Virtual screening and docking studies of identified potential drug target: Polysaccharide deacetylase in *Bacillus anthracis*

K. Zaveri*, A. Krishna Chaitanya, I. Bhaskar Reddy

Department of Biochemistry and Bioinformatics, Institute of Science, GITAM University
Visakhapatnam-530 045, Andhra Pradesh, India

*E-mail address: kunal.zaveri22@gmail.com

ABSTRACT

In recent years, *insilico* approaches have been predicting novel drug targets. The present day development in pharmaceutics mainly ponders on target based drugs and this has been aided by structure based drug designing and subtractive genomics. In the present study, the computational genome subtraction methodology was applied for identification of novel, potential drug target against *Bacillus anthracis,* cause of deadly anthrax. The potential drug target identified through subtractive genomics approach was considered as polysaccharide deacetylase. By virtual screening against NCI database and Drugbank chemical libraries, two potential lead molecules were predicted. Further the potential lead molecules and target protein were subjected for docking studies using Autodock.

Keywords: Anthrax; Bacillus anthracis; Molecular docking; Virtual screenin

1. INTRODUCTION

Drug discovery and development is an 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 [1]. 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 [2].

Bacillus anthracis, a gram-positive rod shaped bacteria belonging to the *Bacillus cereus* group has an extremely monomorphic genome, and has high structural similarity with physiological and *B. cereus* and *B. thuringiensis* [3]. The virulent nature of *B.anthracis* is because it harbors to plasmids, pXO1 of 181 kb and pXO2 of 93.5 kb [4]. *B.anthracis* mostly constitutes two major toxins which are composed of three different proteins, protective antigen, lethal factor and edema factor. The structural genes for these components are located on the virulent plasmid pXO1 [5]. The plasmid pXO2 encodes an antiphagocytic capsule composed of a high molecular-weight polypeptide (poly-D-glutamic acid) which inhibits the

host phagocytosis of the vegetative form of *B.anthracis.* These proteins are individually nontoxic, a fully virulent bacilli requires the presence of both an antiphagocytic capsule and three plasmids encoded toxic components, for producing toxic responses in the host organism like edema and cell death [6].

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 [7]. In the present study, to identify the specific potential drug targets in *B. anthracis*, the subtractive genomics approach was applied. Application of *insilico* subtractive genomics approach nullifies the host toxicity which is common in conventional approaches for drug development against bacterial diseases [8]. Here a series of subtractive genomic methods are used for analyzing the complete proteome of *B. anthracis* to avoid the paralogous proteins and homologous proteins to humans. Structure based designing an important aspect of computer aided drug designing have revolutionized the drug development in pharmaceutics as it have resulted many fruitful success in molecular docking [9].

2. METHODOLOGY

The complete proteome information of *B. anthracis* was available in UniProtKB [10] database which is a central hub for completely annotated functional proteins. A series of subtractive genomics methods are applied to identify the potential drug targets. Subtractive genomics involve two steps, identification of essential genes to pathogen and they being non-human homologous. Primarily protein sequences retrieved from the UniProtKB database were subjected to the Database of Essential Genes (DEG) [11], selecting BLASTP (Protein query vs. Protein Database) with default parameters for analyzing the essential proteins required for survival of pathogen. These identified essential proteins in pathogens to be selected as a potential drug target should be non-human homologs and hence these protein sequences are subjected to H-BLAST [12].

The 3D structure of proteins is essential to understand the active sites of the proteins that cause pathogenesis. The potential drug targets identified based on the subtractive genomics method are subjected for identification of 3D structures to PDB-BLAST. From the dataset obtained, the best potential drug target is identified as Delta-lactam-biosynthetic de-N-acetylase protein. The selection of drug target was based on the following criteria:

o Target structures (experimental data/name based) availability
o Target sequences identities/similarities to that of PDB structures
o Targets possessing the feature of Essentiality and Non-human homologs

The sequence percentage identity of the Delta-lactam-biosynthetic de-N-acetylase (polysaccharide deacetylase) with the PDB-BLAST was 99% and it was also found as essential protein and non-human homologue (absent in humans). Hence, the structure of polysaccharide deacetylase with PDB ID. 2J13 is selected for further studies.

For macromolecules, the active sites and binding sites are associated with structural cavities and pockets and is the region where one molecule binds to another. Active sites of potential drug targets are predicted by CASTp (Computed Atlas of Surface Topography of proteins) [13] plugin in PyMOL for better visualization of the pockets.

PyRx [14] is a graphical user interface for AutoDock 4.2 and AutoDock Vina to perform virtual screening. It provides binding affinities and RMSD scores for each ligand with nine different poses. A library of ligands is generated from NCI Diversity Data Set II and Drug Bank and are subjected for virtual screening against identified potential drug targets using PyRx. Docking is a method that predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. AutoDock [15] is a suite of automated docking tools. It is used to predict how small molecules bind to a receptor of known 3D structure. AutoDock 4 consists of two main programs: *autodock that* which performs docking of a ligand to a set of grids describing the target protein; *autogrid* pre-calculates these grids. For every drug target from virtual screening a set of certain potential lead molecules are identified in this molecule are subjected for docking using AutoDock. Primarily the protein and ligand molecules are prepared, optimized, grid is generated near the active site of protein molecule and then docked. Then the docked poses are analyzed in chimera.

3. RESULTS AND DISCUSSION

Delta-lactam-biosynthetic de-N-acetylase protein (PDB ID: 2J13) plays an essential role in an endospore life cycle of the bacteria. In structure based drug designing studies, the knowledge of protein active site is vital to generate a grid on the surface of protein for efficient docking. Using CASTp server about 35 binding pockets are predicted for 2J13 of which the best binding pocket is represented in figure 1.
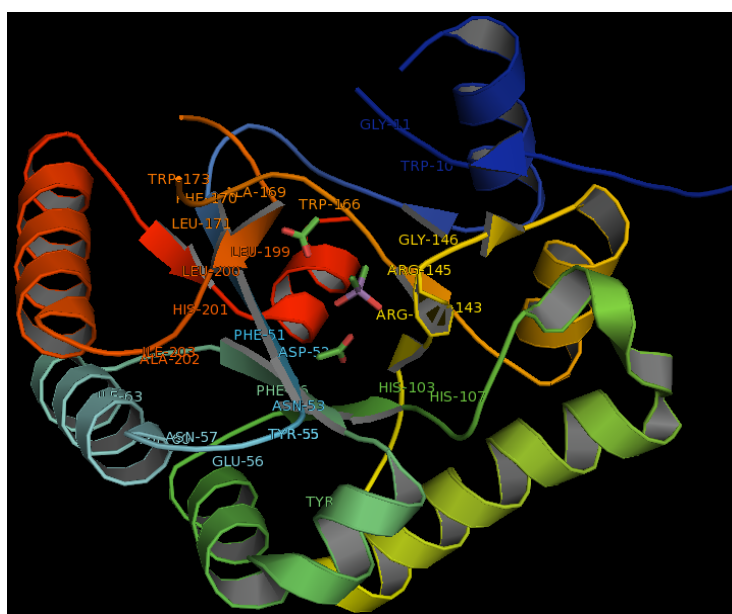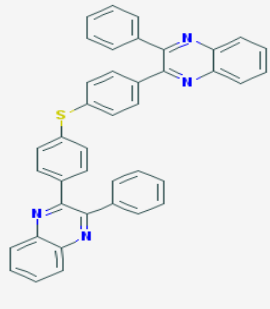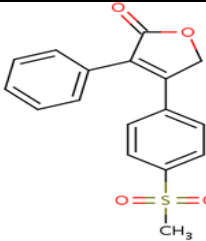

Figure 1. Active site predicted for 2J13 protein using CASTp.

For identification of potential lead molecules, the NCI Diversity dataset and Drugbank ligand libraries were screened using PyRx tool. For each ligand, 10 poses were generated based on binding affinity. The lead molecules with least binding affinity from the ligand library of NCI diversity data set & drugbank database were found as 2-phenyl-3 [4-[4-(3-

phenyl quinoxalin-2-yl) phenyl) sulfanyl phenyl] quinoxaline (NCI_293778) and Rofecoxib respectively (Table 1) which were selected for further docking studies.

Table 1. Structure of Potential Lead Molecule Predicted by Virtual Screening.

Ligand Library	Potential Lead predicted based on Binding Affinity	2 Dimensional representation
NCI Diversity Data set	2-phenyl-3[4-[4-(3-phenyl quinoxalin-2-yl)phenyl) sulfanyl phenyl] quinoxaline (NCI_293778)	
Drug Bank	Rofecoxib	

The potential lead molecules were subjected to molecular docking process using AutoDock program. Prior to docking, proteins and ligands were prepared by adding Kollman and Gasteiger charges. For better docking results in ligand preparation step, the ligand molecule was energy minimized and the number of rotational bonds were set to 6, if the ligand constitutes more than 32 rotational bonds. The ligand molecules features after ligand preparation step were shown in table 2.

Table 2. Ligand Preparation.

Ligand	Gasteiger charges	Merged Non-polar Hydrogen	Aromatic Carbons	Detected Rotatable Bonds	Set TORSDOF
NCI_293778	Added	-	40	6	6
Rofecoxib	Added	-	17	10	10

Table 3. Grid Generated for 2j13 Protein.

Target Protein	Center Grid Box			Spacing Angstroms	No. of Points in Dimensions			Total Grid Points per Map
	X	Y	Z		X	Y	Z	
2J13	28.832	24.045	13.147	0.375 Å	40	40	40	64000

A prepared 2J13 protein molecule with added Gasteiger and Kollman charges was subjected to grid preparation near the binding pocket as predicted by CASTp server. The XYZ coordinates for grid to 2J13 molecule were represented in the table 3. Further, based on the grid data, the ligand molecules were subjected to docking and the results were analyzed using the AutoDock analysis option to visualize ten different docked poses. Based upon the least binding affinity and other parameters, the best pose was selected and the docked structures were visualized in Chimera software (Figure 2) for detailed residue-ligand interactions. Docking of NCI_293778 and Rofecoxib with 2J13 represents that ligands were effectively bound by interacting with ASP-52, ARG-145, GLY-146 and TRP-166 which are the residues in the active site of 2J13. The detailed docking results are shown in the table 4.

Table 4. Docking Results of Target protein 2J13.

Target	2J13	
Ligand	Rofecoxib	NCI_293778
Binding Energy	-1.89	5580.0
Ligand Efficiency	-0.08	124.0
Inhib Constants	48.17mM	-
Inter mol Energy	-4.78	5570.0
VDW_hb_desolv Energy	-4.67	0.31
Electrostatic Energy	-0.11	-
Total Energy	0.00	259.07
Torsional Energy	2.98	1.79
Unbound Energy	0.0	259.07
ClRMS	1.53	0.0
refRMS	42.29	41.91
rseed1 and 2	None	None

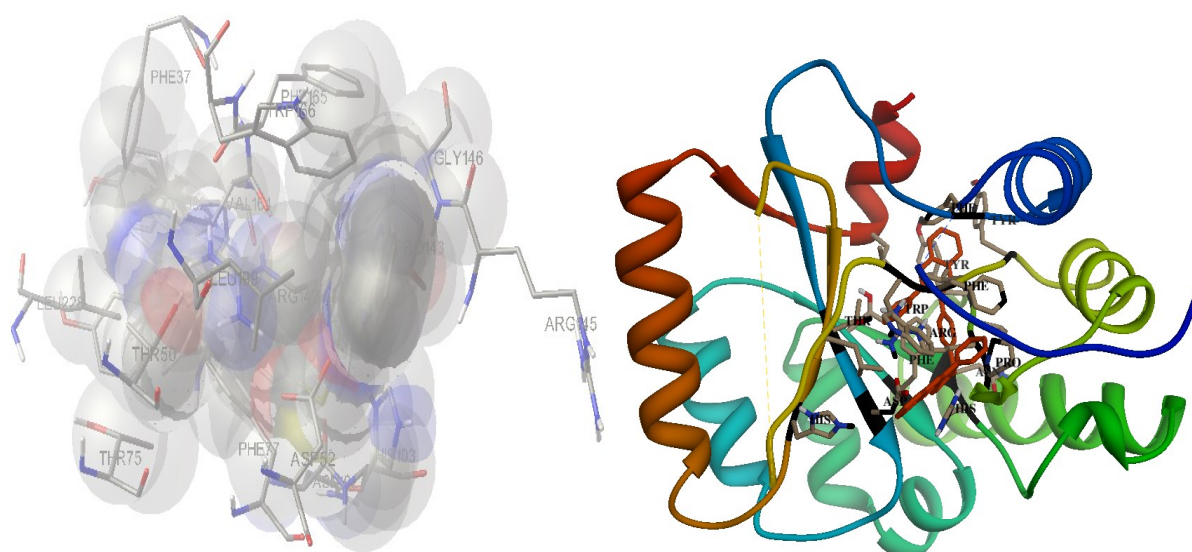**Figure 2.** Visualization of Docking between 2J13 and NCI_293778.
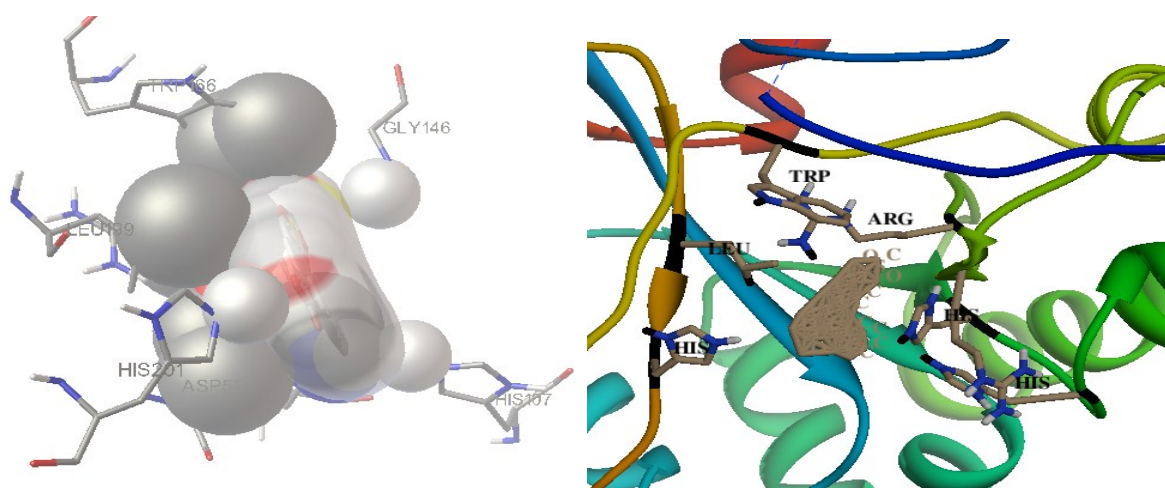

Figure 3. Visualization of Docking between 2J13 and Rofecoxib.

4. CONCLUSION

The modern Computational (Bioinformatics) applications in the field of biology like Subtractive Genomics, Homology modeling, Virtual Screening and Molecular Docking have provided a new path in the field of Drug Discovery, Designing and Development with cost-effective and timely analyzing process in Whole Genome analysis for Target identification, Modeling the 3D structures for the target molecule which doesn't possess experimental data screening of large chemical databases helps in lead identification of ligand-residue interactions. In the present study, the target protein Delta-lactam-biosynthetic de-N-acetylase protein (PDB ID: 2J13) was screened against NCI and DrugBank databases. In this screening process, potential ligands for the target were identified based on least binding affinites and are then subjected further to docking studies. In Docking studies, the 2J13 target protein, the best scored ligands were identified as NCI-293778 and Rofecoxib which share the common active

site residues as that of residues predicted in CASTp, which can be inferred that the docked conformations are good and acceptable.

5. FUTURE PROSPECTS

From the current study, the future prospectives can be directed towards the study of pharmacokinetics (ADMET properties), 2D and 3D QSAR analysis and further validation of identified lead molecules can be done through High Throughput Virtual Screening (HTVS) and can be leaded towards clinical trials if the results are satisfactory.

References

[1]   Cheng T, Li Q, Zhou Z, Wang Y, and Stephen HB. *The AAPS Journal*; (2012), 14(1):133-141

[2]   Milena L. *International Scientific Conference Computer Science*; (2008)

[3]   Mary EP, Mary JP, Fabian C, Stephen EP, Angel AC. Argentina J Microbiol (2011);43(4):294-310.

[4]   Erlendur Helgason, Ole Andreas Okstad, Dominique AC, Henning A. Johansen, Agnes fouet, Miche´ le Mock, Ida Hegna, and Anne-brit kolsto. *Appli and Env Microbiolo* (2000) ;2627-2630

[5]   Sirard JC, Mock M, Fouet A, *J Bacteriol* (1994);176(16):5188-5192

[6]   Stefan Riedel.. *Proc (Bayl Univ Med Cent)* (2005);18(3): 234–243.

[7]   Xingyu Lin, Xi-Ping Huang, Gang Chen, Ryan Whaley, Shiming Peng, Yanli Wang, Guoliang Zhang, Simon X. Wang, Shaohui Wang, Bryan L. Roth, and Niu Huang. *J Med Chem* (2012) ;55(12):5749–5759.

[8]   Hosen MI, Tanmoy AM, Mahbuba DA, Salma U, Nazim M, Islam MT, Akhteruzzaman S.. *Interdiscip Sci.* (2014) ;6(1):48-56.

[9]   Lu Chen, John K. Morrow, Hoang T. Tran, Sharangdhar S. Phatak, Lei Du-Cuny, and Shuxing Zhang. *Curr Pharm Des.* (2012) ;18 (9):1217–1239.

[10]   The UniProt Consortium. Activities at the Universal Protein Resource (UniProt). *Nucl Acids Res* (2014) ;42: 191-198.

[11]   Hao Luo, Yan Lin, Feng Gao, Chun-Ting Zhang and Ren Zhang. *Nucl Acids Res* (2013); 42;574-580.

[12]   Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ. *J. Mol. Biol.* (1990); 215:403-410.

[13] Joe Dundas, Zheng Ouyang, Jeffery Tseng, Andrew Binkowski, Yaron Turpaz, and Jie Liang..*Nucl. Acids Res* (2006) ;34:116-118.

[14] Wolf LK. *C&EN* (2009) ;87:32.

[15] Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS and Olson AJ. *J. Comp Chem* (2009);16:2785-91.