a text file called conf.The PDBQT file of target and the ligand was obtained.

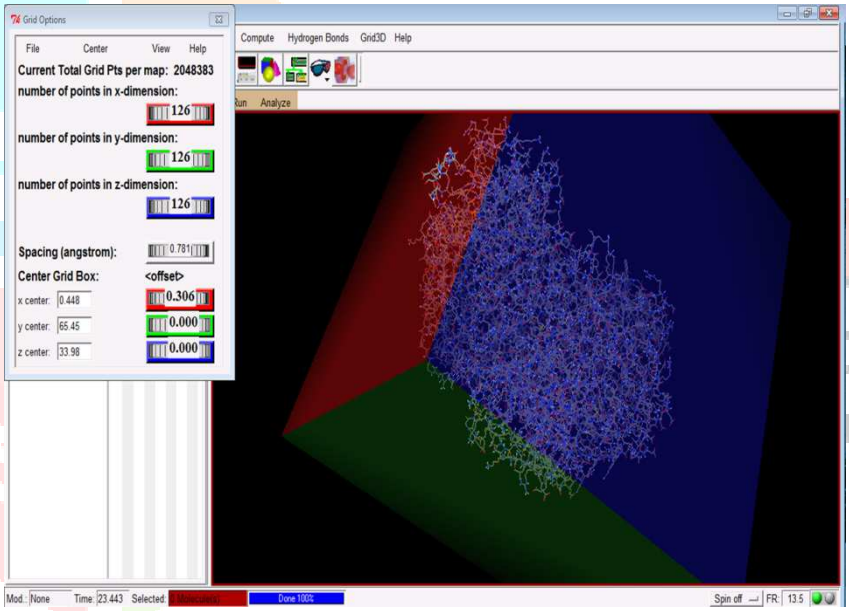


Fig8:Grid box formed by the Autodock for 1KOR.

Similarly the grid box and the configuration was done for other genes.14 ligands were taken for docking purpose, where Josamycine and Rifapentine showed the better result compared to other ligands.

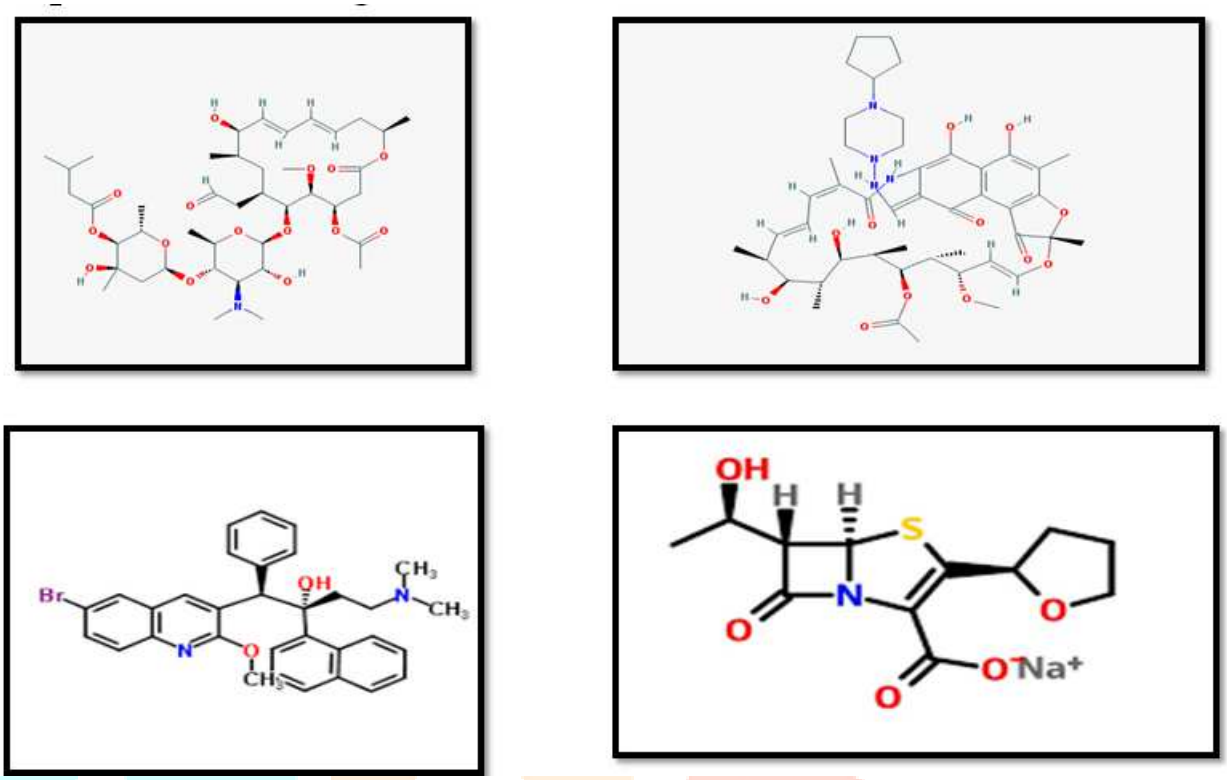


Fig9 : Ligands structure a.Josamycin, b.Rifapentine, c.Bedaquiline, d.Faropenem

Further the command prompt was used to run the program, where the conffile ,gene and ligand pdbqt file was saved in one folder.Further the docking analysis is performed based on the binding energy value and the interaction was analysed using pymol software.

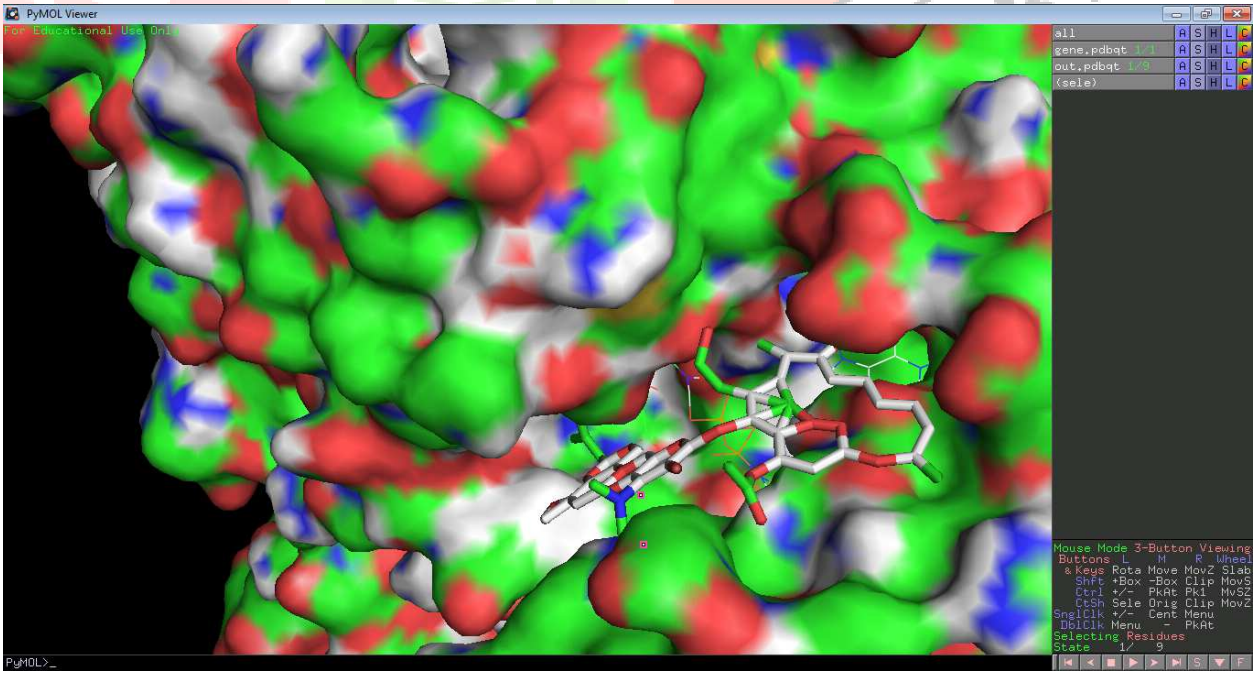


Fig10: Docking result of Josamycin which is docked with the 1KOR

Rifapentine was shown the better result for the gene 2CCA.

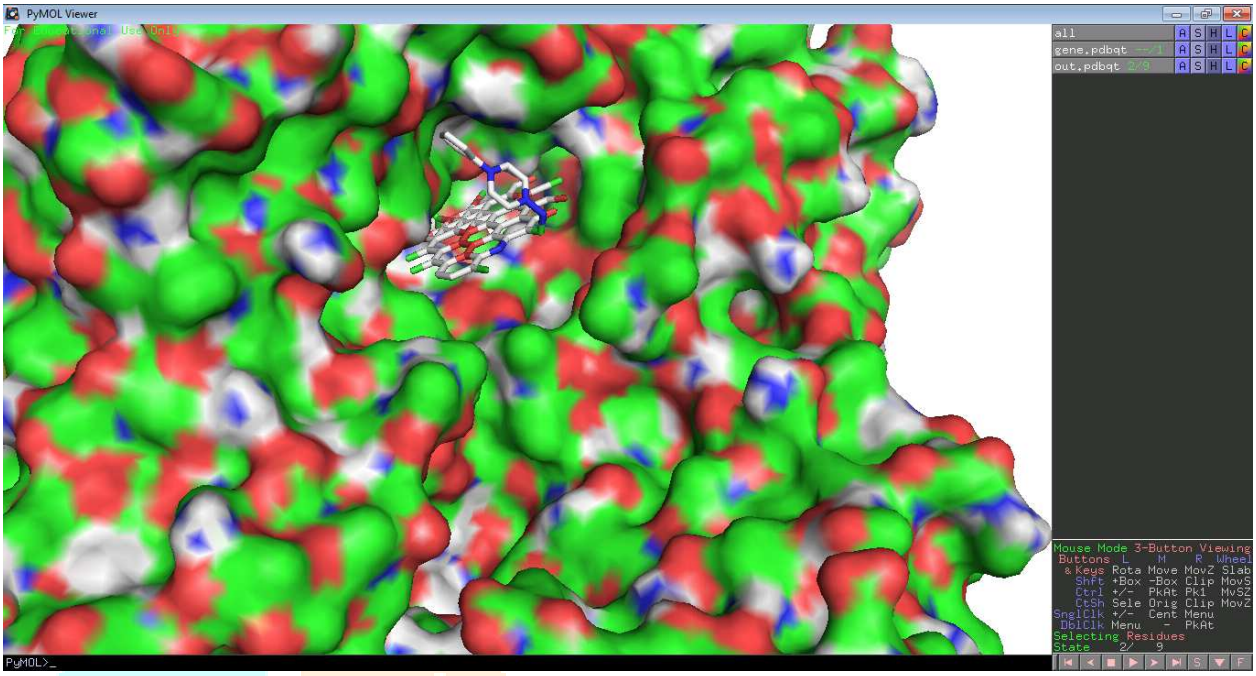


Fig15: 2CCA was docked with the ligand Rifapentine.

Docking was performed between the 6 genes and the listed ligands using the autodockvina. The software will predict the proper binding site. Further the analysis of the result was done based on the values.

	LIGANDS	2CCA	1P44	3VZ1	3IFZ	1KOR	2EYQ
2	5 Amino-1,3,4,thiadiazole -2 thiol	-4.6	-3.7	-4.4	-3.7	-4	-4.1
3	Bedaquiline	-9.8	-8.2	-8.2	-9.9	-9.3	-8.7
4	Delamanid	-8	-9	-9.7	-9.5	-8	-9.2
5	Faropenem	-8.1	-6.6	-6.1	-6.3	-7.1	-7
6	Isoniazid	-5.8	-4.7	-4.9	-4.3	-4.7	-4.7
7	Josamycin	-9.2	-9	-9.1	-9.3	-10.3	-10.1
8	KanamycinA	-8.1	-9.2	-7.4	-7.5	-8.2	-9.1
9	Levofloxacin	-7.4	-8.9	-6.8	-7.1	-7.3	-7.7
10	Pretomanid	-7.2	-7.8	-7.4	-7.3	-7.9	-8.2
11	Protionamide	-5.2	-7	-5.2	-5	-5.3	-5.2
12	Pyrazinamide	-5.1	-4.8	-5.3	-4.2	-4.8	-4.6
13	Rifapentine	-12.7	-10.9	-11.8	-12.4	-11.8	-13.1
14	SQ-109	-5	-5.1	-6.2	-5.1	-6.7	-6.8
15	Terizidone	-7.2	-7	-6.6	-6.4	-8	-6.7

3.5 ADMET PROPERTIES

SL.NO	LIGAND NAME	PROPERTY	MODEL	RESULT	PROBABLIITY
1	REFAPENTINE	ABSORPTION	Blood-Brain Barrier	BBB-	0.9659
			Human Intestinal Absorption	HIA+	0.66848
		DISTRIBUTION	Subcellular localisation	Mitochondria	0.5477
		METABOLISM	CYP450IA2 Inhibitor	Non-Inhibitor	0.8865
		TOXICITY	Carcinogens	Non-carcinogens	0.8147
2	JOSAMYCINE	ABSORPTION	Blood-Brain Barrier	BBB-	0.9659
			Human Intestinal Absorption	HIA+	0.5235
		DISTRIBUTION	Subcellular localisation	Mitochondria	0.5110
		METABOLISM	CYP450IA2 Inhibitor	Non-Inhibitor	0.9070
		TOXICITY	Carcinogens	Non-carcinogens	0.9287

ADMET properties of Rifapentine and Josamycine.

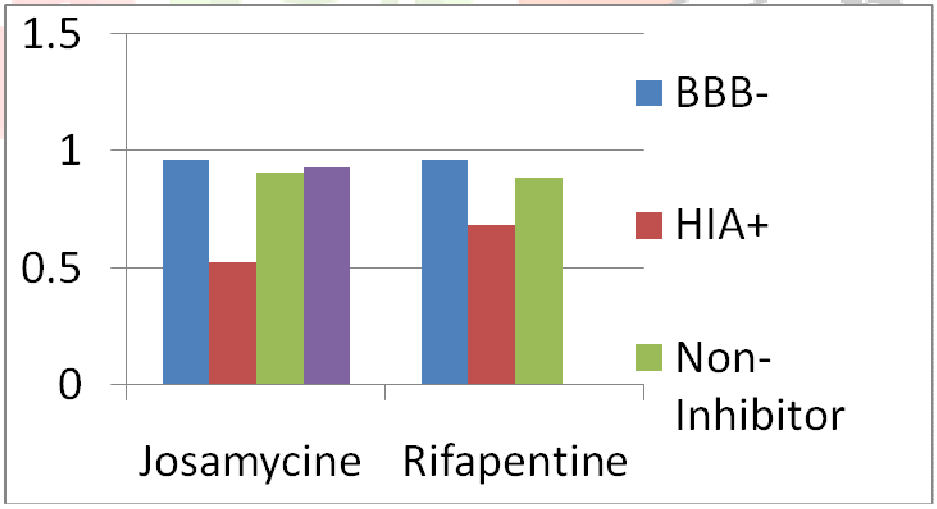


Fig16 : Result of ADMET properties.

4. CONCLUSION

Tuberculosis (TB) is caused by *Mycobacterium tuberculosis*. TB is an infectious disease that usually affects the lungs. Some strains of the TB bacteria developed resistance to the standard drugs through genetic changes and *Mycobacterium tuberculosis* (MTB) is a rod shaped bacteria that can thrive only human beings .TB is often called Multidrug resistance(MDR) or Multi drug resistance is antimicrobial resistance shown by a species of microorganisms to multiple anti-microbial drugs .MDR is most threatening to public health. MDR bacteria that is resist multiple antibiotics

A dataset of genes was reviewed using the TARGET PATHOGEN database where the characterization of the genes were carried out such that separate the characterized genes and uncharacterized genes and we concentrated on the un characterized genes for our project and among the 4000 genes we obtained 1373 characterized gens and 2627 of characterized genes.

These genes were further shortlisted to 30 genes based on their functional characteristics which were suitable for the *Mycobacterium tuberculosis*. Further to know the gene-gene interactions of these 30 genes, these genes were submitted to gene interaction analysis.

In the gene network analysis we used the STRING software to find out the gene-gene interaction where we submitted 30 genes to STRING and gene network was formed from which 6 genes were shortlisted because they had better drug interactions.

Modelling of the protein was carried out using SWISS MODEL .Among the seven genes which had better interactions, two genes did not have the structure so they were modelled using the SWISS MODEL and the template was obtained for these 2 targets.

Validation of these structures obtained from SWISS MODEL was carried out using RAMPAGE and ERRAT tool was also used for the validation of the structures.RAMPAGE showed the allowed regions and favourable regions based on which the modelled structures were validated and in ERAT tool the percentage of error and the warning percentage were given and 99% of the residues were below the threshold and 1% of which were above the threshold. Thus the modelled structures were validated.

Further the docking studies were carried out using AUTODOCK VINA software and the docking results showed that two ligands had better interaction score and those were REFAPENTINE (-13.1 kcal/mol) and JOSAMYCINE (- 10.1 kcal/mol).

ADMET properties were studied using ADMET SAR software and these two ligands also showed better properties compared to other ligands and these ligands are non-carcinogenic and non-toxic.

Hence we conclude that based on the docking studies and ADMET properties ,two ligands REFAPENTINE (-13.1 kcal/mol) and JOSAMYCINE (- 10.1 kcal/mol) have shown better interactions and these can be the potential drug molecules for Tuberculosis.

ACKNOWLEDGEMENT

The Authors would like to thank CCMC-VGST (under KFIST Level 1) Govt of Karnataka , Department of Biotechnology and the management TOCE Bengaluru for the support and Infrastructure provided and for their undiminished encouragement and valuable inputs for the project.

5.REFERENCE

1. Sosa EJ, Burguener G, Lanzarotti E Target-Pathogen: a structural bioinformatic approach to prioritize drug targets in pathogens , journal nucleic acids research. published in 2018.
2. Nisha T,nair,Shefin Nisthar identification of novel drug candidate against Mycobacterium Tuberculosis Inh A protein through computer aided drug discovery, journal, Indian journal of pharmaceutical education and research. published in 2016.
3. V.R.Bollela, E.I.Nambuurete, Detection of kat G and inh A mutations to guide isoniazid and ethionamid use of drug resistant tuberculosis and journal is HHS PUBLIC ACCESS published in 2016.
4. Marva Seifert, Donald Catanzaro, Genetic mutations associated with isoniazid resistance in mycobacterium tuberculosis and the journal is PLOS ONE published in 2015.
5. Danil V Zimenkov, Olga V Antonova, Detection of second line drug resistance in mycobacterium tuberculosis using oligonucleotides microarrays and journal is BMC INFECTIOUS DISEASES and it is published in 2013.
6. Wanil Kang, Yu Pang, Current status of new tuberculosis vaccine in children and the journal is HUMAN VACCINE IMMUNOTHERAPEUTICS and it is published in 2016.
7. Marva V.J, Autodock Vina:Improving the speed and accuracy of docking with scoring function and journal is HHS PUBLIC ACCESS published in 2011.
8. Charles C Wang, System approach to tuberculosis vaccine development and journal is RESPIROLOGY published in 2013.
9. Asad Amir, Khyti Rana, M TB H37 RV , In silico drug targets, Identification by metabolic pathway analysis and the journal is INTERNATIONAL JOURNAL OF EVOLUTIONARY BIOLOGY published in 2014.
10. Mustafa AS, In silico analysis and experimental variation of M TB specific protein and peptides M TB for immunological diagnosis and vaccine development, journal MEDICAL PRINCIPLE AND PRACTICE published in 2013.
11. Beban Kai Sheng Chung, Thomas dick, In silico analysis for discovery of tuberculosis drug targets, journal ANTIMICROBIAL CHEMOTHERAPY published in 2013.
12. Edivi W Tiwemersa, Natural history of tuberculosis duration and fatality of untreated pulmonary TB in HIV negative patients, journal PLOS ONE published in 2011.
13. Molebogeng X Rangaka, Controlling the seed beads of TB, Diagnosis and tereatment of TB infection, journal HSS PUBLIC ACCESS published in 2015.
14. Ragini Singh, Basanhi Ramachandran, In silico based high throughput screening for discovery of novel combinations of TB treatment, journal ANTIMICROBIAL AGENTS AND CHEMOTHERAPY published in 2015.

15. Dipendra Gurnung, String software,intelligent predictive string search algorithm, journal SCIENCE DIRECT published in 2016.

16. V.C Sheng, Comprehensive source and free tool for assessment of chemical ADMET properties, journal CHEMICAL INFORMATION AND MODELLING published in 2012.

17. David S Wishart, Comprehensive resource for in sillico drug discovery and exploration, journal NUCLEIC ACID RESEARCH published in 2006.