JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

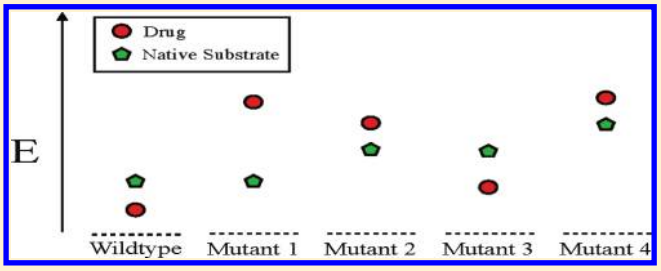
Efficient a Priori Identification of Drug Resistant Mutations Using Dead-End Elimination and MM-PBSA.

Maria Safi and Ryan H. Lilien*

Department of Computer Science, University of Toronto, Toronto, Ontario M5S 3G4, Canada

Supporting Information

ABSTRACT: Active site mutations that disrupt drug binding are an important mechanism of drug resistance. Computational methods capable of predicting resistance a priori are poised to become extremely useful tools in the fields of drug discovery and treatment design. In this paper, we describe an approach to predicting drug resistance on the basis of Dead-End Elimination and MM-PBSA that requires no prior knowledge of resistance. Our method utilizes a two-pass search to identify mutations that impair drug binding while maintaining affinity for the native substrate. We use our method to probe resistance in four drug-target systems: isoniazid-enoyl-ACP reductase (tuberculosis), ritonavir-HIV protease (HIV), methotrexate-dihydrofolate reductase (breast cancer and leukemia), and gleevec-ABL kinase (leukemia). We validate our model using clinically known resistance mutations for all four test systems. In all cases, the model correctly predicts the majority of known resistance mutations.



INTRODUCTION

The emergence of drug resistance remains a significant and frustrating cause of treatment failure. Four molecular mechanisms underly the majority of drug resistance are (1) point mutations in the drug target, (2) alterations in nontarget compensatory genes, (3) increased drug metabolism, and (4) reduction in intracellular concentration through reduced cellular uptake or upregulated small molecule efflux.[1−3] In this paper, we focus on modeling the most common of these methods: the introduction of mutations within the drug binding site. To cause resistance, these mutations must maintain near-native protein function, otherwise they are effectively inherently inhibited. Therefore, in order to confer resistance, a binding site mutation should reduce drug binding while maintaining native substrate binding at near original levels (Figure 1). The resistance problem is therefore similar to that of modeling binding selectivity in that both problems evaluate the binding preferences of the protein receptor for one molecule over another (i.e., substrate vs drug).[4−7]

Knowledge of potential resistance mutations, before they are clinically observed, would be very useful. During lead prioritization, this knowledge may direct the research team away from candidates that are most likely to confer resistance. Knowledge of resistance mutation hotspots would allow pharmaceutical researchers to favor leads that avoid interactions with these problematic regions of the active site. During lead optimization, this approach could complement the idea of respecting the substrate envelope[8] in guiding medicinal chemists toward modifications aimed at evading resistance. In the clinical setting, knowledge of potential resistance mutations could allow the development of treatment regimens, with drug cocktails likely to maximize efficacy.

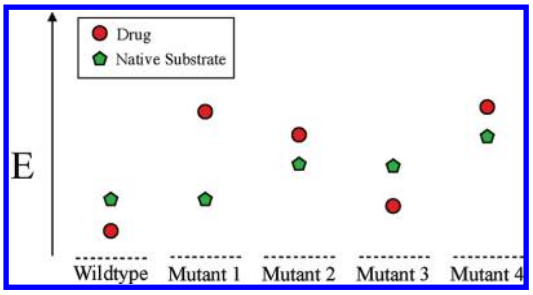


Figure 1. Binding energy shifts of mutant sequences. Hypothetical binding profiles for the native substrate (green polygon) and the drug (red circle). The wild-type protein binds the drug more strongly than the native substrate and is therefore sensitive to the drug. Mutant 1 represents the ideal resistant case: the protein's interaction with the native substrate is no worse than that of the wild type yet binding of the drug is significantly impaired. Mutant 2 represents the more realistic resistant case where both native substrate and drug binding are affected. Mutant 3 preferentially binds the drug over the native substrate and therefore remains sensitive to the inhibitor. Mutant 4 prefers to bind the native substrate; however, the significant decrease in binding energy may result in impaired native function and thus a constitutively inhibited protein.

Methods for modeling and identifying known resistance mutations are slowly emerging. These methods can be divided into two categories: sequence-based methods and structure-

Received: December 27, 2011
Published: May 31, 2012

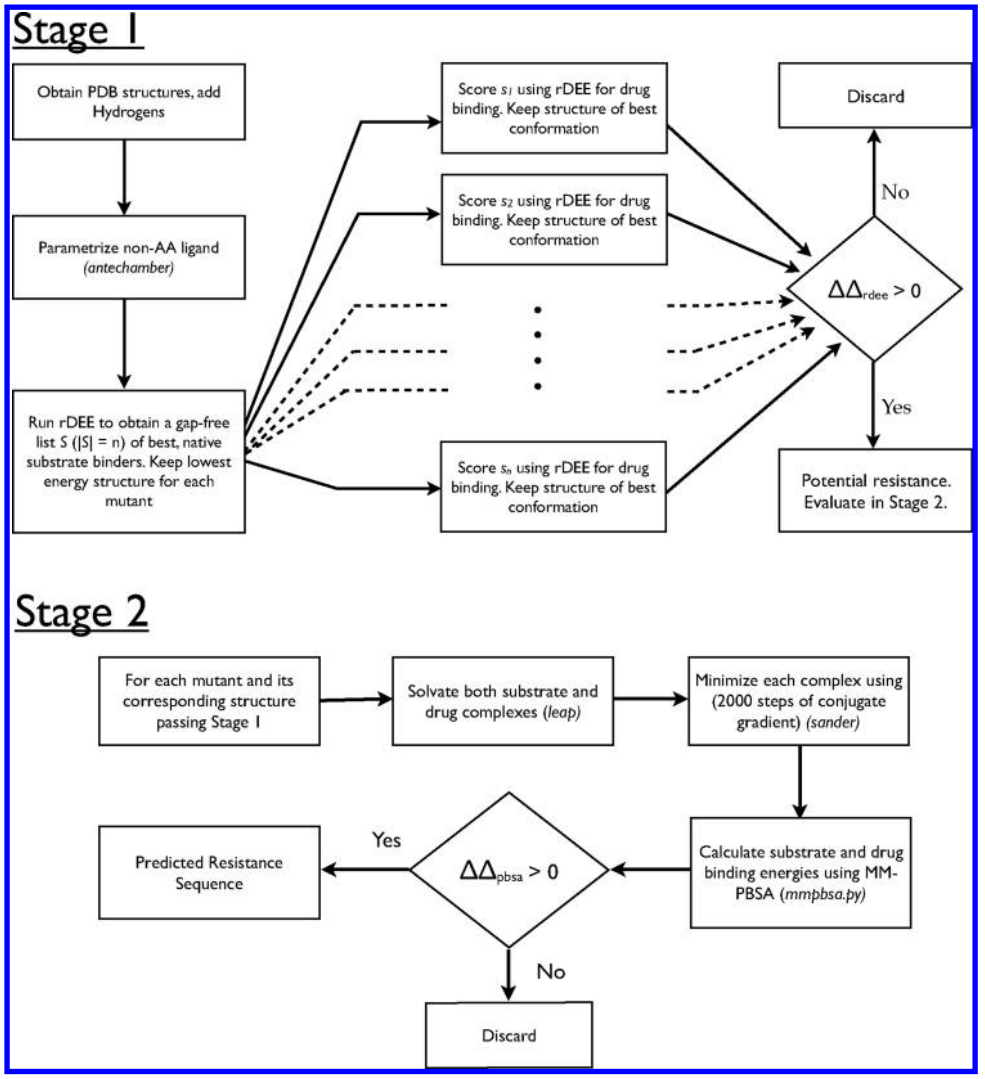


Figure 2. Flowchart of methods. The two-stage approach is displayed. (A): Stage 1. DEE is used to search and score potential mutants. Complex structures with both substrate and drug corresponding to lowest energy conformation for each selected mutant are generated. (B): Stage 2. Mutants that pass Stage 1 are solvated and energy minimized. A PBSA-based approach is used to recalculate binding energies.

based methods. Sequence-based methods currently include both computational and wet lab/clinical analyses. The bulk of sequence-based methods, both computational and experimental, are knowledge based and make use of existing sequence and phenotype data to generate and score potentially resistant sequences. Among the sequence-based methods are the genotypic resistance assays (GRTs) that are primarily used to identify resistant strains in a clinical setting.[9−11] GRTs predictions are based on previously identified molecular markers of resistance. Existing sequence-based computational approaches closely mimic the GRT resistance assays. These approaches employ machine learning and statistical methods such as neural networks and random forests.[12−17] Genetic features indicative of resistance are identified by analyzing known sensitive and resistant sequences. The presence or absence of these features are used to detect resistance in a candidate sequence. These methods are useful in identifying known or combinations of known patterns of resistance; however, these methods are not useful in identifying novel resistance mutations. For systems where knowledge of previously known mutations is small or nonexistent, such as in emerging diseases and new drug targets, the utility of these knowledge-based methods is significantly reduced.

A second class of methods attempts to model the structural effects of known resistance mutations using molecular modeling and molecular dynamics simulations.[18−24] For example, Chen et al. used a docking algorithm to study resistance in several drug targets.[25] They introduced specific, known resistance mutations into the protein target and used docking algorithms to study the effects of these mutations on drug and substrate binding. These molecular modeling approaches have been useful in understanding the structural basis of known resistance mutations, but these algorithms are not designed to search the combinatorial number of potential mutations to identify novel resistant sequences.

Because resistance mutations typically involve a small number of amino acid changes to the active site, it may be possible to predict new resistance mutations in the drug targets before they arise. This involves searching through and ranking a large number of possible candidate solutions. An exciting new direction is the use of protein redesign algorithms to identify novel resistant sequences using first principles (i.e., without the use of known resistance data). The techniques utilized in

protein redesign are extremely efficient at searching exponentially large search spaces and generating a ranked list of candidate redesigns. This approach does not rely on clinical data and can be useful for systems where very little is known about possible resistance. For example, the Donald lab successfully used the K^* ensemble-based protein redesign algorithm[26] to identify novel resistance mutations in dihydrofolate reductase.[27] Their work demonstrates that computational methods can be useful when modeling a priori drug resistance.

In the current work, we propose a general framework to probe active site localized resistance mutations. Our approach uses our restricted Dead-End Elimination (rDEE)-based protein design[28] coupled with a two-pass search and scoring method based on the more biophysically accurate MM-PBSA model. The use of DEE-based methods makes our approach deterministic, fast, and guaranteed to identify the lowest energy solution from the specified search space. DEE was initially described by[29] and has been successfully employed in several redesigns.[30−33] In this work, we tested our model's ability to predict resistance mutations in four drug target systems. These four systems serve as our validation set. For each experiment, the algorithm had no foreknowledge of known resistance mutations. In all cases, the model's predicted mutations have good agreement with the known mutations. We conclude that the use of protein redesign methods, including DEE, has significant potential to identify previously unseen resistance mutations in a range of drug targets.

METHODS

We define the term "mutation sequence" to refer to the sequence of amino acids of a mutated gene variant. Our method consists of two stages. The first stage uses an efficient restricted Dead-End Elimination (rDEE)[28] search of allowable mutation space to identify potential resistance mutations. In the second stage, the more accurate yet computationally expensive MM-PBSA scoring method is used to validate and rank the identified mutation sequences.

STAGE 1: EFFICIENT DEAD-END ELIMINATION-BASED SEARCH

In the first stage, we use a Dead-End Elimination-based search to identify mutation sequences that disrupt drug binding while maintaining sufficient binding of the native substrate. To enforce this constraint, a two-pass search procedure was utilized: the *native substrate pass* and the *drug pass*. To improve computational efficiency, our method utilizes restricted Dead-End elimination (rDEE) in combination with restricted A^* enumeration. These methods are fully described elsewhere.[28] Protein conformations are scored using an efficient pairwise energy function that includes the dihedral, electrostatic, and vdW terms of the AMBER energy function.[34,35] A flowchart depicting the approach appears in Figure 2.

Our use of DEE closely parallels its use in a typical DEE-based protein redesign. Each of the protein's residues is modeled as either rigid and not mutable, flexible and not mutable, or flexible and mutable. Side-chain flexibility is modeled using a discrete rotamer library of low-energy conformations.[36] Rigid residues include those sufficiently far from the region of interest such that that the rigid approximation is not likely to significantly affect the model. A subset of the n active site residues are modeled as flexible and mutable, while the remaining active site residues are modeled as flexible and not mutable. In the context of modeling drug resistance, it is extremely unlikely that a large number of active site residues will mutate simultaneously. In other words, for most drug-target systems, we expect a resistant sequence to contain k mutations where $k < n$ and typically $k \leq 4$. Because the number of mutations in a resistant phenotype is often much smaller than the number of residues in the active site and because traditional DEE algorithms do not have a mechanism for restricting the search to any of the $\binom{n}{k}$ allowed solutions, traditional DEE algorithms are not an efficient choice for resistance prediction. To date, restricted redesign has been accomplished either by running $\binom{n}{k}$ separate searches where in each search only k specific residues are allowed to mutate or by using a single run where solutions are enumerated until a sequence with only k mutations is generated. In the latter case, care must be taken to avoid pruning the desired solution. Recently, we described a restricted version of DEE (rDEE) and A^* specifically tailored to facilitate the search for redesigns with a limited number of allowed mutations.[28] An rDEE A^* search is therefore an ideal choice for predicting resistance mutations. It removes the need to perform $\binom{n}{k}$ separate searches and is faster than searching through the results of an unrestricted run.

Native Substrate Pass. The search for mutation sequences capable of maintaining native substrate binding is termed a *positive* design. A gap free list of mutation sequences ranked on their predicted ability to bind the native substrate is identified using rDEE and restricted A^* enumeration. Mutation sequences whose predicted native substrate binding energies are no more than 1.5 kcal/mol worse than the wild type were identified. Mutation sequences whose predicted native substrate binding energies were worse than this threshold were eliminated from consideration. We emphasize that this threshold is a system parameter that the user can use to control the number of mutant sequences that are further evaluated. We chose a threshold of 1.5 kcal/mol for all reported experiments to allow a compromise between computational efficiency and percentage of the mutation search space evaluated beyond the native substrate pass. However, the user may modify this threshold when evaluating different protein−drug systems.

Drug Pass. The search for mutation sequences with reduced drug binding is termed a *negative* design. The output of the native substrate pass is a set of sequences S predicted to bind the native substrate no worse than 1.5 kcal/mol of the wild type sequence. In the drug pass, all sequences in S are screened for drug binding. Individual DEE searches are used to identify the lowest energy conformation for each mutation sequence $s \in S$. In these DEE searches, residues are modeled as flexible or rigid, but because the mutations away from the wild type are implicit in s, no further mutations are permitted. The best binding energy between the drug and any conformation of s, as computed by the AMBER energy function (see Additional Modeling Details), is saved as the drug-interaction energy of sequence s. We discard any mutation sequence predicted to have stronger drug binding than the wild-type protein because such a mutant is unlikely to cause resistance. Thus, at the end of the Drug Pass, the surviving mutation sequences have the following two properties: first, the binding energy of the mutant

dx.doi.org/10.1021/ci200626m | J. Chem. Inf. Model. 2012, 52, 1529−1541

with the native substrate is no worse than 1.5 kcal/mol of the binding energy of the wild type with the native substrate, and second, the binding energy of the mutant with the drug is worse than that of the wild type with the drug (i.e., the mutant disrupts drug binding). Finally, for each of these surviving mutation sequences, the best interaction energy with the native substrate and the best interaction energy with the drug are used to compute a $\Delta\Delta_{rdee}$ score (see Scoring Function below).

Additional DEE Search Details. For both the positive and negative design passes, protein systems consist of all residues with at least one atom within 10 Å of the native substrate. The subset of these residues, which together compose the active site, are modeled as both flexible and mutable. These active site residues are identified from previous structural studies[37−41] and PDBeMotif.[42] Residues that are included (i.e., those within 10 Å) but are not part of the active site form a steric shell that is modeled as rigid and nonmutable during the DEE based search.

STAGE 2: RESCORING WITH MM-PBSA

The first stage serves as a filter to identify candidate resistant mutation sequences and their corresponding low-energy structures. In the second stage, these candidates are rescored using the more accurate and more computationally expensive MM-XBSA models as implemented in AMBER 11.0[34,35] (XBSA refers to either the PBSA or GBSA methods). While molecular modeling approaches are far from perfect,[43] both the MM-PBSA and the MM-GBSA techniques have demonstrated reasonable accuracy on a number of different test systems.[44−47] In Stage 2 of the current work, we utilized MM-PBSA as it is typically more accurate than MM-GBSA at the cost of added runtime.[48] We performed experiments using both MM-PBSA and MM-GBSA in Stage 2 and found no significant differences in results; however, MM-PBSA takes approximately four times longer to compute. Therefore, if optimizing runtime is crucial, MM-GBSA could be used instead of MM-PBSA in Stage 2.

In Stage 2, we rescore each mutation sequence by considering the two structures generated in Stage 1: the lowest energy (best) conformation of the protein bound with the drug and the lowest energy conformation of the protein bound with the native substrate. The rescoring process utilizes all protein residues, including those beyond the steric shell and active site. There are three steps to this rescoring.

First, the mutant structure generated by rDEE in Stage 1 is parametrized using the *leap* module in AMBER 11.0 and solvated in an octahedral box of TIP3P water molecules extending 12 Å beyond the protein on all sides. Second, the solvated complexes are minimized with 500 steps of steepest descent minimization followed by 1500 steps of conjugate-gradient minimization. No restraints are used during minimization. A residue-based cutoff of 12 Å for nonbonded terms is used. For comparison, we also evaluated longer minimizations with 5000 and 10000 steps; however, in all four protein systems, 2000 steps (500 + 1500) was sufficient for the minimization to converge. Finally, we compute the ligand binding energy of the solvated, minimized complex. Poisson−Boltzmann (PBSA) implementation of mmpbsa.py module in AMBER 11.0 is used to compute the ligand binding energy. The overall workflow evaluates binding energies using AMBER PBSA, an explicit solvent model, and energy minimization (but no molecular dynamics). This pipeline was introduced by Ferrari et al.[44] The authors report promising correlation (∼0.8) between experimentally determined energies and those predicted by AMBER PBSA using explicit solvent for a set of aldose inhibitors.

Using the bound structures generated from Stage 1, each mutation sequence is scored for both native substrate and drug binding. The $\Delta\Delta_{pbsa}$ score is calculated (see below). Mutation sequences with positive values for both ΔΔ scores ($\Delta\Delta_{rdee}$ and $\Delta\Delta_{pbsa}$) are categorized as the predicted resistant mutation sequences. A diagram of the process is shown in Figure 2.

Scoring Function. We defined a scoring function to measure resistance. The scoring function computes the difference in binding energy between the wild-type and mutation sequence for both the native substrate and the drug. An "ideal" resistant mutation sequence would maintain native interaction energy with the native substrate while decreasing the protein's interaction energy with the drug (Figure 1). The improvement in binding energy for the native substrate is measured as $E_{wt,s} - E_{mut,s}$, and the decrease in binding energy for the drug is measured as $E_{wt,d} - E_{mut,d}$. Where $E_{x,y}$ is the interaction energy between protein x (wt, wild-type; mut, mutation sequence) and molecule y (s, native substrate; d, drug). The terms are combined to create a resistance score,

$$\Delta\Delta = (E_{wt,s} - E_{mut,s}) - (E_{wt,d} - E_{mut,d}) \quad (1)$$

Two resistance scores can be computed: one using the interaction energies from Stage 1 and one from the MM-PBSA based energies of Stage 2. The method used to calculate the ΔΔ score is indicated in the subscript; thus scores calculated by rDEE and PBSA are referred to as $\Delta\Delta_{rdee}$ and $\Delta\Delta_{pbsa}$ respectively.

A positive ΔΔ score indicates a possible resistance mutation. A positive score is obtained when a mutation sequence disrupts drug binding more than it disrupts binding the native substrate. A negative score indicates the opposite and is thus unlikely to confer resistance. The scoring function of eq 1 is intuitively similar to the scoring functions used in a number of published binding selectivity studies.[49−51]

In summary, a series of filters and scoring methods are applied to each molecular system. First, a positive design pass applies rDEE to the protein and native substrate system. Sequences with sufficiently "good" energies are then evaluated in a negative design pass to assess drug binding and compute a $\Delta\Delta_{rdee}$ resistance score. Mutation sequences with a positive resistance score are reevaluated in Stage 2 using MM-PBSA. Mutation sequences with both positive $\Delta\Delta_{rdee}$ and $\Delta\Delta_{pbsa}$ scores are output as predicted resistant mutation sequences. All mutations that do not pass Stage 1 or that have a negative $\Delta\Delta_{pbsa}$ score are considered sensitive to the drug.

Molecular Systems. For all four systems, only the active site residues (specified for each system below) were modeled as flexible and mutable. For rDEE runs, a rigid steric shell consisting of residues with at least one atom within 10 Å of the active site (and not explicitly modeled as active site) was included in energy computation. In Stage 2, all protein residues were included in calculations.

Isoniazid-enoyl-ACP Reductase. Isoniazid is a competitive inhibitor of *Mycobacterium tuberculosis* enoyl-ACP reductase. We utilized the structures from PDB IDs: 2IDZ (inhibitor bound complex) and 1BVR (native substrate bound complex). Similar to previous work,[38] the following active site residues were modeled as flexible/mutable: Ile 16, Ile 21, Phe 41, Ile 47, Ser 94, Phe 149, Lys 165, Leu 218, and Trp 222. Enoyl-ACP reductase binds NADH and a fatty acyl substrate for its native

function. The previously known resistance mutations for isoniazid are clustered within the NADH binding site.[38]

Ritonavir-HIV Protease. Ritonavir is a protease inhibitor used against HIV. We utilized the structures from PDB IDs: 1N49 (inhibitor bound complex) and 1F7A (native substrate bound complex). The 1F7A structure contains the D25N amino acid substitution but is often used as the native form of HIV protease in structure-based modeling studie.[52,53] Similar to previous work,[54,55] the following 11 active site residues were modeled as flexible/mutable: Gly 27, Asp 29, Asp 30, Met 46, Gly 48, Ile 50, Ile 54, Val 82, Ile 84, Gly 126, and Ile 146. Mutations in the active site catalytic loop, Asp 25, Thr 26, and Gly 27, are known to adversely impact HIV protease function; therefore, these residues were not allowed to mutate during Stage 1.

Methotrexate-DHFR. Methotrexate is an anticancer drug targeting human dihydrofolate reductase (hDHFR). We utilized the structures from PDB IDs: 3EIG (inhibitor bound complex) and 1DRF (native substrate bound complex). The structure 3EIG contains two active site amino acid substitutions (F31R and Q35E). We reverted these substitutions back to wild type and performed 5000 steps of steepest descent unrestrained energy minimization using the *sander* module in AMBER. The resulting structure was used as our wild-type inhibitor bound complex. Similar to previous work,[40,41] the following 10 human DHFR residues were modeled as flexible/mutable: Ile 7, Leu 22, Glu 30, Phe 31, Arg 32, Phe 34, Gln 35, Leu 67, Val 115, and Thr 136.

Gleevec-ABL Kinase. Gleevec is an anticancer drug that inhibits human ABL kinase. We utilized the structure from PDB ID: 2HYY (inhibitor bound complex) to model drug binding. The gleevec-ABL kinase system presented a modeling challenge as a structure of ABL kinase bound to its native substrate had not been published in the PDB database. In place of the missing structure, we generated an unbound wild-type receptor by removing the inhibitor from 2HYY. The resulting PDB structure then was solvated in *leap* and subjected to 5000 steps of steepest descent unrestrained energy minimization using the *sander* module in AMBER. The resulting unbound ABL kinase structure was used in the native substrate rDEE pass of Stage 1. As the substrate bound complex was missing, the value ($E_{wt,d} - E_{mut,d}$) was used to determine if the mutant disrupted drug binding compared to the wild type (see details below). Similar to previous work,[37,42] the following 14 residues were modeled as flexible/mutable: Tyr 253, Val 256, Lys 271, Glu 286, Met 290, Ile313, Thr 315, Phe 317, Met 318, Ile 360, His 361, Leu 370, Asp 381, and Phe 382.

Additional Modeling Details. As a first step to evaluate our search pipeline, we limited our search to only hydrophobic amino acids (or hydrophobic plus polar neutral for enoyl-ACP reductase), and we evaluated our ability to recover known resistance mutations involving these amino acids in our search space. The approach we took allows us to decouple the performance of our resistance scoring from the individual scoring of molecular interactions of difficult to model residues and provided the best opportunity for validation using known resistance mutations. Residue flexibility in rDEE was modeled using the Lovell, Richardson, and Richardson's side chain rotamer library.[36] A validated implementation of the AMBER energy function (electrostatic, vdW, and dihedral energy terms)[56,57] was used to compute the pairwise energies between residues. The *reduce* module in AMBER 11.0 was used to protonate the input structures in a neutral environment.

RESULTS AND DISCUSSION

We used our two-stage search and scoring method to predict resistance mutations in four target systems. Three of the four searches started with experimentally determined structures of the substrate and drug bound wild-type complexes (the gleevec-ABL kinase system used only an experimental structure of the drug bound wild-type complex). None of the four searches used knowledge of previously published resistance mutations. The goal of these experiments was to demonstrate the ability of our approach to discover resistance mutations. We evaluated the quality of the identified candidate resistance sequences by comparison to known resistance mutations. This validation is only partial; a predicted mutation may indeed be resistant, but to date, it may not have been experimentally verified nor reported in the literature.

Isoniazid Resistance. Isoniazid remains part of the first line treatment for tuberculosis (TB) worldwide. It is a prodrug activated by the *Mycobacterium tuberculosis*'s KatG enzyme to form the acyl-NADH complex (INH-NAD) that binds the target enoyl Acyl-Carrier Protein reductase (enoyl-ACP reductase). Enoyl-ACP reductase is a 270 amino acid long protein involved in type II fatty acid synthesis. For wild-type functionality, the TB enoyl-ACP reductase binds a fatty acyl substrate and an NADH cofactor. The NADH and fatty acyl binding are prevented by the competitive inhibitor isoniazid. Numerous point mutations in enoyl-ACP reductase are known to confer isoniazid resistance; these include I16T, I21T, I21V, I47T, V78A, S94A, and I95P.[58] Five of these mutations (I16T, I21T, I21V, I47T, and S94A) fall within our search space. Our algorithm was used to identify potential resistance mutations in the isoniazid enoyl-ACP reductase system. Most of the isoniazid resistance mutations involve single amino acid changes. We therefore restricted our search to at most one simultaneous mutation. In the Enoyl-ACP reductase system, the nine active site residues (see Methods) were allowed to mutate to any of the following 13 amino acid types: Gly, Ala, Leu, Ile, Val, Phe, Tyr, Trp, Met, Asn, Gln, Ser, and Thr. This set includes the nine hydrophobic amino acids A, F, G, I, L, M, V, W, and Y and four polar neutral amino acids N, Q, S, and T. The polar neutral amino acids were included because a number of known resistance mutations involve mutations to Thr.

In this limited system, the search space contains 117 single-point mutation sequences. A total of 47 sequences passed the native substrate pass (positive design) of the rDEE search indicating that substrate binding was predicted to lie within acceptable limits. These 47 sequences were then scored for isoniazid binding using rDEE (negative design). The $\Delta\Delta_{rdee}$ score was calculated for each sequence. Forty-six sequences showed a positive $\Delta\Delta_{rdee}$ score indicating that these candidate mutations affected drug binding more than they affected the binding of the native substrate. Next, for each of these 46 sequences, the MM-PBSA method was used to calculate the binding energies of both the substrate and isoniazid. Sixteen of the initial 117 single-point mutants had positive $\Delta\Delta$ scores for both scoring methods and were categorized as our predicted resistant sequences (Table 1).

We evaluated the performance of our algorithm in the context of known resistance mutations in the isoniazid enoyl-ACP reductase system. There are five known resistance mutations among the modeled sequences. Four of these five mutation sequences are identified among our list of 16 candidate resistant sequences (I16T, I21T, I21V, and I47T).

Table 1. Predicted Resistance for Isoniazid-TB[a]

Mutation	Comments
I16T, I21T, I21V, I47T	Known resistance mutations
I16V, I21A, I21W, I21F, I21Y, I47V	Plausible, similar to known mutations
K165M, K165Q, F149N	Unlikely, K165, F149 mutants may disrupt function (catalytic triad)
F41M, F41L, L218Y	Less likely, F41 is important for NADH binding

[a]All 16 single mutants predicted resistant by our model are listed. Of the five known mutants, four were predicted as resistant by our approach. Another six of the predicted 16 are highly likely because of their similarity to known mutants.

Of the 12 remaining predicted resistant mutation sequences, six occur at positions I21 (I21A, I21W, I21F, and I21Y), I16 (I16V), or I47 (I47V) and can be considered plausible on the basis of their similarity to the known resistance mutations. Of the remaining six mutations, both F41M and F41L are less likely to confer resistance. F41 is hypothesized to be involved in binding the adenine moiety of both the cofactor NADH and isoniazid. Therefore, a mutation at F41, which disrupts drug binding, may also impair native function. Finally, the remaining three mutations, K165M, K165Q, and F149N, are unlikely to cause resistance because both F149 and K165 have been implicated in the catalytic triad for the enoyl ACP reductase. For example, experimental evaluation of a number of single mutants of K165, including K165A and K165M, indicates that NADH binding is severely affected, inhibiting native function.[58] Structures obtained after energy minimization step of Stage 2 for three known resistance mutations (I21V, I21T and I47T), one plausible mutation (I47V) and one unlikely resistance mutation (F41M) are shown in Figure 3.

Of the five known isoniazid resistant mutation sequences, only one, S94A, was not identified by our method. A review of our analysis identifies that S94A was pruned at the native substrate pass. The mutation was predicted to impair native substrate binding by 3.74 kcal/mol, which was more than the allowed threshold of 1.5 kcal/mol. Interestingly, the S94A mutation is known to confer resistance through the loss of water mediated bonds involving Ser.[59] The native substrate pass likely had difficulty modeling this interaction because although solvent is modeled in Stage 2, there is no explicit water model in our rDEE phase of Stage 1. In summary, four of five known active site resistance mutations were recovered by our method, while an additional six predicted mutations have plausible mechanisms of resistance.

Ritonavir Resistance. HIV protease is an aspartyl protease essential for HIV replication. Most protease inhibitors bind the active site and prevent interaction with the native peptide substrate. Our model consisted of HIV protease, the inhibitor ritonavir, and a native binding peptide (Figure 4). We allowed up to two simultaneous mutations at 11 active site residues (see Methods). Residues were allowed to mutate to the following nine hydrophobic amino acids: Gly, Ala, Leu, Ile, Val, Phe, Tyr, Trp, and Met. A total of 787 mutation sequences out of 3771 passed our native substrate filter. A total of 720 of these sequences made it through the drug pass and were rescored using MM-PBSA. A total of 177 mutation sequences had positive $\Delta\Delta_{rdee}$ and $\Delta\Delta_{pbsa}$ scores and were output as predicted resistant mutation sequences. These sequences are discussed below, and a complete list can be found in the Supporting Information.

The extremely large amount of HIV protease sequence and screening data complicates the evaluation of our model. To simplify comparison, we constructed two validation sets. Our first validation set is derived from the HIV Drug Resistance Database.[60,61] It contains the 28 known single and double residue mutations that lie within our defined search space and confer at least 2.5-fold resistance to ritonavir (Table 2). In this context, *z*-fold resistance indicates that the IC_{50} of ritonavir for the mutant is *z* times higher than that for the wild type. This first set of mutations is the best supported set of known resistance mutations in the literature. We refer to the first set as the Gold Standard Set. Because this is a somewhat conservative list, we also created a second set of plausible mutations using 17 single residue mutations. These single residue mutations are known to confer resistance to at least one protease inhibitor and are within our search space. They include D30GY, M46IL, G48MV, I50V, I54LMV, V82AFLM, and I84FLV.[60−65] In this notation, any single amino acid code after the residue number is a valid single-point mutation (i.e., D30GY indicates that D30G and D30Y are both separate resistance mutations). Our second

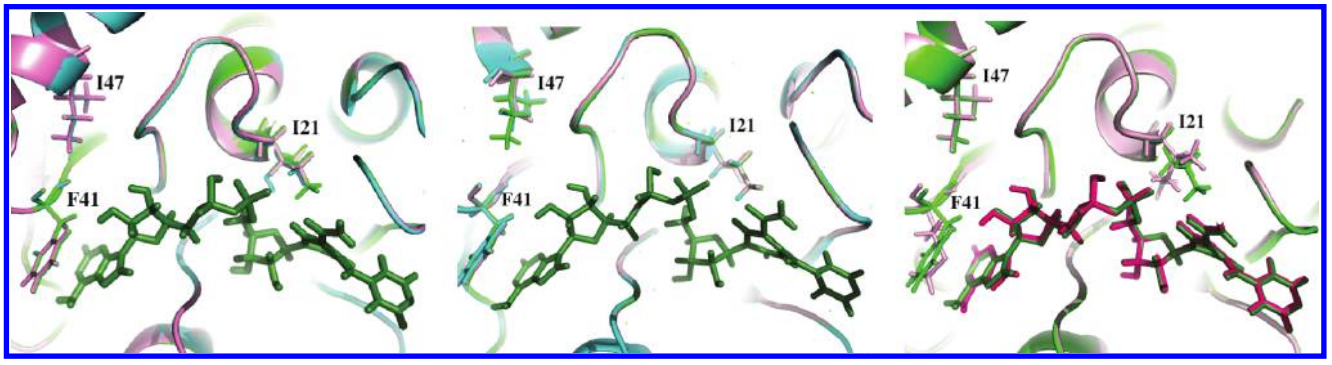


Figure 3. Structures for isoniazid resistant mutations. The enoyl-ACP reductase protein is shown as cartoon with selected residues and isoniazid rendered in stick form. (Left): Two mutations occurring at Ile 21: wild-type sequence (green), I21V (pink), and I21T (cyan). Both mutations are known, and drug binding is predicted to be disrupted by a loss of vdW contacts (~1.2 kcal/mol) in I21V and a loss of electrostatic interactions (~6.5 kcal/mol) in I21T. (Center): Two mutations occurring at Ile 47: wild type (green), predicted and plausible I47V mutation (pink), and known I47T (cyan). A loss of electrostatic interactions (~1 kcal/mol) is predicted to be responsible for the disruption of drug binding. (Right): A mutation at Phe 41: wild type (green) and the predicted F41M (pink). A major loss of both vdW contacts (~4 kcal/mol) and electrostatic interactions (~3 kcal/mol) is predicted compared to the wild type. A single isoniazid molecule in dark green is shown in the left and center panels as the drug does not shift significantly between wild-type and mutant structures. In the right panel, isoniazid's position in the F41M mutant is shown in pink.

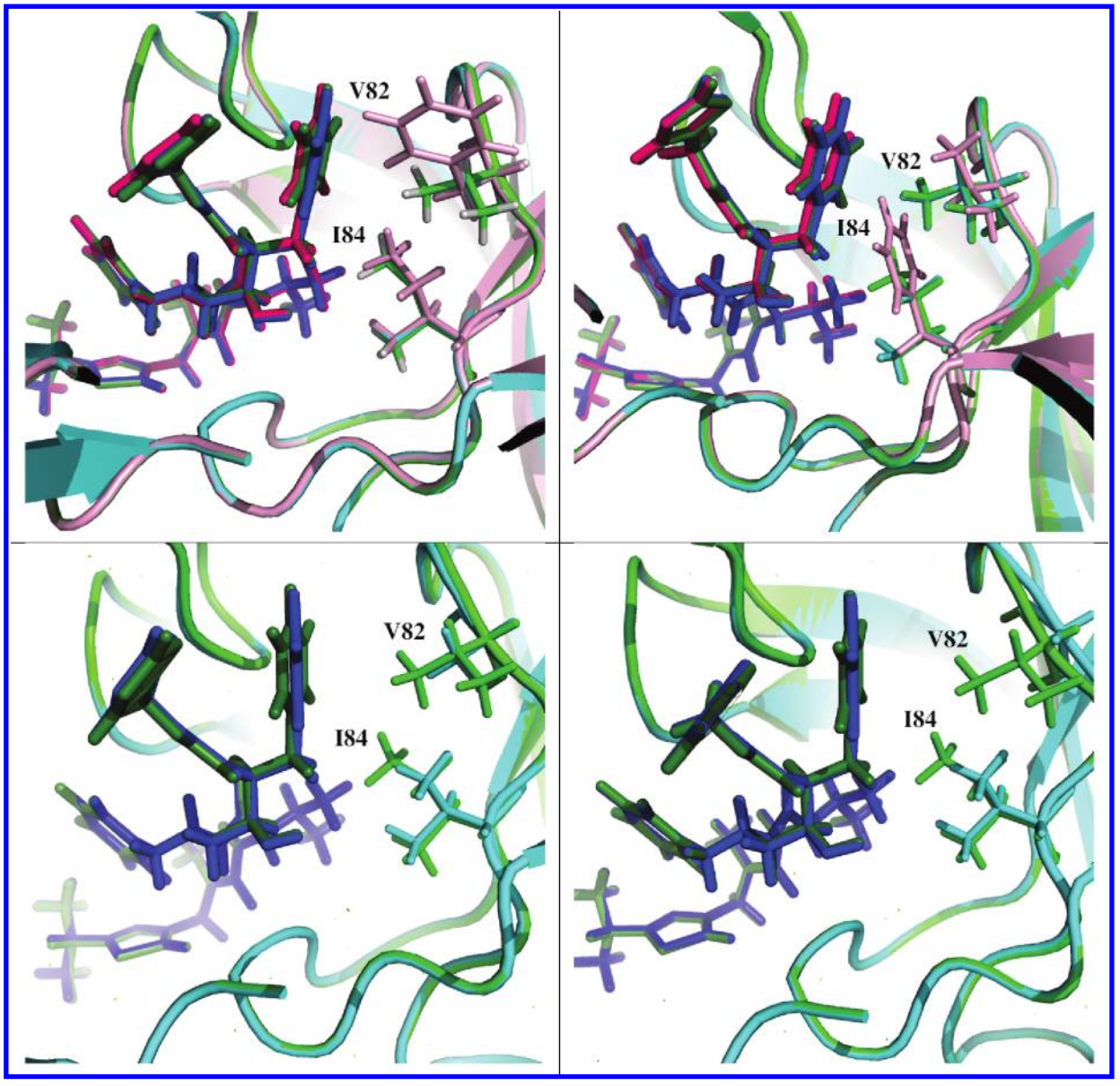


Figure 4. Structures for ritonavir resistance mutants. HIV protease is shown as a cartoon with selected residues and ritonavir in stick form. In all panels, the mutant structures have been superimposed on the wild-type structure (green). In all panels, ritonavir drawn in dark green corresponds to the wild type, otherwise its color reflects the corresponding mutant. (Top Left): Known single-point mutants V82A (cyan) and V82F (pink) are displayed. For V82A, loss of vdW interactions (∼1.5 kcal/mol) is predicted to be the cause of disrupted ritonavir binding. Small changes in both vdW and electrostatic interactions are what cause disrupted binding in V82F. (Top Right): Known single-point mutants I84V (cyan) and I84F (pink) are displayed. For both mutants loss of vdW interactions (∼1 kcal/mol) is the predicted cause of impaired ritonavir binding. (Bottom Left): The structure of known double mutant V82A/I84V (cyan) is shown. Major loss of vdW (∼2.5 kcal/mol) in the mutant structure along with a small loss of electrostatics (∼1 kcal/mol) is predicted to cause disruption of drug binding. (Bottom Right): The structure of predicted double mutant V82G/I84V (cyan) is shown. Major loss of vdW (∼3 kcal/mol) as well as a small loss of electrostatics (∼1 kcal/mol) is predicted to cause disruption in ritonavir binding.

validation set consists of the 138 single and double residue mutations that can be constructed using this set of 17 amino acid substitutions. For example, the double mutant V82A/I84L is in the second set of plausible mutations because both V82A and I84L are resistance conferring. In contrast to the first validation set, not all mutation combinations in the second set have been experimentally verified. The constructed combinations are plausible because the constituent mutations are known to display synergistic resistance. We refer to the second validation set as the Plausible Validation Set.

Gold Standard Validation Set. Of the 28 known resistance mutations in the first validation set, 13 are identified by our method as predicted resistant sequences with positive $\Delta\Delta_{rdee}$ and $\Delta\Delta_{pbsa}$ scores. This represents an enrichment factor of approximately 10 (28/3771 positive in the entire search space compared with 13/177 positive in the search results). The 15 nonidentified mutations were pruned at one of the search stages. Six were eliminated in the native substrate pass because the rDEE-based prediction of native binding energy was affected by more than the allowed 1.5 kcal/mol. Figure 5 shows a recovery plot and ROC curve. At our threshold of 1.5

dx.doi.org/10.1021/ci200626m | J. Chem. Inf. Model. 2012, 52, 1529−1541

Table 2. Gold Standard Validation Set for Ritonavir Resistance in HIV Protease[a]

Mutation	Prediction	Fold Res.	Comment
M46I/V82A	S	400	Does not pass substrate filter.
V82A/I84V	R	400	
I54A/V82A	R	212	
I54V/I84V	S	201	Does not pass Stage 2. I54V/I84FL predicted resistant.
I54V/V82F	R	128	
I54L/V82A	R	118	
M46I/I84A	S	67	Does not pass substrate filter.
M46I/I84V	S	48	Does not pass substrate filter. M46L/I84V predicted Resistant.
M46L/V82A	R	45	
I54L/I84V	S	29	Does not pass Stage 2. I54L/I84Y is resistant.
I54V/V82A	R	22	
G48V/V82A	S	15	Does not pass substrate filter.
M46I/V82F	S	15	Does not pass substrate filter.
V82A	R	11	
I54M/V82A	R	9.6	
I54V	S	8.8	Does not pass Stage 2.
M46L/I84V	R	8.4	
I50V	S	8.2	Does not pass Stage 2.
V82F	R	8.0	
M46L/V82L	R	5.8	
M46L/I54L	S	5.5	Does not pass Stage 2. M46L/I54M predicted Resistant.
M46I/I50V	S	5.2	Does not pass substrate filter.
I54L	S	5	Does not pass Stage 2.
I84V	R	4.5	
I54M	S	4.4	Does not pass Stage 2.
M46I	S	4	Does not pass substrate filter.
V82L	R	3.1	
M46L	S	2.5	Does not pass Stage 1

[a]The validation set of 28 known single- and double-point ritonavir resistance conferring mutations obtained from HIV-DB are listed in order of fold resistance. Only the mutants in modeled residues where the fold-resistance was more than 2.5 are included. The prediction column indicates the prediction result of our algorithm (R, predicted resistance sequence; S, predicted sensitive).

kcal/mol, approximately 75% of the known resistance mutations are screened, yet only 18% of the entire search space makes it through the native substrate pass. By relaxing the native substrate binding threshold from 1.5 to 4.0 kcal/mol, the numbers are approximately 85% of the known resistance mutations and 25% of the entire search space, respectively. Therefore, by modifying this threshold, the user can adjust the number of sequences screened for resistance.

Plausible Validation Set. Of the 138 single and double mutants in the plausible validation set, 40 are identified among our list of candidate mutation sequences. This represents an enrichment factor of 6.2 (138/3771 positive in the entire search space, compared with 40/177 positive in the search results). An interesting phenomenon involves secondary or compensatory mutations. We employ a strict native substrate binding cutoff to ensure that only mutants capable of binding the native substrate are considered for resistance. Some of the known single-point mutants such as M46I and D30Y do not pass this filter. However, allowing the freedom to incorporate a second and potentially compensatory mutation allows double mutants greater ability to maintain near native binding. Thus, numerous double-point mutants that include a known single-point resistance mutation pass the substrate filter and are further screened. A total of 89 mutation sequences or 50.3% of our 177 output sequences fall into this category.

Seven of these mutants are combined with known mutant D30Y and four include known mutant I50V. Fourteen of the 89 compensatory mutation sequences identified by our search include a combination of V82 and I84 (V82A/I84YM, V82Y/I84LF, V82W/I84LF, V82M/I84Y, V82L/I84Y, V82G/I84FLV, V82I/I84F, and V82F/I84YM). There is experimental support for several protease inhibitor resistance mutations involving this pair.[66,67] A total of 48 of the 177 mutation sequences predicted by our model are neither covered by the Plausible Validation Set nor by the described compensatory phenomenon. These 48 sequences may be false positives or unknown novel true positives. A list of all 177 mutants can be found in Supporting Information.

Methotrexate Resistance. Human DHFR is a frequent chemotherapeutic target. It plays an important role in cell proliferation through its involvement in folic acid metabolism and the production of purines. In its native state, DHFR catalyzes the production of tetrahydrofolate from dihydrofolate and the electron donor NADPH. The chemotherapeutic agent methotrexate (MTX) inhibits cell proliferation by binding DHFR approximately one thousand times more tightly than the native folate. Resistance to methotrexate is widespread and arises both from the upregulation of DHFR and from the introduction of point mutations in the DHFR protein. Unfortunately, it is not clear which amino acid substitutions are primarily responsible for conferring resistance. Instead, methotrexate resistance is associated with a range of single and double amino acid mutations at several active site "hotspots".

As in the previous two cases, each active site residue could mutate to the hydrophobic amino acids: Gly, Ala, Leu, Ile, Val, Phe, Tyr, Trp, and Met. We allowed up to two simultaneous mutations at the 10 active site residues (see Methods). This defines a search space of 3258 mutation sequences. Resistance conferring mutations are known to occur at DHFR residues Ile 7, Leu 22, Phe 31, Phe 34, Asp 35, and Val 115. Four of these residues, positions 22, 31, 34, and 35, are mutation hotspots where empirical observation suggests that individual mutations at these residues can confer MTX resistance.[41,68−70] Two-point mutations, where each mutation occurs in these hotspots, are also known to be MTX resistant.[41,70] In the absence of a verified and concise list of known resistant mutations, we define a Plausible Validation Set as the mutation sequences with one or two amino acid substitutions involving only residues 22, 31, 34, and 35. The Plausible Validation Set contains 441 or 13.5% of the search space's 3258 mutation sequences.

Our search produces 272 mutation sequences with a positive $\Delta\Delta_{rdee}$ score. A total of 75 single- and double-point mutations had positive $\Delta\Delta$ scores for both rDEE and MM-PBSA scoring methods and compose our set of predicted resistant sequences. The Plausible Validation Set contains 18 of these 75 sequences (24%). This represents an approximate 2-fold enrichment in identified sequences over their native abundance (13.5%). A complete list of all 75 predicted mutations appears in the Supporting Information. Beyond the mutation hotspots listed above, the profile of MTX resistance becomes a gray area. For example, mutations in Ile 7 are known to pair with hotspot mutations at positions F31, F34, and D35 to produce resistant mutants.[40,41] Of our 75 predicted resistant sequences, 24 contain a mutation in residue 7. This may be a true positive, or

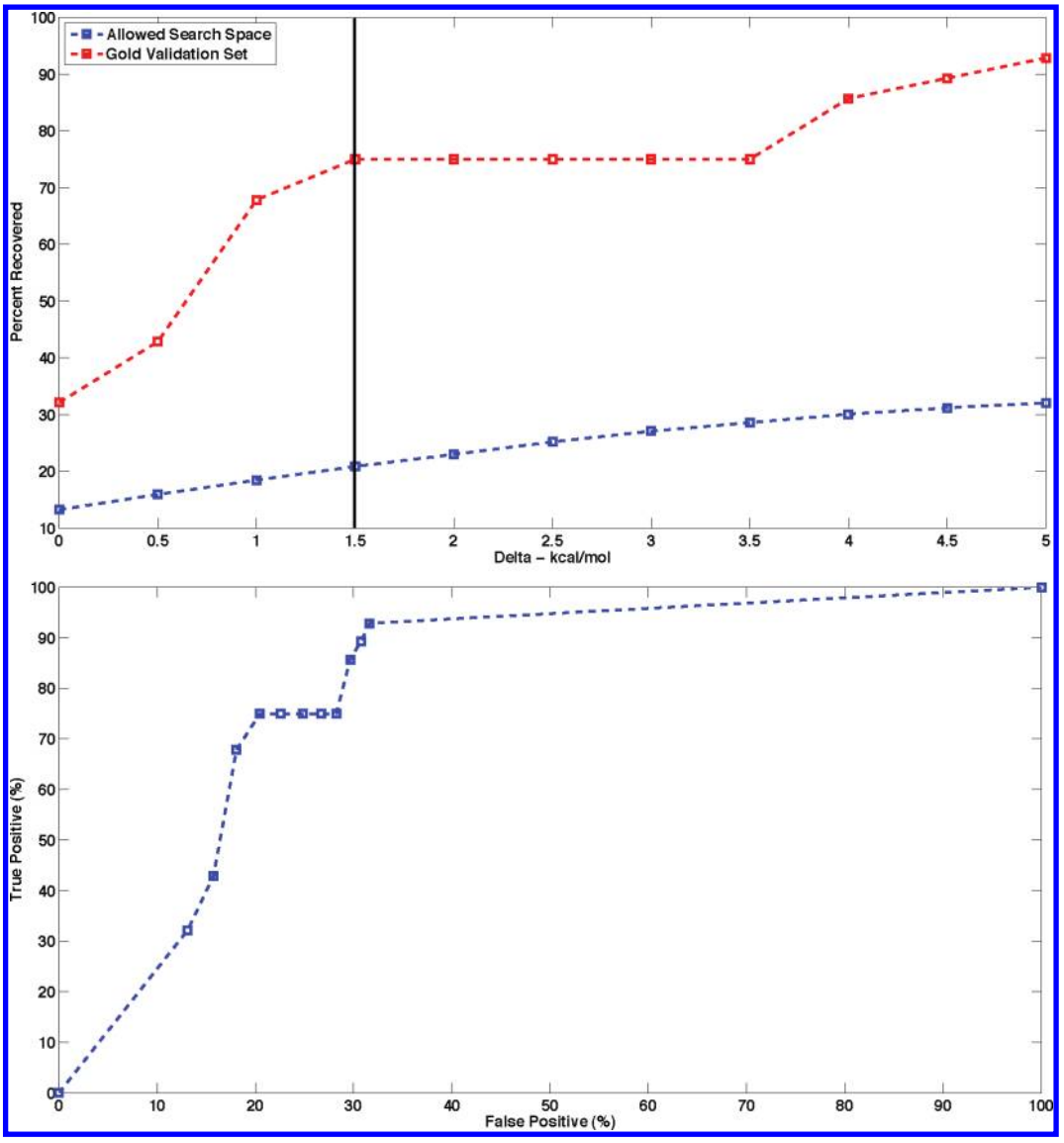


Figure 5. Retrieval of HIV mutants. (Top): Percent of retrieved known mutants from Gold Validation set (red curve) as well as all the mutants included in the search space (blue curve). The *x*-axis represents the change in native substrate binding energy of the mutant compared to the wild type. Also shown is the native substrate pass threshold, set to 1.5 kcal/mol from the wild type (vertical black bar). A higher *x* value indicates a greater loss of binding compared to the wild type. The value at $x = 0$ indicates that these sequences were predicted to have a higher than wild-type affinity for the native substrate in the substrate pass. (Bottom): Percent of true positives (i.e., known mutants from Gold Validation set) is drawn as a function of false positives (i.e., all other mutants from search space).

may simply be noise. In summary, our model was able to identify an enriched set of resistance mutations corresponding to known resistance hotspots in the methotrexate−DHFR system.

Gleevec Resistance. In the treatment of chronic myelogenous leukemia (CML), the tyrosine kinase domain of the BCR-ABL fusion protein is inhibited by the highly selective chemotherapeutic drug gleevec. Point mutations in the kinase domain can cause gleevec resistance. The most well-known of these mutations is the T315I "gatekeeper mutation" that confers resistance to both gleevec and all second generation tyrosine kinase inhibitors. We conducted a search of single-point mutations among the 14 active site residues (see Methods). As in the previous cases, we allowed residues to mutate to the following nine hydrophobic amino acid types: Gly, Ala, Leu, Ile, Val, Phe, Tyr, Trp, and Met.

Unlike the previous three systems, there is no experimentally determined structure of the protein bound to its native substrate. Therefore, the gleevec system serves as a test of the feasibility of our method to identify potential resistance mutations in the not uncommon case where a structure of the native substrate protein complex is unavailable. Our approach for Stage 1 was to replace the use of the native substrate complex with the unbound apo protein. This approach is somewhat conservative. It will only identify mutation sequences that negatively affect drug binding without affecting the inherent structure of the active site. There is no guarantee that the identified mutation sequences will maintain native binding. The list of predicted resistant sequences is therefore likely to contain an increased number of false positives, but the sensitivity for identifying resistant sequences should be approximately the same. In the Stage 1 native substrate pass, a gap free list of mutation sequences with

dx.doi.org/10.1021/ci200626m | J. Chem. Inf. Model. 2012, 52, 1529−1541

energies no worse than 1.5 kcal/mol of the wild-type apo protein is identified. In the Stage 1 drug pass and Stage 2, the ΔΔ score is approximated as $E_{mut,d} - E_{wt,d}$, where a positive score indicates a mutant with weaker affinity for the drug. This is the approach taken in the remainder of this section.

There are 117 sequences considered in the mutation search. A total of 32 sequences passed the apo protein stage of the rDEE search (i.e., by not affecting the energy of the apo protein by more than 1.5 kcal/mol). Surviving sequences were evaluated for gleevec binding, and a total of 19 sequences had a positive $\Delta\Delta_{rdee}$ score. Of these 19 mutants, 13 had a positive ($E_{mut,d} - E_{wt,d}$) score (Stage 2) and were categorized as our predicted resistant sequences (Table 3).

Table 3. Predicted Resistance for Gleevec-ABL Kinase[a]

Mutation	Comments
T315I	Known gleevec resistance mutation
T315V, T315M	Experimentally confirmed to confer resistance in in vitro studies[71,72]
Y253L, Y253M	Plausible, Y253F causes decreased susceptibility to gleevec
M290L, E286LMY	Unlikely, possible ATP binding site[71]
V256L, L370M, H361M, F382W	Unlikely, residue function is not well-known

[a]All 13 single mutants predicted resistant by rDEE and MM-PBSA are given. The clinically well-known T315I gatekeeper mutation is predicted to confer resistance to gleevec by our approach. Two of the predicted mutants are known to be resistant in vitro and an additional two are highly likely due to their similarity to known mutants.

The set of predicted resistant sequences correctly includes the gatekeeper mutation T315I. In addition, the set also contains two sequences, T315V and T315M, reported to confer gleevec resistance in in vitro studies.[71,72] The set also contains four mutations at positions 290 and 286, namely, M290L and E286LMY. There is evidence to suggest that both M290 and E286 are possibly involved in ATP binding; therefore, mutations in these residues, while impairing gleevec binding, can also result in an inactive kinase.[71] This prediction is understandable as our search did not have access to the native substrate complex and therefore only considered gleevec binding.

The T315I gatekeeper mutation is the dominant mechanism of tyrosine kinase resistance; however at least one drug, ponatinib, has been shown to overcome this mutation.[73] As a short second experiment, we performed a single-point mutation resistance analysis for ponatinib. Similar to experimental studies,[73] the following nine residues were modeled as flexible: Tyr 253, Glu 286, Thr 315, Phe 317, Met 318, Ile 360, His 361, Asp 381, and Phe 382 (PDB ID: 3OXZ). Each was allowed to mutate to the following set of nine hydrophobic residues: Ala, Phe, Gly, Ile, Leu, Met, Val, Trp, and Tyr. As in the gleevec experiment, a native substrate bound structure of tyrosine kinase was not available; therefore, we predicted resistance using the same modified scoring procedure we utilized for gleevec. Our approach predicts that T315I is indeed sensitive to ponatinib. A total of six single-point mutants were categorized as resistant to ponatinib and included F317LM, D381I, F382ILM. In summary, despite not having an experimental structure of the protein bound to its native substrate, our method was able to identify several known resistance mutations for gleevec, including the T315I gatekeeper mutation. Furthermore, in additional experiments with ponatinib, our method correctly identified T315I as ponatinib sensitive.

CONCLUSION

We have introduced a two-stage structure-based search and scoring procedure for identifying resistance conferring mutations. Our technique pairs an efficient restricted Dead-End Elimination-based search with the more accurate MM-PBSA scoring method. Positive design ensures that candidate mutations maintain the protein's native function, while negative design identifies mutations with significantly reduced affinity for the inhibitor. The output of our two-pass search is an enriched list of possible resistance mutations. It is more efficient to validate this "short list" than the brute-force approach of testing all possible active site mutations. It is our hope that the type of approach presented in this paper can provide a priori knowledge of drug resistance. The model can also be used to prioritize lead compounds and protein targets on the basis of the ease with which resistance can arise.

Computational methods, like ours, are most effective if they can maximize the number of true positives while minimizing the number of false positives. In our case, this corresponds to not predicting a resistant mutant as being sensitive. Of course, in the absence of exhaustive knowledge of experimentally verified true positives, it is difficult to completely assess performance. In our testing, we constructed a number of validation sets using known or likely mutation sequences. We demonstrated our technique using four well-known systems. In all four cases, our technique produced a set of predicted mutation sequences enriched with known resistant sequences. Our method was successful in both single-point and double-point mutation searches. Three of the four systems utilized experimentally determined structures of the protein in complex with the native substrate as well as with the drug. One system did not have an available experimental structure of the native substrate protein complex. For this system, we replaced the missing complex with the structure of the apo protein and approximated the ΔΔ score. The fact that we were still able to recover well-known resistance mutations suggests that in some cases resistant mutants can be predicted despite having only partial structural information.

Our technique is one step toward a general purpose computational tool. As such, it is not without limitations. Structural models employing approximations of interaction energies should not be interpreted as ground truth. Computational models are generally most useful when used in conjunction with wet lab testing. There are three primary directions for future work. First, we want to refine our ability to gracefully handle the situation where, as in our tyrosine kinase example, an experimental structure of the native substrate protein complex is not available. Second, the ability to model larger scale and allosteric type conformational changes could allow identification of mutations beyond the active site. Finally, it would be useful to close-the-loop by coupling our computational model with feedback from wet lab testing of the predicted mutation sequences.

In conclusion, our results on four diverse drug-target systems indicate that structure-based methods like ours can be useful in identifying resistance mutations in drug targets. Because no prior knowledge of resistance is needed, our approach can be employed as a first step to probe resistance in systems where information regarding drug resistance is minimal or nonexistent.

■ ASSOCIATED CONTENT

Supporting Information

A complete list of mutations for both ritonavir-HIV protease and methotrexate-dihydrofolate reductase. This information is available free of charge via the Internet at http://pubs.acs.org

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: lilien@cs.toronto.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Alan Moses, Dr. Ramgopal Mettu, and Mr. Izhar Wallach for helpful discussions and comments on drafts. This work is supported by a Bill and Melinda Gates (Grand Challenges Exploration) grant and an NSERC Discovery grant to R.H.L.

■ REFERENCES

(1) Blanchard, J. S. Molecular mechanisms of drug resistance in mycobacterium tuberculosis. *Annu. Rev. Biochem.* **1996**, *65*, 215−239.

(2) Borst, P. Genetic mechanisms of drug resistance. A review. *Acta Oncol.* **1991**, *30*, 87−105.

(3) Erickson, J. W.; Burt, S. K. Structural mechanisms of HIV drug resistance. *Annu. Rev. Pharmacol. Toxicol.* **1996**, *36*, 545−571.

(4) Huggins, D. J.; Sherman, W.; Tidor, B. Rational approaches to improving selectivity in drug design. *J. Med. Chem.* **2012**, *55*, 1424−1444.

(5) Noble, M.; Endicott, J.; Johnson, L. Protein kinase inhibitors: insights into drug design from structure. *Science* **2004**, *303*, 1800−1805.

(6) Ohtaka, H.; Muzammil, S.; Schön, A.; Velaquez-Campoy, A.; Vega, S.; Freire, E. Thermodynamic rules for the design of high affinity HIV-1 protease inhibitors with adaptability to mutations and high selectivity towards unwanted targets. *Int. J. Biochem. Cell Biol.* **2002**, *36*, 1787−1799.

(7) Pastor, M.; Cruciani, G. A novel strategy for improving ligand selectivity in receptor-based drug design. *J. Med. Chem.* **1995**, *38*, 4637−4647.

(8) Altman, M.; Ali, A.; Reddy, G.; Nalam, M.; Anjum, S.; Cao, H.; Chellappan, S.; Kairys, V.; Fernandes, M.; Gilson, M.; Schiffer, C.; Rana, T.; 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.

(9) Eboumbou Moukoko, E. C.; Bogreau, H.; Briolant, S.; Pradines, B.; Rogier, C. Molecular markers of plasmodium falciparum drug resistance. *Med. Trop (Mars)* **2009**, *69*, 606−612.

(10) Operario, D. J.; Moser, M. J.; St George, K. Highly sensitive and quantitative detection of the H274Y oseltamivir resistance mutation in seasonal A/H1N1 influenza. *J. Clin. Microbiol.* **2010**, 3517−3524.

(11) Van Laethem, K.; Schrooten, Y.; Lemey, P.; Van Wijngaerden, E.; De Wit, S.; Van Ranst, M.; Vandamme, A. A genotypic resistance assay for the detection of drug resistance in the human immunodeficiency virus type 1 envelope gene. *J. Virol. Methods* **2005**, *123*, 25−34.

(12) Buendia, P.; Cadwallader, B.; DeGruttola, V. A phylogenetic and Markov model approach for the reconstruction of mutational pathways of drug resistance. *Bioinformatics* **2009**, *25*, 2522−2529.

(13) Chen, B. J.; Causton, H. C.; Mancenido, D.; Goddard, N. L.; Perlstein, E. O.; Pe'er, D. Harnessing gene expression to identify the genetic basis of drug resistance. *Mol. Syst. Biol.* **2009**, *5*, 310.

(14) Fjell, C. D.; Jenssen, H.; Hilpert, K.; Cheung, W. A.; Pante, N.; Hancock, R. E.; Cherkasov, A. Identification of novel antibacterial peptides by chemoinformatics and machine learning. *J. Med. Chem.* **2009**, *52*, 2006−2015.

(15) Heider, D.; Verheyen, J.; Hoffmann, D. Predicting bevirimat resistance of HIV-1 from genotype. *BMC Bioinformatics* **2010**, *11*, 37.

(16) Pasomsub, E.; Sukasem, C.; Sungkanuparph, S.; Kijsirikul, B.; Chantratita, W. The application of artificial neural networks for phenotypic drug resistance prediction: Evaluation and comparison with other interpretation systems. *Jpn. J. Infect. Dis.* **2010**, *63*, 87−94.

(17) Zhang, J.; Hou, T.; 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.

(18) Dixit, A.; Torkamani, A.; Schork, N. J.; Verkhivker, G. Computational modeling of structurally conserved cancer mutations in the RET and MET kinases: The impact on protein structure, dynamics, and stability. *Biophys. J.* **2009**, *96*, 858−874.

(19) Frieboes, H.; Edgerton, M. E.; Fruehauf, J.; Rose, F. R.; Worrall, L. K.; Gatenby, R. A.; Ferrari, M.; Cristini, V. Prediction of drug response in breast cancer using integrative experimental/computational modeling. *Cancer Res.* **2009**, *69*, 4484−4492.

(20) Lapins, M.; Wikberg, J. E. Proteochemometric modeling of drug resistance over the mutational space for multiple HIV protease variants and multiple protease inhibitors. *J. Chem. Inf. Model.* **2009**, *49*, 1202−1210.

(21) Pricl, S.; Fermeglia, M.; Ferrone, M.; Tamborini, E. T315I-mutated Bcr-Abl in chronic myeloid leukemia and imatinib: insights from a computational study. *Mol. Cancer Ther.* **2005**, *4*, 1167−1174.

(22) Velazquez-Campoy, A.; Muzammil, S.; Ohtaka, H.; Schön, A.; Vega, S.; Freire, E. Structural and thermodynamic basis of resistance to HIV-1 protease inhibition: implications for inhibitor design. *Curr. Drug Targets Infect. Disord.* **2003**, *3*, 311−328.

(23) Wahab, H. A.; Choong, Y. S.; Ibrahim, P.; Sadikun, A.; Scior, T. Elucidating isoniazid resistance using molecular modeling. *J. Chem. Inf. Model.* **2009**, *49*, 97−107.

(24) Zhu, X. L.; Ge-Fei, H.; Zhan, C. G.; Yang, G. F. Computational simulations of the interactions between acetyl-coenzyme-A carboxylase and clodinafop: Resistance mechanism due to active and nonactive site mutations. *J. Chem. Inf. Model.* **2009**, *49*, 1936−1943.

(25) Chen, Y. Z.; Gu, X. L.; Cao, Z. W. Can an optimization/scoring procedure in ligand−protein docking be employed to probe drug-resistant mutations in proteins? *J. Mol. Graph. Model.* **2001**, *19*, 560−570.

(26) Lilien, R.; Stevens, B.; Anderson, A.; Donald, B. A novel ensemble-based scoring and search algorithm for protein redesign, and its application to modify the substrate specificity of the gramicidin synthetase A phenylalanine adenylation enzyme. *J. Comput. Biol.* **2005**, *12*, 740−761.

(27) Frey, K. M.; Georgiev, I.; Donald, B. R.; Anderson, A. C. Predicting resistance mutations using protein design algorithms. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 13707−13712.

(28) Safi, M.; Lilien, R. H. Restricted dead-end elimination: Protein redesign with a bounded number of residue mutations. *J. Comput. Chem.* **2010**, *31*, 1207−1215.

(29) Desmet, J.; Maeyer, M.; Hazes, B.; Lasters, I. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* **1992**, *356*, 539−542.

(30) Gielens, C.; Idakieva, K.; De Maeyer, M.; Van de Bergh, V.; Siddiqui, N. I.; Compernolle, F. Conformational stabilization at the active site of molluskan (*Rapana thomasiana*) hemocyanin by a cysteine-histidine thioether bridge A study by mass spectrometry and molecular modeling. *Peptides* **2007**, *28*, 790−797.

(31) Looger, L.; Dwyer, M.; Smith, J.; Hellinga, H. Computational design of receptor and sensor proteins with novel functions. *Nature* **2003**, *423*, 185−190.

(32) Maglia, G.; Jonckheer, A.; De Maeyer, M.; J.M., F.; Engelborghs, Y. An unusual red-edge excitation and time-dependent Stokes shift in the single tryptophan mutant protein DD-carboxypeptidase from streptomyces: The role of dynamics and tryptophan rotamers. *Protein Sci.* **2008**, *17*, 352−361.

(33) Novoa de Armas, H.; Dewilde, M.; Verbeke, K.; De Maeyer, M.; Declerck, P. J. Study of recombinant antibody fragments and PAI-1

complexes combining protein−protein docking and results from site-directed mutagenesis. *Structure* **2007**, *15*, 1105−1116.
(34) Case, D. A.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M. J.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The AMBER biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(35) Pearlman, D.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E. I.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(36) Lovell, S.; Word, J.; Richardson, J.; Richardson, D. The penultimate rotamer library. *Proteins* **2000**, *40*, 389−408.
(37) Cowan-Jacob, S. W.; Fendrich, S.; Floersheimer, A.; Furet, P.; Liebetanz, J.; Rummel, G.; Rheinberger, P.; Centeleghe, M.; Fabbro, D.; Manley, P. W. Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63*, 80−93.
(38) Dias, M. V.; Vasconcelos, I. B.; Prado, A. M.; Fadel, V.; Basso, L. A.; de Azevedo, W. F.; Santos, D. S. Crystallographic studies on the binding of isonicotinyl-NAD adduct to wild-type and isoniazid resistant 2-trans-enoyl-ACP (CoA) reductase from mycobacterium tuberculosis. *J. Struct. Biol.* **2007**, *159*, 369−380.
(39) Golovin, A.; Henrick, K. MSDmotif: Exploring protein sites an motifs. *BMC Bioinf.* **2008**, *9*, 312.
(40) Schweitzer, B. I.; Srimatkandada, S.; Gritsman, H.; Sheridan, R.; Venkataraghavan, R.; Bertino, J. R. Probing the role of two hydrophobic active site residues in the human dihydrofolate reductase by site-directed mutagenesis. *J. Biol. Chem.* **1989**, *264*, 20786−20795.
(41) Volpato, J. P.; Yachnin, B. J.; Blanchet, J.; Guerrero, V.; Poulin, L.; Fossati, E.; Berghuis, A. M.; Pelletier, J. Multiple conformers in active site of human dihydrofolate reductase F31R/Q35E double mutant suggest structural basis for methotrexate resistance. *J. Biol. Chem.* **2009**, *284*, 20079−20089.
(42) Protein Data Bank Europe. *PDBeMotif*. http://www.ebi.ac.uk/pdbe-site/pdbemotif/ (accessed September 10, 2011).
(43) Pearlman, D. A. Evaluating the molecular mechanics poisson-boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase. *J. Med. Chem.* **2005**, *48*, 7796−7807.
(44) Ferrari, A.; Degliesposti, G.; Sgobba, M.; Rastelli, G. Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. *Bioorg. Med. Chem.* **2007**, *15*, 7865−7877.
(45) Guimaraes, C.; Cardozo, M. MMGB/SA rescoring of docking poses in structure based lead optimization. *J. Chem. Inf. Model.* **2008**, *48*, 958−970.
(46) Raju, R.; Burton, N.; Hillier, I. Modeling the binding of HIV-reverse transcriptase and nevirapine: An assessment of quantum mechanical and force field approaches and predictions of the effect of mutations on binding. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7117−7125.
(47) Wang, J.; Hou, T.; Xu, X. Recent advances in free energy calculations with a combination of molecular mechanics and continuum models. *Curr. Comput-Aid. Drug.* **2006**, *2*, 95−103.
(48) 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*, 69−82.
(49) Cheng, A.; Eksterowicz, J.; Geuns-Meyer, S.; Sun, Y. Analysis of kinase inhibitor selectivity using a thermodynamics-based partition index. *J. Med. Chem.* **2010**, *53*, 4502−4510.
(50) Kangas, E.; Tidor, B. Electrostatic specificity in molecular ligand design. *J. Chem. Phys.* **2000**, *112*, 9120−9132.
(51) Sherman, W.; Tidor, B. Novel method for probing the specificity binding profile of ligands: Applications to HIV protease. *Chem. Biol. Drug. Des.* **2008**, *71*, 387−407.
(52) Altman, M.; Nalivaika, E.; Prabu-Jeyabalan, M.; Schiffer, C.; Tidor, B. Computational design and experimental study of tighter binding peptides to an inactivated mutant of HIV-1 protease. *Proteins* **2007**, *70*, 678−694.
(53) King, N. M.; Prabu-Jeyabalan, M.; Nalivaika, E.; Wigerinck, P.; Bethune, M.; Schiffer, C. Structural and thermodynamic basis for the binding of TMC114, a next-generation human immunodeficiency virus type 1 protease inhibitor. *J. Virol.* **2004**, *78*, 12012−12021.
(54) Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J. Mol. Biol.* **2000**, *301*, 1207−1220.
(55) Weber, I. T.; Miller, M.; Jaskólski, M.; Leis, J.; Skalka, A. M.; Wlodawer, A. Molecular modeling of the HIV-1 protease and its substrate binding site. *Science* **1989**, *243*, 928−931.
(56) Weiner, S.; Kollman, P.; Case, D.; Singh, U.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.
(57) Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(58) Parikh, S.; Moynihan, D.; Xiao, G.; Tonge, P. The role of Tyrosine 158 and Lysine 165 in the catalytic mechanism of InhA, the enoyl-ACP reductase from mycobacterium tuberculosis. *Biochemistry* **1999**, *38*, 13623−13634.
(59) Pantano, S.; Alber, F.; Lamba, D.; Carloni, P. NADH interactions with WT- and S94A-acyl carrier protein reductase from mycobacterium tuberculosis: An ab initio study. *Proteins* **2002**, *47*, 62−68.
(60) Rhee, S.; Gonzales, M.; Kantor, R.; Betts, B.; Ravela, J.; Shafer, R. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* **2003**, *31*, 298−303.
(61) Shafer, R. Rationale and uses of a public HIV drug-resistance database. *J. Infect. Dis* **2006**, *194*, S51−S58.
(62) Brenner, B.; Wainberg, M.; Salomon, H.; Rouleau, D.; Dascal, A.; Spira, B.; Sekaly, R.; Conway, B.; Routy, J. Resistance to antiretroviral drugs in patients with primary HIV-I infection. Investigators of the Quebec Primary Infection Study. *Int. J. Antimicrob. Agents* **2000**, *16*, 429−434.
(63) Rhee, S.; Taylor, J.; Fessel, W.; Kaufman, D.; Towner, W.; Troia, P.; Ruane, P.; Hellinger, J.; Shrivani, V.; Zolopa, A.; Shafer, R. HIV-1 protease mutations and protease inhibitor cross-resistance. *Antimicrob. Agents Chemother.* **2010**, *54*, 4253−4261.
(64) Stoffler, D.; Sanner, M.; Morris, G.; Olson, A.; Goodsell, D. Evolutionary analysis of HIV-1 protease inhibitors: Methods for design of inhibitors that evade resistance. *Proteins* **2002**, *48*, 63−74.
(65) Wang, X.; Tong, X.; Tang, H.; Liu, P.; Zhang, W.; Yang, R. Study on genotypic resistance mutations to antiretroviral drugs on HIV strains of treated and treatment-naive HIV-1 infectious patients in Hubei province. *Zhonghua Liu Xing Bing Xue Za Zhi* **2007**, *11*, 1112−1115.
(66) Boden, D.; Markowitz, M. Resistance to human immunodeficiency virus type 1 protease inhibitors. *Antimicrob. Agents Chemother.* **1998**, *42*, 2775−2783.
(67) 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*, 1177−1188.
(68) Ercikan-Abali, E. A.; Mineishi, S.; Tong, Y. Active site-directed double mutants of dihydrofolate reductase. *Cancer Res.* **1996**, *56*, 4142−4145.
(69) Fossati, E.; Volpato, J.; Poulin, L.; Guerrero, V.; Dugas, D.; Pelletier, J. 2-Tier bacterial and in vitro selection of active and methotrexate-resistant variants of human dihydrofolate reductase. *J. Biomol. Screen.* **2008**, *13*, 504−514.
(70) Volpato, J.; Fossati, E.; Pelletier, J. Increasing methotrexate resistance by combination of active-site mutations in human dihydrofolate reductase. *J. Mol. Biol.* **2007**, *373*, 599−611.

(71) Corbin, S.; Buchdunger, E.; Pascal, F.; Druker, B. Analysis of the structural basis of specificity of inhibition of the Abl kinase by STI571. *J. Biol. Chem.* **2002**, *277*, 32214−32219.
(72) Detection of gleevec resistant mutations, U.S. Patent 7326534, February 5, 2008. http://www.patents.com/us-7416873.html.
(73) Zhou, T.; Commodore, L.; Huang, W.; Wang, Y.; Thomas, M.; Keats, J.; Xu, Q.; Rivera, V.; Shakespeare, W.; Clackson, T.; Dalgarno, D.; Zhu, X. Structural mechanism of the Pan-BCR-ABL inhibitor ponatinib (AP24534): Lessons for overcoming kinase inhibitor resistance. *Chem. Biol. Drug. Des.* **2011**, *77*, 1−11.