Systematic Investigation on Interactions for HIV Drug Resistance and Cross-Resistance among Protease Inhibitors

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.7243/2050-2273-1-2</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:14169380">http://nrs.harvard.edu/urn-3:HUL.InstRepos:14169380</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Systematic Investigation on Interactions for HIV Drug Resistance and Cross-Resistance among Protease Inhibitors

Jing Zhang1*, Tingjun Hou3†, Yang Liu1, Gang Chen1, Xiao Yang1, Jun S. Liu4 and Wei Wang2*

Abstract
We systematically investigate the interactions for HIV drug resistance and cross-resistance of three protease inhibitors. Using Bayesian probabilistic models, we first detect mutation combinations related to drug resistance in the HIV-1 protease or reverse transcriptase (RT) of patients with single, double and triple drug treatments respectively, and then infer detailed interaction structures of these mutations. Moreover, we compare the resistance patterns of single-drug treatments and multiple-drug treatments and study the selection patterns of mutations from single-drug treatments and multiple-drug treatments, using molecular dynamics simulation and free energy calculations, in order to draw inference for giving advice on the design of multiple-drug treatments.

Introduction
HIV drug-resistance, which is caused by mutations of viral proteins that disrupt the drugs’ binding, enables the virus to survive and reproduce in the presence of antiretroviral drugs, hence substantially diminishes the efficacy of HIV drug therapy [1-3]. Furthermore, HIV drug-resistance induced by the genetic variation in virus replication is preserved and flourished by selection pressure during therapy [3-5]. Although it has been shown that physicians with access to drug resistance data can provide better-responded therapy to patients, the complicated mutation patterns and unequal importance of the viral protein mutations make it hard for genotypic data to be helpful in designing therapeutic regimen[7-9].

Zhang [13] proposed an innovative method for investigating mutation interactions of HIV induced by a certain drug treatment [13]. In this study, the Bayesian variable partition (BVP) model is applied to the data of three different drug treatments (Indinavir, zidovudine, and nevirapine) respectively to select mutations associated with drug resistance, and then the recursive model selection (RMS) procedure is employed to infer the dependence structure among the interacting mutation positions found by the BVP model. This paper rediscovers the majority of known resistant mutations to the aforementioned three drugs [8] and uncovers several interacting structures for these mutations. In particular, a conditional independence structure among the mutations M46I, I54V and V82A agrees with previous experimental results.

However, combinations of drugs from more than one drug classes, which are believed to be able to prolong virus suppression and immunologic reconstitution, lead to the problem of cross-resistance, i.e., drug resistance of the HIV virus when two or more drugs are used in the therapy for the same patient [11-15]. Currently there are sixteen antiretroviral drugs that have been approved for the HIV-1 infection:

seven nucleoside/nucleotide reverse transcription (RT) inhibitors (NRTI), six protease inhibitors (PIs), and three nonnucleoside RT inhibitors (NNRTI) [11-12]. Yet previous studies provide little insight on this issue and often give inconsistent results when analyzing the same input mutation data [5,10].

In this study, we proposed a systematic procedure to apply BVP and RMS on the data of single, double and triple treatments of three PIs: IDV, NFV and SQV from Stanford HIVdb database to detect the multiple-drug treatment effects and compare them to the single-drug treatment effects. We then use Molecular Dynamic (MD) Simulations and Free Energy Calculations to explore the selection pattern of drug-resistance effects from single-drug profile to multiple-drug profile. We find that only a few mutation positions of single-drug treatment showing up in multiple-drug treatment, and that the existence of these mutations depends on the competing outcome of the interaction effects on the drugs.

Materials and methods
Using BVP and RMS for cross-resistance for multiple drug
Here we first used the Bayesian variable partition (BVP) model [13,14] to search for the mutation positions (1 to 99) associated with the cross-resistance of IDV, NFV and SQV interactively. After detecting the interaction mutation positions, we applied Recursive Model Selection [13] (RMS) to infer the detailed dependence structure among the interacting positions.

BVP and RMS are both designed to explore mutations for single-drug resistance rather than cross-resistance for multiple drugs. We now show our procedure of using BVP and RMS to study cross-resistance. We applied them sequentially to the data from single-drug treated, double-drug treated and triple-drug treated patients respectively.

From Stanford database, we collected 4146 untreated patients’ sequences, and single-drug treated sequences (949 IDV-treated, 714 NFV-treated, 240 SQV-treated), double-drug treated sequences (280 IDV-NFV, 143 IDV-SQV, 96 NFV-SQV), and finally triple-drug treated patients (201 IDV-NFV-SQV).

*correspondence: jing.zhang.jz349@yale.edu
1Department of Statistics, Yale University, 24 Hillhouse Avenue, New Haven, CT.
Full list of author’s information is available at the end of the article.

© 2012 Zhang & Wang et al; licensee Herbert Publications Ltd. This is an open access article distributed under the terms of Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), This permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
First we applied BVP to all single-drug-treated datasets (IDV, NFV and SQV) versus 4146 untreated dataset, and searched for interactively associated mutation positions (see Figure 1 for results). Then RMS was used to figure out the dependence structure among the interacting positions (see Figure 2). These are just single-drug resistance results.

Second, we applied BVP and RMS to 3 double-drug datasets individually (Figure 4 shows the results for IDV-NFV, Figure S1 and S2 in Supplementary Materials show the results for IDV-SQV, NFV-SQV). Comparing this result with the single-drug resistance results, we can see how the resistant mutation interactions change when two of the three protease inhibitors were used.

Third, we applied BVP and RMS to triple-drug dataset. Figure 3 shows the result.

**Results**

We downloaded the data from Stanford HIVdb database (http://www.hivdb.stanford.edu). Among all protease inhibitors (PI), NRTI, NNRTI, only several PIs and RTs have sufficient sample sizes, namely IDV, NFV, SQV, AZT and NVP. Since we are interested in cross-resistance within drug classes, we also need good sample sizes for multiple-drug treatments. Only IDV, NFV and SQV have large enough sample sizes for all single-drug treatments (949 IDV-treated sequences, 714 NFV-treated, 240 SQV-treated), double-drug treated sequences (280 IDV-NFV, 143 IDV-SQV, 96 NFV-SQV), and triple-drug treated patients (201 IDV-NFV-SQV). Hence we focus our investigation on these three PIs, with the understanding that given enough data, the same methods can be applied to other PIs and NRTI, NNRTI, etc.

**Interactions for single-drug resistance of IDV, NFV and SQV**

Zhang [13] proposed a systematic approach for detecting and understanding combinatorial mutation patterns responsible for HIV single-drug resistance. In order to find interaction pattern for drug resistance of IDV, Zhang [13] applied Bayesian variable partition (BVP) and Recursive Model Selection (RMS) on HIV-1 type B protease sequence data - 4146 sequences from untreated patients and 949 sequences from IDV-treated patients. Similarly, we applied BVP and RMS on HIV-1 type B protease sequence data, from NFV-treated patients (714 sequences) and SQV-treated patients (240 sequences). Figure 1 and Figure 2 show the interacting positions detected by BVP and the dependence structure inferred by RMS for IDV, NFV and SQV. Interestingly, even from this single-drug resistance results we can see there are some positions associated with more than one drug treatments. 54, 82 and 90 (in red circles in Figure 1) are associated with all of the three drugs, and 10, 71 and 73 are associated with IDV and SQV (but not NFV), while 88 associated with NFV and SQV (but not IDV). Totally, there are 10 positions interactively associated with IDV, 6 with NFV, and 12 with SQV.

**Dependent structures for IDV, NFV and SQV**

After applying RMS, we were able to figure out some of the dependent structures among the interacting positions. Zhang [13] identified two independent groups: one is composed of 46, 54 and 82, the other composed of 73 and 90 (Figure 2a). In Figure 2b and 2c, there are three independent groups identified for NFV: 54 and 82, 20 and 90, 30 and 88. While for SQV, two independent groups were identified:
Figure 2. The inference of detailed dependence structure for IDV (a), NFV (b) and SQV (c) using RMS. The result of IDV was reported in Zhang [13]. For NFV (b), three independent interaction groups were inferred with high posterior probabilities (>0.9): 54 and 82, 20 and 90, 30 and 88. For SQV (c), two independent interaction groups were inferred with high posterior probability (>0.9): 48, 54 and 82; 71, 73, 79 and 90, although other details of the structure were not able to be inferred with high confidence.

Figure 3. The posterior probabilities (a) of each position of mutation to be associated interactively with triple-drug treatment (IDV, NFV and SQV, applying BVP to triple-drug treated dataset) and the inferred detailed dependence structure (b) among interacting positions (applying RMS to triple-drug treated dataset). In (a), totally eight positions are interactively associated: 24, 30, 46, 54, 73, 82, 88 and 90. In (b), three independent interaction groups were inferred: 30 and 88; 73 and 90; 24, 46, 54 and 82. The detailed structure inside the last group was ambiguous.

Conjecture of drug-resistance mutation interaction profiles for multiple-drug treatments based on profiles from single-drug treatments

Figure 1 and Figure 2 are the drug-resistant mutation interaction profiles for IDV, NFV and SQV based on single-drug treatment data. Currently, the state of art in cross-resistance is based on single-drug resistance profiles. If the drug-resistant effects in a multiple-drug treatment are simply additive (independent), i.e. mutation interactions from one drug will not affect the one from another drug in the same treatment, then we would expect to see most of the mutation interactions in single-drug treatments show up in multiple-drug treatment. For example, there are 10 mutation positions involved in resistant interactions of IDV treatment and 6 positions involved in NFV treatment (Figure 1 and 2A), totally 13 non-redundant positions forming about 4 interaction groups (one for 30 and 88, one for 10 and 71, one for 73, 90 and 20, one for 54, 82 and others) for IDV and NFV. Based on the independent assumption, if we apply BVP and RMS to double-drug treatment of IDV and NFV, we would expect to see about 13 positions for mutation interactions and 4 interaction groups either from IDV or NFV. Similar arguments could also be applied for double-drug treatments of IDV-SQV and NFV-SQV. Finally, in triple-drug treatment...
of IDV-NFV-SQV, we would expect to see totally 18 non-redundant positions for resistant interactions.

**Interactions for double-drug resistance of IDV, NFV and SQV**

In order to study the interaction patterns for double-drug resistance of three PIs, we applied BVP and RMS to sequences from double-PI treatments: 280 from IDV-NFV treatment, 143 from IDV-SQV treatment, 96 from NFV-SQV treatment. Figure 4 shows the results for IDV-NFV data, while Figure S1 and S2 in the Supplementary Materials show those for IDV-SQV data and NFV-SQV data. For IDV-NFV treatment, in Figure 4, all of the 7 positions BVP detected are either from IDV single drug resistance [13], or from NFV single drug resistance (see Figure 1). For example, 54, 82 and 90 are interactively associated with both IDV and NFV treatments, while 24 and 73 are only associated with IDV treatment, 30 and 88 from NFV treatment. But there are only 7 mutation positions (Figure 4), instead of 13, as conjectured, from single-drug treatments of IDV and NFV. In other words, 6 positions (10, 20, 32, 46, 47, 71) disappeared comparing with single-drug treatment profiles.

Similar things happen to the dependent structure of double-drug resistance inferred by RMS. For IDV-NFV double-drug treatment, we found three resistant interaction groups: 24, 54 and 82 are from one independent interaction group in IDV resistance [13], while 73 and 90 are from another independent group of IDV [13]; 30 and 88 are from NFV independent group. All of these independence structures remain the same for double-drug resistance. However, we did not see the conditional independence of 46, 54 and 82 (which was observed in IDV drug resistance) in IDV-NFV treatment. In NFV single-drug treatment, 90 is in the group of 20, but in IDV-NFV double-drug treatment, 20 is not even detected by BVP (Figure 4). So again comparing with our conjecture from single-drug profiles, the group of 10 and 71 from IDV disappeared.

As for IDV-SQV data, we observed similar results (Figure S1): most of the positions detected are from IDV or SQV single-drug resistance, and two independent groups (46, 54, 82, and 73, 90) are also the same as the ones from IDV single-drug resistance. However, due to the small sample size (143 sequences), we were not able to infer the details of each group. 20, 36, 61, 63 are not in single-drug profiles.
of IDV or SQV, but in double-drug treatment profile of IDV-SQV. 10, 22, 23, 48, 79 and 84 are in single-drug profiles but not double-drug profile. For NFV-SQV (Figure S2), again the sample size is too small (96 sequences), so the posterior probabilities shown in Figure S2a are very moderate, thus the details of the dependence structure cannot be inferred (Figure S2b).

Interactions for triple-drug resistance of IDV, NFV and SQV
We also studied triple-drug resistance of the three drugs by applying BVP and RMS to 201 sequences from three PI treated patients (note that this sample size is even larger than double-drug treatments of IDV-SQV and NFV-SQV). Figure 3 shows the associated positions and the detailed dependent structure. The results are very similar to double-drug IDV-NFV resistance (Figure 4) except that 46 is observed again. However, only 8 positions are involved in resistant interactions, much smaller than what we conjectured from single-drug profiles 18. It seems that only the ones repetitively detected in single-drug, double-drug profiles are detected in triple-drug profile. In Figure 3b, three independent interaction groups were inferred: 30 and 88; 73 and 90; 24, 46, 54 and 82. The detailed structure inside the last group (24, 46, 54 and 82) was ambiguous due to the small sample size. These interaction groups are also showing up repetitively in single- and double-drug profiles.

Multiple-drug treatment effect
Observing the results of double-drug treatments and triple-drug treatment, we see they are very different from what we conjectured from single-drug resistance profiles. In double-drug IDV-NFV treatment profile, there are only 7 mutation positions (Figure 4) in interactions, not 13 as conjectured from single-drug treatments of IDV and NFV. In triple-drug IDV-NFV-SQV treatment profile, only 8 are in Figure 3, not 18, as conjectured from single-drug profiles. Thus in multiple-drug treatment, the resistance profile is not simply additive from single-drug profiles. There are some multi-drug treatment effects that select mutation interactions from single-drug profiles. In order to figure out what is this selection effect, we further did Molecular Dynamic (MD) Simulations and Free Energy Calculations of two interaction groups: 54 and 82; 30 and 88 (since it is well-known that the resistance caused by 73 and 90 cannot be explained by the binding free energy analysis 13, so we did not do MD simulation or free energy analysis on this group). Both are repetitively detected in single-, double- and triple-drug resistance profiles. We hope to figure out what makes them selected in multiple-drug profiles from single-drug profile.

Molecular Basis of Interacting Mutations Revealed by MD Simulations and Free Energy Calculations
From above statistical analysis we can see the three independent interaction groups: 54 and 82; 30 and 88; 73 and 90, repeatedly being detected in single-, double- and triple-drug resistance. To further investigate the molecular implications of the mutation interactions within the three groups across three different PIs, we conducted MD simulations to analyze the binding free energies of the protease/IDV/NFV/SQV complexes. The results are shown in Table1, from which we can see that the interaction effects of 54V/82A in all of the three drugs are very strong but in different directions: 54V has individual resistance in neither IDV nor NFV but it amplifies the one from 82A in IDV while neutralizes 82A in NFV; in SQV 54V does have some intermediate resistance but 82A does not (even more susceptible than the wild type), so 82A actually neutralizes 54V for 54V/82A. Because the interaction effects are strong in all the drugs, the interaction group of 54 and 82 are detected in all single-, double- and triple-drug treatments. Figure S5 shows the difference between the drug-protease residue interactions of the wild-type complex and those of the L82A mutated complex for (a) IDV, (b) NFV and (c) SQV. Figure S6 shows the structure alignment of the wild-type and the L82A mutated complexes of IDV and the structural distributions of eight important residues labeled in Figure S5a.

The interaction group of 30 and 88 behaves very differently from 54 and 82, since there are two possible mutations at 88: 88D or 88S. The interaction effect is most evident when we compare 30N/88S and 30N/88D in NFV: neither 30N nor 88S has individual resistance to NFV, but 88D does; more importantly, 30N/88D has intermediate resistance but 30N/88S is more susceptible to NFV than the wild type. This strong interaction effect is the main reason that 30 and 88 are detected in any single-, double- and triple-drug treatment where NFV is involved.

Figure S5 shows the difference between the drug-protease residue interactions of D30N/88S and those of D30N/88D for (a) IDV, (b) SQV and (c) NFV. It is quite clear that for IDV, there is very minimum difference; for SQV, some intermediate level of difference were observed; however, for NFV, very strong difference were observed at positions 50 and 50’. This observation is consistent with our statistical inference for IDV, SQV and NFV. Figure 6 shows the structural distribution of important residues labeled in Figure S5b for SQV and Figure Sc for NFV. The D30N/88S HIV-1 protease mutant is shown as blue strand and SQV or NFV is shown as green stick. The favorable residues to the binding of SQV or NFV to the D30N/88S mutant are shown as the yellow CPK model, those of the unfavorable residues are the red CPK model, and the residues at positions 30 and 88 are the green CPK model. In both cases of SQV (Figure 6a) and NFV (Figure 6b), the distributions of favorable and unfavorable residues are very unbalanced.

![Figure S2](image-url)
Table 1. The binding free energies and the energy components calculated by MM/GBSA (kcal/mol)

<table>
<thead>
<tr>
<th></th>
<th>∆E_vdw</th>
<th>∆E_ele</th>
<th>∆G_gb</th>
<th>∆G_sa</th>
<th>∆E_ele+∆G_gb</th>
<th>∆E_vdw+∆G_sa</th>
<th>∆G_cal</th>
<th>∆∆G_cal</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild</td>
<td>-74.44</td>
<td>0.78</td>
<td>-18.38</td>
<td>0.35</td>
<td>26.67</td>
<td>0.08</td>
<td>-10.01</td>
<td>0.02</td>
</tr>
<tr>
<td>30N</td>
<td>-69.55</td>
<td>1.10</td>
<td>-20.9</td>
<td>1.49</td>
<td>25.68</td>
<td>0.38</td>
<td>-9.92</td>
<td>0.12</td>
</tr>
<tr>
<td>54V</td>
<td>-79.12</td>
<td>1.36</td>
<td>-8.42</td>
<td>0.39</td>
<td>22.17</td>
<td>0.11</td>
<td>-9.57</td>
<td>0.61</td>
</tr>
<tr>
<td>73S</td>
<td>-77.92</td>
<td>0.85</td>
<td>-17.16</td>
<td>0.53</td>
<td>24.52</td>
<td>0.55</td>
<td>-10.22</td>
<td>0.06</td>
</tr>
<tr>
<td>82A</td>
<td>-72.26</td>
<td>0.00</td>
<td>-31.67</td>
<td>1.07</td>
<td>34.85</td>
<td>1.11</td>
<td>-9.34</td>
<td>0.58</td>
</tr>
<tr>
<td>88S</td>
<td>-71.66</td>
<td>0.95</td>
<td>-16.16</td>
<td>0.72</td>
<td>25.84</td>
<td>0.73</td>
<td>-9.83</td>
<td>0.16</td>
</tr>
<tr>
<td>88D</td>
<td>-77.23</td>
<td>0.60</td>
<td>-17.82</td>
<td>1.94</td>
<td>26.89</td>
<td>1.69</td>
<td>-9.43</td>
<td>0.68</td>
</tr>
<tr>
<td>90M</td>
<td>-73.44</td>
<td>1.64</td>
<td>-23.66</td>
<td>5.89</td>
<td>29.87</td>
<td>3.53</td>
<td>-10.18</td>
<td>0.10</td>
</tr>
<tr>
<td>30N/88S</td>
<td>-73.97</td>
<td>3.00</td>
<td>-20.42</td>
<td>1.50</td>
<td>27.25</td>
<td>0.57</td>
<td>-9.93</td>
<td>0.03</td>
</tr>
<tr>
<td>30N/88D</td>
<td>-74.06</td>
<td>2.14</td>
<td>-14.13</td>
<td>0.20</td>
<td>23.11</td>
<td>1.19</td>
<td>-10.01</td>
<td>1.07</td>
</tr>
<tr>
<td>54V/82A</td>
<td>-75.15</td>
<td>1.31</td>
<td>-18.86</td>
<td>2.01</td>
<td>27.30</td>
<td>2.19</td>
<td>-9.76</td>
<td>0.05</td>
</tr>
<tr>
<td>73S/90M</td>
<td>-69.29</td>
<td>0.14</td>
<td>-24.12</td>
<td>1.12</td>
<td>30.02</td>
<td>0.18</td>
<td>-10.53</td>
<td>0.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>∆E_vdw</th>
<th>∆E_ele</th>
<th>∆G_gb</th>
<th>∆G_sa</th>
<th>∆E_ele+∆G_gb</th>
<th>∆E_vdw+∆G_sa</th>
<th>∆G_cal</th>
<th>∆∆G_cal</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild</td>
<td>-67.48</td>
<td>0.50</td>
<td>-24.48</td>
<td>2.65</td>
<td>27.91</td>
<td>2.21</td>
<td>-9.48</td>
<td>0.47</td>
</tr>
<tr>
<td>30N</td>
<td>-69.98</td>
<td>0.44</td>
<td>-18.81</td>
<td>0.11</td>
<td>24.11</td>
<td>0.09</td>
<td>-9.69</td>
<td>0.66</td>
</tr>
<tr>
<td>54V</td>
<td>-68.93</td>
<td>0.58</td>
<td>-23.04</td>
<td>2.18</td>
<td>26.84</td>
<td>2.24</td>
<td>-10.24</td>
<td>0.06</td>
</tr>
<tr>
<td>73S</td>
<td>-69.49</td>
<td>0.59</td>
<td>-20.31</td>
<td>0.74</td>
<td>25.30</td>
<td>0.91</td>
<td>-8.33</td>
<td>0.53</td>
</tr>
<tr>
<td>82A</td>
<td>-73.07</td>
<td>0.81</td>
<td>-20.65</td>
<td>0.44</td>
<td>-24.94</td>
<td>1.66</td>
<td>-8.43</td>
<td>0.65</td>
</tr>
<tr>
<td>88S</td>
<td>-70.93</td>
<td>1.19</td>
<td>-14.67</td>
<td>0.72</td>
<td>21.91</td>
<td>0.01</td>
<td>-9.15</td>
<td>0.16</td>
</tr>
<tr>
<td>88D</td>
<td>-67.13</td>
<td>0.27</td>
<td>-23.34</td>
<td>0.62</td>
<td>29.71</td>
<td>1.01</td>
<td>-9.53</td>
<td>0.50</td>
</tr>
<tr>
<td>90M</td>
<td>-69.03</td>
<td>0.64</td>
<td>-21.81</td>
<td>0.45</td>
<td>26.54</td>
<td>0.80</td>
<td>-8.53</td>
<td>0.63</td>
</tr>
<tr>
<td>30N/88S</td>
<td>-71.18</td>
<td>1.05</td>
<td>-18.62</td>
<td>0.16</td>
<td>23.78</td>
<td>0.38</td>
<td>-8.57</td>
<td>0.59</td>
</tr>
<tr>
<td>30N/88D</td>
<td>-69.37</td>
<td>0.91</td>
<td>-18.88</td>
<td>0.06</td>
<td>23.71</td>
<td>0.25</td>
<td>-7.12</td>
<td>0.56</td>
</tr>
<tr>
<td>54V/82A</td>
<td>-66.79</td>
<td>1.73</td>
<td>-24.07</td>
<td>1.31</td>
<td>28.09</td>
<td>0.36</td>
<td>-10.07</td>
<td>0.17</td>
</tr>
<tr>
<td>73S/90M</td>
<td>-69.12</td>
<td>2.10</td>
<td>-19.79</td>
<td>0.33</td>
<td>24.27</td>
<td>0.09</td>
<td>-8.48</td>
<td>0.60</td>
</tr>
</tbody>
</table>

*a∆G_cal* is the difference between the binding free energy of the mutated complex and that of the wild-type (WT).

*bStandard deviations were estimated from two block average values.

Figure 5. The difference between the drug-protease residue interactions of D30N/N88S and those of D30N/N88D for (a) IDV, (b) SQV and (c) NFV. The residues with absolute value larger than 1.00 kcal/mol are labeled. (∆∆G equals to the interactions between drug and the D30N/N88S protease mutant minus those between drug and the D30N/N88D protease mutant).

Figure 6. Structural distributions of important residues labeled in (a) Figure 5b and (b) Figure 5c. The D30N/N88S HIV-1 protease mutant is shown as blue strand and SQV (a) or NFV (b) is shown as green stick. The favorable residues to the binding of SQV to the D30N/N88S mutant are shown as the yellow CPK model, those of the unfavorable residues as the red CPK model, and the residues at positions 30 and 88 as the green CPK mode.
Figure S4. The difference between the drug-protease residue interactions of the wild-type complex and those of the N88S mutated complex for (a) IDV, (b) NFV and (c) SQV. More positions are affected in SQV compared with IDV and NFV by N88S single mutation.

Figure S5 shows the difference between the drug-protease residue interactions of the wild-type complex and those of the L88A mutated complex for (a) IDV, (b) NFV and (c) SQV. The residues with absolute value larger than 1.00 kcal/mol are labeled. ($\Delta G$ difference equals to the interactions between drug and the wild-type protease minus those between drug and the mutant).

Figure S6. Structure alignment of the wild-type and the L82A mutated complexes of IDV and the structural distributions of eight important residues labeled in Figure S5a. The wild-type complex is shown as red strand, the mutated complex as green strand and IDV is shown as red or green stick. The favorable residues to the binding of IDV with the wild-type protease are shown as the red CPK model, and the unfavorable residue as the green CPK model. The structures are averaged from 0.5 to 3 ns. The root mean square deviation (RMSD) of the main-chain atoms for the two averaged structures after structure alignment is 0.97 Å.
simulations, we were able to detect and analyze, beyond the single-drug level, the complex interactions of drug resistance mutations and cross-mutations of the HIV-1 protease and reverse transcriptase. Most significantly, we applied the said method to all possible combinations of single-, double-, and triple-drug treatment data on IDV, NFV, and SQV from the Stanford HIV-1db database to study the problem of cross-resistance, revealing the selection pattern of the puzzled problem of multiple-drug treatment effects. We find that in general, the interactions effect is likely to be strengthened when it has strong interaction with at least one other drug in the same treatment. Hence it seems that simultaneous multi-drug treatment is more effective than sequential multi-drug treatment.

There are, admittedly, many complexities not addressed in this paper. For example, the HIV data from Stanford HIV db are collected from a global sample and thus multiple subpopulations in both treated and untreated population are possible, which may render our inference biased. Another issue may lie in the quasi-species nature of HIV-1. The HIV-1 population within an individual consists of innumerable variants and minor variants that often go undetected and hence there exists the possibility that our data underrepresent those variants. Furthermore, HIV-1 drug resistance can be not only acquired (developing in a person receiving antiretroviral treatment) but also transmitted (occurring because a virus with drug-resistance mutations was transmitted to a drug-naive person). In recent years, the transmitted resistance occurrence has been increasing due to scaled-up antiretroviral treatments. In Europe, North America, and Brazil, it has been reported that the prevalence of drug resistance ranges from 5–15% in newly diagnosed individuals. Because our untreated sequence data were collected from 1982 to 2005, it is possible that there are several transmitted drug resistant sequences in the untreated group that may affect both the sensitivity of our BVP algorithm and the power of our Bayesian model structure inference method. Because there are many antiretroviral drugs, cross-resistance is a severe and practical problem.

Despite all the possibilities that may emerge, this study has verified the original findings of HIV drug resistance and demonstrated the long-puzzled selection pattern of HIV multiple drug treatment effects. We are positive that the method and results presented here will enlighten new and more accurate ways to decipher the myths behind drug resistance of HIV and other related diseases.

Competing interests
The authors declare that they have no competing interests.

Acknowledgement
J Zhang was supported by start-up funding and Sesseel Award from Yale University. The computation was done with the help from the Yale University Biomedical High Performance Computing Center, which was supported by the NIH grant RR19895.

Author information
1. Department of Chemistry and Biochemistry 4254 Urey Hall, UCSD, 9500 Gilman Drive La Jolla, CA 92039-0359.
2. Functional Nano & Soft Materials Laboratory, Soochow University, Suzhou 215123, P. R. China.
3. Department of Statistics, Harvard University, 1 Oxford street, Cambridge, MA 02138 USA.

References