EFFICIENT IDENTIFICATION OF NOVEL ANTIFUNGAL COMPOUNDS TARGETING CYTOCHROME-B OF FUSARIUM OXYSPORUM F. SP. LYCOPERSICI BY VIRTUAL SCREENING AND STATISTICAL ANALYSIS

[1]Reetu Gour, [2]A.P. Garg and [3]Ajay Kumar Shukla
[1]Research scholar, [2]Professor, [3]Lecturer
[1]Department of Microbiology
Ch. Charan Singh University, Meerut (250004), India

Abstract - Fusarium oxysporum f. sp. lycopersici is a soil borne plant pathogen belonging to Class Hyphomycetes. It causes fusarial wilt specifically in tomato. Cytochrome b protein plays an important role in electron transport chain during respiration and is considered as crucial target for developing anti-fungal compounds. We predicted the protein model of cytochrome b in Fusarium oxysporum based on Saccharomyces cerevisiae cytochrome b template. Chain-A of cytochrome b in Fusarium oxysporum f. sp. lycopersici is identical to the N-chain of cytochrome b of Saccharomyces cerevisiae. Predicted cytochrome b structure was used to screen ZINC database compounds through DOCKBlaster which yielded 200 top scoring compounds. In this paper, we report best two compounds namely, ZINC53788 and ZINC18029029 which can be further evaluated for their anti-fungal activities through field or laboratory trials. The binding affinity scores for these two compounds have been calculated as -37.55 and -37.09 kcal/mol respectively using SAS. They are expected to inhibit the functions of cytochrome-b at ubiquinol binding site based on higher predicted binding affinity. This inhibition is expected to inactivate the toxic effects of F. oxysporum f. sp. lycopersici by preventing binding of ubiquinol on cytochrome-b protein by interfering in its metabolic process. The mode of action can further be elucidated and proved using the potential of ZINC53788 and ZINC18029029 compounds as anti-fungal agents.

Keywords - Binding affinity, Comparative modeling, Cytochrome-b, Fusarium oxysporum f. sp. lycopersici, Virtual Screening.

1- INTRODUCTION

Fusarium oxysporum f. sp. radicis-lycopersici is a soilborne plant pathogen in the class Hyphomycetes a causal agent of tomato foot and root rot, a disease of worldwide economic importance in commercial tomato production. The disease results in severe economic losses in the greenhouse, open field crops, and hydroponic cultures. Like all formae speciales of *F. oxysporum*, this fungus is a soil inhabitant and extremely difficult to control. Detailed investigations on the *in vivo* interactions between the pathogenic fungus and the plant could lead to the discovery of more efficient ways to manage the disease. This disease was first described by Massee in England in 1895. It is of worldwide importance where its severity has been reported in at least 32 countries with warm climate. At one time, the disease nearly destroyed tomato production in parts of Florida and the southeastern states of United States. In order to identify novel antifungal compounds for controlling fusarial wilt, specifically in tomato, cost and time effective strategies are required. Computational approaches like docking (virtual screening) are widely used in present day bioinformatics. In this study, natural compounds have been screened for their potency against Cytochrome-b. The ubiquinol binding site is targeted to identify novel competitive inhibitors of the cytochrome-b. We also predicted the protein model of cytochrome-b in *F. oxysporum* f. sp. *lycopersici* based on the structure of *Saccharomyces* cytochrome-b by homology modeling. *S. cerevisiae* has been used as a model system to characterize the effect of cytochrome-b. Stigmatellin, another inhibitor was retrieved in the bound form to the same site. Template-based modeling is the most successful computational prediction technique till current time. The F-chain of cytochrome-b in *Fusarium oxysporum* f. sp. *lycopersici* is identical to the N-chain of cytochrome-b of *Saccharomyces cerevisiae*. Then the modelled cytochrome-b structure of *Fusarium* was used to screen all compounds of the ZINC database. Top 200 chemical compounds were obtained which can be used for drug discovery to inhibit the function of cytochrome-b in plants on the basis of binding affinity of these compounds which was calculated through statistical analysis.

2- MATERIALS AND METHODS

Protein target and template selection

The target protein, cytochrome-b of *Fusarium oxysporum* f. sp. *radicis-lycopersici* [genpept ID: 50429189] was retrieved from the National Centre for Biotechnology Information (NCBI). The homology model of cytochrome-b of *F. oxysporum* f. sp. *radicis-lycopersici* was predicted through i-Tasser. The model was visualized through Chimera. The crystal structure of the template, protein cytochrome-b of *Saccharomyces cerevisiae* (cyt-b) [PDB ID: 1KYO] was retrieved from the Protein Data Bank (PDB). Cytochrome-b was modified to get the bound ligand and water molecules removed using Viewerlite 5.0. This structure was used as the template structure for the F-chain of cytochrome-b protein of *Fusarium oxysporum* f. sp. *radicis-lycopersici.*

Homology modeling and validation

The term homology modeling also called the comparative modeling refers to the modeling of protein 3D structure based on an experimentally determined structure of homologous protein as the template. To study various functions of protein, dynamics of interaction between ligand and targets and even in the drug discovery community, knowledge of protein structure are of crucial importance. After the prediction of the 3D protein model, this model was validated through PSVS which generated Ramachandran plot on the behalf of allowed PHI-PSI angle values in the protein structure. The stereo-chemical validation of model structures of proteins is an important part of the comparative molecular modeling process. The predicted protein model of *Fusarium oxysporum* f. sp. *radicis-lycopersici* [cyt_b & the S. cerevisiae PDB model] were structurally superimposed for comparison. We found out that the N-chain of *Saccharomyces cerevisiae* was identical to the F-chain of *Fusarium oxysporum* f. sp. *radicis-lycopersici* through Chimera. Ubiquinol binding site is present on F-chain (*F. oxysporum* f. sp. *radicis-lycopersici*) where the stigmatelin was retrieved in bound state with cytochrome as an inhibitor (at ubiquinol binding site).

Virtual screening

Molecular docking is a computational technique to elucidate the mode of action of binding of two molecules to form a stable complex and also its preferred conformation. The simulation of the possible interaction between two molecules can answer important biological queries and can be used to draw breakthrough conclusions. The extent of binding is measured as the binding affinity and usually has the unit of kcal/mol. In certain software the binding affinity can be converted to a binding score. This quantity is helpful while ranking a number of compounds for their affinity towards the target. This is commonly termed as virtual screening. After the prediction of protein model of F-chain of *Fusarium oxysporum* f. sp. *radicis-lycopersici*, it was chosen as the target for docking analysis. The site at which a molecule with specific shape fits to interact is known as the active site. The active site is defined on the basis of lock and key model. Because of secondary and tertiary folding, proteins adopt specific structures. DOCK Blaster provided the top 10 results for active site search. In this study, we selected one binding site on the basis of similarity of residues which were present in interaction with previously bound ligand at the ubiquinol binding site. Docking was done using the online server DOCKblaster which automatically screens the ZINC database compound library to search for probable inhibitors. The output is in the form of top 200 scoring candidates ranked according to their binding affinity (in kcal/mol).

Descriptor selection using statistical analysis:

The descriptors selected were input in the SAS software (Version 9.2) for regression analysis and selection of the most significant descriptors using stepwise regression. The significant descriptors are selected based on their probability at 5% level of significance.

QSAR modeling:

A linear regression model is the form of $Y = b_0 + b_1X + \varepsilon$, where b0 is the intercept and is called the regression coefficient; b1 is the parameter estimate of an independent variable. There can be a number of independent variables with the parameter estimate as their coefficients and their positive or negative sign depicting the type of contribution they have in the model & is the error term depicting the residual which cannot be explained by the selected descriptors. The model was developed using 50 compounds. The values of R-square and Adj-Rsq are taken as a measure of robustness of the QSAR model. A value of more than 0.7 is considered to be good. If a model has more number of variables, the value of Adj-Rsq should be considered.

Regression analysis:

These all descriptors are submitted with its yield into Statistical Analysis Software (SAS). It provided the output which contained R-square value F-test value, standard error value; Mean square value etc .Then generated the 2D QSAR model on the basis of the highest r-Square value. The variables which have a greater r-Square value that variable predicted the best QSAR 2D model. On the basis of the predicted value and the dependent value of the variable generated the different types of plot-

Fitness plot- This plot is generated on the basis of the predicted value of the dependent variable.

Contribution plot- contribution plot is generated between the partial r-square value and variable here partial r-square value was treated as a contribution of the dependent variable.

Radar plot - Radar plot is generated between the predicted value and the dependent value of the variable which show the relation between the predicted value by SAS and the dependent value of the variable.

Model Validation

Many statistical parameters like n (number of compounds in regression), k (number of variables), degree of freedom, optimum component (number of optimum PLS components in the model), r2 (squared correlation coefficient), F-test (Fischer's value), q2 (cross-validated correlation coefficient), pred_r2 (r2 for external test set) etc are used to validate the protein model statistically.

3. RESULTS AND DISCUSSION

In order to screen potent antifungal compounds, we carried out the basic virtual screening protocol on the ZINC database. The ZINC database contains over a million compounds with drug-like properties. The protein target chosen in the study is mitochondrial cytochrome b, a crucial factor in the respiratory electron transport chain. Out of the whole complex we specifically targeted the N- chain of Saccharomyces cerevisiae to which ubiquinol is known to bind as the Q0 binding site. The homology model of F-chain of Fusarium oxysporum f. sp. radicis-lycopersici is given in Figure 1. In this structure prediction procedure, the 2 top-ranked alignments between the target sequence and the structures in the template library are used to build the protein structure. The first ranked alignment with template 3KYO representing S. cerevisiae Cytochrome b Nchain yielded a sequence identity of 38% % with the query and thus the final model was selected for docking. The protein model obtained majorly constitutes alpha helical secondary structural motifs. These helices span the trans-membrane region of mitochondria. Majorly all chains of Cytochrome b consists of alpha chains.

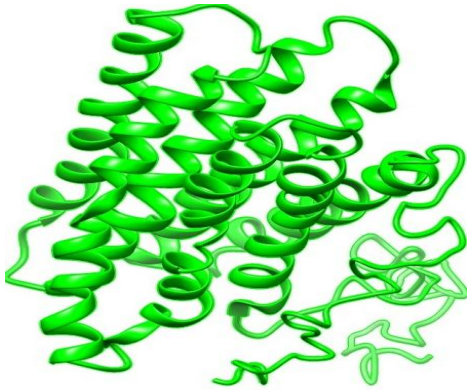


Figure- 1 Three dimensional structure of Fusarium cytochrome b chain A

Validation

A protein model obtained through in silico procedure should be validated before further analysis for its conformational correctness. This is based on the number of residues falling under the Ramchandran criteria for phi-psi dihedral angle values. These dihedral angles also called as torsion angles in a protein define the degree of protein folding process possible under steric laws. Also given is the Ramchandran plot for Cytochrome b A-chain (Figure 2).

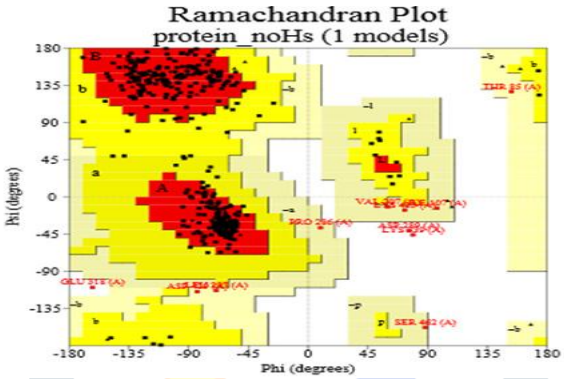


Figure- 2 Shows 86.6% residues in most favoured region; 11.1% residues in additional allowed region; 1.7% residues **in additional allowed region and 0.7%** in d**isallowed region.**

The colored regions on the plot represent different combinations of phi-psi dihedral angles, such as; the most favorable combinations of phi-psi values for corresponding residue conformations have been represented by red regions while the white areas represent the conformations where atoms of the polypeptide come even closer than the sum of their van der Waals radii. These regions are sterically excluded and 0.0% of modelled Cyt_b (Fusarium oxysporum f. sp. radicis-lycopersici) amino acid residues fall in this category. Thus, the red regions correspond to conformations namely the α-helical and β-sheet without any steric clashes. About 86.6% of the residues fall in the most allowed region. The yellow areas represent the allowed regions if the atoms are allowed to come a little closer together (slightly shorter van der Waals radii are considered). 1.7 % of the residues were seen to fall in this region. An additional region corresponding to the left-handed α-helix is obtained. 11.1% of the residues fall under this region. There were a total of 11 outliers (residues) in the protein model. Another quantity that is calculated is the G-factor that has been represented in Figure 3. The G-factor can be defined as a log odd score based on the observed distributions of phi-psi dihedral angles. A low G-factor score ranging from -1 to -4 indicates that a given structure corresponds to a low-probability conformation. A high G-factor score ranging from 0 to 2 indicates the high-probability conformation. The amino acid residues majorly fall in the lower G-score range and thus indicate a low probability conformation. However, the graph shows good positive G score values for a few residues both for all dihedral angles and phi-psi angles. These residues show good structural conformation in the structure.

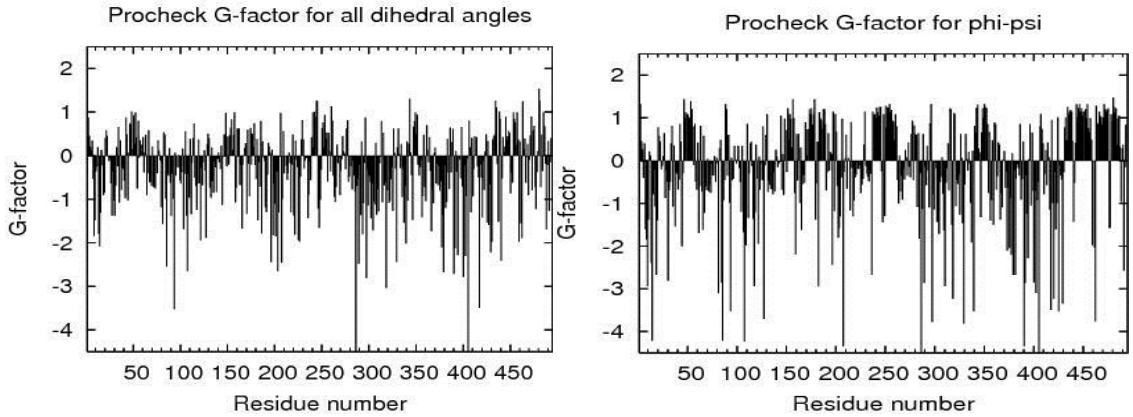
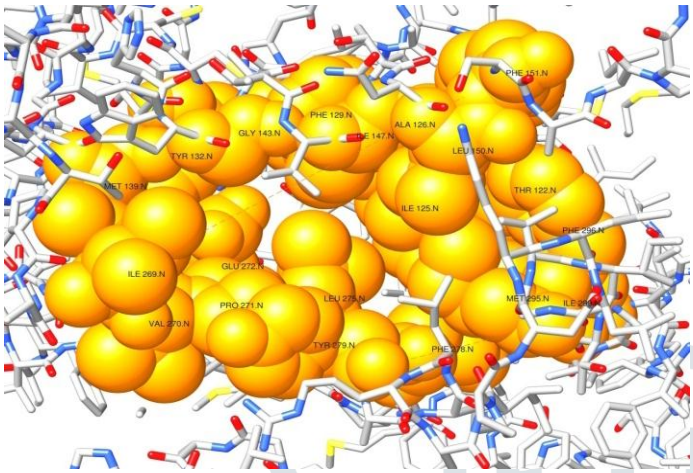


Figure-3 G-factor graphs for phi-psi and all dihedral angles

Docking Analysis

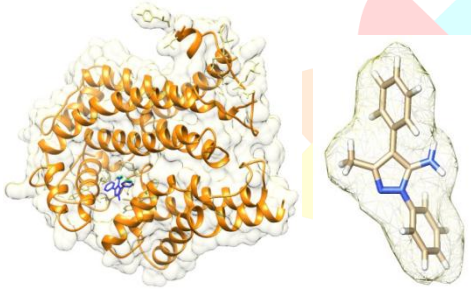
For predicting the interaction of protein target with that of ZINC database compounds, docking was performed by DOCKBlaster software. Result of active site search step through DOCKBlaster provides the residues involved in putative binding sites. It resulted in 12 active sites in which one site was selected on the basis of the similarity of residues which are present surrounded the ubiquinol binding site where the stigmatellin was also bound to inhibit the function of cyt_b. Here we selected only common residues constituting the active site for more accuracy. Residues constituting the active site are MET139, TYR132, GLY143, PHE129, ILE147, ALA126, PHE151, LEU150, ILE125, THR122, PHE296, MET295, ILE299, PHE278, TYR279, LEU275, PRO271, GLU272, ILE269, and VAL270 are reported in (figure-4).



MET 139	PHE 296
TYR 132	MET 295
GLY 143	ILE 299
PHE 129	PHE 278
ILE 147	TYR 279
ALA 126	LEU 275
PHE 15	PRO 271

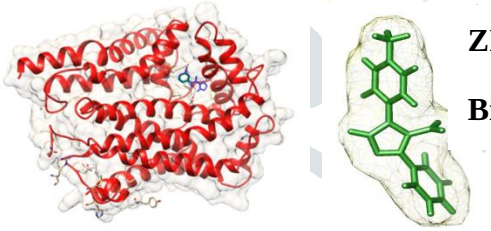
Figure -4 Active site cavity of protein Sequence and list of residues

Two hundred energetically stable inhibitor configurations were obtained by DOCK Blaster. Out of these configurations, the top 10 with lowest docking energy score (in kcal/mol) have been selected from which best 2 scoring compounds along with their binding affinities and other properties are reported below.



ZINC ID =53798

Binding Score = -37.55 kcal/mol



ZINC ID = 18029019

Binding Score = -37.05 kcal/mol

Figure -5 The top scoring compounds and their properties

QSAR Modeling

Here we report a QSAR model that looks for structure-activity relationship for the reported docked chemical compound targeting the Q_o site of Cyt b subunit of the multi-unit mitochondrial transmembrane protein Cytochrome bc1 which plays a crucial role in electron transfer in Saccharomyce cerevisiae & Fusarium oxysporum f. sp. Lycopersici.

This study involves 48 compounds which are those chemical compounds that act as a inhibitor in cytochrome b at the ubiquinol binding site in case of Fungi (Fusarium oxysporum f. sp. Lycopersici) or yeast (Saccharomyces cerevisiae) using regression analysis. These compounds were evaluated for the various properties that directly or indirectly impart chemical properties and the presence or absence of which may enhance their biological action. These properties or molecular descriptors are quantitative and can be 2D (spatial, geometrical). These quantitative descriptors are taken as independent variables and Binding affinity (BA) & Molecular Weight (MW) are taken as the dependent variables for a regression analysis that results in a linear model. This linear model can be used to predict the value of dependent variable (in this case, BA & MW) for unknown compounds using the values of same set of independent variables. The descriptor have been selected on the basis on the binding score of chemical compounds The complete statistical stepwise regression analysis on the basis of binding score yielded top 6 significant descriptors are ;JG17, Nx, ATS6m, n Heavy Atom.

Model evaluation and validation

Model is evaluated on the basis of certain statistical parameters using both internal and external validation. The number of compounds in the training set was specified by N which is 48. Considering the regression coefficient, r2 (0.4357) the model can be stated not to be a robust one

although it has good predictive power for training set compounds. Other important statistical parameters are presented in Table-1 and JG18 is highlighted to emphasize its importance in QSAR model validation.

Table -1 The values of final descriptors calculated for G-QSAR model development are provided in Table 3.

Dependent variable	JG17	nx	JG14	ATS6m	n H Atom	**JG18**
BA	0.1008	0.2065	0.2509	0.3235	0.3704	**0.4357**

Table 2. Observed values and predicted values of binding affinity for the docked chemical compound (from ZINC DB)

OUTPUT STATISTICS			
Obs.	Dependent value	Predicted value	Residual
1	37.75	36.0287	1.7213
2	37.09	36.3195	0.7705
3	36.71	35.857	0.853
4	36.51	36.6258	-0.1158
5	36.43	35.5153	0.9147

The robustness of the model is better understood through the linear graphical representation between actual and predicted activities of the 46 compounds (Figure-6) and radar plots for predicted value and dependent value of the variable (Figure-7).The linear graphical representation shows the extent of variation between the actual and predicted activities of the concentric set. The larger the distance of training and test set points from the regression line, more is the difference between the actual and the predicted activity values. The radar graphs depict the difference in the actual and predicted activities for the training and the test sets separately by the extent of overlap between blue (actual activity) and red (predicted activity) lines. The radar plot for training set represents a good r2 value if the two lines show a good overlap while for the test set a good overlap represents high pred_r2 value

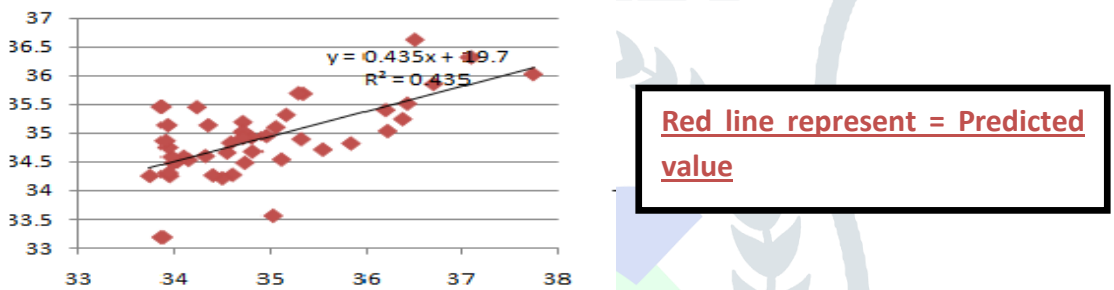


Figure- 6 Graphical representation of observed vs. predicted activity and Dependent variable.

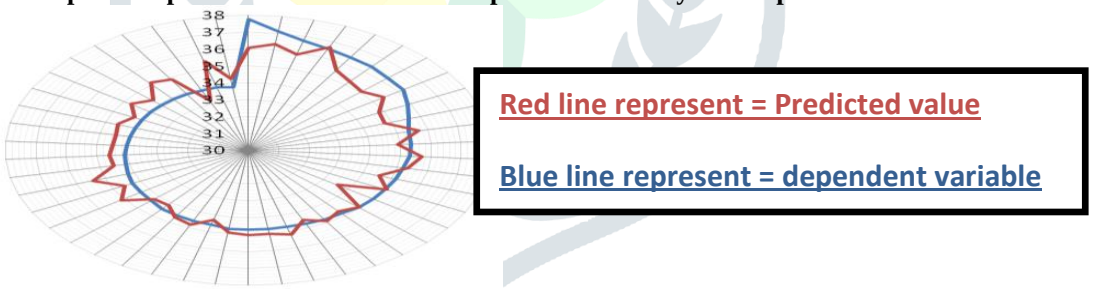
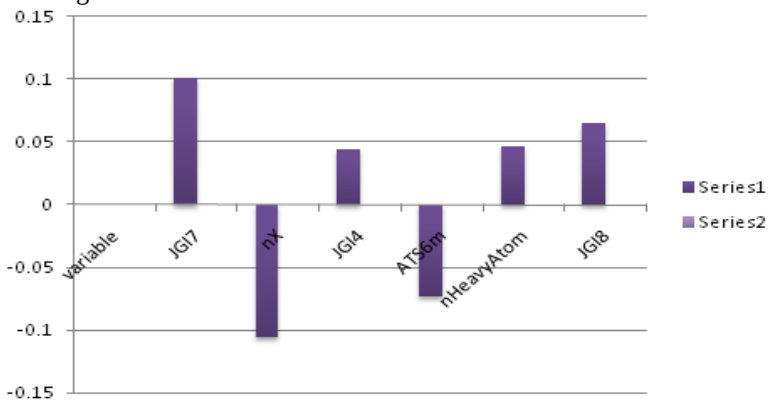


Figure- 7 Radar graph representing the similarity b/w dependent & predicted value of variable

The contribution plot for each descriptor is given in Figure -9. The contribution of each descriptor specifies the properties that should be present or absent in the drug lead for its enhancing its inhibitory activity. Presence of descriptors with positive contribution increases its inhibitory activity while descriptors with negative contribution decrease the same



Contribution Plot of the final descriptors for the generated 2D QSAR model

4. CONCLUSIONS

The study reports screening of ZINC database compounds based on docking and statistical analysis through SAS with the Fusarium oxysporum f. sp. radicis-lycopersici cytochrome b as the target protein. We also report the three dimensional structure of Fusarium oxysporum f. sp. radicis-lycopersici predicted by comparative modeling approach. Out of 11 putative active sites of the modeled protein predicted by DockBlaster, only one was chosen based on the residues involved in ubiquinol binding site. The ZINC database compounds were docked to the modeled protein to search for the most potent ligand. We report the top two scoring ligands that were seen to bind with good affinity. These compounds can be evaluated further through experimental investigation for their potency as anti-fungal agents. The structural aspects needed for these compounds to be active can also be studied and enhanced by modification of chemical structure.

5. REFERENCES

[1] M. Kontoyianni , P. Madhav, E. Suchanek, W. Seibel: Theoretical and practical considerations in virtual screening: a beaten field? Current medicinal chemistry, vol.15(2),pp:107-116, 2008.

[2] AW.Sousa da Silva, WF. Vranken: ACPYPE - AnteChamber Python Parser interfacE. BMC Res Notes, 5:367, 2012.

[3] DB Kitchen, H Decornez, JR Furr, J Bajorath: Docking and scoring in virtual screening for drug discovery: methods and applications. Nature reviews Drug discovery, vol. 3(11), pp:935-949, 2004.

[4] C Acharya, A Coop, JE Polli, Jr AD MacKerell: Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. Current computer-aided drug design,vol. O7(1),pp:10, 2011.

[5] M Akamats: Current state and perspectives of 3D-QSAR. Cur Top Med Chem.,vol. 2(12), pp :1381-1394, 2002.

[6] S Ajmani, K Jadhav, SA Kulkarni: Group-Based QSAR (G-QSAR): Mitigating Interpretation Challenges in QSAR. QSAR & Combinatorial Science,vol. 28(1), pp:36-51,2009.

[7] W. C. Snyder and H. N. Hansen, J. Bot. Amer.: The species concept in Fusarium, vol.-27,pp:64-67, 1940.

[8] Alan O. Sykes, An Introduction to Regression Analysis, vol. 46,pp:219-312, 2010.

[9] N. Dwivedi, M. Bhartandu Nath, K. Vishwa Mohan, 2D-QSAR model development and analysis on variant groups of anti-tuberculosis drugs, Bioinformation, vol.-7(2),pp: 82–90,2011;.

[10] W.R. Widger, W.A. Cramert, R.G. Herrmann AND A. Trebsts, Sequence homology and structural similarity between cytochrome b of mitochondrial complex III and the chloroplast b6-f complex: Position of the cytochrome b hemes in the membrane, Proc. Natl. Acad. Sci.,vol.- 81,pp:674-678, USA 1984.