Molecular Modeling and Docking Study of 2-Nitropropane Dioxygenase of Mycobacterium tuberculosis

Abid AM*, Ibrahim BS1, Yadav PK1, Arya H1 and Rasool A2
1Centre for Bioinformatics, Pondicherry University, Puducherry
2Indian institute of Chemical Technology, CSIR-IICT, Hyderabad

Abstract

Mycobacterium tuberculosis is infectious bacteria and causes tuberculosis in humans. M. tuberculosis infects the immune deficient human and shows the symptoms of the infection. Also bacteria stay in latent phase inside the human body and can be active in suitable conditions. One third of the total population of the world is infected by M. tuberculosis; therefore it is very important to have potential drugs against tuberculosis. Mycobacteria have reported multi-drug resistance to the available drugs for tuberculosis. Hence, there is a need to find a new target for the drugs. Fatty acid synthase II (FAS II) is the enzyme that catalyzes the synthesis of fatty acid and is not found in humans. It is a multifunctional polypeptide, composed of different domains in which can be targeted individually to inhibit the function of FAS II. 2-Nitropropane Dioxygenase is a part of enoyl reductase domain in FAS II and can be potentially targeted. In this study, the homology modeling of 2NPD from M. tuberculosis has been done and small molecules that have the potential to bind and inhibit the function of the enzyme have been identified. Also the stability of protein-ligand complex was determined.

Keywords: Mycobacterium tuberculosis; Fatty acid synthase II; 2-Nitropropane dioxygenase

Introduction

TB is an abbreviation of Tuberculosis. It is an infection caused by the bacterium Mycobacterium tuberculosis. Research has revealed many strains of this pathogen, the most infectious strain of Mycobacterium tuberculosis being - H37Rv strain [1]. TB is contagious that can spread through the lymph nodes and bloodstream to any organ in our body, and is mostly found in the lungs. One time exposure to this disease prevents from being affected again, since the bacteria stays back in the body but in an inactive form. But if the immune system weakens, such as in people with HIV or elderly adults, TB bacteria can become active [2].

M. tuberculosis is transmitted through the air, and this contagious disease is more likely to occur if a person is exposed to someone else with TB on a day-to-day basis, such as by living or working in close quarters with someone who has the active disease. Even then, because the bacteria generally stay inactive after they attack the body, just a little number of individuals infected with TB will ever have the active disease. The remaining will have what's called latent TB infection where they show no signs of infection and won't be able to spread the disease to others, unless the bacterium becomes active. Because these latent infections can eventually become active, even people without symptoms should go for the medical treatment. Medication can have the chance to get rid of the inactive bacteria before they become active.

Tuberculosis can also affect other parts of the body, including the kidneys, spine or brain. When tuberculosis found at outer part of your lungs, signs and symptoms can be differ according to the organs involved. For example, tuberculosis of the spine may give back pain, and that in kidneys might cause blood in urine.

Mycobacterium tuberculosis

The mycobacteria are a sort of germ. It is in wide range. In which the most common one causes the tuberculosis. Mycobacteria are aerobic and non-motive bacteria (except for the species Mycobacterium marinum, which has been appeared to be motile inside macrophages) that are characteristically acid-fast bacteria.

All species of mycobacteria have rope like structures of peptidoglycan that are arranged in such a way to give them property of acid fast bacteria [3].

Mycobacteria abundant in soil and water, but Mycobacterium tuberculosis is mainly identifies as a pathogen that lives in the host. The most commonly used strain of Mycobacterium tuberculosis is the H37Rv strain.

Target protein

Fatty acid synthase (FAS): FAS is an enzyme that catalyzes the formation of palmitate from acetyl-CoA and malonyl-CoA in the presence of NADPH into saturated fatty acids. There are two classes of FAS, They are- FAS I, which is a multifunctional polypeptide and found mainly in mammals and fungi and FAS II, being a group of discrete mono functional enzymes for the synthesis of fatty acids. These are found in bacteria and archaea. Although the structural organizations of FAS I and FAS II are different, chemical reactions and catalytic mechanisms for fatty acid synthesis are essentially the same. The absence of FAS II system in humans, it is considered as a potential drug target against M. tuberculosis. FAS in the bacteria have different domains. In the N-terminal there are three catalytic domains: ketoacyl synthase, malonyl/acyltransferase, and dehydrase. Whereas in C-terminal there are four catalytic domains: enoyl reductase, ketoacyl reductase, acyl carrier protein and thioesterase. 2-Nitropropane dioxygenase is a part of the enoyl reductase domain [4].
Nitropropane dioxygenase (2NPD): 2NPD belongs to the family of oxidoreductase and catalyzes the conversion of 2-nitropropane into acetone and nitrite in the presence of oxygen. It is essential for fatty acid synthesis and hence a potential target for the drugs against tuberculosis. The structure of Mycobacterium tuberculosis 2NPD enzyme is unknown [5,6].

Materials and Methods

Tools

**Modeller:** One of the most widely used tool for homology or comparative modeling of protein three-dimensional structures is MODELLER which is a program for automated protein Homology Modeling. The only con of it is that it is command line based where the users find it a bit difficult to start with it [7].

**PyMol:** PyMOL is free copyrighted software for all to use and modify. It is used to understand the structural biology, visualization of complex macromolecular structure of protein at its amino acid level, meaning a protein sequence and its structures can be explored in research using this tool. The viewing of Secondary and tertiary structures of proteins, also viewing the amino acids and nucleic acids can be done using PyMol [8].

**GROMACS:** GROMACS stand for Groningen Machine for Chemical Simulations. The GROMACS project was originally started in 1991 at Department of Biophysical Chemistry, University of Groningen, Netherlands (1991-2000). The goal was to construct a dedicated parallel computer system for molecular simulations, based on ring architecture. The molecular dynamics specific routines were rewritten in the C computer system for molecular simulations, based on ring architecture.

**Schrodinger:** Schrodinger is a software which includes various applications for different *in silico* drug design process. Schrodinger drug designing software uses both ligand and structure based methods. In which Maestro provides a powerful molecular modelling environment.

Molecular docking

All docking studies were conducted using the Glide module of maestro software by Schrodinger. Docking consists of four steps: Protein Preparation, Ligand Preparation, Receptor Grid Generation, and Ligand Docking.

**Chimera:** An excellent molecular graphics package that supports a wide range of operations, including flexible molecular graphics, high resolution images for publication, user-driven analysis, multiple sequence alignment analysis, multiple model analysis, docking, to see the interaction between protein and ligand [10].

**Clustal X:** Clustal X is a windows interface for the Clustal W, a multiple sequence alignment program. It provides an integrated environment for performing multiple sequence and profile alignments and analysing the results. The sequence alignment is displayed in a window on the screen. A versatile coloring scheme has been incorporated allowing you to highlight conserved features in the alignment. The pull-down menus at the top of the window allow you to select all the options required for traditional multiple sequence and profile alignment.

**TreeView X:** TreeView X is an open source program to display phylogenetic trees on Linux, Unix, Mac OS X, and Windows platforms. It can read and display NEXUS and Newick format tree files (such as those output by PAUP*, ClustalX, TREE-PUZZLE, and other programs). The program was written by Rod Page using the wxWidgets C++ library.

**Phylogenetic tree**

Phylogenetic trees are used or rather created to infer evolutionary relationships among various biological species or other entities. Phylogeny is based upon similarities and differences in their physical or genetic characters. The horizontal lines are branches representing evolutionary lineages changing over time. The longer the branch in the horizontal dimension, the larger the amount of change. The vertical dimension is used simply to lay out the tree visually with the labels being evenly spaced. Phylogenetic tree have two kinds nodes known as external and internal, external nodes shows the different biological species and internal nodes shows the putative ancestor for the different biological species.

**SAVES:** SAVES is Structure Analysis and Verification Server. It runs 6 programs for checking and validating protein structures during and after model refinement.

**PROCHECK:** Checks the stereo chemical quality of a protein structure by analysing residue-by-residue geometry and overall structure geometry.

**WHATIF:** This does extensive checking of many stereo chemical parameters of the residues in the model.

**ERRAT:** Analyzes 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.

**VERIFIED 3D:** Determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D) by assigning a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar etc) and comparing the results to good structures.

**PROVE:** Calculates the volumes of atoms in macromolecules using an algorithm which treats the atoms like hard spheres and calculates a statistical Z-score deviation for the model from highly resolved (2.0 Å or better) and refined (R-factor of 0.2 or better) PDB-deposited structures.

**Ramachandran plot:** Produce an interactive Ramachandran plot.

**Methodology**

**NCBI BLAST:** The NCBI Protein Blast has been used in this study to find similar a protein; Table 1 lists the top five templates, on the basis of query coverage and identity that were found after the NCBI blast.

**Modelling:** Modeller software has been used to predict the structure of 2-Nitropropane Dioxygenaseprotein from *Mycobacterium tuberculosis*, in which first FASTA format sequence was converted into PIR format (“2NPD.ali” given below) that is the modeller readable format.

**2NPD sequence in PIR format >**:P1:2NPD sequence:2NPD::0.00: 0.00
The first line in the above PIR sequence format file contains the sequence code, in the format "">PI2.2NPY". The second line with ten fields separated by colons generally contains information about the structure file, if applicable. The rest of the file contains the sequence of the Schrodinger Suite 9.2 Maestro.

Simulation: After modelling the simulation for the target protein, the protein structure was optimized using the default parameters. Then the ligand receptor docking was performed in SP and XP manner. The output files are generated in the form of the "out_pv.maegz" form. Then they are saved and scores viewed per form of the "out_pv.maegz" form. Then they are saved and scores viewed.

Docking

Protein preparation: Protein preparation is an important step in any in-silico docking process, because most of the protein structures that we take as the input are x-ray crystal structures. They are static and don't include H atoms. They have implicit water molecules which are helpful for the crystal ability to study any one of the biological process that might be helpful or not. So in order to make the structure dynamic and functional protein has to be prepared. In protein preparation we have to assign bond orders, add hydrogen and missing side chain using prime.

Steps involved in protein preparation: Modeled structure of protein was first imported using input option in table menu. The modeled structure of 2-Nitropropane Dioxygenase protein was given as the input in the text field and displayed in workspace. After that missing side chain was filled using prime and pre-processes were performed using the default parameters. Then the protein structure was optimized at neutral pH using optimized option. It was assign the proper boned orders and the other charge distribution. After that protein structure was minimized by using the "Impref minimization" by taking OPLS 2005 as force field to submit the minimize option. In this way protein preparation was done by using "protein preparation wizard" workflow of the Schrodinger Suite 9.2 Maestro.

Steps involved in ligand preparation: Ligand preparation is an application of the Schrodinger Suite 9.2 Maestro, which is used to repara the ligand by using the various force fields (OPLS and MMFs). Ligand preparation will generate all the possible conformation of the ligand (isomers). Ligand can be in .SDF or .mol format. In virtual screening, a single file of all ligands needs to be made and saved. Here natural chemical compound was used to inhibit the target. For the conformation of result, both the force fields (OPLS and MMFFs) were used then virtual screening was done. In Ligprep application ligand molecules are given in .SDF format. Then the ionizer was used to generate the possible ionic states of all the molecules. Then OPLS 2005 and MMFFs were taken as the force field for minimizing the structures. Both force fields were used separately to prepare ligands. In stereoisomers the structure "Retaill specified chiralities" was selected to get the possible number of ligand states which are having same chiral centres. Then Ligprep was performed by starting the ligprep program.

Virtual screening process: Virtual screening is the process of Molecular docking in which the protein molecules act as your receptor and small molecules as Ligand. In the glide program we can do the docking in three ways: HTVS (High Throughput Virtual Screening), SP (Standard Precession) and XP (Extra Precession).

HTVS is useful for the process of virtual screening in which large number of small molecules or ligands can be screened virtually to identify whether they can be interacted with the Receptor or not, it is a fast method to analyze the molecules and can be useful in screening a large number of small molecule libraries.

SP virtual screening is the standard form of docking; it takes a bit longer time to give the accurate interaction result of the docked molecules of the receptors.

XP virtual screening is the precise method of docking where it gives the molecular interactions very precisely and accurately. So it is a slow method when compared with the other method.

Here we used both SP and XP method for virtual screening. These methods are present as an option in the Docking menu of the virtual screening workflow. The virtual screening workflow can be performed by the following steps given below:

Steps: Virtual screening can be initiated by giving the ligand file (.sdf) in the option "source of ligand" which is present in "Input" option of virtual screening workflow. In "Virtual Screening Workflow" receptor is given as receptor grid file (.zip) as input file. For this go to "Receptor" option and click on "Add" button and add the receptor file. Then all the default parameters are set. Then the ligand receptor docking was performed in SP and XP manner. The output files are generated in the form of the "out_pv.maegz" form. Then they are saved and scores viewed in the project table and analysed and saved for the future purposes. The result of virtual screening gives the list of compounds on the basis of Docking and GlideG score. In this way the virtual screening was done. By the analysis of these results and scores and the interaction of protein and the ligand we can conclude which ligand can be said as the best possible drug like compound for the taken receptors. The results are analysed and discussed in the Results and discussion section.

Results and Discussion

Template selection (NCBI blast)

Using NCBI blast server 2NPD protein blast done, in which five templates were selected those are listed below in Figure 1 as well as in Table 1.

<table>
<thead>
<tr>
<th>PDB-Id</th>
<th>Query Coverage</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>3BO9 A</td>
<td>99%</td>
<td>35%</td>
</tr>
<tr>
<td>Z6J A</td>
<td>98%</td>
<td>32%</td>
</tr>
<tr>
<td>4IQL A</td>
<td>71%</td>
<td>39%</td>
</tr>
<tr>
<td>2GJL A</td>
<td>96%</td>
<td>33%</td>
</tr>
<tr>
<td>2Z6I A</td>
<td>98%</td>
<td>32%</td>
</tr>
</tbody>
</table>

Table 1: List of template.
From among the above listed templates, PDB ID 2GJLA was selected based on the phylogenetic tree even though its query coverage and identity are 97% and 31%, respectively which is less than the other templates in the table. The phylogenetic tree is based on the evolutionary relationship. This PDB protein name is 2-Nitropropane Dioxygenase present in *Pseudomonas aeruginosa* organism. In phylogenetic tree the 2NPD protein was near to 2GJL protein. So, evolutionary relation of 2NPD protein is closer to 2GJL than others which are depicted in phylogenetic tree in Figure 2.

**Prediction of protein structure**

Using Modeller software structure of 2NPD were predicted and listed in Figure 3. In this predicted structure alpha-helix, beta-sheet and coil combination is present [11].

**Ramachandran plot**

Ramachandran Plot was taken after the modeling of 2NPD protein. The 2NPD protein sequence length is 355 residues and in ramachandran plot favored region is showing 326 residues. Only 9 residues is showing outlier region (Figure 4 and Table 2).

**Result of dynamic simulation**

Simulation was done at 5ns with protein and protein-ligand complex using gromacs software.

**RMSD graph**

RMSD stand for root-mean-square-deviation. It is used to measure of the average distance between the atoms specially protein backbone atoms. Figure 5 shows that RMSD value of protein were showing stability between 4 ns to 5 ns. While protein-ligand complex were showing stability between 4.25 ns to 5 ns.

The initial increase of RMSD values can be seen as an equilibration phase, during which the protein model should achieve a more correct arrangement optimizing its structure and this time period is called equilibration period.

**RMSF graph**

RMSF stands for root-mean-square-fluctuation. RMSF graph shows fluctuation in protein and protein-ligand complex around 265 to 300 residues which form loop region in the structure (Figure 6).

**Radius of gyration graph**

Radius of gyration shows the compactness of protein and protein-ligand complex i.e., how much protein is folded or unfolded with or without ligand. In radius of gyration graph of protein-ligand complex, the curve fluctuates till 2600 ps and then tends to stabilize and curve lies between 1.92 to 1.96 nm while radius of gyration graph of protein, the curve fluctuates till 2700 ps and then tends to stabilize and curve lies between 2.02 to 2.06 nm which infers the protein-ligand complex is more compact and have
Figure 3: Model structure of 2NPD.

Figure 4: Ramachandran plot of SwissPDB viewer.

<table>
<thead>
<tr>
<th>Region</th>
<th>Number of Residues</th>
<th>Percentages (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favoured region</td>
<td>326</td>
<td>92.4</td>
</tr>
<tr>
<td>Allowed region</td>
<td>18</td>
<td>5.1</td>
</tr>
<tr>
<td>Outlier region</td>
<td>9</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Table 2: Ramachandran plot descriptions.

better interaction in between them (Figure 7).

Potential energy graph: Potential graph shows that the modeled protein was stabled in the range of -1.65 e+06 kJ/mol and the ligand is stabilizing in the range of -1.61 e+06 kJ/mol (Figure 8).
Figure 5: RMSD graph of protein and protein+ligand.

Figure 6: RMSF graph of protein and protein+ligand.
Result of molecular docking

For virtual screening there are 1355 small compound (ligands) was taken from tuberculosis database (A webserver for predicting inhibitor against drug tolerant M. Tuberculosis) out of 1355 small compound five Ligands were showed best interaction with 2NPD protein (Tables 3 and 4).

The results of the docking studies can be studied by the analysis of number of Hydrogen bonds, the ligand is forming with the receptor active site residues and the number of hydrophobic interactions also
<table>
<thead>
<tr>
<th>Inhibitor Name</th>
<th>Pub Chem CID</th>
<th>Glide Score</th>
<th>Glide Energy</th>
<th>IUPAC Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inhibitor 1</td>
<td>623657</td>
<td>-10.439</td>
<td>-71.918</td>
<td>4-{3-[(5Z)-5-[[4-fluorophenyl]methylidene]-2,4-dioxo-1,3-thiazolidin-3-yl]propanoylamino}-2-hydroxybenzoic acid</td>
</tr>
<tr>
<td>Inhibitor 3</td>
<td>18432</td>
<td>-9.388</td>
<td>-53.953</td>
<td>(5Z)-5-[[1-[[2-(4-chlorophenoxo)ethyl]pyrrol-2-yl]methylidene]-1-(furan-2-ylmethyl)-1,3-diazinan-2,4,6-trione</td>
</tr>
<tr>
<td>Inhibitor 4</td>
<td>621743</td>
<td>-9.274</td>
<td>-65.787</td>
<td>2-[2-methoxy-4-[[E]-2-(5-nitroquinolin-2-yl)ethenyl]phenoxy]acetic acid</td>
</tr>
</tbody>
</table>

Table 3: Top five dock ligand.
plays an important role in the complex between protein and ligand. The strength of the interaction can also be analyzed by the scoring of the each possible interaction. That score is called as the glide score or GScore. The more the negative score the better the interaction. The ligand or compound giving highest negative score is choose as a best ligand and further analysis is performed by taking it with the receptor protein. The interaction diagram of these ligands with protein is visualized below (Figures 9-12).

**Conclusion**

In this study, the 2NPD protein structure was modeled using Modeller software. Molecular dynamics studies were performed for the model protein 2NPD, which showed its stability. The predicted protein model of 2NPD of *Mycobacterium tuberculosis* was used further in docking studies to obtain the set of lead molecules which could inhibit the protein function. Ligand based docking studies were performed which resulted in obtaining hits like CID 6236537, CID 9646169, CID 1843492CID, 6217473, and CID 6895994 which have high potency to form stable complex.

These inhibitor molecules can be used as lead molecules and further drugs can be designed. This could help in curing tuberculosis.
References


