Antiviral Therapy **10**:157–166

Prediction of HIV-1 protease inhibitor resistance using a protein–inhibitor flexible docking approach

*Ekachai Jenwitheesuk and Ram Samudrala**

Computational Genomics Group, Department of Microbiology, University of Washington School of Medicine, Seattle, WA, USA

*Corresponding author: Tel: +1 206 732 6122; Fax: +1 206 732 6055; E-mail: ram@compbio.washington.edu

Emergence of drug resistance remains one of the most challenging issues in the treatment of HIV-1 infection. Here we focus on resistance to HIV-1 protease inhibitors (PIs) at a molecular level, which can be analysed genotypically or phenotypically. Genotypic assays are based on the analysis of mutations associated with reduced drug susceptibility, but are problematic because of the numerous mutations and mutational patterns that confer drug resistance. Phenotypic resistance or susceptibility can be experimentally evaluated by measuring the amount of free drug bound to HIV-1 protease molecules, but this procedure is expensive and time-consuming. To overcome these problems, we have developed a docking protocol that takes protein–inhibitor flexibility into account to predict phenotypic drug resistance. For six FDA-approved PIs and a total of 1792 HIV-1 protease sequence mutants, we used a combination of inhibitor flexible docking and molecular dynamics (MD) simulations to calculate protein–inhibitor binding energies. Prediction results were expressed as fold changes of the calculated inhibitory constant (K_i), and the samples predicted to have fold-increase in calculated K_i above the fixed cut-off were defined as drug resistant. Our combined docking and MD protocol achieved accuracies ranging from 72–83% in predicting resistance/susceptibility for five of the six drugs evaluated. Evaluating the method only on samples where our predictions concurred with established knowledge-based methods resulted in increased accuracies of 83–94% for the six drugs. The results suggest that a physics-based approach, which is readily applicable to any novel PI and/or mutant, can be used judiciously with knowledge-based approaches that require experimental training data, to devise accurate models of HIV-1 PI resistance prediction.

Introduction

Human immunodeficiency virus type 1 (HIV-1) drug-resistant mutations that arise during treatment have significantly affected patient management and the long-term effectiveness of antiretroviral therapy [1,2]. Evaluating the effects of these mutations has become an important factor in developing treatment strategies for infected patients [3,4]. Several methodologies have been developed to either phenotypically or genotypically determine HIV-1 drug susceptibility [5–10].

Genotypic testing is relatively rapid and particularly useful when a strong correlation between specific mutations and drug resistance exists [11,12]. However, the interpretation of genotypic information is difficult when complex mutational patterns and large numbers of polymorphisms interact to cause resistance, cross-resistance or resistance reversal [12–15].

To predict the resistant behaviour of HIV-1 from genotypic data, several groups have developed computational tools based on molecular dynamics (MD) simulations [16,17], neural networks [18,19], linear regression [20], decision trees [21,22], support vector machines (SVM) [22,23], molecular docking [24–27] and rule-based interpretation systems [28]. Among these tools, the physics-based approach, that is, docking and MD simulations that evaluate the energy of protein–inhibitor complexes, is one of the techniques commonly used to study the interaction of HIV-1 protease and its inhibitors at the atomic level. The binding energies produced by this approach have successfully identified key mutations in parts of the enzyme that confer drug resistance and show high correlation with the inhibitory concentration (IC_{50}) values determined by the phenotypic susceptibility tests [16–17,24–27].

In this study, we used protein–inhibitor docking with a molecular dynamics protocol that takes protein–inhibitor flexibility into account [29] to determine the correlation between the experimentally determined inhibitory concentration (phenotype) and the computer calculated protease inhibitor binding

affinities based on the HIV-1 protease gene mutations (genotype). We further compared our predictions with two other established HIV-1 genotypic interpretation systems: a rule-based method (Stanford HIV drug resistance database; http://hivdb.stanford.edu/) and an SVM-based method (geno2pheno; http://195.37.60.133/cgi-bin/geno2pheno.pl/). Finally, we analysed the advantage of generating consensus results from all these methods to achieve better prediction accuracy.

Materials and methods

Dataset

A total of 1792 HIV-1 protease sequences and their corresponding IC_{50} values were retrieved from the Stanford HIV drug resistance database. We selected the data for which genotypic and phenotypic evaluations were performed by ViroLogic, Inc., using population sequencing and PhenoSense HIV assays, respectively. The phenotypic results were categorized as resistant or susceptible based on the fold change cut-off values recommended by the manufacturers (10-fold for lopinavir and 2.5-fold for other drugs).

Generation of protease mutant three-dimensional structures

The X-ray crystal structures of the wild-type HIV-1 protease–inhibitor complexes used for the binding energy determination were obtained from the Protein Data Bank (PDB): 1HPV for amprenavir, 1HSG for indinavir, 1MUI for lopinavir, 1OHR for nelfinavir, 1HXW for ritonavir and 1HXB for saquinavir. The three-dimensional structures of the protease mutants were constructed according to protease gene mutation data using the wild-type X-ray crystallography structures as templates. The wild-type side chains were substituted with the mutant side chains based on the backbone-dependent side chain rotamer library and a linear repulsive steric energy term provided by SCWRL version 2.95 [30].

Molecular dynamics simulations

MD simulations of all the wild-type and the mutant structures were carried out using NAMD version 2.5b1 [31] with the X-PLOR force field [32]. The van der Waals, bond, angle, dihedral and improper dihedral parameters for all the inhibitors were adopted from the Hetero-compound Information Centre-Uppsala (HIC-Up) (http://xray.bmc.uu.se/hicup). Protein protonation states were modelled as in a related HIV-1 protease MD simulations study [33]. All protein residues were modelled in their charged state except for one of the two aspartic acid groups (Asp 25 and Asp 25′) in the active site since previous studies have shown that at least one of these two aspartic acids is protonated [34–36]. We used a protonated Asp 25′ and deprotonated Asp 25 for all HIV-1 protease–inhibitor complexes. The terminal residues of both monomers were also protonated (Pro 1, Pro 1′, Phe 99 and Phe 99′).

The complexes were immersed in a 20 Å radius sphere of TIP3–water using the program Solvate [37] to allow the protein–inhibitor complexes to relax in an aqueous environment. Initially, 100 steps of energy minimization of the protein–inhibitor–water system were performed using default conjugate gradient energy minimization parameters, followed by 0.1 picoseconds (ps) of MD simulations at 300 K with periodic boundary conditions, an atom-based shifted distance-dependent dielectric constant $\varepsilon = 4r$, a switch function on the van der Waals interaction and a time step of 1 femtosecond (fs). The non-bonded interaction list was updated every 20 time-steps. The van der Waals interactions were truncated at a distance of 12 Å. The structures at 0.1 ps were recorded and processed in the docking step.

Preparation of protease and inhibitor structures for docking

Preparation of protein and inhibitor structures was carried out using utility programs in AutoDock version 3.0.5 [38]. Each trajectory of the protein–inhibitor complex recorded from the MD simulations was first separated into two files, one containing the inhibitor and the other containing the protein. For the preparation of the protein structures, only polar hydrogens were added to the protein and AMBER united-atom partial charges were assigned using the Protonate utility from AMBER [39]. The solvation parameters were added to the protein coordinates using the Addsol utility (both utilities are supplied with the program package) [38]. We then generated three-dimensional affinity and electrostatic grid boxes that were big enough to cover the entire active site using the AutoGrid program, which utilizes Lennard-Jones 12-10 and 12-6 parameters. The number of grid points in x, y, z-axes were $60 \times 110 \times 60$ with each point separated by 0.375 Å. The water molecules solved by X-ray crystallography were retained in our model for the docking step.

Inhibitor structures were treated as all atom entities, for which all hydrogens were added to fill the empty valences, followed by the addition of the Kollman partial atomic charges. The root atom and the rotatable dihedrals in the inhibitors were defined using the AutoTors program. All the rotatable bonds were allowed to rotate freely. The partial charges of the non-polar hydrogens are added to the charge of the carbons they bond to, and then these hydrogens are deleted from the molecule. The atom type for the aromatic carbons was reassigned to be handled by the aromatic carbon grid map.

Determination of protease inhibitor binding energies by docking

Docking calculations were carried out using AutoDock version 3.0.5 with a Larmarckian genetic algorithm and a 'Solis & Wets' local search [38]. The number of docking runs was set to 100. The maximum number of energy evaluations before the LGA run termination was 1 500 000 and the maximum number of generations of the LGA run before termination was 27 000. Other docking parameters were set to the default values provided by the software. The details of the docking parameters used to set up the docking experiments are given in Table 1.

At the end of each binding energy calculation, docking solutions with inhibitor all-atom RMSDs within 1.0 Å of each other were clustered together and ranked according to the lowest energy representative. AutoDock generates three energy terms: intermolecular energy, internal energy of the ligand and torsional free energy. The final docked energy was calculated from the sum of the intermolecular energy and the internal energy of the ligand. The free energy of binding was calculated from the sum of the intermolecular and the torsional free energies, and consequently converted into an inhibitory constant (K_i) according to Hess's law. The lowest-energy solution was accepted as the calculated binding energy and its K_i value was used to define the binding affinity of the inhibitors.

To find the cut-off value for each inhibitor, the samples were divided into two groups: phenotypically susceptible and phenotypically resistant. We first measured the fold-changes in the calculated K_i relative to the wild-type for both groups separately for each inhibitor. The mean and the standard deviation of the fold-changes in the calculated K_i of the phenotypically susceptible group were calculated and the cut-off value was set at two standard deviations above the mean (Table 2). The docking results that had a fold-increase in the calculated K_i above this cut-off were defined as drug resistant.

Table 1. Details of the docking parameters used in this study

Parameter	Value
Quaternion	Random
Number of the initial dihedral angles	10
Initial value of the initial dihedral angles	Random
Translation step sizes	2.0
Orientation step sizes	50.0
Torsion step sizes	50.0
Number of individuals in the population for LGA job	50
Maximum number of energy evaluations before LGA run is terminated	1 500 000
Maximum number of generations of LGA run before it is terminated	27 000
Number of the best inhibitor conformations that survive into the next generation	1
Gene mutation rate	0.02
Crossover rate	0.8
Number of generations for picking worst conformations	10
Mean of Cauchy distribution for gene mutation	0.0
Variance of Cauchy distribution for gene mutation	1.0
Maximum number of local search iterations	300
Size of the local search space to sample (ρ)	1.0
Number of consecutive successes before changing ρ	4
Number of consecutive failures before changing ρ	4
Number of step size before the local search is terminated	0.01
Number of docking runs	100

Comparison of the accuracy in resistance prediction with other methods

We analysed the same set of protease sequences with two well-known resistance prediction servers: the Stanford HIV drug resistance database (rule-based) and geno2pheno (SVM). These methods match the mutation information taken from the query sequences with their phenotypic–genotypic resistance database. The resistance scores were assigned to each matched mutant sequence. The final resistance score was generated by adding together all the matched mutation resistance scores. A score of 30 for the Stanford HIV drug resistance database and a score of 3.5 for geno2pheno were used to define a sample as resistant or susceptible. If a prediction (by any method) of resistant or susceptible matched the experimental observation (based on the manufacturer's IC_{50} cutoffs given above), then it was counted as correct prediction.

Results and discussion

Correlation of experimentally determined and calculated binding energies

We have previously described the improvement of protein–inhibitor binding energy prediction by incorporating protein flexibility through the use of MD simulations and an inhibitor flexible docking technique (AutoDock) for 25 HIV-1 protease–inhibitor complexes [29]. The correlation coefficient of the experimentally determined and calculated binding energies for 25 complexes was 0.38 when the inhibitor flexible docking was utilized alone (Figure 1A). The correlations improved significantly when the MD simulation was performed (up to 10 ps) prior to docking and the highest correlation coefficient of 0.87 was obtained after 0.1 ps of dynamics simulations (Figure 1B). This result indicates that our protocol has high accuracy in binding energy prediction, which

Table 2. Cutoff values used to define a prediction as drug resistant or susceptible

	Amprenavir	Lopinavir	Indinavir	Nelfinavir	Ritonavir	Saquinavir
Cutoff						
Phenotyping	2.5	10	2.5	2.5	2.5	2.5
Docking (mean + 2SD)	1.11	1.20	1.01	1.10	1.06	1.56
Mean	1.01	0.90	0.91	0.84	0.94	1.22
Standard deviation (SD)	0.05	0.15	0.05	0.08	0.06	0.17

The cutoff values of docking with dynamics are calculated from the mean of the calculated K_i of the phenotypically susceptible samples and are set at two standard deviations above the mean. Predictions that have fold increase in calculated K_i above this cutoff are defined as drug resistant.

allows us to apply this technique to interpret HIV-1 genotypic susceptibility data.

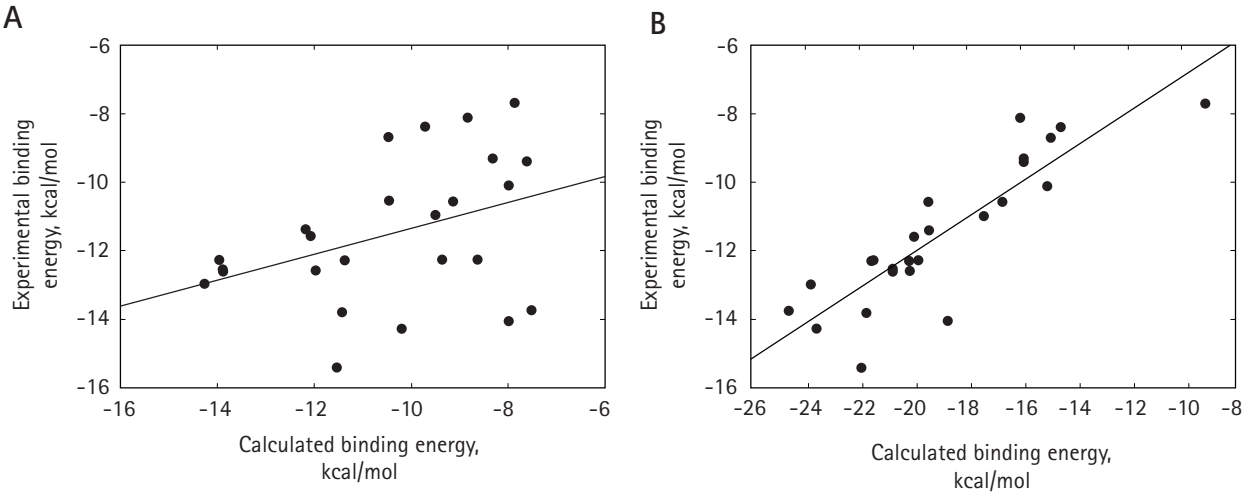
Accuracy of the docking with dynamics protocol

The overall accuracy of the docking with dynamics protocol was 64% for samples in the phenotypically resistant group and 83% for samples in phenotypically susceptible group. Specifically for each inhibitor, the combined resistance/susceptible prediction accuracies were 83% for amprenavir, 74% for lopinavir, 76% for indinavir, 58% for nelfinavir, 80% for ritonavir and 72% for saquinavir (Figure 2).

Although the protocol reported in 2003 by Shenderovich *et al.* [27] is similar to our own, the prediction accuracies are not comparable. There are two major enhancements in this work relative to that presented by Shenderovich *et al.* Firstly, we analysed 1792 HIV-1 protease sequences as opposed to 50 in [27]; our sample size is therefore larger and includes all reported resistant mutations and polymorphisms in the protease gene. Secondly, we included a protein–inhibitor relaxation feature in our protocol. The mutant structure used to estimate the binding energy was generated based on a graph-theory algorithm of SCWRL software followed by short energy minimization and 0.1 ps MD simulation. The trajectories after 0.1 ps MD simulation were recorded and used in the docking step. This protein–inhibitor relaxation feature significantly improves the accuracy of predicted HIV-1 protease inhibitor binding energies [29].

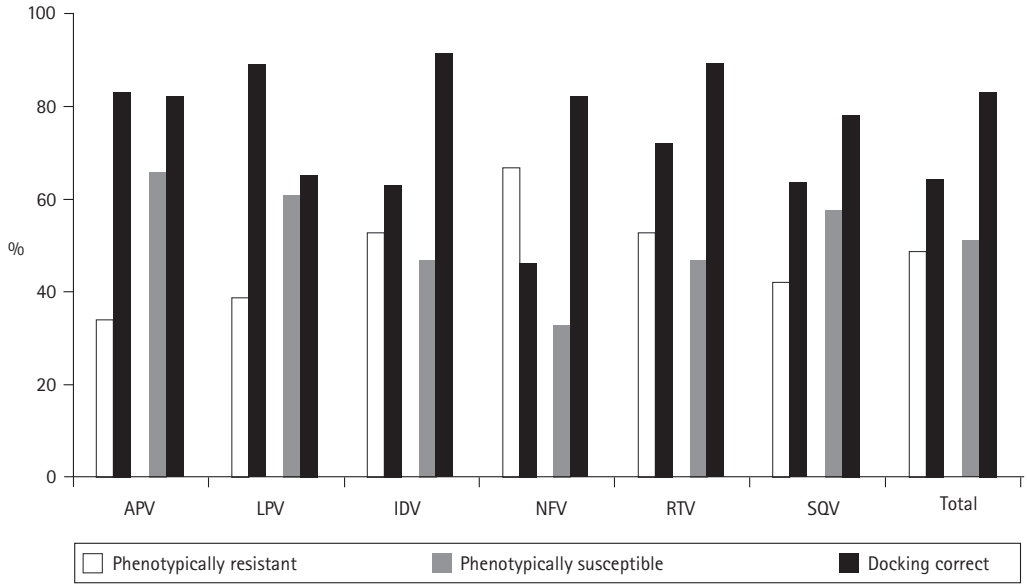
The mutant protease sequences used to generate the models contain a wide variety of polymorphisms and mutation patterns that do not match the sequences of the X-ray crystallography structure of the protease–inhibitor complexes in the PDB. However, we showed in our previous study that the structures obtained at the end of 0.1 ps of MD simulation were not significantly different from the original X-ray crystallography

Figure 1. Plots comparing experimentally determined and calculated binding energies for 25 HIV-1 protease–inhibitor complexes



Protein-rigid docking produced a poor correlation (0.38) between the experimental and calculated energies **(A)**, while the best correlations (0.87) were obtained after 0.1 ps MD simulations **(B)**.

Figure 2. Accuracy of docking with dynamics for drug-resistant and drug-susceptible mutations



The white and grey bars represent the fraction of drug-resistant and drug-susceptible mutations (determined by phenotypic susceptibility testing results). The black bar associated with each fraction bar represents the percentage correct prediction made by docking with dynamics for that set. On average, the accuracy is 64% for drug-resistant mutations and 83% for drug-susceptible mutations.

structures for 25 protease–inhibitor complexes (average all-atom RMSD was 0.35 Å) [29]. This indicates that the trajectories used to estimate the binding energy were not distorted by the simulation. Our protocol allows rearrangement of the side chain especially on the active site surface, which improves the interacting surface complementarities of the protein–inhibitor complex. The time scale of 0.1 ps is only sufficient to allow minor re-adjustment of side chains; larger structural changes caused by the mutations are not modelled.

The short MD simulation time (0.1 ps) did not effect movement of the main chain, although the flap region moved away from the binding pocket (average all-atom RMSD of the flap region was 0.54 Å) as the MD simulation time increased. This movement could affect rearrangement of binding pocket-lining side chains only. We assume that the main chains of the simulated mutant structures were not significantly different from those of the template wild-type X-ray crystal structures.

Accuracy of the consensus predictions generated from the three methods

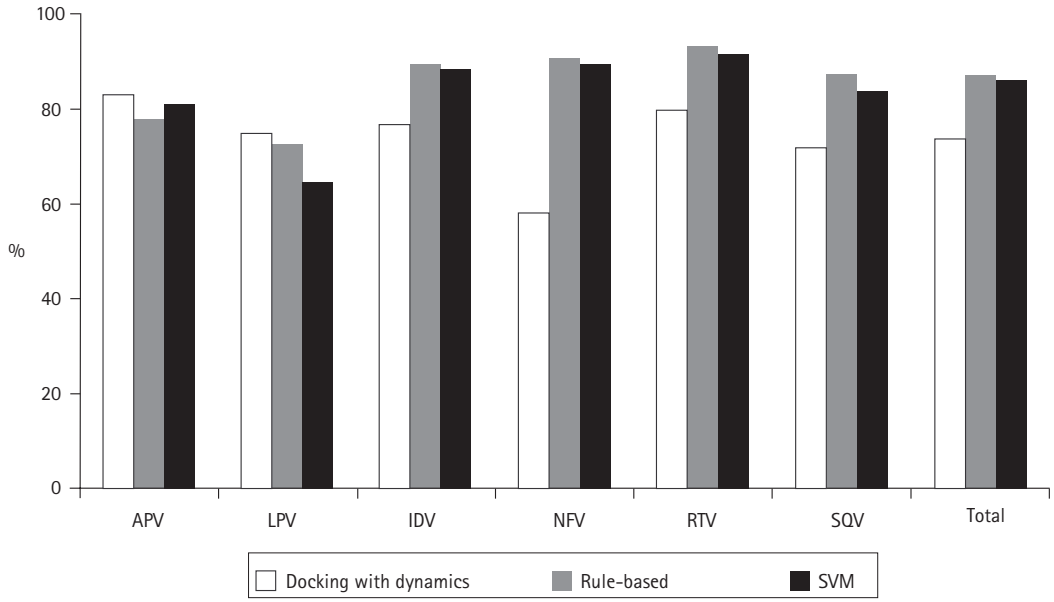
Figure 3 compares the accuracy of three prediction methods: docking with dynamics, rule-based and SVM. Overall, the accuracy was 74% for docking with dynamics, 87% for rule-based and 86% for SVM. Prediction results from the rule-based and SVM methods show higher accuracy for indinavir, nelfinavir, ritonavir and saquinavir compared with the accuracy of the docking with dynamics protocol. However, docking with dynamics modestly outperforms the other two methods for amprenavir and lopinavir.

Since the prediction results of the rule-based and the SVM had almost identical accuracy, we generated consensus predictions of the physics-based and machine-learning approaches. We expected that these two approaches would complement each other to improve the accuracy of HIV-1 genotypic susceptibility interpretation.

We generated four types of consensus predictions based on the concordant predictions from the docking with dynamics and rule-based methods (consensus 1), the docking with dynamics and SVM methods (consensus 2), the docking with dynamics method and either rule-based or SVM methods (consensus 3) and the concordant predictions from all three methods (consensus 4). On average, the docking with dynamics protocol had a 75% concordance with the other methods (that is, all the methods used in the consensus calculations produced the same result 75% of the time). The overall accuracy for these mutants was 90%. The concordance and the accuracy of each consensus prediction with respect to the inhibitors are shown in Figure 4.

The best consensus prediction type in terms of the percent concordance was between docking with dynamics and either rule-based or SVM (consensus 3). The percent concordance was 84% for amprenavir,

Figure 3. Comparison of the accuracy of three HIV-1 resistance prediction methods



The overall accuracy of our docking with dynamics protocol was 73%, the Stanford rule-based method was 87% and the geno2pheno SVM method was 86%. The accuracy of the rule-based and SVM methods are higher than the docking with dynamics protocol for indinavir, nelfinavir, ritonavir and saquinavir. However, our protocol outperforms the other two methods for resistance prediction of amprenavir and lopinavir, which are the two most recently approved HIV-1 protease inhibitors.

76% for lopinavir, 83% for indinavir, 64% for nelfinavir, 85% for ritonavir and 74% for saquinavir. The best consensus prediction type in terms of the prediction accuracy is one where all methods are included (consensus 4). The resistance/susceptibility prediction accuracies were 89% for amprenavir, 78% for lopinavir, 93% for indinavir, 86% for nelfinavir, 96% for ritonavir and 94% for saquinavir. These results indicate that the consensus predictions generated from different methods have higher overall accuracies than any of the methods considered individually.

In this study, there are only two key mutations, Asp30Asn and Gly48Val, which docking with dynamics always failed to identify as a cause of drug resistance. However, the rule-based and the SVM constantly determined Asp30Asn and Gly48Val mutations as nelfinavir and saquinavir resistance, respectively. Generating the consensus from the three methods would improve the accuracy of genotypic susceptibility interpretation for sequences containing these mutations.

This suggests that genotypic interpretations should not rely on a single system and that decisions about therapeutic regimens may be undertaken with higher confidence when consensus results are obtained. Discordant predictions indicate that expert interpretation that takes into account the mutation patterns and the treatment histories of the patient is necessary for proper diagnosis.

Analysis of resistant mutations in the context of docking with molecular dynamics simulations

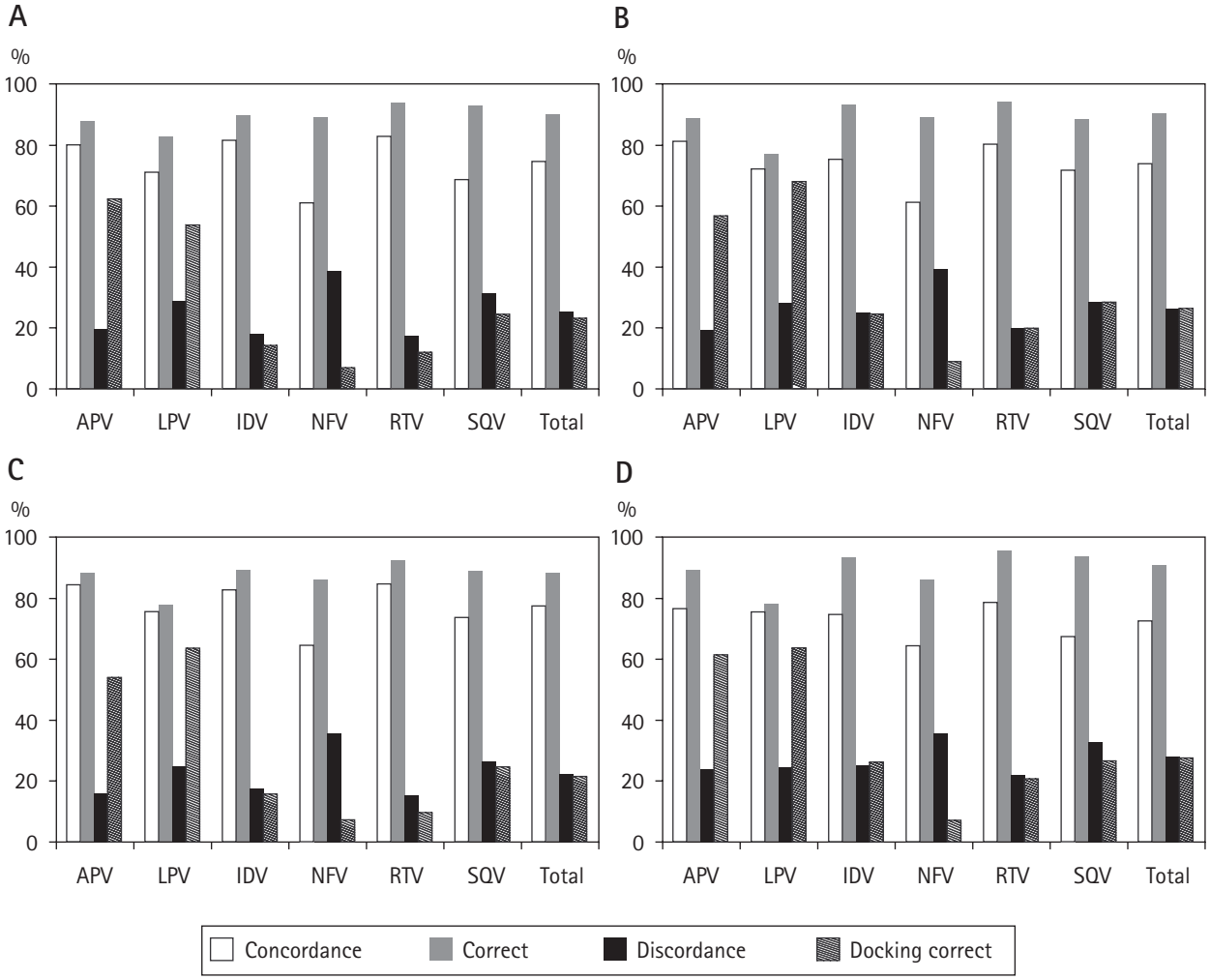
Analyses of the protease mutant structures showed that mutations around the binding pocket significantly reduce the binding affinity of the inhibitors to the mutant active site (Figure 5A,B). On the other hand, mutations that occur far away from the binding pocket do not directly affect the binding affinity but contribute to reducing the binding affinity in conjunction with the binding pocket mutations.

Residue 82 provides a major contribution to the binding pocket for the inhibitor group at residues P1 and P1′, which are generally hydrophobic. Val82 is part of the hydrophobic S1/S1′ pocket, which also contains residues Leu23, Gly27, Ile50, Pro81 and Ile84. Thus Val82/82′ helps stabilize the binding of hydrophobic residues at P1/P1′.

The docking with dynamics results show that a side-chain substitution at position 82 from $–CH_3$ of valine to –OH of threonine (Val82Thr), without any substitution at residue 84, increases sensitivity of the mutant to all drugs. In contrast to Val82Thr, the bulky side chains at position 82 and position 84 physically modify the conformation of the S1 and S1′ subsites of the active site and weaken the van der Waals interaction with the inhibitors.

In addition, the loss of the side chain $–CH\text{-}CH_3$ group in Val82Ala and $–CH_2$ in Ile84Val substitution by itself

Figure 4. Comparison of the percentage concordance and accuracy of four consensus prediction approaches



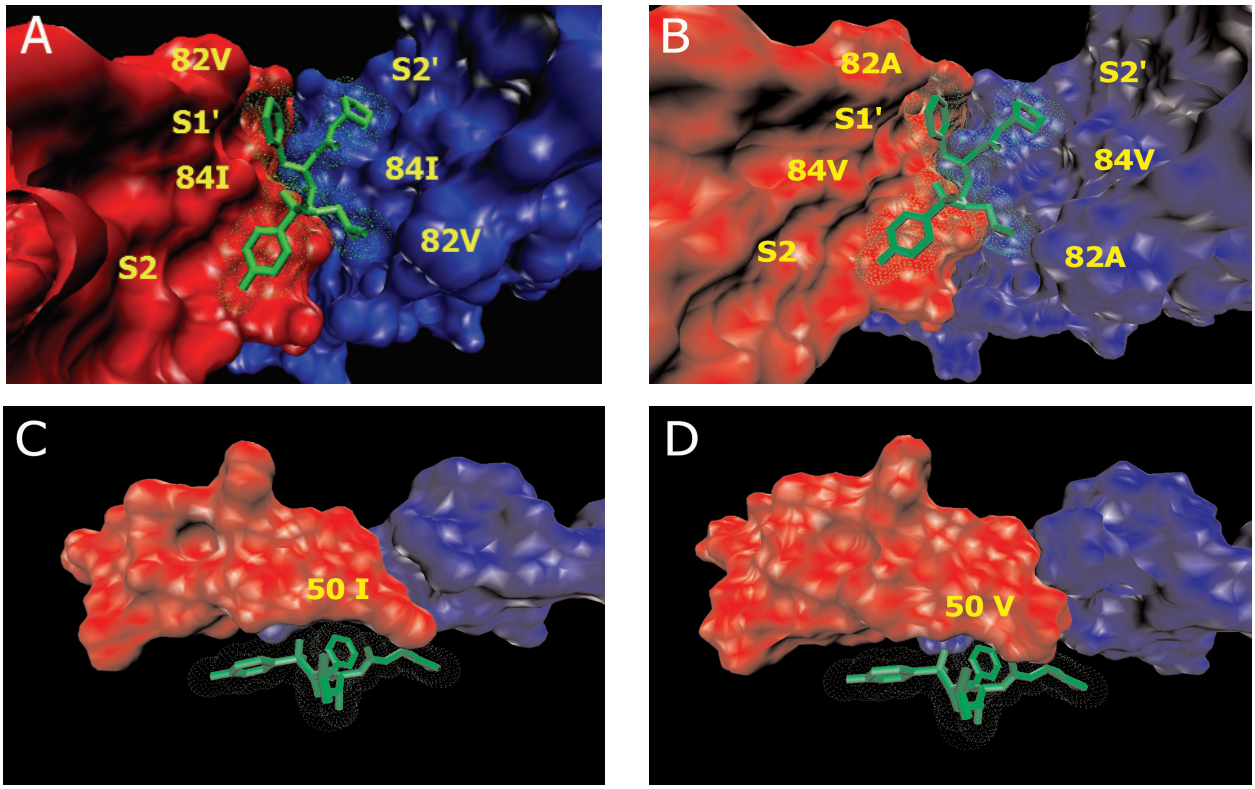
The white and grey bars represent percentage concordance among prediction methods (consensus) and the overall accuracy for the concordant predictions, respectively. The black and cross-hatched bars represent percent discordance among prediction methods and the accuracy of the docking with dynamics protocol for the discordant predictions, respectively. Each approach examines the concordant predictions between **(A)** docking and rule-based (consensus 1), **(B)** docking and SVM (consensus 2), **(C)** docking and either rule-based or SVM (consensus 3) and **(D)** docking, rule-based and SVM (consensus 4). The best accuracies are obtained when concordant predictions by all three methods are considered (consensus 4).

or present with other mutations at residues 82, 84 and/or 90 always increases the calculated K_i value of all drugs. The effect of these mutations on the reduction of the binding affinity for indinavir is explained as follows: the pyridyl and piperidine group of indinavir occupy the S3 and S1 subsites and the benzyl moiety occupies the S1′ subsite. The tertiary-butyl group and the indanol moiety occupy the S2 and S2′ subsites, respectively. The loss in binding affinity is mainly due to the loss of van der Waals contacts at residues 82 and 84 between the protease and indinavir as a result of these mutations.

The important amino acid substitutions that cause high-level amprenavir resistance were Ile50Val and a combination of Ile84Val + Leu90Met and Ile54Val + Val82Ala + Ile84Val + Leu90Met. All these mutations were also cross-resistant to indinavir, nelfinavir, ritonavir and saquinavir. The single mutations, Asp30Asn, Ile50Leu, Asn88Asp and Leu90Met, and double mutations Asp30Asn + Asn88Asp or Ile54Val/Met + Leu90Met did not display reduced affinity for amprenavir.

The docking with dynamics protocol shows that the Ile50Leu mutation did not affect the calculated K_i of amprenavir, indinavir or saquinavir. However, the mutation interferes with binding to nelfinavir and ritonavir and results in an increased calculated K_i value. The Ile50Val mutation reduced the hydrophobic contact between the side chain of valine residue 50 and the P2′ phenyl ring of the benzenesulphonamide

Figure 5. Comparison of the three-dimensional conformation of the binding site and flaps region of the HIV-1 protease wild-type and mutant



(A) The binding pocket of the protease wild-type. The side chain of isoleucine residue 84 protrudes inside the binding cavity, generating two small subsites (S1′ and S2). **(B)** The loss of the $-CH_2$ group when Ile84 mutates to Val and the loss of the $-CH-CH_3$ group when Val82 mutates to Ala; this creates a hole in the S1′ subsite of the enzyme and reduces the hydrophobic and van der Waals contacts with the inhibitor, resulting in weakened potency. **(C)** The flaps region (residue 45 to 55) of the protease wild-type bound with the amprenavir (shown in green). **(D)** The Ile50Val mutation that reduces the hydrophobic contact between the side chain of valine residue 50 and the phenyl ring of the benzenesulphonamide moiety of amprenavir. This mutation also increases the calculated inhibitory constant for nelfinavir and ritonavir.

moiety of amprenavir (Figure 5C,D). The results also show that all Ile50Val/Leu mutant structures are involved in increasing the calculated K_i for nelfinavir and ritonavir.

Our docking with dynamics protocol failed to identify Asp30Asn as a cause of nelfinavir resistance. Of the 38 Asp30Asn mutant structures evaluated, only three structures were correctly identified as nelfinavir-resistant. Correct predictions for these structures are probably due to the combination of Asp30Asn, Val82Ala and Ile84Val mutations. The remaining structures, which were predicted as nelfinavir susceptible, had only Asp30Asn mutation without any other primary mutations. Docking with dynamics also failed to define mutations at residue 48 as saquinavir-resistant. We noticed that Asp30Asn and Gly48Val mutations are located at the edge of the protease active site. The small conformational changes seen in these mutations did not significantly alter the mutant inhibitor binding energies that our docking protocol incorrectly identified as the wild-type.

The important advantage of the docking with dynamics protocol is that it does not need a large phenotype–genotype database for training, since it determines the interaction between the inhibitors and protein structures that are generated directly from the amino acid sequences. Amino acid substitutions that distort the binding pocket conformation highly affect the binding energies of inhibitors, which proportionally vary with the fold-change of the IC_{50} values. Since the K_i values are calculated solely from the binding energy of the protease–inhibitor complex, our protocol outperforms the rule-based and SVM methods for newly approved drugs, that is, amprenavir and lopinavir, for which there are fewer phenotype–genotype results available in a database.

Another advantage of the docking with dynamics over the machine-learning method is that it provides

structural details on how the inhibitor binds to its binding pocket. This structural information would allow us to understand resistance mechanisms and may ultimately be used in deriving a new generation of antiretroviral drugs against resistant strains.

A web server where a mutant protease sequence may be submitted and have its resistance/susceptibility to various drugs predicted using our docking with dynamics protocol, is available at http://protinfo.compbio.washington.edu/pirspred. The docking with dynamics calculations take up to 6 h on a single AMD 1.7 GHz CPU to generate the prediction results for one drug. The web server returns the results immediately if our protocol has already been applied to the mutant sequence, or places the sequence in a queue, which is processed gradually (due to the computational cost of our protocol).

Conclusion

In this study, we present a new docking with dynamics technique that integrates protein relaxation through the use of MD simulations, and applies it to predict resistance/susceptibility of HIV-1 protease to six different inhibitors. The docking with dynamics protocol achieved high accuracy for mutations that alter the protease active site conformation. Results from our protocol were complementary to those obtained from two well-established knowledge-based methods: the rule-based method derived from the Stanford HIV drug resistance database and the geno2pheno SVM-based method. The consensus results generated from all three methods significantly improved the accuracy and enable comprehensive assessment of HIV-1 drug resistance.

Acknowledgements

This work was supported in part by a Searle Scholar Award to Ram Samudrala. We thank Dr John E Mittler and members of the Samudrala group for helpful comments.

References

1. Hammer SM, Vaida F, Bennett KK, Holohan MK, Sheiner L, Eron JJ, Wheat LJ, Mitsuyasu RT, Gulick RM, Valentine FT, Aberg JA, Rogers MD, Karol CN, Saah AJ, Lewis RH, Bessen LJ, Brosgart C, DeGruttola V & Mellors JW; AIDS Clinical Trials Group 398 Study Team. Dual vs single protease inhibitor therapy following antiretroviral treatment failure: a randomized trial. *Journal of the American Medical Association* 2002; **288:**169–180.
2. Gallant JE. Strategies for long-term success in the treatment of HIV infection. *Journal of the American Medical Association* 2000; **283:**1329–1334.
3. Carpenter CC, Cooper DA, Fischl MA, Gatell JM, Gazzard BG, Hammer SM, Hirsch MS, Jacobsen DM, Katzenstein DA, Montaner JS, Richman DD, Saag MS, Schechter M, Schooley RT, Thompson MA, Vella S, Yeni PG & Volberding PA. Antiretroviral therapy in adults: updated recommendations of the International AIDS Society-USA Panel. *Journal of the American Medical Association* 2000; **283:**381–390.
4. Hirsch MS & Richman DD. The role of genotypic resistance testing in selecting therapy for HIV. *Journal of the American Medical Association* 2000; **284:**1649–1650.
5. Rusconi S, La Seta Catamancio S, Sheridan F & Parker D. A genotypic analysis of patients receiving zidovudine with either lamivudine, didanosine or zalcitabine dual therapy using the LiPA point mutation assay to detect genotypic variation at codons 41, 69, 70, 74, 184 and 215. *Journal of Clinical Virology* 2000; **19:**135–142.
6. Frenkel LM, Wagner LE 2nd, Atwood SM, Cummins TJ & Dewhurst S. Specific, sensitive, and rapid assay for human immunodeficiency virus type 1 pol mutations associated with resistance to zidovudine and didanosine. *Journal of Clinical Microbiology* 1995; **33:**342–347.
7. Sugiura W, Shimada K, Matsuda M, Chiba T, Myint L, Okano A & Yamada K. Novel enzyme-linked minisequence assay for genotypic analysis of human immunodeficiency virus type 1 drug resistance. *Journal of Clinical Microbiology* 2003; **41:**4971–4979.
8. Garcia Lerma J, Schinazi RF, Juodawlkis AS, Soriano V, Lin Y, Tatti K, Rimland D, Folks TM & Heneine W. A rapid non-culture-based assay for clinical monitoring of phenotypic resistance of human immunodeficiency virus type 1 to lamivudine (3TC). *Antimicrobial Agents & Chemotherapy* 1999; **43:**264–270.
9. Gonzalez R, Masquelier B, Fleury H, Lacroix B, Troesch A, Vernet G & Telles JN. Detection of human immunodeficiency virus type 1 antiretroviral resistance mutations by high-density DNA probe arrays. *Journal of Clinical Microbiology* 2004; **42:**2907–2912.
10. Holodniy M, Katzenstein D, Winters M, Montoya J, Shafer R, Kozal M, Ragni M & Merigan TC. Measurement of HIV virus load and genotypic resistance by gene amplification in asymptomatic subjects treated with combination therapy. *Journal of Acquired Immune Deficiency Syndromes* 1993; **6:**366–369.
11. Schuurman R, Demeter L, Reichelderfer P, Tijnagel J, de Groot T & Boucher C. Worldwide evaluation of DNA sequencing approaches for identification of drug resistance mutations in the human immunodeficiency virus type 1 reverse transcriptase. *Journal of Clinical Microbiology* 1999; **37:**2291–2296.
12. Race E, Gilbert SM, Sheldon JG, Rose JS, Moffatt AR, Sitbon G, Dissanayeke SR, Cammack N & Duncan IB. Correlation of response to treatment and HIV genotypic changes during phase III trials with saquinavir and reverse transcriptase inhibitor combination therapy. *AIDS* 1998; **12:**1465–1478.
13. Setti M, Bruzzone B, Ansaldi F, Borrelli P, Indiveri F & Icardi G. Identification of key mutations in HIV reverse transcriptase gene can influence the clinical outcome of HAART. *Journal of Medical Virology* 2001; **64:**199–206.
14. Sturmer M, Staszewski S, Doerr HW, Larder B, Bloor S & Hertogs K. Correlation of phenotypic zidovudine resistance with mutational patterns in the reverse transcriptase of human immunodeficiency virus type 1: interpretation of established mutations and characterization of new polymorphisms at codons 208, 211, and 214. *Antimicrobial Agents & Chemotherapy* 2003; **47:**54–61.
15. Kijak GH, Carobene MG & Salomon H. A highly prevalent polymorphism at codon 72 of HIV-1 reverse transcriptase in Argentina prevents hybridization reaction at codon 74 in the LiPA genotyping test. *Journal of Virological Methods* 2001; **94:**87–95.
16. Rick SW, Topol IA, Erickson JW & Burt SK. Molecular mechanisms of resistance: free energy calculations of mutation effects on inhibitor binding to HIV-1 protease. *Protein Science* 1998; **7:**1750–1756.
17. Wang W & Kollman PA. Computational study of protein specificity: the molecular basis of HIV-1 protease drug

resistance. *Proceedings of the National Academy of Sciences, USA* 2001; **98:**14937–14942.

18. Draghici S & Potter RB. Predicting HIV drug resistance with neural networks. *Bioinformatics* 2003; **19:**98–107.
19. Wang D & Larder B. Enhanced prediction of lopinavir resistance from genotype by use of artificial neural networks. *Journal of Infectious Disease* 2003; **188:**653–660.
20. Wang K, Jenwitheesuk E, Samudrala R & Mittler J. Simple linear model provides highly accurate genotypic predictions of HIV-1 drug resistance. *Antiviral Therapy* 2004; **9:**343–352.
21. Quigg M, Frost SD, McDonagh S, Burns SM, Clutterbuck D, McMillan A, Leen CS & Brown AJ. Association of antiretroviral resistance genotypes with response to therapy comparison of three models. *Antiviral Therapy* 2002; **7:**151–157.
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. *Proceedings of the National Academy of Sciences, USA* 2002; **99:**8271–8276.
23. Cai CZ, Han LY, Ji ZL, Chen X & Chen YZ. SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucleic Acids Research* 2003; **31:**3692–3697.
24. Chen YZ, Gu XL & Cao ZW. Can an optimization/scoring procedure in ligand-protein docking be employed to probe drug-resistant mutations in proteins? *Journal of Molecular Graphics & Modelling* 2001; **19:**560–570.
25. Schaffer L & Verkhivker GM. Predicting structural effects in HIV-1 protease mutant complexes with flexible ligand docking and protein side-chain optimization. *Proteins* 1998; **33:**295–310.
26. Joseph-McCarthy D, Thomas BE 4th, Belmarsh M, Moustakas D & Alvarez JC. Pharmacophore-based molecular docking to account for ligand flexibility. *Proteins* 2003; **51:**172–188.
27. Shenderovich MD, Kagan RM, Heseltine PN & Ramnarayan K. Structure-based phenotyping predicts HIV-1 protease inhibitor resistance. *Protein Science* 2003; **12:**1706–1718.
28. Shafer RW, Stevenson D & Chan B. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Research* 1999; **27:**348–352.
29. Jenwitheesuk E & Samudrala R. Improved prediction of HIV-1 protease-inhibitor binding energies by molecular dynamics simulations. *BMC Structural Biology* 2003; **3:**2.
30. Bower M, Cohen FE & Dunbrack RL Jr. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *Journal of Molecular Biology* 1997; **267:**1268–1282.
31. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K & Schulten K. NAMD2: greater scalability for parallel molecular dynamics. *Journal of Computational Physics* 1999; **151:**283–312.
32. Brunger AT. *X-PLOR version 3.1, A System for X-ray Crystallography and NMR*. 1992. New Haven, CT: Yale University Press.
33. Geller M, Miller M, Swanson SM & Maizel J. Analysis of the structure of HIV-1 protease complexed with a hexapeptide inhibitor. Part II: Molecular dynamic studies of the active site region. *Proteins* 1997; **27:**195–203.
34. Piana S & Carloni P. Conformational flexibility of the catalytic Asp dyad in HIV-1 protease: An *ab initio* study on the free enzyme. *Proteins* 2000; **39:**26–36.
35. Hyland LJ, Tomaszek TA Jr & Meek TD. 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:**8454–8463.
36. Yamazaki T, Nicholson LK, Torchia DA, Wingfield P, Stahl SJ, Kaufman JD, Eyermann CJ, Hodge CN, Lam PYS, Ru Y, Jadhav PK, Chang CH & Weber PC. NMR and X-rays evidence that the HIV protease catalytic aspartyl groups are protonated in the complex formed by the protease and a non-peptide cyclic urea-based inhibitor. *Journal of the American Chemical Society* 1994; **116:**10791–10792.
37. Grubmuller H, Heymann B & Tavan P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* 1996; **271:**997–999.
38. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK & Olson AJ. Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. *Journal of Computational Chemistry* 1998; **19:**1639–1662.
39. Case DA, Darden TA, Cheatham, TE, III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS & Kollman PA. AMBER 8. 2004. San Francisco: University of California.

Received 25 August 2004, accepted 22 November 2004