Study of Interaction between Wild type and Mutant HIV-1 protease and Cyclic Urea Inhibitor Using In Silico Techniques

Saba Sheikh*, Tanvee Pardeshi, Shama Mujawar, Krishna Deshmukh, Nidhi Sood, Preety Sharma and Sameer Chaudhary

*RASA Life Science Informatics, 301, Dhanashree Apartments, opposite Chitaranjan Vatika, Model colony, Shivaji Nagar Pune- MH 411016, INDIA. Phn: +912065600408

*Corresponding author: Saba Sheikh; email: sheikhsaba391@gmail.com

ABSTRACT
HIV-1 protease is a 99 amino acid aspartyl protease which is responsible for the cleavage of the viral polyprotein into functional constituent proteins. Inhibition of HIV protease causes the release of immature and noninfectious particles. Cyclic Urea inhibitors are seven member ring structure of cyclic urea compound. These inhibitors strongly bind with HIV-1 protease and are responsible for the formation of mature HIV which stops the cleavage of long viral protein into small peptide segment. Computational approaches were used to dock ligand (cyclic urea inhibitors) with HIV-1 proteases (mutated and wild type) to study their binding interaction and other properties. An ADME Toxicity property shows the best possible choice of drug against HIV-1 protease. Our in silico results approach that the mutated protein can bind with some of the Cyclic Urea inhibitors derivatives, which can block the functioning of HIV-1 protease, can be taken into consideration for further studies.

Keywords: HIV-1 proteases; Cyclic Urea inhibitors; in silico study; virtual screening; Docking; Drug Designing.

1. INTRODUCTION
It has now been two decades since acquired immunodeficiency syndrome (AIDS) was first reported by the US Center for Diseases Control (CDC). According to WHO almost 78 million people have been infected with the HIV virus and about 39 million people have died of HIV. Globally, 35.0 million people were living with HIV at the end of 2013.

HIV viruses are known to be the retroviruses which carry their hereditary information in the form of RNA. T cells carry CD4 antigen on their surface which is the initial targets of HIV virus. As the virus infects a cell, with the help of reverse transcriptase the RNA gets converted to DNA. Without reverse transcriptase the viral genome would not be able to incorporate into the host cell, resulting in failure to replicate. Enzyme integrase incorporates viral DNA into the DNA of the infected cell. Viral DNA inside the host cell transcribes into messenger RNA, which is then translated into viral proteins. The hydrolysis of these viral polyproteins by HIV 1 proteases make them functional and thus are essential for the viral assembly and its activation. On maturation, virion buds are formed from the host and these are targeted for HIV enzyme inhibitor design.

After the entry of HIV virus in the body within a month or two 40% to 90% people experience flu like symptoms known as “acute retroviral syndrome” or “primary HIV infection”. Sometimes there is no appearance of symptoms for years even after the infection. Common symptoms include fever, swollen gland, sore throat, rash, fatigue, cough, shortness of breath and many more.
HIV-1 Protease is aspartyl protease which functions as a dimer of C2 symmetry with each monomer consisting of 99 amino acid residues. Each monomer has one α helix and two antiparallel β sheets in the secondary structure. The enzyme active site is a catalytic triad composed of Asp25-Thr26-Gly27 from each monomer. It is gated by two extended β hairpin loops (residue 46-56) known as flaps [1].

HIV-1 protease is inhibited by protease inhibitors and contains three main genes: gag, pol and env. The gag and env genes codes for the viral membrane components i.e nucleocapsid and glycoprotein. Similarly pol gene codes for three essential enzymes (reverse transcriptase, integrase and protease) as well as other proteins. Gag and pol genes are translated as long polypeptide chain precursors (polyprotein); pol is a gag-pol fusion protein [2]. For the production of functional proteins there should be cleavage of gag and gag-pol at specific points.

HIV-1 protease is responsible for the cleavage of the gag-pol polyproteins into their functional constituent proteins, including the release of the protease itself from the gag-pol precursor. This key step in the maturation process of HIV-1 occurs during the final stages of the HIV life cycle as the virion buds from host cells. A regulation of HIV protease activity in the virus-replicated cycle is critical for proper assembly and maturation of HIV polyproteins to produce the infectious virus [3]. Thus, inhibition of HIV protease causes the release of immature and noninfectious particles.

For the production of infectious virions proteases needs to be activated by retroviruses. An impair mutation in the protease results in the production of non-infectious virus particle and analysis of the proteins from these mutant virions reveals the unprocessed Gag and Gag-Pol precursor proteins which are the viral protease substrates. In an experiment, each amino acid of the HIV-1 protease was individually mutated using a simple mutagenesis procedure which is capable of introducing and identifying missense mutations in each residue of a protein. Phenotypic screening of these mutants in a heterologous assay system reveals three regions within the protease where multiple consecutive amino acid residues are sensitive to mutation. These results show that random mutagenesis can be used to identify functionally important regions within a protein. Mutants with conditional phenotypes have also been identified within this collection [4].

With the study of new targets against HIV-1 protease, it was stated that cyclic urea inhibitors play an important role in halting of immature HIV-1 protease [5]. Cyclic Urea Inhibitors are seven member ring structure of cyclic urea compound. The main compound in derivatives of cyclic urea inhibitors is cyclic urea whereas the functional groups are different. Low binding energy compound (CUI) have more tendency to interact with protein [5]. Cyclic Urea inhibitors bind strongly with HIV-1 protease at active site. These inhibitors stop the cleavage of long viral protein into small peptide segment.

![Figure 1: Structure of Cyclic Urea Inhibitors](http://bioinfo.aizeonpublishers.net/content/2015/3/bioinfo659-666.pdf)
Now a day, in-silico techniques are used for novel approaches with great importance in target identification and drug-design visualization. These techniques employ drug development from the preclinical discovery stage to late stage clinical development. The study of computational biology approaches for analyzing the HIV-1 protease active interaction predicts many host cellular factors and discovers new pathways. The main benefit of the in-silico drug design is cost effectiveness in research and drug development.

There are different techniques used in in-silico drug design such as homology, molecular dynamic, energy minimization molecular docking etc. As such, docking studies can be used to identify the structural features that are important for binding and in-silico screening efforts in which suitable binding partners can be identified. Here we describe a practical approach for setting up docking simulations.

Research shows that binding of drugs alter the function of reverse transcriptase, integrate and HIV 1 protease. It was found that several inhibitors that may block the function of HIV 1 protease and these drug also having less toxicity effect [6, 7]. As Human Immunodeficiency Virus (HIV) cleaves viral polyproteins during the replication of virus and so this process is necessary for the life cycle of the virus and thus, inhibitor of HIV-1 PR are widely used in the treatment of HIV/AIDS.

In our extensive study of cyclic urea inhibitors, we aim to show the importance of cyclic urea inhibitors to stop the cleavage of viral polyproteins into functional proteins by HIV-1 protease. Efforts have been made to identify interactions that might increase their potency and their therapeutic benefits. Docking of these inhibitors with wild type and mutant HIV-1 protease was done to evaluate the interactions between the protein and the inhibitors. The successful results are believed to provide a new insight for understanding the binding of inhibitors with HIV-1 protease at molecular level.

2. MATERIALS AND METHODS

2.1 Literature Review and Protein Selection:
The structures of six cyclic urea inhibitors namely DMP323, DMP450, SD146, XK216, XK263 and XV638 were downloaded from the Pubchem in SDF format. All the drugs were clean in Marvin application and save in Mol2 format. The complexes of clean drugs were made in Discovery Studio. After that the complexes were submitted in PharmaGist to generate pharmacophores. A pharmacophore is the spatial arrangement of features that enables a molecule to interact with a target receptor in a specific binding mode. Once identified, a pharmacophore can serve as a powerful model in versatile applications for rational drug design, such as virtual screening, de novo design, lead optimization and ADME/Tox studies [8]. PharmaGist, is the first websever for elucidating 3D pharmacophores from a set of drug-like molecules that are known to bind to a target receptor. The method efficiently searches for possible pharmacophores and reports the highest-scoring ones [9].

While generating the pharmacophores on PharmaGist, each complex was set on none and first ligand in input with minimum number of features in pharmacophore 3& 4 and number of output Pharmacophores on 5. After comparing the PharmaGist results of six complexes the best two pharmacophores were selected based on highest alignment score. These pharmacophores were structurally similar to the cyclic urea inhibitors. The selected pharmacophores were then submitted to ZINCpharmer under subset Zinc Drug Database and numbers of hits were obtained. ZINCpharmer (http://zincpharmer.csb.pitt.edu) is an online interface for searching the purchasable compound of the ZINC database using the Pharmer pharmacophore search technology. After analyzing the pharmacophores in different subsets, the ligands obtained were downloaded. The downloaded ligands were in duplicates and hence using an in-house JAVA tool, duplicate ligands were removed. This reduced the number of ligands obtained from 394 to 103.

Three proteins were used in this project. Two were mutant type and one wild type. Two mutant types were selected from the Protein Data Bank (PDB) and one wild type was modeled using SWISS-MODEL and Phyre2. The two mutated proteins HIV-1 mutant (I84V) protease complexed with DMP323 (1MES) and HIV-1 mutant (I84V) protease complexed with DMP450 (1MER) were directly selected from the PDB. The wild type protein was modeled by selecting the FASTA sequence of HIV-1 protease of Homo sapiens (Q9YQ30) from Universal Protein Resource (Uniprot) as it provides the scientific community with a comprehensive, high quality and freely accessible resource of protein sequence and functional information (http://www.ebi.ac.uk/uniprot). The FASTA sequence of the selected HIV-1 protease was submitted in SWISS-MODEL and Phyre2 to model the wild type protein and the results was send on mentioned Email Id. SWISS-MODEL and Protein Homology/analogy Recognition Engine (Phyre2) are the protein structure modeling servers. After evaluating the results of both the server the Phyre2 results was best so it was selected as wild type protein for the further work.

2.2 Binding site:
The binding site information was found out differently for wild type and mutant protein. The amino acid interaction of the mutated proteins was checked on ligplot. Ligplot automatically generates the schematic diagrams of the protein-ligand interactions of the protein present in the protein data bank (http://www.ncbi.nlm.nih.gov/pubmed/7630882) [10]. For wild type modeled protein the amino acid interaction was unknown so pockets were found out to decide the amino acid interaction. For this purpose
different servers were used such as Computer Atlas of Surface Topography of proteins (Castp) which provides an online resource for locating, delineating and measuring concave surface regions on three-dimensional structures of proteins (https://bioinformatictools.wordpress.com/tag/castp/). COFACTOR is a structure-based method for biological function annotation of protein molecules (http://zhanglab.ccmb.med.umich.edu/COFACTOR/) [11]. COACH is a meta-server approach to protein-ligand binding site prediction (http://zhanglab.ccmb.med.umich.edu/COACH/) [12]. After comparing the results of all the described sites the interacting amino acid for the wild type protein was finalized after docking it against predicted pockets and analyzing their binding interaction. X, Y and Z co-ordinates of the interacting amino acid of all the receptors were noted and the average was calculated. These proteins were further cleaned in Discovery Studio by removing the heteroatom’s and prepared for docking.

![Image of Wild type Protein](http://bioinfo.aizeonpublishers.net/content/2015/3/bioinfo659-666.pdf)

**Figure 2:** Image of Wild type Protein

### 2.3 Docking Studies:

*In-silico* docking technique was performed on the molecules obtained from ZINCPharmer, with mutant protein 1MES and 1MER and wild type protein. To dock this large number of molecules AutoDock Vina was used. The required script and parameters were taken from the (http://autodock.scripps.edu/) website. AutoDock Vina is multiple docking open source software and this software has Lamarckian genetic algorithm (LGA) to search globally optimized conformations. Autogrid generates the maps to evaluate the ligand–protein interaction and grid map was set on 60 x 60 x 60 and the centre was calculated as per different receptors. Then docking was started.

### 2.4 Loop docking:

Based on the results of docking an in-house docking i.e. loop docking was done. The purpose behind the loop docking was to validate the Autodock Vina results. Most of the time the best docked results does not represent the best docking orientation so therefore loop docking has to be done. This automated “loop docking” will continue until a threshold value (d) is reached. The threshold value (d) is the difference between the docking binding energy of the last run and the preceding one. A threshold value of 0.05 was found to be appropriate. When this imposed value reached, the docking was stopped and the best-docked structure is selected. Autodock software is used for docking calculations [13].

### 2.5 ADMET:

Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) properties of drug candidates, pesticides, and industrial chemicals play key roles in drug discovery and environmental hazards assessment [13]. An ADMET structure-activity relationship database, abbreviated as admetSAR is an online server which gives the information about the toxicity, carcinogenicity of the drugs and shows whether the drugs follows all the Lipinski Rule or not. The loop docking results obtained were further checked by ADMET. The smiles of the finalized ligand were submitted in the admetSAR to check its toxicity.

### 3. RESULTS AND DISCUSSION

#### 3.1 Binding site identification:

Two mutant proteins of HIV-1 protease i.e. 1 MER and 1MES were downloaded and the ligplot was checked in PDBsum, whereas the wild type protein which was modeled in SWISS-MODEL having no ligplot. Using ligplot information ligand binding site was estimated and the X, Y and Z co-ordinates were noted down. Using
this co-ordinates mean was calculated and used for docking. To find out the interaction for wild type protein different online pocket finding servers were used. CASTp the online web server gives the 7 pockets finder results, each pocket with different amino acid. Similarly COACH and COFACTOR gives the different results. After comparing the results of all the sites pocket no. 6 of CASTp with residue ASP was selected as it shows the best results when docked with the ligands.

![Figure 3: Ligplot Images of 1MER and 1MES](http://bioinfo.aizeonpublishers.net/content/2015/3/bioinfo659-666.pdf)

### 3.2 Docking Analysis

Two selected pharmacophores i.e. pharmacophore 1 and pharmacophore 2 were docked with both mutated protein 1MER and 1MES and wild type protein in Autodock Vina. The results of Autodock vina were analyzed in Autodock. Analysis was done to check the hydrogen bonding and the binding energy.

From the total results obtained we analyzed top 25% results. Results were then sorted out on the basis of lowest binding energy and top first results selected as given below:

**Table 1: top 25% results**

<table>
<thead>
<tr>
<th>Protein name</th>
<th>ZINC Id</th>
<th>Binding Energy</th>
<th>Interacting amino acid</th>
<th>Conformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1MER</td>
<td>ZINC03831089</td>
<td>-8.1</td>
<td>ASP30:HN1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ZINC03831089</td>
<td>-8.1</td>
<td>ASP30:HN1</td>
<td>1</td>
</tr>
<tr>
<td>1MES</td>
<td>ZINC11592789</td>
<td>-7.8</td>
<td>ASP30:HN1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ZINC11592789</td>
<td>-7.8</td>
<td>ASP30:HN1</td>
<td>1</td>
</tr>
<tr>
<td>Wild type</td>
<td>ZINC03831271</td>
<td>-6.2</td>
<td>GLY48:N1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>ZINC14879985</td>
<td>-6.3</td>
<td>THR80:OG1</td>
<td>1</td>
</tr>
</tbody>
</table>

### 3.3 Loop Docking Analysis

After analyzing the results of Autodock Vina the ligand showing the lowest or minimum binding energy was selected for loop docking. Loop docking was done to validate the results obtained from Autodock Vina. The ligands having the zinc id ZINC03831089, ZINC11592789 and ZINC14879985 were loop docked with 1MER, 1MES and wild type protein respectively. The results obtained after loop docking shows an average binding energy as given below:

---

http://bioinfo.aizeonpublishers.net/content/2015/3/bioinfo659-666.pdf
### Table 2: Loop Docking

<table>
<thead>
<tr>
<th>Protein name</th>
<th>ZINC Id</th>
<th>Binding Energy</th>
<th>Interacting amino acid</th>
<th>Conformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1MER</td>
<td>ZINC03831089</td>
<td>-8.1</td>
<td>ASP30:HN 1</td>
<td>1</td>
</tr>
<tr>
<td>1MES</td>
<td>ZINC11592789</td>
<td>-7.8</td>
<td>ASP30:HN 1</td>
<td>1</td>
</tr>
<tr>
<td>Wild type</td>
<td>ZINC14879985</td>
<td>-6.3</td>
<td>THR80:OG1 1</td>
<td>1</td>
</tr>
</tbody>
</table>

### 3.4 ADMET

ADMET analysis has done to check the toxicity, carcinogenicity and many other factors of the drug. A set of test used in drug discovery to provide insight into how a pharmaceutical drug interacts with body as a whole. Drug should not contain any carcinogenic substance or mutagen which may alter the genetic material. It has to be check whether the drug is following the Lipinski Rule or not. The Lipinski rule states that there should not be more than 5 hydrogen bond donors, 10 hydrogen acceptors, molecular mass should be less than 500 daltons and log P not greater than 5. So referring to the Lipinski’s rule we concluded that these two ligands followed it completely and hence can be considered to be druggable molecules. The result is given below:

### Table 3: Lipinski’s rule two ligands

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>ZINC Id</th>
<th>ZINC03831089</th>
<th>ZINC11592789</th>
</tr>
</thead>
<tbody>
<tr>
<td>1MER</td>
<td>ZINC03831089</td>
<td>374.4706 g/mol</td>
<td>443.42662 g/mol</td>
</tr>
<tr>
<td>Molecular Weight</td>
<td>374.4706 g/mol</td>
<td>443.42662 g/mol</td>
<td></td>
</tr>
<tr>
<td>Molecular Formula</td>
<td>C_{22}H_{20}O_{5}</td>
<td>C_{22}H_{22}N_{2}O_{7}</td>
<td></td>
</tr>
<tr>
<td>Hydrogen Acceptors</td>
<td>5</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Hydrogen Donors</td>
<td>3</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>XLogP3</td>
<td>1.9</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Carcinogen</td>
<td>Not-Carcinogen</td>
<td>Non-Carcinogen</td>
<td></td>
</tr>
<tr>
<td>Biodegradation</td>
<td>Not ready biodegradable</td>
<td>Not ready biodegradable</td>
<td></td>
</tr>
</tbody>
</table>
After analyzing the ADME properties, toxicity screening was carried out using the admetSAR server. Following images show the result for toxicity screening done. After analyzing the ligands by admetSAR it was found that the ZINC03831089 and ZINC11592789 having popular name 11,17,21-trihydroxy-6-methylpregna-1,4-diene-3,20-dione (http://zinc.docking.org/substance/3831089) and (4R,4aR,5s,5aR,12aR)-4-dimethylamino-3,5,10,12,12a-pentahydroxy-6-methyl-1,11-dioxo-4,4a,5,5a-tet (http://zinc.docking.org/substance/11592789) respectively are non-carcinogenic and non-AMES toxic. So these can be used for further studies as probable drug candidates in the treatment of HIV infection by targeting protease enzyme.
4. CONCLUSION

The in silico study was done to check the interaction of Cyclic Urea inhibitors with the mutated and wild type HIV-1 protease. After performing the computational docking with the two mutated and one wild type ligand it was found that that the cyclic urea inhibitors having the best binding affinity. Mutated ligands showed the amino acids of our point of interest whereas the wild type ligand showed the totally different one. So, the admetSAR was done only of mutated proteins. Both the mutated proteins follow the Lipinski rule of five. This in silico study shows that the 1MER and 1MES can be taken into consideration for further studies for developing inhibitors against HIV-1 protease.

5. REFERENCES