RESEARCH ARTICLE

Target identification and drug interaction studies of *Bacillus anthracis*

Shyam Ingle* and Yogita Rokade

P.G. Department of Bioinformatics, Shri Shivaji Science College, Shivaji Nagar, Amravati-444603.
*Correspondence author Email: bioinfo.sham@gmail.com

Manuscript details:

Available online on
http://www.ijlsci.in

ISSN: 2320-964X (Online)
ISSN: 2320-7817 (Print)

Editor: Dr. Arvind Chavhan

Cite this article as:
Shyam Ingle and Yogita Rokade (2017) Target identification and drug interaction studies of *Bacillus anthracis*, *Int. J. of Life Sciences*, Special Issue, A8:180-187.

ABSTRACT

Target based or structural based drug designing is the rapidly growing area. The explosion of genomic, proteomic and structural information has provided hundreds of new target and opportunity to find new drug lead compounds. Based on burgeoning structural data of *Bacillus anthracis* we focused on anthrax disease. *Bacillus anthracis* is a gram positive, rod shaped bacteria that caused serious infectious anthrax disease. In this study, we have taken 1669 protein sequences from NCBI protein database for which protein structure is available at PDB database. From these 1669 proteins, we have detected 10 druggable and 3 virulence protein sequences using TiD tool. Out of detected virulence protein, we have chosen N5-carboxyaminoimidazole ribonucleotide synthetase as a target. Against this target protein, we have screened Ropinirole and Isatin as a lead molecule for docking studies.

Keywords: *Bacillus anthracis*, drug designing, TiD tool, Chemoinformatics, ligand screening, target identification, Autodock

INTRODUCTION

The oldest isolate of Bacillus anthracis, the causative agent of anthrax, dates back to 1917 (Redmond et al.,1998). Bacillus anthracis, a gram-positive rod shaped bacteria and belongs to the Bacillus cereus group has an extremely monomorphic genome and has high structural similarity with physiological and B. cereus and B. thuringiensis (Pavan et al., 2011). The proteome data is available on The Universal Protein Resource (UniProt) and Protein database at National center for biotechnological information (NCBI) database. UniProt provides a central resource for protein sequences and functional annotation with three database components, each addressing a key need in protein bioinformatics (Wu et al., 2006). Among the total proteome data of B. anthracis, few key proteins are identified as a target. Due to the burgeoning of protein data, there are many proteins are still remaining to explore as a drug target. Insilico approaches are good to explore this data. The target identification is the key step of Insilico drug designing

One of the key protein shortlisted for our study is N 5-carboxyaminoimidazole ribonucleotide synthetase (PurK). Recent research has shown that de novo purine biosynthesis in microbes is different from that in humans. PurK is an enzyme in the purine-biosynthetic pathway that converts 5-aminoimidazole ribonucleotide to N 5-carboxyaminoimidazole ribonucleotide and has been suggested as a potential antimicrobial drug target and it is unique to prokaryotes (Tuntland et al., 2011). The differences in the pathways are centered around the synthesis of 4-carboxyaminoimidazole ribonucleotide (CAIR) which requires the enzyme N(5)-carboxyaminoimidazole ribonucleotide (N(5)-CAIR) synthetase. Humans do not require and have no homologs of this enzyme. In recent study, 2-(2,3-dioxoindol-1-yl)-N,N-diethyl-acetamide compound shows inhibition activity against prokaryots but no inhibition to human bifunctional enzymatic activity (Firestine, S et al., 2009). We have chosen 2-(2,3-dioxoindol-1-yl)-N,N-diethylacetamide compound as a reference molecule for lead screening.

Drug discovery and development is a computational approach which is very intense, lengthy, time-consuming and an interdisciplinary endeavor. In early stages of drug discovery, identification of potential leads with specific interaction to target is very essential and conventionally pharmaceutics adapts wet-lab high-throughput screening (HTS) methods which are high cost and time taking process, an alternative is a computational approach (Cheng et al., 2012). With the advent of genomics, proteomics, metabolomics, Bioinformatics and efficient technologies like combinatorial chemistry, virtual screening, de novo designing and structure-based drug designing have highly revolutionized the process of drug discovery (Lazarova et al., 2008). Present day drug discovery mainly ponders on target based drug designing, which is broadly defined as "single compound acting on a single target to a single disease". Single target based drugs are designed such that lead molecules can promisingly bind to its specific target, reducing the off-target side effects
(Lin et al., 2012).

Chemo-informatics tools present a tremendous potential to advance in silico drug design and discovery, as they serve the integration of information in several levels to enhance the reliability of data outcomes. To name a few, chemical structure similarity searching, data mining/machine learning, panel docking, and bioactivity spectra based algorithms have been routinely and successfully implemented (Katsila et al., 2016). Some examples are the ligand-based interaction fingerprint (LIFt) approach (Cao et al., 2015). in predicting potential targets for small-molecule drugs using physics-based docking and sampling methods and the protein ligand interaction fingerprints (PLIF) method (Eberini et al., 2011). for summarizing interactions between ligands and proteins using a fingerprint scheme.

TiD is a standalone application, which relies on the basic assumption that a protein must be essential for pathogens survival and non-homologous with the host to qualify as a putative target. With an input bacterial proteome, TiD removes paralogous proteins, picks essential ones, and excludes proteins homologous with host organisms. The targets illustrate non-homology with at least 40 out of 84 gut microbes, considered safe for human. TiD classifies proposed targets as known, novel and virulent. Users can perform pathway analysis, choke point analysis, interactome analysis, subcellular localization and functional annotations through web servers cross-referenced with the application. Drug targets identified by TiD for Listeria monocytogenes, Bacillus anthracis, and Pseudomonas aeruginosa have revealed significant overlaps with previous studies (Gupta et al., 2017).

After target identification ligand screening is the next step. For this purpose, we refer ExPASy web server. ExPASy server is the online portal which provides access to scientific databases and software tools (i.e., resources) in different areas of life sciences including proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics etc. Swiss Similarity is online web based tool for ligand screening using the similarity based method. In this work, two ligand molecules were predicted against a target protein. The last step of drug designing is target ligand interaction studies. The autodock is the best freely available docking software.

Since its release in 1990 (Goodsell et al., 1990), AutoDock has proven to be an effective tool capable of quickly and accurately predicting bound conformations and binding energies of ligands with macromolecular targets (Morris et al., 1998; Huey et al., 2007; Goodsell et al., 1996). In order to allow searching of the large conformational space available to a ligand around a protein, AutoDock uses a grid-

based method to allow rapid evaluation of the binding energy of trial conformation
(Morris et al., 2009).

MATERIAL AND METHODS:

Data Collection

Protein sequence data was collected from NCBI protein database. NCBI database is available at https://www.ncbi.nlm.nih.gov. At NCBI database select protein database and search the query as organism name i.e. *Bacillus anthracis*. After that, the PDB protein sequences were sorted by filtering sequences from filter option and we got 1669 protein sequences. All selected 1669 sequence was downloaded in multiple sequence FASTA file format.

Methodology

The primary aim of this work is to identify the Target proteins from collected *Bacillus anthracis* proteome. Target proteins are those proteins which should be druggable and capable to produce virulence. To find target proteins TiD tool was used which is available at http://bmicnip.in/TiD/. This tool is a standalone tool and no need to do any installation. For this tool, the direct executable file is available only need to manage some data files. Other software's like.Net framework, Python2.7.10, biopython-1.65 for python2.7.10, NCBI BLAST 2.2.31+ were installed.

Following steps were used for target identification using TiD tool.

1. Paralog analysis.
2. Essentiality analysis.
3. Nonhomologous analysis.
4. Target Prioritization.

Paralog analysis:

Paralog analysis was performed CD-HIT at 60% identity to remove redundant paralogous sequences. FASTA file containing 1669 protein sequences were uploaded for this step. Paralog analysis step gives 250 sequences.

Essentiality analysis:

Essentiality analysis looks for pathogen specific essential genes in DEG, CEG and common from both DEG & CEG based on threshold E-values and bit scores.

1. Database of Essential Genes (DEG): It finds essential genes of the pathogen from DEG database.
2. Clustering of Essential Genes (CEG): It finds essential genes of the pathogen from CEG database.
3. Common of DEG and CEG: It finds common essential genes of the pathogen from DEG and CEG database.

In this step common of DEG and CEG analysis were performed. For this step, Essential analysis output file was used. E-value and Bit Score was selected as the default value. The default E-value is 10-5 and Bit Score is 100(same for CEG & Common for DEG and CEG). After common of DEG and CEG analysis, 101 essential gene sequences were obtained.

Nonhomologous analysis:

Nonhomologous analysis helps in identifying non-homolog proteins of pathogenic bacteria in different hosts and gut flora based on threshold E-values and bit scores. The default E-value is 0.005 and Bit Score is NONE. For this step select human as host and 40% threshold value was used for gut flora. After this analysis, 34 protein sequences were obtained.

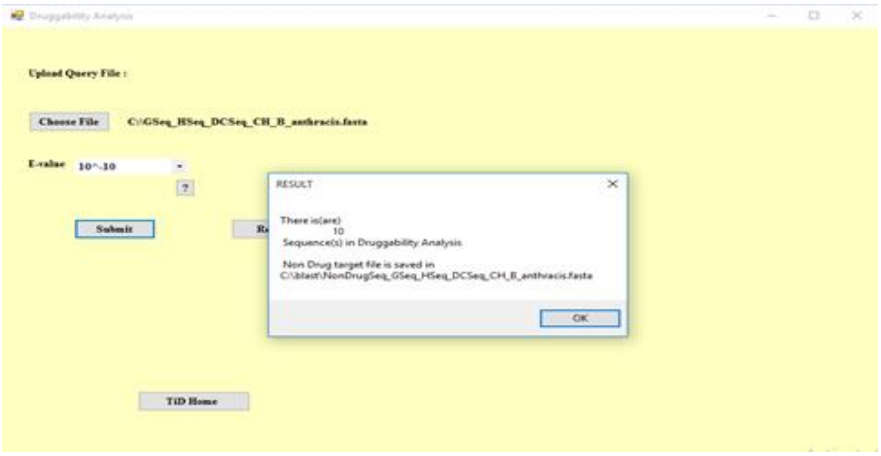

Fig1: Druggability analysis results

Fig2: Identified Druggable Target sequences

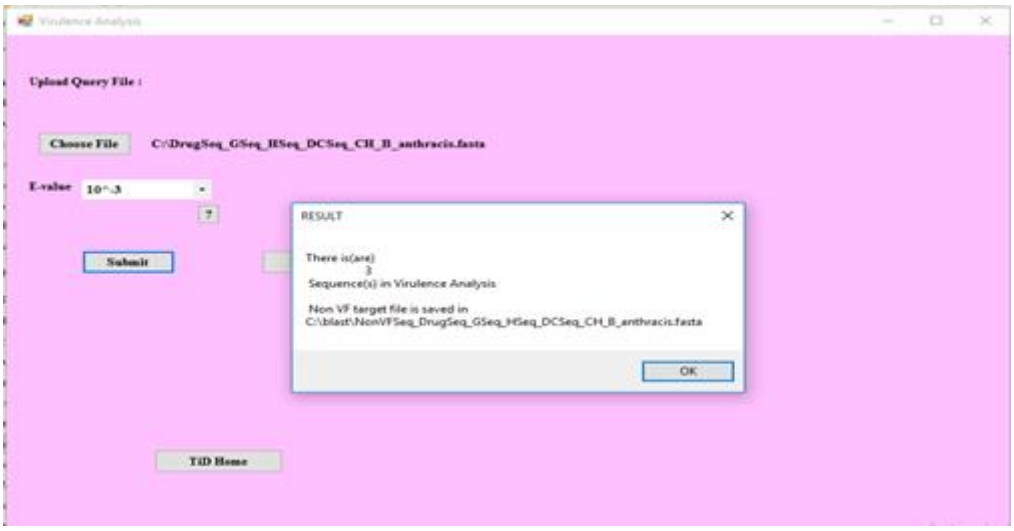

Fig3: Virulence analysis result

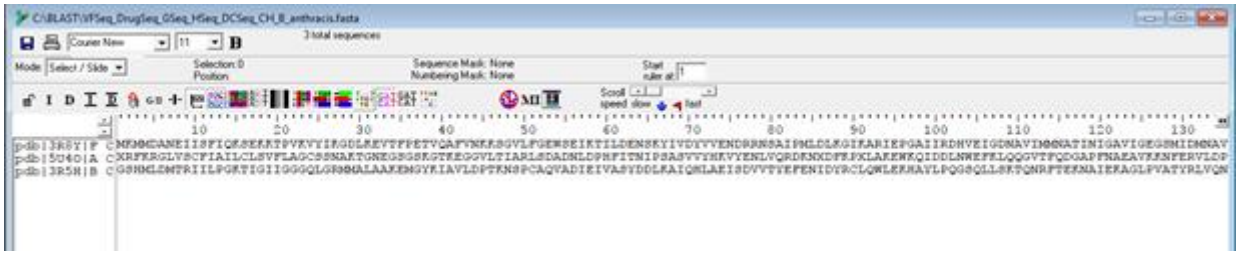

Fig4: Virulence target sequences

Target Prioritization:

Target Prioritization finds homology with druggable target and virulence target. For druggable analysis and virulence target, the default E-value is 10^-10 was used to optimize the result. After analysis 10 druggable and 3, virulence targets was Identified

Docking Studies:

Binding site Prediction:

Among three druggable targets and only one virulent, *N*5-carboxyaminoimidazole ribonucleotide synthetase protein was selected for docking suites. Primarily Protein structure was collected from PDB database and structure file saved as a 3R5H.pdb file. To check the binding site or active site for this protein POCASA tool was used. Using this tool 5 top score binding pockets were identified and top scored pocket coordinates were selected for drug interaction studies.

Protein Preparation:

Downloaded pdb file contained some salts, compounds and water molecules. Protein preparation is the key step in docking study in which all water molecules and other compounds are removed from protein structure file. LidDig online tool was used to remove all salts and

compounds from pdb file. This polish file further cleaned by removing of a water molecule using Auto dock tool.

Lead Identification:

Selected 2-(2,3-dioxoindol-1-yl)-N,N-diethylacetamide compound shows inhibition activity against prokaryots but no inhibition to human bifunctional enzymatic activity. We have chosen 2-(2,3-dioxoindol-1-yl)-N,N-diethylacetamide compound as a reference molecule for lead screening.

Lead compound was identified based on structure similarity search. The 2-(2,3-dioxoindol-1-yl)-N,N-diethylacetamide molecule was screened using ExPaCy's SwissSimilarity tool for which fingerprint algorithm was used against approved and experimental drug database. The highest similar ligand with reference compound was identified and selected for docking. Ropinirole (Approved) and Isatin (Experimental) drug compound were selected based on fingerprint algorithm. Similarity score for these compounds are 0.57 and 0.75 respectively.

Docking Protocol:

The AutoDock suite is used in numerous laboratories: a recent search on PubMed yielded over 1000 citations in the past year. For this work, we used Autodock 4.0 and following protocol was applied to check target ligand interaction.

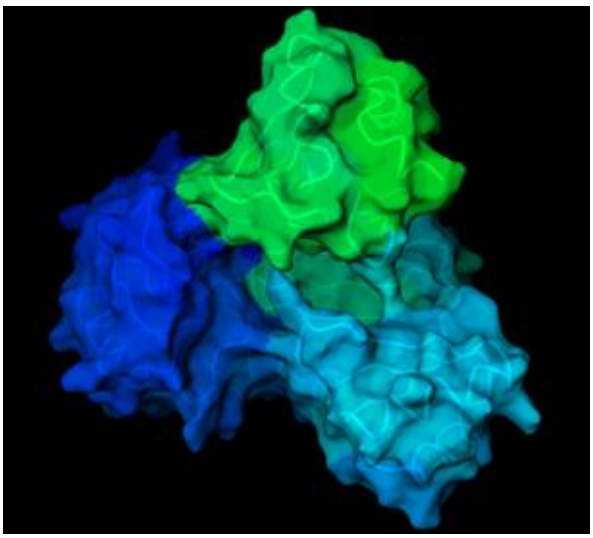

Fig5:N5-carboxyaminoimidazole ribonucleotide synthetase protein cavities

Protocol

1) Ligand preparation

Ligand → input → open (choose lead.pdb)
Ligand → torsion tree → set number of torsion (fever number set to 6)
Ligand → output → save as PDBQT (lead.pdbqt)

2) Protein preparation

Read molecule → open (3R5H.pdb)
Select → select from string (residue HOH*; atom *)
Edit → delete → delete selected atoms
Edit → hydrogen → add (add hydrogen)
Grid → macromolecule → choose (3R5H.pdb)
→ save (3R5H.pdbqt)

3) Grid parameter file

Grid → Grid box (60,60,60) (-12,-8,35) → file → close saving current
Grid → set map types → choose ligand → select ligand
Grid → output → save GPF (3R5H.gpf)

4) Docking parameter file

Docking → macromolecule → set rigid → filename
→ open (3R5H.pdbqt)
Docking → ligand → choose → directly
Docking → search parameters → genetic algorithm → accept
Docking → output → Lamarckian GA → save (lead.dpf)

5) Running docking program in Command prompt.

Autogrid4.exe –p 3R5H.gpf -l 3R5H.glg
Autodock4.exe –p lead.dpf –l lead.dlg

6)

7) Analyzing docking result

Analyze ➔ docking ➔ open (lead.dlg)
Analyze ➔ micromolecule ➔ open (3R5H.pdbqt)
Analyze ➔ conformations ➔ play ranked by energy
Analyze ➔ docking ➔ show interactions

)

RESULTS

In this docking studies best docking poses and score were obtained for Ropinirole and Isatin are shown in fig 6-9. We also compared obtained score of Ropinirole and Isatin with reference compound score and it shows almost similar binding energy score.

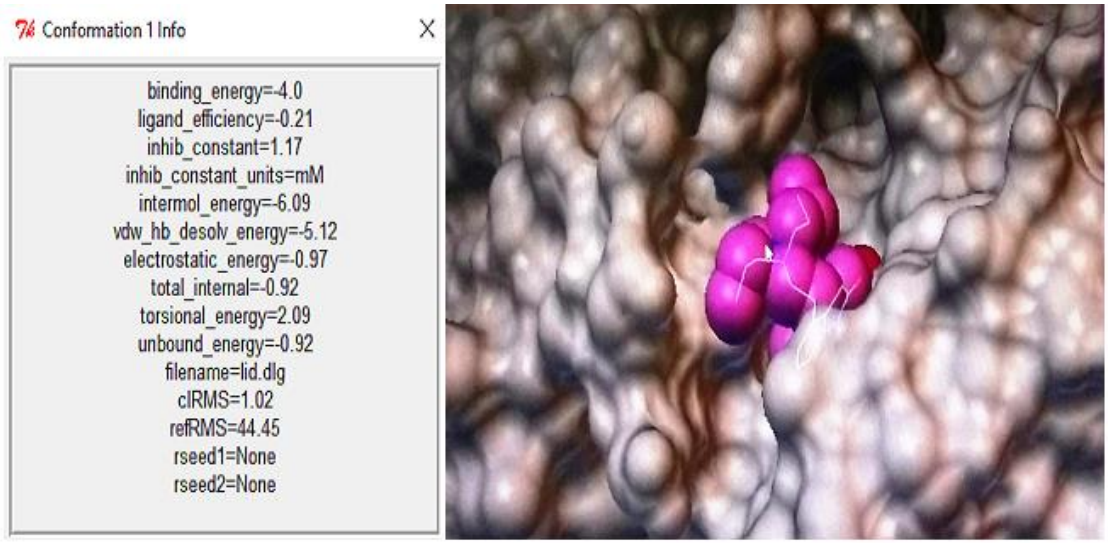

Fig6: Ropinirole - Best Docking pose 1 and docking score

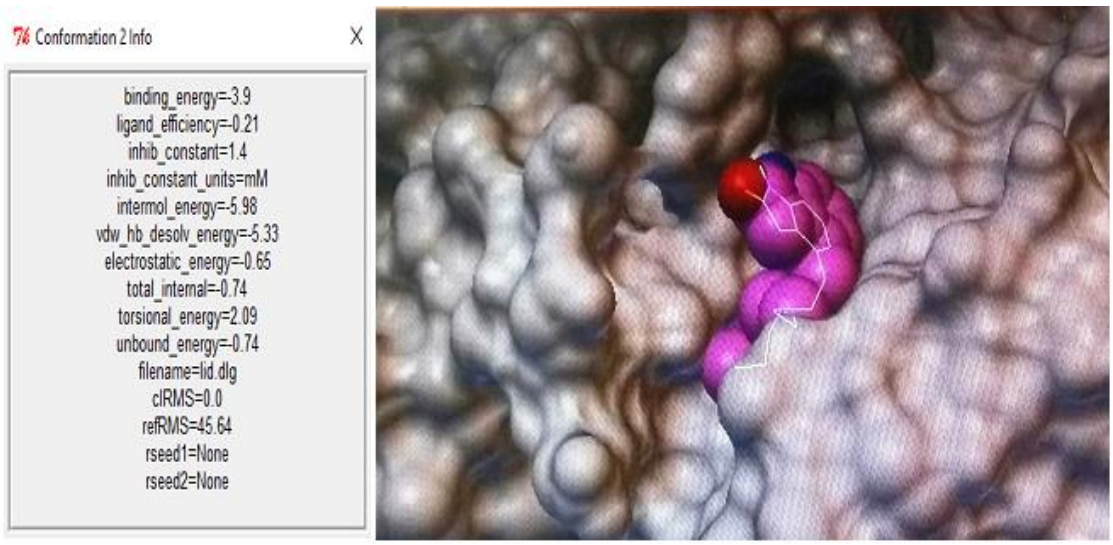

Fig7: Ropinirole - Best Docking pose 2 and docking score

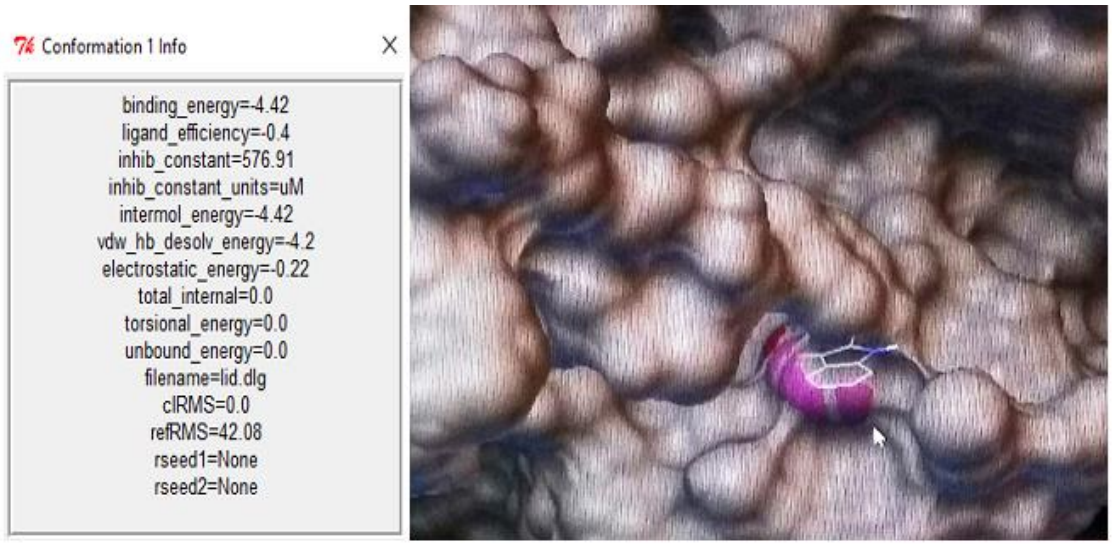

Fig8: Isatin - Best Docking pose 1 and docking score

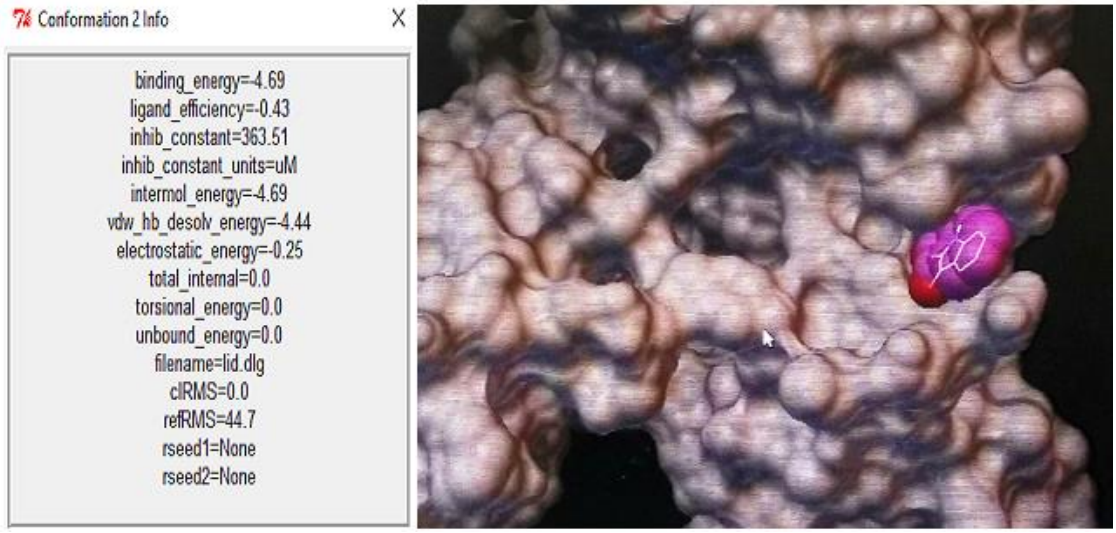

Fig9: Isatin - Best Docking pose 2 and docking score

CONCLUSION

Antibiotic resistance has been a significant increase during the past decade. The increasing frequency of drug resistant bacterial infection as amplified the need for novel antimicrobial agents. Previous study has shown that N 5-carboxyaminoimidazole ribonucleotide synthetase is a key intermediate in purine biosynthesis in bacteria , fungi but not in humans. In our study we indentified N 5-carboxyaminoimidazole ribonucleotide synthetase as a virulent as well as druggable protein using novel approached i.e. TiD. For this target 2-(2,3-dioxoindol-1-yl)-N,N-diethylacetamide compound screened against approved drug and experimental database using fingerprint algorithem. In the screening procedure Ropinirole (approved) and Isatin (experimental) drug were indentify with good similarity score. Further, we have studied interaction of selected target and screened lead compound which results good score with lowest possible binding energy and ligand efficiency. This result shows both Ropinirole and Isatin compounds can be acts as a good antimicrobial drug candidate. The activity of this ligand against target protein can be optimized by wet lab experiment.

Conflicts of interest: The authors stated that no conflicts of interest.

REFERENCES

Redmond, C, Pearce, MJ, Manchee, R J, & and Berdal, B P (1998) Deadly relic of the Great War Nature; 393(6687): 747.

Pavan ME, Pettinari MJ, Cairó F, Pavan EE, Cataldi AA (2011) Bacillus anthracis: a molecular look at a famous pathogen Rev Argent Microbiol; 43(4):294-310.

Wu, C H, Apweiler, R, Bairoch, A, Natale, DA, Barker, WC, Boeckmann, B, & Magrane, M (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information Nucleic acids research; 34(suppl_1): D187-D191.

Tuntland, ML, Johnson, ME, Fung, M, & Santarsiero, BD (2011) Structure of N5-carboxyaminoimidazole ribonucleotide synthase (PurK) from Bacillus anthracis Acta Crystallographica Section D: Biological Crystallography; 67(10): 870-874.

Firestine, SM, Paritala, H, Mcdonnell, JE, Thoden, JB, & Holden, HM (2009) Identification of inhibitors of N 5-carboxyaminoimidazole ribonucleotide synthetase by high-throughput screening Bioorganic & and medicinal chemistry; 17(9): 3317-3323.

Cheng, T, Li, Q, Zhou, Z, Wang, Y, & and Bryant, SH (2012) Structure-based virtual screening for drug discovery: a problem-centric review The AAPS journal; 14(1): 133-141.

Lazarova, M (2008) Virtual screening models; methods and software systems In International Scientific Conference Computer Science (Vol 1; p 55).

Lin, X, Huang, X P, Chen, G, Whaley, R, Peng, S, Wang, Y, & and Huang, N (2012) Life beyond kinases: structure-based discovery of sorafenib as nanomolar antagonist of 5-HT receptors Journal of medicinal chemistry; 55(12): 5749-5759.

Katsila, T, Spyroulias, GA, Patrinos, GP, & and Matsoukas, MT (2016) Computational approaches in target identification and drug discovery Computational and structural biotechnology journal; 14; 177-184.

Cao, R, & and Wang, Y (2015) In silico study of polypharmacology with ligand-based interaction fingerprint Receptors & and Clinical Investigation; 2(4).

Eberini, I, Daniele, S, Parravicini, C, Sensi, C, Trincavelli, ML, Martini, C, & and Abbracchio, MP (2011) In silico identification of new ligands for GPR17: a promising therapeutic target for neurodegenerative diseases Journal of computer-aided molecular design; 25(8): 743.

Gupta, R, Pradhan, D, Jain, AK, & and Rai, CS (2017) TiD: Standalone software for mining putative drug targets from bacterial proteome Genomics; 109(1): 51-57.

Goodsell, DS, & and Olson, AJ (1990) Automated docking of substrates to proteins by simulated annealing Proteins: Structure; Function; and Bioinformatics; 8(3): 195-202.

Morris, GM, Goodsell, DS, Halliday, RS, Huey, R, Hart, WE, Belew, RK, & and Olson, AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function Journal of computational chemistry; 19(14): 1639-1662.

Huey, R, Morris, GM, Olson, AJ, & and Goodsell, DS (2007) A semiempirical free energy force field with charge-based desolvation Journal of computational chemistry; 28(6): 1145-1152.

Goodsell, DS, Morris, GM, & and Olson, AJ (1996) Automated docking of flexible ligands: applications of AutoDock Journal of Molecular Recognition; 9(1): 1-5.

Morris, GM, Huey, R, Lindstrom, W, Sanner, MF, Belew, RK, Goodsell, DS, & and Olson, AJ (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility Journal of computational chemistry; 30(16): 2785-2791.