International Journal of Basic and Applied Biology
Print ISSN: 2394-5820, Online ISSN: 2349-2539, Volume 2, Number 3; October-December, 2014 pp. 130-136

http://www.krishisanskriti.org/ijbab.html

Drug Discovery against Unknown Function Protein Rv2623 from Mycobacterium Tuberculosis Via Molecular Docking and Dynamics

Corresponding Author Ashwani Sharma[1,#], Indu Cahturvedi[2], Chandrmohan Jana[3] and Ophelie Robine[4,#]

[1,2,3]New Era Proteomics Drug Discovery Research C-1/31, Yamuna Vihar New Delhi-110053, India
[2]Department of Animal Nutrition Central Institute for Research on Goats Makhdoom, Mathura-281 122, UP, India
[3]New Era Proteomics Drug Discovery Research C-1/31, Yamuna Vihar New Delhi-110053, India
Former: University of Rouen, Normainde France
[4]Current: New Era Proteomics Drug Discovery Research 56 Route de Darnetal, 76000, Rouen, France
E-mail: [1]ashwansharma@gmail.com, [2]indu_0012@yahoo.co.in, [3]chandramohanjana@gmail.com,[4]ophrob@hotmail.fr

Abstract: *Structural Genomics (SG) proteins are the unknown function proteins with little knowledge about their function and binding sites residues. Many of these proteins are regulating stress mediated and disease mediated pathways in the pathogenic microorganism such as Mycobacterium, influenza and H1N1 virus. Therefore, these proteins serve as novel drug targets for designing potential drugs against these microorganisms. We selected SG protein Rv2623 from Mycobacterium tuberculosis regulating the stress response pathways. Molecular docking and Molecular dynamics approaches were employed for finding natural potential inhibitor against Rv2623 protein. Plant origin compound Naringin from citrus fruit shown higher affinity for Rv2623 as compare to first line anti-tuberculosis compounds. Therefore, Naringin compound can act as potential inhibitor for Rv2623 protein which may indirectly inhibit the growth of Mycobacterium tuberculosis in the host cell.*

1. INTRODUCTION

In the post-genomic era, a large number of protein sequences are available, which have not been assigned any function. Three dimensional structures of a many of these proteins are being solved with substantial investment of resources with the objective to assign function. With this paradigm shift, the Structural Genomics (SG) projects around the world are rapidly producing many novel protein structures with little knowledge about their function [1-3]. Currently, 104125 (www.rcsb.org) proteins three-dimensional structures are available in the protein data bank and few of them are still uncharacterized and classified as "structural genomics unknown function protein (SG)". Several of these SG proteins are available in the pathogenic microorganism such as Rv2717, Rv3472 and Rv0098 and Rv2623 etc. in *Mycobacterium tuberculosis*. These pathogenic unknown function proteins may perform specific function during the growth cycle of tuberculosis bacteria and participate in its cellular metabolism. Therefore, these M. tuberculosis SG proteins may act as novel drug targets against tuberculosis disease. Here we propose the strategy for functional classification and active site detection along with substrate or inhibitor finding against these SG proteins from *Mycobacterium tuberculosis*. Sequence and structure comparison methods failed to assign function and functional sites to these pathogenic unknown function proteins as less sequence similarity obtained during sequence alignments [4]. Therefore, several structures based methods such as DALI [5], PINTS [6] and PROFUNC [7] etc. have been developed for finding the function in these structural genomics proteins. However, most of these methods have been failed to provide information about potential ligand or inhibitor against these pathogenic unknown function proteins from *Mycobacterium tuberculosis*.

We presume that M. tuberculosis SG proteins can act as potential novel drug targets against tuberculosis disease. Several anti-tuberculosis drugs are available in the market against few enzymes of Mycobacterium [8-11]. However, resistance developed by the bacterium against these drugs always perceives to search for novel anti-tuberculosis drug targets. Additionally, due to severe complications caused by the existing anti-tuberculosis drugs, there is needed to develop new drug therapy against tuberculosis [10, 12-13]. Several plant metabolites have been explored in order to design anti tuberculosis drug therapy. Artonin F [14] and linaroside [15] are two flavonoids from the root bark of *Artocarpus rigidus* (Moraceae) and from *Lantana camara* Linn (Verbenaceae) showing antimycobacterial activity against *Mycobacterium tuberculosis*. Xanthones is another sub-group of phenolics. Artonol B which belongs to these particular compounds show antimycobacterial activity against *Mycobacterium tuberculosis* [14] but also cytotoxicity against cancer cell lines [16-17]. Curcuminoids belong to phenolics too. For this sub-group,

demethoxycurcumin is one of the main active compounds. It is isolated from the rhizomes of *Curcuma longa* (Zingiberaceae) and exhibits antitubercular activity against *Mycobacterium tuberculosis* [18] as well as anti-inflammatory and anticancer activities [19]. Nordicentrine is an aporphine which belongs to quinoline alkaloids class. It is having antimicrobial activity against *Mycobacterium tuberculosis* [20]. These plant origin compounds are easily available, less cost effective and can be effective to inhibit the growth of tuberculosis bacteria. Therefore, we are proposing application of Molecular Dynamics [21] and docking [11] approaches to find potential natural inhibitor for unknown function protein Rv2623 from *Mycobacterium tuberculosis*. We already published information about novel commercial anti-tuberculosis compounds against Cytochrome P450 mono-oxygenase enzyme of *Mycobacterium tuberculosis* [11].

2. MATERIALS AND METHODS

2.1 Protein file

Three dimensional structure of the unknown function protein Rv2623 from *Mycobacterium tuberculosis* was targeted for our Molecular docking and Molecular dynamics studies. Its X-ray crystalline structure was obtained from (PDB Code: 3CIS) RCSB protein data bank in the pdb format. Further, its chain A was separated and seeded for our docking and molecular dynamics studies.

2.2 Anti-tuberculosis compounds library

Several plant origin anti-tuberculosis compounds have been reported in the literatures. Their potential applications have been summarized to treat tuberculosis disease and works have been determined to target different proteins from *Mycobacterium tuberculosis.* We collected information about these plant origin anti-tuberculosis compounds and made a library. Total 160 compounds were downloaded from CHEMDB, PUBCHEM database and their SMILES strings were converted into 3D structures via CORINA server (http://www.molecular-networks.com/online_demos/corina_demo). In addition, different first line chemical anti-tuberculosis compounds were also considered for our docking study as reference. These drugs are currently used for treating the tuberculosis disease. Therefore affinities of the plant origin compounds were compared with these first line compounds for the Rv2623 protein from *Mycobacterium tuberculosis*. These compounds were downloaded from DRUG BANK database.

2.3 Active site residues prediction

Structure of Rv2623 protein (PDB Code: 3CIS, chain A) was submitted to functional site prediction servers such as PROFUNC, Q-site finder and PINTS for putative binding site residues prediction. These residues were chosen as potential docking target against the anti-tuberculosis compound library.

2.4 Molecular Docking

The anti-tuberculosis compounds library was screened against the pdb structure of Rv2623 protein (PDB Code: 3CIS, chain A) by computational molecular docking method. The molecule docking was performed by iGemdockv2.1 software. Note that the binding site residues predicted by the active site prediction servers used as docking target. The first line anti-tuberculosis compounds were also docked with the Rv2623 protein. The *Drug Screening* platform of iGemdock was selected for our computational docking study with parameters such as Population size: 200, Number of generations: 70 and Number of solutions: 3. In the output file, the compounds were ranked based on their interaction energies and fitness values produced by the docking via iGemdock software. The most stable conformation of the plant origin anti-tuberculosis compound was selected based on the lowest fitness value and further submitted for Rv2623-ligand complex molecular dynamics simulation for refinement of compound binding at the active site of protein.

2.5 Statistical analysis

Distributed statistical analysis was performed by histogram analysis and Gaussian distribution. It determined the distribution of the interaction energies after the iGemdock docking. Furthermore, statistical test was performed by measuring standard score known as the Z scores for the interaction energies produced by iGemdock. The Z score is able to measure uniqueness and confidence in the docking result and to find unique compound showing affinity for Rv2623 which was unlikely to produce by other compounds in random population. Later P values were estimated for each Z score to detect confidence in the docking results. The standard P value was set as cut off that is 0.05 at 95% confidence interval. The best compound was selected with based on more negative Z score of interaction energy and P value less than < 0.05.

2.6 Molecular Dynamics of Rv2623-ligand complex

The top hit plant origin anti-tuberculosis compound from the docking step was submitted to Rv2623-ligand complex molecule dynamics simulation by GROMACS.v.4.5.5 software. First, the GROMOS9643A1 atoms force field was chosen for preparing topology file for Rv2623 protein by using *pdb2gmx* command. Next, the pdb coordinates of the top hit (lowest interaction energy from iGemdock) plant origin compound was obtained after the docking process and furnished to PRODRG server for generating the compound GROMACS topology file. Note that during PRODRG processing, the chirality was on and full charges were provided to the compound structure. Later, partial charges were adjusted for the compound in the topology file according to the force field. Furthermore, the Rv2623 protein and

compound topology files were combined and made protein-ligand topology file and saved in *.gro* format. The solvation of Rv2623 protein-compound system was performed by using SPC water model with 25598 number of water molecules within cubic box of size 1.0 nm by using *editconf* and *genbox* commands.

Molecular dynamics was performed with neutralized system charge. Therefore, the system was kept neutral by adding 8 number of Na^+ ions using *genion* command. In the first step of molecular dynamics simulation, the system was pre-minimized for 1000 steps using steepest descent method by *grompp* tool. Furthermore, the minimized system was furnished to heating process where system was slowly heated to constant temperature of 300 K via NVT (constant number of particles, volume and temperature (300 K) ensemble dynamics for 100 ps (time steps, dt=0.002 and nsteps =50000) with leap-frog integrator. Note that Andersen thermostat was used for maintaining the constant temperature of 300 K. Next, the NVT was continued with NPT (constant number of particles, pressure and temperature) ensemble dynamics for 100 ps where the pressure coupling of 1 bar was provided with Berendsen algorithm. Furthermore, the NVT and NPT equilibrated system was subjected to final production molecular dynamics for 3 ns with time steps of 2fs and number of steps of 1500000.

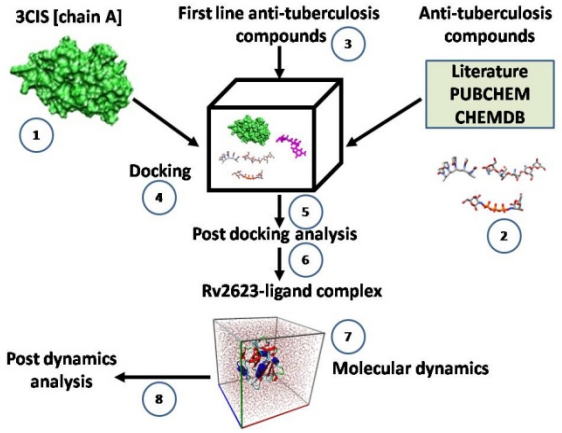

Fig. 1: It shows the schematic for the overall method.

(1): Obtained the Structure of Rv2623 protein (PDB Code: 3CIS, chain A) from protein data bank. (2): Downloaded the structures of the plant origin Anti-tuberculosis compounds from PUBCHEM and CHEMDB databases. (3): Obtained first line chemical anti-tuberculosis compounds from DRUG BANK database. (4): Docking of compounds library against the active site residues of Rv2623. (5): Post docking analysis for binding position, binding pose and docking energy of the compounds. (6): made Rv2623-ligand complex with the ligand produced lowest docking energy. (7): Complex refinement by molecular dynamics. (8): Post dynamics analysis for hydrogen bond analysis and stabilization of ligand interactions with Rv2623.

GROMACS utility commands such as *g_energy* (for Total energy, Potential energy and Temperature), *g_gyrate* (for radius of gyration) and *g_rms* (for root mean square deviation (rmsd) of the complex structure from the original complex structure) were utilized for the output analysis. Additionally, *g_hbond* was utilized to calculate the number of hydrogen bonds formed between Rv2623 and compound. Study was further extended to analyze the hydrogen bond distances between their Donor (D) and Acceptor (A) atoms. *g_rdf* command was used to measure the radial distribution function between Rv2623 and compound. Finally, the obtained trajectories were clustered using *g_cluster* module of GROMACS package to calculate the RMS clusters of Rv2623 backbone-backbone conformations with RMSD cutoff of 0.1nm.

3. RESULTS

Library of 160 plant origin natural compounds was docked with the Rv2623 protein (PDB Code: 3CIS, chain A) via iGemdock. Compounds were screened against the functional site residues predicted by the functional site prediction servers. Furthermore, the compounds are ranked based on their interaction energies (fitness values) with the protein Rv2623. Plant compound Naringin produced lowest interaction energies (fitness value) -126.41 kcal/mole as compared to other natural compounds. Second and third ranks are obtained by Delphinidin and Artonol B with interaction energies of -119.24 kcal/mole and -116.84 kcal/mole. These docking results show that Naringin compound has very high affinity for the active site of Rv2623 as compare to other compounds (Fig. 2).

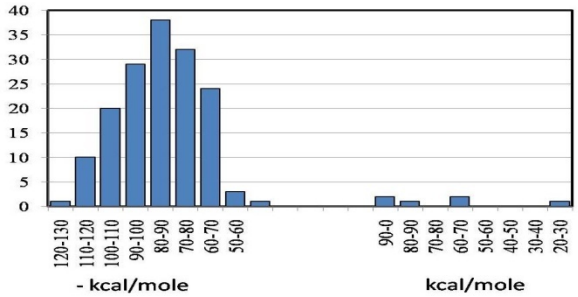

Fig. 2: Histogram showing the distribution of the interaction energies produced after molecular docking of the anti-tuberculosis compounds with the Rv2623 (chain A) by iGemdock software.

The interaction energy of Naringin with the Rv2623 produced Z score of -1.40 and located at the left tail of the distribution (Fig. 3). The Z score analysis revealed that the affinity produced by Naringin for the Rv2623 is very unique and this strong binding is not able to detect by other plant compounds. The corresponding P value for the Z score of Naringin interaction energy is 0.01which is lesser than the standard P value cut off of 0.05 (95% confidence interval) and has

confidence of 99% on the docking affinity of Naringin for Rv2623.

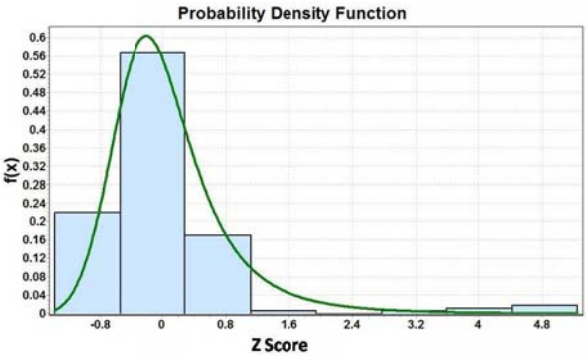

Fig. 3: Distribution of the Z scores of the docking interaction energies with respect to their probabilities distribution (P values).

The docking affinity of Naringin for Rv2623 is compared with the first line anti tuberculosis compounds. The values of interaction energies of first line anti-tuberculosis compounds are listed in Table 1. It shows that streptomycin compound produced lowest interaction energy of -106.40 kcal/mole and shows best affinity for the RV2623 as compared to other compounds. However this affinity is very low as compare to the affinity of Naringin for Rv2623. Even, the Z scores produced by the first line anti-tuberculosis compounds are not significant and unique as compare to Naringin Z score which shows that Naringin can act as potential substitute for first line anti-tuberculosis compounds.

Table 1: Interaction energies of top plant origin and first line chemical anti-tuberculosis compounds.

Rank	Compounds	Interaction Energy (kcal/mole)	Z Score
1	Naringin	-126.412	-1.37
2	Delphinidin	-119.239	-1.16
3	Artonol B	-116.842	-1.08
First line anti-tuberculosis compounds			
4	Streptomycin	-106.401	-0.77
5	Ethambutol	-73.528	0.20
6	Isoniazid	-65.053	0.45
7	Pyrazinamide	-64.705	0.46
8	Rifampicin	27.659	3.20

Naringin compound is a flavanone-7-*O*-glycoside between the flavanone compound Naringenin and the disaccharide neohesperidose. The flavonoids naringenin and hesperetin, which form the aglycones of naringin and hesperidin, occur naturally in citrus fruits, especially in inner peel and pulp of grapefruit, where naringin is responsible for the fruit's bitter taste.

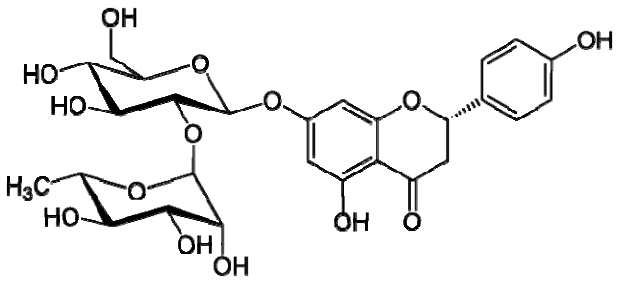

Fig. 4: 2D structure of Naringin compound.

3.1. Binding site Residues

Residues were extracted that surrounding the Naringin at the binding site of Rv2623 protein. Fig. 5 shows that the amino acid residues Gly 13, Asp 15, Asp 16,Ser 17, Ala 20, Gln 21, Gly 117, Gly 120, Ser 121, Gly 122, Arg 123 Trp 124 and Arg 127 directly surround the Naringin. The residues Ser 17, Ser 121, Trp 124 and Arg 127 make hydrogen bonds with the Naringin. Also Asp 15, Ala 20, Gln 21 and Gly 122 make van der waals interaction with Naringin.

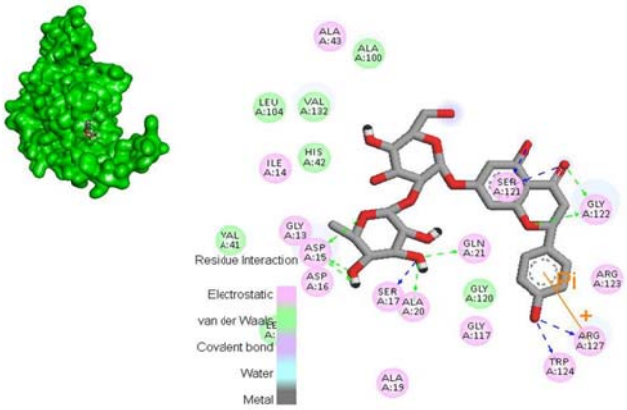

Fig. 5: Binding site residues surrounding the Naringin at the binding site of Rv2623 protein.

3.2.Molecular Dynamics of Rv2623-Naringin complex

The complex of Rv2623 protein (PDB Code: 3CIS, chain A)-Naringin plant compound was refined by Molecular dynamics simulation. The complex was submitted to 3 ns of simulation and stabilized after 3 ns of simulation. The system produced average potential energy of -1.099 x10^{-6} kJ/mole.

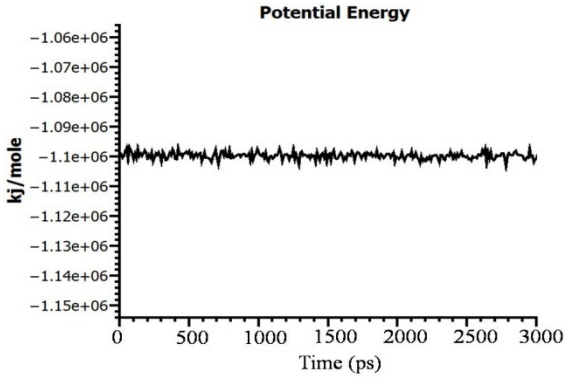

Fig. 6: Potential Energy of the system during the simulation.

Radius of gyration analysis revealed that Rv2623-Naringin complex was remain at constant distance and predicted that the system was very compact during the course of MD simulation for 3 ns (Fig. 7). This shows that there is strong force of interactions between Naringin and Rv2623 protein.

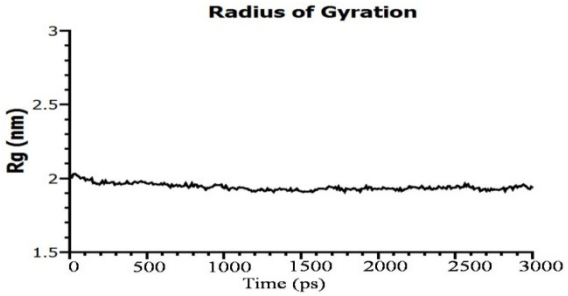

Fig. 7: Radius of gyration for Rv2623-Naringin complex.

Fig. 8 shows the backbone-backbone conformations RMSD which determines amount of atomic fluctuations occur for the Rv2623-Naringin complex during the course of MD simulation. It represents that from 0 ps to 1500 ps there is more atomic fluctuations for Rv2623-Naringin complex however after 1500 ps, the system becomes stable and has average RMSD of 0.27 nm (Fig. 8).

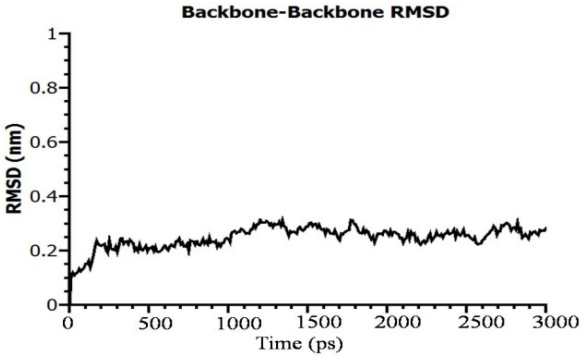

Fig. 8: Backbone-Backbone conformations RMSD for the Rv2623-Naringin complex during the MD simulation of 3 ns.

Furthermore, the formation of number of hydrogen bonds was analyzed between Rv2623 and Naringin during the MD simulation. Fig. 9 shows that during most of the simulation time of 3 ns, the Naringin interacted with the Rv2623 protein by forming three numbers of hydrogen bonds. Sometimes one or two extra number of hydrogen bonds is formed. However after 2600 ps, the three numbers of hydrogen bonds are formed between Naringin and Rv2623 protein that maintained until end of the simulation.

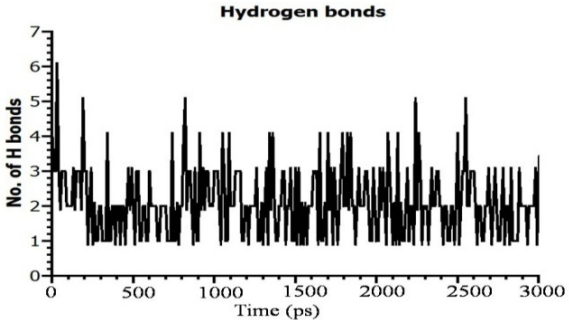

Fig. 9: Number of hydrogen bonds formed between Naringin and Rv2623 protein during the MD simulation of 3 ns.

Fig. 10 depicts the radial distribution function between Naringin and Rv2623 protein. It reveals that the density of Naringin compound vary from the center of Rv2623 protein within distance of 1Å.

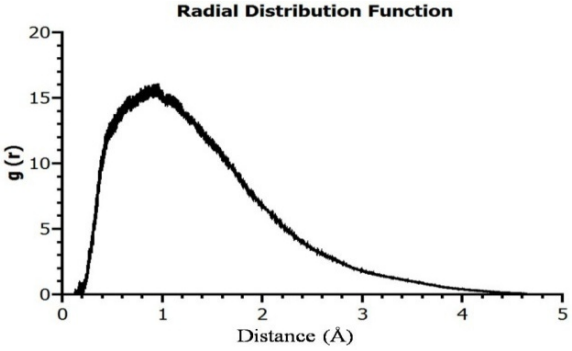

Fig. 10: Radial Distribution Function between Naringin and Rv2623 protein during the MD simulation of 3 ns.

Clustering of the MD conformations was performed using single linkage (nearest neighbor) method within RMSD cut off 0.1 nm from the centre structure. Fig. 11 depicts that three clusters are formed for the conformations of the complex system. For initial 60 ps of simulation, all conformations are stored in cluster 2. Later from 80 ps to 140 ps all conformations are stored in cluster 4. After 150 ps, the Rv2623-Naringin complex system becomes stable and all the conformations are stored in cluster 5. We extracted several snapshots after 500 ps in order to analyze the formation of hydrogen bonds between Naringin and Rv2623 protein.

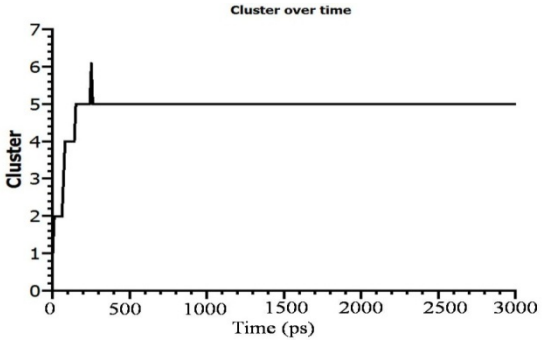

Fig. 11: Cluster over time for the conformations of Rv2623-Naringin complex during the MD simulation of 3 ns.

Snapshots were extracted at starting point of 0 ps and end of the simulation at 3000 ps. At 0 ps, the residues Ser 121, Met 107, Ser 133, Ser 131, Arg 127 and Gly 120 make interaction with the Naringin compound. However, at 3000 ps, the stabilized hydrogen bonds and van der waals interactions are formed where His 42, Ala 43, Gly 117 and Gly 120 make direct interactions with the Naringin. The Molecular dynamics is able to find stabilized complex and refined the interaction between Naringin and Rv2623 protein.

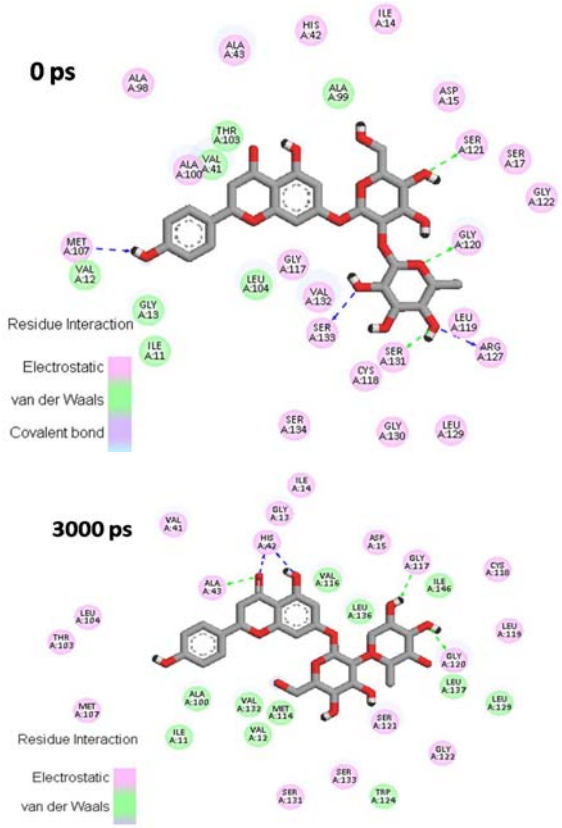

Fig. 12: Snapshots at 0 ps and 3000 ps from the simulation of Rv2623-Naringin complex.

4. DISCUSSION

In our work we targeted Structural genomics proteins Rv2623 from pathogenic microorganism *Mycobacterium tuberculosis* for drug discovery. The *Mycobacterium tuberculosis* is generating resistance against the already existing chemical drugs therefore there is need to find alternative strategy for treating the tuberculosis disease. We selected Rv2623 protein which is a stress protein and protecting the mycobacterium bacteria from adverse environmental conditions. In our work, we predicted that the plant compound Naringin produced lowest interaction energy and has higher affinity for Rv2623 as compare to first line tuberculosis compounds. This reveals that Naringin can be act as potential inhibitor for Rv2623 protein which indirectly reduces its activity to protect the tuberculosis bacteria from adverse conditions. Also as the plant origin of Naringin such as from grape fruit, citrus fruit, it is easily available and has no side effect as compare to chemical anti-tuberculosis drugs. In addition, our computational docking and molecular dynamics studies may help in reducing the time and cost for screening and selecting the best compounds for future drug designing. Therefore, we propose that the plant origin compounds such as Naringin can be act as potential anti-tuberculosis compound for treating the tuberculosis disease.

5. CONCLUSION

Our computational drug discovery study against the Rv2623 protein from *Mycobacterium tuberculosis* is very helpful in designing or finding novel drugs for treating tuberculosis disease. We detected that plant origin compounds have more affinities for Rv2623 as compare to first line chemical anti-tuberculosis compounds which provides a clue about replacing these chemical compounds by plant origin compounds for treating tuberculosis. Our application of molecular docking and molecular dynamics for drug discovery may help the experimental biologist for fast screening of compounds and may reduce the load of testing of large number of compounds which is very expensive. Also our computational approach is able to determine binding site residues for Naringin at the active site of protein. Therefore, the Naringin compound may act as good model for designing natural medicine for treatment of tuberculosis disease.

6. ACKNOWLEDGEMENTS

Indu Chaturvedi greatly acknowledges Indian Science and Technology Organization (ISTO) and CIRG, Mathura for providing all financial supports. IC and CJ are very thankful to New Era Proteomics, Delhi and France for providing computer facilities for running computational calculations.

REFERENCES

[1] Pazos, F., and Sternberg, M. J. E., "Automated prediction of protein function and detection of functional sites from structure", *Proceedings of the National Academy of Sciences of the United States of America,* 101, 2004, pp. 14754-14759.

[2] Baker, D., and Sali, A., "Protein structure prediction and structural genomics", *Science,* 294, 2001, pp. 93-96.

[3] Stark, A., *et al.*, "Finding functional sites in structural genomics proteins", *Structure,* 12, 2004, pp. 1405-1412.

[4] Bryliński, M., *et al.*, "Prediction of functional sites based on the fuzzy oil drop model", *PLoS Computational Biology,* 3, 2007, pp. 0909-0923.

[5] Holm, L., and Rosenström, P., "Dali server: Conservation mapping in 3D", *Nucleic Acids Research,* 38, 2010, pp. W545-W549.

[6] Stark, A., and Russell, R. B., "Annotation in three dimensions. PINTS: Patterns in Non-homologous Tertiary Structures", *Nucleic Acids Research,* 31, 2003, pp. 3341-3344.

[7] Laskowski, R. A., *et al.*, "ProFunc: A server for predicting protein function from 3D structure", *Nucleic Acids Research,* 33, 2005, pp. W89-W93.

[8] Leys, D., *et al.*, "Atomic structure of Mycobacterium tuberculosis CYP121 to 1.06 Å reveals novel features of cytochrome P450", *Journal of Biological Chemistry,* 278, 2003, pp. 5141-5147.

[9] McLean, K. J., *et al.*, "The preponderance of P450s in the Mycobacterium tuberculosis genome", *Trends in Microbiology,* 14, 2006, pp. 220-228.

[10] Tripathi, R. P., *et al.*, "Fighting tuberculosis: An old disease new challenges", *Medicinal Research Reviews,* 25, 2005, pp. 93-131.

[11] Sharma, A., *et al.*, "Computational finding of potential inhibitor for Cytochrome P450 Mono-oxygenases Enzyme of Mycobacterium tuberculosis", *Bioinformation,* 8, 2012, pp. 931-7.

[12] Donald, P. R., "Antituberculosis drug-induced hepatotoxicity in children", *Pediatric Reports,* 3, 2011.

[13] Moldoveanu, B. C., *et al.*, "Risk factors for drug side effects during first-line tuberculostatic treatment", *Archives of the Balkan Medical Union,* 47, 2012, pp. 45-49.

[14] Namdaung, U., *et al.*, "Bioactive constituents of the root bark of Artocarpus rigidus subsp. rigidus", *Chem Pharm Bull (Tokyo),* 54, October 2006, pp. 1433-6.

[15] Begum, S., *et al.*, "Antimycobacterial activity of flavonoids from Lantana camara Linn", *Nat Prod Res,* 22, 15, April 2008, pp. 467-70.

[16] Ee, G. C., *et al.*, "Artomandin, a new xanthone from Artocarpus kemando (Moraceae)", *Nat Prod Res,* 25, June 2011, pp. 995-1003.

[17] Hakim, E. H., *et al.*, "Artoindonesianin P, a new prenylated flavone with cytotoxic activity from Artocarpus lanceifolius", *Fitoterapia,* 73, December 2002, pp. 668-73.

[18] Agrawal, D. K., *et al.*, "Demethoxycurcumin and its Semisynthetic Analogues as Antitubercular Agents", *Planta Med,* 74, 15, December 2008, pp. 1828-1831.

[19] Yodkeeree, S., *et al.*, "Demethoxycurcumin suppresses migration and invasion of MDA-MB-231 human breast cancer cell line", *Eur J Pharmacol,* 627, 10, February 2010, pp. 8-15.

[20] Lekphrom, R., *et al.*, "Bioactive styryllactones and alkaloid from flowers of Goniothalamus laoticus", *J Ethnopharmacol,* 125, 17, August 2009, pp. 47-50.

[21] Glazer, D. S., *et al.*, "Improving Structure-Based Function Prediction Using Molecular Dynamics", *Structure,* 17, 2009, pp. 919-929.