JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

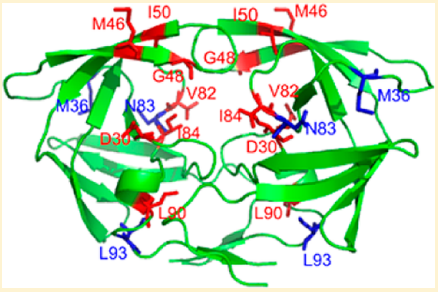
Characterizing Binding of Small Molecules. II. Evaluating the Potency of Small Molecules to Combat Resistance Based on Docking Structures

Bo Ding,[†] Nan Li,[†] and Wei Wang*[,†,‡]

[†]Department of Chemistry and Biochemistry, [‡]Department of Cellular and Molecular Medicine, UCSD, La Jolla, California 92093-0359, United States

Supporting Information

ABSTRACT: Drug resistance severely erodes the efficacy of therapeutic treatments for many diseases. Assessing the potency of a drug lead to combat resistance is no doubt critical for designing new drugs or new therapeutic combinations. Virtual screening is often the first step in drug discovery and a challenging problem is to accurately predict the resistant profile of an inhibitor based on the docking structures. Using a well studied system HIV-1 protease, we have illustrated the success of a computational method called MIEC-SVM on tackling this problem. We computed molecular interaction energy components (MIECs) between the ligand and the protease residues to characterize the docking poses, which were input to support vector machine (SVM) to distinguish resistant from nonresistant mutants. More importantly, the method is able to predict resistant profiles for new drugs based on the docking structures as indicated by its satisfactory performance in leave-one-drug-out and leave-drug/mutants-out tests. Therefore, the MIEC-SVM method can also facilitate designing effective therapeutic combinations by combining drugs with complementary resistant profiles.



INTRODUCTION

Drug resistance is one of the key roadblocks to successful treatment for many diseases,[1−4] especially when the drug is against targets with rapid replication and unstable genome, such as cancer cells[5] and viruses.[6−9] Resistant cancer cells or viral strains are selected in the presence of the drug to become dominant in the population, which severely impairs the therapeutic efficacy of a drug that is developed for years at a cost of millions of dollars.[10−12] Cocktail therapy, in which a patient is treated with a mixture of drugs, can often delay but not completely abolish resistance.[13−15] Obviously, it is invaluable for drug development if the potency of a drug lead to combat resistance can be evaluated at an early stage, which remains a great challenge as there exist diverse mechanisms to cause resistance.[1,16−19]

One common mechanism of drug resistance is that the selected mutations disrupt the drug binding but do not impair the function of the drug target. For example, the majority of resistance mutations of the HIV-1 protease significantly interfere with the binding of the drug but not that of the substrates so that viral polypeptides can still be cleaved into individual viral proteins by the protease.[20] HIV-1 protease is a primary target in AIDS therapy, and nine drugs against it have been approved by FDA.[10] Thousands of clinical samples have been accumulated on the HIV-1 protease, which makes it an ideal testing system for methods to predict the resistance profile of a drug at the early stage of drug development.

Many attempts have been made to predict resistant mutations that impede the drug binding. These methods are either based on sequence[21−25] or structure.[26−29] The sequence-based methods rely on the availability of the genotypes labeled to be resistant or nonresistant, from which they identify the mutation patterns discriminating the two groups. The limit of these methods is that they cannot predict whether a new mutant not existing in the available genotypes is resistant or not. In contrast, the structure-based methods make *ab initio* evaluation of the binding between the mutants and the drug based on the active or inactive form of the 3-D structures.[30,31] These methods normally do not need training data and can identify new resistant mutants, but they often require extensive knowledge of the binding mode and are too computationally expensive to evaluate the potential ligands' potency against a large number of mutants.

In order to combine the advantages of these two approaches, we previously reported a method to model and predict the susceptibility of mutants against a drug based on the energetic pattern of the protein−ligand interaction that is characterized by molecular interaction energy components (MIECs).[32,33] We showed that this method achieved superior performance on the FDA approved drugs against the HIV-1 protease. Because the crystal structures of these drugs complexed with the wild-type protease were already solved, we only needed to model the complex structures of the mutants from the crystal structures. In virtual screening or at the early stage of drug development, complex structures are not necessarily available; thus, the

Received: January 6, 2013

 © XXXX American Chemical Society dx.doi.org/10.1021/ci400011c | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

question arises as to how well our method would work if the template complex structure is obtained from modeling and may be noisy.

In the present study, we aim to address this question by mimicking the real case in virtual screening and we use the HIV-1 protease as the testing system. Since the crystal structure of the drug target is often available, we docked seven drugs to the wild-type HIV-1 protease and then optimized the complex structures using molecular dynamics. Mutant complex structures were then modeled from the optimized wild-type structures. To robustly characterize the binding patterns of the protein−ligand based on the docking structures, we applied the MIECs using MM/GBSA models that include van der Waals, electrostatics, and solvation energies estimated by General Born model and accessible surface areas.[34−36] The MIECs of resistant and nonresistant mutants were used to train a support vector machine (SVM) that captures the distinct patterns causing resistance. We showed that the MIEC-SVM model achieved excellent performance in leave-one-drug-out test, which is to predict the drug-resistant profile of a ligand that is not included in the training data set. We also analyzed the mechanism for drug resistance, which sheds light on designing resistance-evading drugs.

RESULTS

Predicting Resistant Mutants with SVM. We first evaluated the performance of the MIEC-SVM model on predicting drug susceptibility (Table 1) using all 99 protease

Table 1. Drug Susceptibility Data Set Collected from the Stanford HIV Drug Resistant Database[37]

	resistant samples	nonresistant samples	resistant ratio
Amprenavir (APV)	388	1939	16.7%
Atazanavir (ATV)	359	437	45.1%
Indinavir (IDV)	783	1,661	32.0%
Lopinavir (LPV)	620	991	38.4%
Nelfinavir (NFV)	1,064	1,400	43.2%
Ritonavir (RTV)	841	1,566	34.9%
Saquinavir (SQV)	653	1,792	26.7%
total	4,708	9,786	32.5%

residues and 5 MIEC terms (ΔE_{vdW}, ΔE_{elec}, ΔG_{GB}, ΔA_{pho}, and ΔA_{phi}). As described in Experimental Methods, we docked each of the drugs into the wildtype protease and generated the mutant complex structures by mutating the wild-type protease residues. Regardless whether the wild-type complex structure was optimized using molecular dynamics or not, the performance of the model was satisfactory in 500 10-fold cross validations, as indicated by high positive prediction value Q_+ (0.905/0.894 with/without MD optimization), negative prediction value Q_- (0.954/0.950), sensitivity (0.902/0.895), specificity (0.956/0.950), and MCC (0.859/0.844) (Table 2a). Such a performance is comparable to the model trained and tested on mutant complex structures generated from the crystal structure of the wild-type (MCC = 0.853), which illustrates the robustness of our method to successfully capture the energetic patterns that discriminate resistant and nonresistant mutations. Compared to the model in our previous study[63] that used MIECs of ΔE_{vdW}, ΔE_{elec}, and ΔG_{GB}, inclusion of ΔA_{pho} and ΔA_{phi} improved the performance, e.g. MCC improved from 0.795 to 0.853 (Table 2a).

Table 2. Comparison of the Universal MIEC-SVM Models with (a) Cross Validation and (b) "LODO" Tests

	a			
	dock-MD[a]	dock[b]	crystal[c]	crystal-no SASA[d]
Q_+	0.905	0.894	0.897	0.810
Q_-	0.954	0.950	0.955	0.960
sensitivity	0.902	0.895	0.903	0.923
specificity	0.956	0.950	0.951	0.896
MCC	0.859	0.844	0.853	0.795

	b		
	dock-MD[a]	dock[b]	crystal[c]
Q_+	0.738	0.418	0.762
Q_-	0.896	0.898	0.880
sensitivity	0.750	0.787	0.735
specificity	0.882	0.502	0.861
MCC	0.607	0.290	0.625
AUC	0.901	0.824	0.922

[a]The complex structures of mutants were generated from docking models with molecular dynamics optimization. [b]The complex structures of mutants were generated from docking models without molecular dynamics optimization. [c]Crystal structures. [d]The performance of the MIEC-SVM model used in ref 32 (without the SASA terms) and the model were trained on crystal structures. All the values are the average of 500 cross validation runs.

To mimic the scenario of predicting resistant profile for a new inhibitor at the early stage of drug development, we performed a more rigorous test than cross validation, which is dubbed as "leave one drug out" (LODO). Namely, we trained our model on six drugs and tested it in the left-out drug. We conducted the LODO test on each of the seven drugs with the mutant complex structures modeled from docking structures with MD optimization using all the 99 residues and the 5 MIECs. The MIEC-SVM models in LODO tests still achieved an average MCC of 0.607 and area under curve of ROC (AUC) of 0.901 (Table 2b), which indicates a great performance considering the difficulty of this test. The slight drop of the average MCC in LODO compared to cross validation is not unexpected because the testing drug was not included in the training data in LODO test.

We noticed that the models obtained from the MD-optimized and crystal template structures showed comparable performance (MCC = 0.607 and 0.625 respectively) but significantly out-performed that obtained from docking without MD optimization (MCC = 0.290) (Table 2b). This observation is not unexpected as the docking structures were not relaxed in rigid docking, which contributes additional errors to the model. This error is ligand-specific and can be reduced by relaxation, which is important in the LODO test.

Feature Selection Boosts the Performance. We next conducted feature selection to remove noisy MIECs and protease residues. Considering the large number of possible combinations of MIECs and residue positions, it is a great challenge to find the optimal configuration. We implemented a heuristic strategy that first searched for the optimal combination of MIECs and then the optimal position composition. There are five MIECs in the initial model and a total of 31 possible combinations. We evaluated the performance of these possible combinations of MIECs when including all the residues in the MIEC-SVM models on LODO tests. The top 10 combinations with the best average MCC of the 7

dx.doi.org/10.1021/ci400011c | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

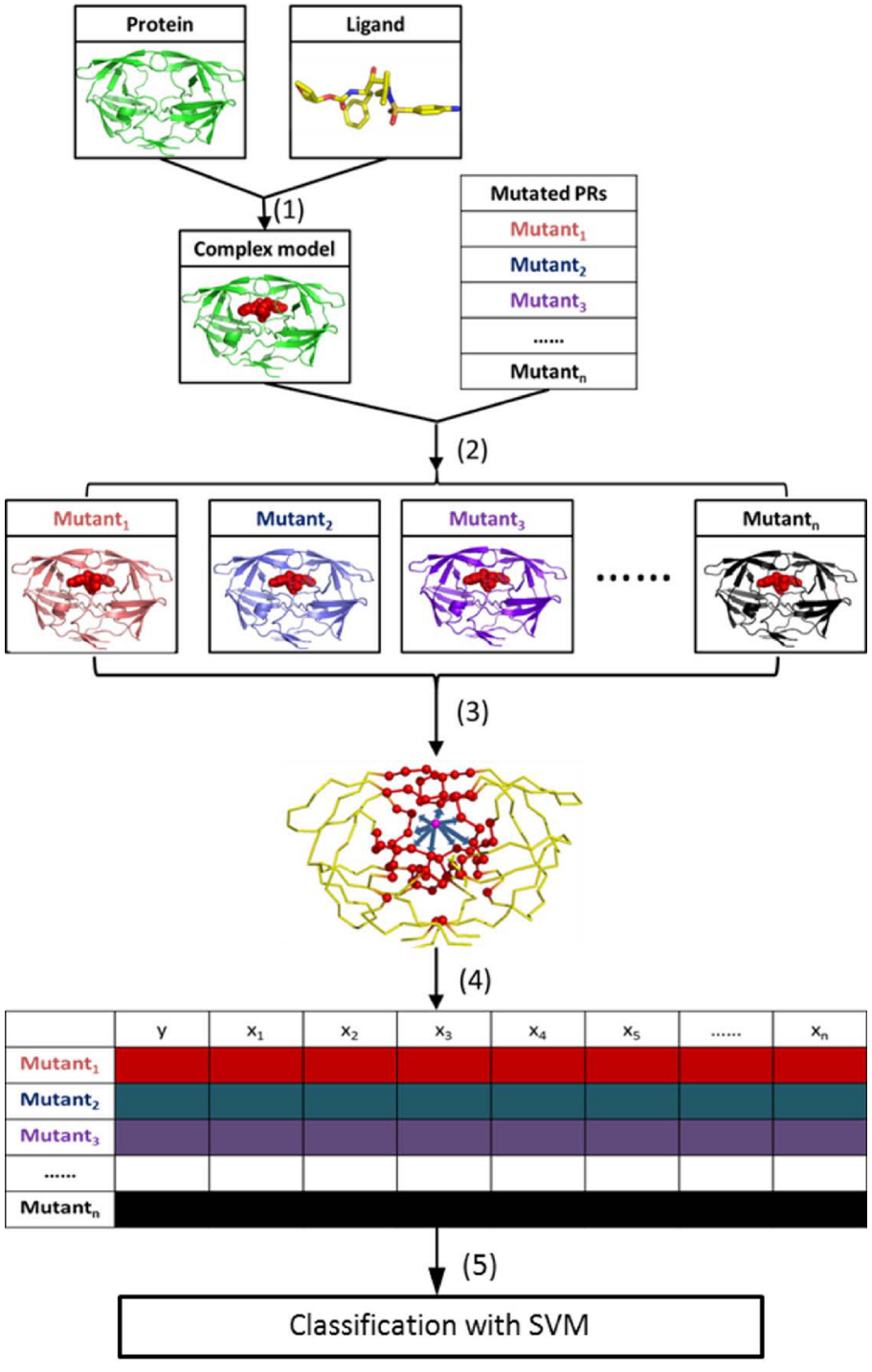


Figure 1. Flowchart of the MIEC-SVM method to predict resistant mutants based on docking. (1) Model the wildtype complex structures using docking and molecular dynamics. (2) Model the mutant complexes by mutating the side chains from the template structure followed by molecular mechanics minimization. (3) Calculate the MIECs between the ligand and the protease residues using MM/GBSA. The ligand is shown as a purple ball, and the residues are shown as red balls. (4) Generate the fully filled MIEC matrix, where y is the binary response variable that represents resistant or not and x_i is the ith MIEC. (5) Train the SVM to classify mutants as resistant or not.

LODO tests were used for selecting protease residues in the next step.

Considering the huge number of possible combinations of the 99 residues, we took a greedy search strategy to remove uninformative residues from the model. Starting with all the 99 residues, we left out one residue and its associated MIECs in the training (leave-one-residue-out, LORO), and assessed the trained model using LODO test. We kept the best performing model in the 99 possible models in the first round and repeated the LORO process until all residues were removed. As a result, for each of the 10 tested combinations of MIECs, we were able to find the most informative composition of residues (Figure 2). It is obvious that the model using four MIECs (ΔE_{vdW}, ΔE_{ele}, ΔA_{pho}, and ΔA_{phi}) and 47 protease residues achieved the highest average MCC of 0.761 in LODO tests.

dx.doi.org/10.1021/ci400011c | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

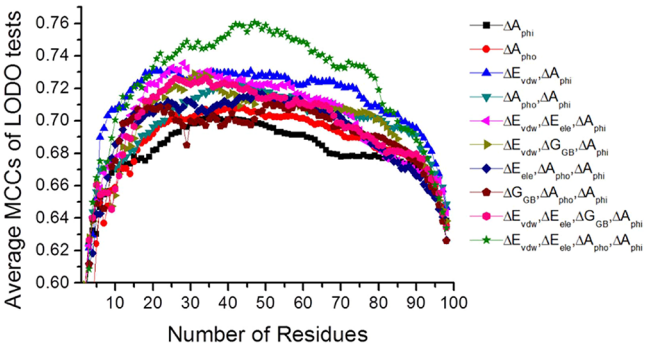


Figure 2. Feature selection of the most informative residues. The *Y* axis is the average MCC of the model in the seven LODO tests. The combination of ΔE_{vdW}, ΔE_{ele}, ΔA_{pho}, and ΔA_{phi} with 47 residues gave the best average MCC of 0.76.

The model using the optimal configuration of MIECs and protein positions showed significant improvement over the model without feature selection (the average MCCs increased from 0.607 to 0.761) (Table 3a). The model performed

Table 3a. The performance of MIEC-SVM in LODO tests using mutant complex structures generated from docking followed by MD optimization

Drug	Q_+	Q_-	Sensitivity	Specificity	MCC	AUC
APV	0.481	0.990	0.961	0.794	0.596	0.904
ATV	0.843	0.931	0.916	0.869	0.780	0.935
IDV	0.801	0.968	0.936	0.893	0.799	0.966
LPV	0.879	0.965	0.943	0.923	0.855	0.937
NFV	0.853	0.928	0.910	0.882	0.786	0.950
RTV	0.880	0.947	0.901	0.933	0.828	0.944
SQV	0.819	0.902	0.695	0.948	0.681	0.921
Average	0.794	0.947	0.895	0.892	0.761	0.937

significantly worse on APV (MCC=0.596) than on the other drugs whose MCCs were larger than 0.680. This may be due to the fact that APV has a more unbalanced data set as the ratio between resistant and nonresistant mutants is 1:5.04 compared to a ratio of 1:1.88 for the other drugs (Table 1). It also explains why APV showed a much smaller Q_+ (=0.481) than the other five drugs (>0.8), but comparable Q_-, sensitivity and specificity. Consistently, when down-sampling the nonresistant mutants to generate 500 random data sets with a resistant/nonresistant mutant ratio of 1:1.30−2.95, similar to that for the other drugs (1:2.02−2.39), the average MCC was improved to 0.722, Q+ to 0.713, Q- to 0.990, Sensitivity to 0.961, Specificity to 0.794, AUC to 0.915.

The model were further tested by "Leave-Drug/Mutants-Out" cross validations, in which randomly selected 90% of drug mutant data were used to train the model and the remaining 10% data were used for test. Table 3b shows that our method still achieved a satisfactory performance in the 500 runs of such tests, comparable to the LORO or LODO tests.

Previous analyses have clustered drug resistance mutations of the HIV protease against a specific drug into two classes: major mutations, which are often selected first in the presence of drugs and can significantly reduce the drug binding by themselves, and minor mutations, which do not have a substantial effect by themselves but often co-occur with major mutations to help replication of the mutated virus. For the 7 drugs, we chose to cluster the resistant positions into the

Table 3b. The performance of MIEC-SVM in Leave-drug/mutants-out tests using mutant complex structures generated from docking followed by MD optimization

Drug	Q_+	Q_-	Sensitivity	Specificity	MCC	AUC
APV	0.479	0.988	0.953	0.792	0.589	0.892
ATV	0.837	0.923	0.906	0.865	0.765	0.928
IDV	0.807	0.961	0.920	0.899	0.793	0.965
LPV	0.850	0.966	0.945	0.900	0.830	0.914
NFV	0.857	0.917	0.892	0.888	0.776	0.934
RTV	0.884	0.939	0.885	0.939	0.824	0.925
SQV	0.793	0.898	0.685	0.939	0.656	0.911
Average	0.787	0.941	0.884	0.889	0.748	0.924

following four groups: major position, minor position, conserved mutation and other position. Major positions, at least one major mutation against one drug can be found on these positions, such as 30, 90; minor positions, no major mutations can be found on these positions, but at least one minor mutation against one drug can be found on these positions, such as 71, 73; conserved positions, any position with less than 0.1% mutations documented in the database, such as 1,2 ; other position, which do not belong to any of the first three groups. The list of four clusters can be found in Table 4. In the previous studies,[37,64−66] there are 15 major positions, 23 minor positions, 25 conserved positions and 36 other positions. The 47 informative positions selected by feature selection include 12 major positions, 11 minor positions, 7 conserved positions and 17 other positions (Table 4). Indeed, we found that the 10 most informative positions among the 47 with MIECs of ΔE_{vdw},ΔE_{ele},ΔA_{pho} andΔA_{phi} could achieve an average MCC of 0.7 in the LODO tests. These ten positions consist of 7 major (D30, M46, G48, I54, V82, I84 and L90), and 3 minor positions (M36, N83, and I93). The feature selection results coincided with the previous analysis[64,66,67] on the HIV-1-infected patients as the major positions are well enriched in both 47 selected positions (p-value=0.0005 as calculated by hypergeometric distribution) and the 10 most informative positions (p-value=4.0×10^{-5}).

Analysis of resistant profiles for the drugs. Equipped by the capability of evaluating whether a mutant is resistant to a drug based on docking structures, we are able to compare the resistant profile of a new inhibitor with those of the existing ones, which would facilitate design of effective combinations targeting the protease. For comparing resistant profiles of drugs, we need to obtain the resistant data for each drug using the same set of mutations, which is largely unavailable in the public databases. For example, Table 1 lists experimental measurements of 14,494 mutant-drug pairs between 2,382 mutants and 7 drugs while 2,180 pairs out of the total 16,674 possible mutant-drug pairs (2,382*7) had no experimental data. To design an effective combination, the resistant profiles of component drugs should be complementary to each other. Therefore, we predicted the resistance for the remaining 2,180 mutant-drug pairs using each of the 500 cross-validation models and then these individual predictions were integrated using Bayesian statistics (see Material and Methods). This way, the resistant profiles of all the drugs are comparable. In addition, as there was no susceptibility data available for DRV (darunavir),[68] we also predicted the resistant profile of DRV against the 2,382 mutants. For this purpose, we trained the model with all experiment data to predict the resistant profile of DRV (Figure 4).

dx.doi.org/10.1021/ci400011c | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 4. The best performance group related to the observations in the patients

Observations in the patients	Positions selected by feature selection	Positions removed by feature selection
Major positions	**D30**,V32,L33,**M46**, **G48**,I50,**I54**,T74,L76,**V82**,**I84**,**L90**[a]	I47,Q58,N88
Minor positions	L10,V11,I13,L24,**M36**, K43,L63,A71,G73,**N83**,**I93**[a]	G16,K20,E34,E35,F53,D60,I62,I64,H69,V77,G86,L89
Conserved positions	L5,D25,W42,Y59, G78,P81,N98,F99	P1,Q2,W6,G17,E21,T26,G27,D29,T31,G40,P44,G49,G51,V56,T80,R87,G94,L97
Other positions	T4,K14,I15,Q18,L19,A22, A28,L38,R41,K55, R57,C67,G68,I72,T91, Q92	I3,Q7,R8,P9,T12,L23,S37,P39,K45,G52, Q61,E65,I66,K70,V75,P79,I85,C95,T96

[a]The positions in bold are the most informative 10 positions.

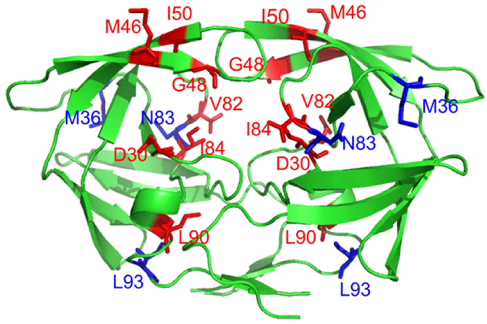


Figure 3. Distribution of the key residues selected by the feature selection. The red and blue residues belong to the major and minor resistant mutations.[63,65]

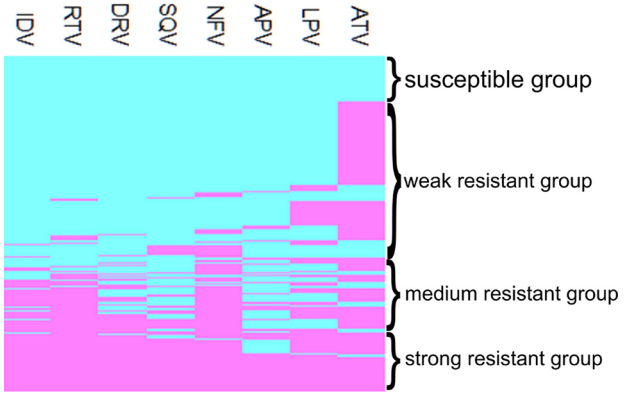


Figure 4. The mutant resistant profile for drugs. Each row is for a mutant. Cyan and pink represent that mutant is susceptible and resistant to the drug, respectively.

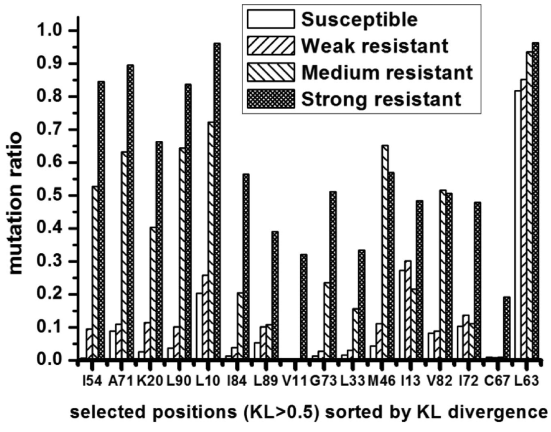


Figure 5. The mutation ratio of the selected positions in the four mutation groups.

The resistant profiles of the 8 drugs against the 2,382 protease mutants are clustered using hierarchical clustering method (Figure 4). For individual drugs, APV and ATV have the least (19.8%) and most number of resistant mutants (70.5%), respectively. It is obvious that ATV has a very distinct resistant profile from the other drugs, which is not totally unexpected as ATV has the strongest binding affinity among all the drugs. Obviously, no two drugs have complementary resistant profiles, which is not surprising as the cocktail regimens always include nonprotease drugs. On the other hand, this observation highlights the urgent need to develop drugs with resistant patterns complementary to the existing drugs to abolish drug resistance in AIDS therapy.

To identify the resistant mutations that are frequently observed against the existing drugs, we divided the 2,382 mutations into four groups based on the number of drugs to which they are resistant: susceptible (not resistant to any drug, 319 mutants); weak resistant (resistant to 1−2 drugs, 1,114 mutants); medium resistant (resistant to 3−6 drugs, 528 mutants) and strong resistant (resistant to 7−8 drugs, 421 mutants). Next, we compared the mutations of each position in the four groups (Figure S1 in Supporting Information). We calculated the probability of each amino acid occurring on a position in any of the four groups as $P_i = (N_i+Pseudo_i)/(\sum_i^{aa}(N_i+Pseudo_i))$, where P_i is the probability of the amino acid type i, N_i is the number of the amino acid i found at the position, $Pseudo_i$ is the pseudo count for amino acid i and set to 0.1 to avoid zero count, and aa represents all 20 amino acids. The occurring probability of all 20 amino acid at a position forms a occurring vector. Positions that show significantly different occurring vectors in the strong resistant group from the susceptible group are important positions for drug resistance and need attention in designing new drugs. To identify these positions, we computed the average Kullback−Leibler (KL) divergence[69] between the occurring vectors of the strong resistant and susceptible groups.

The 16 positions with larger than 0.5 KL distance are shown in Figure 5. The position with the largest KL distance is a well-known resistant position I54 that has no mutation in the susceptible group and a significant portion (84.6%) mutated to V/L in the strong resistant group. These 16 positions also include several other known major resistant positions, L90, I84, M46 and V82. Interestingly, mutations at I54, L90 and M46 have been shown to be dependent[33] and they are all included in this group. Surprisingly, positions A71, K20 and L10 also showed significantly different mutations in the strong resistant group from the susceptible group (Figure S1). Based on this analysis, new drugs with reduced resistance from the above positions would be well complementary to the existing drugs in cocktail therapy.

dx.doi.org/10.1021/ci400011c | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

DISCUSSION AND CONCLUSION

Here we present an approach to evaluating a ligand's potency to combat resistance caused by mutations disrupting drug binding. Different from other approaches that rely on ligand-specific descriptors, our method focuses on characterizing the binding site of the target protein and the interaction between protein residues and the entire ligand, which extends the generality of the model, as shown by the satisfactory performance of our model of LODO tests and "Leave-Drug/Mutants-Out" tests. A significant advancement over our previous studies[32] is that our models showed superior performance based on complex structures generated from docking rather than crystal structures, which makes it possible to apply the MIEC-based methods to screen for potent drug leads at the early stage of drug development.

We also demonstrated the capability of our method on comparing resistant profiles against the same set of mutants. Although the current cocktail treatments often target multiple viral proteins such as protease and reverse transcriptase, developing therapeutics composing of drugs with complementary resistant profiles for one target can serve as an alternative and our method can help design such therapeutic regimes. In addition, analysis of the resistant mutants against all the HIV-1 protease drugs provides a way to systematically identify the positions that are critical for drug resistance. Our analyses shed light on rational design of new drugs that should minimize resistance caused by mutations on the major resistant positions of I54, L90, I84, M46 and V82 as well as positions of A71, K20 and L10 whose roles in drug resistance have not been well appreciated.

In this and previous studies,[70] using the HIV-1 protease as a testing system, we have illustrated a strategy to search for inhibitors with strong binding affinity against the drug target and evaluate its potency against possible resistant mutations based on docking structures. First, this strategy starts with an iterative procedure to search from new inhibitors in virtual screening using a MIEC-based model trained on a set of known inhibitors.[70] The number of known inhibitors can be as small as 50, which is not uncommon for important drug targets. The newly identified inhibitors are then added to the training set for the next iteration. Second, the newly identified inhibitors are then evaluated by its potency against possible resistant mutations using the strategy presented in the current study. Third, since the strong binding inhibitors such as ATV may suffer from strong resistance, an iteration going through the first two steps is necessary to simultaneously improve binding strength and potency to combat resistance of inhibitors.

EXPERIMENTAL SECTION

Data Set. The experimental data were collected from the Stanford HIV drug resistance database,[37] including the resistant data for seven drugs (ATV,[38] APV,[39] IDV,[40] LPV,[41] NFV,[42] SQV,[43] and RTV[44]). The binding affinities between the mutants and the ligand were measured by IC50. On the basis of the ratio of IC_{50} between the mutants and the wide type, the mutants were classified into resistant (positive samples), indicated by a ratio of IC_{50} larger than 10^{27}, and nonresistant (negative samples) (Table 1).

Modeling the Complex Structures of the Wild Type and Drugs. *A. Generate the Template Complex Structures from Docking.* The complex structures of the wild type and drugs were constructed by docking followed by molecular dynamics optimization. The unbounded crystal structure of the HIV-1 protease was obtained from the PDB database (pdb ID 1hpv[39]). The structure was prepared for docking by the Maestro's protein preparation wizard package in Glide.[45,46] The preparation procedure included removing waters, assigning bond orders, and adding hydrogen atoms. This process also optimized the protonation and tautomeric state of HIS, orientation of amide in ASN and GLN, hydroxyl in SER, THR, and TYR, and thiol groups in CYS sequentially by hydrogen position optimization and exhaustive sampling options. The final step was finished by restrained minimization of ligand/protein complexes using the OPLS_2001 force field.[17,47]

The ligands were docked to the protease by Glide using precomputed grids. The grids were generated by taking the centroid of the ligand in the pdb structure of 1HPV as the center. The grid box was centered at the centroid of the docking ligands and had a size of 10 Å × 10 Å × 10 Å. The grid box for all the atoms in the ligands was 30 Å × 30 Å × 30 Å. We set the van der Waals radius scaling to 0.8 and the partial charge cutoff to 0.25. We also turned on the options of ring conformation sampling and nitrogen inversion in docking. For each ligand, the pose with the best score was saved as the initial complex structure for the follow-up optimization.

B. Molecular Dynamics Optimization of the Template Complex Structures. (a) System Setup. All the missing hydrogen atoms were added to the template complex structures generated from docking with the *leap* module in the Amber10 software package.[48] We used monoprotonated state for the catalytic Asp25/25′, as determined previously.[29,49,50] The proton was placed at OD1 oxygen of Asp25, which is closer to the drugs in the crystal structures. For each ligand, the partial charges on the atoms were calculated using the RESP fitting technique based on the electrostatic potentials obtained from Gaussian 03[51] with Hartree-Fock(HF)/6-31G*. The parameters for the atoms in the ligands were generated using the *antechamber* module in Amber10.[48] We used Amber03[52] and general Amber force field[53] (gaff) in our simulations. The template complex structures generated from docking were first solvated in the truncated octahedron periodic box of TIP3P waters. The size of the box was chosen so that any solute atom was at least 10 Å away from the boundary. An appropriate number of Cl^- counterions were added to neutralize the system.

(b) Molecular Dynamics Simulation. We run 4000 steps of steepest descent minimization and 6000 steps of conjugate gradient minimization to relax the template complex structures. Then, the systems were heated to 300 K at an interval of 50 K over 5 ps with a 0.5 fs time step. Finally, the systems were equilibrated for 100 ps with a 2 fs time step at 300 K and the molecular dynamics simulations were continued until the convergence of the root-mean-square-deviation (RMSD) of C-alpha atoms of the protease and the ligand was achieved. The simulation length ranged from 3 to 5 ns. The last conformation of each template complex structure from the simulation was optimized using 10 000 steps of steepest descent minimization and saved for modeling the complex structures between mutants and the ligands.

Modeling the Complex Structures of the Mutants/Ligands. The mutants/ligands complex structures were modeled from the templates. The backbone of the protease and the ligand were fixed and the side-chains were mutated using Scrwl4.[54] Then the models were minimized using the

dx.doi.org/10.1021/ci400011c | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

sander program in Amber10 with 500 steps of steepest descent and 3500 steps of conjugate gradient minimization. The minimization would stop before the maximum number of steps was reached if the root-mean-square (rms) of the Cartesian elements of the energy gradient was less than 0.1 kcal/(mol Å).

Calculation of MIECs. In the MM/GBSA method,[55,56] the binding free energies between the mutants and the ligands were calculated as

$$\begin{aligned}\Delta G_{\text{binding}} &= \Delta G_{\text{complex}} - \Delta G_{\text{ligand}} - \Delta G_{\text{protein}} \\ &= \Delta E_{\text{MM}} + \Delta G_{\text{GB}} + \Delta G_{\text{nonpolar}} - T\Delta S \end{aligned} \quad (1)$$

Where ΔE_{MM} is the binding energy between the protein and the ligand in vacuum and it is the sum of van der Waals energy (ΔE_{vdW}) and electrostatic energy (ΔE_{elec}). ΔG_{GB} is the polar part of the solvation energy obtained by the General Born model using the parameters from ref 57. $\Delta G_{\text{nonpolar}}$ is the nonpolar part of the solvation energy, which is estimated by the solvent accessible surface area (SASA) as $\Delta G_{\text{nonpolar}} = \sum_i \gamma_i \Delta A_i$, where γ_i is a parameter and ΔA_i is the loss of SASA upon binding for atom i. Considering the different contribution of hydrophobic and hydrophilic surface, we have γ_{pho} for all the hydrophobic atoms and γ_{phi} for all the hydrophilic atoms. Therefore, $\Delta G_{\text{nonpolar}} = \gamma_{\text{pho}}\Delta A_{\text{pho}} + \gamma_{\text{phi}}\Delta A_{\text{phi}}$, while ΔA_{pho} is the loss of the hydrophobic surface area and ΔA_{phi} is the loss of the hydrophilic surface area. $-T\Delta S$, the change of conformational entropy, was not included in this study due to high computational burden for a large number of mutants. The binding energy can be decomposed to residue-based components as[58,59]

$$\begin{aligned}\Delta G_{\text{binding}} &= \Delta G_{\text{complex}} - \Delta G_{\text{ligand}} - \Delta G_{\text{protein}} \\ &= \Delta E_{\text{MM}} + \Delta G_{\text{GB}} + \Delta G_{\text{nonpolar}} - T\Delta S \\ &= \Delta E_{\text{vdW}} + \Delta E_{\text{elec}} + \Delta G_{\text{GB}} + \gamma_{\text{pho}}\Delta A_{\text{pho}} \\ &\quad + \gamma_{\text{phi}}\Delta A_{\text{phi}} \\ &= \sum_i (\Delta E^i_{\text{vdW}} + \Delta E^i_{\text{elec}} + \Delta G^i_{\text{GB}} + \gamma_{\text{pho}}\Delta A^i_{\text{pho}} \\ &\quad + \gamma_{\text{phi}}\Delta A^i_{\text{phi}}) + (\gamma_{\text{pho}}\Delta A^{\text{ligand}}_{\text{pho}} + \gamma_{\text{phi}}\Delta A^{\text{ligand}}_{\text{phi}}) \end{aligned} \quad (2)$$

where ΔE^i_{vdW}, ΔE^i_{elec}, and ΔG^i_{GB} are the van der Waals, electrostatic, and polar contribution to solvation energy between residue i and the ligand. ΔA^i_{pho} and ΔA^i_{phi} are the loss of hydrophobic and hydrophilic surface area for residue i upon binding. $\Delta A^{\text{ligand}}_{\text{pho}}$ and $\Delta A^{\text{ligand}}_{\text{phi}}$ are the loss of hydrophobic and hydrophilic surface area for the ligand upon binding. As a result, the MIECs include van der Waals interaction energy (ΔE_{vdW}), electrostatic (Coulombic) interaction energy (ΔE_{elec}), polar contribution to solvation energy (ΔG_{GB}), the loss of hydrophobic surface area (ΔA_{pho}), and the loss of hydrophilic surface area (ΔA_{phi}) upon binding. ΔE_{vdW}, ΔE_{elec}, and ΔG_{GB} were calculated using MM/GB protocol in AMBER10.[60] The three terms for individual residues were computed by *gleap* program[61] in Amber10 using default parameters, including the cutoff of 18.0 Å for both ΔE_{vdW} and ΔE_{elec}, 1 for the interior dielectric constant, and 80 for the exterior dielectric constant in the GB model and the charges from Amber03 force-field.[52] ΔA_{pho} and ΔA_{phi} were calculated using the method developed by Street and Mayo[36] with the vdw radii from Amber03[52] or gaff as the atom radii and a probe radii of 1.4 Å.

As the judgment of drug resistance depends on the ratios of IC_{50} between the mutant and the wild type, we calculated the difference of MIECs between each protease mutant and the wild type interacting with the inhibitor as the features input to the support vector machine (SVM) (Figure 1). In total, there are 497 items, i.e. five energy items for each of the 99 protease residues and 2 energy items, ΔA_{pho}, ΔA_{phi} for the ligand.

Training Models with Support Vector Machine. We trained support vector machine (SVM) to identify drug-resistance mutations. The LIBSVM[62] program was employed in the training and test of the SVM models. The models were validated with 10-fold cross-validations and a "leave-one-drug-out" test (LODO test). Cross-validations tested the models' prediction ability on the new mutants against the drugs that were already included in the training set. The generality of the model on predicting resistant mutations against new inhibitors was assessed by the LODO test, in which the model was trained on all but one drug and tested on the left-out drug. For both tests, TP (true positive), FP (false positive), TN (true negative), and FN (false negative) were counted, and the performances of the models were evaluated using the standard criteria: prediction accuracy of positive samples $Q_+ = \text{TP}/(\text{TP} + \text{FP})$; prediction accuracy of negative samples $Q_- = \text{TN}/(\text{TN} + \text{FN})$; sensitivity $\text{SE} = \text{TP}/(\text{TP} + \text{FN})$; specificity $\text{SP} = \text{TN}/(\text{TN} + \text{FP})$; and Matthews correlation coefficient $C = (\text{TN} \times \text{TP} - \text{FN} \times \text{FP})/[(\text{TP} + \text{FN})(\text{TP} + \text{FP})(\text{TN} + \text{FN})(\text{TN} + \text{FP})]^{1/2}$.

Feature Selection Based on the LODO Test. As the LODO test mimics the scenario in virtual screening better than cross-validation, we conducted feature selection on both MIEC terms and protein positions to optimize the model performance on LODO tests. To achieve the best overall performance of the model on the seven drugs, we used the average MCC of the seven LODO tests as the metric in feature selection. We first selected the MIEC terms by examining the performance of all the possible combinations. As we considered five types of MIECs (ΔE_{vdW}, ΔE_{elec}, ΔG_{GB}, ΔA_{pho}, and ΔA_{phi}), there are $2^5 - 1 = 31$ possible combinations. The top 10 best performed combinations were kept for further analysis. Next, for each of these combinations, we selected protease positions that are most informative in the model using a greedy algorithm because exhaustive assessment of all the possible combinations of the 99 protease residues is impossible. We used a "leave-one-residue-out" test (LORO test), which is to train and test the model without the MIECs associated with one residue at a time, to select informative residues. We started with the full model that considered all the residues and removed a residue if its left-out gave the best performance in the LORO test. The performance was evaluated by the average MCC of 7 LODO tests. This procedure iterated until all the residues were removed (Figure 2). The best performing combination of MIEC terms and protease positions were then selected as the final model.

Prediction with Different Models. We obtained seven MIEC-SVM models for the seven drugs. When predicting the resistant profiles of a new drug or a combination of existing drugs, we chose a model-average approach that combines the prediction power of all the MIEC-SVM models using Bayesian statistics with the assumption that the models are independent. If there are N models and their predictions of the probability for a protease mutant resistant to a combination of existing

dx.doi.org/10.1021/ci400011c | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

drugs or a new inhibitor are $\{B_1, B_2, ..., B_N\}$, then the probability of this mutant to be resistant or not (A = true or false) is

$$P(\text{Output} = A|\text{Pred}_1 = B_1, \text{Pred}_2 = B_2, ..., \text{Pred}_N = B_N)$$
$$= [P(\text{Pred}_1 = B_1, \text{Pred}_2 = B_2, ..., \text{Pred}_N = B_N |\text{Output} = A)*P(\text{Output} = A)]$$
$$/[P(\text{Pred}_1 = B_1, \text{Pred}_2 = B_2, ..., \text{Pred}_n = B_n)]$$
$$= \frac{\prod_{i=1}^{N} P(\text{Pred}_i = B_i|\text{Output} = A)*\text{P}(\text{Output} = A)}{\prod_{i=1}^{N} P(\text{Pred}_i = B_i)}$$

For each model, $P(\text{Pred} = B|\text{Output} = A)$ can be easily estimated from the performance on the test data set, following the formula below:

$$P(\text{Pred} = T|\text{Real} = T) = \text{TP}/(\text{TP} + \text{FN})(\text{Sensitivity})$$

$$P(\text{Pred} = F|\text{Real} = T) = 1 - \text{TP}/(\text{TP} + \text{FN})$$

$$P(\text{Pred} = F|\text{Real} = F) = \text{TN}/(\text{TN} + \text{FP})(\text{Specificity})$$

$$P(\text{Pred} = T|\text{Real} = F) = 1 - \text{TN}/(\text{TN} + \text{FP})$$

A mutant is resistant if $[P(\text{Output} = \text{True}|\text{Pred}_1 = B_1, \text{Pred}_2 = B_2, ..., \text{Pred}_N = B_N)]/[P(\text{Output} = \text{False}|\text{Pred}_1 = B_1, \text{Pred}_2 = B_2, ..., \text{Pred}_N = B_N)] > 1$; not resistant, otherwise.

■ ASSOCIATED CONTENT

Supporting Information

Mutation profiles of the four resistant groups. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

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

Author Contributions

B.D. and W.W. conceived and designed the study. B.D. performed the experiments and analyzed the results. N.L. contributed to developing the computational models. B.D. and W.W. wrote the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This project was partially supported by NIH (R01GM085188 to W.W.).

■ ABBREVIATIONS

MIEC, molecular interaction energy component; SVM, support vector machine; SASA, solvent accessible surface area; TP, true positive; FP, false positive; TN, true negative; FN, false negative; Q_+, accuracy of positive samples; Q_-, accuracy of negative samples; SP, specificity; SE, sensitivity; MCC, Matthews correlation coefficient; ROC, receiver operating characteristic; AUC, area under the curve

■ REFERENCES

(1) Gottesman, M. M. Mechanisms of cancer drug resistance. *Annu Rev Med* **2002**, *53*, 615−27.

(2) Calmy, A.; Pascual, F.; Ford, N. HIV drug resistance. *New Engl. J. Med.* **2004**, *350* (26), 2720−1.

(3) Yoon, K. S.; Kwon, D. H.; Strycharz, J. P.; Hollingsworth, C. S.; Lee, S. H.; Clark, J. M. Biochemical and molecular analysis of deltamethrin resistance in the common bed bug (Hemiptera: Cimicidae). *J. Med. Entomol.* **2008**, *45* (6), 1092−101.

(4) Li, X. Z.; Nikaido, H. Efflux-mediated drug resistance in bacteria: an update. *Drugs* **2009**, *69* (12), 1555−623.

(5) Shekhar, M. P. Drug resistance: challenges to effective therapy. *Curr. Cancer Drug Targets* **2011**, *11* (5), 613−23.

(6) Andreoni, M.; Perno, C. F. Positioning of HIV-protease inhibitors in clinical practice. *Eur. Rev. Med. Pharmacol. Sci.* **2012**, *16* (1), 10−8.

(7) Margeridon-Thermet, S.; Shafer, R. W. Comparison of the Mechanisms of Drug Resistance among HIV, Hepatitis B, and Hepatitis C. *Viruses* **2010**, *2* (12), 2696−739.

(8) Goldberg, D. E.; Siliciano, R. F.; Jacobs, W. R., Jr. Outwitting evolution: fighting drug-resistant TB, malaria, and HIV. *Cell* **2012**, *148* (6), 1271−83.

(9) Le, Q. M.; Kiso, M.; Someya, K.; Sakai, Y. T.; Nguyen, T. H.; Nguyen, K. H.; Pham, N. D.; Ngyen, H. H.; Yamada, S.; Muramoto, Y.; Horimoto, T.; Takada, A.; Goto, H.; Suzuki, T.; Suzuki, Y.; Kawaoka, Y. Avian flu: isolation of drug-resistant H5N1 virus. *Nature* **2005**, *437* (7062), 1108.

(10) Perrin, L.; Telenti, A. HIV treatment failure: testing for HIV resistance in clinical practice. *Science* **1998**, *280* (5371), 1871−3.

(11) Jhaveri, K.; Modi, S. HSP90 Inhibitors for Cancer Therapy and Overcoming Drug Resistance. *Adv. Pharmacol.* **2012**, *65*, 471−517.

(12) Ballif, M.; Harino, P.; Ley, S.; Coscolla, M.; Niemann, S.; Carter, R.; Coulter, C.; Borrell, S.; Siba, P.; Phuanukoonnon, S.; Gagneux, S.; Beck, H. P. Drug resistance-conferring mutations in Mycobacterium tuberculosis from Madang, Papua New Guinea. *BMC Microbiol.* **2012**, *12* (1), 191.

(13) Henkel, J. Attacking AIDS with a 'cocktail' therapy? *FDA Consum.* **1999**, *33* (4), 12−7.

(14) Kahrstrom, C. T. New TB drug cocktail. *Nat. Rev. Microbiol.* **2012**, *10* (9), 594.

(15) Shiffman, M. L. Interferon-free regimens: the near future, the likely and the not so likely. *Clin. Liver Dis.* **2011**, *15* (3), 665−75.

(16) Foulkes-Murzycki, J. E.; Scott, W. R.; Schiffer, C. A. Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. *Structure* **2007**, *15* (2), 225−33.

(17) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105* (28), 6474−6487.

(18) Gillet, J. P.; Gottesman, M. M. Mechanisms of multidrug resistance in cancer. *Methods Mol. Biol.* **2010**, *596*, 47−76.

(19) Gotte, M. The distinct contributions of fitness and genetic barrier to the development of antiviral drug resistance. *Curr. Opin. Virol.* **2012**, *2*, 644−650.

(20) 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. USA* **1988**, *85* (13), 4686−90.

(21) Sa-Filho, D. J.; Costa, L. J.; de Oliveira, C. F.; Guimaraes, A. P.; Accetturi, C. A.; Tanuri, A.; Diaz, R. S. Analysis of the protease sequences of HIV-1 infected individuals after Indinavir monotherapy. *J. Clin. Virol.* **2003**, *28* (2), 186−202.

(22) Beerenwinkel, N.; Schmidt, B.; Walter, H.; Kaiser, R.; Lengauer, T.; Hoffmann, D.; Korn, K.; Selbig, J. Diversity and complexity of HIV-1 drug resistance: a bioinformatics approach to predicting phenotype from genotype. *Proc. Natl. Acad. Sci. USA* **2002**, *99* (12), 8271−6.

(23) Rhee, S. Y.; Taylor, J.; Wadhera, G.; Ben-Hur, A.; Brutlag, D. L.; Shafer, R. W. Genotypic predictors of human immunodeficiency virus type 1 drug resistance. *Proc. Natl. Acad. Sci. USA* **2006**, *103* (46), 17355−60.

(24) Beerenwinkel, N.; Sing, T.; Lengauer, T.; Rahnenfuhrer, J.; Roomp, K.; Savenkov, I.; Fischer, R.; Hoffmann, D.; Selbig, J.; Korn,

K.; Walter, H.; Berg, T.; Braun, P.; Fatkenheuer, G.; Oette, M.; Rockstroh, J.; Kupfer, B.; Kaiser, R.; Daumer, M. Computational methods for the design of effective therapies against drug resistant HIV strains. *Bioinformatics* **2005**, *21* (21), 3943−50.
(25) Saigo, H.; Uno, T.; Tsuda, K. Mining complex genotypic features for predicting HIV-1 drug resistance. *Bioinformatics* **2007**, *23* (18), 2455−62.
(26) Chen, X.; Weber, I. T.; Harrison, R. W. Molecular dynamics simulations of 14 HIV protease mutants in complexes with indinavir. *J. Mol. Model.* **2004**, *10* (5−6), 373−81.
(27) Wang, W.; Kollman, P. A. Computational study of protein specificity: the molecular basis of HIV-1 protease drug resistance. *Proc. Natl. Acad. Sci. USA* **2001**, *98* (26), 14937−42.
(28) Hou, T.; McLaughlin, W. A.; Wang, W. Evaluating the potency of HIV-1 protease drugs to combat resistance. *Proteins* **2008**, *71* (3), 1163−74.
(29) Hou, T.; Yu, R. Molecular dynamics and free energy studies on the wild-type and double mutant HIV-1 protease complexed with amprenavir and two amprenavir-related inhibitors: mechanism for binding and drug resistance. *J. Med. Chem.* **2007**, *50* (6), 1177−88.
(30) Sadiq, S. K.; Wright, D. W.; Kenway, O. A.; Coveney, P. V. Accurate ensemble molecular dynamics binding free energy ranking of multidrug-resistant HIV-1 proteases. *J. Chem. Inf. Model.* **2010**, *50* (5), 890−905.
(31) Wu, K. W.; Chen, P. C.; Wang, J.; Sun, Y. C. Computation of relative binding free energy for an inhibitor and its analogs binding with Erk kinase using thermodynamic integration MD simulation. *J. Comput. Aided Mol. Des.* **2012**, *26* (10), 1159−69.
(32) Hou, T.; Zhang, W.; Wang, J.; Wang, W. Predicting drug resistance of the HIV-1 protease using molecular interaction energy components. *Proteins* **2009**, *74* (4), 837−46.
(33) Zhang, J.; Hou, T.; Wang, W.; Liu, J. S. Detecting and understanding combinatorial mutation patterns responsible for HIV drug resistance. *Proc. Natl. Acad. Sci. USA* **2010**, *107* (4), 1321−6.
(34) Chothia, C. Structural invariants in protein folding. *Nature* **1975**, *254* (5498), 304−8.
(35) Eisenberg, D.; McLachlan, A. D. Solvation energy in protein folding and binding. *Nature* **1986**, *319* (6050), 199−203.
(36) Street, A. G.; Mayo, S. L. Pairwise calculation of protein solvent-accessible surface areas. *Fold Des* **1998**, *3* (4), 253−8.
(37) 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* (1), 298−303.
(38) Muzammil, S.; Armstrong, A. A.; Kang, L. W.; Jakalian, A.; Bonneau, P. R.; Schmelmer, V.; Amzel, L. M.; Freire, E. Unique thermodynamic response of tipranavir to human immunodeficiency virus type 1 protease drug resistance mutations. *J. Virol.* **2007**, *81* (10), 5144−54.
(39) Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. Crystal-Structure of Hiv-1 Protease in Complex with Vx-478, a Potent and Orally Bioavailable Inhibitor of the Enzyme. *J. Am. Chem. Soc.* **1995**, *117* (3), 1181−1182.
(40) Chen, Z.; Li, Y.; Chen, E.; Hall, D. L.; Darke, P. L.; Culberson, C.; Shafer, J. A.; Kuo, L. C. Crystal structure at 1.9-A resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases. *J. Biol. Chem.* **1994**, *269* (42), 26344−8.
(41) Stoll, V.; Qin, W.; 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* (8), 2803−6.
(42) Kaldor, S. W.; Kalish, V. J.; Davies, J. F., 2nd; Shetty, B. V.; Fritz, J. E.; Appelt, K.; Burgess, J. A.; Campanale, K. M.; Chirgadze, N. Y.; Clawson, D. K.; Dressman, B. A.; Hatch, S. D.; Khalil, D. A.; Kosa, M. B.; Lubbehusen, P. P.; Muesing, M. A.; Patick, A. K.; Reich, S. H.; Su, K. S.; Tatlock, J. H. Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* **1997**, *40* (24), 3979−85.
(43) Krohn, A.; Redshaw, S.; Ritchie, J. C.; Graves, B. J.; Hatada, M. H. Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere. *J. Med. Chem.* **1991**, *34* (11), 3340−2.
(44) Kempf, D. J.; Marsh, K. C.; Denissen, J. F.; McDonald, E.; Vasavanonda, S.; Flentge, C. A.; Green, B. E.; Fino, L.; Park, C. H.; Kong, X. P.; et al. ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proc. Natl. Acad. Sci. USA* **1995**, *92* (7), 2484−8.
(45) 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* (7), 1750−9.
(46) 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* (7), 1739−49.
(47) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225−11236.
(48) Case, D. A.; D., T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; W., R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, I. G.; Kolossváry, K. F. W.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; T., C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, a. P. A. *Amber 10*; University of California, San Francisco, 2008.
(49) Hyland, L. J.; Tomaszek, T. A., Jr.; Meek, T. D. Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry* **1991**, *30* (34), 8454−63.
(50) Ido, E.; Han, H. P.; Kezdy, F. J.; Tang, J. Kinetic studies of human immunodeficiency virus type 1 protease and its active-site hydrogen bond mutant A28S. *J. Biol. Chem.* **1991**, *266* (36), 24359−66.
(51) Frisch, M. J. T.; G., W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2003.
(52) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; 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* (16), 1999−2012.
(53) Wang, J.; 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* (9), 1157−74.
(54) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L., Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **2009**, *77* (4), 778−95.

(55) Wang, W.; Kollman, P. A. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *J. Mol. Biol.* **2000**, *303* (4), 567−82.
(56) Hou, T.; Wang, J.; Li, 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* (1), 69−82.
(57) Bashford, D.; Case, D. A. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−52.
(58) Chong, L. T.; Duan, Y.; Wang, L.; Massova, I.; Kollman, P. A. Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proc. Natl. Acad. Sci. USA* **1999**, *96* (25), 14330−14335.
(59) Wang, W.; Kollman, P. A. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and continuum solvent model. *J. Mol. Bol.* **2000**, *303* (4), 567−582.
(60) Case, D. A.; Cheatham, T. E., 3rd; 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* (16), 1668−88.
(61) Zhang, W.; Hou, T.; Qiao, X.; Xu, X. Some basic data structures and algorithms for chemical generic programming. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (5), 1571−5.
(62) Chang, C. C.; L. C. *LIBSVM: a library for support vector machine*. http://www.csie.ntu.edu.tw/∼cjlin/libsvm, 2001.
(63) Hou, T.; Zhang, W.; Wang, J.; Wang, W. Predicting drug resistance of the HIV-1 protease using molecular interaction energy components. *Proteins* **2009**, *74* (4), 837−846.
(64) 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. *Antivir. Ther.* **2002**, *7* (2), 123−9.
(65) Meynard, J. L.; Vray, M.; Morand-Joubert, L.; Race, E.; Descamps, D.; Peytavin, G.; Matheron, S.; Lamotte, C.; Guiramand, S.; Costagliola, D.; Brun-Vezinet, F.; Clavel, F.; Girard, P. M. Phenotypic or genotypic resistance testing for choosing antiretroviral therapy after treatment failure: a randomized trial. *AIDS* **2002**, *16* (5), 727−36.
(66) Reid, C.; B., R.; Day, S.; Larder, B.; DeGruttola, V.; Winslow, D. A Dynamic Rules-Based Interpretation System Derived by an Expert Panel Is Predictive of Virological Failure. *Antivir Ther.* **2002**, *7* (Suppl 1), S91.
(67) Wensing, A. M.; van Maarseveen, N. M.; Nijhuis, M. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral Res.* **2010**, *85* (1), 59−74.
(68) Kovalevsky, A. Y.; Liu, F.; Leshchenko, S.; Ghosh, A. K.; Louis, J. M.; Harrison, R. W.; Weber, I. T. Ultra-high resolution crystal structure of HIV-1 protease mutant reveals two binding sites for clinical inhibitor TMC114. *J. Mol. Biol.* **2006**, *363* (1), 161−73.
(69) Kullback, S.; Leibler, R. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22* (1), 79−86.
(70) Ding, B.; J. W.; 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.