Automated docking analysis provides new insights on protease inhibitor selection for antiretroviral therapy in HIV-1 infection

Carlos Famíliaa, João Santosb, Nuno Taveiraa,b and Alexandre Quintasa

a) Instituto Superior de Ciências da Saúde Egas Moniz; b) URIA, CPM, Faculdade de Farmácia da Universidade de Lisboa; *) e-mail: carlosfamilia@gmail.com

Abstract

Antiretroviral drugs may target four key proteins on the HIV life cycle: reverse transcriptase, protease, transmembrane envelope glycoprotein and integrase. Protease inhibitors (PI) were introduced in clinical use in 1995 immediately leading to the emergence of new drug resistant variants. Due to the high mutation rate of the HIV reverse transcriptase and the strong selective pressure imposed by protease inhibitors there are numerous genetic variants of the HIV protease at the individual and population level. Selection of which protease inhibitors to include in a new therapeutic regimen may be well informed by sequencing analysis of the pol gene. However, in heavily treated patients the increasing genetic complexity of the resistant proteases limits the usefulness of genotypic resistance testing to inform therapeutic decisions.

The main goal of this study was to develop an in silico method that allows the choice of PIs based on the structural differences found between drug resistant and drug susceptible proteases. Three dimensional (3D) models of the HIV-1 protease were produced for eight sequences with known PI resistance mutations selected from the Stanford HIV Resistance Database and for one control protease without resistance mutations. Ligand-protein automated docking with flexible ligand was subsequently performed for the binding site of each of the proteases with some of the PIs commercially available in Portugal.

Using as an example the resistant mutation 48V to analyze 1) protease-PI binding affinity, 2) complexity of the pathway the drugs take to reach the binding site, 3) number of occurrences with free binding energies below zero and their distribution, we show that amprenavir and tipranavir are the PIs of choice to inhibit the replication of HIV strains bearing this mutation. Our preliminary results suggest that this new method may help to inform difficult therapeutic decisions in heavily treated HIV-1 patients.
Introduction

Drug resistance represents a foremost barrier to successful treatment of HIV patients and clinical awareness of HIV drug resistance is a major issue towards an efficient antiretroviral therapy. HIV protease (PR) is an essential enzyme in the viral life cycle and its unique features were crucial in the choice of a HIV drug targets [1-3].

Protease inhibitors (PI) were introduced in clinical use in 1995, immediately leading to the emergence of new drug-resistant strains, as a result of the high levels of virus production, the high mutation rate and the strong selective pressure imposed by protease inhibitors [4]. Emerging PI resistant strains significantly affect the long-term effectiveness of antiretroviral therapy [5]. The evaluation of the effects of resistant mutations is therefore a critical factor for the elucidation of the most efficient PI to include in a new therapeutic regimen. Usually this can be done either by genotypic or phenotypic assays [6].

Genotypic assays are based on the determination of the nucleotide sequence and analysis of the pol gene [7]. However, in heavily treated patients the increasing complexity of mutation patterns, the large number of existent polymorphisms that interact to form resistant-strains, cross-resistance and reversal resistance, limits the usefulness of genotypic resistance testing to inform therapeutic decisions [5, 7].

Phenotypic drug susceptibility assays determine the inhibitory concentration of an anti-retroviral agent necessary to reduce HIV-1 replication by 50% (IC50). Although they can reproducibly determine phenotypic susceptibility, they also have several limitations as they are labour intensives, extremely time consuming, expensive, and furthermore, these assays may impose significant in vitro selective pressures that may favour the outgrowth of certain virus isolates and not other from the original heterogeneous viral population [6, 8].

In the present study we propose a new *in silico* approach to evaluate the best anti-protease therapy, keeping in mind three key objectives: (i) To evaluate the Pol mutations which lead to PI resistance; (ii) To choose the most adequate PI for anti-protease therapy (iii) to circumvent the limitations of the currently available methods (iv) to achieve the previous objectives in a short period of time. The methodology is based on the structural differences found between drug resistant and drug susceptible variants using molecular modelling and molecular docking.

Among all current computational approaches for protein molecular modelling, homology modelling is the only method that can reliably generate three-dimensional (3D) models of a protein (target) from its amino acid sequence [9]. Swiss-Model allows an automated comparative modelling of 3D protein structures, resorting to a database of predetermined experimental structures (NMR or X-ray crystallography), from which, the structure that shares the most similarity with the protein sequence in study, is used has template for the new structure [10, 11]. The higher the similarity between the target and the template, the more accurate will be the constructed model [10].
Molecular docking tries to predict the globally most favorable site of binding of a trial molecule [12] and its preferred orientation when bound to another molecule to form a stable complex, with its many degrees of conformational freedom, using the most realistic scoring function and having a large number of applications in the field of medicine [13].

Thus, we propose an in silico approach based on both the minimum and the 250 docking runs median of ligand binding free energies, corresponding to the best ligand docking conformation and the difficulty for the ligand to find the docking site, respectively. Additionally the percentage of occurrence below zero was also considered has a parameter in the data analysis. The obtained results are mostly in agreement to the Stanford database of HIV protease resistance mutations, but additionally allow us to predict the best PI for a given mutation.

Materials and Methods

Dataset

X-ray crystal structures of proteases were retrieved from the Protein Data Bank (PDB), based on the structure resolution, after which, their sequence were submitted for genotypic testing through Stanford HIV drug resistance database (http://hivdb.stanford.edu/). The structure with PDB code 1ZPA was selected because it didn’t show resistance to any PI commercially available.

Generation of protease mutant three-dimensional structures

From the PDB 1ZPA sequence several mutants were generated by in silico directed mutagenesis each having a singular amino acid replacement that, accordingly to Stanford HIV drug resistance database, confer high resistance to one of the eight commercially available PI used in this study (Amprenavir, Atazanavir, Indinavir, Lopinavir, Nelfinavir, Ritonavir, Saquinavir and Tipranavir).

The three-dimensional structures of the protease mutants were generated resorting to Swiss-Model, using the 1ZPA X-ray crystallographic structure as template.

Preparation of proteases and protease inhibitor structures for docking.

Preparation of protein and inhibitor structures was carried out using AutodockTools version 1.4.5.

For the preparation of the protein structures, only polar hydrogens were added to the protein and Gasteiger charges were automatically assigned to each atom. Afterwards a three-dimensional affinity and electrostatic grid boxes was generated which could cover the entire active site using the AutoGrid version 4 [14] program. The number of grid points in x, y, z-axes was 100 x 100 x 100 with each point separated by 0.1917 Å (Fig. 1).
Inhibitor structures were treated as all atom entities, for which all hydrogens were added to fill the empty valences, followed by the addition of Kollman partial atomic charges. The root atom and the rotatable dihedrals in the inhibitors were defined using the AutoTors program. All the rotatable bonds were allowed to rotate freely. The partial charges of the non-polar hydrogens were added to the charge of the carbons they bonded to, being deleted afterwards. The atom type for the aromatic carbons was reassigned to be handled by the aromatic carbon grid map.

**Determination of protease inhibitor binding energies by docking**

Docking calculations were carried out using AutoDock version 4 with a Lamarkian Genetic Algorithm (LGA) and a ‘Solis & Wets’ local search [20]. The number of docking runs was set to 50. The maximum number of energy evaluations before LGA run termination was 2,500,000 and the maximum number of generation of the LGA run before termination was 27000. Other docking parameters were left in the default values provided by AutoDockTools. Each docking was repeated five times, having in the end a total of 250 docking runs.

The final docked energy was calculated during the docking procedure from the sum of the intermolecular energy and the internal energy of the ligand. Both the minimum and the median free energy of ligand binding of all docking runs were used on results analysis of the present study.
Results and Discussion

Molecular docking evaluates the energy of protein–inhibitor complexes. The binding energies produced by this approach often identify mutations conferring resistance in a target protein. HIV protease is a homodimeric protein in which the active site is localized between the two subunits (Fig. 2) [1].

Figure 2 - Schematic structure of 1ZPA HIV-1 protease bounded to Amprenavir (A) and to Tipranavir (B) both in the conformation corresponding to the lowest free energy of binding [15-19].

The minimum values of free energy of binding estimated for each docking process is a measure of the affinity of the inhibitor for the active site of the protease. Table 1 shows an increase of the free energy of inhibitor binding of 0.65, 0.87, 0.24, 1.84 and 0.69 for the mutations 30N, 47A, 48V, 50L and 84V, respectively. These results are in good agreement with the Stanford genotypic resistance database, considering that the previous mutations correspond, by the same order, to the genotypic resistance for Nelfinavir, Lopinavir, Saquinavir, Atazanavir and Indinavir.

However, in every mutation either a decrease or an increase in the minimum energy of binding for some inhibitors was found when compared to the values obtained for 1ZPA, suggesting that some mutations of known resistance PI could lead to a higher susceptibility, or lead to cross-resistance to others protease inhibitors.

Table 1 – Minimum estimated free energy of binding (kcal·mol⁻¹) for every docking procedure of PI inhibitors in 1ZPA HIV-1 protease and the mutations in study.

<table>
<thead>
<tr>
<th></th>
<th>Amprenavir</th>
<th>Atazanavir</th>
<th>Indinavir</th>
<th>Lopinavir</th>
<th>Nelfinavir</th>
<th>Ritonavir</th>
<th>Saquinavir</th>
<th>Tipranavir</th>
</tr>
</thead>
<tbody>
<tr>
<td>30N</td>
<td>-9.60</td>
<td>-9.95</td>
<td>-10.66</td>
<td>-9.91</td>
<td>-10.42</td>
<td>-10.09</td>
<td>-10.31</td>
<td>-11.09</td>
</tr>
<tr>
<td>47A</td>
<td>-8.65</td>
<td>-9.06</td>
<td>-10.79</td>
<td>-10.03</td>
<td>-10.28</td>
<td>-8.63</td>
<td>-10.11</td>
<td>-10.02</td>
</tr>
<tr>
<td>50L</td>
<td>-9.43</td>
<td>-6.91</td>
<td>-10.34</td>
<td>-9.65</td>
<td>-10.92</td>
<td>-9.40</td>
<td>-10.81</td>
<td>-11.64</td>
</tr>
<tr>
<td>84V</td>
<td>-9.06</td>
<td>-9.75</td>
<td>-10.77</td>
<td>-10.00</td>
<td>-10.99</td>
<td>-10.29</td>
<td>-11.03</td>
<td>-10.80</td>
</tr>
</tbody>
</table>
The high number of docking runs used should give a more approximate value to the real free energy of binding, allowing a reliable comparison of values. Nevertheless, the free energy binding differences between the 1ZPA protease and the mutated proteases are still inside the error margin.

In order to have a more robust measurement of PI binding capacity, the median value of estimated free binding energy of all docking events was calculated. The median values of free binding energy represent the ligand affinity to the whole region defined by the grid, being an indication of how straightforward the trajectory of the ligand will be until it successfully attaches to the active site. Table 2 shows the median values for the free binding energy of each PI to wild-type protease (1ZPA) and to the mutated proteases. For protease mutations 47A, 50L and 84V corresponding to the genotypic resistance for Nelfinavir, Atazanavir and Indinavir respectively, the pathways of docking runs are associated with a higher energy when compared to the median value of 1ZPA, indicating that it is more difficult for the inhibitor to reach the binding site of the mutated proteases. In the mutation 48V for Amprenavir, Nelfinavir and Tipranavir, and in the mutation 84V for Nelfinavir, nonetheless, there was a lower median value of estimated free energy of inhibitor binding, suggesting that the path may be more favorable than the one encountered for this inhibitors to 1ZPA structure binding site. Therefore, a higher median value was detected in all mutations for most of the PIs, suggesting that the path leading to the binding site is less favorable for these inhibitors.

Table 2 – Median value of estimated free energy of binding (kcal·mol⁻¹) calculated from all of the docking runs effectuated in each docking procedure of PI inhibitors in 1ZPA HIV-1 protease and the mutations in study.

<table>
<thead>
<tr>
<th>PI Inhibitor</th>
<th>1ZPA</th>
<th>30N</th>
<th>47A</th>
<th>48V</th>
<th>50L</th>
<th>84V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amprenavir</td>
<td>-7.53</td>
<td>-7.33</td>
<td>-6.94</td>
<td>-7.68</td>
<td>-6.96</td>
<td>-7.22</td>
</tr>
<tr>
<td>Atazanavir</td>
<td>5.02</td>
<td>10.61</td>
<td>6.85</td>
<td>15.41</td>
<td>15.01</td>
<td>8.68</td>
</tr>
<tr>
<td>Indinavir</td>
<td>-7.53</td>
<td>-6.91</td>
<td>-6.75</td>
<td>-7.43</td>
<td>-6.41</td>
<td>-7.36</td>
</tr>
<tr>
<td>Lopinavir</td>
<td>-5.84</td>
<td>-5.04</td>
<td>-4.68</td>
<td>-4.65</td>
<td>-3.51</td>
<td>-5.71</td>
</tr>
<tr>
<td>Nelfinavir</td>
<td>-8.45</td>
<td>-2.94</td>
<td>-7.97</td>
<td>-8.57</td>
<td>-7.92</td>
<td>-8.54</td>
</tr>
<tr>
<td>Saquinavir</td>
<td>-7.10</td>
<td>-6.38</td>
<td>-6.32</td>
<td>-6.25</td>
<td>-6.33</td>
<td>-6.56</td>
</tr>
</tbody>
</table>

In order to establish a new therapeutic regimen, we analyzed both minimum and median values of the estimated free binding energy of the mutant protein compared to the ones obtained for the wild-type protease (1ZPA) (Table 3), as well as the number and distribution of the observed occurrences with a free energy of binding lower than zero (Fig. 3).

Looking first at the difference between the values of the minimum free binding energy, all PIs, except for Nelfinavir, showing a loss of affinity for the binding site could be selected for therapy. Analyzing the variation of the median value of free binding energy, all PIs seem to be a good choice, except for Atazanavir and Ritonavir.
Table 3 – Minimum estimated free energy of binding (kcal·mol⁻¹). Median value of estimated free energy of binding (kcal·mol⁻¹) calculated from all of the docking runs effectuated in each docking procedure of PI inhibitors in 1ZPA HIV-1 protease and its mutations 48V.

<table>
<thead>
<tr>
<th></th>
<th>Amprenavir</th>
<th>Atazanavir</th>
<th>Indinavir</th>
<th>Lopinavir</th>
<th>Nelfinavir</th>
<th>Ritonavir</th>
<th>Saquinavir</th>
<th>Tipranavir</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum Energy of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min Energy of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆</td>
<td>-0.16</td>
<td>-0.58</td>
<td>-0.27</td>
<td>-0.09</td>
<td>0.13</td>
<td>-0.38</td>
<td>0.24</td>
<td>-0.48</td>
</tr>
<tr>
<td>Median Energy of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binding 1ZPA</td>
<td>-7.53</td>
<td>5.02</td>
<td>-7.53</td>
<td>-5.84</td>
<td>-8.45</td>
<td>-5.39</td>
<td>-7.10</td>
<td>-8.40</td>
</tr>
<tr>
<td>Median Energy of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆</td>
<td>-0.15</td>
<td>10.39</td>
<td>0.11</td>
<td>0.10</td>
<td>-0.12</td>
<td>2.25</td>
<td>0.85</td>
<td>-0.23</td>
</tr>
</tbody>
</table>

The PIs that showed a best relation between the inhibitor affinity to the binding site and the trajectory to the active site are Amprenavir and Tipranavir. The analysis of the number of occurrences below zero and its distribution (Fig. 3) indicates that Amprenavir, Nelfinavir or Tipranavir have the best response inhibiting the mutated Protease when compared to 1ZPA.

Figure 3– Percentage of occurrences for each value of estimated free energy of binding (kcal·mol⁻¹) below zero kcal·mol⁻¹, darker gray corresponds to the docking results analysis of 1ZPA and lighter gray to the 48V mutation.

Overall, the results (minimum, median, percentage of occurrences below zero and their distribution), point for a better efficiency of inhibition for the 48V mutation of HIV-1 protease, when using Amprenavir or Tipranavir.

Conclusions

We have developed an in silico approach to identify mutations that lead to PI resistance and to help selecting the most adequate PI for HIV therapy. This method is based on protein homology modeling and molecular docking. We show that amprenavir and tipranavir are the PIs of choice to inhibit the
replication of HIV strains bearing the 84V resistance mutation. Our preliminary results suggest that this new method may help to inform difficult therapeutic decisions in heavily treated HIV-1 patients.

**Future Perspectives**

Molecular modeling and molecular docking methods still have a long way to run before producing completely reliable results *per se*. This could be achieved by increasing the number of parameters considered and using new scoring functions both of which require the use of advanced computational capabilities.

**Acknowledgments**

This work was supported by Egas Moniz Cooperativa de Ensino Superior grant EMAQ07. This study was made possible by Prof. Doutor Mário Polido and Drª. Margarida Lima, the Student’s Association of ISCSEM, Engº Luis Lopes and Bruno Costa, that yields computational resources for the execution of this project.

**Bibliography**


