Structure-Based QSAR Analysis of a Set of 4-Hydroxy-5,6-dihydropyrones as Inhibitors of HIV-1 Protease: An Application of the Receptor-Dependent (RD) 4D-QSAR Formalism

Osvaldo A. Santos-Filho*,‡ and Anton J. Hopfinger†

Center for Pharmaceutical Biotechnology, College of Pharmacy, and Bioinformatics Program (M/C 563),
Department of Bioengineering, College of Engineering, The University of Illinois at Chicago,
835 S. Wolcott Street, Chicago, Illinois 60612-7340

Receptor-dependent (RD) 4D-QSAR models were constructed for a set of 39 4-hydroxy-5,6-dihydropyrone analogue HIV-1 protease inhibitors. The receptor model used in this QSAR analysis was derived from the HIV-1 protease (PDB ID 1d4s) crystal structure. The bound ligand in the active site of the enzyme, also a 4-hydroxy-5,6-dihydropyrone analogue, was used as the reference ligand for docking the data set compounds. The optimized RD 4D-QSAR models are not only statistically significant ($r^2 = 0.86$, $q^2 = 0.80$ for four- and greater-term models) but also possess reasonable predictivity based on test set predictions. The proposed "active" conformations of the docked analogues in the active site of the enzyme are consistent in overall molecular shape with those suggested from crystallographic studies. Moreover, the RD 4D-QSAR models also "capture" the existence of specific induced-fit interactions between the enzyme active site and each specific inhibitor. Hydrophobic interactions, steric shape requirements, and hydrogen bonding of the 4-hydroxy-5,6-dihydropyrone analogues with the HIV-1 protease binding site model dominate the RD 4D-QSAR models in a manner again consistent with experimental conclusions. Some possible hypotheses for the development of new lead HIV-1 protease inhibitors can be inferred from the RD 4D-QSAR models.

INTRODUCTION

Acquired immune deficiency syndrome (AIDS) is one of the most serious health problems today. Currently, there is no cure for the disease, and it is considered one of the most challenging subjects of the medical sciences. AIDS is characterized by a susceptibility to infection with opportunistic pathogens or by the occurrence of an aggressive form of Kaposi's sarcoma or B-cell lymphoma, accompanied by a profound decrease in the number of CD4 T cells. AIDS is caused by at least two types of human immunodeficiency virus (HIV), HIV-1 and HIV-2. The first is the most virulent type of HIV and, consequently, is responsible for the majority of cases worldwide. HIV-2 has largely been endemic in West Africa but is now spreading in India. Both viruses seem to have spread to humans from other primate species.[1]

Two viral enzymes, reverse transcriptase and HIV protease, have been the key targets of drugs designed to arrest viral replication, and a combination of inhibitors of both enzymes has become the standard front-line AIDS therapy. Despite promising results from clinical trials for such inhibitors, several problems still persist, especially with the protease inhibitors, which include low bioavailability,[2] side effects,[3] toxicity,[4] drug interactions with other common medications,[5,6] and viral resistance.[7] Consequently, the search for new lead inhibitors remains an active area of investigation.

Peptidomimetic compounds are among the most potent HIV-1 protease (HIV-1 PR) inhibitors. However, these inhibitors have shown low bioavailability and poor pharmacokinetics and normally possess multiple stereocenters. Hence, efforts have increasingly focused upon identifying nonpeptidic HIV-1 PR inhibitors. 4-hydroxy-5,6-dihydropyrone analogues appear to be one promising and novel class of nonpeptidic protease inhibitors.[8,9]

In a previous study, we investigated a set of 4-hydroxy-5,6-dihydropyrone HIV-1 PR inhibitors[9] and constructed QSAR models using the *receptor-independent* (RI) 4D-QSAR paradigm.[10] Those models provided some preliminary pharmacophore insights for the selection of new 4-hydroxy-5,6-dihydropyrone HIV-1 PR lead inhibitors. However, the absence of the 3D structure of the "receptor", in this case, HIV protease, can be a major limitation in the generality and reliability of such 4D-QSAR models.

In this paper, we report the construction of new QSAR models for an expanded set of 4-hydroxy-5,6-dihydropyrone analogues using the *receptor-dependent* (RD) version of the 4D-QSAR paradigm.[11] As will be shown later, the use of the 3D structure of the enzyme in constructing the 4D-QSAR models considerably improves the overall quality of the models. Moreover, the proposed "active" conformations of the analogues docked into the active site of the enzyme seem to be in accordance with the limited experimental data.

METHODS

Data Set of 4-Hydroxy-5,6-dihydropyrone HIV-1 PR Analogues. The chemical structures and the biological activities of the 4-hydroxy-5,6-dihydropyrone derivatives,

---

* Corresponding author e-mail: sanfil@uic.edu.
† Current address: College of Pharmacy, MSC09 5360, 1 University of New Mexico, Albuquerque, NM 87131-0001. E-mail: hopfingr@unm.edu.
‡ Current address: Department of Medicinal Chemistry and Pharmacognsy (MC 781), College of Pharmacy, The University of Illinois at Chicago, 833 S. Wood Street, Chicago, IL 60612-7231.
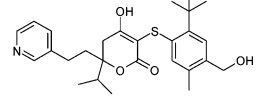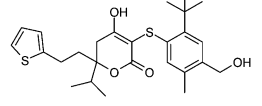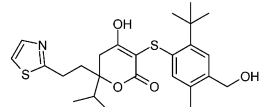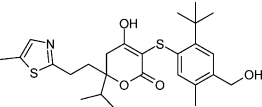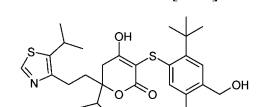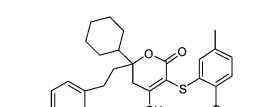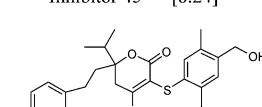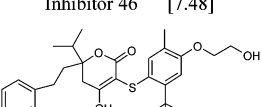
Table 1. Chemical Structures and Inhibition Potencies [PIC$_{50}$ (M)] of the Training Set Compounds

Inhibitor 1-(S)   [9.96]   Inhibitor 2   [8.92]   Inhibitor 3   [8.82]   Inhibitor 4   [9.55]

Inhibitor 5   [7.15]   Inhibitor 6   [9.11]   Inhibitor 7   [9.47]   Inhibitor 8   [8.30]

Inhibitor 9-(S)   [9.89]   Inhibitor 9-(R)   [7.94]   Inhibitor 11   [8.89]   Inhibitor 12   [9.62]

Inhibitor 13   [9.39]   Inhibitor 14   [9.24]   Inhibitor 15   7.85   Inhibitor 16   9.72

Inhibitor 17   [9.35]   Inhibitor 18   [7.89]   Inhibitor 19   [8.89]   Inhibitor 20   [7.85]

Inhibitor 21   [8.80]   Inhibitor 22   [9.10]   Inhibitor 23   [7.74]   Inhibitor 24   [9.02]

Inhibitor 25   [9.01]   Inhibitor 26   [9.14]   Inhibitor 27   [8.85]   Inhibitor 28   [7.52]

Inhibitor 29   [8.34]   Inhibitor 30   [6.47]   Inhibitor 31   [7.67]   Inhibitor 32   [8.23]

Inhibitor 33   [8.48]   Inhibitor 34   [9.00]   Inhibitor 35   [8.19]   Inhibitor 36   [10.52]

Inhibitor 37   [8.54]   Inhibitor 38   [8.19]   Inhibitor 39   [8.42]   Inhibitor 40   [8.96]

which compose both the training and the test sets, are given in Tables 1 and 2. Details about the biological assays to determine the reported molar inhibition potency [pIC$_{50}$ (M)] are published elsewhere.[8,9]

QSAR ANALYSIS OF 4-HYDROXY-5,6-DIHYDROPYRONES

J. Chem. Inf. Model., Vol. 46, No. 1, 2006 **347**

Table 2. Chemical Structures and the Inhibition Potencies of the Test Set Compounds

Inhibitor 41 [9.03] Inhibitor 42 [9.46]

Inhibitor 43 [9.24] Inhibitor 44 [8.72]

Inhibitor 45 [8.24] Inhibitor 46 [7.48]

Inhibitor 47 [8.44] Inhibitor 48 [8.59]

Hagen and collaborators reported that the enolic hydroxyl group of the dihydropyrone core forms hydrogen bonds with the aspartate residues (Asp-25 and Asp-125) in the active site of the HIV-1 PR. Moreover, the lactone moiety interacts with the NH moiety of the isoleucine residues (Ile-50 and Ile-150). Those researchers also predicted that hydrophobic groups appended to the 6 position of the analogues would occupy the S1 and S2 pockets of HIV-1 PR and that the phenyl substituents on the S-phenyl moiety at C-3 would fill the S1′ and S2′ pockets.[8,9] Figure 1 shows a schematic representation of these molecular interactions and binding modes.

The Crystal Structure of the HIV-1 Protease. The "receptor" geometry employed in this work is the crystallographic structure of the HIV-1 protease (HIV-1 retropepsin; E.C. 3.4.23.16; PDB ID 1d4s).[12] Its structure is shown in Figure 2. This structure was obtained from the RCSB Protein Data Bank (PDB).[13] The reason for choosing this particular structure is due to the chemical structure of the docked ligand, that is, also a 4-hydroxy-5,6-dihydro-2-pyrone ana-
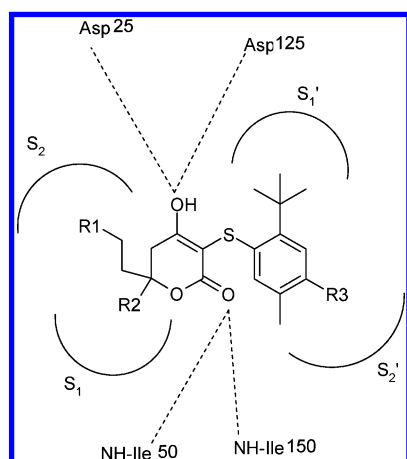
Figure 1. Schematic representation of the possible hydrogen-bonding interactions and the binding pose of a prototypical dihydropyrone in the active site of HIV-1 protease.

logue. As will be shown later, such a structure was used to define the reference alignment during the docking of the training set and test set ligands into the enzyme active site.

The RD 4D-QSAR Method. The 4D-QSAR method was originally developed to investigate structure−activity data sets where the geometry of the receptor was *not* available.[10,14−19] This form of the 4D-QSAR paradigm is now defined as RI 4D-QSAR analysis. Recently, the 4D-QSAR paradigm has been extended to explicitly include the geometry of the receptor in building a QSAR model. That is, the methodology now includes the capacity to do *quantitative* structure-based design. This form of 4D-QSAR study is called RD 4D-QSAR analysis.[11]

The RD 4D-QSAR analysis consists of 12 steps, which can be summarized as shown in Table 3. A detailed description of the method is given in ref 11. The principle feature of a RD 4D-QSAR analysis is that the resultant pharmacophore sites of the QSAR models generated in the analysis are explicitly dependent upon the combined geometries of the (bound) ligand and the receptor.

RESULTS

The geometry of 4-hydroxy-5,6-dihydro-2-pyrone bound in the active site of HIV-1 PR (PDB ID 1d4s)[12] was used as the reference structure in the docking procedure for the training set and test set compounds. That is, these new ligands were docked by superimposing selected atoms from them onto corresponding atoms of the ligand of the reference structure (4-hydroxy-5,6-dihydro-2-pyrone) as it is aligned in the active site of the enzyme. The reference ligand was deleted after each new ligand was posed using this binding alignment. This type of docking procedure is known as docking by superimposition.[20] After obtaining the enzyme−ligand complexes, each ligand was energy-minimized in sito using steepest descent, conjugate gradient, and Newton−Raphson methods,[21,22] as implemented in the HyperChem program.[23] Partial atomic charges of the ligand inhibitors were computed using the semiempirical MNDO method.[24]

The structure of HIV-1 PR consists of 3128 atoms (without the inhibitor and water molecules), organized into two symmetrical polypeptide side chains. Receptor pruning, using the Tokarski−Hopfinger pruning method,[25] was performed for the complete HIV-1 PR−ligand complex in order to scale down the protein to a manageable size effectively containing only the residues of the "lining" of the binding site. The total inhibitor-interaction energy for the largest inhibitor, ligand 38 in Table 1, shows no significant change after the size (radius) of the pruned HIV-1 PR receptor increases beyond 14 Å. Hence, residues more than 14 Å away from this center were neglected in building the receptor binding site model. However, residues having at least one atom in the 14 Å region were completely included in the pruned receptor (the binding lining) model. A 2 Å outer layer (14 − 12 Å) of the pruned receptor model was treated as a "frozen shell" in order to maintain the conformational integrity of the composite binding site environment of the receptor. Isolated residues, separated by less than five intervening residues, were connected through the missing residues. Zero-mass and zero-partial-charge hydrogen atoms were used to complete the open ends of the residue fragments of the pruned receptor model. Ligand 38 is shown docked
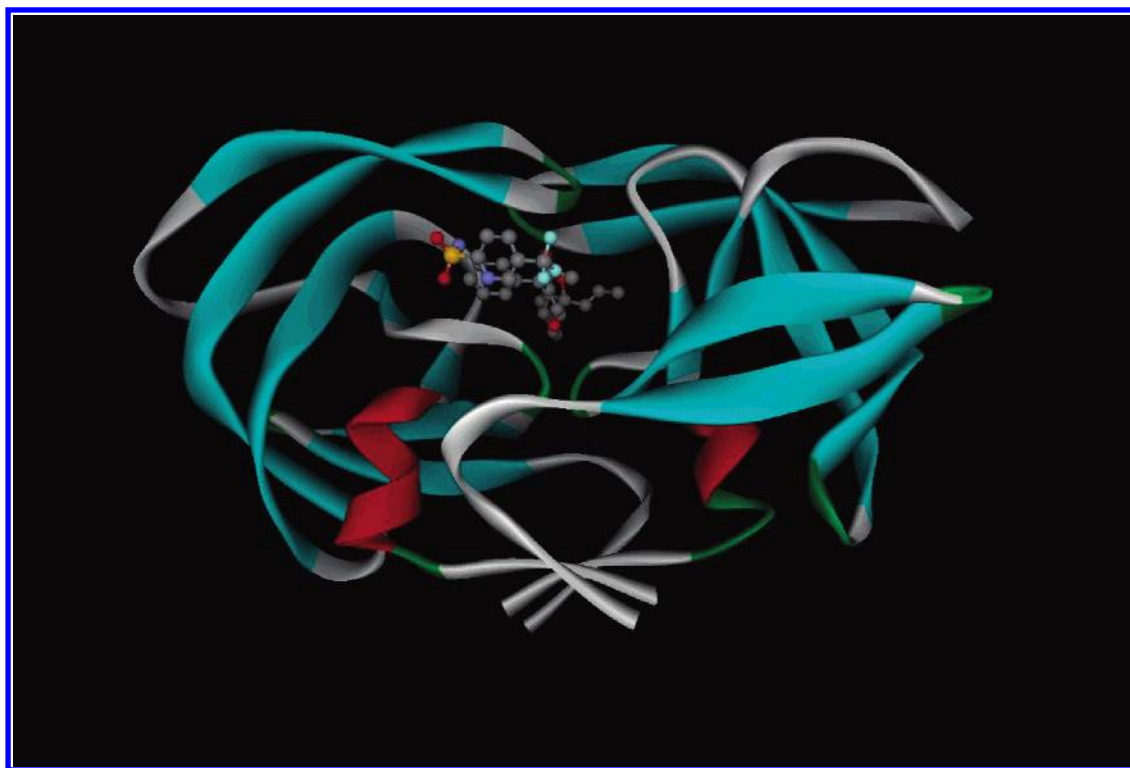
Figure 2. Structure of HIV-1 protease (HIV-1 retropepsin, E.C. 3.4.23.16; PDB ID 1d4s). The docked structure of 4-hydroxy-5,6-dihydro-2-pyrone is shown in ball-and-stick representation.

Table 3. The Twelve Operational Steps in Performing a RD 4D-QSAR Analysis[11]

step	description of the step operation
1	receptor pruning and atom charge assignment of the "receptor"
2	modeling of the data set of compounds
3	ligand docking
4	select the trial set of interaction pharmacophore elements, IPEs
5	constraint of selective receptor and inhibitor atoms
6	molecular dynamics simulations of each pruned "receptor"−inhibitor complex
7	alignment in a binding site
8	grid analysis
9	trial descriptor pool generation
10	partial least-squares (PLS) regression analysis
11	construction of RD 4D-QSAR models/GFA−MLR−GFA
12	"active" conformation postulation of the ligands

into the pruned inhibitor binding site of the HIV-1 PR receptor geometry in Figure 3.

Five trial alignments were selected to systematically explore the principle alignment sites on each ligand (Table 4). The alignments were chosen to span across the entire ligand and to focus on the "sides" and the "center" of the ligand. All selected alignments, except alignment 1, led to RD 4D-QSAR models possessing large numbers of outlier compounds and, consequently, having low correlation coefficients of fit (r^2) and cross-validation correlation coefficients (q^2). Consequently, alignment 1 was identified as the best alignment to use in building the RD 4D-QSAR models, and only QSAR models related to this alignment are discussed in this paper.

RD 4D-QSAR analysis permits the classification and characterization of enzyme−ligand interactions by the assignment interaction pharmacophore elements. Seven interaction pharmacophore elements (IPEs) are currently used. These IPEs are defined as follows: (a) any type of atom

(*any*), (b) nonpolar atom (*np*), (c) polar atom of positive partial charge ($p+$), (d) polar atom of negative partial charge ($p-$), (e) hydrogen-bond acceptor (*hba*), (f) hydrogen-bond donor (*hbd*), and (g) aromatic atoms (*ar*).

Geometry optimizations of each of the pruned ligand−receptor complexes were performed using molecular mechanics (steepest descent and conjugate gradient methods) to obtain the lowest potential energy for each pruned enzyme−ligand complex. After the energy minimization step, molecular dynamics simulations (MDS) were carried out for each system to generate the conformational ensemble profile (CEP) of each pruned ligand−receptor complex. The applied MDS protocol is as follows:

(a) A total of 20 000 simulation steps were used, with the step size equal to 0.001 ps, that is, 20 ps MDS.

(b) The atomic coordinates of each complex's conformation and its energy were recorded every 20 simulation steps. Thus, a total of 1000 frames were obtained for constructing the CEP and corresponding grid cell occupancy descriptor (GCOD) trial descriptor set of each pruned HIV-1 PR−ligand complex.

(c) A simulation temperature of 310 K, which is the temperature used in the enzyme inhibition essay, was used.[8,9] The simulation temperature was kept constant by coupling the system to a temperature bath according to the method proposed by Berendsen and colleagues.[26]

(d) The molecular dielectric was held fixed at 3.5.

All MDSs were done using the Molsim package.[27]

Each CEP conformation of a given pruned receptor−ligand complex was placed in a reference grid cell lattice according to each alignment under consideration. In this work, the resolution of the grid cell lattice was defined as 1 Å. The normalized absolute occupancy of each grid cell by each IPE atom type over the CEP for the alignment provided the trial
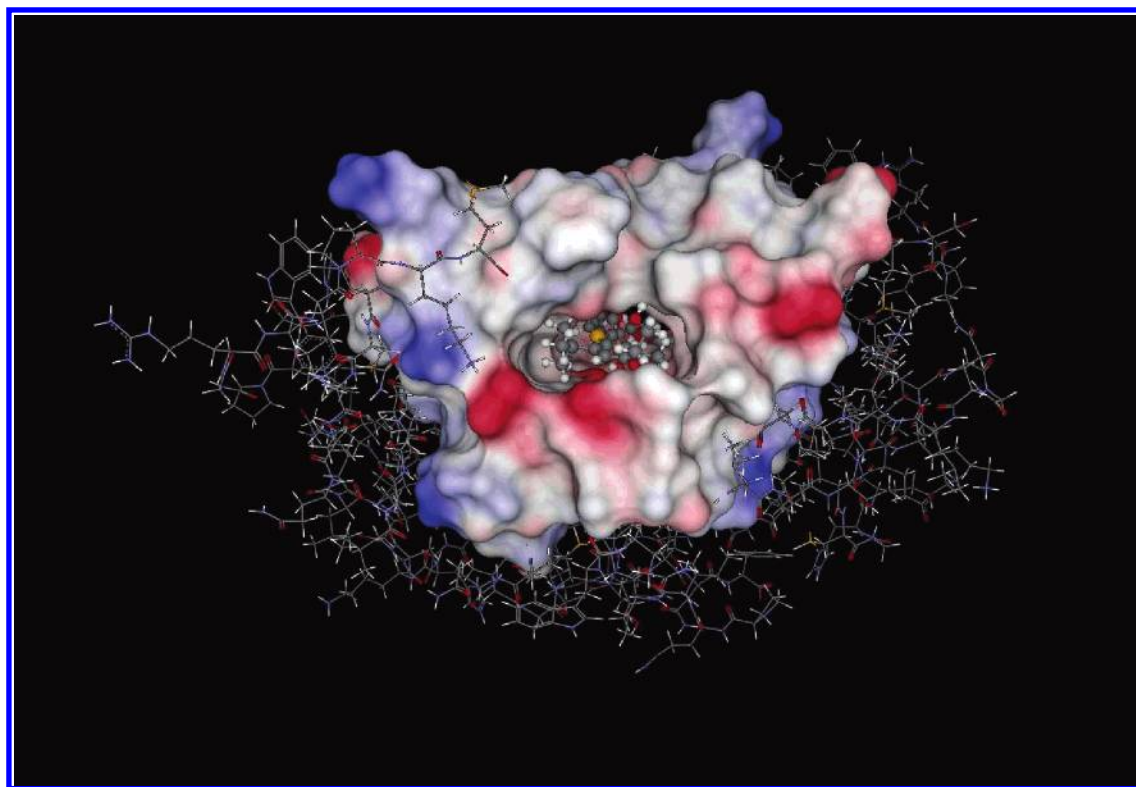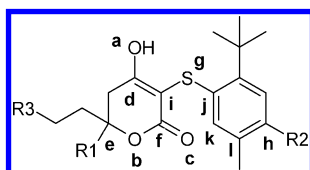
QSAR Analysis of 4-Hydroxy-5,6-dihydropyrones

J. Chem. Inf. Model., Vol. 46, No. 1, 2006 **349**

Figure 3. Complex of the pruned HIV-1 receptor model with ligand 38 docked. The entire protein, the 14 Å pruned region, and the ligand are shown in stick, surface, and ball-and-stick representations, respectively.

Table 4. Set of Trial Alignments Used in Constructing the RD 4D-QSAR Models

alignment	atoms
1	a, b, c
2	d, e, f
3	e, g, h
4	i, g, j
5	j, k, l

pool of RD 4D-QSAR independent variables referred to as the GCODs. Figure 4 shows, by way of example, a general reference grid cell space containing the pruned HIV-1 PR ligand 38 complex.

Partial least-squares (PLS)[28] regression was employed as a data reduction tool to identify the most highly weighted GCODs from this entire set of descriptors for the training set. The 200 most highly weighted PLS GCODs were used to form the trial descriptor basis set for subsequent QSAR model fitting and optimization.

Multiple linear regression and the genetic function approximation algorithm (GFA−MLR)[29] were employed in combination for fitting and optimizing the RD 4D-QSAR models. A mutation frequency during the crossover optimization cycle was set to 25% for each new generation of RD 4D-QSAR models. In this study, analogues were considered outliers if the difference in the predicted and observed activities (residue) exceeded 2.0 standard deviations from the mean. In this context, ligand number 8 was found to be the only outlier inhibitor.
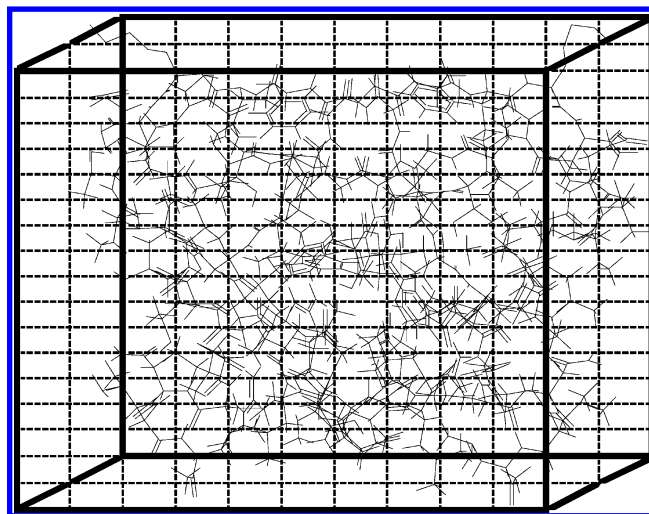

Figure 4. Schematic representation of the grid cell lattice superimposed upon the pruned HIV-1 PR−ligand 38 complex. To simplify the figure, only non-hydrogen atoms are shown.

In this study, the diagnostic measures used to characterize the RD 4D-QSAR models were the linear cross-correlation among a model's descriptors and measures of model significance, including the correlation coefficient (r^2) and leave-one-out cross-validation correlation coefficient (q^2).

The optimized one- to six-term RD-4D-QSAR models for the HIV-1 PR inhibitor training set of Table 1 are reported below.

$$pIC_{50}(M) = 6.97 + 3.33\,GC1(2, -3, 3, \textit{any})$$

$$(n = 39, r^2 = 0.57, q^2 = 0.51) \tag{1}$$

350 *J. Chem. Inf. Model., Vol. 46, No. 1, 2006*

SANTOS-FILHO AND HOPFINGER

$$\text{pIC}_{50}\,(\text{M}) = 6.56 + 2.81\,\text{GC1}(2, -3, 3, \textit{any}) + \\ 2.45\,\text{GC2}(5, 3, 4, \textit{np})$$

$$(n = 39, r^2 = 0.69, q^2 = 0.63) \qquad (2)$$

$$\text{pIC}_{50}\,(\text{M}) = 6.89 + 5.51\,\text{GC1}(11, 8, -2, \textit{any}) + \\ 3.44\,\text{GC2}(2, -3, 3, \textit{any}) - 3.57\,\text{GC3}(0, 5, 3, \textit{np})$$

$$(n = 39, r^2 = 0.76, q^2 = 0.65) \qquad (3)$$

$$\text{pIC}_{50}\,(\text{M}) = 6.69 + 3.79\,\text{GC1}(3, 4, 5, \textit{np}) - \\ 1.36\,\text{GC2}(0, -2, 5, \textit{np}) - 3.20\,\text{GC3}(0, 5, 3, \textit{np}) + \\ 3.64\,\text{GC4}(2, -3, 3, \textit{any})$$

$$(n = 39, r^2 = 0.86, q^2 = 0.80) \qquad (4)$$

$$\text{pIC}_{50}\,(\text{M}) = 6.09 + 2.54\,\text{GC1}(0, 12, 4, \textit{np}) + \\ 1.96\,\text{GC2}(2, 4, 5, \textit{np}) + 2.91\,\text{GC3}(2, 6, 9, \textit{any}) - \\ 4.93\,\text{GC4}(0, 5, 3, \textit{np}) + 3.36\,\text{GC5}(2, -3, 3, \textit{any})$$

$$(n = 39, r^2 = 0.86, q^2 = 0.78) \qquad (5)$$

$$\text{pIC}_{50}\,(\text{M}) = 6.60 + 1.83\,\text{GC1}(2, 6, 9, \textit{any}) - \\ 4.01\,\text{GC2}(0, 5, 3, \textit{any}) - 1.15\,\text{GC3}(0, -2, 5, \textit{any}) + \\ 3.27\,\text{GC4}(2, -3, 3, \textit{any}) + 0.73\,\text{GC5}(1, -2, 15, \textit{any}) + \\ 3.14\,\text{GC6}(3, 4, 5, \textit{np})$$

$$(n = 39, r^2 = 0.88, q^2 = 0.82) \qquad (6)$$

where pIC_{50} (M) is the molar HIV-1 PR inhibition potency and GCi(x, y, z, IPE) are the grid cell occupancy descriptors (GCODs), where x, y, and z are the Cartesian coordinates of the reference grid cell and IPE is the specific interaction pharmacophore element of the GCOD. For the statistical measures of model fit and significance, n is the number of sampled inhibitors and r^2 and q^2 are the regression and leave-one-out cross-validation correlation coefficients, respectively.

The cross-correlation matrix of the residuals of fit between pairs of the models given by eqs 1–6 was computed and is given in Table 5. The reason for determining the cross-correlation of the residuals of fit among the pairs of models is that two equivalent models will have near-identical residuals of fit, while a distinct pair of models should essentially have noncorrelated residuals of fit.[30] Table 5 indicates the following pairs of equivalent models: 1 and 2, 1 and 3, 4 and 5, 4 and 6, and 5 and 6. Using the Ockham's razor principle,[31] which proposes that one should make no more assumptions than needed, that is, the simplest QSAR models (with a lower number of terms) would be preferred, leads to the selection of two relevant models, namely, models 1 and 4.

Table 6 lists the observed and the test set predictions using models 1 and 4. The test set predictions using model 1 exhibit higher residuals of fit to the observed values than the corresponding predictions using model 4. However, the r^2 and q^2 values of model 4 are both markedly superior to those of model 1, and a multiple-site pharmacophore (multiple-GCOD model) is to be expected for the structural diversity of the training set inhibitors as opposed to a single pharmacophore site as expressed by model 1. Probably, the modest difference in predictive accuracy between models 1 and 4 for the two RD 4D-QSAR models, eqs 1 and 4, is due to the particular selected test set inhibitors and does not compromise the conclusion that eq 4 is the preferred RD

Table 5. Linear Correlation Matrix of Residuals of Fit for the Top One-, Two-, Three-, Four-, Five-, and Six-Descriptor-Term Models. High Cross-Correlations Are Shown in Italic, Bold Type

	model 1	model 2	model 3	model 4	model 5	model 6
model 1	1.00					
model 2	*0.86*	1.00				
model 3	*0.89*	0.64	1.00			
model 4	0.57	0.60	0.53	1.00		
model 5	0.57	0.58	0.61	*0.85*	1.00	
model 6	0.51	0.55	0.52	*0.94*	*0.88*	1.00

Table 6. Test Set Predictions of the pIC_{50} (M) Values Using the One- and Four-Term RD 4D-QSAR Models

test set (inhibitors)	observed inhibition potencies pIC_{50} (M)	predicted inhibition potencies pIC_{50} (M)		residues pIC_{50} (M)	
		model 1	model 4	model 1	model 4
41	9.03	9.17	9.91	−0.14	−0.88
42	9.46	8.77	8.47	0.69	0.99
43	9.24	9.17	9.66	0.07	−0.42
44	8.72	9.07	8.95	−0.35	−0.23
45	8.24	8.80	8.60	−0.56	−0.36
46	7.48	8.37	9.20	−0.89	−1.72
47	8.44	8.07	8.00	0.37	0.44
48	8.59	7.87	7.61	0.72	0.98

4D-QSAR model. Moreover, the GC1(2, −3, 3, *any*) descriptor of eq 1 is found as a descriptor with about equal weighting (regression coefficient) in all the other RD 4D-QSAR models.

Remarkably, the one-term model (eq 1), that is, a single pharmacophore site, captures a relatively large amount (57% of the variance) of the SAR information of the training set. That is the first time *one* pharmacophore site has been seen to be so significant in a 4D-QSAR model. Correspondingly, this pharmacophore site needs to be strongly considered in the design of new 4-hydroxy-5,6-dihydropyrone analogue inhibitors.

As the final step of the RD 4D-QSAR analysis, the "active" complex geometry of each HIV-1 PR−inhibitor pair was calculated. The active complex geometry was achieved by first identifying all complex states sampled for each complex that were within ΔE of the global minimum energy complex of the CEP. The ΔE was set as 2 kcal/mol in this study. Then, that single complex geometry within ΔE that predicts the greatest inhibition potency for a particular RD 4D-QSAR model is chosen as the active complex geometry with respect to that model. Obviously, each active complex geometry is composed of a distinct "active" conformation for both the inhibitor and the pruned enzyme model. Figures 5 and 6 show, respectively, the predicted "active" conformations for the most active inhibitor (number 36 in Table 1) and the least active inhibitor (number 30 in Table 1) as a function of the RD 4D-QSAR models given by eqs 1 and 4.

To explore if the RD 4D-QSAR models "capture" any type of induced-fit interactions between the enzyme and the inhibitors, two structural analyses of the HIV-1 pruned receptor model were done before and after binding interactions for inhibitors 30 and 36. First, the superimposed pruned receptor models before and after interaction with inhibitors 36 and 30 were determined and are shown in parts a and b of Figure 7, respectively. For inhibitor 36, the root-mean-square deviation (RMSD) between corresponding receptor atoms before and after ligand binding is 9.20 Å, and for ligand 30, the RMSD is 9.26 Å.
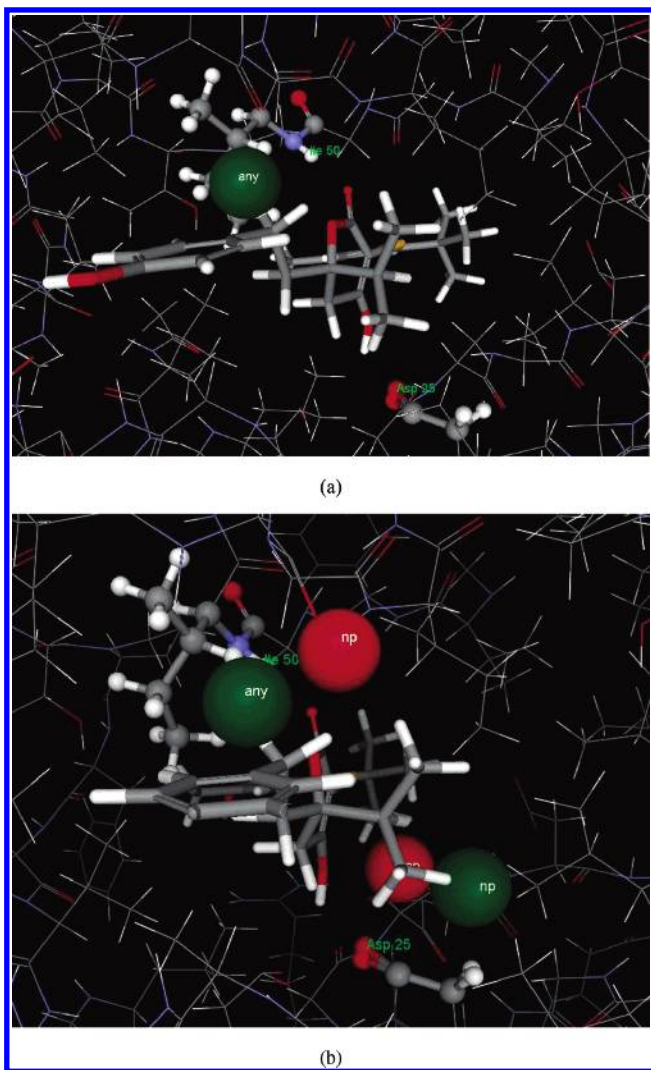
QSAR ANALYSIS OF 4-HYDROXY-5,6-DIHYDROPYRONES

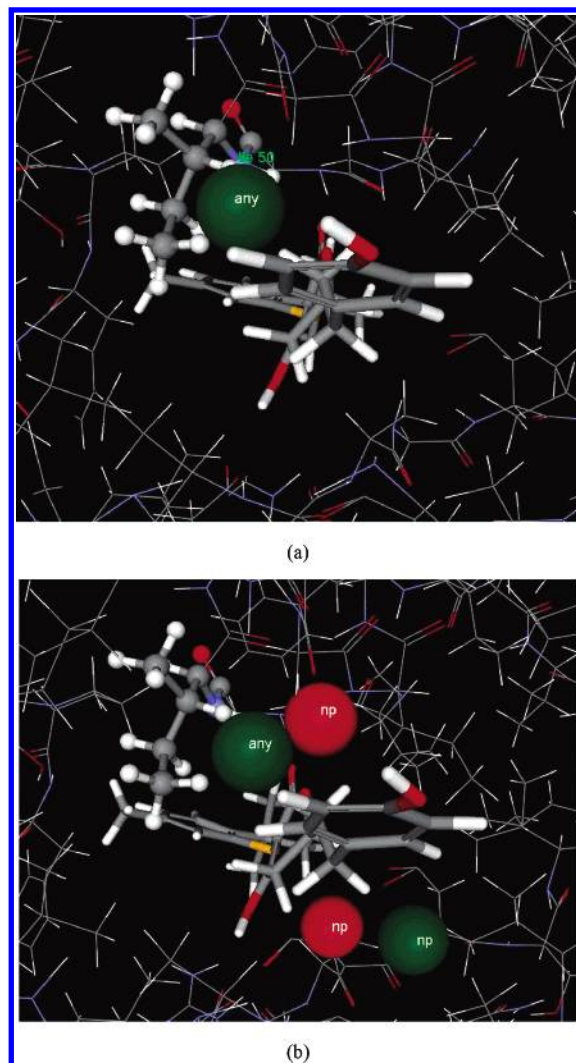*J. Chem. Inf. Model., Vol. 46, No. 1, 2006* **351**

Figure 5. Graphical representations of the HIV-1 PR−inhibitor 36 complex (in its "active" conformation) based upon the models given by (a) eq 1 and (b) eq 4. Inhibition-enhancing and inhibition-diminishing grid cells are shown, respectively, in green and red spheres.

Figure 6. Graphical representations of the HIV-1 PR−inhibitor 30 complex (in its "active" conformation) based upon the models given by (a) eq 1 and (b) eq 4. Inhibition-enhancing and inhibition-diminishing grid cells are shown, respectively, in green and red spheres.

Figure 8 shows the final bound geometry at the enzyme binding site due to the presence of inhibitor 36 as compared to that of inhibitor 30. The RMSD between the induced binding site geometries for these two inhibitors is 1.44 Å. Overall, these two sets of conformational changes seen in the receptor binding site geometry support specific ligand-sponsored induced fits upon ligand−receptor binding.

DISCUSSION

The shape of the predicted "active" conformations of the inhibitors using the RD 4D-QSAR models of this study are different from those found in preliminary RI 4D-QSAR analysis.[10] The inhibitors adopt "flatlike" active conformations using the RD 4D-QSAR models instead of the "foldlike" active conformations predicted by the RI 4D-QSAR models. Flatlike conformations are consistent in molecular shape with the geometries found in crystallographic (structural) works.[8] Thus, the explicit inclusion of the structure of the enzyme binding site is needed in a 4D-QSAR analysis to make a meaningful estimate of the active/bound conformation of an inhibition in the 4-hydroxy-5,6-

dihydropyrone analogue class. However, the RD 4D-QSAR models do not show the *simultaneous* formation of all of the hydrogen bonds identified in the crystallographic studies. Rather, the RD 4D-QSAR analysis suggests that both the number and types of hydrogen bonds formed between the ligand and receptor vary with the choice of ligand and, for a given ligand, may be broken and formed in a correlated manner over the time course of the bound ligand state. Examples of these ligand−receptor binding features are shown in Table 7 for inhibitors 1, 30, and 36.

The current study reveals a preponderance of nonpolar GCODs found in the RD 4D-QSAR models. This finding suggests that the variance in the inhibition potency across the training set of HIV-1 PR−4-hydroxy-5,6-dihydropyrone inhibitors is largely governed by hydrophobic interactions. This finding is consistent with our previous (receptor-independent) 4D-QSAR investigation,[10] and also with other structural investigations[8,9] (see Figure 1). Moreover, despite the time-dependent making and breaking of individual ligand−receptor hydrogen bonds, there is a baseline "net constancy" of hydrogen bonding and polar interaction across all the ligands with the receptor which serves as the "anchor interac-
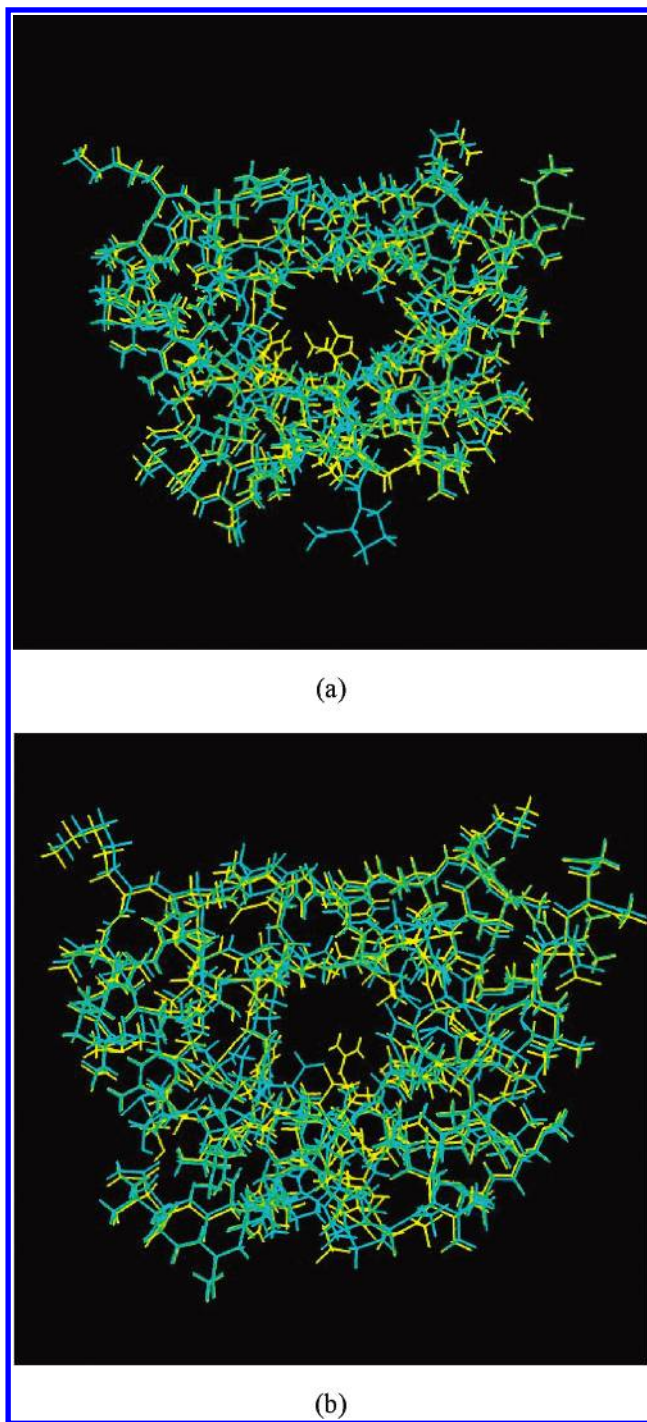
(a)

(b)

Figure 7. Superimposition of the HIV-1 pruned receptor model before and after its MDS binding interaction with (a) inhibitor 36 and (b) inhibitor 30. The pruned receptor modeling before interacting with the ligand is shown in yellow, and that at the end of the MDS binding interaction is shown green.

tion" to provide the relatively strong ligand−receptor inhibition for all of these inhibitors.

In model 4 (eq 4), three of the four GCODs are nonpolar. The remaining GCOD, GC1(2, −3, 3, *any*), also present in model 1 (eq 1), seemingly is related to inhibitor−HIV-1 PR interactions involving Ile-50 located near the entrance of the S1 hydrophobic pocket of the enzyme.

For eqs 1 and 4, the descriptors GC1(2, −3, 3, *any*) and GC4(2, −3, 3, *any*), respectively, positively contribute to the inhibitory potency of the inhibitors when occupied. It appears that this GCOD is capturing hydrophobic interactions
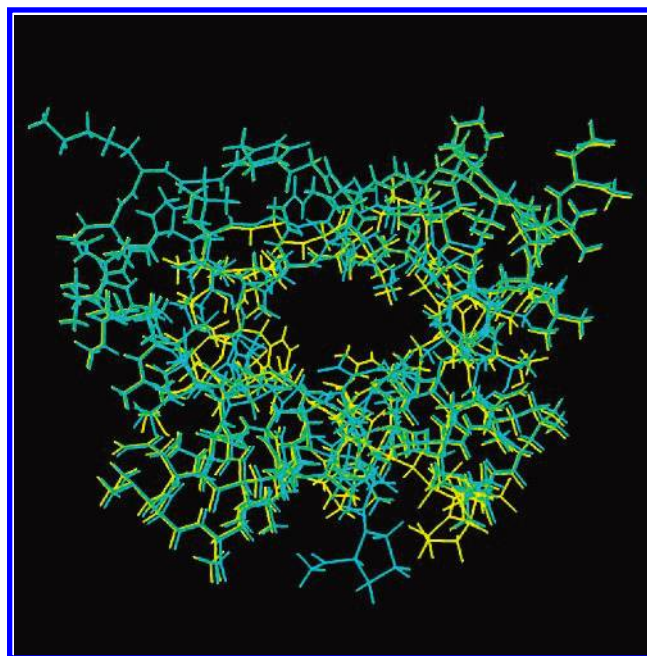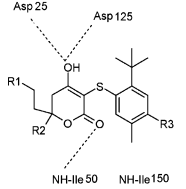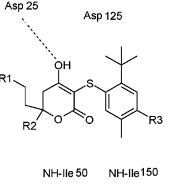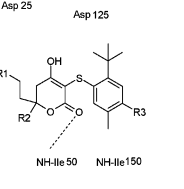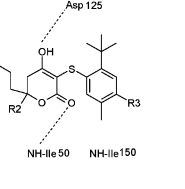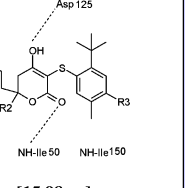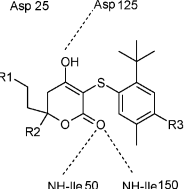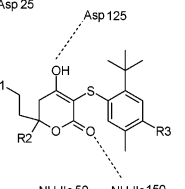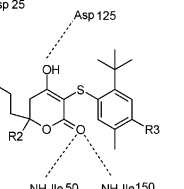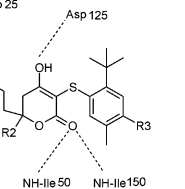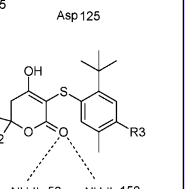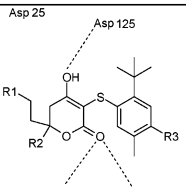
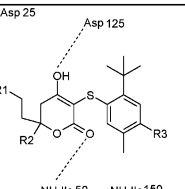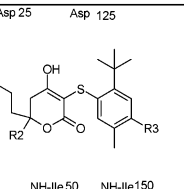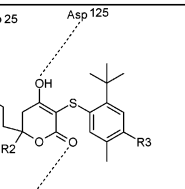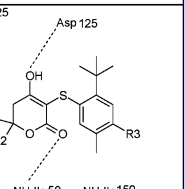**Figure 8.** Induced geometric effect of the inhibitor on the conformation of the pruned HIV-1 model receptor. The lowest-energy MDS conformations of the pruned model receptor as a function of its interactions with inhibitors 30 and 36 are shown in yellow and green, respectively.

involving the ethyl spacer found between the aromatic and the dihydropyrone ring, and this GCOD would be related to any HIV-1 PR−inhibitor hydrophobic interactions that occur in the S2 pocket of the enzyme (Figure 1).

In model 4, besides GC4(2, −3, 3, *any*), described above, three other GCODs—GC1(3, 4, 5, *np*), GC2(0, −2, 5, *np*), and GC3(0, 5, 3, *np*)—are found and shown in Figures 5b and 6b. GC1 is a biological-activity-enhancing pharmacophore site descriptor related to the hydrophobic interactions between inhibitors and the S2′ pockets of the enzyme, whereas GC2 and GC3 are both inhibition-potency-diminishing GCODs. These two pharmacophore site descriptors seem to be present in the model to capture the drawback of possibly having interfering nonpolar groups located in regions where polar or hydrogen-bond interactions can occur and stabilize the complex. GC2, for example, is relatively near the hydroxyl group of the dihydropyrone ring. This may be an indication that a nonpolar group located in GC2 may compromise the ability of the hydroxyl group to form a hydrogen bond to one of the aspartic acid residues (Asp-25 and Asp-125) of the enzyme. An interesting finding in the current work is the absence of a hydrogen bond to an aspartic acid residue for the bound state of inhibitor 30, a relatively weak inhibitor. In the case of the most-active inhibitor (inhibitor 36), however, a strong hydrogen bonding to the aspartic acid residue is observed.

A similar argument can be applied to the pharmacophore site of GC3, but with respect to the ability of the lactone moiety to hydrogen bond to a backbone NH of Ile-50 or Ile-150 of the enzyme. Probably, the presence of a relatively bulky group, such as the isopropyl substituent, near the lactone moiety would restrict ligand−receptor hydrogen bonding in this region of the complex and, consequently, could be "facilitating" the formation of a hydrogen bond to the aspartic acid residues of the enzyme. This could be a

Table 7. Variations in Both the Number and Type of Hydrogen Bonds Made and Broken by Some Representative Inhibitors with Members of the Receptor Binding Site over the Course of the RD 4D-QSAR MD Simulations[a]

[a] A dotted line indicates that a hydrogen bond is formed between the indicated chemical groups.

354 *J. Chem. Inf. Model., Vol. 46, No. 1, 2006*

SANTOS-FILHO AND HOPFINGER

way to "induce" specific ligand–receptor binding modes that could be tested in a future investigation.

The RD 4D-QSAR analysis of HIV-1 PR–inhibitor SARs seems to provide a range of specific ligand–receptor binding information and, consequently, makes a considerable improvement over our previous RI 4D-QSAR analysis. First, the use of the structural geometry of the enzyme leads to very significant QSAR equations, as given by their r^2 and q^2 measures, and the test set predictions are remarkably good, even using a one-term model (eq 1). There is also a tendency to be able to generate better models with fewer descriptor terms than those found for the corresponding RI 4D-QSAR models. Interestingly, models with more than six terms were not more significant than the four- through six-term models presented here.

The preponderance of nonpolar GCODs in all of the optimized models is in accordance with previous predictions.[8,9] The validity and worth of these RD 4D-QSAR models could be further explored by virtual high-throughput screening (VHTS). A subsequent VHTS study would not only provide more hints about the HIV-1 PR–4-hydroxy-5,6-dihydropyrone analogues but would also perhaps lead to proposing new classes of potential HIV-1 PR inhibitors.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Janeway, C. A., Jr.; Travers, P.; Walport, M.; Shlomchik, M. J. *Immunobiology: the immune system in health and disease*, 6th ed.; Garland Science Publishing: New York, 2005.

(2) Barry, M.; Gibbons, S.; Back, D.; Mulcahy, F. Protease inhibitors in patients with HIV disease. *Clin. Pharmacokinet.* **1997**, *32*, 194–209.

(3) Vigoroux, C.; Gharakhanian, S.; Salhi, Y.; Nguyen, T. H.; Adda, N.; Rozenbaum, W.; Capeau, J. Adverse metabolic disorders during highly active retroviral treatment (HAART) of HIV disease. *Diabetes Metab.* **1999**, *25*, 383–392.

(4) Sommadossi, J. P. HIV protease inhibitors: pharmacologic and metabolic distinctions. *AIDS (London, U. K.)* **1999**, *13* (Suppl. 1), S29–S40.

(5) Erickson, J. W. The not-so-great escape. *Nat. Struct. Biol.* **1995**, *2*, 523–529.

(6) Mascolini, M. Protease inhibitors and nucleoside combinations, or, Alice in Gaithersburg. *J. Int. Assoc. Physicians AIDS Care* **1996**, *4*, 23–32.

(7) Ohtaka, H.; Freire, E. Adaptative inhitibors of the HIV-1 protease. *Prog. Biophys. Mol. Biol.* **2005**, *88*, 193–208.

(8) Vara Prasad, J. V. N.; Boyer, F. E.; Domagala, J. M.; Ellsworth, E. L.; Gajda, C.; Hamilton, H. W.; Hagen, S. E.; Markoski, L. J.; Steinbaugh, B. A.; Tait, B. D.; Humblet, C.; Lunney, E. A.; Pavlovsky, A.; Rubin, J. R.; Ferguson, D.; Graham, N.; Holler, T.; Hupe, D.; Nouhan, C.; Tummino, P. J.; Uromov, A.; Zeikus, E.; Zeikus, G.; Gracheck, S. J.; Saunders, J. M.; VanderRoest, S.; Brodfuehrer, J.; Iyer, K.; Sinz, M.; Gulnik, S. V.; Erickson, J. W. Nonpeptidic HIV protease inhibitors possessing excellent antiviral activities and therapeutic indices. PD 178390: A lead HIV protease inhibitor. *Bioorg. Med. Chem.* **1999**, *7*, 2775–2800.

(9) Hagen, S. E.; Domagala, J. M.; Gajda, C.; Lovdahl, M.; Tait, B. D.; Wise, E.; Holler, T.; Hupe, D.; Nouhan, C.; Uromov, A.; Zeikus, G.; Zeikus, E.; Lunney, E. A.; Pavlovsky, A.; Gracheck, S. J.; Saunders: J. M.; VanderRoest, S.; Brodfuehrer, J. 4-Hydroxy-5,6-dihydropyrones as inhibitors of HIV protease: the effect of heterocyclic substituents at C-6 on antiviral potency and pharmacokinetic parameters. *J. Med. Chem.* **2001**, *44*, 2319–2332.

(10) Santos-Filho, O. A.; Hopfinger, A. J. The 4D-QSAR paradigm: application to a novel set of nonpeptidic HIV protease inhibitors. *Quant. Struct.-Act. Relat.* **2002**, *21*, 369–381.

(11) Pan, D.; Tseng, Y.; Hopfinger, A. J. Quantitative structure-based design: formalism and application of receptor-dependent RD-4D-QSAR analysis to a set of glucose analogue inhibitors of glycogen phosphorylase. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1591–1607.

(12) Thaisrivongs, S.; Skulnick, H. I.; Turner, S. R.; Strohbach, J. W.; Tommasi, R. A.; Johnson, P. D.; Aristoff, P. A.; Judge, T. M.; Gammill, R. B.; Morris, J. K.; Romines, K. R.; Chrusciel, R. A.; Hinshaw, R. R.; Chong, K. T.; Tarpley, W. G.; Poppe, S. M.; Slade, D. E.; Lynn, J. C.; Horng, M. M.; Tomich, P. K.; Seest, E. P.; Dolak, L. A.; Howe, W. J.; Howard, G. M.; Schewende, F. J.; Toth, L. N.; Padbury, G. E.; Wilson, G. J.; Shiou, L.; Zipp, G. L.; Wilkinson, K. F.; Rush, B. D.; Ruwart, M. J.; Koeplinger, K. A.; Zhao, Z.; Cole, S.; Zaya, R. N.; Kakuk, T. J.; Janakiraman, M. N.; Watenpaugh, K. D. Structure-Based Design of HIV Protease inhibitors: sulfonamide-containing 5,6-dihydro-4-hydroxy-2-pyrones as nonpeptidic inhibitors. *J. Med. Chem.* **1996**, *39*, 4349–4353.

(13) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(14) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.* **1997**, *119*, 10509–10524.

(15) Venkatarangan, P.; Hopfinger, A. J. Prediction of ligand–receptor binding free energy by 4D-QSAR analysis: application to a set of glucose analogue inhibitors of glycogen phosphorylase. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1141–1150.

(16) Albuquerque, M. G.; Hopfinger, A. J.; Barreiro, E. J.; Alencastro, R. B. Four-dimensional quantitative structure–activity relationship analysis of a series of interphenylene 7-oxabicycloheptane oxazole thromboxane A₂ receptor antagonists. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 925–938.

(17) Ravi, M.; Hopfinger, A. J.; Hormann, R. E.; Dinan, L. 4D-QSAR analysis of a set of ecdysteroids and a comparison to CoMFA modeling. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1587–1604.

(18) Krasowski, M. D.; Hong, X.; Hopfinger, A. J.; Harrison, N. L. 4D-QSAR analysis of a set of propofol analogues: mapping binding sites for an anesthetic phenol on the GABA_A receptor. *J. Med. Chem.* **2002**, *45*, 3210–3221.

(19) Hong, X.; Hopfinger, A. J. 3D-pharmacophores of flavonoid binding at the benzodiazepine GABAA receptor site using 4D-QSAR analysis. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 324–336.

(20) Lambert, M. H. Docking conformationally flexible molecules into protein binding sites. In *Practical applications of computer-aided drug design*; Charifson, P. S., Ed.; Marcel Dekker: New York, 1997; pp 243–303.

(21) Jensen, F. *Introduction to computational chemistry*; John Wiley & Sons: West Sussex, U. K., 1999.

(22) Cramer, C. J. *Essentials in computational chemistry*, 2nd ed.; John Wiley & Sons: West Sussex, U. K., 2004.

(23) *HyperChem Reference Manual*; Hypercube, Inc.: Gainesville, FL, 1996.

(24) Dewar, M. J. S.; Thiel, W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907.

(25) Tokarski, J. S.; Hopfinger, A. J. Constructing protein models for ligand–receptor binding thermodynamic simulations: an application to a set of peptidometic renin inhibitors. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 779–791.

(26) Berendsen, H. J. C.; Postman, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(27) Doherty, D. C. *MOLSIM Package*, version 3.2; The ChemBats21 Group, Inc.: Lake Forest, IL.

(28) Glen, W. G.; Dunn, W. J., III; Scott, D. R. Principal components analysis and partial least squares. *Tetrahedron Comput. Methodol.* **1989**, *2*, 349–354.

(29) Rogers, D. G.; Hopfinger, A. J. Application of genetic function approximation to quantitative structure–activity relationships and quantitative structure–property relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854–866.

(30) *4D-QSAR Package,* version 2.0; The ChemBats21 Group: Lake Forest, IL.

(31) Ockham's Razor is a principle attributed to the 14th century English logician and Franciscan friar William of Ockham (ca. 1285–1349) that forms the basis of methodological reductionism, also called the principle of parsimony or law of economy. In its simplest form, Ockham's razor states that one should make no more assumptions than needed. The principle is most often expressed as "Entia non sunt multiplicanda preaeter necessitatem", or "Entities are not to be multiplied beyond necessity".