JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Using Hierarchical Virtual Screening To Combat Drug Resistance of the HIV-1 Protease

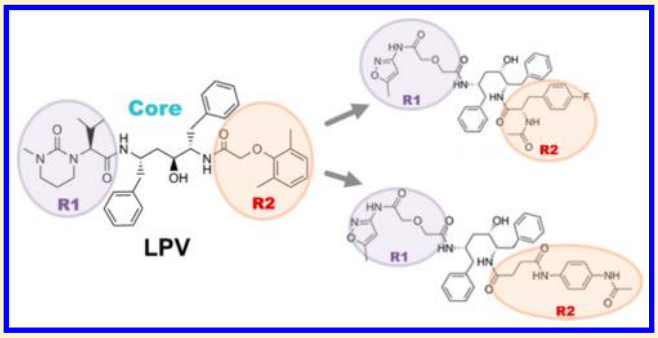
Nan Li,[†,§] Richard I. Ainsworth,[†,§] Bo Ding,[†] Tingjun Hou,*[,‡] and Wei Wang*[,†]

[†]Department of Chemistry and Biochemistry University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0359, United States

[‡]College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China

Supporting Information

ABSTRACT: Human immunodeficiency virus (HIV) protease inhibitors (PIs) are important components of highly active anti-retroviral therapy (HAART) that block the catalytic site of HIV protease, thus preventing maturation of the HIV virion. However, with two decades of PI prescriptions in clinical practice, drug-resistant HIV mutants have now been found for all of the PI drugs. Therefore, the continuous development of new PI drugs is crucial both to combat the existing drug-resistant HIV strains and to provide treatments for future patients. Here we purpose an HIV PI drug design strategy to select candidate PIs with binding energy distributions dominated by interactions with conserved protease residues in both wild-type and various drug-resistant mutants. On the basis of this strategy, we have constructed a virtual screening pipeline including combinatorial library construction, combinatorial docking, MM/GBSA-based rescoring, and reranking on the basis of the binding energy distribution. We have tested our strategy on lopinavir by modifying its two functional groups. From an initial 751 689 candidate molecules, 18 candidate inhibitors were selected using the pipeline for experimental validation. IC_{50} measurements and drug resistance predictions successfully identified two ligands with both HIV protease inhibitor activity and an improved drug resistance profile on 2382 HIV mutants. This study provides a proof of concept for the integration of MM/GBSA energy analysis and drug resistance information at the stage of virtual screening and sheds light on future HIV drug design and the use of virtual screening to combat drug resistance.



INTRODUCTION

The human immunodeficiency virus type-1 aspartyl protease (HIV-1 PR) generates mature infectious virions through clevage of the viral Gag and GagPol precursor proteins[1] and is a major therapeutic target for HIV/AIDS. The HIV-1 PR is a homodimer with 99 residues in each monomer and contains the highly conserved catalytic triads D25/25′−T26/26′−G27/27′. The competitive inhibition of HIV-1 PR, in combination with other inhibitors (reverse transcriptase, integrase, and fusion) and coreceptor antagonists, forms the basis of highly active anti-retroviral therapy (HAART). This combination HIV-1 chemotherapy suppresses viral replication and can reduce the plasma HIV-1 viral load to <50 RNA copies mL^{-1}, resulting in a significant reconstitution of the immune system. Five HIV-1 protease inhibitors (PIs) were approved by the FDA between 1995 and 1999, including saquinavir (SQV), ritonavir (RTV), indinavir (IDV), nelfinavir (NFV), and amprenavir (APV). These drugs suffer from limitations such as low bioavailability and short half-life, necessitating multiple daily dosing. Other problems include extensive toxicity, severe gastrointestinal symptoms, and the development of broad class resistance from multiple HIV-1 PR resistance mutations leading to therapy failure. Several PIs were approved by the FDA between 2000 and 2006, including lopinavir (LPV), atazanavir (ATV), the amprenavir prodrug fosamprenavir (FPV), tipranavir (TPV), and darunavir (DRV). All of these PIs except for TPV and DRV are competitive peptidomimetic inhibitors in that they contain the nonscissile hydroxyethylene [$-CH_2-CHOH-$] bond instead of the readily hydrolyzable peptide [$-CO-NH-$] bond[2] and mimic the natural HIV-1 PR substrate. Their associated therapies are developed and administered so as to increase the level of PI in the plasma. For example, "boosting" with a subtherapeutic dose of RTV (which also acts as a cytochrome P450 3A4 enzyme inhibitor) leads to reduced liver metabolism of the given PI and thus to increased PI plasma levels and half-life.

The poor proofreading capabilities of the viral reverse transcriptase introduces an estimated one mutation for every 10^3-10^4 nucleotides synthesized, representing 1−10 mutations in each viral genome, with every replication cycle. Under the selection pressure of a given drug regimen, the virus rapidly evolves a reduced susceptibility while maintaining substrate recognition, thus decreasing the efficacy of the treatment.

Received: February 2, 2015
Published: May 20, 2015

DOI: 10.1021/acs.jcim.5b00056
J. Chem. Inf. Model. 2015, 55, 1400−1412

Previous studies[3−5] have classified the HIV-1 PR residues into four classes according to their importance with regard to drug resistance across all PIs. There are 15 "major" positions, which are sufficient to cause drug resistance when mutated. In addition, there are 23 "minor" positions, which usually co-occur with major mutations and have no effect when mutated alone. Recent reviews have proposed a stepwise drug resistance mechanism to explain the temporal relationship between major and minor mutations.[6,7] Twenty-six positions are "conserved" throughout the HIV mutant database (typically lower than 0.1% mutation rate) and are closely related to protease folding, dimerization, and catalytic activity. Finally, there are 35 positions defined as "other", which show an unclear relationship to drug resistance on the basis of currently available mutant isolates. It is suggested that designing drugs to target conserved residue positions will reduce the ability of the virus to evolve into drug-resistant strains.[8]

LPV (formally ABT-378) is commonly regarded as the first "second-generation" PI approved for clinical use and was developed to be active against RTV-resistant isolates of HIV-1. LPV/r or ATV/r are used as first-line options for PI-based HAART. According to the latest information in the HIV Drug Resistance Database,[4] as of March 9, 2015, mutations leading to reduced susceptibility or virological response for LPV/r (RTV-boosted) are V32I, I47V, I47A, G48M, I50V, I54V, I54T, I54A, I54L, I54M, L76V, V82A, V82F, V82S, V82T, and I84V. These mutations lower the free energy difference between separated LPV + HIV-1 PR and the bound complex.

Virtual screening detects lead compounds on target proteins in a fast and efficient manner, thus narrowing down the list of potential compounds for timely and costly wet-lab tests. It is feasible to dock millions of ligands to the target protein in an affordable time. Several molecular docking methods have been designed to date, such as DOCK,[9] AUTODOCK,[10,11] FlexX,[12,13] Glide,[14−16] and GOLD.[17,18] They have been widely used in the search for lead molecules as an early step in drug design.[19] For HIV-1 PR, there have been several successful virtual screening applications for novel inhibitors using molecular docking methods,[20−24] the pharmacophore model,[22,25] and molecular dynamics (MD)-based methods.[26] However, most studies do not include the drug resistance information on HIV-1 PR at the inhibitor design stage.

The molecular mechanics/generalized Born surface area (MM/GBSA) method is an energy calculation method that is widely used in binding energy analysis of protein−ligand and protein−protein systems.[27,28] The energy decomposition contains the molecular mechanics terms, which include the van de Waals and electrostatics energies. Also described are the solvation energy, including the surface area (SA) nonpolar solvation term, the generalized Born (GB) polar solvation term, and the entropy. MM/GBSA can also be decomposed per residue or per residue−residue pair. MM/GBSA has been widely applied to binding energy analysis of protein−ligand systems, such as binding energy calculations,[29−32] drug resistance analysis,[5,8,33−35] binding specificity analysis,[36−39] and docking rescoring.[40−42] In constrast to the linear interaction energy (LIE) method, MM/GBSA does not rely on the availability of experimental binding affinity data that are crucial for training LIE models. The conductor-like screening model (COSMO)[43] performs better than MM/GBSA in some cases.[44,45] However, the heavy computational requirement of COSMO makes it less applicable in screening pipelines. Furthermore, the prediction reliability of MM/GBSA has been well-validated by our previous studies on the HIV-1 PR systems.[5,34]

In this study, we purpose a hierarchical virtual screening pipeline that integrates combinatorial library construction, combinatorial docking, MM/GBSA rescoring, and drug-resistant analysis based on MM/GBSA decomposition. We started with the second-generation HIV-1 PI LPV and modified its two functional groups. From an initial 751 689 ligands, we identified 18 candidate inhibitors whose scores topped the list ranked to consider both binding affinity and binding energy contribution to conserved residues of HIV-1 PR. We verified the inhibitory activities by measuring the binding affinities of the candidate inhibitors. Three inhibitors show HIV-1 PR inhibition abilities in the micromolar range. This successful result shows the effectiveness of the virtual screening pipeline and also emphasizes that drug resistance should be considered in future virtual screening pipelines.

METHODS

1. Virtual Combinatorial Library. According to the position of the peptide bonds, the structure of LPV can be divided into three parts: R1, the core, and R2 (Figure 1). On

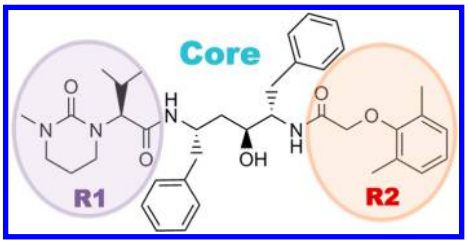


Figure 1. Fragmentation of the structure of lopinavir (LPV).

the basis of the core structure of LPV, the combinatorial library was generated by attaching functional carboxylic groups to the R1 and R2 positions. An initial 3D structure of the core scaffold was created using the SYBYL molecular simulation package[46] (two hydrogen atoms were used as the attachment points for R1 and R2). The 3D structures of the R1 and R2 groups were selected from the ZINC database[47] and used as input to the Legion program in SYBYL for the generation of the combinatorial library. The chirality in all of the structures was maintained during the process of library generation. A total of 867 carboxylic acids with molecular weights between 150 and 200 g mol^{-1} and availability from the vendors Sigma-Aldrich and Maybridge were used. This resulted in a combinatorial library of 751 689 candidate ligands, which was saved as SDF mol and Tripos mol2 library files for subsequent calculations.

2. Combinatorial Docking. The crystal structure of wild-type HIV-1 PR complexed with LPV (PDB accession code 1mui)[48] was used as the template for molecular docking. All of the water molecules in the crystal structure were removed. The PR residues within 8 Å of LPV were defined as the binding pocket for molecular docking. Each molecule in the combinatorial library was subsequently docked to the binding site of the PR using the CombiFlexX combinatorial docking approach in SYBYL. The combinatorial docking starts with a set of core structure placements (aligning the core structure in the ligand with the predefined core in the binding site) and builds up the full compound by adding the variants at the R1 and R2 positions consecutively. Here, single core placement (placement of the core of LPV in the crystal structure of the

DOI: 10.1021/acs.jcim.5b00056
J. Chem. Inf. Model. 2015, 55, 1400−1412

complex) was adopted in the combinatorial docking, since only the flexibility of the functional groups at R1 and R2 needs to be considered. The binding affinities of the docked molecules were estimated using the FlexX scoring function.[12]

3. Reranking of the Docking Results by MM/GBSA. The MM/GBSA method was applied to rescore the top 50 000 molecules from FlexX docking. All of the ligands were optimized using the Broyden−Fletcher−Goldfarb−Shanno (BFGS) technique. A subsequent AM1-BCC partial charge derivation was performed for all of the ligand atoms using the semiempirical quantum method implemented in the *divcon* program[49] in AMBER9.[50] AM1-BCC is significantly faster than HF/6-31G* charge calculation.[51] Previous studies also showed that MM/PBSA using AM1-BCC and HF/6-31G* RESP charges gave compatible binding free energies.[52] The AMBER03 force field was used for proteins,[53] and the general AMBER force field (*gaff*) was used for ligands.[54] The calculations of the partial charges and the generation of the force field parameters for all of the molecules were accomplished using the *antechamber* program in AMBER9.[55] The solvent effect was considered using the GB model (*igb* = 2) implemented in *sander*.[56] The docking complex was minimized using the *sander* program in AMBER9.[50] The maximum number of minimization steps was set to 4000, and the convergence criterion for the root-mean-square (rms) of the Cartesian elements of the energy gradient was 0.05 kcal mol^{-1} Å^{-1}. The first 500 steps were performed with the steepest-descent algorithm and the remaining steps with the conjugate-gradient algorithm.

For each minimized protein/ligand complex, the binding free energy was estimated using the MM/GBSA technique according to eq 1:[8,57,58]

$$\begin{aligned}\Delta G_{\text{bind}} &= G_{\text{complex}} - G_{\text{protein}} - G_{\text{ligand}} \\ &= \Delta E_{\text{MM}} + \Delta G_{\text{GB}} + \Delta G_{\text{SA}} \\ &= \Delta E_{\text{vdW}} + \Delta E_{\text{ele}} + \Delta G_{\text{GB}} + \Delta G_{\text{SA}} \end{aligned} \quad (1)$$

where ΔE_{MM} is the gas-phase interaction energy between the protein and the ligand, including the electrostatic energy (ΔE_{ele}) and the van der Waals energy (ΔE_{vdW}), and ΔG_{GB} and ΔG_{SA} are the polar and nonpolar components of the desolvation free energy, respectively. The electrostatic solvation energy (ΔG_{GB}) was calculated using the modified GB model developed by Onufriev and co-workers.[56] The values 80 and 2 were used for the exterior and interior dielectric constants, respectively. The nonpolar contribution was determined from the solvent-accessible surface area (SASA) using the LCPO method:[59] $\Delta G_{\text{SA}} = 0.0072 \times \Delta(\text{SASA})$. Conformational entropy was not evaluated because of its extremely high computational cost and its ineffectiveness in improving the binding energy correlation.[40] The 50 000 protease/ligand complexes were evaluated by MM/GBSA and ranked on the basis of the binding energy.

4. MD Simulations for Top-Ranked Candidates. The top 300 molecules ranked by MM/GBSA in the previous step were chosen for the following MD simulations. Each wild-type protein/ligand complex was immersed in a rectangular box of TIP3P water molecules.[60] The water box was extended 10 Å away from any solute atom. The particle mesh Ewald (PME) method[61] was employed to calculate the long-range electrostatic interactions. The complexes were first relaxed using 2000 cycles of the MM minimization procedure (500 cycles of steepest-descent minimization and 1500 cycles of conjugate-gradient minimization). Next, the system was gradually heated in the *NVT* ensemble from 10 to 300 K over 20 ps. Initial velocities were assigned from a Maxwellian distribution at the starting temperature. Subsequently, 500 ps MD simulations in the *NPT* ensemble with a target temperature of 300 K and a target pressure of 1 atm were performed. The SHAKE procedure was employed to constrain all of the hydrogen atoms,[62] and the time step was set to 2.0 fs. During the sampling process, MD snapshots were saved every 2 ps after the systems were well-equilibrated. The MM optimizations and MD simulations were accomplished using *sander* in AMBER9.[50] Finally, for conformational sampling, the binding free energies were calculated using the MM/GBSA method averaged over 75 snapshots evenly extracted from the MD trajectory between 200 and 500 ps using the *mm_pbsa.pl* program in AMBER9.

5. Drug Resistance Analysis Using Free Energy Decomposition. The MM/GBSA energy decomposition analysis was applied to calculate the interactions between each ligand and each HIV-1 PR residue (ΔG_{bind}) averaged over 75 snapshots extracted from the trajectory between 200 and 500 ps using the *mm_pbsa* program in AMBER9.[37,63] ΔG_{bind} is the sum of the van der Waals (ΔE_{vdW}), electrostatic (ΔE_{ele}), and solvation ($\Delta G_{\text{solvation}}$) energies. We focused on the binding energy contributions from 14 drug resistance sites (V32, L33, M46, I47, G48, I50, F53, I54, G73, L76, V82, I84, N88, and L90), which were collected from drug resistance data in the Stanford HIV database[64] and covered all major and minor PR drug resistance mutation positions at the time this work was carried out. On the basis of these drug-resistant properties of the HIV-1 PR residues, the binding free energies contributed by the 14 drug-resistant residues were classified as $\Delta G_{\text{ligand-res}}$. It is well-known that PR drug resistance is primarily caused by mutations that significantly affect the interaction between PR and the drug.[8,33] Therefore, a drug that forms weaker interactions with the resistance residues will have a better chance to be an efficient drug to combat resistance; that is to say, weaker $\Delta G_{\text{ligand-res}}$ is favorable for decreased drug resistance. Here we defined the index $I_{\text{resistance}}$ to estimate quantitatively the capability of a drug to combat drug resistance:

$$I_{\text{resistance}} = \text{Rank}_{\Delta G_{\text{ligand-res}}} + \text{Rank}_{\Delta G_{\text{vdW-ligand-res}}} \quad (2)$$

where $\text{Rank}_{\Delta G_{\text{ligand-res}}}$ is the rank sorted in increasing order on the basis of the total binding free energy contributed by the drug-resistant residues ($\Delta G_{\text{ligand-res}}$) and $\text{Rank}_{\Delta G_{\text{vdW-ligand-res}}}$ is the rank sorted in increasing order on the basis of the total van der Waals binding free energy contributed by the drug-resistant residues ($\Delta G_{\text{vdw-ligand-res}}$), thus further weighting the resistance metric toward the nonpolar interactions. Smaller $I_{\text{resistance}}$ values represent better drug resistance properties, i.e., drugs that have lower binding energy contributions to drug resistance sites.

It should be noted that drug resistance is not the only important factor for a real drug, as strong binding affinity is also very important. Therefore, a good drug is expected to have a good balance between binding free energy and drug-resistant capability. Here we defined another index, $I_{\text{bind-resistance}}$, to estimate drug resistance and binding according to eq 3:

$$I_{\text{bind-resistance}} = \text{Rank}_{\Delta G_{\text{bind}}} - I_{\text{resistance}} + 600 \quad (3)$$

DOI: 10.1021/acs.jcim.5b00056
J. Chem. Inf. Model. 2015, 55, 1400−1412

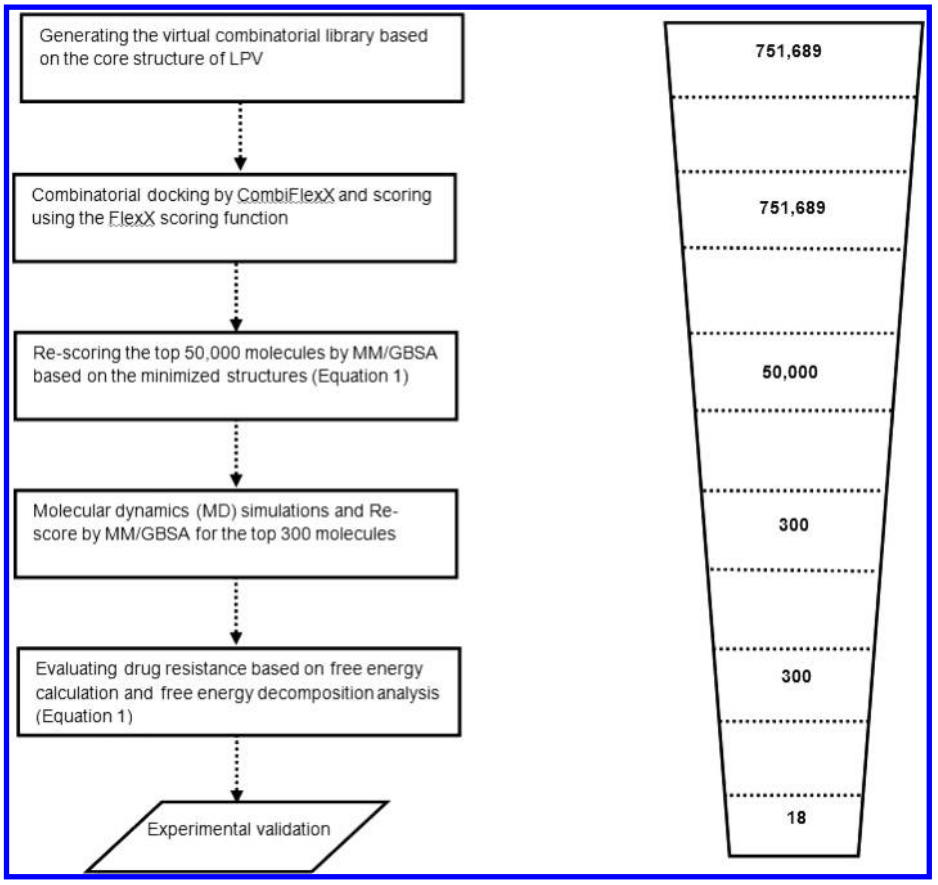


Figure 2. Hierarchical strategy to design the drug candidates to combat drug resistance.

where $\text{Rank}_{\Delta G_{\text{bind}}}$ is the rank sorted in increasing order by the total binding free energy (ΔG_{bind}). The additive integer 600 ensures that all of the $I_{\text{bind-resistance}}$ values are positive. The ligands with the 18 lowest $I_{\text{bind-resistance}}$ values were chosen for experimental validation.

6. Inhibitory Activity Measurements for Candidate Inhibitors. All of the candidate inhibitors were synthesized by Medicilon Inc. (Shanghai, China). The inhibitory activity of each candidate inhibitor was measured by PhenoSense assay, which uses a retroviral vector that contains the HIV genome without PR and reverse transcriptase (RT) and includes a luciferase reporter inserted into *env* coding region. Mutated PR or RT genes can be integrated into the vector to evaluate the inhibitory activity. In our study, the wild-type PR and RT coding region were integrated into a retroviral vector. The vector was then transfected into human embryonic kidney (HEK) 293 cell cultures in the presence of the candidate inhibitor. The viral solution collected from transfection cells was used to infect fresh HEK 293 cell cultures. The luciferase signal was collected to quantify the inhibitory activity of the candidate inhibitors. A series of inhibitor concentrations were applied to the cell cultures, and all of the collected data were fitted to a sigmoidal function for determination of the IC_{50} values. More details can be found in the paper by Petropoulos et al.[65]

7. Binding Energy Evaluation for HIV-1 Protease Mutants. The 2382 HIV-1 PR mutants were taken from a previous study.[5] Structures of the complexes of the mutants and seven ligands with inhibitory activity were generated on the basis of the post-MD complex structures of the wild-type protease and ligands. Side-chain conformations of mutated sites were predicted by SCWRL 4.0.[66] The modeled complex structures were solvated by a TIP3P water box extended to 10 Å from the solute. The mutant complex structures were then optimized for 10 000 steps with 4000 steps of steepest-descent minization followed by 6000 steps of conjugate-gradient minimization using *sander* in AMBER9.[50] Binding energies of mutants were evaluated using *mm_pbsa.pl* in AMBER9 on minimized structures. The GB model (*igb* = 2) was used for the polar part of the solvation effect and the LCPO model was used for the nonpolar part of solvation with $\Delta G_{\text{SA}} = 0.0072 \times \Delta(\text{SASA})$. Conformational entropy was not evaluated because of its extremely high computational cost.

Further analysis included the calculation of the difference in average residue binding energy contribution between LPV and the candidate ligands ($\Delta G^{\text{res}}_{\text{bind-diff}}$) according to eq 4:

$$\Delta G^{\text{res}}_{\text{bind-diff}} = \frac{1}{N_{\text{mut}}}\left\{\sum_{\text{mut}} (\Delta G^{\text{res A}}_{\text{bind-LPV}} - \Delta G^{\text{res A}}_{\text{bind-lig}}) + \sum_{\text{mut}} (\Delta G^{\text{res B}}_{\text{bind-LPV}} - \Delta G^{\text{res B}}_{\text{bind-lig}})\right\} \quad (4)$$

where $\Delta G^{\text{res A}}_{\text{bind-LPV}}$ is the binding energy contribution at the specific residue position (res) on the A-chain monomer (res A) when LPV is in complex with the given protease mutant and $\Delta G^{\text{res A}}_{\text{bind-lig}}$ is the same value for the drug candidate ligand (lig) in complex with the protease mutant. These values were also calculated for residues at equivalent positions on the B-chain

DOI: 10.1021/acs.jcim.5b00056
J. Chem. Inf. Model. 2015, 55, 1400−1412

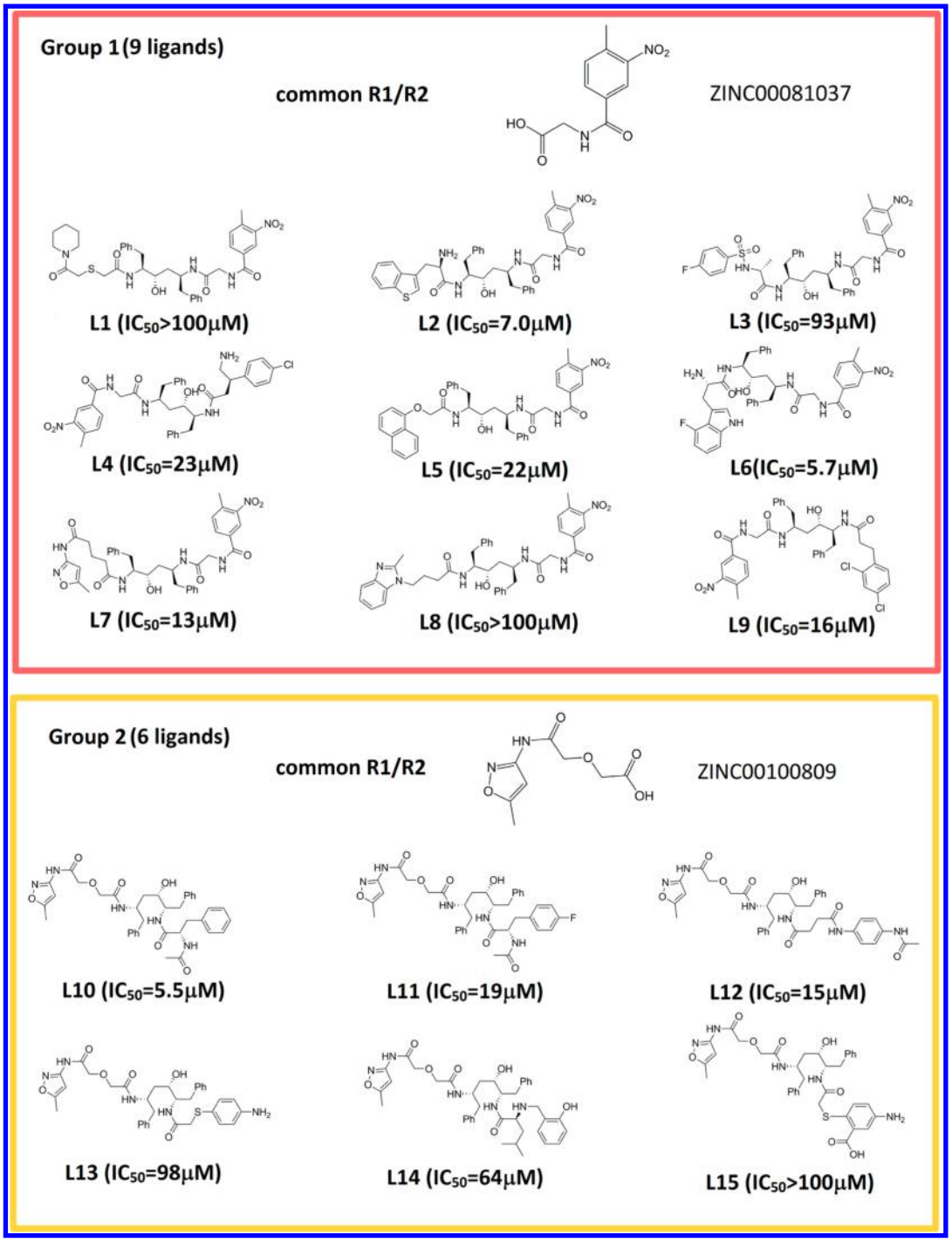


Figure 3. Skeletal formulas and IC_{50} values for L1−L15. Ph stands for the benzene ring in the LPV core structure.

monomer (res B). The summations are taken over all mutants (mut) considered, and N_{mut} is the total number of mutants.

8. Drug Resistance Ratio. We define the drug resistance ratio for a given PI as the percentage of resistant mutants for a given drug taken over all of the available mutants. Therefore, the smaller the resistance ratio is, the better is the drug's performance to combat drug resistance. The definition of a resistant mutant is taken from our previous study:[5] if the IC_{50} value for a given ligand in complex with an HIV-1 PR mutant is 10-fold weaker than that of the ligand with wild-type HIV-1 PR, the mutant is a resistant mutant with respect to the ligand.

RESULTS

1. MM/GBSA Performance on Drug Resistance. Since MM/GBSA serves as the primary scoring function in this study, we validated the performance of MM/GBSA in two different tests: the relationship between $\Delta G_{ligand\text{-}res}$ and the drug resistance ratio and the ability of ΔG_{bind} to discriminate between drug-resistant and nonresistant mutants.

First, $\Delta G_{ligand\text{-}res}$ values for binding of seven FDA-approved drugs with wild-type HIV-1 PR were evaluated and compared with their drug resistance ratios[5] (Table S1 in the Supporting Information). Of the seven drugs, four have better drug resistance ratios than LPV. When all seven drugs are ranked by

DOI: 10.1021/acs.jcim.5b00056
J. Chem. Inf. Model. 2015, 55, 1400−1412

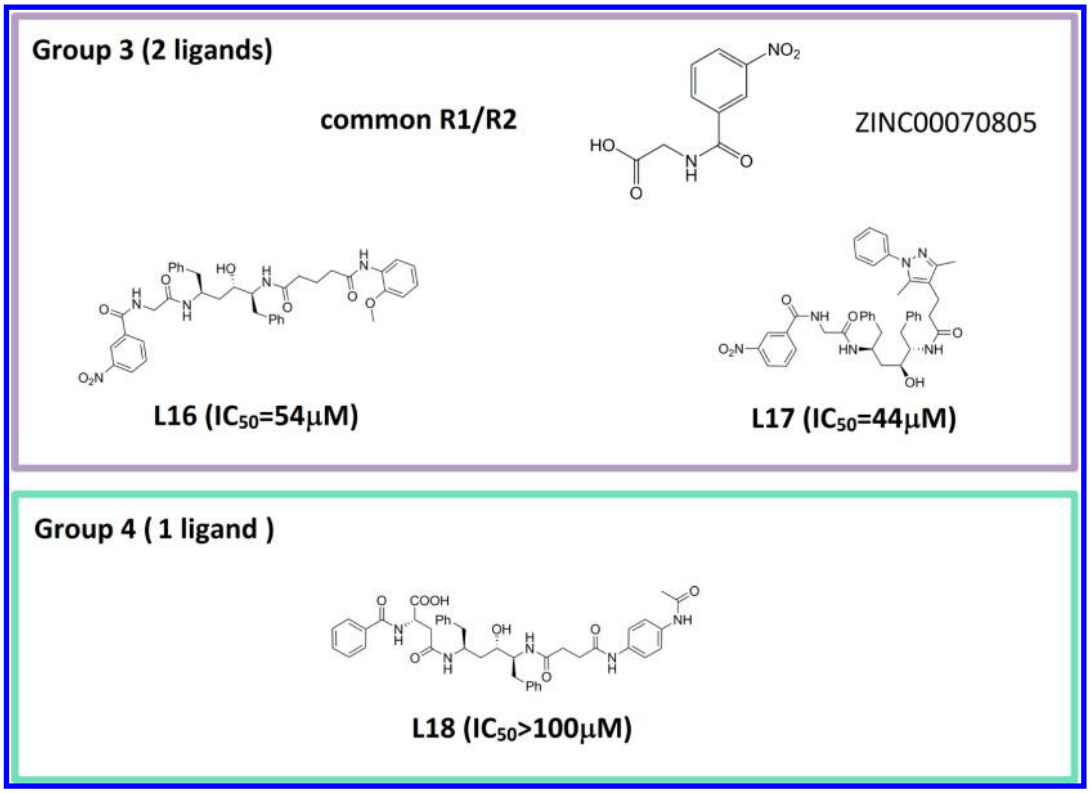


Figure 4. Skeletal formulas and IC_{50} values for L16−L18. Ph stands for the benzene ring in the LPV core structure.

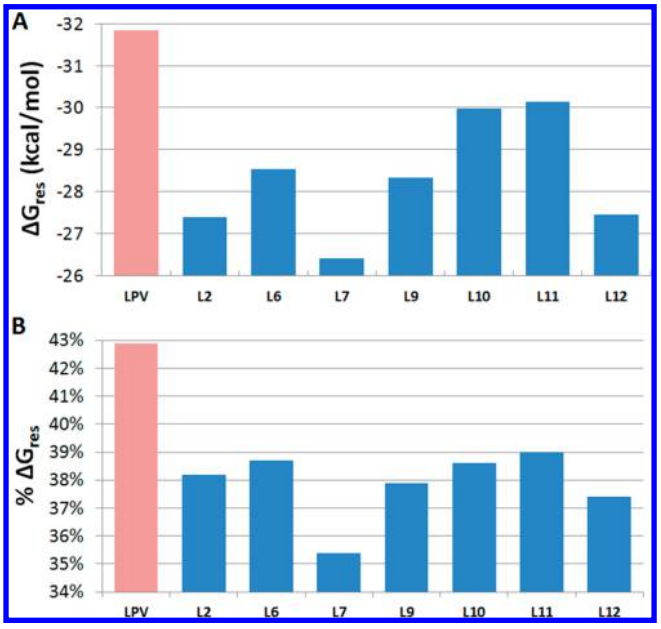


Figure 5. Binding energy contributions of drug resistance residues in LPV and ligands with inhibitory activity: (A) absolute binding energy contributions; (B) relative binding energy contributions.

$\Delta G_{ligand\text{-}res}$, for the drugs with weaker $\Delta G_{ligand\text{-}res}$ than LPV, three out of four drugs (APV, RTV, and SQV) have smaller (better) drug resistance ratios than LPV, which corresponds to an enrichment ratio of 1.3 (3/4 vs 4/7) if drugs are selected using $\Delta G_{ligand\text{-}res}$, i.e., drugs having weaker interactions with the resistance residues will likely be better candidates to combat resistance, making $\Delta G_{ligand\text{-}res}$ a good metric for HIV PR inhibitor screening.

Second, we used LPV as a sample case to evaluate the discrimination ability of ΔG_{bind} on resistant and nonresistant mutants. Complexes of 1611 HIV-1 PR mutants and LPV were modeled (see Methods 7), and ΔG_{bind} for each complex was calculated. We found that the area under the curve (AUC) of the receiver operating characteristic (ROC) curve is good (0.86) when ΔG_{bind} is used as a predictor in classification of resistance and nonresistance mutants, which indicates that resistant and nonresistant mutants can be well-separated by this single predictor (Matthews correlation coefficient (MCC) = 0.58 using the cutoff −2.614 kcal mol^{-1}; Figure S1 in the Supporting Information).

2. Hierarchical Virtual Screening Strategy. In this study, we purpose a novel concept for combating drug resistance by selecting inhibitors with a smaller binding energy contribution from highly mutable residues that contribute to resistance. The concept was tested through combinatorial screening by altering the R1 and R2 groups of LPV (see Methods 1), which was implemented by a hierarchical strategy (Figure 2). CombiFlexX combinatorial docking was applied for conformation sampling, and MM/GBSA was used for rescoring and binding energy estimation. The key part of the screening strategy is the integration of the mutation information from drug-resistant HIV strains with MM/GBSA energy decomposition analysis in order to quantify the binding energy contribution from the resistance-related protease residues. The top 50 000 ligands ranked by the FlexX docking score (see Methods 2) were selected for MM/GBSA rescoring (see Methods 3). Subsequently, the top 300 ligands were selected for 500 ps MD simulations (see Methods 4) and MM/GBSA energy decomposition analysis (see Methods 5) to evaluate the binding energy contribution from each protease residue.

The 300 ligands were ranked (see Methods 5) using three different keys: total binding energy ($Rank_{\Delta G_{bind}}$), binding energy

DOI: 10.1021/acs.jcim.5b00056
J. Chem. Inf. Model. 2015, 55, 1400−1412

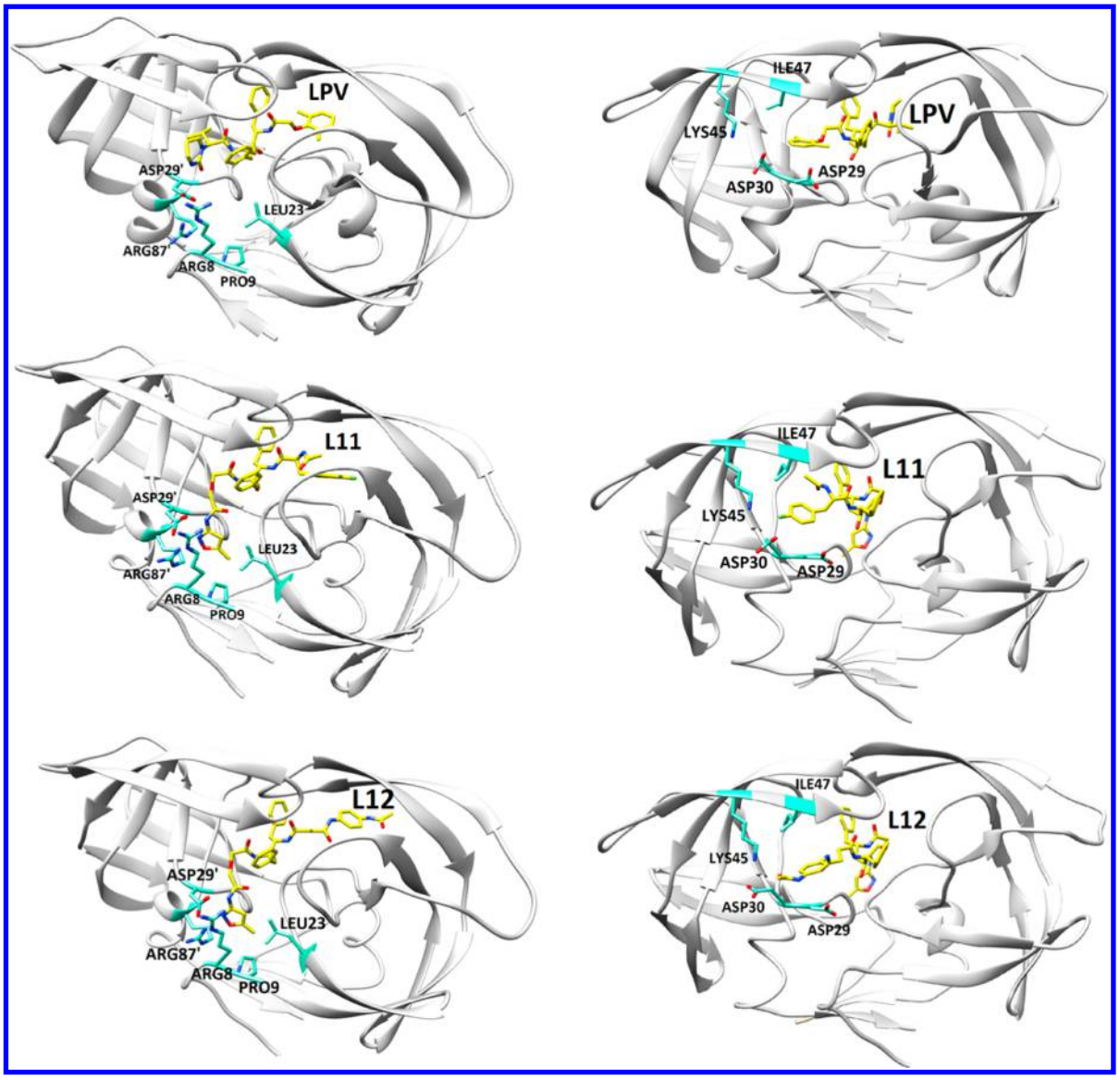


Figure 6. Binding conformations of L11 and L12 with wild-type HIV-1 PR. The left column shows the view on the R1 side; the right column shows the view on the R2 side (see Figure 1 for the definitions of R1 and R2). Highlighted residues are contact points for R1 and R2 groups. The crystal structure of LPV in complex with wild-type HIV-1 PR (PDB accession code 1mui)[48] is given for comparison.

contribution from drug resistance sites ($Rank_{\Delta G_{ligand\text{-}res}}$), and van der Waals energy contribution from drug resistance sites ($Rank_{\Delta G_{vdW\text{-}ligand\text{-}res}}$). The combination of the three ranks post-MD (eq 3) was used to select the final top 18 ligands for experimental validation (Table S2 in the Supporting Information).

The post-MD rank is similar to the pre-MD rank (r = 0.874; Figure S2 in the Supporting Information), though the MD simulation does slightly change $Rank_{\Delta G_{bind}}$ as a result of conformational averaging, with a ligand average absolute rank difference of 33. Although MD simulation improves $Rank_{\Delta G_{bind}}$ for all 18 selected ligands, from a largest improvement of 136 (L14) to a smallest of 5 (L16), the post-MD $Rank_{\Delta G_{bind}}$ still shows a considerable rank range from the best rank of 18 (L10) to the worst of 163 (L4). For $Rank_{\Delta G_{ligand\text{-}res}}$ and $Rank_{\Delta G_{vdW\text{-}ligand\text{-}res}}$, the range is smaller for the 18 ligands. The ranks of all 18 ligands range from 218 to 295, which is only ~60% of the range of $Rank_{\Delta G_{bind}}$ (5 to 136).

3. Experimental Validation of Candidate Protease Inhibitors. IC_{50} values for the 18 candidate ligands (Figures 3 and 4) were measured using PhenoSense. In total, we successfully identified seven ligands showing inhibitory activity to PR. Three ligands have IC_{50} values less than 10 μM (L2, L6, and L10), and four ligands have values between 10 and 20 μM (L7, L9, L11, and L12). The 18 ligands can be further divided into four groups according to the common R1/R2 groups in the ligand. All seven ligands showing inhibitory activity come from the two largest groups (Figure 3). Group 1 ligands share a common nitrobenzoyl carboxylic group (ZINC00081037), and group 2 ligands share a common isoxazole-containing carboxylic group (ZINC00100809).

Further observation of the top three inhibitors from groups 1 and 2 showed that the inhibitory activity of the ligands is affected by the combined effect of side-group size and polarity. In group 1, L2 (IC_{50} = 7.0 μM) and L6 (IC_{50} = 5.7 μM) both

DOI: 10.1021/acs.jcim.5b00056
J. Chem. Inf. Model. 2015, 55, 1400−1412

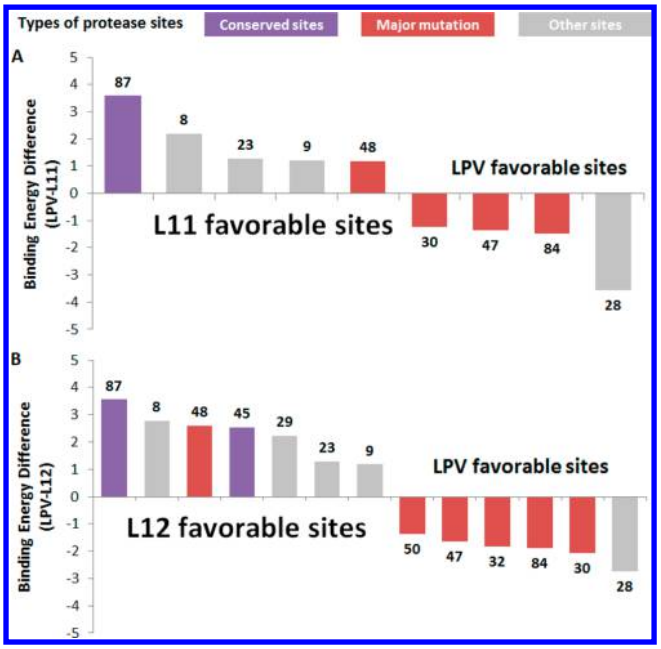


Figure 7. Average residue contributions to the binding energy difference (see eq 4) between (A) LPV and L11 and (B) LPV and L12 summed over all protease mutants and both protease monomers. Residue positions are indicated as numbers above or below the bars. Only residues that show significant variation ($|\Delta G^{res}_{bind\text{-}diff}| > 1$ kcal mol^{-1}) are presented.

contain double-ring heterocyclic groups, while L5 (IC_{50} = 22 μM) contains a more nonpolar naphthalene side group than L2 and L6, resulting in deterioration of the inhibitory activity. In group 2, L10 (IC_{50} = 5.5 μM) has a side group with both moderate size and appropriate polarity. Either altering the size of the side group, such as in L11 (IC_{50} = 19 μM) and L12 (IC_{50} = 15 μM), or altering the polarity of the side group, such as in L13 (IC_{50} = 98 μM), L14 (IC_{50} = 64 μM), and L15 (IC_{50} > 100 μM), decreases the inhibitory activity.

For the seven ligands with inhibitory activity, we re-evaluated the binding energy contributions from the drug resistance sites ($\Delta G_{ligand\text{-}res}$) and compared them with the contribution for LPV. For binding energy (Figure 5A), the $\Delta G_{ligand\text{-}res}$ contributions of all seven ligands are 2−6 kcal mol^{-1} less than that of LPV. For the proportion of $\Delta G_{ligand\text{-}res}$ to total energy (ΔG_{bind}) (Figure 5B), all seven ligands show 4−8% less contribution than LPV does. For the designed ligands, such a variation in the contribution of drug-resistant residues is well-expected, since it was carefully considered in our hierarchical strategy. A relatively low contribution from drug-resistant residues should enhance the drug effectiveness against drug-resistant HIV-1 mutants. We did not observe a strong correlation between the $\Delta G_{ligand\text{-}res}$ contribution and the IC_{50} value since the non-drug-resistant residues can compensate for the energy contribution decrement from the drug resistance residues. Nor was there a strong correlation between ΔG_{bind} and the IC_{50} value observed, since the IC_{50} is measured in vivo and is not affected by the binding affinity alone. However, we identified that ΔG_{bind} is still a good scoring function for HIV PR inhibitor screening. It can successfully select ligands with protease inhibition actitivity ($IC_{50} \leq 20$ μM), as shown by the satisfactory prediction performance (AUC = 0.805; Figure S3 in the Supporting Information) when ΔG_{bind} was used to predict ligands with protease inhibitory activity among the 18 selected ligands.

4. Drug Resistance Profiles of Candidate Protease Inhibitors. In our previous study, we investigated the drug resistance profiles of seven FDA-approved HIV-1 PI drugs via a classification of HIV-1 PR mutants as resistant (mutant binding is 10 or more times weaker than wild-type) or nonresistant. The investigation of the drug resistance profiles for candidate ligands is crucial to gain a better understanding of their effectiveness on known HIV-1 PR mutants. For this analysis, we selected all of the ligands with IC_{50} values less than 20 μM. We used two different computational methods to evaluate the drug resistance profiles of these seven selected ligands (L2, L6, L7, L9, L10, L11, and L12). First, we evaluated the total binding energy for each combination of selected ligand and protease mutant. We found that the nonresistant mutant proportions of L10 (63%), L11 (67%), and L12 (94%) are better than that of LPV (61% from the previous study[5]). We also used a more sophisticated drug resistance model, MIEC-SVM,[5] to predict drug resistance profiles for L10, L11, and L12. L12 (64%) still showed better nonresistant mutant proportions than LPV (61%), while the nonresistant mutant proportion of L11 (61%) is comparable to that of LPV. L10 (60%) has a slightly worse nonresistance proportion than LPV. Taking a combination of the two prediction results, we selected L11 and L12 as ligands with the most favorable drug resistance profiles and top candidates for future drug leads.

5. Structural Analysis of the L11 and L12 Binding Conformations. The preferred binding conformations of L11 and L12 are given in Figure 6 along with the crystal structure of LPV in complex with HIV-1 PR. For L11, L12, and LPV, we calculated the average residue contribution to the binding energy using eq 4. Protease sites were divided into ligand-favorable sites ($\Delta G^{res}_{bind\text{-}diff} > 1$ kcal mol^{-1}) and LPV-favorable sites ($\Delta G^{res}_{bind\text{-}diff} < 1$ kcal mol^{-1}). The protease sites were also classified into three categories according to their roles in PI drug resistance from the previous study: conserved sites, major mutation sites, and other sites.[5] Conserved sites are important for protease−substrate recognition and protease function, and therefore, mutations at such sites are seldom observed. Mutations at major mutation sites are closely related to drug resistance and lead to drug-resistant mutants. Sites defined as "other" show an unclear relationship to drug resistance on the basis of currently available mutant isolates.

Figure 7 shows all of the protease sites that have significant variation ($|\Delta G^{res}_{bind\text{-}diff}| > 1$ kcal mol^{-1}) between LPV and the two ligands (L11 and L12) averaged over all 2382 available mutants. L11-favorable sites contain one conserved site at position 87, while LPV-favorable sites contain more major mutation sites than L11-favorable sites (Figure 7A). In the case of LPV and L11, the energy contribution shifts from the major mutation sites to the conserved sites, thus enhancing the ligand's binding energy stability against HIV-1 PR mutants and improving the ligand's drug resistance profile compared to LPV. For L12, a more significant trend was observed (Figure 7B). L12-favorable sites were enriched with two conserved sites (45 and 87), compared with none in LPV-favorable sites. LPV-favorable sites were enriched with five major mutation sites (30, 32, 47, 50, and 84) compared with only one major mutation site (48) in L12-favorable sites.

For each of the LPV- or L11/L12-favorable sites, the total energy difference was decomposed into the nonpolar energy difference (van der Waals and surface area) and polar energy

DOI: 10.1021/acs.jcim.5b00056
J. Chem. Inf. Model. 2015, 55, 1400−1412

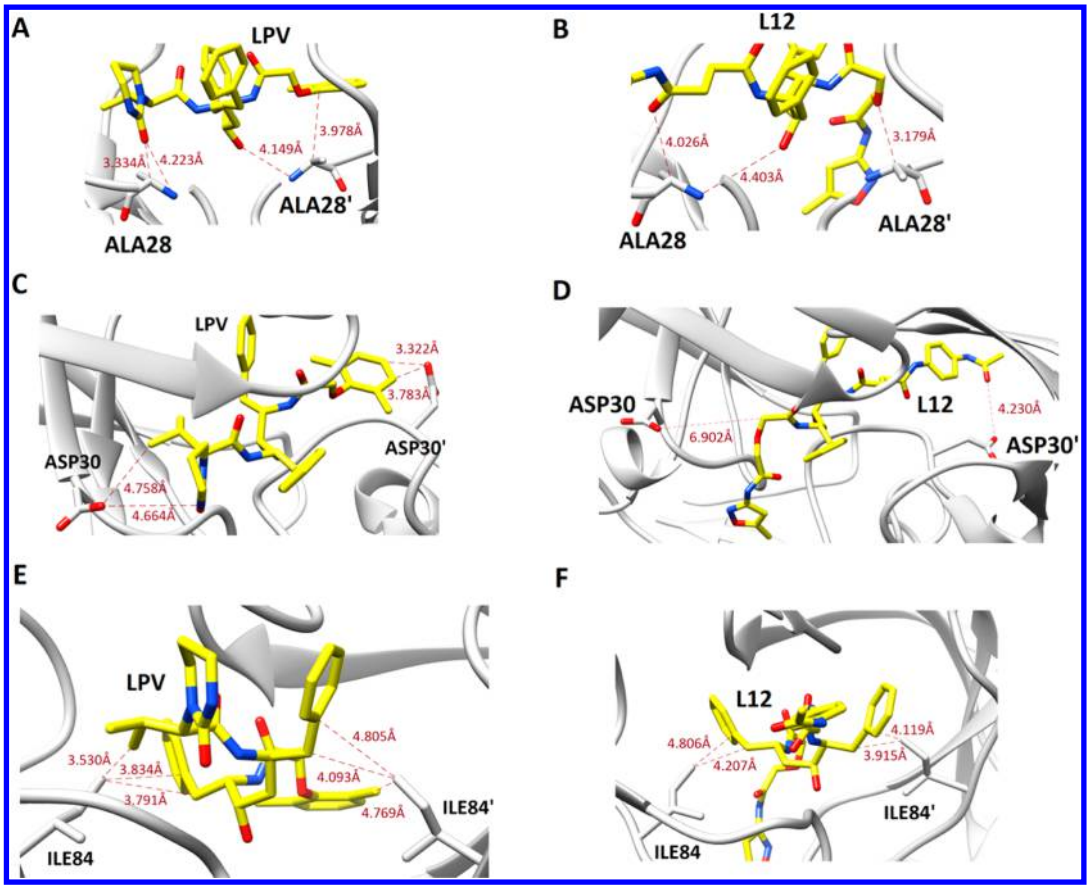


Figure 8. Interactions between LPV-favorable protease sites and inhibitor ligands LPV and L12. Wild-type HIV-1 PR is used as a representative structure. (A) LPV interactions with site 28. Representative interactions are shown with distances between related heavy atoms; the oxygen (red) and nitrogen (blue) atoms of site 28 are backbone atoms of alanine. (B) L12 interactions with site 28. (C) LPV interactions with site 30. (D) L12 interactions with site 30. (E) LPV interactions with site 84. (F) L12 interactions with site 84.

difference (electrostatics and GB) (Table S3 in the Supporting Information). Three different classes can be defined from the nonpolar and polar contributions to the total energy difference: nonpolar-dominant sites, polar-dominant sites, and dual sites. Most of the LPV- and L11/L12-favorable sites are dominated by nonpolar interactions. Among the nine sites for L11 (five L11-favorable sites and four LPV-favorable sites), six are nonpolar-dominant sites, one is a polar-dominant site, and two are dual sites (Figure 7A). Among the 13 sites for L12 (seven L12-favorable sites and six LPV-favorable sites), eight are nonpolar-dominant sites, two are polar-dominant sites, and three are dual sites (Figure 7B). It is worth noting that nonpolar interactions almost completely determine the LPV-favorable sites in both the L11 and L12 cases, while none of the LPV-favorable sites are polar-dominant, which indicates that the hierarchical screening works well to weaken the nonpolar interactions between the ligand and drug resistance sites as designed.

We examined the structural details of the top three LPV-favorable and L12-favorable sites (Figures 8 and 9, respectively) using wild-type HIV-1 PR as a representative structure. Of the three LPV-favorable sites (28, 30, and 84), site 28 is a dual site with both favorable nonpolar and polar interactions for LPV compared with L12. LPV has a larger number of contacts with site 28 than does L12. The R2 group of LPV contributes to the nonpolar component of the energy via a hydrophobic interaction between the benzene ring and the side chain of residue 28 (3.978 Å between the benzene carbon atom of LPV and the C^{α} atom of wild-type alanine 28; Figure 8A). The polar component of the energy comes from the close contact of the oxygen atoms of LPV with the backbone nitrogen atom of site 28 (Figure 8A,B). Sites 30 and 84 are both nonpolar-dominant sites. LPV forms many closer contacts with D30 than does L12, for both residues in protease chains A and B (Figure 8C,D). For site 84, LPV forms more close contacts with the residues than L12 does (Figure 8E,F).

For the top three L12-favorable sites (87, 8, and 48), site 87 is a dual site showing obvious closer contact with L12 than LPV, contributing both nonpolar and polar components to the ligand−protein binding energy (Figure 9A,B). From Figure 9B we can see that the oxygen and nitrogen atoms of the functionalized isoxazole ring of L12 are in close proximity (3.509 and 3.950 Å, respectively) to the guandinium side group of the R87 and may disrupt the stabilizing effects of R87. The sequence conservation of site 87 is closely related to its biological function, as residue R87 is known to contribute to dimer stability[67] via the formation of an intramonomer hydrogen bond between the side-chain guanidinium group of residue R87 and one of the side-chain D29 carboxylate oxygen atoms.[68] The R87 guanidinium also forms intermonomer hydrogen bonds to L5 and W6.[69] Therefore, the strong energy contribution from site 87 will benefit the ligand's drug resistance profile. Site 8 is a nonpolar-dominant site, which shows closer contact to L12 than to LPV (Figure 9C,D).

Site 48 is the only unfavorable ligand-preferred site from our comparative analysis of L11/L12 and LPV. Thus, a reduction of

DOI: 10.1021/acs.jcim.5b00056
J. Chem. Inf. Model. 2015, 55, 1400−1412

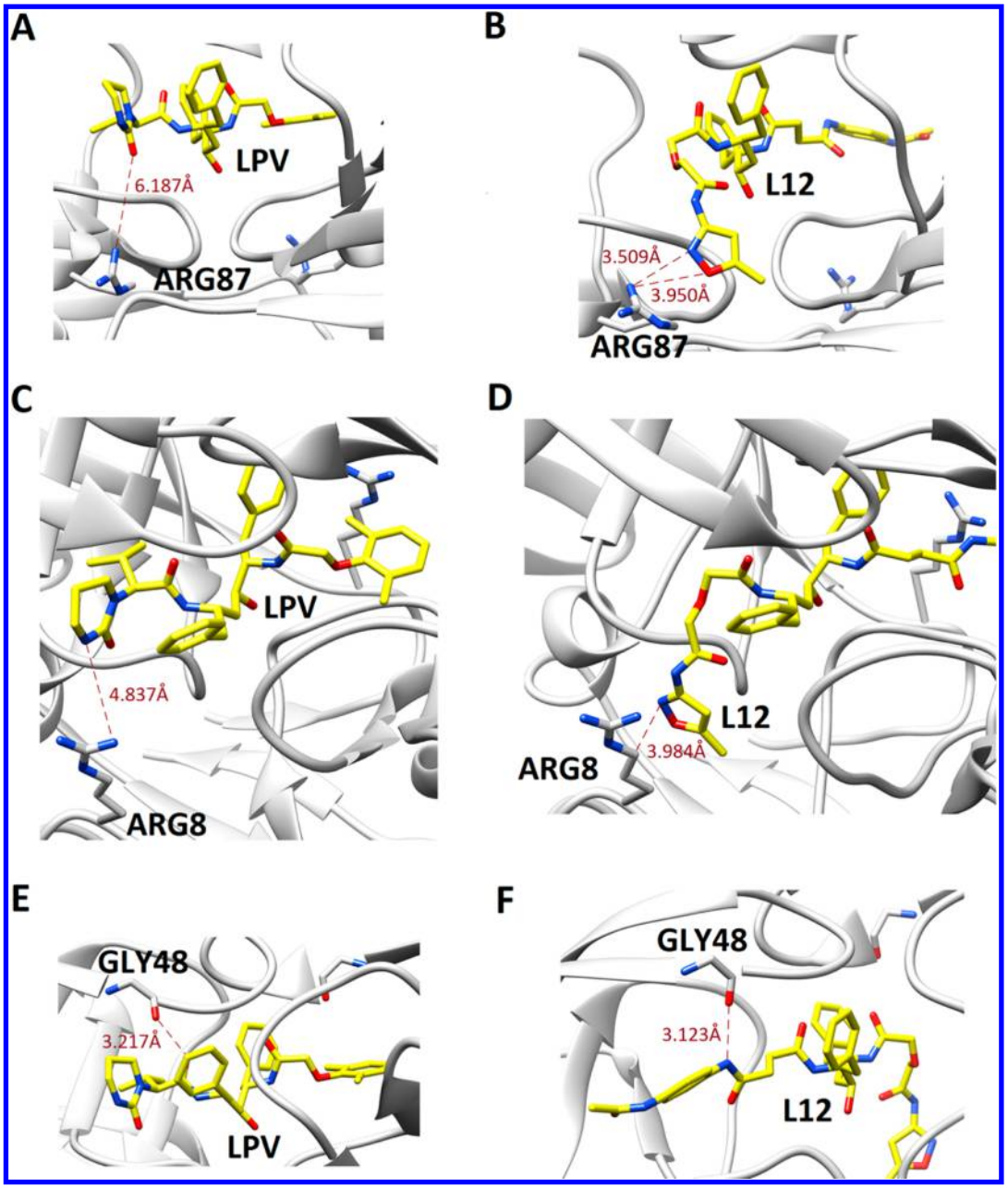


Figure 9. Interactions between L12-favorable protease sites and inhibitor ligands LPV and L12. Wild-type HIV-1 PR is used as a representative structure. (A) LPV interactions with site 87. (B) L12 interactions with site 87. (C) LPV interactions with site 8. (D) L12 interactions with site 8. (E) LPV interactions with site 48. (F) L12 interactions with site 48.

the ligand interaction with site 48 could form the basis for further structural refinement of these ligands. Site 48 is a polar-dominant site with similar contact residues for both LPV and L12. L12 forms a favorable polar interaction between the backbone nitrogen atom of L12 and the backbone oxygen atom of site 48 (glycine has no side chains). This compares to the contact between the benzene carbon atoms and the backbone oxygen atom in binding of LPV (Figure 9E,F).

Of further note, L12 has highly specific interactions with site 29, whose mutations inactivate the function of HIV-1 PR.[70] This is encouraging since L12 forms even more favorable interactions with this highly conserved site than does LPV. Furthermore, site 47 appears as an LPV-favorable site. The I47A major mutation leads to the loss of van der Waals interactions in the S2/S2′ subsites, caused by the loss of three side-chain methyl groups and by structural changes in the A47 main chain that lead to structural changes in the flap antiparallel β-strand.[71] L12 has a reduced binding energy contribution of -2 kcal mol^{-1} compared with LPV at this site.

Finally, our previous work[8] using a combination of conservation analysis and free energy calculations identified five residues (including most importantly D29) that may enhance the ability of a PI to combat drug resistance. Our current results show that L12 has improved binding compared with LPV for three of these residue sites, namely, L23, D29, and R87.

We conclude that the improved drug resistance ratios of L11 and L12 come from the variations in residue contributions to the binding energy, which more precisely show the shift of energy contributions from major mutation sites to conserved

DOI: 10.1021/acs.jcim.5b00056
J. Chem. Inf. Model. 2015, 55, 1400−1412

sites and sites not related to drug resistance. For both L11 and L12, nonpolar interactions play an important role in the shift of the energy contributions, as originally designed in the pipeline. No positional enrichment (e.g., core region or flap region) of significantly favorable sites is observed.

CONCLUSIONS

In this study, we provide a proof of concept for HIV-1 drug design that selects ligands with dominant binding contributions from conserved protease residues. In contrast to traditional HIV-1 drug design strategies, which target existing drug-resistant mutants, our strategy uses an alternative method that targets conserved protease residues. In this way, we can design inhibitors that are effective against incoming HIV-1 mutants. IC_{50} measurements successfully identified seven ligands with PR inhibitory activity, thus illustrating the preliminary success of this strategy. Two ligands (L11 and L12) showed a better predicted drug resistance profile than that of LPV. Our analysis showed that a shift of binding energy contributions from major mutation sites to more conserved sites is crucial to obtain better drug resistance profiles for protease inhibitors.

All of the selected ligands identified in this study still have the opportunity to undergo further optimization to boost their inhibitory activities to the magnitude of current FDA-approved drugs. However, these potential technical improvements do not trivialize the success of identifying ligands with inhibitory activity. Since functional residues of HIV-1 PR are not always mutable, they could become excellent targets for the design of the next generation of HIV-1 PR inhibitors.

ASSOCIATED CONTENT

Supporting Information

ROC curve for drug resistance prediction (Figure S1), MD simulation reranking (Figure S2), ROC curve for prediction performance (Figure S3), binding energy contributions and resistance ratios for seven FDA-approved HIV-1 PR inhibitors (Table S1), ranks of the 18 selected ligands (Table S2, XLS), decomposition of total energy differences (Table S3, XLS), and a zip file containing docking poses of 18 ligands to wild-type HIV in PDB format. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00056.

AUTHOR INFORMATION

Corresponding Authors

*Tel: 858-822-4240. Fax: 858-822-4236. E-mail: wei-wang@ucsd.edu.

*E-mail: tingjunhou@zju.edu.cn.

Author Contributions

§N.L. and R.I.A. contributed equally.

Author Contributions

Conceived and designed experiments: N.L., T.H., R.I.A., and W.W. Performed calculations: N.L., T.H., and B.D. Data analysis: N.L. and T.H. Manuscript preparation: N.L., R.I.A., T.H., and W.W.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Some of the molecular dynamics simulations were conducted in the Center for Theoretical Biological Physics (CTBP) computing cluster at the University of California, San Diego. This work was partially supported by the National Institutes of Health (GM085188 to W.W.).

REFERENCES

(1) Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A.; Scolnick, E. M.; Sigal, I. S. Active Human Immunodeficiency Virus Protease Is Required for Viral Infectivity. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 4686−4690.

(2) De Clercq, E. The Design of Drugs for HIV and HCV. *Nat. Rev. Drug Discovery* **2007**, *6*, 1001−1018.

(3) Van Laethem, K.; De Luca, A.; Antinori, A.; Cingolani, A.; Perna, C. F.; Vandamme, A. M. A Genotypic Drug Resistance Interpretation Algorithm That Significantly Predicts Therapy Response in HIV-1-Infected Patients. *Antiviral Ther.* **2002**, *7*, 123−129.

(4) Rhee, S. Y.; Gonzales, M. J.; Kantor, R.; Betts, B. J.; Ravela, J.; Shafer, R. W. Human Immunodeficiency Virus Reverse Transcriptase and Protease Sequence Database. *Nucleic Acids Res.* **2003**, *31*, 298−303.

(5) Ding, B.; Li, N.; Wang, W. Characterizing Binding of Small Molecules. II. Evaluating the Potency of Small Molecules To Combat Resistance Based on Docking Structures. *J. Chem. Inf. Model.* **2013**, *53*, 1213−1235.

(6) Arts, E. J.; Hazuda, D. J. HIV-1 Antiretroviral Drug Therapy. *Cold Spring Harbor Perspect. Med.* **2012**, *2*, No. a007161.

(7) Wensing, A. M.; Van Maarseveen, N. M.; Nijhuis, M. Fifteen Years of HIV Protease Inhibitors: Raising the Barrier to Resistance. *Antiviral Res.* **2010**, *85*, 59−74.

(8) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. Biomolecular Simulations: Recent Developments in Force Fields, Simulations of Enzyme Catalysis, Protein−Ligand, Protein−Protein, and Protein−Nucleic Acid Noncovalent Interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211−243.

(9) Coleman, R. G.; Carchia, M.; Sterling, T.; Irwin, J. J.; Shoichet, B. K. Ligand Pose and Orientational Sampling in Molecular Docking. *PloS One* **2013**, *8*, No. e75992.

(10) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455−516.

(11) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A Semiempirical Free Energy Force Field with Charge-Based Desolvation. *J. Comput. Chem.* **2007**, *28*, 1145−1197.

(12) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(13) Rarey, M.; Kramer, B.; Lengauer, T. The Particle Concept: Placing Discrete Water Molecules during Protein−Ligand Docking Predictions. *Proteins* **1999**, *34*, 17−28.

(14) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(15) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750−1759.

(16) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177−6273.

(17) Jones, G.; Willett, P.; Glen, R. C. Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *J. Mol. Biol.* **1995**, *245*, 43−53.

(18) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727−775.

DOI: 10.1021/acs.jcim.5b00056
J. Chem. Inf. Model. 2015, 55, 1400−1412

(19) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935−984.
(20) Yanuar, A.; Suhartanto, H.; Munim, A.; Anugraha, B. H.; Syahdi, R. R. Virtual Screening of Indonesian Herbal Database as HIV-1 Protease Inhibitor. *Bioinformation* **2014**, *10*, 52−57.
(21) Yedidi, R. S.; Liu, Z.; Kovari, I. A.; Woster, P. M.; Kovari, L. C. P1 and P1′ *para*-Fluoro Phenyl Groups Show Enhanced Binding and Favorable Predicted Pharmacological Properties: Structure-Based Virtual Screening of Extended Lopinavir Analogs against Multi-Drug Resistant HIV-1 Protease. *J. Mol. Graphics Modell.* **2014**, *47*, 18−24.
(22) Shityakov, S.; Dandekar, T. Lead Expansion and Virtual Screening of Indinavir Derivate HIV-1 Protease Inhibitors Using Pharmacophoric - Shape Similarity Scoring Function. *Bioinformation* **2010**, *4*, 295−299.
(23) Kunze, J.; Todoroff, N.; Schneider, P.; Rodrigues, T.; Geppert, T.; Reisen, F.; Schreuder, H.; Saas, J.; Hessler, G.; Baringhaus, K. H.; Schneider, G. Targeting Dynamic Pockets of HIV-1 Protease by Structure-Based Computational Screening for Allosteric Inhibitors. *J. Chem. Inf. Model.* **2014**, *54*, 987−1078.
(24) Chellappan, S.; Reddy, G. S. K. K.; Ali, A.; Nalam, M. N. L.; Anjum, S. G.; Cao, H.; Kairys, V.; Fernandes, M. X.; Altman, M. D.; Tidor, B.; Rana, T. M.; Schiffer, C. A.; Gilson, M. K. Design of Mutation-Resistant HIV Protease Inhibitors with the Substrate Envelope Hypothesis. *Chem. Biol. Drug Des.* **2007**, *69*, 298−313.
(25) Pandit, D.; So, S. S.; Sun, H. Enhancing Specificity and Sensitivity of Pharmacophore-Based Virtual Screening by Incorporating Chemical and Shape Features—A Case Study of HIV Protease Inhibitors. *J. Chem. Inf. Model.* **2006**, *46*, 1236−1280.
(26) Altman, M. D.; Ali, A.; Reddy, G. S. K. K.; Nalam, M. N. L.; Anjum, S. G.; Cao, H.; Chellappan, S.; Kairys, V.; Fernandes, M. X.; Gilson, M. K.; Schiffer, C. A.; Rana, T. M.; Tidor, B. HIV-1 Protease Inhibitors from Inverse Design in the Substrate Envelope Exhibit Subnanomolar Binding to Drug-Resistant Variants. *J. Am. Chem. Soc.* **2008**, *130*, 6099−6113.
(27) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314−3321.
(28) Massova, I.; Kollman, P. A. Computational Alanine Scanning To Probe Protein−Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *J. Am. Chem. Soc.* **1999**, *121*, 8133−8143.
(29) Hou, T. J.; Wang, J. M.; Li, Y. Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69−82.
(30) Sun, H. Y.; Li, Y. Y.; Tian, S.; Xu, L.; Hou, T. J. Assessing the Performance of MM/PBSA and MM/GBSA Methods. 4. Accuracies of MM/PBSA and MM/GBSA Methodologies Evaluated by Various Simulation Protocols Using PDBbind Data Set. *Phys. Chem. Chem. Phys.* **2014**, *16*, 16719−16729.
(31) Sun, H. Y.; Li, Y. Y.; Shen, M. Y.; Tian, S.; Xu, L.; Pan, P. C.; Guan, Y.; Hou, T. J. Assessing the Performance of MM/PBSA and MM/GBSA Methods. 5. Improved Docking Performance Using High Solute Dielectric Constant MM/GBSA and MM/PBSA Rescoring. *Phys. Chem. Chem. Phys.* **2014**, *16*, 22035−22045.
(32) Xu, L.; Sun, H. Y.; Li, Y. Y.; Wang, J. M.; Hou, T. J. Assessing the Performance of MM/PBSA and MM/GBSA Methods. 3. The Impact of Force Fields and Ligand Charge Models. *J. Phys. Chem. B* **2013**, *117*, 8408−8421.
(33) Hou, T. J.; McLaughlin, W. A.; Wang, W. Evaluating the Potency of HIV-1 Protease Drugs To Combat Resistance. *Proteins* **2008**, *71*, 1163−1174.
(34) Hou, T. J.; Zhang, W.; Wang, J.; Wang, W. Predicting Drug Resistance of the HIV-1 Protease Using Molecular Interaction Energy Components. *Proteins* **2009**, *74*, 837−846.
(35) Zhang, J.; Hou, T. J.; Wang, W.; Liu, J. S. Detecting and Understanding Combinatorial Mutation Patterns Responsible for HIV Drug Resistance. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 1321−1326.
(36) Hou, T. J.; Zhang, W.; Case, D. A.; Wang, W. Characterization of Domain−Peptide Interaction Interface: A Case Study on the Amphiphysin-1 SH3 Domain. *J. Mol. Biol.* **2008**, *376*, 1201−1214.
(37) Hou, T. J.; Xu, Z.; Zhang, W.; McLaughlin, W. A.; Case, D. A.; Xu, Y.; Wang, W. Characterization of Domain−Peptide Interaction Interface: A Generic Structure-Based Model To Decipher the Binding Specificity of SH3 Domains. *Mol. Cell Proteomics* **2009**, *8*, 639−649.
(38) Hou, T. J.; Li, N.; Li, Y. Y.; Wang, W. Characterization of Domain−Peptide Interaction Interface: Prediction of SH3 Domain-Mediated Protein−Protein Interaction Network in Yeast by Generic Structure-Based Models. *J. Proteome Res.* **2012**, *11*, 2982−2995.
(39) Li, N.; Hou, T. J.; Ding, B.; Wang, W. Characterization of PDZ Domain−Peptide Interaction Interface Based on Energetic Patterns. *Proteins* **2011**, *79*, 3208−3220.
(40) Hou, T. J.; Wang, J. M.; Li, Y. Y.; Wang, W. Assessing the Performance of the Molecular Mechanics/Poisson Boltzmann Surface Area and Molecular Mechanics/Generalized Born Surface Area Methods. II. The Accuracy of Ranking Poses Generated from Docking. *J. Comput. Chem.* **2011**, *32*, 866−877.
(41) Ding, B.; Wang, J.; Li, N.; Wang, W. Characterization of Small Molecule Binding. I. Accurate Identification of Strong Inhibitors in Virtual Screening. *J. Chem. Inf. Model.* **2013**, *53*, 114−122.
(42) Graves, A. P.; Shivakumar, D. M.; Boyce, S. E.; Jacobson, M. P.; Case, D. A.; Shoichet, B. K. Rescoring Docking Hit Lists for Model Cavity Sites: Predictions and Experimental Testing. *J. Mol. Biol.* **2008**, *377*, 914−934.
(43) Klamt, A.; Schuurmann, G. COSMO—A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and Its Gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799−805.
(44) Anisimov, V. M.; Cavasotto, C. N. Quantum Mechanical Binding Free Energy Calculation for Phosphopeptide Inhibitors of the Lck SH2 Domain. *J. Comput. Chem.* **2011**, *32*, 2254−2263.
(45) Ucisik, M. N.; Zheng, Z.; Faver, J. C.; Merz, K. M. Bringing Clarity to the Prediction of Protein−Ligand Binding Free Energies via "Blurring". *J. Chem. Theory Comput.* **2014**, *10*, 1314−1325.
(46) *SYBYL Molecular Simulation Package*; Tripos: St. Louis, MO, 2004.
(47) Irwin, J. J.; Shoichet, B. K. ZINC—A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45*, 177−182.
(48) Stoll, V.; Qin, W. Y.; Stewart, K. D.; Jakob, C.; Park, C.; Walter, K.; Simmer, R. L.; Helfrich, R.; Bussiere, D.; Kao, J.; Kempf, D.; Sham, H. L.; Norbeck, D. W. X-ray Crystallographic Structure of ABT-378 (Lopinavir) Bound to HIV-1 Protease. *Bioorg. Med. Chem.* **2002**, *10*, 2803−2806.
(49) Dixon, S. L.; Merz, K. M., Jr. Semiempirical Molecular Orbital Calculations with Linear System Size Scaling. *J. Chem. Phys.* **1996**, *104*, 6643−6649.
(50) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(51) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(52) Wang, W.; Lim, W. A.; Jakalian, A.; Wang, J.; Wang, J. M.; Luo, R.; Bayly, C. T.; Kollman, P. A. An Analysis of the Interactions between the Sem-5 SH3 Domain and Its Ligands Using Molecular Dynamics, Free Energy Calculations, and Sequence Analysis. *J. Am. Chem. Soc.* **2001**, *123*, 3986−3994.
(53) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. A Point-charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.

DOI: 10.1021/acs.jcim.5b00056
J. Chem. Inf. Model. 2015, 55, 1400−1412

(54) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(55) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.
(56) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins* **2004**, *55*, 383−394.
(57) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(58) Wang, J. M.; Hou, T. J.; Xu, X. J. Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Curr. Comput.-Aided Drug Des.* **2006**, *2*, 287−306.
(59) Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217−230.
(60) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(61) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An *N*·log(*N*) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(62) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(63) Gohlke, H.; Kiel, C.; Case, D. A. Insights into Protein−Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras−Raf and Ras−RaIGDS Complexes. *J. Mol. Biol.* **2003**, *330*, 891−913.
(64) Shafer, R. W.; Stevenson, D.; Chan, B. Human Immunodeficiency Virus Reverse Transcriptase and Protease Sequence Database. *Nucleic Acids Res.* **1999**, *27*, 348−352.
(65) Petropoulos, C. J.; Parkin, N. T.; Limoli, K. L.; Lie, Y. S.; Wrin, T.; Huang, W.; Tian, H.; Smith, D.; Winslow, G. A.; Capon, D. J.; Whitcomb, J. M. A Novel Phenotypic Drug Susceptibility Assay for Human Immunodeficiency Virus Type 1. *Antimicrob. Agents Chemother.* **2000**, *44*, 920−928.
(66) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L., Jr. Improved Prediction of Protein Side-Chain Conformations with SCWRL4. *Proteins* **2009**, *77*, 778−873.
(67) Ishima, R.; Ghirlando, R.; Tozser, J.; Gronenborn, A. M.; Torchia, D. A.; Louis, J. M. Folded Monomer of HIV-1 Protease. *J. Biol. Chem.* **2001**, *276*, 49110−49116.
(68) Louis, J. M.; Ishima, R.; Nesheiwat, I.; Pannell, L. K.; Lynch, S. M.; Torchia, D. A.; Gronenborn, A. M. Revisiting Monomeric HIV-1 Protease: Characterization and Redesign for Improved Properties. *J. Biol. Chem.* **2003**, *278*, 6085−6092.
(69) Ishima, R.; Gong, Q. G.; Tie, Y. F.; Weber, I. T.; Louis, J. M. Highly Conserved Glycine 86 and Arginine 87 Residues Contribute Differently to the Structure and Activity of the Mature HIV-1 Protease. *Proteins* **2010**, *78*, 1015−1025.
(70) Ishima, R.; Torchia, D. A.; Lynch, S. M.; Gronenborn, A. M.; Louis, J. M. Solution Structure of the Mature HIV-1 Protease Monomer: Insight into the Tertiary Fold and Stability of a Precursor. *J. Biol. Chem.* **2003**, *278*, 43311−43319.
(71) Saskova, K. G.; Kozisek, M.; Lepsik, M.; Brynda, J.; Rezacova, P.; Vaclavikova, J.; Kagan, R. M.; Machala, L.; Konvalinka, J. Enzymatic and Structural Analysis of the I47A Mutation Contributing to the Reduced Susceptibility to HIV Protease Inhibitor Lopinavir. *Protein Sci.* **2008**, *17*, 1555−1564.

DOI: 10.1021/acs.jcim.5b00056
J. Chem. Inf. Model. 2015, 55, 1400−1412