Genotype Dependent QSAR for HIV-1 Protease Inhibition

Carlo W. Boutton,*,† Hendrik L. De Bondt,† and Marc R. De Jonge‡

Tibotec BVBA, Gen. De Wittelaan L11 B3, B-2800 Mechelen, Belgium, and Center for Molecular Design, Janssen Pharmaceutica NV, Antwerpsesteenweg 37, B-2350 Vosselaar, Belgium

Received May 28, 2004

The development of drug-resistant viruses limits the therapeutic success of anti-HIV therapies. Some of these genetic HIV-variants display complex mutational patterns in their *pol* gene that codes for protease and reverse transcriptase, the most investigated molecular targets for antiretroviral therapy. In this paper, we present a computational structure-based approach to predict the resistance of a HIV-1 protease strain to amprenavir by calculating the interaction energy of the drug with HIV-1 protease. By considering the interaction energy per residue, we can identify what residue mutations contribute to drug-resistance. This approach is presented here as a structure-based tool for the prediction of resistance of HIV-1 protease toward amprenavir, with a view to use the drug−protein interaction-energy pattern in a lead-optimization procedure for the discovery of new anti-HIV drugs.

Introduction

The human immunodeficiency virus type 1 aspartic protease (HIV-1 PR) is one of the most important enzymes in anti-AIDS drug design. Inactivation of this enzyme causes the production of immature, noninfectious viral particles and hence blocks further HIV-infection. Currently, seven drugs are approved for treatment of HIV infection by blocking the activity of HIV-1 protease. Despite the enormous investigations in viral therapy and combination of therapies, treatment success is limited due to the evolution of drug-resistant variants.[1] In infected patients, viral populations can rapidly develop resistance to drugs under the antiretroviral pressure of the patients' treatment regimens. In addition, because of the high degree of cross-resistance to HIV-1 protease drugs, infecting viruses may develop resistance to drugs other than those to which they have been exposed. In this context, resistance testing has become an important diagnostic tool in the management of HIV infections.[2,3] Resistance information can be either directly assessed, by phenotypic assays in which recombinant virus techniques directly measure viral replication in the presence of increasing drug concentrations,[4,5] or deduced from genotypic assays that are based on sequencing of the relevant parts of the viral genome.[3] Interpretation of the sequence information of genotypic assays is becoming increasingly difficult because the influence of a certain mutation on drug resistance cannot be considered independently of other mutations, but different types of interactions must be taken into account.[6,7] In the past few years, several attempts have been published to predict in silico the phenotypic behavior of HIV-1 protease mutants. Some algorithms are based upon carefully chosen rules[8,9] extracted from large databases or pattern searches on such large datasets.[10] Other methods use more statistically advanced techniques such as neural networks,[11,12] support vector machines,[13] cluster analysis,[14] decision trees,[15,16] or linear discriminant analysis.[17] All these prediction methods are based upon statistical techniques and their accuracy is largely dependent on the size and complexity of the training set. They are retrospective in nature and must frequently be updated to accommodate new mutational patterns and new antiviral drugs. Especially for new or rare mutational patterns, the results of these methods may be inaccurate. To overcome the disadvantages of statistical-based prediction methods, several attempts have been made to develop computational technologies based on molecular modeling of HIV-1 protease complexes and to build quantitative prediction models for HIV-1 protease inhibitor activity.[18] Given recent improvements in search algorithms and energy functions, computational docking methods have become a valuable tool to probe the interaction between an enzyme and its inhibitors.[19] Several studies have been published about the implications of mutations on the three-dimensional structure of HIV-1 protease,[20−22] mostly based on molecular dynamics studies. The three-dimensional structures of resistant HIV-1 protease strains often show rearrangements, resulting in a decrease in van der Waals interactions with the inhibitor and, consequently, weaker binding to the mutated HIV-1 protease strains. While a few years ago only qualitative predictions were made for resistant mutations,[23] the last publications show a change toward more quantitative structure-based predictions.[24,25] These methods not only lead to valuable predictions of drug-resistant mutations but also improve the understanding of the structural and energetic basis for enzyme−substrate interactions.[26] Furthermore, the development of successful strategies for structure-based molecular design of new drugs requires the ability to accurately predict binding affinities from structural considerations.

In this study, we present a structure-based computational method to quantitatively predict the resistance or sensitivity of an HIV-1 protease strain toward amprenavir (Figure 1). Our approach is a variation of the linear interaction energy method,[27] augmented with

* Corresponding author. Phone: 0032−15293167. Fax: 0032−15286346. E-mail: cboutton@tibbe.jnj.com.
† Tibotec BVBA.
‡ Janssen Pharmaceutica NV.

10.1021/jm049596h CCC: $30.25
Published on Web 11/16/2004

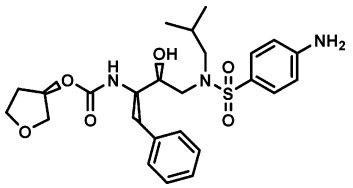

Figure 1. Structure of amprenavir.

protein residue dependence for its activity prediction. It was originally developed for wild-type activity prediction of HIV-1 protease[25] and HIV-1 reverse transcriptase.[28] The method differs from other structure-based prediction attempts in two ways. First, instead of using the total interaction energy between the enzyme and the drug, we consider the interaction energy at residue level; i.e., for each residue of HIV-1 protease, we count its individual interaction energy with the drug. Second, since we consider the interaction energy at the residue level, we take the asymmetry of the complex into account. HIV-1 protease is a symmetric homodimer, with the possible exception of the protonation state of the catalytic site,[29] but after binding an asymmetrical drug, the interaction with the drug is different for both chains[30] and consequently the energy pattern becomes also asymmetrical. These two key points are crucial for our method and open the possibility to use the results with a view to lead optimization purposes.

We do not only predict whether a given HIV-1 protease strain is resistant toward a given drug or not, but can also provide additional information as what residues are responsible for resistance and need attention in structure-based drug design. We believe that this information will prove its value in future drug discovery projects. Since a good quantitative prediction of resistance is crucial in the first place, we limit the scope of this paper to this subject.

Experimental Section

We used the X-ray structure of HIV-1 protease complexed with amprenavir[31] (PDB code 1hpv) as a template for the generation of 2980 closely related homology models. Each of these models is built by modifying amino acids in the X-ray structure according to point mutations found in one of the corresponding 2980 patient HIV-1 protease strains used in this study. The geometry of the mutated amino acids is kept as close as possible to the X-ray structure, by maintaining the torsional angles where applicable and by setting sterically feasible values otherwise. The protein structure is subsequently allowed to accommodate the induced geometry differences by a local, whole complex, conjugated gradient, and truncated Newton optimization. After the minimization, the nonbonded interaction energy of each residue with the inhibitors is calculated. This energy is computed as separate Coulomb, van der Waals, and hydrogen-bond contribution. Moreover, we distinguish between main-chain and side-chain interactions. This results in six energy terms per residue. These terms are computed with a consistent force field parametrization that is similar to MMFF94[32,33] in its functional form but slightly modified to allow for directional H-bonds.[28] The phenotype generation, geometry optimization, and energy computations were performed on an 80-processor cluster and required 60 h of run time. Important to note is also the asymmetry of the interactions. Since both the catalytic site of protease and amprenavir are asymmetrical, the interactions with both chains of the HIV-1 protein will be different, resulting in a different interaction pattern for the two HIV-1 protease monomers constituting the biologically active form.

```
          10        20        30        40        50
          |         |         |         |         |
PQVTLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
          60        70        80        90
          |         |         |         |
GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
```

Figure 2. Wild-type sequence of HIV-1 protease (GenBank ID: NP 705926).

Table 1. Distribution of the Number of Mutations in the Dataset of 2980 HIV-1 Protease Strains

no. of mutations	no. in dataset	no. of mutations	no. in dataset
1	1	15	187
2	20	16	140
3	50	17	135
4	107	18	90
5	137	19	84
6	158	20	50
7	186	21	27
8	185	22	21
9	208	23	14
10	214	24	4
11	260	25	1
12	248	26	2
13	257	27	1
14	193		

Results and Discussion

Composition of the Dataset. The dataset is a phenotypically stratified selection of 2980 randomly chosen HIV-1 protease strains, extracted from an in-house database of HIV-1 protease sequences. Each of the genotypes is unique in the dataset and has at least one, but mostly more, mutations compared to the wild-type sequence (Figure 2).

The number of mutated positions compared to the wild-type sequence of HIV-1 protease (Figure 2) varies from 1 to 27. The complete distribution of the number of mutations in the dataset is listed in Table 1.

For each of the considered genotypes, the susceptibility to amprenavir is compared to the inhibition of amprenavir to the wild-type sequence. We call the fold resistance (FR) for a given HIV-1 protease sequence the relative increase or decrease in inhibition by amprenavir compared to the inhibitory effect of the same drug to the reference sequence. For statistical reasons, and because EC_{50} and binding constants depend on the exponential of energy differences, we will work with the pFR or the negative base-10 logarithm of the fold resistance:

$$\mathrm{pFR} = -\log_{10}(\mathrm{FR}) = -\log_{10}\left(\frac{(\mathrm{EC}_{50})_{\mathrm{mutant}}}{(\mathrm{EC}_{50})_{\mathrm{WT}}}\right) \quad (1)$$

The EC_{50} values are measured with Antivirogram, a commercially available phenotypic assay.[34] A positive pFR value reflects a lower EC_{50} of amprenavir to the mutant sequence, compared to the EC_{50} of amprenavir to the wild-type sequence. It indicates an increased susceptibility of the HIV-protease strain toward amprenavir, while a negative pFR points to a more amprenavir-resistant HIV-1 protease strain. In the whole dataset of 2980 unique strains, the pFR ranges from hypersensitive values (pFR > +0.94) to very resistant values (pFR < −2.1). A total of 1085 or 36% of the genotypes have a pFR value larger than 0 and are thus hypersensitive. The distribution of pFR in the dataset is shown in Figure 3.

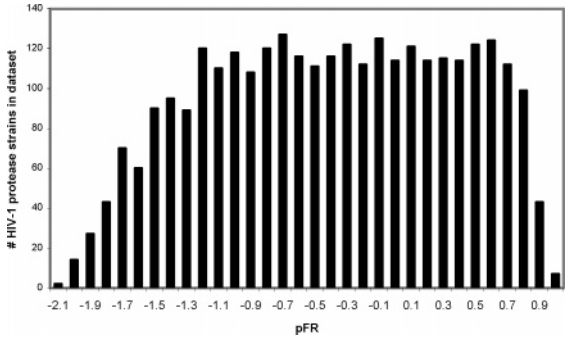

Figure 3. Distribution of the resistance in the dataset of 2980 HIV-1 protease strains.

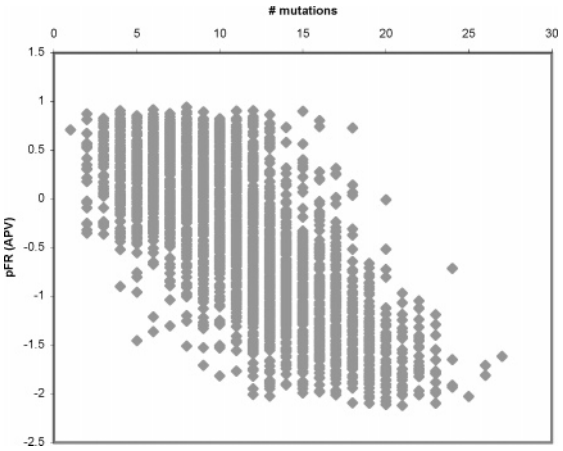

Figure 4. Amprenavir resistance as function of the number of mutations in HIV-1 protease.

Intuitively, one could expect that more mutations lead to a higher resistance, and by analyzing the dataset, we indeed see that there is a tendency that highly mutated sequences show more resistance (Figure 4). However, to become resistant the virus needs to find a compromise between inhibition and fitness. Highly mutated genotypes can loose their activity or show a decreased fitness by which the normal enzymatic activity of the HIV-1 protease is disturbed. This balance between inhibition and activity/fitness is the reason there is no perfect correlation between resistance and the number of mutations. Moreover, in the dataset we see resistant strains only having two mutations. For the virus it is crucial to find those positions that can be mutated to lead to resistance and have a minimal effect on the fitness. The power of our approach here is to point out those key mutations in HIV-1 protease to circumvent the inhibitory effect of amprenavir.

Each HIV-1 protease strain in the dataset is transformed into a three-dimensional complex with amprenavir by the protocol described in the Experimental Section. Afterward, a drug−protein interaction energy map is derived from each 3D complex of each HIV-1 protease strain. The interaction energy is split in three terms: a Coulombic, electrostatic, and hydrogen-bond energy contribution. We further distinguish these three energy terms for backbone and side-chain contributions, resulting in six energy contributions for each residue. The total energy matrix thus consists of six columns, one for each counted energy contribution, and 198 rows, a row for each residue in the HIV-1 protease dimer. We want to underline that in contrast to all other prediction

residue	residue type	Ecsc	Ecbb	Evsc	Evbb	Ehsc	Ehbb
20	LYS	-0.001	0.008	-0.001	-0.001	0.000	0.000
21	GLU	-0.003	-0.003	-0.001	-0.002	0.000	0.000
22	ALA	0.000	0.035	-0.003	-0.007	0.000	0.000
23	LEU	-0.016	-0.036	-0.652	-0.030	0.000	0.000
24	LEU	0.000	0.078	-0.006	-0.026	0.000	0.000
25	ASP	-4.365	-0.335	1.156	-0.106	-1.267	0.000
26	THR	0.134	0.338	-0.042	-0.153	0.000	0.000
27	GLY	0.000	-16.988	0.000	3.577	0.000	-2.469
28	ALA	-0.033	-6.822	-0.569	-1.652	0.000	0.000
29	ASP	-0.323	-7.845	-0.497	0.240	0.000	-0.678
30	ASP	0.073	1.896	-0.298	-0.231	0.000	-0.516
31	THR	-0.016	1.107	-0.024	-0.168	0.000	0.000
32	VAL	-0.093	-0.355	-0.946	-0.173	0.000	0.000
33	LEU	-0.001	-0.169	-0.007	-0.016	0.000	0.000
34	GLU	0.001	-0.024	-0.004	-0.005	0.000	0.000
35	GLU	0.000	-0.011	-0.001	-0.002	0.000	0.000
36	ASN	0.014	-0.008	-0.001	0.000	0.000	0.000
37	SER	-0.001	-0.002	0.000	0.000	0.000	0.000
38	LEU	0.000	-0.002	0.000	0.000	0.000	0.000
39	PRO	0.005	-0.007	0.000	0.000	0.000	0.000
40	GLY	0.000	0.000	0.000	0.000	0.000	0.000

Figure 5. Energy matrix of amprenavir with a randomly chosen HIV-1 protease strain.

algorithms, we can distinguish all 198 residues. We know that HIV-1 protease is a homodimer with two symmetrical chains, but although the sequence similarity is conserved, the 2-fold symmetry disappears after binding an asymmetrical ligand. As a consequence, the drug-interaction energy of a residue in one chain will be different than the drug-interaction energy of its counterpart in the other chain. It is also known that structural water molecules can play a crucial role in the binding of HIV-1 protease inhibitors.[35] Although they are important and despite the fact that their presence is taken into account when deriving the three-dimensional models, their interaction energy with the drug is not part of the prediction model. Since they cannot be mutated and therefore not be directly used by the protein to circumvent inhibition, we chose not to use them as variables for prediction, but their presence in the three-dimensional models will have an influence on the drug-interaction energy of neighboring residues. When modifying this approach for drug discovery purposes, structural water molecules can be considered either as part of the protein and incorporated as extra residues in the drug-interaction energy matrix or as part of the ligand. An example of an interaction-energy matrix is given in Figure 5.

All 2980 energy matrices are merged into one dataset of 2980 data points (HIV-1 protease strains) and $198 \times 6 = 1188$ variables (energy terms). Before extracting a linear prediction model, we removed 596 data points, or 20% of the dataset, to use them as a blind test afterward. The strains in this test set were chosen on a random base, so that the distribution of pFR and number of mutations is representative for the rest of the sample. From the 2384 (80% of the total data set) kept data points in the training set, we searched for the best linear prediction function. This is done in SAS,[36] by a stepwise variable selection approach. The result is a linear sum of energy contribution each with a weighing factor:

$$\mathrm{pFR}_{\mathrm{predicted}} = \beta_0 + \sum_{j=1}^{198} \sum_{k=1}^{6} \beta_{jk} E_{jk} \qquad (2)$$

Table 2. Comparison of Two Models, with Different Number of Variables

	model 1		model 2	
	training set	test set	training set	test set
no. of variables	136		40	
r^2	0.84	0.76	0.794	0.783
rmse	0.31	0.38	0.345	0.348

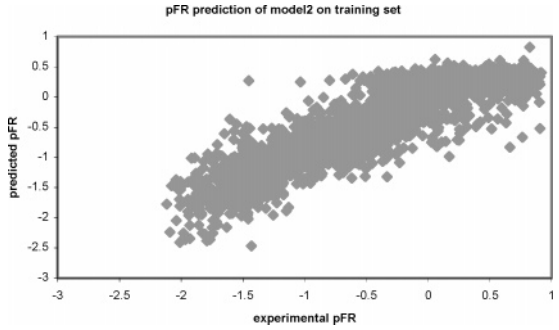

Figure 6. Correlation between predicted and measured resistance for the training set.

where E_{jk} is the drug-interaction energy term k of residue j, β_{jk} is the weight on the interaction energy contribution k of residue j, and β_0 is the intercept.

The prediction on the training set resulted in a squared correlation coefficient of 0.84 and a root mean square error (rmse) of 0.31, which is close to the Antivirogram reproducibility error.[34] Afterward, we applied the model on the test set, resulting in a correlation coefficient of 0.76 and a rmse of 0.38. We recognize that the prediction accuracy on the test set decreased and thus we suspect our prediction model of overfitting, but the correlation on the test set is still high enough to perform reliable predictions. The prediction model used 136 variables from the 1188 that maximally can be chosen. We agree that this number of variables is rather large, and therefore, we derived a new model with a limited number of input variables. Moreover, we saw that only 41 variables are responsible for a squared correlation coefficient of 0.80 or 41 variables can explain 95% of the prediction power of the total model. Therefore, we derived a new model with only these 41 most important parameters as input. The new prediction equation contains 40 significant variables and has a squared correlation coefficient on the training set of 0.794 with a corresponding rmse of 0.345. On the test set, the squared correlation coefficient only slightly drops to 0.783, while the rmse remains almost constant (0.348). A summary of the statistics of both models is given in Table 2.

It is clear that the bias of the first model (model 1) with the 136 variables disappeared after reducing the number of variables in the second model (model 2). The almost constant correlation coefficient and rmse for the training and test set prove the robustness of model 2. The predicted pFR versus the experimental value for this model is shown in Figures 6 and 7 for the training and test set, respectively.

The contribution of the variables to the total prediction power of the linear model is shown in Figure 8 for the 10 most important variables.

The combination of these 10 variables contributes for more than 90% of the total predictive power of the

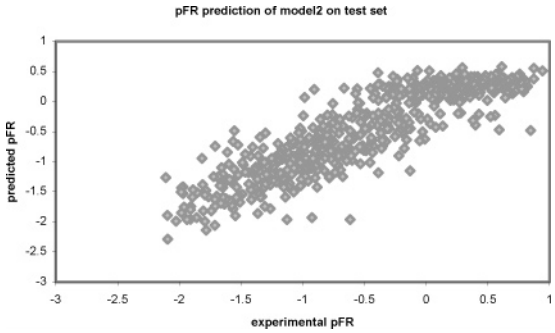

Figure 7. Correlation between predicted and measured resistance for the test set.

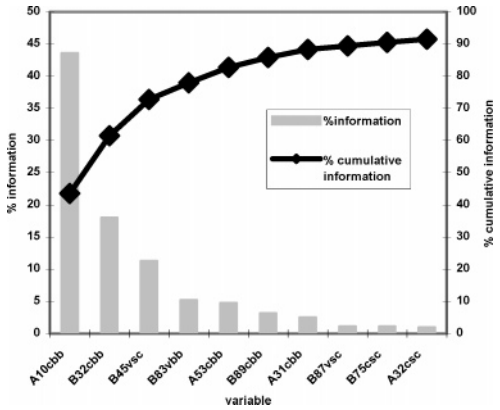

Figure 8. Information contribution of the 10 most important interaction energy values.

model. The weight factors for the input variables (energy contributions) are not always 1, indicating that a selection of variables instead of an equation with weighted energy terms would not give a good prediction model. A perfect correlation is not to be expected between the resistance (pFR) and a simple full weight sum of the energy variables for several reasons. First, our interaction terms only consider enthalpy effects instead of free interaction energy; second, our experimental values result from a cell-based assay that is less accurate than an enzymatic assay; and third, the inhibitory effect of amprenavir on each HIV-protease strain will also be influenced by the changed kinetics of the different enzyme genotypes.

We further want to focus on three striking observations that further support the unique strength of structure-based resistance predictions. First, by considering only the 10 most important contributions, six are interactions where the backbone is involved. This can be surprising since most prediction algorithms focus on the nature of the side-chain mutations, but it is logical that also backbone energies are affected by side-chain mutations. What we see here reflects the results of a drug-design process. Enzyme inhibitors are chemically constructed to have their crucial interactions with the protein backbone to prevent escaping inhibition by mutations of the side chains. From the 40 variables used in the prediction model, 22 involve backbone interactions and 18 are side-chain interactions with the drug. This again emphasizes the difference or our approach to the classical prediction algorithms. The nature of the mutations is transformed to a list of energy changes, and instead of focusing on the side-chain mutations themselves, we rather look at the energy changes at the

mutated positions and their environment. Second, if we consider the 10 most important residues positions, then eight are known as positions that cause resistance to amprenavir.[37] If we compare the whole list of most important residue interactions, almost 60% of the variables are on positions that have a sequence shift of at most one residue to positions that are known resistance mutations for amprenavir. The fact that we do not always predict the position itself is logical, since a mutation on a particular position will have an effect on the neighboring positions. This again confirms the value of the structure-based approach and its use for lead optimization purposes. Third, contributions of residues in one chain are not the same as their counterpart in the other chain. If we annotate the 2 HIV-1 protease chains A and B, then from the 10 most important contributions six are related to residues in the A-chain, while the other four are contributions from other residues in the B-chain. This asymmetry in our prediction model reflects the asymmetry of HIV-1 protease after binding an asymmetrical drug. To the best of our knowledge, this is the first time a prediction tool is described that is able to point out the asymmetry of the HIV-1 protease after drug binding.

The method described here is developed for lead optimization by structural knowledge extraction from resistance predictions. A necessary condition to reach this goal is to perform reliable resistance predictions, which was the topic of this paper. We showed that by considering the drug-interaction energy per residue, an accurate and reliable prediction model can be derived.

Acknowledgment. The authors wish to thank Kurt Hertogs and his team for providing the experimental data. Herwig Van Marck, Tim Van den Bulcke, and Hans Vermeiren are acknowledged for the stimulating discussions.

Note Added after ASAP Publication. This manuscript was released ASAP on 11/16/2004 with an error in the location of a data point in Table 2. The correct version was posted on 11/19/2004.

References

(1) Shafer, R. W.; Kantor, R.; Gonzales, M. J. The genetic basis of HIV-1 resistance to reverse transcriptase and protease inhibitors. *AIDS Rev.* **2000**, *2*, 211−228.
(2) Perrin, L.; Telenti, A. HIV treatment failure: Testing for HIV resistance in clinical practice. *Science* **1998**, *280*, 1871−1873.
(3) Vandamme, A. M.; Van Laethem, K.; De Clerq, E. Managing resistance to anti-HIV drugs: An important consideration for effective disease management. *Drugs* **1999**, *57*, 337−361.
(4) Hertogs, K.; de Bethune, M.-P.; Miller, V.; Ivens, T.; Schel, P.; van Cauwenberghe, A.; van den Eynde, C.; van Gerwen, V.; Azijn, H.; van Houtte, M. A rapid method for simultaneous detection of phenotypic resistance to inhibitors of protease and reverse transcriptase in recombinant human immunodeficiency virus type I isolates form patients treated with antiretroviral drugs. *Antimicrob. Agents Chemother.* **1998**, *42*, 269−276.
(5) Walter, H.; Schmidt, B.; Korn, K.; Vandamme, A. M.; Harrer, T.; Überla, K. Rapid, Phenotypic HIV-1 drug sensitivity assay for protease and reverse transcriptase inhibitors. *J. Clin. Virol.* **1999**, *13*, 71−80.
(6) Mahalingam, B.; Boross, P.; Wang, Y.-F.; Louis, J. M.; Fischer, C. C.; Tozser, J.; Harrison, R. W.; Weber, I. T. Combining Mutations in HIV-1 Protease to Understand Mechanisms of Resistance. *Proteins* **2002**, *48*, 107−116.
(7) Tisdale, M.; Kemp, S. D.; Parry, N. R.; Larder, B. A. Rapid in vitro selection of human immunodeficiency virus type 1 resistant to 3′-thiacytidine inhibitors due to a mutation in the YMDD region of reverse transcriptase. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5653−5656.
(8) Schmidt, B.; Walter, H.; Moschik, B.; Paatz, C.; van Vaerenbergh, K.; Vandamme, A. M.; Schmitt, M.; Harrer, T.; Überla, K.; Korn, K. Simple Algorithm derived from a geno-/phenotypic database to predict HIV-1 protease inhibitor resistance. *Aids* **2000**, *14*, 1731−1738.
(9) Ravela, J.; Betts, B. J.; Brun-Vézinet, F.; Vandamme, A. M.; Descamps, D.; Van Laethem, K.; Smith, K.; Schapiro, J. M.; Winslow, D. L.; Reid, C.; Shafer, R. W. HIV-1 Protease and Reverse Transcriptase Mutation Patterns Responsible for Discordances Between Genotypic Drug Resistance Interpretation Algorithms. *J. Acquir. Immune Defic. Syndr.* **2003**, *33*, 8−14.
(10) Shafer, R. W.; Jung, D. R.; Betts, B. J. Human immunodeficiency virus type 1 reverse transcriptase and protease mutation search engine for queries. *Nat. Med.* **2000**, *6*, 1290−1292.
(11) Drâghici, S.; Potter, R. B. Predicting HIV drug resistance with neural networks. *Bioinformatics* **2002**, *19*, 98−107.
(12) Wang, D. J.; Larder, B. A.; Revell, A.; Harrigan, R.; Montaner, J. A neural network model using clinical cohort data accurately predicts virological response and identifies regimens with increased probability of success in treatment failures. *Antiviral Ther.* **2003**, *8*, S112.
(13) Beerenwinkel, N.; Schmidt, B.; Walter, H.; Kaiser, R.; Lengauer, T.; Hoffman, D.; Korn, K.; Selbig, J. Quantitative phenotype prediction by support vector machines. *Antiviral Ther.* **2002**, *7*, S97.
(14) Beerenwinkel, N.; Kaiser, R.; Schmidt, B.; Walter, H.; Korn, K.; Hoffmann, D.; Lengauer, T.; Selbig, J. Clustering resistance factors: Identification of complex categories of drug resistance. *Antiviral Ther.* **2001**, *6*(S1), 105.
(15) Beerenwinkel, N.; Kaiser, R.; Rahnenführer, J.; Däumer, M.; Hoffmann, D.; Selbig, J.; Lengauer, T. Tree models for the evolution of drug resistance 2003, XII International HIV Drug Resistance Workshop: Basic Principles and Clinical Implications, Abstract97, Los Cabos, Mexico.
(16) Beerenwinkel, N.; Schmidt, B.; Walter, H.; Kaiser, R.; Lengauer, T.; Hoffmann, D.; Korn, K.; Selbig, J., Diversity and complexity of HIV-1 drug resistance: A bioinformatics approach to predicting phenotype from genotype. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 8271−8276.
(17) Sevin, A. D.; DeGruttola, V.; Nijhuis, M.; Schapiro, J. M.; Foulkes, A. S.; Para, M. F.; Boucher, C. A. Methods for investigation of the Relationship between Drug-Susceptibility Phenotype and Human Immunodeficiency Virus Type 1 Genotype with Applications to AIDS Clinical Trials Group 333. *J. Infect. Dis.* **2000**, *182*, 59−67.
(18) Wang, W.; Kollman, P. A. Computational study of protein specificity: The molecular basis of HIV-1 protease drug resistance. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14937−14942.
(19) Schapira, M.; Totrov, M.; Abagyan, R. Prediction of the binding energy of small molecules, peptides and proteins. *J. Mol. Recogn.* **1999**, *12*, 177−190.
(20) Rose, B.; Craik, C. S.; Stroud, R. M. Domain flexibility in Retroviral Proteases: Structural Implications for Drug Resistant Mutations. *Biochemistry* **1998**, *37*, 2607−2621.
(21) Mahalingam, B.; Louis, J. M.; Hung, J.; Harrison, R. W.; Weber, I. T. Structural Implications of Drug-Resistant Mutants of HIV-1 Protease: High-Resolution Crystal Structures of the Mutant Protease/Substrate Analogue Complexes. *Proteins: Struct. Funct. Genet.* **2001**, *43*, 455−464.
(22) Piana, S.; Carloni, P.; Rothlisberger, U. Drug resistance in HIV-1 protease: Flexibility assisted mechanism of compensatory mutations. *Protein Sci.* **2002**, *11*, 2393−2402.
(23) 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*, 560−670.
(24) Shenderovich, M. D.; Kagan, R. M.; Heseltine, P. N.; Ramnarayan, K. Structure-based phenotype predicts HIV-1 protease inhibitor resistance. *Protein Sci.* **2003**, *12*, 1706−1718.
(25) Vinkers, H. M.; De Jonge, M. R.; Daeyaert, F. D.; Heeres, J.; Koymans, L. M. H.; van Lenthe, J. H.; Lewi, P. J.; Timmerman, H.; Janssen, P. A. J. Inhibition and substrate recognition-a computational approach applied to HIV protease. *J. Comput. Aided Mol. Des.* **2003**, *17*, 567−581.
(26) Stoffler, D.; Sanner, M. F.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. Evolutionary Analysis of HIV-1 Protease Inhibitors: Methods for Design of Inhibitors That Evade Resistance. *Proteins: Struct. Funct. Genet.* **2002**, *48*, 63−74.
(27) Aqvist, J.; Medina, C.; Samuelson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385−391.
(28) De Jonge, M. R.; Koymans, L. M. H.; Vinkers, H. M.; Daeyaert, F. D. Structure based activity prediction of HIV-1 reverse transcriptase inhibitors. *J. Med. Chem.* **2004**, submitted.

(29) Chen, X.; Tropsha, A. Relative binding free energies of peptide inhibitors of HIV-1 protease: The influence of the active site protonation state. *J. Med. Chem.* **1995**, *38*, 42−48.
(30) 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.
(31) Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. Crystal Structure of HIV-1 protease in complex with VX-478, a portent and orally bioavailable inhibitor of the enzyme. *J. Am. Chem. Soc.* **1995**, *117*, 1181−1182.
(32) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parametrization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 450−619.
(33) Halgren, T. A. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *17*, 720−729.
(34) Hertogs, K.; de Bethune, M.-P.; Miller, V.; Ivens, T.; Schel, P.; van Cauwenberghe, A.; van den Eynde, C.; van Gerwen, V.; Azijn, H.; van Houtte, M. A rapid method for simultaneous detection of phenotypic resistance to inhibitors of protease and reverse transcriptase in recombinant human immunodeficiency virus type I isolates form patients treated with antiretroviral drugs. *Antimicrob. Agents Chemother.* **1998**, *42*, 269−276.
(35) Mahalingam, B.; Louis, J. M.; Hung, J.; Harrison, R. W.; Weber, I. T. Structural Implications of Drug-Resistant Mutants of HIV-1 Protease: High-Resolution Crystal Structures of the Mutant Protease/Substrate Analogue Complexes. *Proteins* **2001**, *43*, 455−464.
(36) SAS is a registered trademark of SAS institute Inc., Cary, NC.
(37) Stanford HIV Drug Resistance Database, http://hivdb.stanford.edu/cgi-bin/PIResiNote.cgi.

JM049596H