

## Structure Based Prediction of Binding Affinity of Human Immunodeficiency Virus-1 Protease Inhibitors

Santosh S. Kulkarni and Vithal M. Kulkarni\*

Pharmaceutical Division, Department of Chemical Technology, University of Mumbai,  
Matunga, Mumbai 400 019, India

Received March 10, 1999

A series of computations were performed to derive a strategy for the prediction of binding affinities of non-peptidic human immunodeficiency virus-1 (HIV-1) protease inhibitors. This paper describes the development of a 3D quantitative structure–activity relationship (3D-QSAR) methodology by using receptor information of HIV-1 protease. The docking and molecular dynamics simulations were performed on a model ligand/enzyme complex to optimize the variables involved in the generation of ligand/enzyme models. The protonation scheme of the active site aspartic acid residues of HIV-1 protease was derived from a computational study. The active site aspartate is monoprotonated with a proton placed on the OD1 atom of the ASP B25. This protocol of docking and molecular dynamics (MD) simulation was then used to derive the ligand–enzyme complexes of the molecules used in the present study. The molecular mechanics interaction descriptors were calculated from these ligand/enzyme models. A partial least squares (PLS) method was used to derive a linear correlation between the interaction descriptors and the biological activity. A good correlation was observed when the change in the energy of the ligand upon complex formation and the electrostatic contributions to the solvation energy of the ligand were included in the QSAR analysis. A highest cross-validated  $q^2$  value of 0.649 was observed. This model had a conventional  $r^2$  of 0.725, and when this model was used to predict the activity of the external test set, it produced a  $r^2_{\text{pred}}$  of 0.761. The total interaction energy was partitioned into interactions in different subsites and interactions with each of the amino acid residues of the enzyme. The PLS analysis using these descriptors helped to identify the important interactions which can be exploited for the design of HIV-1 protease inhibitors.

### INTRODUCTION

Structure based drug design (SBDD) is an approach for designing ligands using the information of the receptor geometry.<sup>1</sup> Availability of the three-dimensional structure of the biologically important therapeutic target proteins at atomic resolution has assisted in designing newer and potent analogues.<sup>2</sup> In SBDD it was hoped to generate compounds that are more active than those previously synthesized and tested. There has been a pressing need in the modern medicinal chemistry program to predict the binding affinity of the ligand before it is synthesized.<sup>3</sup> The predicted binding affinities can help to rank the ligand candidates. This information is essential for accepting or rejecting the design hypothesis and for the synthetic prioritization of the accepted ligand candidates. To assess the structural modification of the ligand in terms of its affinity, it is quite essential to understand the ligand receptor complex at the molecular level. A quantitative knowledge of the forces governing the interaction and their contributions toward these interactions is essential for successful prediction of binding affinity.<sup>4–6</sup> There are various methods developed to predict the binding affinities. Free energy perturbation (FEP) is a method for accurate prediction of binding affinity.<sup>7</sup> FEP by thermodynamic integration involves a nonphysical stepwise conversion between pairs of similar ligands in the free and bound states. In this method solvent molecules are treated explicitly. But large computational demand and the restricted change in the structures preclude its use in a priori prediction of binding affinity. The FEP is a promising technique with immense

potential in the future to be useful in a priori prediction of binding affinity.

Another method of affinity prediction is quantitative structure–activity relationships (QSARs).<sup>8</sup> In these methods, correlations are derived between the binding affinities and some physicochemical descriptors of a set of ligands. The thermodynamics of binding is represented in the form of these physicochemical descriptors. These correlations are based upon a fundamental assumption that a correlation exists between the enthalpy of binding for a congeneric series of ligands with similar size and flexibility with the physicochemical descriptors. The newer methods of QSAR, i.e., 3D-QSAR, and comparative molecular field analysis (CoMFA) in particular derive a correlation between the binding affinities and the energetic fields around the molecule.<sup>9–11</sup> These energetic fields have contributions from steric, electrostatic, and sometimes hydrophobic interactions. But these methods make little use of the receptor geometry or the structural information of ligand–receptor interactions.

The physicochemical information derived from the ligand–receptor interaction have long been used in deriving QSAR. The molecular mechanics energies of interaction between the ligand and receptor were used as independent variables. Purely force field based molecular mechanics methods are too simplistic for predictions of binding affinities. But subjecting these interactions to statistical analysis can help to reduce the noise due to force field potential inaccuracies. Also such analysis would help to identify functionally important interactions that can be further exploited in future

ligand design. With increasing use of the receptor geometry to calculate independent variables in QSAR it has been formally called receptor dependent quantitative structure–activity relationship (RD-QSAR).<sup>12</sup> A first such application was on DNA intercalating anticancer anthracyclines.<sup>13</sup>

A highly correlative 3D-QSAR model was derived from the intermolecular interaction energy of HIV-1 protease peptidomimetic inhibitors.<sup>14</sup> The RD-QSAR equation was then used for predicting the activity of the newly designed molecules. However, this analysis neglected the energetic consequences of ligand/receptor binding and the solvation/desolvation phenomenon. Recently a comparative binding energy analysis on the same dataset of HIV-1 protease inhibitors was reported.<sup>15</sup> This analysis contained all energetic terms and the continuum solvation model to account for the solvation and desolvation phenomenon. Highly correlative and predictive models were derived.

In another application of the comparative binding energy analysis Ortiz et al. have derived a RD-QSAR model on 26 inhibitors of human synovial fluid phospholipase A2.<sup>16</sup> The molecular mechanics energy terms were subjected to variable selection procedure during statistical analysis to select the energetic signals from the background noise. The statistical analyses were performed using the PLS method with the GOLPE variable selection procedure.

Hopfinger et al. have described a new RD-QSAR methodology called free energy force field (FEFF) formalism for 3D-QSAR analysis.<sup>12</sup> In this application all the enthalpic and entropic contributions to the ligand/receptor interactions in solvent were considered, and they were treated separately as independent variables in the QSAR development. They have used a genetic algorithm to derive an optimum model. This method was applied to a set of peptidomimetic renin inhibitors.

Marshall et al. have developed a hybrid approach, VALIDATE, to predict the binding affinity of novel ligands.<sup>17</sup> The enthalpy of binding was calculated from the molecular mechanics, and entropic estimations were performed heuristically from the complementary hydrophilic surface area. The QSAR models were derived using 51 crystallographic complexes representing a wide range of interaction size, type, activity, and physicochemical features. Statistical analyses were performed using partial least squares (PLS) and neural network analysis. This method predicted the activity for HIV-1 protease inhibitors satisfactorily but was not able to predict the binding affinities for thermolysin inhibitors.

In this paper we describe an empirical method for binding affinity prediction. In this application we have made use of a hybrid approach of flexible docking and simulations as described by Luty and Stouten to derive ligand receptor models.<sup>18,19</sup> A similar protocol was used initially to develop a model of the receptor to economize the computation and to derive a protonation state of the active aspartates of HIV-1 protease. The independent variables were calculated from the ligand/enzyme geometries. These variables were then used for PLS analysis to develop the RD-QSAR models. This method was applied to a series of nonpeptidic human immunodeficiency virus-1 protease inhibitors. HIV protease is a prototype enzyme involved in the processing of gag and gag-pol gene polyproteins. Inhibition of HIV protease is known to produce an immature and noninfectious virus. Therefore HIV-1 protease is an attractive target for the

treatment of acquired immunodeficiency syndrome (AIDS).<sup>20,21</sup> Various inhibitors of HIV-1 protease are in human clinical use, and some are in advanced clinical trials. Availability of large amounts of structural data on HIV-1 protease has helped to design novel ligands.<sup>22,23</sup> Using a mass screening of corporate database warfarin and phenprocoumon were identified as non-peptide HIV-1 protease inhibitors.<sup>24</sup> These ligands were subsequently used as non-peptidic templates to design novel inhibitors in the structure based drug design program.

## METHODOLOGY

**1. Ligands.** The structures and HIV-1 protease inhibitory activities of the ligands are given in Table 1. These ligands were reported by an Upjohn research group as novel nonpeptidic HIV-1 protease inhibitors developed in a SBDD program.<sup>24–32</sup> We have modeled 40 ligands considering the structural diversity and the range of activity. Out of the 40 ligands 31 were used as a training set and 9 ligands were set aside as an external test set. The test set compounds were selected from the whole dataset by considering the fact that these molecules were designed on the basis of the information derived from the molecules of the training set. This ideally could serve to check the predictiveness of the models in a priori prediction approach.

Thaisrivongs et al. have reported a cocrystal structure of a non-peptidic inhibitor in complex with HIV-1 protease (pdb entry 2upj).<sup>26</sup> All molecular modeling studies were performed on these structural data. The structure building and model generation calculations were performed using the molecular modeling tools within the Insight-II package,<sup>33</sup> with CFF91 force field implemented in the Discover program for energy calculations. The QSAR analyses were performed using SYBYL 6.22<sup>34</sup> running on a Silicon Graphics Indigo2 (R4400) computer.

The structures of the inhibitors were built using the coordinates of a model inhibitor as it appears in the crystal structure. The novel molecules were built using standard bond lengths and angles. The conformational analysis of the inhibitors was performed by rotating the nonterminal bonds by 15° steps for 360°. The low-energy conformers were minimized using the conjugate gradient minimizer until a convergence criterion of 0.001 kcal/mol Å was achieved. Atom centered charges for all the inhibitors were assigned from a single point semiempirical AM1 calculation in MOPAC (version 6.0).<sup>35</sup>

**2. Enzyme.** All hydrogen atoms were added to the enzyme molecule. Protonation of the active site aspartates was modeled as described below. No crystallographic water molecules were used in these calculations. The structure optimizations were performed using CFF91 force field with a nonbonded cutoff of 10 Å and a distance dependent dielectric constant of 4. Initially only hydrogen atoms were minimized while the heavy atoms were kept fixed. The constraints on the heavy atoms were released stepwise during a series of minimizations. The energy of the whole complex was minimized until a convergence criterion of 0.01 kcal/mol Å was achieved.

**3. Docking of the Ligands.** We have applied a flexible docking algorithm as implemented in the affinity module in Insight-II.

The docking procedure in affinity involves first placing the ligand in the active site of the enzyme. This initial model

**Table 1.** Structures and Biological Activities of the Training and Test Set Compounds

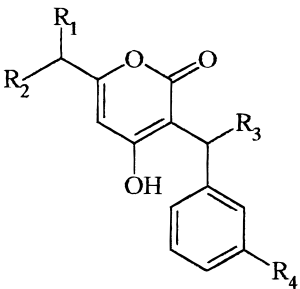
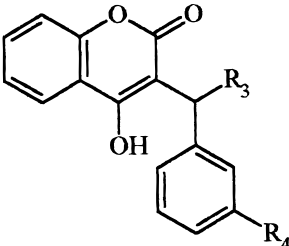
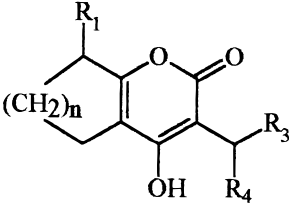
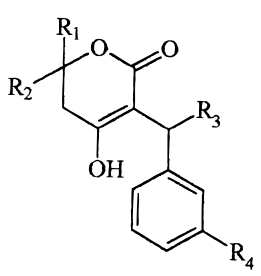
compd no.	R1	R2	R3	R4	p <i>K<sub>i</sub></i> <sup>a</sup>	ref
Training Set						
						
1	PhCH <sub>2</sub> –	H	–Et	H	0.3010	24
2	PhCH <sub>2</sub> –	–Et	–Et	H	1.8539	24
3	H	H	– <i>c</i> -propyl	–NHSO <sub>2</sub> (3-Me-5-imidazolyl)	1.8539	29
4	– <i>c</i> -propyl	– <i>c</i> -propyl	– <i>c</i> -propyl	–NHSO <sub>2</sub> (4-CNPh)	2.2676	29
5	PhCH <sub>2</sub> –	–Et	– <i>c</i> -propyl	–NHSO <sub>2</sub> (3-Me-5-imidazolyl)	3.2840	29
6	–CH <sub>2</sub> - <i>c</i> -propyl	–CH <sub>2</sub> - <i>c</i> -propyl	– <i>c</i> -propyl	–NHSO <sub>2</sub> (3-Me-5-imidazolyl)	3.0757	29
7	PhCH <sub>2</sub> –	–Et	– <i>c</i> -propyl	–NHSO <sub>2</sub> (4-MeOph)	1.6576	29
8	PhCH <sub>2</sub> –	–Et	– <i>c</i> -propyl	–NHCO(CH <sub>2</sub> ) <sub>2</sub> NHCOO- <i>t</i> -bu	1.3872	26
9	PhCH <sub>2</sub> –	–Et	– <i>c</i> -propyl	–NHCOCH(CH <sub>2</sub> (4-imidazolyl))NHCOO- <i>t</i> -bu	2.3979	26
10	–CH <sub>2</sub> - <i>c</i> -propyl	–CH <sub>2</sub> - <i>c</i> -propyl	– <i>c</i> -propyl	–CH <sub>2</sub> SO <sub>2</sub> (2-pyridinyl)	2.6383	30
11	–CH <sub>2</sub> - <i>c</i> -propyl	–CH <sub>2</sub> - <i>c</i> -propyl	– <i>c</i> -propyl	–NHSO <sub>2</sub> (2-pyridinyl)	3.1805	30
12	–CH <sub>2</sub> - <i>c</i> -propyl	–CH <sub>2</sub> - <i>c</i> -propyl	– <i>c</i> -propyl	–CH <sub>2</sub> SO <sub>2</sub> (1-Me-2-imidazolyl)	2.5086	30
						
13			–Et	H	0.0969	24
14			–CH <sub>2</sub> COCH <sub>3</sub>	H	–1.4771	24
15			–Et	–NHCOCH <sub>2</sub> NHCOO- <i>t</i> -bu	0.7959	26
16			–Et	–NHCO(CH <sub>2</sub> ) <sub>3</sub> NHCOO- <i>t</i> -bu	–0.2041	26
17			–Et	–NHCOCH(CH <sub>2</sub> ) <sub>2</sub> NCOO- <i>t</i> -bu	–0.5563	26
18			–Et	–CONHCH <sub>2</sub> (2-benzimidazolyl)	–0.6434	26
						
19	H		–Et	–Ph	$\frac{n}{2}$ –0.0414	25
20	H		–Et	–Ph	4 1.1249	25
21	H		– <i>c</i> -propyl	–Ph	4 1.8861	25
22	H		–Ph	–Ph	4 0.7447	25
23	H		– <i>c</i> -propyl	– <i>c</i> -propyl	4 1.2441	25
24	PhCH <sub>2</sub> –		–Et	–Ph	4 0.5751	27
25	<i>n</i> -pr–		– <i>c</i> -propyl	–Ph	4 1.5850	27
26	H		– <i>c</i> -propyl	– <i>m</i> (NHCOPh)Ph	4 1.3767	27
27	H		– <i>c</i> -propyl	– <i>m</i> (NHCO-cinnamyl)Ph	4 0.3279	27
28	H		– <i>c</i> -propyl	– <i>m</i> (NHCO(CH <sub>2</sub> ) <sub>2</sub> NHCOO- <i>t</i> -bu)Ph	4 2.3979	27
29	H		– <i>c</i> -propyl	– <i>m</i> (NHCOO-Et)	4 1.7212	27
30	H		– <i>c</i> -propyl	– <i>m</i> (NHSO <sub>2</sub> Ph)	4 2.5229	28
31	H		– <i>c</i> -propyl	– <i>m</i> [NHSO <sub>2</sub> (4-CN-Ph)]	4 3.1539	28

Table 1 (Continued)

compd no.	R1	R2	R3	R4	$pK_i^a$	ref
<div style="text-align: center;">  </div>						
32	–Et	–Ph	–Et	H	0.7959	31
33	Ph(CH <sub>2</sub> ) <sub>2</sub> –	<i>n</i> -pr–	–CH=CH <sub>2</sub>	H	1.7959	31
34	Ph(CH <sub>2</sub> ) <sub>2</sub> –	<i>n</i> -pr–	–CH=CH <sub>2</sub> Ph	H	1.3010	31
35	<sup>t</sup> bu–	<sup>t</sup> bu–	–Et	H	1.1249	31
36	Ph(CH <sub>2</sub> ) <sub>2</sub> –	<i>n</i> -pr–	–Et	–NHSO <sub>2</sub> Ph	1.7695	32
37RR	Ph(CH <sub>2</sub> ) <sub>2</sub> –	<i>n</i> -pr–	–Et	–NHSO <sub>2</sub> (3-Me-5-imidazolyl)	4.2218	32
38SS	Ph(CH <sub>2</sub> ) <sub>2</sub> –	<i>n</i> -pr–	–Et	–NHSO <sub>2</sub> (3-Me-5-imidazolyl)	3.0000	32
39RR	Ph(CH <sub>2</sub> ) <sub>2</sub> –	<i>n</i> -pr–	–Et	–NHSO <sub>2</sub> (5-CF <sub>3</sub> -2-pyridinyl)	5.0969	32
40SS	Ph(CH <sub>2</sub> ) <sub>2</sub> –	<i>n</i> -pr–	–Et	–NHSO <sub>2</sub> (5-CF <sub>3</sub> -2-pyridinyl)	3.6576	32

<sup>a</sup>  $pK_i$  is expressed as negative logarithm of  $K_i$  ( $\mu$ M).

was minimized to remove bad contacts in the structure. Then the ligand was moved by a random combination of translation, rotation, and torsional changes. This random move of the ligand samples both the orientational and conformational spaces of the ligand with respect to the enzyme active site. Subsequently the energy of the randomly moved structure was checked. If the energy was within the specified tolerance parameter of the previous structure, it was saved and subjected to energy minimization. The final minimized structure was accepted or rejected on the basis of the energy and similarity criteria. We have used Metropolis criterion for the energy check. In this criterion structures whose energy was lower than that of the last accepted structure or whose Boltzmann factor is greater than a random number between 0 and 1 are accepted. The structure similarity was checked by calculating the root mean square distances of the ligand between the current structure and structures found so far. The accepted ligand/enzyme structure was then subjected to molecular dynamics simulations to sample the conformational space around each local minimum energy structure.

The starting bound conformation and orientation of the model ligand **8** in the crystal structure complex provided a good starting model to dock the rest of the ligands. The HIV-1 protease enzyme contains 3121 atoms including protons. Comparison of the apo and ligand bound enzyme structure shows a large change in the position of the atoms in the flap region. However, several cocrystal structures of HIV-1 protease with different peptidic and non-peptidic ligands have shown that there is no large conformational change in the active site atoms of the enzyme. Hence a reduced size receptor model centered around the active site was used for all docking and model generation studies. The residues that lie in a 10 Å radius of the model ligand were used in the calculations. Compound **8** was used as a model ligand to define the reduced receptor site model because this ligand is the largest and, hence, could include more residues of the enzyme. The ligands were docked by superimposing the atoms of ligands to compound **8** in the active site. This

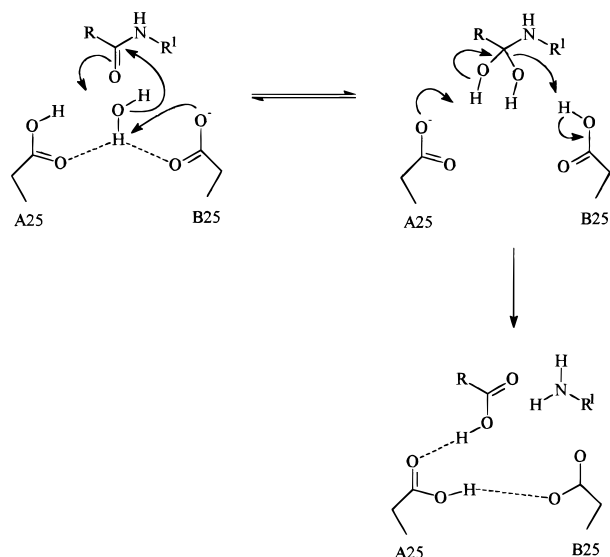
initial model was then subjected to affinity docking. The ligand moves and explores the enzyme active site. The accepted ligand/enzyme model was then subjected to molecular dynamics simulations. A series of computations were performed on the model ligand **8**/enzyme complex to optimize the docking and molecular dynamics parameters.

The temperature of the simulation was evaluated by subjecting the ligand **8**/enzyme complex model to variable temperature MD simulations for 50 ps. MD simulations at 200, 300, and 1000 K were performed on a ligand/enzyme model with a similar starting geometry. The MD simulation had a time step of 1 fs, and the trajectories of the ligand/enzyme complex were saved every 100 steps. The potential energy of the complex and the root-mean-square deviations (RMSD) of the ligand and moving atoms (active site) of the enzyme from the crystal structure were monitored throughout the MD simulations.

The time span required for successful docking of the ligands was evaluated by subjecting the ligand **8**/enzyme model for 100 ps MD simulation at 300 K. A time step of 1 fs and an update frequency of 100 was used for MD simulation. A movie of the trajectory was played to visualize the movement of the ligand in the enzyme active site during MD simulation. The trajectory of the whole MD run was evaluated by dividing the whole trajectory into windows of 2 ps each. The potential energy of the ligand **8**/enzyme complex and the RMSD of the ligand and moving atoms of the enzyme from the crystal structure were calculated in each of these 2 ps windows.

The initial position of the ligand in the active site (starting geometry of ligand **8**/enzyme complex) can have an effect on the final model generated. To study the effect of initial geometry on the successful docking of the ligand, the model ligand **8**/enzyme complex was subjected to repeated MD simulations at 300 K for 50 ps starting from different positions. The RMSD of ligand **8**/enzyme models from X-ray crystal structure geometry was calculated.





**Figure 1.** Proposed amide bond hydrolysis mechanism of aspartic protease.

This docking protocol was then validated by docking other HIV-1 protease inhibitors. These model ligands had peptidic inhibitors U-75875, MVT-101, and A-76928 and non-peptidic inhibitor XK-263. The X-ray coordinates of these ligands with HIV-1 protease were used as reference for all comparisons.

This docking and MD simulations protocol was then applied to all the ligands to generate the ligand/enzyme model required for our present study.

**4. Protonation States of the Active Site Aspartates.** The active site aspartates of HIV-1 protease assume an opposite role in general acid–base catalysis. In this, unprotonated aspartyl residue deprotonates the catalytic water and the protonated aspartyl residue polarizes the carbonyl oxygen. Therefore the catalytically competent form of the enzyme is “monoprotonated”, but the exact position of the proton is still elusive.

The proposed hydrolysis mechanism of HIV-1 protease is depicted in Figure 1. This mechanism is based on the three-dimensional structural data and from the kinetic measurements of pH dependence on catalysis.<sup>36–38</sup> The Asp B25 polarizes the lytic water molecule by general base deprotonation, and Asp A25 elicits general acid protonation on the carbonyl carbon of the substrate. This leads to the formation of the tetrahedral intermediate. This is followed by deprotonation of a hydroxy group of the tetrahedral intermediate by Asp A25 and simultaneous protonation of the amide nitrogen by Asp B25, leading to the hydrolysis products. The binding of a substrate can be affected by the positions of protons in the active site. Also, ligand binding may alter the protonation state of the active site aspartates. This emphasizes the importance of proton assignments in the active site during ligand/enzyme modeling. The placement of the protons on the enzyme and ligand is critical to the alignment of an inhibitor in the active site, and this will influence the calculation of free energy of binding.

Many experimental studies and molecular dynamics simulations of HIV-1 protease inhibitor complexes have been performed to assign the protonation state of the active site aspartates. In a series of studies on pH rate behavior and

solvent kinetic effect it was observed that one of the aspartyl groups is protonated and the other is unprotonated when an inhibitor containing hydroxyethylene transition state isostere binds.<sup>37,38</sup> Whereas in the case of inhibitors with reduced amide transition state isostere preferential binding occurs when both the aspartates are unprotonated. In an analysis of the X-ray crystal structure of HIV-1 protease and hydroxyethylene isostere containing inhibitor it has been proposed that only one of the aspartates is protonated.<sup>39</sup>

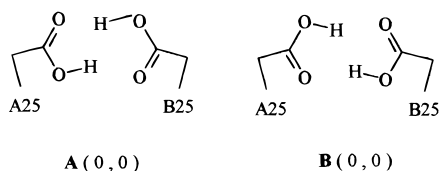
In a computational approach using free energy perturbation (FEP) theory the relative binding affinities of *S* and *R* isomers of peptidomimetic inhibitor U85548E was evaluated under different protonation states of the active site aspartates.<sup>40</sup> Molecular mechanics energies were computed from a MD simulation of inhibitor/enzyme complex models. The proton was placed successively on all four oxygens. This study suggests that catalytically relevant enzyme is monoprotonated with the proton placed on Asp B25. On the contrary a diprotonated model was proposed from the MD simulation of ligand enzyme complex of U85548E.<sup>41</sup> In the same study the protonation state for an inhibitor containing reduced amide transition state isostere MVT-101 was also studied. A dianionic model was found to be favorable on the basis of geometric comparison with the crystal structure.

Free energy perturbation theory was used to discriminate the protonation states of the active site aspartates for the hydroxyethylamine containing inhibitor JG-365.<sup>42</sup> A monoprotonated model with a proton on Asp B25 yields the calculated free energy difference in the range of those observed experimentally.

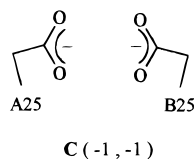
Using the NMR chemical shift titrations and H/D isotope shift measurements,  $pK_a$  of all Asp and Glu residues in HIV-1 protease in complex with a cyclic urea based non-peptidic inhibitor were calculated.<sup>43</sup> The results suggest that both the active site aspartates are protonated and provide a strong network of hydrogen bonds with the diols of the inhibitor molecule.

The above discussion shows that there is no clear understanding regarding the choice of the protonation states. Each ligand presents a different environment to the active site aspartate. It is quite essential to assign the correct protonation states to the active site aspartates to calculate the binding energy. In this regard we have performed a series of calculations to derive a suitable protonation state for the ligands used in the present study. A series of mono-, di-, and unprotonated states of the active site aspartates were considered in complex with a representative model ligand. The protonation states are shown in Figure 2. All the protonation models were created by successively placing the protons on the heteroatoms. The initial locations of the protons of the aspartates were optimized by 300 steps of conjugate gradient minimization. To allow the atoms of the active site and the ligand to explore the energy space, we have performed a molecular dynamics simulation at 300 K for 50 ps. The motion of the active site atoms and the ligands under the influence of protonation states were studied by calculating the RMS deviation for the active site atoms from the crystal structure. The intermolecular interaction energies and intramolecular energies of the enzyme and ligand were monitored throughout the MD run. The torsion angles of the active site aspartates and the atomic distances between the active site atoms of the enzyme and ligand were monitored.

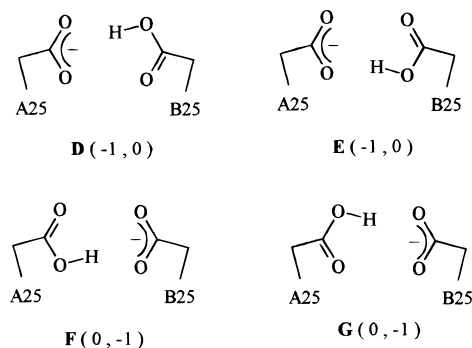
## a) Diprotonated



## b) Unprotonated



## c) Monoprotonated



**Figure 2.** Protonation models of the active site aspartic acid residues considered in the present study.

The suitable protonation model was selected on the basis of the results obtained from the MD runs.

**5. Calculation of Interaction Descriptors.** The ligand receptor interactions involve many aspects. The structural and energetic complementarity aspects of the ligand and receptor as high complementarity between the ligand and receptor lead to high binding affinity. The aspect of conformational change in the ligand and/or receptor upon complex formation as large conformational changes in the ligand and receptor away from their low-energy conformations leads to a decrease in binding affinity. And the thermodynamic aspects describe the transfer of the ligand and receptor and loss of rotational and translational entropy of the ligand. For successful prediction of binding affinities, these aspects must be quantitated and their contributions to the free energy of binding scaled.

**Structural and Energetic Complementarity.** Structural and energetic complementarity of the ligand and receptor is an absolute prerequisite for specific binding. The ligand/enzyme interaction energy is calculated on a representative conformation of the ligand/enzyme complex using the molecular mechanics force field. The total interaction energy was partitioned into each interaction module (described below) and for each individual amino acid. These each interaction energy terms were used as separate variables for QSAR analysis.

**Conformational Energy of the Ligand and Enzyme.** The energetic consequences of the ligand and enzyme upon complex formation were calculated as the strain energy of the ligand and enzyme. The ligand molecule was extracted

from the complex and minimized to a local minima. Similarly side chains of the active site amino acids of the enzyme were minimized after isolating the ligand molecule. In the case of HIV-1 protease it is known that the enzyme undergoes a considerable change in its flap conformation upon ligand binding.<sup>23</sup> But during strain energy calculations for the enzyme, we have ignored these motions. The strain energy for the ligand and enzyme was defined using the formula.

$$E_{\text{strain}} = E_{\text{bound}} - E_{\text{free}}$$

where  $E_{\text{bound}}$  is the energy of the ligand or enzyme in the ligand/enzyme complex conformation.  $E_{\text{free}}$  is the energy of the ligand or enzyme in its nearest local minima.

**Solvation and Desolvation Energies.** The solvation and desolvation energies were calculated by solving the nonlinear Poisson–Boltzmann equation by finite difference method.<sup>44</sup> All these calculations were performed using the Delphi module in Insight-II. The Delphi program utilizes the finite difference solution to the nonlinear Poisson–Boltzmann equation to calculate the electrostatic potential in and around macromolecules.<sup>45</sup> This calculation can be used to determine the electrostatic contribution to the solvation and binding energies.<sup>46</sup> In the finite difference approximation, the molecule is mapped onto a three-dimensional cubical grid, with a grid spacing of unit size. The Poisson–Boltzmann equation was then used at each grid point to calculate the potential at each grid point.

The ligand/enzyme models obtained from docking studies were used as input for the Delphi program. The interior of the protein was considered a dielectric constant of  $\epsilon = 4$ , the surrounding solvent was considered a dielectric constant of  $\epsilon = 80$  with ionic strength of 0.145 M, and for vacuum calculation the dielectric constant was set to 1. The cubic grid was centered around the molecule with a grid spacing of 0.5 Å. The solvent accessible surfaces were calculated using a spherical probe of 1.4 Å radius. The potentials at the grid points delimiting the box were calculated analytically by treating each charge atom as a Debye–Huckel sphere. The electrostatic contribution to the solvation free energy was calculated by subtracting the total electrostatic free energy in vacuo from the total electrostatic free energy in water. The electrostatic contribution to the desolvation free energy of the ligand was calculated by subtracting the total electrostatic free energy in water from the total electrostatic free energy in protein environment. To calculate the total electrostatic free energies in different conditions, a series of calculations were performed with identical grid mappings with the same interior dielectric ( $\epsilon = 4$ ). The exterior dielectric was set to 1, 4, and 80 to simulate vacuum, protein, and aqueous environments, respectively.

**6. Partitioning of the Interaction Energy and Variable Selection.** The total ligand/enzyme interaction energy in the representative ligand/enzyme complex model is subdivided into six different interaction modules.<sup>47</sup> The amino acid residues in the 10 Å reduced receptor model were divided into six subsites. Table 2 defines the amino acid residues used in each site. The ligand molecule was considered as a single fragment during the interaction energy calculations. The interaction energies between the inhibitor and each interaction module were treated as separate descriptors in QSAR.

The total interaction energy was also partitioned on a per residue basis to study the effect of partitioning of energy

**Table 2.** Partitioning of Active Site Amino Acid Residues into Each Interaction Modules

S3	S2	S1	S1'	S2'	S3'
A7, A8, A9, A10, A21, A22, A23, A82, B53	A50, A51, B28, B29, B30, B31, B32, B45, B46, B47, B48, B52, B54, B56, B76, B78, B79, B87, B88, B90	A24, A25, A33, A34, A80, A81, A83, A84, A85, A86, B26, B27, B49	A26, A27, B24, B25, B33, B34, B80, B81, B83, B84, B85, B86, A49	A28, A29, A30, A31, A32, A45, A46, A47, A48, A52, A54, A56, A76, A78, A79, A87, A88, A90, B50, B51	A53, B7, B8, B9, B10, B21, B22, B23, B82

terms on the final QSAR results. The amino acid residues in the 10 Å reduced receptor model were used for interaction energy calculation. This analysis contains many energy terms, and some energy terms may not contribute to the differences in binding and may therefore add “noise” to the matrix of energy terms. To reduce the noise level, variables with a standard deviation below a certain value were deleted. But this type of variable selection can also lead to loss of a useful information “signal” in the energy matrix. Hence a series of analyses were performed in which variables with standard deviations of the column of less than 0.1, 0.2, 0.3, ... 1.0 kcal/mol were deleted sequentially. The optimum value of standard deviation for variable selection was chosen on the basis of the results obtained from the PLS analysis.

All energy calculations were performed using the custom written commands from the biosym command language (BCL) protocol.

**7. Partial Least Squares Analysis.** A stable QSAR can be obtained from the highly underdetermined matrix using the partial least squares (PLS) method.<sup>48</sup> PLS “solves” the equation by performing rotations of energy terms that maximize the linear correlation between the independent (energy terms) and the dependent (activities) variables. The QSAR models were developed with leave one out (LOO) cross-validation analysis. In this method, one compound is deleted and the model is developed. The activity of the deleted compound was predicted. This analysis was repeated until all the compounds in the training set are predicted once. The QSAR results depict the predictiveness of the model rather than the fitness of the model. The results from the cross-validation analysis are expressed as  $q^2$ ,

$$q^2 = 1 - \frac{\text{PRESS}}{\sum(Y - Y_{\text{mean}})^2}$$

where  $\text{PRESS} = \sum(Y - Y_{\text{pred}})^2$ .

The optimum number of components is the one that gives the lowest standard error of prediction. The model was also validated using an external prediction set. The best QSAR model was used to predict the activity of nine compounds that were not used in the model generation. The results from the predictions are reported as  $r^2_{\text{pred}}$ ,

$$r^2_{\text{pred}} = \frac{\text{SD} - \text{PRESS}}{\text{SD}}$$

where SD is the sum of the squared deviation between the biological activities of the test set and the mean of the training set molecules and PRESS is the sum of the squared deviations between the predicted and actual activity values for every molecule in the test set.

## RESULTS AND DISCUSSION

Various HIV protease inhibitors have been designed using the structural information of HIV-1 protease. Generally these

ligands contained a structural fragment to mimic the tetrahedral intermediate of the transition state. Pioneering work done by the Upjohn and Dupont researchers has helped to realize the non-peptidic inhibitors of HIV-1 protease.<sup>49</sup> Using an analogues lead, we have designed a series of non-peptidic HIV-1 protease inhibitors.<sup>50</sup> With this experience in structure based drug design of HIV-1 protease inhibitors we felt the need to develop a computational methodology to forecast the activity of the newly designed molecules. Such a computational method would be of help in accessing a new design hypothesis and in the prioritization of synthetic targets.

We have selected a dataset of 40 compounds reported by the Upjohn group as HIV-1 protease inhibitors. These compounds represent a uniform dataset with quantitative activity data for in vitro enzyme inhibition. This biological data can be directly correlated with the free energy of binding. The ligands selected contained various structural fragments appended to a common core structure. The biological activity values of the selected compounds were on the order of 6.5 log units. The QSAR models were derived on a training set of 31 compounds. Nine compounds were set aside as an external prediction set to validate the derived QSAR models. The test set molecules appear in the later part of the SBDD program; i.e., they were designed using the information derived from the training set ligands. This creates a unique set to validate the present QSAR models.

In SBDD docking calculations are needed to predict how new, hypothetical compounds will bind to the protein.<sup>51</sup> Currently various methods are known to dock a conformationally flexible ligand into the active site of protein. The algorithms involved in these methods use different techniques such as molecular dynamics, metropolis Monte Carlo, Monte Carlo minimization, genetic algorithms, distance geometry, and tree searching to explore the conformational space of the ligand and enzyme during docking calculations.<sup>52</sup> In the present study we have used a flexible docking method to derive the ligand/enzyme geometry. A docking calculation should be able to produce a crystal geometry of the ligand/protein complex with reasonable accuracy. Also docking calculation should be sufficiently fast to be used in a priori prediction of binding affinities. Successful docking of the ligand would depend on the optimum choice of the variables involved in docking calculations. The conformational flexibility of each enzyme and ligand could be different for each particular type of ligand protein complexes. We have performed preliminary calculations to optimize the docking variables for HIV-1 protease/non-peptide inhibitor complex.

The temperature of the simulation affects the exploration of the conformational space of the ligand/enzyme complex. The temperature of the MD simulation should allow the ligand/enzyme complex to explore all the possible conformational space and to traverse the energy barriers in different local minimas. Also the temperature of MD simulation should not allow the ligand/receptor complex to drift away

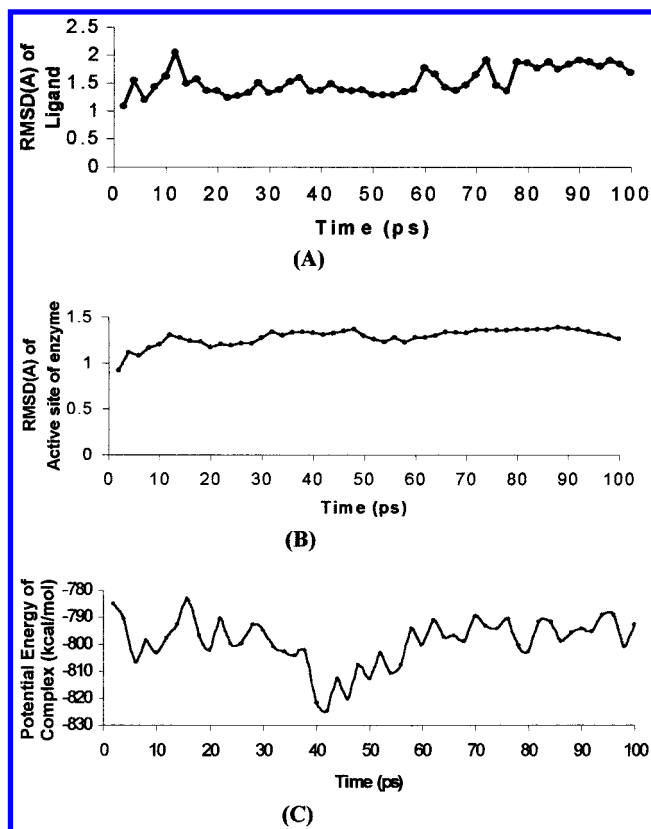


**Table 3.** Effect of Temperature of Simulation on Root Mean Square Deviations (RMSD) of Ligand and Active Site of Enzyme

temp (K)	ligand			active site		
	max	av	std dev	max	av	std dev
200	1.246	0.766	0.157	1.082	0.926	0.068
300	3.881	1.729	0.754	1.370	1.144	0.105
1000	21.892	10.878	4.574	5.707	4.470	0.770

from the crystal structure. A series of trial MD runs on a model ligand was performed at various temperatures starting from the same geometry for 50 ps. RMS deviations of each frame from the crystal structure were calculated. Table 3 depicts the mean and standard deviation of RMS deviation observed for the ligand and the moving atoms of the enzyme. The average RMS deviation of the ligand over the MD simulation shows that at 200 K the ligand stays close to the crystal structure. At 300 K, the ligand moves and explores the possible conformational space. The high RMS deviation at 1000 K suggests that at this temperature ligand and active site atoms move away from the crystal structure geometry. Analysis of the movie of the trajectory shows that at 300 K the ligand and the active site of the enzyme explore the conformational space. The potential energies of trajectories from the 1000 K MD simulation run were in higher energy than those from the 300 K trajectories. Overall 300 K MD simulation provides a reasonable choice of temperature for obtaining low-energy trajectories.

A 100 ps MD simulation on a model ligand was performed at 300 K to determine the time space required for MD simulation. Analysis of the movie of the trajectory shows that the ligand moves and explores the possible conformations in order to find a sterically allowed binding with the receptor. The RMS deviations of the ligand and active site atoms of the enzyme were calculated for each frame using the crystal structure as the reference frame. A high RMS deviation values of the ligand show that the ligand explored wide conformational space at the expense of moving away from the crystallographic position. The whole trajectory was divided into 50 windows of 2 ps each, and an average RMS deviation of ligand in each 2 ps window was calculated. A plot of the average RMSD versus time (Figure 3A) shows that the ligand moves quite away from the crystallographic position in the first 12 ps with an average RMSD in that window of 2.0461 Å, and then it comes close to the crystal structure. The ligand maintains a low RMSD until 55 ps after which the ligand starts to drift away from the crystallographic positions. A similar plot (Figure 3B) of RMSD of the active site atoms of the enzyme in a 2 ps window against time shows that, for first 12 ps, the active site atoms have a high RMS deviation; as simulation proceeds it decreases and remains stable for some time and again increases. During this time the side chains of the amino acid residues of the active site change their conformation to have a favorable interaction with the moving ligand. The active site atoms move close to the crystallographic positions at 50 ps and remain there until 65 ps, and later they move away from the crystallographic position. A similar analysis on the potential energy of each frame was performed. An average value in each 2 ps window was calculated. A plot of average potential energy versus time (Figure 3C) shows that there is a wide range in potential energy of the complex but around 50 ps the complex could achieve a low-energy state of the whole

**Figure 3.** Effect of the time space of the simulations.

100 ps trajectory. Also these low-energy frames had low RMSD for ligand and enzyme from the crystal structure. Hence all further calculations were performed with MD simulations for 50 ps as this time space allows a reasonable conformational sampling with low potential energy of the ligand/enzyme complex.

The initial geometry for all these calculations was similar to the crystal structure. For new ligands the choice of initial position of the ligand could affect final ligand/enzyme geometry. Hence a series of MD simulations were performed on a model ligand **8** starting from different initial positions. A MD simulation for 50 ps at 300 K was performed. This computation was performed from five different initial positions. The RMS deviations of the ligand and enzyme complex from different runs was calculated using the crystal structure as a reference frame. These five different initial positions were generated by slight rotation or translation of the ligand in the active sight. All the runs produced a low RMSD (<1.50 Å) with close interaction energy values, suggesting that initial position/placement of the ligand has no effect on the final model generated.

Thus the docking protocol involved placing the ligand in the active site and then subjecting the ligand/enzyme complex for minimization followed by MD simulation at 300 K for 50 ps. This docking protocol was validated by docking a series of HIV-1 protease inhibitors. The RMSD of heavy atoms of the complexes from the crystal structure are shown in Table 4. Also the interaction energies of the docking models were compared with the values observed in the X-ray crystal structure. Low values of RMSD in all cases suggest that the present protocol for docking can produce the ligand/receptor geometry of sufficient accuracy for binding energy estimations.



**Table 4.** RMSD of the Complexes after Docking

ligand	$E_{\text{total}}^a$	$E_{\text{total}}^b$	RMSD (Å)
A76928	-169.25	-167.80	1.301
MVT101	-153.39	-150.14	1.422
U75875	-131.72	-129.24	1.112
XK263	-120.47	-121.77	1.194
compd <b>8</b>	-110.21	-112.32	1.194

<sup>a</sup> Total interaction energy between the ligand and enzyme as observed in the X-crystal structure. <sup>b</sup> Total interaction energy between the ligand and enzyme as observed in the docked model.

A computational approach was used to derive a suitable protonation state (scheme) of the active site aspartates. Analysis of the trajectories from the MD simulations of various protonation models suggests that some models show more stability than others. A particular model was considered stable on the basis of how closely the structure obtained from the simulation resembles the available experimental geometry. The ligand molecule is flexible and could adopt a suitable conformation based on the protonation state of the active site aspartates. Hence RMSDs of the ligands and active site aspartates from the crystal structure were compared. Also the potential energy of the ligand/enzyme complex and the intermolecular interaction energy between ligand and enzyme and intramolecular energies of ligand and enzyme were compared. Experimental geometries of most of the aspartyl protease show that the aspartic acid residues are coplanar. Hence coplanarity of the aspartates was monitored by measuring the torsion angles of the active site aspartates throughout the MD simulations. A large variation in torsion angles during MD simulation indicates instability of the protonation model. Key ligand/enzyme interatomic distances were monitored during MD simulations. Average values and the standard deviation around the average values were also calculated. A large difference from the crystal structure in these interatomic distances could indicate unsuitability of that particular protonation model. During docking and subsequent MD simulation the ligand moves in the active site and explores the favorable interactions. In some cases the ligands move away from the binding groups. And large movements of the ligand away from these interacting groups of the enzyme indicate an instability of the model.

The results obtained (Table 5) suggests that only few protonation models are stable for these types of ligands. The monoproteination models with a proton on the Asp B25 OD1 and Asp A25 OD1 (models E and F) and the diproteination model (model A) have less total energy of the complex. The intermolecular interaction energies suggest that the unprotonated model (model C) is more favorable since this model had the lowest interaction energy. However, the intramolecular energy of the enzyme was highest for this model. A good interaction was observed for model C between the ligand and enzyme at the expense of high enzyme strain energy. The intermolecular interaction energies also suggest that the monoproteinated models (models E and F) are more favorable than the diproteinated model (model A). The monoproteinated models (models E and F) are similar since HIV-1 protease is a symmetrical enzyme. But the ligands used in the present study are nonsymmetrical; therefore only one protonation model would be preferred. A comparison of the RMSD of the heavy atoms of the complex (Table 5) shows that the protonation model E produces the experi-

mental geometry more accurately than model F. Also the comparison of the RMSD of the ligand shows that model E is more accurate in producing the experimental geometry than model F. Monitoring the torsion angles over the trajectory space shows that only a few models could maintain the coplanarity of the aspartates. (Table 6). Models E and F could maintain the coplanarity whereas other models (models C and D) showed large fluctuations in the torsion angles. Model E produced the key interatomic distances with reasonable accuracy (Table 6). This model also could maintain the intermolecular hydrogen bonds. Considering the structural and energetic criteria accumulated over the MD simulations, the monoproteinated model (model E) with a proton on the Asp B25 OD1 was selected as the preferred protonation state of the active site aspartates for the present set of ligands.

The ligand/enzyme models were generated by docking the ligands on the monoproteinated enzyme model. These ligand/enzyme models were then used to calculate the energy descriptors. The intermolecular interaction energy between the ligand and each interaction module was calculated. The interaction energy between the ligand and each amino acid residue of the reduced size receptor model was also calculated. The strain energies of the enzyme and ligand were calculated. The solvation and desolvation energies were calculated using the nonlinear Poisson–Boltzmann equation. All the energy values were then exported to the SYBYL QSAR program for statistical analysis. The interaction descriptors of the ligands are tabulated in Table 7.

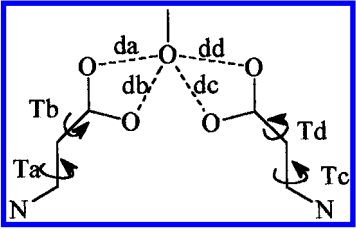
**Correlation of the Biological Activities with the Interaction Descriptors.** The statistical results from the PLS analysis are reported in Table 8. The correlation derived using the total interaction energy as a sole descriptor gave a poor model. This model was found to be less predictive internally ( $q^2 = 0.226$ ) and also correlative ( $r^2 = 0.315$ ). A good correlation was reported in the literature with the total interaction energy alone and the biological activity for a series of HIV-1 protease peptidic inhibitors.<sup>22</sup> The molecular mechanics interaction energy may not be a sole descriptor to describe the variation in the biological activity of the HIV-1 protease inhibitors. The poor correlation observed in the present study could be due to the underlying dataset. Our dataset had a broad range of activity and wide structural variation. Some additional regressors have to be derived to include the information of the whole dataset. The ligands used in the present study had some flexible derivatives. The ligand strain energies were calculated and included in the data table. The QSAR model derived showed a good internal predictivity with  $q^2 = 0.649$  and correlative property  $r^2 = 0.725$ . This is in contrast with what is observed in the literature in which inclusion of such descriptors did not improve the significance of the model. When this model was used to predict the activity of the external test set, it performed well. The external predictivity measure,  $r^2_{\text{pred}}$ , of 0.761 was observed. The enzyme strain energy descriptor was added to the QSAR table (data not shown). But no improvement in the correlations was observed. In an attempt to improve the significance of the QSAR models still further, we have included the desolvation and solvation energies in the QSAR table. Inclusion of desolvation energy of the ligand decreased the  $q^2$  by 0.19 unit, but the external predictivity of the model was slightly improved. The inclusion electro-

**Table 5.** Energetics of the Complex Structures of Compound **8** Possessing Different Protonation States

protonation model	$E_{\text{total}}^a$	$E_{\text{enz}}^b$	$E_{\text{lig}}^c$	$E_{\text{int}}^d$	RMSD		
					heavy <sup>e</sup>	aspartates <sup>f</sup>	ligands <sup>g</sup>
A	-1207.52	-1014.84	-89.78	-103.40	0.575	0.517	0.807
B	-1193.88	-996.88	-90.92	-106.61	0.498	0.227	1.217
C	-1126.26	-925.58	-80.88	-120.31	0.516	0.333	1.523
D	-1202.30	-1008.33	-81.00	-113.46	0.660	1.159	2.384
E	-1205.24	-1015.18	-78.26	-112.32	0.484	0.570	0.552
F	-1209.98	-1018.57	-78.71	-113.22	0.636	0.625	1.182
G	-1203.70	-1015.42	-83.38	-105.44	0.535	0.912	1.284

<sup>a</sup> Total energy of the complex. <sup>b</sup> Intramolecular energy of the enzyme in the complex. <sup>c</sup> Intramolecular energy of the ligand in the complex. <sup>d</sup> Intermolecular interaction energy of the ligand and enzyme. <sup>e</sup> RMS deviation of the heavy atoms of the enzyme. <sup>f</sup> RMS deviation of the two active site aspartates. <sup>g</sup> RMS deviation of the ligand.

**Table 6.** Average Interatomic Distances and Torsions of the Ligand/Enzyme Complex Models with Different Protonation Models Obtained from MD Simulations

protonation model								
	da	db	dc	dd	Ta	Tb	Tc	Td
A	3.612	2.996	3.556	3.021	-167.14	110.16	-173.66	-163.50
B	2.925	3.210	3.502	3.050	-170.40	164.15	-165.49	141.39
C	3.301	3.090	3.529	3.820	-172.08	148.61	-173.45	-96.38
D	3.722	2.795	4.062	3.707	-169.83	-90.38	-164.85	-105.86
E	2.991	2.573	3.410	3.039	-167.98	162.93	-167.48	165.47
F	4.046	3.335	3.642	2.769	-164.49	153.28	-173.72	-158.36
G	4.347	3.801	4.267	2.993	-173.17	90.15	-173.73	-162.07

static contribution to the solvation energy of the ligands in the QSAR analysis shows good statistical significance. This QSAR model exhibited a good external predictivity with a  $r^2_{\text{pred}}$  of 0.781. The QSAR equation obtained was

$$\text{p}K_i = -4.62 - (0.09)E_{\text{total}} - (0.31)E_{\text{lig}}^{\text{strain}} + (0.05)E_{\text{lig}}^{\text{solv}} \quad (1)$$

$$N = 31, \quad q^2 = 0.630, \quad r^2 = 0.727, \quad F = 23.971$$

The correlations derived so far were from the total or global interaction energy of the ligand/enzyme. But in SBDD it is worthwhile to know the functionally important interactions to guide the iterative design cycle. The interaction energy was partitioned into six different interaction energies for each binding subsite. The correlations derived from these energies shows an improvement over those from total interaction energies. The cross-validated  $q^2$  was 0.340, and non-cross-validated  $r^2$  was 0.488. The relative contributions of each energy value show that variance in biological activity is mainly explained by interactions at the S1 and S2 subsites. This does not mean that interactions in other subsites are not important, but, in the present dataset of the compounds, interactions at the S1 and S2 subsites mainly govern the activities of the compounds. The S3 and S3\_1 subsite interaction energies contributed to only 1.8 and 0.8%, respectively. The correlations derived after exclusion of S3 and S3\_1 subsite energies improved the predictions  $q^2$  (0.346) slightly. On considering the fact that the test set compounds may have some ligands, which can have interaction in the S3 and S3\_1 pockets, we have not excluded these

two interaction terms from the QSAR table. Additional regressors were added to improve the correlations. The addition of the change in the energy of the ligand upon binding ( $E_{\text{lig}}^{\text{strain}}$ ) shows a large improvement in the statistical significance of the QSAR model. This model has an internal predictivity ( $q^2$ ) of 0.566 and  $r^2$  of 0.748. This model performed reasonably well when used to predict the activity of the compounds from the test set. The  $r^2_{\text{pred}}$  of 0.682 was observed for the prediction of the binding affinity for the 15 compounds. The inclusion of the electrostatic contributions to the solvation energy of the ligand in the QSAR analysis shows an improvement in the external predictivity ( $r^2_{\text{pred}} = 0.781$ ). The relative contributions of each energy terms show that only a few subsite interactions are important. The change in the energy of the ligand upon complex formation contributes to a large extent for the QSAR model. The QSAR equation obtained was

$$\text{p}K_i = -5.54 - (0.18)S1 - (0.03)S2 - (0.04)S3 - (0.05)S1\_1 - (0.08)S2\_1 - (0.03)S3\_1 - (0.28)E_{\text{lig}}^{\text{strain}} + (0.03)E_{\text{lig}}^{\text{solv}} \quad (2)$$

$$N = 31, \quad q^2 = 0.515, \quad r^2 = 0.740, \quad F = 25.593$$

To find the exact amino acid residues that contribute for the binding energy, we have partitioned the total interaction energy into each residue level interaction terms. The interaction energy with all the 84 amino acid residues of the reduced site receptor model were calculated and entered as separate columns in the QSAR table. The correlation derived using

**Table 7.** Interaction Energy Descriptors of the HIV-1 Protease Inhibitors Used in This Study

comp no.	$E_{\text{total}}^a$	$E_{\text{lig strain}}^b$	$E_{\text{lig solv}}^c$	$E_{\text{lig desolv}}^d$	predicted activity <sup>e</sup>
1	-71.68	2.22	-16.08	6.05	0.40
2	-78.27	0.21	-16.10	6.14	1.36
3	-83.31	0.28	-21.87	7.50	1.75
4	-98.30	4.47	-20.05	6.97	2.80
5	-107.97	5.37	-24.57	8.93	3.05
6	-98.44	3.51	-21.40	7.39	2.52
7	-101.48	4.57	-21.55	7.67	2.48
8	-112.32	9.46	-27.88	9.99	1.26
9	-118.48	4.42	-40.04	14.44	2.55
10	-111.32	3.31	-21.72	7.56	2.90
11	-107.18	1.11	-25.34	9.05	3.41
12	-99.63	0.68	-22.27	7.72	2.93
13	-68.86	1.53	-14.52	5.72	0.52
14	-63.97	4.66	-16.37	5.75	-0.36
15	-95.83	6.60	-29.05	10.05	0.11
16	-93.74	6.63	-32.42	11.62	-0.81
17	-85.62	3.34	-24.60	8.23	-0.27
18	-76.50	5.70	-24.62	8.76	0.14
19	-73.35	3.45	-12.18	4.29	0.28
20	-71.12	0.54	-12.61	4.65	0.64
21	-71.64	-0.11	-11.94	4.16	1.30
22	-67.71	0.13	-14.43	5.18	0.32
23	-63.69	-0.04	-8.65	2.53	0.85
24	-78.85	0.92	-14.81	5.46	0.83
25	-69.35	-0.81	-12.19	4.40	1.30
26	-78.50	0.37	-23.80	8.63	1.61
27	-94.69	8.41	-21.67	7.30	1.04
28	-95.22	4.64	-29.14	9.95	1.98
29	-84.22	0.41	-16.30	5.61	1.75
30	-87.34	1.04	-20.88	7.15	2.42
31	-86.31	0.32	-21.27	6.94	2.00
32	-69.98	-1.76	-14.30	5.09	0.39
33	-82.60	3.74	-17.02	6.58	1.90
34	-85.01	5.21	-19.50	7.37	1.58
35	-75.18	-0.05	-12.63	4.63	0.41
36	-104.39	5.46	-24.90	9.01	3.33
37	-107.22	2.23	-27.56	10.02	3.38
38	-95.31	0.63	-28.68	10.42	1.99
39	-113.80	1.29	-29.09	10.29	3.25
40	-99.51	3.24	-29.36	10.52	1.29

<sup>a</sup> Total interaction energy between the ligand and enzyme. <sup>b</sup> Strain energy of the ligand. <sup>c</sup> Electrostatic contribution to the solvation energy of the ligand calculated by solving the nonlinear Poisson–Boltzmann equation. <sup>d</sup> Desolvation energy of the ligand. <sup>e</sup> Predicted activity of the ligands from the QSAR analysis involving the interaction descriptors with standard deviation of 0.5 and above (32 variables).

**Table 8.** Statistical Results from the PLS Analysis

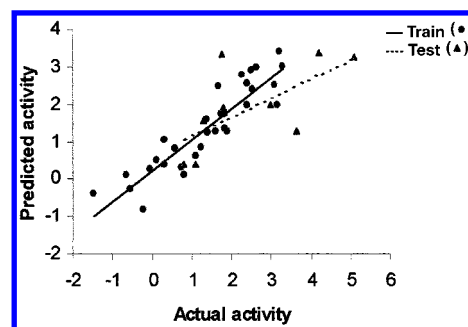
descriptors	no. of variables	$q^2$	$r^2$	$F$	$r^2_{\text{pred}}$
1. $E_{\text{total}}^a$	1	0.226	0.315	13.308	
2. $E_{\text{total}} + E_{\text{lig strain}}^b$	2	0.649	0.725	36.836	0.761
3. $E_{\text{total}} + E_{\text{lig strain}}^b + E_{\text{lig solv}}^c$	3	0.630	0.727	23.971	0.781
4. $E_{\text{total}_1}^d$	6	0.340	0.488	13.364	
5. $E_{\text{total}_1}^d + E_{\text{lig strain}}^b$	7	0.566	0.748	19.318	0.682
6. $E_{\text{total}_1}^d + E_{\text{lig strain}}^b + E_{\text{lig solv}}^c$	8	0.515	0.740	25.293	0.781

<sup>a</sup> Total interaction energy between ligand and enzyme. <sup>b</sup> Strain energy of the ligand. <sup>c</sup> Electrostatic contribution to solvation energy of the ligand calculated by solving the nonlinear Poisson–Boltzmann equation. <sup>d</sup> Total interaction energy subdivided into six interaction terms for interactions in each subsite.

these interaction descriptors shows that the internal predictivity has decreased with  $q^2$  of 0.186, but the correlation coefficient of  $r^2$  is 0.720. To improve the internal as well as external prediction, additional regressors were added. Addition of strain energy of the ligand shows a good improve-

**Table 9.** Statistical Results after Variable Selection

	std dev of descriptor column	no. of variables	$q^2$	$r^2$	$F$	$r^2_{\text{pred}}$
1.	0.0	86	0.534	0.831	31.973	0.535
2.	0.1	50	0.534	0.830	31.819	0.536
3.	0.2	43	0.534	0.829	31.535	0.536
4.	0.3	37	0.525	0.825	30.710	0.535
5.	0.4	33	0.530	0.826	32.842	0.530
6.	0.5	32	0.534	0.827	31.036	0.539
7.	0.6	27	0.546	0.825	30.709	0.526
8.	0.7	23	0.539	0.811	27.861	0.610
9.	0.8	21	0.521	0.795	25.211	0.592
10.	0.9	19	0.519	0.773	22.181	0.606
11.	1.0	17	0.524	0.762	20.785	0.680

**Figure 4.** Graph of actual versus predicted activity from the QSAR analysis with the interaction descriptors having standard deviations of 0.5 and above (32 variables).

ment in the internal predictivity ( $q^2 = 0.557$ ). This model exhibited a moderate external predictivity ( $r^2_{\text{pred}} = 0.436$ ). The electrostatic contribution to the desolvation energy of the ligand marginally reduces  $q^2$  and  $r^2$ . The electrostatic contribution to the solvation energy of the ligand improves the external predictivity of the model ( $r^2_{\text{pred}} = 0.535$ ). In these 84 interaction energy terms some may have “signal” for the correlation but some terms may contribute to “noise” in the QSAR analysis. To improve the signal to noise ratio, the energy data matrix was pruned by deleting the energy terms which have the standard deviation less than a certain value. The QSAR table was pruned progressively to derive a statistically significant QSAR model. The results obtained from this analysis are reported in Table 9. At standard deviation of 0.1 kcal/mol the QSAR model derived shows essentially the same statistical significance as the one with no variable selection. The model has 50 (48 + 2) energy descriptors. Out of 84 only 48 interactions are sufficient in explaining the variance in the biological activity. The variable selection and statistical analysis are repeated by deleting the interaction descriptors. At 0.5 kcal/mol the QSAR model shows a reasonable balance of internal as well as external predictivity. A graph depicting the actual and the predicted activity of the training and test compounds is depicted in Figure 4. This model has 32 interaction descriptors. Some of the amino acid residues involved in this QSAR model are Asp A25, Pro A81, Ile A84, Gly B27 in S1 pocket Gly A27, Asp B25, Pro B81, Ile B84 in S1' pocket Ile A50, Ala B28, Asp B30, Ile B47, Gly B48 in S2 pocket Ala A28, Asp A29, Asp A30, Ile A47, Gly A48, Ile B50 in S2' pocket Val A82 in S3 pocket, and Arg B8, Val B82 in S3' pocket. The interactions with these amino acid residues mainly explain the variance in biological activity in the present series of compounds. Most of the amino acid residues used in present correlation make a direct contact with the inhibitor,



suggesting that short-range interactions are more important than long-range interactions. The present RD-QSAR study showed that not only interactions at the active site but also the properties of the ligand itself such as solvation/desolvation and the conformational energies are bound to play an important role in determining the activity. A similar analysis reported in the literature shows that a good correlation can be obtained with the total interaction energy alone. However the present study may not be compared with this, as our dataset is different than what is reported in the literature. It may not be possible to generalize the technique of deriving valid QSAR using the receptor information, as it would vary with each dataset and each ligand/receptor system used. But the present approach of using interaction energy terms in QSAR may lead to some newer general technique for 3D-QSAR using the receptor information. One way to do this would be to divide the active site of the receptor in a box of grids and calculate the interaction energy at each of these grid points. These energies can be used as interaction descriptors for the statistical analysis. This approach can have broad applicability not only in HIV-1 protease but also for other molecular targets, and also in the case of prediction of activity of a ligand on two or more targets as in the case of HIV-1 protease wild types and mutants and also for the design of selective inhibitors when homologues proteins with similar active site are known to exist.

### CONCLUSIONS

The ligand/enzyme models of the nonpeptidic HIV-1 protease inhibitors were built using the docking and molecular dynamics simulations. The protonation scheme of the active site aspartic acid residues was derived using a set of computations. The active site aspartates are monoprotinated with a proton placed on the OD1 atom of the Asp B25. The ligand/enzyme models were derived and used to generate the interaction descriptors. The total interaction energy was calculated and used as an independent descriptor to develop a QSAR. With just total interaction energy a poor correlation was obtained. To improve the statistical significance of the QSAR models, additional descriptors were used. Inclusions of the ligand strain energy and solvation energy improve the correlative and predictive properties. The molecular mechanics interaction energy was partitioned into six different interaction energies considering the different interaction subsite of the HIV-1 protease. A similar trend was observed wherein inclusion of consequences of ligand binding and solvation effects improves the statistical significance of the QSAR model. Each amino acid residue level interaction was also calculated. This large descriptor matrix was then used for PLS analysis to derive a linear correlation. From this analysis it was found that interaction at some amino acids is essential to exhibit tight binding. Future ligand design efforts should be directed toward these interactions; also the ligands should be made more conformationally rigid to reduce the entropic penalty upon ligand binding. The present study shows the application of various computational strategies for the prediction of binding affinities of the ligands in SBDD. The descriptors generated and the correlations derived are rather empirical, but the statistical results obtained show that the present approach will be useful for the prediction and ranking of binding affinities of the ligands. The use of ligand/enzyme interactions in QSAR is not only useful in prediction

of binding affinities for a molecular target, but such information can also be useful in predicting the binding affinities for the homologues molecular targets. Presently research is underway in our laboratory to develop a general approach for the use of receptor information in developing QSAR and utilizing this information to develop selective chemotherapeutic agents.

### ACKNOWLEDGMENT

The authors thank the University Grants Commission (UGC), New Delhi, for the financial support under its COSIST program. S.S.K. thanks Dr. Tanaji and Mr. Vijay for useful discussions.

### REFERENCES AND NOTES

- (1) Whittle, P. J.; Blundell, T. L. Protein Structure-Based Drug Design. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 349–375.
- (2) Greer, J.; Erickson, J. W.; Baldwin, J. J.; Verney, M. D. Application of the Three-Dimensional Structures of Protein Target Molecules in Structure-Based Drug Design. *J. Med. Chem.* **1994**, *37*, 1035–1054.
- (3) Ajay; Murcko, M. A. Computational Methods to Predict Binding Free Energy in Ligand–Receptor Complexes. *J. Med. Chem.* **1995**, *38*, 4953–4967.
- (4) Williams, D. H.; Cox, J. P. L.; Doig, A. J.; Gardner, M.; Gerhard, U.; Kaye, P. T.; Lal, A. R.; Nicholls, I. A.; Salter, C. J.; Mitchell, R. C. Toward the Semiquantitative Estimation of Binding Constants. Guides for Peptide–Peptide Binding in Aqueous Solution. *J. Am. Chem. Soc.* **1991**, *113*, 7020–7030.
- (5) Searle, M. S.; Williams, D. H. The Cost of Conformational Order: Entropy Changes in Molecular Associations. *J. Am. Chem. Soc.* **1992**, *114*, 10690–10697.
- (6) Searle, M. S.; Williams, D. H.; Gerhard, U. Partitioning of Free Energy Contributions in the Estimation of Binding Constants: Residual Motions and Consequences for Amide–Amide Hydrogen Bond Strengths. *J. Am. Chem. Soc.* **1992**, *114*, 10697–10704.
- (7) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomenon. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (8) Martin, Y. C. *Quantitative Drug Design: A Critical Introduction*; Marcel Dekker Inc.: New York, Basel, 1978; p 425.
- (9) Cramer, R. D., III.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
- (10) DePriest, S. A.; Mayer, D.; Naylor, C. B.; Marshall, G. R. 3D-QSAR of Angiotensin-Converting Enzyme and Thermolysin Inhibitors: A Comparison of CoMFA Models Based on Deduced and Experimentally Determined Active Site Geometries. *J. Am. Chem. Soc.* **1993**, *115*, 5372–5384.
- (11) Kulkarni, S. S.; Kulkarni, V. M. Three-Dimensional Quantitative Structure Activity Relationship of Interleukin 1 $\beta$  Converting Enzyme Inhibitors: A Comparative Molecular Field Analysis Study. *J. Med. Chem.* **1999**, *42*, 373–380.
- (12) Tokarski, J. S.; Hopfinger, A. J. Prediction of Ligand–Receptor Binding Thermodynamics by Free Energy Force Field (FEFF) 3D-QSAR Analysis: Application to a Set of Peptidomimetic Renin Inhibitors. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 792–811.
- (13) Hopfinger, A. J.; Nakata, Y.; Max, N. Quantitative Structure–Activity Relationships of Anthracycline Antitumor Activity and Cardiac Toxicity Based Upon Intercalation Calculations. In *Intermolecular Forces*; Pullman, B., Ed.; D. Reidel Publishing: Dordrecht, The Netherlands, 1981; p 431.
- (14) Holloway, M. K.; Wai, J. M.; Halgren, T. A.; Fitzgerald, P. M. D.; Vacca, J. P.; Dorsey, B. D.; Levin, R. B.; Thompson, W. J.; Chen, L. J.; deSolms, S. J.; Gaffin, N.; Ghosh, A. K.; Giuliani, E. A.; Graham, S. L.; Guare, J. P.; Hungate, R. W.; Lyle, B. A.; Sanders, W. M.; Tucker, T. J.; Wiggins, M.; Wiscout, C. M.; Woltersdorf, O. W.; Young, S. D.; Darke, P. L.; Zugay, J. A. *A Priori* Prediction of Activity for HIV-1 Protease Inhibitors Employing Energy Minimization in the Active Site. *J. Med. Chem.* **1995**, *38*, 305–317.
- (15) Perez, C.; Pastor, M.; Ortiz, A. R.; Gago, F. Comparative Bonding Energy Analysis of HIV-1 Protease Inhibitors: Incorporation of Solvent Effects and Validation as a Powerful Tool in Receptor-Based Drug Design. *J. Med. Chem.* **1998**, *41*, 836–852.
- (16) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of Drug Binding Affinities by Comparative Binding Energy Analysis. *J. Med. Chem.* **1995**, *38*, 2681–2691.



- (17) Head, R. D.; Smythe, M. L.; Oprea, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. R. VALIDATE: A New Method for the Receptor-Based Prediction of Binding Affinities of Novel Ligands. *J. Am. Chem. Soc.* **1996**, *118*, 3959–3969.
- (18) Luty, B. A.; Wasserman, Z. R.; Stouten, P. F. W.; Hodge, C. N.; Zacharias, M.; McCammon, J. A. A Molecular Mechanics/Grid Method for Evaluation of Ligand–Receptor Interactions. *J. Comput. Chem.* **1995**, *16*, 454–464.
- (19) Wasserman, Z. R.; Hodge, C. N. Fitting an Inhibitor into the Active Site of Thermolysin: A Molecular Dynamics Case Study. *Proteins: Struct., Funct. Genet.* **1996**, *24*, 227–237.
- (20) Debouck, C.; Metcalf, B. W. Human Immunodeficiency Virus Protease: A Target for AIDS Therapy. *Drug. Dev. Res.* **1990**, *21*, 1–17.
- (21) Huff, J. R. HIV Protease: A Novel Chemotherapeutic Target for AIDS. *J. Med. Chem.* **1991**, *34*, 2305–2314.
- (22) Blundell, T. L.; Lapatto, R.; Wilderspin, A. F.; Hemmings, A. M.; Hobart, P. M.; Danley, D. E.; Whittle, P. J. The 3-D Structure of HIV-1 Protease and the Design of Antiviral Agents for the Treatment of AIDS. *Trends Biol. Sci.* **1990**, *15*, 425–430.
- (23) Appelt, K. Crystal Structures of HIV-1 Protease-Inhibitor Complexes. *Perspect. Drug Discovery Des.* **1993**, *1*, 23–48.
- (24) Thaisrivongs, S.; Tomich, P. K.; Watenpaugh, K. D.; Chong, K. T.; Howe, J. W.; Yang, C. P.; Strohhach, J. P.; Turner, S. R.; McGrath, J. P.; Bohanon, M. J.; Lynn, J. C.; Mulichak, A. M.; Spinelli, P. A.; Hinshaw, R. R.; Pagano, P. J.; Moon, J. B.; Ruwart, M. J.; Wilkinson, K. F.; Rush, B. D.; Zipp, G. L.; Dalga, R. J.; Schwende, F. J.; Howard, G. M.; Padbury, G. E.; Toth, L. N.; Zhao, Z.; Koeplinger, K. A.; Kakuk, T. J.; Cole, S. L.; Zaya, R. M.; Piper, R. C.; Jeffrey, P. R. Structure-Based Design of HIV Protease Inhibitors; 4-Hydroxycoumarins and 4-Hydroxy-2-pyrones as Non-Peptidic Inhibitors. *J. Med. Chem.* **1994**, *37*, 3200–3204.
- (25) Romines, K. R.; Watenpaugh, K. D.; Tomich, P. K.; Hoew, W. J.; Morris, J. K.; Lovasz, K. D.; Mulichak, A. M.; Finzel, B. C.; Lynn, J. C.; Horng, M. M.; Schwende, F. J.; Ruwart, M. J.; Zipp, G. L.; Chong, K. T.; Dolak, L. A.; Toth, L. N.; Howard, G. M.; Rush, B. D.; Wilkinson, K. F.; Possert, P. L.; Dalga, R. J.; Hinshaw, R. R. Use of Medium-Sized Cycloalkyl Rings To Enhance Secondary Binding: Discovery of a New Class of Human Immunodeficiency Virus (HIV) Protease Inhibitors. *J. Med. Chem.* **1995**, *38*, 1884–1891.
- (26) Thaisrivongs, S.; Watenpaugh, K. D.; Howe, W. J.; Tomich, P. K.; Dolak, L. A.; Chong, K. T.; Tomich, C. S. C.; Tomasselli, A. G.; Turner, S. R.; Strohhach, J. W.; Mulichak, A. M.; Janakiraman, M. N.; Moon, J. B.; Lynn, J. C.; Horng, M. M.; Hinshaw, R. R.; Curry, K. A.; Rothrock, D. J. Structure-Based Design of Novel HIV Protease Inhibitors: Carboxamide-Containing 4-Hydroxycoumarins and 4-Hydroxy-2-pyrones as Potent Nonpeptidic Inhibitors. *J. Med. Chem.* **1995**, *38*, 3624–3637.
- (27) Romines, K. R.; Watenpaugh, K. D.; Howe, W. J.; Tomich, P. K.; Lovasz, K. D.; Morris, J. K.; Janakiraman, M. N.; Lynn, J. C.; Horng, M. M.; Chong, K. T.; Hinshaw, R. R.; Dolak, L. A. Structure-Based Design of Nonpeptidic HIV Protease Inhibitors from a Cyclooctylpyranone Lead Structure. *J. Med. Chem.* **1995**, *38*, 4463–4473.
- (28) Skulnick, H. I.; Johnson, P. D.; Howe, W. J.; Tomich, P. K.; Chong, K. T.; Watenpaugh, K. D.; Janakiraman, M. N.; Dolak, L. A.; McGrath, J. P.; Lynn, J. C.; Horng, M. M.; Hinshaw, R. R.; Zipp, G. L.; Ruwart, M. J.; Schwende, F. J.; Zhong, W. Z.; Padbury, G. E.; Dalga, R. J.; Shiou, L.; Possert, P. L.; Rush, B. D.; Wilkinson, K. F.; Howard, G. M.; Toth, L. N.; Williams, M. G.; Kakuk, T. J.; Cole, S. L.; Zaya, R. M.; Lovasz, K. D.; Morris, J. K.; Romines, K. R.; Thaisrivongs, S.; Aristoff, P. A. Structure-Based Design of Sulfonamide-Substituted Non-Peptidic HIV Protease Inhibitors. *J. Med. Chem.* **1995**, *38*, 4968–4971.
- (29) Thaisrivongs, S.; Janakiraman, M. N.; Chong, K. T.; Tomich, P. K.; Dolak, L. A.; Turner, S. R.; Strohhach, J. W.; Lynn, J. C.; Horng, M. M.; Hinshaw, R. R.; Watenpaugh, K. D. Structure-Based Design of Novel HIV Protease Inhibitors: Sulfonamide-Containing 4-Hydroxycoumarins and 4-Hydroxy-2-pyrones as Potent Non-Peptidic Inhibitors. *J. Med. Chem.* **1996**, *39*, 2400–2410.
- (30) Schwartz, T. M.; Bundy, G. L.; Strohhach, J. W.; Thaisrivongs, S.; Johnson, P. D.; Skulnick, H. I.; Tomich, P. K.; Lynn, J. C.; Chong, K. T.; Hinshaw, R. R.; Raub, T. J.; Padbury, G. E.; Toth, L. N. Synthesis and Pharmacological Evaluation of Sulfone Substituted HIV Protease Inhibitors. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 399–402.
- (31) Thaisrivongs, S.; Romero, D. L.; Tommasi, R. A.; Janakiraman, M. N.; Strohhach, J. W.; Turner, S. R.; Biles, C.; Morge, R. R.; Johnson, P. D.; Aristoff, P. A.; Tomich, P. K.; Lynn, J. C.; Horng, K. T.; Hinshaw, R. R.; Howe, W. J.; Finzel, B. C.; Watenpaugh, K. D. Structure-Based Design of HIV Protease Inhibitors: 5,6-Dihydro-4-hydroxy-2-pyrones as Effective, Nonpeptidic Inhibitors. *J. Med. Chem.* **1996**, *39*, 4630–4642.
- (32) Thaisrivongs, S.; Skulnick, H. I.; Turner, S. R.; Strohhach, 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.; Trapley, 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.; Schwende, 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. M.; 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 Non-Peptidic Inhibitors. *J. Med. Chem.* **1996**, *39*, 4349–4353.
- (33) *Insight II*, Release 97.0; Molecular Simulations Inc.: San Diego, CA, 1997.
- (34) *SYBYL 6.22 molecular modeling software*; Tripos Associates Inc.: St. Louis, MO, 1995.
- (35) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (36) Veerapandian, B.; Cooper, J. B.; Sali, A.; Blundell, T. L.; Rosati, R. L.; Dominy, B. W.; Damon, D. B.; Hoover, D. J. Direct Observation by X-ray Analysis of the Tetrahedral “Intermediate” of Aspartic Proteinases. *Protein Sci.* **1992**, *1*, 322–328.
- (37) Hyland, L. J.; Tomaszek, T. A., Jr.; Roberts, G. D.; Carr, S. A.; Magaard, V. W.; Bryan, H. L.; Fakhoury, S. A.; Moore, M. L.; Minnich, M. D.; Culp, J. S.; DesJarlais, R. L.; Meek, T. D. Human Immunodeficiency Virus-1 Protease. 1. Initial Velocity Studies and Kinetic Characterization of Reaction Intermediates by <sup>18</sup>O Isotope Exchange. *Biochemistry* **1991**, *30*, 8441–8453.
- (38) Hyland, L. J.; Tomaszek, T. A., Jr.; Meek, T. D. Human Immunodeficiency Virus-1 Protease. 2. Use of pH Rate Studies and Solvent Kinetic Isotope Effects To Elucidate Details of Chemical Mechanism. *Biochemistry* **1991**, *30*, 8454–8463.
- (39) Jaskolski, M.; Tomaselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrichson, R. L.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. Structure at 2.5 Å Resolution of Chemically Synthesized Human Immunodeficiency Virus Type 1 Protease Complexed with a Hydroxyethylene Based Inhibitor. *Biochemistry* **1991**, *30*, 1600–1609.
- (40) 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.
- (41) Harte, W. E., Jr.; Beveridge, D. L. Prediction of the Protonation State of the Active Site Aspartyl Residues in HIV-1 Protease-Inhibitor Complexes via Molecular Dynamics Simulation. *J. Am. Chem. Soc.* **1993**, *115*, 3883–3886.
- (42) Ferguson, D. M.; Radmer, R. J.; Kollman, P. A. A Determination of the Relative Binding Energies of Peptide Inhibitors of the HIV-1 Protease. *J. Med. Chem.* **1991**, *34*, 2654–2659.
- (43) Yamazaki, T.; Nicholson, L. K.; Torchia, D. A.; Wingfield, P.; Stahl, S. J.; Kaufman, J. D.; Eyermann, C. J.; Hodge, C. N.; Lam, P. Y. S.; Ru, Y.; Jadhav, P. K.; Chang, C.; Weber, P. C. NMR and X-ray Evidence that the HIV Protease Catalytic Aspartyl Groups Are Protonated in the Complex Formed by the Protease and a Non-Peptidic Cyclic Urea-Based Inhibitor. *J. Am. Chem. Soc.* **1994**, *116*, 10791–10792.
- (44) Nicholls, A.; Honig, B. A. Rapid Finite Difference Algorithm, Utilizing Successive Over-Relaxation to Solve the Poisson–Boltzmann Equation. *J. Comput. Chem.* **1991**, *12*, 435–445.
- (45) Gilson, M. K.; Sharp, K. A.; Honig, B. H. Calculating the Electrostatic Potential of Molecules in Solution: Method and Error Assessment. *J. Comput. Chem.* **1987**, *9*, 327–335.
- (46) Gilson, M. K.; Honig, B. H. Calculation of the Total Electrostatic Energy of a Macromolecular System: Solvation Energies, Binding Energies, and Conformational Analysis. *Proteins* **1988**, *4*, 7–18.
- (47) Schechter, I.; Berger, A. On the Size of the Active Site in Proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157–162.
- (48) Wold, S. Partial Least Squares Analysis. In *3D-QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM Science Publishers: Leiden, The Netherlands, 1993.
- (49) Lam, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C. H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Vitanen, S. E. Rational Design of Potent, Bioavailable, Nonpeptide Cyclic Ureas as HIV Protease Inhibitors. *Science* **1994**, *263*, 380–384.
- (50) Hariprasada, V.; Talele, T. T.; Kulkarni, V. M. Design and Synthesis of a Novel Series of Nonpeptidic HIV-1 Protease Inhibitors. *Pharm. Pharmacol. Commun.* **1998**, *4*, 365–372.
- (51) Kuntz, I. D. Structure-Based Strategies for Drug Design and Discovery. *Science* **1992**, *257*, 1078–1082.
- (52) Cherfils, J.; Janin, J. Protein Docking Algorithms: Simulating Molecular Recognition. *Curr. Opin. Struct. Biol.* **1993**, *3*, 265–269.