Identification of HIV1 Protease Inhibitor through Molecular Modelling and Structure Based Virtual Screening Approach

Dipika Rungta, Nutan Chauhan and Koel Mukherjee*
Laboratory of Bioinformatics, Department of Biotechnology,
Birla Institute of Technology, Mesra, Ranchi, Jharkhand- 835215, India
*Corresponding author: Email: koelmukherjee@bitmesra.ac.in

[Received-30/04/2014, Accepted-01/08/2014]

ABSTRACT:
HIV-1 protease, a homodimer, has attracted the interest of many researchers due to its essential role in HIV replication and subsequent functional activities. It hydrolyses different viral proteins into their functional form to help in maturing the virus for further extending the disease propagation. The present workflow was designed to identify potential inhibitors for HIV-1 protease that is essential for the life-cycle of HIV. The in silico binding affinities of existing inhibitors namely Atazanavir and Ritonavir were compared using Glide module in Schrodinger suit 2013. Atazanavir was found to have the highest affinity (G-score= -12.117) towards HIV-1 protease. The structure based virtual screening on the basis of the binding modes of best inhibitor (Atazanavir) was performed and best scoring hits were identified. The structural behaviour of the best inhibitor-protease complex was validated through molecular dynamics simulation studies for a time stretch of 5ns. Simulation study revealed that the inhibitor-protease complex had structural stability over the concerned time period. The present structure-based drug designing approach can further be used to enhance more refined inhibitory property of novel lead molecules against HIV-1 protease that will aid knowledge in combating AIDS.

Keywords: HIV-1 Protease, Atazanavir, Glide, AIDS, G-score, binding affinity

[I] INTRODUCTION
From an unprecedented wealth of information regarding the molecular biology and virology of HIV collected in recent years, it became possible to identify numerous intervention points in the viral life cycle that could be exploited in the development of drugs for AIDS therapy [1-3]. Among these, the virally-encoded enzymes, in particular protease and reverse transcriptase have emerged as the most popular targets [4]. HIV-1 protease is a retroviral aspartyl protease (retropepsin) that is essential for the life-cycle of HIV, the retrovirus that causes AIDS. HIV protease cleaves newly synthesized polyproteins at the appropriate places to create the mature protein components of infectious HIV virions. Without effective HIV protease, HIV virions remain uninfected. Thus, mutation of HIV protease's active site or inhibition of its activity disrupts HIV’s ability to replicate and infect additional cells, making HIV-1 Protease inhibition the subject of considerable pharmaceutical research. HIV protease's protein structure has been investigated using X Ray Crystallography [5]. It exists as a homodimer, with each subunit made up
of 99 amino acids. The active site lies between the identical subunits and has the characteristic Asp-Thr-Gly (Asp25, Thr26 and Gly27) sequence common to aspartic proteases. The two Asp25 residues (one from each chain) act as the catalytic residues [6-7]. The conserved active site residues form a symmetrical and highly hydrogen bonded arrangement virtually identical to that described for pepsin [8]. According to the mechanism for HIV protease protein cleavage proposed by Mariusz Jaskolski and colleagues, water acts as a nucleophile, which acts in simultaneous conjunction with a well-placed aspartic acid to hydrolyze the scissile peptide bond [9]. Additionally, HIV protease has two molecular "flaps" which move a distance of up to 7 Å when the enzyme becomes associated with a substrate [10]. HIV protease inhibitors work by efficiently binding to the active site by mimicking the tetrahedral intermediate of its substrate and essentially becoming “stuck,” disabling the enzyme. This results in the production of immature proteins, which cannot assemble into infectious virions. Several protease inhibitors have been licensed for HIV therapy [11]. However, due to the high mutation rates of retroviruses, and considering that changes to a few amino acid within HIV protease can render it much less visible to an inhibitor, the active site of this enzyme can change rapidly when under the selective pressure of replication-inhibiting drugs [12]. One approach of minimizing the development of drug-resistance in HIV is to administer a combination of drugs which inhibit several key aspects of the HIV replication cycle simultaneously, rather than one drug at a time. Other drug therapy targets include reverse transcriptase, virus attachment, membrane fusion, cDNA integration and virion assembly [13]. One of the most common antiretroviral drugs of the protease inhibitor (PI) is Atazanavir (formerly known as BMS-232632). Like other antiretrovirals, it is used to treat infection of human immunodeficiency virus (HIV) [14].

Atazanavir is distinguished from other PIs in that it can be given once-daily (rather than requiring multiple doses per day) and has lesser effects on the patient’s lipid profile (the amounts of cholesterol and other fatty substances in the blood). Like other protease inhibitors, it is used only in conjunction with other HIV medications. The U.S. Food and Drug Administration (FDA) approved atazanavir on June 20, 2003. Traditionally, scientists identify new drugs either by fiddling with existing drugs or by testing thousands of compounds in a laboratory [15]. If we think of the target molecule, HIV protease in this case as a lock, this approach is rather like trying to design a key perfectly shaped to the lock [16]. The present paper carries virtual screening, docking and molecular dynamics approach for the identification of most potent inhibitor of HIV-1 Protease. There are various in silico softwares available that assure the dynamicity and stability of macromolecules. Here, we focus on virtual screening of the ligands, hits identification, docking and validation of the complex structure.

[II] MATERIALS AND METHODS
2.1. Selection of target and lead compound
Crystallographic structures of the target (HIV-1 Protease) was obtained from PDB (Protein Data bank) and saved in standard 3D coordinate format and was docked with inhibitors such as Ritonavir and Atazanavir, in order to get the most potential inhibitor based on their binding affinity. The best binding inhibitor was obtained and was searched for its 50% identical compounds. Small molecule libraries of 50 compounds, that are structurally similar to the drug Atazanavir were obtained from zinc database [17] and was docked with the enzyme, HIV-1 Protease. The potential hits were discovered and the lead compound was obtained for inhibition of the enzyme.

2.2. Protein and Ligand preparation
In order to be used as receptor for docking, protein structures should be processed. This was carried
out using the protein preparation wizard of maestro. In the protein preparation, various operations were done like hydrogen atoms were added, atomic charges were assigned and the water molecules that were not involved in ligand docking were eliminated. After the protein was prepared, the structure was optimized and minimized. The ligands selected as hits were prepared using LigPrep application of Maestro. The receptor grid was generated using the Glide application of maestro [18].

2.3. Molecular Docking

Docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. The docking was done by using Glide application of maestro. Glide is docking software that offers the full spectrum of speed and accuracy from high-throughput virtual screening of millions of compounds to extremely accurate binding mode predictions, providing consistently high enrichment at every level. The enzyme was first docked with the two inhibitors namely, Atazanavir and Ritonavir. The preliminary results were analyzed that showed that Atazanavir had the highest binding affinity against HIV-1 Protease. Further, 50% similar compounds of Atazanavir were searched and were also docked with the same enzyme. The results were saved.

2.4. Graphical Visualization

The docked PDB structures of each inhibitor against the target were visualized and inspected for their goodness of fit and orientation inside the active site. This was done with Pymol (http://www.pymol.org/) and Chimera (http://www.cgl.ucsf.edu/chimera/). Also the conformation and contacts with all amino acids were checked manually.

2.5. MD Simulation

Many marvellous biological functions in proteins and DNA and their profound dynamic mechanisms, such as switch between active and inactive states, cooperative effects, allosteric transition, intercalation of drugs into DNA, and assembly of microtubules, can be revealed by studying their internal motion. Likewise, to really understand the action mechanism of a receptor with its ligand, we should consider not only the static structures concerned but also the dynamical information obtained by simulating their internal motions or dynamic process. In order to examine whether the designed inhibitor remains bound in the presence of explicit solvent from a dynamic point of view, the molecular dynamic simulation was performed with GROMACS 96-43a force fields with the periodic boundary conditions (PBC) by using GROMACS 4.0 package [19]. The topology files and charges for the ligand atoms were generated by the Dundee PRODRG2.5 Server (beta) [20]. The 3D structure of the protein complexed with the best binding inhibitor was taken in a cubic box with a 1.5 Å edge length. To solvate the condition the “SPC” water model (spc216.gro file) was used to fill up the box. Before simulation, energy minimization was performed by steepest descent method and conjugant gradient method. MD simulation was performed for a period of 5 ns. All simulations were run under the periodic boundary condition with NVT ensemble by using Berensen's coupling algorithm for keeping the temperature at 300 K and pressure at 1atm. All bonds were constrained by using the LINCS algorithm. After completion of simulation, the trajectory files were generated, which were analyzed with different tools of GROMACS.

[III] RESULTS AND DISCUSSIONS

For infectious diseases, the development of a single therapeutically useful compound is by no means the end of the story. Such a compound, when used on a large scale, will almost always lead to the occurrence of pathogens which have cleverly developed one or more methods to avoid the harmful effects of the drug [21]. The constant
threat of resistance is a major source of concern. The only way to combat resistant pathogens with new drugs with long-lasting utility is to develop cocktails containing multiple compounds acting in diverse ways [22]. The protein HIV-1 Protease with already bound inhibitor Noa-His Hch psi [CH(OH)CH(OH)] Vam-Ile-Amp (U-75875) was downloaded from PDB database (www.rcsb.org) and was docked with inhibitors such as Ritonavir and Atazanavir, in order to get the most potential inhibitor based on their binding affinity. The best binding inhibitor was obtained and was searched for its 50% identical compounds for which 50 compounds were searched through the zinc database and were docked with the same. Docking was done with the Glide module of Schrodinger suit-2013. The best compound with highest affinity towards HIV-1 Protease, as obtained through glide, was further validated using MD Simulation through Gromacs.

For the docking study, protein (HIV-1 Protease) was prepared with potential energy of -821.977 and Root mean square deviation (RMSD) of 0.153511, which were obtained through the protein preparation wizard of Schrodinger Suit Version-9.7. Simultaneously, the zinc compounds taken from zinc database were prepared through LigPrep application. Table1 gives the information about the top 10. The table reveals that all the inhibitors satisfy the Lipinski’s Rule. According to this rule, the compounds should have

- No more than 5 hydrogen bond donors.
- Not more than 10 hydrogen bond acceptors.
- A molecular mass less than 500 daltons.
- Octanol-water partition coefficient log P not greater than 5.

Among these top 10 inhibitors, after Atazanavir, ZINC03914596 was found to have higher number of hydrogen bond acceptors and donors as compared to others. This means that this compound will have a good inhibitory activity against HIV-1 Protease. Table2 gives the docking results of top 10 compounds showing the binding affinity against the enzyme HIV-1 Protease. The glide scores and the hydrogen bond count for all the 10 inhibitors bound complexes were determined.

The structure with least glide score was considered to be the best docked result. The analysis of the docking results illustrates that Atazanavir had the lowest glide as well as docking score as compared to Ritonavir, and the other 50 compounds (similar to atazanavir) taken from zinc database. The result analysis inferred that zinc compound ZINC03914596 had second highest binding affinity with a glide score = -11.8457. The inhibitory activity of this compound can further be enhanced through mutations and modifications, which will make it stronger inhibitor against HIV-1 Protease. Docking was followed by molecular visualization of the docked complexes having best scores using the softwares like Pymol and Chimera. Fig 1 (a & b) represents the docked structure of Atazanavir and ZINC03914596 with the protein HIV-1 Protease. It gives the pictorial representation of binding orientation of ligands with their receptor at the active site. Fig 2 illustrates the interactions between the enzyme and best binding inhibitor (Atazanavir). The interaction study illustrates that the binding site on the enzyme for the ligand is polar in nature because there are higher number of polar interactions between Atazanavir and the active site of the enzyme having residues like Asp, Arg and H2O molecules. Fig 3 shows the interactions between the protein and the compound (ZINC03914596) with a good binding affinity towards the enzyme HIV-1 Protease. The figure infers that the binding pocket on HIV-1 Protease for the compound ZINC03914596 has higher number of interactions with polar residues like Asp, Arg and H2O. This analysis gives an idea that may be this binding pocket on the enzyme is hydrophilic in nature and has more polar interactions as compared to the non polar ones.

Through the docking studies, we came to know that compound having zinc ID ZINC03914596 had strong binding affinity and hence can be treated as
a competitor of Atazanavir (taken as a reference during virtual screening). This inference was further validated through MD Simulation. That means, the complex containing HIV-1 Protease and ZINC03914596 was energy minimized through molecular dynamics simulation using Gromacs software. The energy was calculated and graphs for RMSD, RMSF, Radius of Gyration and Potential energy were obtained. Fig 4 shows RMSD graph of protease as well as protease-ZINC03914596 complex. The figure infers that after a period of 1000 ps, the RMSD for protease gradually increased for the next thousands ps, thereby leading to a stable protein structure, while the protease-ZINC03914596 complex had less fluctuations in its RMSD. Here, the RMSD increased slightly after 1000 ps, after which the structure got stabilized in a uniform manner with constant RMSD values. Fig 5 shows plot of Radius of gyration (Rg) for protease as well as protease-ZINC03914596 complex. Rg plot shows distance between the atoms and their centre of axis. In the figure, the structure of protease as well as its complex show fewer fluctuations in their Rg values. Fig 6 shows Root square mean fluctuation (RMSF) plot of protease, ZINC03914596 as well as their complex. RMSF values show the flexibility of different segments of a protein. The figure shows that there is less flexibility in the structures of protease, ZINC03914596 and their complex. RMSF plot indicates that the flexibility in the complex remains less in the initial period but it leads to a sudden increase in the RMSF value, while, the protein’s and the ligand’s structure had fewer fluctuations. Fig 7 represents a plot of potential energy of protease. The plot infers that the overall potential energy (-785000 kj/mol) of the protease enzyme remains almost constant during simulation. Snapshots of the complex structure were shown in Fig 8 showing the changes that occurred gradually in the structure of HIV-1 Protease, during the procedure of simulation for a period of 5 ns.

**Table 1:** Properties of top 10 inhibitors

<table>
<thead>
<tr>
<th>S.No</th>
<th>Zinc Compounds</th>
<th>xlogP</th>
<th>H-bond Donors</th>
<th>H-bond Acceptors</th>
<th>Net Charge</th>
<th>Mol. Wt. (kcal/mol)</th>
<th>PE</th>
<th>RMSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Atazanavir</td>
<td>7.97</td>
<td>5</td>
<td>13</td>
<td>0</td>
<td>704.869</td>
<td>132.42334</td>
<td>0.009255</td>
</tr>
<tr>
<td>2.</td>
<td>ZINC03914596</td>
<td>4.26</td>
<td>5</td>
<td>10</td>
<td>1</td>
<td>671.863</td>
<td>241.112</td>
<td>0.039716</td>
</tr>
<tr>
<td>3.</td>
<td>ZINC17329035</td>
<td>1.65</td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>388.427</td>
<td>90.65235</td>
<td>0.002983</td>
</tr>
<tr>
<td>4.</td>
<td>ZINC36674146</td>
<td>2.33</td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>422.872</td>
<td>104.9878</td>
<td>0.026005</td>
</tr>
<tr>
<td>5.</td>
<td>ZINC07909533</td>
<td>3.86</td>
<td>3</td>
<td>8</td>
<td>0</td>
<td>418.453</td>
<td>80.23801</td>
<td>0.005219</td>
</tr>
<tr>
<td>6.</td>
<td>ZINC36704004</td>
<td>3.86</td>
<td>3</td>
<td>8</td>
<td>0</td>
<td>418.453</td>
<td>93.73813</td>
<td>0.015502</td>
</tr>
<tr>
<td>7.</td>
<td>ZINC22896769</td>
<td>6.57</td>
<td>3</td>
<td>8</td>
<td>0</td>
<td>418.453</td>
<td>68.96145</td>
<td>0.013696</td>
</tr>
<tr>
<td>8.</td>
<td>ZINC22896773</td>
<td>5.80</td>
<td>3</td>
<td>8</td>
<td>0</td>
<td>418.453</td>
<td>82.30933</td>
<td>0.011992</td>
</tr>
<tr>
<td>9.</td>
<td>ZINC02171919</td>
<td>4.47</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>389.455</td>
<td>43.3547</td>
<td>0.040068</td>
</tr>
<tr>
<td>10.</td>
<td>ZINC30888252</td>
<td>4.87</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>403.482</td>
<td>42.54979</td>
<td>0.010866</td>
</tr>
</tbody>
</table>
Identification of HIV1 Protease Inhibitor through Molecular Modelling and Structure Based Virtual Screening Approach

**Table 2**: Docking Results of top 10 inhibitors

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Compounds</th>
<th>Glide Score</th>
<th>Docking Score</th>
<th>H-Bonds</th>
<th>PE</th>
<th>RMSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Atazanavir</td>
<td>-12.1172</td>
<td>-9.54061</td>
<td>82</td>
<td>132.4233</td>
<td>0.009235</td>
</tr>
<tr>
<td>2.</td>
<td>ZINC03914596</td>
<td>-11.8457</td>
<td>-11.3212</td>
<td>4</td>
<td>203.8365</td>
<td>0.017757</td>
</tr>
<tr>
<td>3.</td>
<td>ZINC17329035</td>
<td>-10.4968</td>
<td>-10.4968</td>
<td>5</td>
<td>90.65235</td>
<td>0.002983</td>
</tr>
<tr>
<td>4.</td>
<td>ZINC36674146</td>
<td>-10.0973</td>
<td>-10.0973</td>
<td>6</td>
<td>104.9878</td>
<td>0.026005</td>
</tr>
<tr>
<td>5.</td>
<td>ZINC07909533</td>
<td>-10.4354</td>
<td>-9.91094</td>
<td>5</td>
<td>225.9158</td>
<td>0.04124</td>
</tr>
<tr>
<td>6.</td>
<td>ZINC36704004</td>
<td>-9.79982</td>
<td>-9.79982</td>
<td>4</td>
<td>80.23801</td>
<td>0.005219</td>
</tr>
<tr>
<td>7.</td>
<td>ZINC22896769</td>
<td>-9.64116</td>
<td>-9.64116</td>
<td>4</td>
<td>93.73813</td>
<td>0.015502</td>
</tr>
<tr>
<td>8.</td>
<td>ZINC22896773</td>
<td>-9.42446</td>
<td>-9.42446</td>
<td>4</td>
<td>68.96145</td>
<td>0.013696</td>
</tr>
<tr>
<td>9.</td>
<td>ZINC02171919</td>
<td>-9.12507</td>
<td>-9.12507</td>
<td>6</td>
<td>82.30933</td>
<td>0.011992</td>
</tr>
<tr>
<td>10.</td>
<td>ZINC30888252</td>
<td>-9.0539</td>
<td>-9.0303</td>
<td>4</td>
<td>43.3547</td>
<td>0.040068</td>
</tr>
</tbody>
</table>

**Fig 1**: Docked structure of (a) Atazanavir and (b) ZINC03914596

**Fig 2**: HIV1-Protease Interaction with Atazanavir
Identification of HIV1 Protease Inhibitor through Molecular Modelling and Structure Based Virtual Screening Approach

Fig: 3. HIV1-Protease Interaction with ZINC03914596

Fig: 4. Graph showing RMSD of the HIV-1 Protease as well as the Protease–ZINC03914596 complex

Fig: 5. Graph showing Radius of Gyration of HIV-1 Protease as well as Protease-ZINC03914596 complex
Identification of HIV-1 Protease Inhibitor through Molecular Modelling and Structure Based Virtual Screening Approach

**Fig: 6.** Graph showing RMSF of the HIV-1 Protease, Ligand (ZINC03914596) as well as the Protease – ZINC03914596 complex

**Fig: 7.** Graph showing PE of HIV-1 Protease, after Simulation of 5ns

**Fig: 8.** Ribbon Representation of 3D model of HIV-1 Protease after MD simulation of (a) 1 ns (b) 2ns (c) 3ns (d) 4ns and (e) 5ns
[IV] CONCLUSIONS
The present work emphasizes the study of protein inhibitor interactions which has a vital role in Computer Aided Drug Designing. Through virtual screening, hit compounds against HIV-1 Protease were identified, which were used to obtain the lead compound via molecular docking. Simulation studies conclude that the observed conformational changes in the protease structure might have occurred due to the binding of the inhibitor molecule. This structure-based drug designing approach can be used to enhance more refined inhibitory property of novel lead molecules against HIV-1 protease that will aid knowledge in combating AIDS. This would hence be helpful to reduce the troubles of antibiotic resistance and also the drug’s side effects as discussed earlier. In other words, through the above experimental study, we came to know that ZINC03914596 is a strong binding inhibitor of HIV-1 Protease enzyme. Hence, through this work, we got an idea that the above compound can be used as a strong inhibitor against the enzyme HIV-1 Protease and will fit better in combating HIV AIDS, which can be achieved by inserting mutations in its structure leading to its energy minimization. The minimized structure can further be taken into wet lab for its synthesis and can be hence used as a better drug against HIV AIDS.

[V] ACKNOWLEDGEMENT
The authors acknowledge BTISnet, Department of Biotechnology, Government of India, New Delhi for providing Infrastructure facility for Bioinformatics and Schrodinger for providing the glide software.

[VI] REFERENCES
12. Watkins T, Resch W, Irlbeck D, Swanson R (February 2003). Selection of high-level resistance to human immunodeficiency virus...


22. Pande Lab (2012). Folding at home Open source FAQ.