Computational Studies on Tetrahydropyrimidine-2-one HIV-1 Protease Inhibitors: Improving Three-Dimensional Quantitative Structure–Activity Relationship Comparative Molecular Field Analysis Models by Inclusion of Calculated Inhibitor- and Receptor-Based Properties

Anil C. Nair,[†,‡] Philippa Jayatilleke,[†,§] Xia Wang,[†,‖] Stanislav Miertus,[⊥] and William J. Welsh*,[†]

Department of Chemistry & Biochemistry and Center for Molecular Electronics, University of Missouri–St. Louis, St. Louis, Missouri 63121, and International Center for Science and High Technology, UNIDO, Area Science Park, Padriciano 99, Trieste 34012, Italy

A computational chemistry study has been performed on a series of tetrahydropyrimidine-2-ones (THPs) as HIV-1 protease (HIV-1 PR) inhibitors. The present investigation focuses on the correlation of inhibitor–enzyme complexation energies (E_{compl}), inhibitor solvation energies $E_{solv}[I]$, and both polar and nonpolar buried surface areas (BSAs) with the observed values of the binding affinity (pK_I). Various combinations of these specific inhibitor- and receptor-based properties were also evaluated as additional descriptors to three-dimensional quantitative structure–activity relationship (3D-QSAR) models constructed using comparative molecular field analysis (CoMFA). Linear regression of the observed pK_I values with E_{compl}, $E_{solv}[I]$, and the BSAs yielded a strong correlation in terms of both self-consistency ($r^2 \approx 0.90$) and internal predictive ability ($r_{cv}^2 > 0.50$). The 3D-QSAR models obtained from CoMFA using standard partial least-squares (PLS) analysis also yielded a strong correlation between the CoMFA fields and the experimental pK_i ($r^2 = 0.96$; $r_{cv}^2 = 0.58$). Various "enhanced" 3D-QSAR models were constructed in which different combinations of the E_{compl}, $E_{solv}[I]$, and BSAs were added as additional descriptors to the default steric–electrostatic CoMFA fields. Inclusion of $E_{solv}[I]$ in particular yielded significant improvement in the predictive ability ($r_{cv}^2 \approx 0.80$) of the resultant 3D-QSAR model.

Introduction

HIV-1 protease (HIV-1 PR) plays a key role in the post-translational processing of gag and gag-pol viral gene products and thereby is essential for virion maturation.[1,2] Consequently, HIV-1 PR is an attractive target for developing anti-AIDS drugs because its inhibition results in the production of noninfectious virus.[3–6] Many HIV-1 PR inhibitors, including indinavir, nelfinavir, saquinavir, and ritonavir, have been approved as anti-AIDS drugs.[6,7] These drugs were found to be very useful in reducing the viral load and improving the CD4 cell counts in AIDS patients. However, rapid emergence of drug resistance has been reported for almost all protease inhibitors currently in clinical use due to site-specific mutations in the enzyme.[7–9] The bioavailability and toxicity profiles of protease inhibitors are also of importance; thus, there exists an urgent need to discover a new generation of protease inhibitors that are more potent against these mutant forms of the virus and, at

the same time, that exhibit low toxicity and high bioavailability.

Various computational methods, including, for example, a priori prediction of enzyme–inhibitor binding affinities[10–16] and free energy perturbation (FEP) methods,[17–19] have been successfully employed to help guide the rational design and optimization of novel HIV-1 PR inhibitors. Previous studies in our laboratory[16] have shown that enzyme–inhibitor complexation energies calculated using molecular mechanics can provide an efficient and highly useful approach for the rational design of new inhibitor candidates and for rapid screening of moderately sized ligand databases. Three-dimensional quantitative structure–activity relationship (3D-QSAR) methods such as comparative molecular field analysis (CoMFA) are also useful in the rapid prediction of the binding affinities of new inhibitor candidates.[16,20–22] CoMFA is a versatile and powerful tool in rational drug design and related applications. CoMFA samples the steric and electrostatic fields surrounding a set of ligands and constructs a 3D-QSAR model by correlating these 3D steric and electrostatic fields with the corresponding observed activities. One of the unique features of CoMFA is its ability to represent the 3D-QSAR model in terms of color contour maps that depict locations on the ligands where structural modifications might enhance their biological activity (e.g., binding affinity). These maps can serve as guides for designing analogues within the same series

* To whom correspondence should be addressed. Present address: Department of Pharmacology, Robert Wood Johnson Medical School, University of Medicine & Dentistry of New Jersey, 675 Hoes Lane, Piscataway, NJ 08854-5635. Phone: (732) 235-5313. Fax: (732) 235-4073. E-mail: WELSHWJ@UMDNJ.EDU.
† University of Missouri–St. Louis.
‡ Present address: Aventis Pharmaceuticals, 1580 E. Hanley Boulevard, Tucson, AZ 85737.
§ Present address: Tripos, Inc., 1699 S. Hanley Road, St. Louis, MO 63144.
‖ Present address: AstraZeneca Pharmaceuticals, 1800 Concord Pike, Wilmington, DE 19850.
⊥ UNIDO.

of compounds that might possess more desirable biological properties.

The present study reports a computational study of a series of six-membered cyclic ureas known as tetrahydropyrimidine-2-ones (THPs) for use as HIV-1 PR inhibitors. Our primary aim was to explore the utility of calculated enzyme–inhibitor complexation energies (E_{compl}) as well as inhibitor buried surface areas (BSAs) and inhibitor solvation energies (E_{solv}) as descriptors for predicting experimental enzyme inhibition constants. An additional aim was to evaluate whether the statistical quality of standard CoMFA models could be improved, and to what degree, by inclusion of these property values as additional descriptors. Various CoMFA models were developed by including the complexation energies, BSAs, and solvation energies in various combinations as additional descriptors to improve the model's predictive ability. Such "enhanced" CoMFA models can be used to predict the activity of new inhibitors, and the CoMFA contour maps are extremely useful as visual guides in the rational design of new inhibitors.

Computational Methods

Complexation Energies. The binding of an inhibitor (I) to HIV-1 protease (PR) to form the enzyme–inhibitor complex (PR–I) can be represented by the following equilibrium:

$$PR + I \rightleftharpoons PR-I \quad (1)$$

The complexation energy (E_{compl}) of an inhibitor for this reversible inhibition process can be calculated from the energies of the free enzyme $E[PR]$, the free inhibitor $E[I]$, and the enzyme–inhibitor complex $E[PR-I]$ employing the relation

$$E_{compl} = E[PR-I] - (E[PR] + E[I]) \quad (2)$$

Values of $E[PR]$, $E[I]$, and $E[PR-I]$ were obtained from molecular mechanics minimization using the consistent valence force field (CVFF) and default partial atomic charges available within the InsightII/Discover molecular modeling package.[23] The dielectric shielding in proteins was accounted for by setting the dielectric constant ϵ equal to 4, and the nonbonded energy cutoff was set to ∞ (i.e., no cutoff). Energy minimization of the enzyme–inhibitor complexes and of the free enzyme was carried out gradually in steps, viz., only hydrogens initially, then only the side chains, and finally full relaxation of all atoms in the system. This procedure was utilized to minimize unrealistic motions of the protein arising from computational artifacts. In all cases, geometry optimization was performed starting with 200 cycles of steepest descent followed by conjugate gradient energy minimization until a convergence criterion of 0.01 kcal mol^{-1} Å^{-1} for the average energy gradient between successive iterations was achieved. The enzyme structure employed for the calculation was the InsightII/Delphi coordinate file kindly provided by Perez et al.[11] The two active-site aspartate residues A25 and B25 were modeled as protonated and unprotonated, respectively, in accordance with experimental observations.[6] The structural "flap" H_2O molecule located within the ligand binding site was removed because the carbonyl group within the cyclic urea moiety of this series of THPs mimic is intended to mimic the oxygen atom of this water molecule by forming hydrogen-bonding interactions with the backbone NH groups of Ile A50 and Ile B50. Inhibitor binding will displace and, presumably, induce release of this H_2O molecule to the bulk aqueous environment, thus providing an additional favorable entropic contribution to the free energy of binding.

Buried Surface Area. The buried surface area (BSA$_{Tot}$) of a bound ligand inside the active site of the enzyme has been shown in various applications[6] to give insight into the free energy of binding. The BSA$_{Tot}$ can be partitioned into polar buried surface area (BSA$_{Pol}$) and nonpolar buried surface area (BSA$_{NonPol}$). The values of BSA$_{Pol}$ and BSA$_{NonPol}$ are useful, respectively, for quantifying the relative contribution of polar vs nonpolar interactions to enzyme–ligand binding. The BSAs were computed from the solvent-accessible surface areas (SASAs) of each free inhibitor (SASA[I]), the free enzyme (SASA[PR]), and each enzyme–inhibitor complex (SASA[PR–I]) using the following equations:

$$BSA_{Tot} = SASA_{Tot}[PR] + SASA_{Tot}[I] - SASA_{Tot}[PR-I] \quad (3)$$

$$BSA_{Pol} = SASA_{Pol}[PR] + SASA_{Pol}[I] - SASA_{Pol}[PR-I] \quad (4)$$

$$BSA_{NonPol} = SASA_{NonPol}[PR] + SASA_{NonPol}[I] - SASA_{NonPol}[PR-I] \quad (5)$$

The subscripts Tot, Pol, and NonPol indicate the total, polar, and nonpolar components of the surface area, respectively. Values of SASA for all molecular species were calculated using the ProStat option in the homology module of the Insight II modeling package.[23] ProStat provides a set of protein structure analysis functions and allows the computation of the absolute per residue solvent-accessible surface area or that relative to the Gly-X-Gly tripeptide model.

Solvation Energy. Inclusion of solvent effects in the theoretical prediction of inhibition constants should be expected to improve the predictability of the regression model. The solvation energy E_{solv} associated with formation of the ligand–enzyme complex (eq 1) is given by the expression

$$E_{solv} = E_{solv}[PR-I] - (E_{solv}[PR] + E_{solv}[I]) \quad (6)$$

where the individual terms are defined by analogy to eq 2. Preliminary calculations revealed that differences between $E_{solv}[PR-I]$ and $E_{solv}[PR]$ are fairly constant across a series of structurally related inhibitors so that, according to eq 6, E_{solv} scales almost monotonically with $E_{solv}[I]$. Accordingly, it was concluded that $E_{solv}[I]$ aptly serves as a surrogate for E_{solv} at little cost in accuracy and with appreciable gain in computational speed. In fact, the present results show that $E_{solv}[I]$ correlated quite strongly with inhibition constants for this series of THP compounds.

When the GB/SA (generalized Born/surface area) method and MMFF (Merck molecular force field) contained within the Macromodel 6.0 modeling package were used,[24] values of $E_{solv}[I]$ were calculated by the following expression:

$$E_{solv}[I] = E_{MMFF}{}^{Wat}[I] - E_{MMFF}{}^{Gas}[I] \quad (7)$$

where $E_{MMFF}{}^{Wat}[I]$ and $E_{MMFF}{}^{Gas}[I]$ are the geometry-optimized energy of the free (i.e., unbound) inhibitor in water and in the gas phase, respectively. Computation of $E_{solv}[I]$ with inclusion of energy minimization will be superior to the corresponding single-point calculation (i.e., no energy minimization) because it includes the strain energy associated with solvation and desolvation of the free inhibitor.

Comparative Molecular Field Analysis (CoMFA). The CoMFA module was accessed using Sybyl, version 6.6,[25] and partial atomic charges were calculated using the Gasteiger–Marsili[26] formalism. Each inhibitor molecule was first optimized inside the binding pocket of the enzyme to obtain the putative bioactive conformation, then aligned using molecule **1** (see Table 1) as a template. Molecular alignment (superimposition) was carried out using the *atom fit* option in Sybyl with respect to the central six-membered ring core of the molecules. This six-membered ring was present in every compound within this series of THP inhibitors, thus making the alignment procedure a more straightforward process.

The CoMFA region was defined to extend 5 Å beyond the van der Waals radii of the assembly of superimposed molecules, along each of the principal axes of a Cartesian

Table 1. Inhibition Constants (pK_I) from Experiment and from Linear Regression Models I−VIII[a] Involving E_{compl}, $E_{solv}[[I]$, BSA_{Pol}, and BSA_{NonPol} for the Training Set Inhibitors

Compound No.	Structure	pK_I (exp.)	pK_I (predicted)							
			I	II	III	IV	V	VI	VII	VIII
1		7.82	8.18	7.04	7.76	7.55	7.43	7.12	7.01	7.08
2		7.64	9.84	7.69	9.05	9.18	8.90	7.84	7.72	7.87
3		6.01	6.03	6.58	6.53	5.81	6.07	6.40	6.52	6.30
4		6.73	8.03	6.86	7.95	7.37	7.41	6.76	6.82	6.71
5		6.27	7.51	7.34	7.41	7.70	7.63	7.35	7.37	7.41
6		6.49	7.51	6.91	7.27	7.51	7.40	6.96	6.95	7.01
7		7.18	7.76	7.81	7.68	7.79	7.75	7.85	7.81	7.85
8		7.82	7.66	7.64	7.47	8.06	7.92	7.73	7.70	7.81
9		8.59	7.34	7.43	7.48	7.56	7.60	7.37	7.46	7.40
10		6.80	7.68	7.58	7.62	7.98	7.92	7.59	7.63	7.66
11		8.21	8.82	8.64	8.68	8.22	8.23	8.69	8.56	8.59
12		7.96	8.93	8.41	8.72	8.73	8.66	8.45	8.41	8.45
13		9.04	9.21	8.81	8.89	8.73	8.64	8.93	8.76	8.87
14		9.31	9.53	8.88	9.46	9.26	9.25	8.85	8.88	8.84
15		10.05	9.48	9.77	9.46	9.61	9.59	9.82	9.78	9.83

Table 1 (Continued)

Compound No.	Structure	pK$_I$ (exp.)	pK$_I$ (predicted) I	II	III	IV	V	VI	VII	VIII
16		9.51	9.53	9.93	9.54	9.85	9.82	9.97	9.96	10.01
17		9.82	9.73	10.30	9.64	10.55	10.41	10.40	10.39	10.53
18		8.85	9.80	9.39	10.02	10.33	10.37	9.23	9.49	9.36
19		10.70	9.41	10.45	9.43	9.28	9.30	10.56	10.37	10.47
20		10.70	9.52	10.54	9.47	9.69	9.65	10.68	10.51	10.65
21		11.00	9.90	11.18	10.29	10.30	10.43	11.14	11.17	11.12
22		10.22	9.83	10.52	10.21	10.43	10.53	10.42	10.57	10.47
23		8.77	8.49	8.66	8.50	8.19	8.23	8.67	8.60	8.60
24		8.10	8.87	8.00	9.18	8.94	9.08	7.76	8.07	7.81
25		8.31	8.98	8.82	8.68	8.37	8.31	8.95	8.73	8.85
26		10.00	10.32	9.93	9.72	9.85	9.63	10.18	9.88	10.13
27		10.22	9.56	9.74	9.69	9.49	9.56	9.69	9.72	9.66
28		10.00	10.14	9.79	10.21	10.08	10.12	9.71	9.80	9.72
29		10.52	9.705	10.91	10.38	10.05	10.32	10.72	10.89	10.68
30		9.39	9.790	9.85	10.12	10.36	10.44	9.69	9.93	9.79

Table 1 (Continued)

Compound No.	Structure	pK_I (exp.)	pK_I (predicted)							
			I	II	III	IV	V	VI	VII	VIII
31		7.64	8.979	8.68	9.08	8.20	8.34	8.60	8.58	8.47
32		9.52	8.538	8.75	8.46	8.43	8.40	8.81	8.72	8.78
33		7.92	8.019	8.42	8.13	7.77	7.85	8.41	8.36	8.33
34		7.96	8.261	8.26	8.76	7.56	7.87	8.02	8.15	7.86
35		6.96	7.713	7.78	7.66	8.02	7.96	7.82	7.83	7.87
36		8.25	7.934	8.26	7.93	8.29	8.24	8.30	8.30	8.35
37		7.85	7.891	7.86	7.84	8.44	8.35	7.88	7.95	8.00
38		9.62	8.097	8.60	8.23	8.54	8.55	8.59	8.64	8.64
39		8.85	8.055	8.45	8.11	8.69	8.64	8.47	8.53	8.56
40		9.42	8.140	9.35	8.19	8.78	8.72	9.46	9.38	9.50
41		8.92	7.756	8.25	7.70	8.08	8.02	8.33	8.27	8.36
42		7.48	8.481	7.66	8.19	8.12	8.03	7.71	7.65	7.70
43		7.40	8.503	7.61	8.28	8.07	8.02	7.61	7.59	7.59
44		9.19	9.420	8.98	9.64	9.38	9.49	8.82	9.00	8.84
45		9.55	9.015	9.48	8.86	8.99	8.92	9.61	9.45	9.58

Table 1 (Continued)

Compound		pK_I	pK_I (predicted)							
No.	Structure	(exp.)	I	II	III	IV	V	VI	VII	VIII
46		10.00	9.399	9.38	9.68	9.37	9.50	9.24	9.37	9.22
47		9.89	9.022	9.22	8.85	8.92	8.86	9.32	9.19	9.29
48		8.80	8.695	8.61	8.58	8.26	8.25	8.66	8.55	8.58
49		10.00	9.847	9.87	10.17	10.13	10.24	9.72	9.91	9.76

[a] See Results and Discussion for the description of the models.

Table 2. Statistical Parameters Associated with Linear Regression Models I–VIII[a] Constructed from E_{compl}, $E_{solv}[I]$, BSA_{Pol}, BSA_{NonPol}, and BSA_{Total} Descriptors for the Training Set of THP Compounds

model	r^2	r_{cv}^2	F^b	SEE^c	descriptors				
					E_{compl}	$E_{solv}[I]$	BSA_{Pol}	BSA_{NonPol}	BSA_{Total}^d
I	0.56	0.52	59.0	0.83	✓				
II	0.84	0.82	121.0	0.50	✓	✓			
III	0.60	0.57	35.0	0.79	✓		✓		
IV	0.67	0.63	47.0	0.72	✓			✓	
V	0.68	0.62	32.0	0.72	✓				✓
VI	0.85	0.83	84.0	0.50	✓	✓	✓		
VII	0.84	0.82	80.0	0.51	✓	✓		✓	
VIII	0.85	0.84	63.0	0.50	✓	✓			✓

[a] See Results and Discussion for description of the models. [b] F ratio: defined as $r^2/(1 - r^2)$, representing the ratio of properties explained by the QSAR model to those not explained by it. [c] SEE: standard error of estimate. [d] $BSA_{Total} = BSA_{Pol} + BSA_{NonPol}$.

coordinate system. The grid spacing was set to 2 Å units in all *x*, *y*, and *z* directions, and a C sp^3 atom with a formal charge of +1 and a van der Waals radius of 1.52 Å served as the probe. The steric term represents the van der Waals (dispersion-like) interactions, while the Coulombic term represents the electrostatic interactions for which a distance-dependent dielectric expression $\epsilon = \epsilon_0 R_{ij}$ with $\epsilon_0 = 1.0$ was adopted. The maximum field values were truncated to 30 kcal/mol for the steric field energies and ±30.0 kcal/mol for the electrostatic field energies. To improve efficiency and to reduce "noise", a column filter was employed to exclude the columns with a variance smaller than 2.0 kcal/mol.

Partial least-squares (PLS) regression[27] was employed to correlate these field values with the observed pK_I values. The "leave-one-out" cross-validation procedure[28] was repeated on the training-set compounds to yield the highest cross-validated r^2 (r_{cv}^2) and to determine the optimum number of principal components (PCs). The PLS analysis was then repeated without cross-validation to obtain a predictive model and associated conventional r^2 values from which the CoMFA coefficient contour plots for the steric and electrostatic potentials were generated. Linear regression analysis in all cases was carried out using the QSAR module of the Sybyl 6.6 program.[25]

Results and Discussion

A training set consisting of 49 THP inhibitors was selected from a recent experimental study by De Lucca et al. for which values of the inhibition constant (K_I) were determined.[29,30] These inhibition constants (K_I) were converted to logarithmic pK_I values for the regression studies. The molecular structures of these com-

pounds are shown in Table 1. Altogether, 17 QSAR models (designated by Roman numeral) were developed using different combinations of receptor-based and ligand-based descriptors, which include E_{compl}, $E_{solv}[I]$, BSA_{Pol}, BSA_{NonPol}, and CoMFA steric and electrostatic fields. A test set of compounds was selected from the data set to validate the predictive ability of the various regression models. This test set comprised 11 THPs (18% of total data set) that included compounds exhibiting poor, moderate, and high activity.

Individual linear regression models were first constructed between pK_I and E_{compl} plus the following combinations of other calculated descriptors: (I) "none", i.e., E_{compl} only, (II) $E_{solv}[I]$, (III) BSA_{Pol}, (IV) BSA_{NonPol}, (V) BSA_{Pol} and BSA_{NonPol}, (VI) $E_{solv}[I]$ and BSA_{Pol}, (VII) $E_{solv}[I]$ and BSA_{NonPol}, and (VIII) $E_{solv}[I]$, BSA_{Pol}, and BSA_{NonPol}. Predicted inhibition constants (pK_I) for the models are reported in Table 1. The correlation between pK_I and E_{compl} (model I) was mixed in terms of statistical performance measures. The model's internal predictive ability ($r_{cv}^2 = 0.52$) was above the $r_{cv}^2 \geq 0.50$ criterion of acceptability; however, its goodness of fit ($r^2 = 0.56$) to the experimentally observed pK_I values fell well below the corresponding $r^2 \geq 0.90$ criterion. Inclusion of $E_{solv}[I]$ (model II) produced sharp improvements in both r_{cv}^2 (0.82) and r^2 (0.84), indicative of the importance of inhibitor solvation effects for these inhibitors. Compared with model I, slight improvements were obtained for model III ($r_{cv}^2 = 0.57$, $r^2 = 0.60$), which includes BSA_{pol},

for model IV ($r_{cv}^2 = 0.63$, $r^2 = 0.67$), which includes BSA$_{NonPol}$, and for model V ($r_{cv}^2 = 0.62$, $r^2 = 0.68$), which includes both BSA$_{pol}$ and BSA$_{NonPol}$. Models VI, VII, and VIII, which correspond respectively to models III, IV, and V with the addition of E_{solv}[I], displayed significant improvements in statistical quality ($r_{cv}^2 > 0.82$, $r^2 > 0.84$) in every case. These improvements primarily reflect the inclusion of E_{solv}[I] alone, since it is evident that the statistical quality of models VI–VIII (which also include the BSA terms) is virtually identical to that of model II, which includes E_{solv}[I] but excludes any BSA terms.

A summary of the statistical parameters for models I–VIII is given in Table 2. Comparison of these models reveals that improved predictive ability (high r_{cv}^2) is obtained primarily by inclusion of E_{solv}[I] beyond E_{compl}. The strong linear correlation observed for models II, VI, VII, and VIII indicates that E_{solv}[I] and, to a lesser degree, both BSA$_{Pol}$ and BSA$_{NonPol}$ are valuable factors for consideration in predicting the binding affinity for this series of HIV PR inhibitors. By virtue of their ease of computation, they would also work well as simple terms in "docking and scoring" routines used to rank the free energy of binding of a ligand for a protein of known three-dimensional (3D) structure.

The pK_I values for the test set of 11 THP inhibitors were predicted using models I–VIII. The best agreement between the experimentally observed and predicted pK_I values was obtained for models II, VI, VII, and VIII. Models VI–VIII and II were nearly equivalent in terms of their statistical quality (r^2, r_{cv}^2) and performance in predicting the test-set compounds, providing further evidence of the significance of the solvation term E_{solv}[I]. For the sake of brevity, results for the test-set compounds are presented only for model II along with models I and V for reference (Table 3).

The differences (i.e., residuals) between the corresponding experimentally observed and predicted pK_I values from models II, VI, VII, and VIII are acceptably small (<1.5 log units) for all test-set compounds with the exception of compounds **55** and **56**, where model II (as well as models VI and VIII, not included in Table 3) overestimates the actual pK_I values by >2.25 log units. Nevertheless, these models are consistent with experimental results in predicting high pK_I values for these two compounds. It is interesting to note that both model IV ($r^2 = 0.67$; $r_{cv}^2 = 0.63$; not included in Table 3) and model V ($r^2 = 0.62$; $r_{cv}^2 = 0.68$), although inferior to model II ($r^2 = 0.84$; $r_{cv}^2 = 0.82$) in terms of statistical quality, were superior to model II in terms of predicting the pK_I values for compounds **55** and **56**. Model IV, which includes only the BSA$_{NonPol}$ term besides E_{compl}, predicted pK_I values of 10.89 for compound **55** (pK_I-[exptl] = 10.10) and 11.59 for compound **56** (pK_I[exptl] = 10.52). Model V, which includes the BSA$_{NonPol}$ and BSA$_{Pol}$ terms besides E_{compl}, predicted pK_I values of 10.83 for compound **55** and 11.51 for compound **56**. These results suggest that models that do not include E_{solv}[I], but do include the BSA terms, more accurately predict the pK_I values for these two compounds. A plausible explanation is that the magnitude of the E_{solv}-[I] term is overemphasized for these two compounds, both of which are structurally unique in some ways compared with the other THPs in this series. In

Table 3. Predicted pK_I Values for the Test-Set Inhibitors Using Models I, II, and V,[a] from Regressions Involving E_{compl}, E_{solv}[[I], BSA$_{pol}$, BSA$_{NonPol}$

Compound		pK_I	pK_I (predicted)		
No.	Structure	(exp.)	I	II	V
50		10.70	9.49	9.73	9.59
51		6.34	8.60	7.27	7.31
52		7.00	9.97	8.30	9.57
53		7.85	8.93	8.41	8.66
54		9.60	9.21	8.30	9.47
55		10.10	11.40	12.66	10.83
56		10.52	11.22	12.84	11.51
57		7.36	8.10	8.36	8.32
58		10.10	8.84	9.13	8.73
59		10.22	8.89	9.12	8.69
60		10.70	9.30	9.41	9.22

[a] See Results and Discussion for the description of the models.

particular, both compounds contain elongated, sterically bulky substituents located on either side of their symmetry axis that incorporate the Ph–NH–C(=O)–Ar moiety, where Ar represents the pyridine ring in **55** and the thiophene ring in **56**. The extended delocalization of these conjugated sequences could affect the true solvation energy and/or the GB/SA treatment of solvation for these two compounds, thus leading to the apparent overestimations in E_{solv}[I]. This issue is a subject for further investigation. For example, it might be informative to explore other solvation models besides the GB/SA solvation model employed in the present
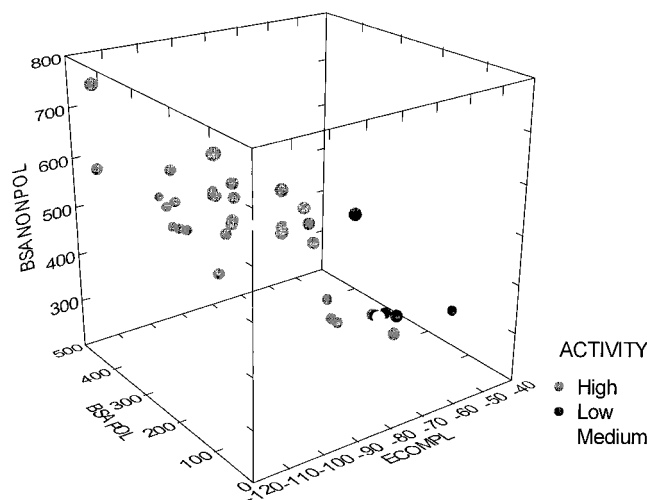
Figure 1. Plot of training-set THP inhibitors in 3D descriptor space according to their calculated values of E_{compl} (X axis), BSA$_{Pol}$ (Y axis), and BSA$_{NonPol}$ (Z axis). Inhibitory activity was divided into three categories: high activity (p$K_I \geq 8.5$), moderate activity (7.0 < pK_I < 8.5), and low activity (p$K_I \leq 7.0$). Each THP inhibitor was color-coded according to the following scheme: high activity (green), moderate activity (yellow), and low activity (red). Note that most of the inhibitors were separated into three distinct clusters according to activity.

study in order to determine whether the present results are dependent on the choice of solvation model.

To further explore the utility of the receptor-based descriptors E_{compl}, BSA$_{Pol}$, and BSA$_{NonPol}$ for predicting inhibitor binding affinity, each THP inhibitor of the training set was plotted in 3D descriptor space described by E_{compl} (X axis), BSA$_{Pol}$ (Y axis), and BSA$_{NonPol}$ (Z axis) using the statistical package SYSTAT.[31] Inhibitory activity was divided into three categories: high activity (p$K_I \geq 8.5$), moderate activity (7.0 < pK_I < 8.5), and low activity (p$K_I \leq 7.0$). In the 3D graph (Figure 1), each THP inhibitor was color-coded according to the following scheme: high activity (green), moderate activity (yellow), and low activity (red). Within this 3D descriptor space, it is seen that most of the inhibitors were separated into three distinct clusters according to activity. The only exceptions are compounds **9** and **41**, perhaps because of their missing P2′ substitutions, which may in turn explain why E_{compl} was underestimated in these two cases. This clustering analysis provides evidence that E_{compl}, BSA$_{Pol}$, and BSA$_{NonPol}$ might prove valuable as descriptors in classification studies for rapid screening of large compound libraries to identify promising candidates.

Various "enhanced" 3D-QSAR models were constructed to determine if, and to what extent, the inclusion of E_{compl}, BSA$_{Pol}$, BSA$_{NonPol}$, and E_{solv}[I] as additional descriptors to the default CoMFA steric–electrostatic fields would improve the statistical and predictive performance of the resultant 3D-QSAR models. Beginning with the conventional CoMFA-PLS model (model IX), eight such 3D-QSAR models were built from the training-set THP compounds by addition of the following descriptors: (model X) E_{compl}; (model XI) BSA$_{Pol}$; (model XII) BSA$_{NonPol}$; (model XIII) BSA$_{Pol}$ and BSA$_{NonPol}$; (model XIV) E_{solv}[I]; (model XV) E_{solv}[I] and BSA$_{Pol}$; (model XVI) E_{solv}[I] and BSA$_{NonPol}$; and (model XVII) E_{solv}[I], BSA$_{Pol}$, and BSA$_{NonPol}$. The 3D-QSAR

models were constructed as described above using PLS analysis accessed through the Sybyl 6.6 program.[25]

The corresponding predicted and experimentally observed pK_I values of the training-set compounds for models IX–XVII are listed in Table 4, and various statistical parameters are summarized in Table 5. Inclusion of E_{compl} alone as an additional descriptor to the CoMFA fields (viz., model X) resulted in a moderate improvement in internal predictive ability (i.e., from $r_{cv}^2 = 0.58$ for model IX to $r_{cv}^2 = 0.67$ for model X). Models XIV–XVII were uniformly superior among these 3D-QSAR models in terms of every calculated measure of statistical quality including goodness of fit ($r^2 \geq 0.96$), internal predictive ability ($r_{cv}^2 \geq 0.77$), and fewest number of PCs (two to four). As found above when comparing models I–VIII, inspection of these 3D-QSAR models (models IX–XVII) reveals that inclusion of E_{solv}-[I] as an additional descriptor led to this dramatic improvement in statistical quality. This is demonstrated most clearly by comparing model IX ($r_{cv}^2 = 0.58$), which contained only the default CoMFA steric–electrostatic field descriptors, with model XIV ($r_{cv}^2 = 0.80$), which included the addition of only E_{solv}[I]. Similar improvement (i.e., $r_{cv}^2 \approx 0.80$) was found for models XV–XVII, which include the various BSA terms besides E_{solv}[I] as additional descriptors. At the same time, inclusion of the BSA terms by themselves did little or nothing to improve the statistical quality of the 3D-QSAR models. This is apparent by comparing model IX (i.e., CoMFA fields alone) with models XI–XIII in which the various BSA terms were added sequentially and then cumulatively (Table 5).

Models IX–XVII were further validated by predicting the pK_I values of the 11 test-set compounds (Table 6). Small residuals between the predicted and experimentally observed pK_I values were obtained in all cases, thus strengthening our confidence in the predictive ability of these models. Model IX, which contained only the default CoMFA field descriptors, overestimated the pK_I value for compound **51**. Models XIV–XVII, all of which share in common the addition of the E_{solv}[I] term, predict pK_I values for this compound that are in closer agreement with experimental values. Somewhat paradoxically, models XIV–XVII tended to overestimate the pK_I values for the highly active test-set compounds **55** and **56**. As discussed earlier with respect to models II, VI, and VIII, which similarly overestimated pK_I for the same two compounds, it is plausible that the unique characteristics of these molecules' structure and/or the GB/SA solvation model may play a role. In this context, it is worth noting that the pK_I values for these two compounds are more accurately predicted by models IX and X–XIII in which the E_{solv}[I] term is absent.

Conclusions

Results from the present study demonstrate the utility of ligand-based and receptor-based computed descriptors as predictors of ligand–receptor binding affinity. Calculated values of inhibitor–receptor complexation energy (E_{compl}), alone and together with various combinations of the inhibitor solvation energy E_{solv}[I] and both polar and nonpolar buried surface areas (BSA$_{Pol}$ and BSA$_{NonPol}$, respectively), were found to perform well ($r_{cv}^2 > 0.5$) in predicting the observed

Table 4. Inhibition Constants (pK_I) from Experiment and from Linear Regression Models IX–XVII[a] Involving CoMFA Fields, E_{compl}, E_{solv}, BSA$_{Pol}$, and BSA$_{NonPol}$ for the Training-Set Inhibitors

compd[b]	pK_I (exptl)	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII
						pK_I (predicted)				
1	7.82	8.14	8.12	8.14	8.34	8.04	7.87	7.79	7.89	7.80
2	7.64	7.36	7.64	7.57	7.38	7.47	7.40	7.27	7.43	7.45
3	6.01	5.94	6.16	6.05	6.06	6.04	5.99	6.24	6.05	6.20
4	6.73	6.64	6.66	6.85	6.75	6.84	6.73	6.59	6.69	6.57
5	6.27	6.72	6.48	6.80	6.69	6.72	6.74	6.69	6.65	6.50
6	6.49	7.20	7.08	7.20	7.23	7.34	7.29	7.33	7.22	7.33
7	7.18	7.15	7.32	7.35	7.19	7.46	7.59	7.66	7.61	7.75
8	7.82	8.03	7.92	7.82	7.80	7.58	8.09	7.87	7.96	7.82
9	8.59	7.20	7.08	7.33	7.27	7.54	7.16	6.99	6.99	7.13
10	6.80	8.04	8.16	8.29	8.04	8.14	8.46	8.31	8.31	8.27
11	8.21	7.96	7.96	7.87	7.99	8.00	7.86	7.86	7.86	8.20
12	7.96	7.58	7.51	7.32	7.42	7.39	7.62	7.58	7.58	7.52
13	9.04	8.79	8.83	8.70	8.75	8.69	8.71	8.61	8.61	8.79
14	9.31	9.23	9.45	9.33	9.29	9.54	9.18	9.07	9.07	9.19
15	10.05	9.81	9.84	9.60	9.63	9.48	9.75	9.88	9.88	9.77
16	9.51	9.99	9.82	9.88	10.02	9.90	9.83	9.84	9.84	9.80
17	9.82	9.88	9.82	9.70	9.80	9.66	9.89	9.88	9.88	9.77
18	8.85	9.10	9.17	9.08	9.02	9.19	8.99	9.01	9.01	8.93
19	10.70	10.69	10.66	10.55	10.51	10.60	10.83	10.81	10.81	10.86
20	10.70	10.76	10.71	10.56	10.62	10.63	10.60	10.51	10.51	10.56
21	11.00	10.77	10.84	11.07	11.133	10.85	10.97	10.99	10.99	10.82
22	10.22	10.22	10.23	10.37	10.26	10.21	10.42	10.37	10.37	10.26
23	8.77	8.83	8.88	8.95	8.96	9.03	8.86	8.84	8.84	8.89
24	8.10	7.96	7.88	7.78	7.96	7.94	7.99	8.10	8.10	7.93
25	8.31	8.56	8.60	8.60	8.62	8.63	8.75	8.65	8.65	8.82
26	10.00	10.16	10.24	10.30	10.16	10.16	9.98	10.03	10.03	9.97
27	10.22	10.34	10.37	10.33	10.45	10.43	10.48	10.61	10.61	10.56
28	10.00	9.90	9.96	9.99	9.87	10.15	9.94	9.90	9.90	9.99
29	10.52	10.63	10.66	10.57	10.69	10.64	10.53	10.55	10.55	10.65
30	9.39	9.22	9.19	9.13	9.19	9.11	9.11	9.19	9.19	9.14
31	7.64	7.69	7.57	7.50	7.63	7.45	7.60	7.76	7.76	7.74
32	9.52	9.03	8.96	9.07	9.11	9.12	9.21	9.20	9.20	9.12
33	7.92	7.72	7.91	7.97	7.85	7.96	8.07	8.09	8.09	8.07
34	7.96	8.17	8.06	8.01	8.14	8.12	7.85	7.98	7.98	7.84
35	6.96	7.71	7.69	7.71	7.65	7.65	7.72	7.73	7.73	7.76
36	8.25	8.42	8.42	8.33	8.28	8.23	8.27	8.31	8.31	8.29
37	7.85	8.10	8.04	7.89	7.98	7.86	7.70	7.68	7.68	7.67
38	9.62	9.29	9.29	9.53	9.46	9.48	9.56	9.54	9.54	9.54
39	8.85	8.81	8.78	8.89	8.95	8.88	8.99	8.92	8.92	8.98
40	9.42	9.49	9.49	9.19	9.29	9.40	9.18	9.13	9.13	9.27
41	8.92	8.97	8.98	8.92	9.00	8.95	8.64	8.71	8.71	8.69
42	7.48	7.77	7.84	7.64	7.79	7.67	7.71	7.72	7.72	7.63
43	7.40	7.69	7.57	7.66	7.83	7.60	7.46	7.54	7.54	7.47
44	9.19	8.99	9.09	9.24	9.07	9.06	9.10	9.12	9.12	9.17
45	9.55	9.51	9.49	9.32	9.16	9.62	9.69	9.55	9.55	9.81
46	10.00	10.05	10.15	10.30	10.03	10.27	10.05	10.07	10.07	10.02
47	9.89	9.77	9.62	9.61	9.74	9.21	9.60	9.69	9.69	9.54
48	8.80	8.84	8.75	8.78	8.79	8.89	8.95	8.94	8.94	8.88
49	10.00	9.99	9.97	10.16	10.00	10.03	10.13	10.10	10.10	10.13

[a] See Results and Discussion for the description of the models. [b] See Table 1 for the structures of compounds.

Table 5. Statistical Parameters Associated with Linear Regression Models IX–XVII[a] Constructed from CoMFA Fields Plus Various Combinations of E_{compl}, E_{solv}[I], BSA$_{Pol}$, BSA$_{NonPol}$, and BSA$_{Total}$ Descriptors for the Training Set of THP Compounds

model	r^2	r_{cv}^2	no. of PCs	F^b	SEE^c	E_{compl}	E_{solv}[I]	BSA$_{Pol}$	BSA$_{NonPol}$	BSA$_{Total}$[d]
						descriptors added to CoMFA fields				
IX	0.96	0.58	9	94.0	0.29					
X	0.96	0.67	9	103.0	0.28	√				
XI	0.95	0.60	8	79.0	0.31	√		√		
XII	0.95	0.51	9	75.0	0.32	√			√	
XIII	0.94	0.56	9	64.0	0.35	√				√
XIV	0.97	0.80	2	141.0	0.24	√	√			
XV	0.96	0.78	3	103.0	0.28	√	√	√		
XVI	0.97	0.79	2	121.0	0.26	√	√		√	
XVII	0.96	0.77	4	100.0	0.28	√	√			√

[a] See Results and Discussion for description of the models. [b] F ratio: defined as $r^2/(1 − r^2)$, representing the ratio of properties explained by the QSAR model to those not explained by it. [c] SEE: standard error of estimate. [d] BSA$_{Total}$ = BSA$_{Pol}$ + BSA$_{NonPol}$.

inhibition constants (as pK_I) for this series of THP HIV-1 protease inhibitors (models I–VIII). The most dramatic improvement in predictive ability ($r_{cv}^2 > 0.8$) was gained by inclusion of E_{solv}[I] (models II, VI–VIII).

E_{solv}[I] was demonstrated to serve as a justified and computationally thrifty surrogate for E_{solv} (see eq 6) in capturing the salient features of solvation and desolvation effects. Since E_{solv}[I] pertains exclusively to the

Table 6. Predicted pK_I Values for the Test-Set Inhibitors Using Models IX, X, XIII, XIV, and XVII[a] from Regressions Involving CoMFA Fields, E_{compl}, $E_{solv}[I]$, BSA_{pol}, and BSA_{NonPol}

compd[b]	pK_I (exptl)	pK_I (predicted)				
		IX	X	XIII	XIV	XVII
50	10.70	10.37	10.34	10.55	10.65	10.63
51	6.34	8.29	8.77	8.01	7.39	7.71
52	7.00	7.96	8.53	7.92	7.69	7.59
53	7.85	7.89	7.72	7.82	7.61	7.68
54	9.60	8.99	9.28	9.40	8.39	8.34
55	10.10	10.04	11.16	10.49	12.19	12.70
56	10.52	10.22	10.98	10.81	12.28	12.55
57	7.36	7.96	8.15	8.73	8.90	9.20
58	10.10	8.82	9.06	9.18	9.29	9.43
59	10.22	9.08	8.69	8.79	8.94	8.94
60	10.70	9.61	9.64	9.63	9.15	9.02

[a] See Results and Discussion for the description of the models.
[b] See Table 3 for the structures of the test-set compounds.

small-molecule ligand and judiciously neglects the large protein molecule, its computation even for a sizable library of compounds would be straightforward and fast. The advantages of $E_{solv}[I]$ over E_{solv} in terms of speed, convenience, and accessibility suggest its utility in "docking and scoring" schemes that are employed in drug-design programs for high-throughput screening of massive small-molecule libraries in search of new leads.

Consistent with our findings from a similar study on a separate class of HIV-1 protease inhibitors,[16] the present study demonstrates the utility of E_{compl} as a descriptor for assessing drug-receptor binding affinity. Computation of E_{compl} requires knowledge of the three-dimensional (3D) structure of the receptor or, preferably, the ligand—receptor complex. Since it is a receptor-based descriptor, calculation of E_{compl} is more computer-intensive than $E_{solv}[I]$ and sometimes, such as in cases where the structure of the receptor is unknown, inaccessible. As shown in both the present study and the aforementioned study,[16] calculation of E_{compl} even for moderately large data sets (<500 compounds) can be performed in a fast and efficient manner on a single workstation.

Correlation of pK_I with the various BSA terms (polar, nonpolar, total) was less apparent than with E_{compl} and $E_{solv}[I]$. This conclusion is supported by comparison of model I with models III–V (Table 2), where it is seen that the statistical parameters r^2 and r_{cv}^2 improve only marginally by addition of the BSA_{Pol}, BSA_{NonPol}, or BSA_{Total} terms to E_{compl}. Nevertheless, addition of the BSA terms did improve the predictive ability of these models for several of the test-set compounds (Table 3). For example, the pK_I value (exptl = 6.34) for the weakly active compound **51** was more accurately predicted by model V (7.31) than by model I (8.60).

The correlation of pK_I with $E_{solv}[I]$ and, to a much lesser degree, the BSA descriptors for this series of THP inhibitors are reflected in the CoMFA models when these properties were included as additional descriptors to CoMFA's default steric—electronic field descriptors (models IX–XVII). The addition of $E_{solv}[I]$ values augmented the statistical quality of these 3D-QSAR models dramatically. For example, the statistical quality of model XIV ($r_{cv}^2 = 0.80$, using only two PCs) is vastly superior to that of model IX ($r_{cv}^2 = 0.58$, using nine PCs).

References

(1) Henderson, L. E.; Benveniste, R. E.; Sowder, R.; Copeland, T. D.; Schultz, A. M.; Oroszlan, S. Molecular characterization of gag proteins from simian immunodeficiency virus (SIVMne). *J. Virol.* **1988**, *62*, 2587–2595.

(2) Loeb, D. D.; Hutchison, C. A.; Edgell, M. H., III; Farmerie, W. G.; Swanstrom, R. Mutational Analysis of Human Immunodeficiency Virus Type 1 Protease Suggests Functional Homology with Aspartic Proteinases. *J. Virol.* **1989**, *63*, 111–121.

(3) Crawford, S.; Goff, S. P. A Deletion Mutation in the 5′ Part of the pol Gene of Moloney Murine Leukemia Virus Blocks Proteolytic Processing of the gag and pol Polyproteins. *J. Virol.* **1985**, *53*, 899–907.

(4) Peng, C.; Ho, B. K.; Chang, T. W.; Chang, N. T. Role of Human Immunodeficiency Virus Type 1-Specific Protease in Core Protein Maturation and Viral Infectivity. *J. Virol.* **1989**, *63*, 2550–2556.

(5) McQuade, T. J.; Tomasselli, A. G.; Liu, L.; Karacostas, V.; Moss, B.; Sawyer, T. K.; Heinrikson, R. L.; Tarpley, W. G. A Synthetic HIV-1 Protease Inhibitor with Antiviral Activity Arrests HIV-like Particle Maturation. *Science.* **1990**, *247*, 454–456.

(6) Babine, R. E.; Bender, S. L. Molecular Recognition of Protein—Ligand Complexes: Application to Drug Design. *Chem. Rev.* **1997**, *97*, 1359–1472.

(7) Boden, D.; Markowitz, M.; Resistance to Human Immunodeficiency Virus Type 1 Protease Inhibitors. *Antimicrob. Agents Chemother.* **1998**, *42*, 2775–2783.

(8) Condra, J. H.; Schleif, W. A.; Blahy, O. M.; Gadryelski, L. J.; Graham, D. J.; Quintero, J. C.; Rhodes, A.; Robbins, H. L.; Roth, E.; Shivaprahash, M.; Titus, D.; Yang, T.; Teppler, H.; Squires, K. E.; Deutsch, P. J.; Emini, E. A. In Vivo Emergence of HIV-1 Variants Resistant to Multiple Protease Inhibitors. *Nature* **1995**, *374*, 569–571.

(9) Ala, P. J.; Huston, E.; Klabe, M.; McCabe, D. D.; Duke, J. L.; Rizzo, C. J.; Korant, B. D.; DeLoskey, R. J.; Lam, P. Y. S.; Hodge, N. C.; Chang, C.-H. Molecular Basis of HIV-1 Protease Drug Resistance: Structural Analysis of Mutant Proteases Complexed with Cyclic Urea Inhibitors. *Biochemistry* **1997**, *36*, 1573–1580.

(10) Holloway, M. C.; Wai, J. M.; Halgren, T. A.; Fitzgerald, P. M. D.; Vacca, J. P.; Dorsay, 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, T. A.; Sanders, W. M.; Tucker, T. J.; Wiggins, M.; Wiscount, 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.

(11) Perez, C.; Pastor, M.; Ortiz, R. A.; Gago, F. Comparative Binding 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.

(12) Miertus, S.; Antcheva, N.; Tossi, A.; Bizik, F.; Benedetti, F.; Gennaro, R.; Romeo, D. CADD of New Pseudo-Peptide Inhibitors of HIV-1 Aspartic Protease. In *QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications*; Sanz, F., Giraldo, J., Manaut, F., Eds.; Prous Science Publishers: Barcelona, 1995; pp 579–582.

(13) Tossi, A.; Antcheva, N.; Romeo, D.; Miertus, S. Development of Pseudopeptide Inhibitors of HIV-1 Aspartic Protease: Analysis and Tuning of the Subsite Specificity. *Pept. Res.* **1995**, *8*, 328–334.

(14) Miertus, S.; Furlan, M.; Tossi, A.; Romeo, D. Design of New Inhibitors of HIV-1 Aspartic Protease. *Chem. Phys.* **1996**, *204*, 173–180.

(15) Nair, A. C.; Miertus, S.; Tossi, A.; Romeo, D. A Computational Study of the Resistance of HIV-1 Aspartic Protease to the Inhibitors ABT-538 and VX-478 and Design of New Analogues. *Biochem. Biophys. Res. Commun.* **1998**, *242*, 545–551.

(16) Jayatilleke, P. R. N.; Nair, A. C.; Zauhar, R.; Welsh, W. J. Computational Studies on HIV-1 Protease Inhibitors: Influence of Calculated Inhibitor—Enzyme Binding Affinities on the Statistical Quality of 3D-QSAR Comparative Molecular Field Analysis (CoMFA) Models. *J. Med. Chem.* **2000**, *43*, 4446–4451.

(17) Viswanadhan, V. N.; Reddy, M. R.; Wlodawer, A.; Varney, M. D.; Weinstein, J. N. An Approach to Rapid Estimation of Relative Binding Affinities of Enzyme Inhibitors: Application to Peptidomimetic Inhibitors of the Human Immunodeficiency Virus Type 1 Protease. *J. Med. Chem.* **1996**, *39*, 706−712.

(18) McCarrick, M. A.; Kollman, P. A. Predicting Relative Binding Affinities of Non-peptide HIV Protease Inhibitors with Free Energy Perturbation Calculations. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 109−121.

(19) Reddy, M. R.; Erion, M. D. Structure-Based Drug Design Approaches for Predicting Binding Affinities of HIV1 Protease Inhibitors. *J. Enzymol. Inhib.* **1998**, *14*, 1−14.

(20) Hilgeroth, A.; Fleischer, R.; Wiese, M.; Heinemann, F. W. Comparison of Azacyclic Urea A-98881 as HIV-1 Protease Inhibitor with Cage Dimeric *N*-Benzyl 4-(4-methoxyphenyl)-1,4-dihydropyridine as Representative of a Novel Class of HIV-1 Protease Inhibitors: A Molecular Modeling Study. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 233−242.

(21) Debnath, A. K. Three-Dimensional Quantitative Structure−Activity Relationship Study on Cyclic Urea Derivatives as HIV-1 Protease Inhibitors: Application of Comparative Molecular Field Analysis. *J. Med. Chem.* **1999**, *42*, 249−259; **2000**, *43*, 764.

(22) Debnath, A. K. Comparative Molecular Field Analysis (CoMFA) of a Series of Symmetrical Bis-Benzamide Cyclic Urea Derivatives as HIV-1 Protease Inhibitors. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 761−767.

(23) *Insight II Molecular Modeling Package* and *Discover 2.98 Simulation Package*; MSI, Inc.: San Diego, CA.

(24) *Macromodel 6.0 Modeling Package*; Schrödinger, Inc.: Portland, OR.

(25) *Sybyl 6.6 Modeling Package*; Tripos, Inc.: St. Louis, MO.

(26) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativities: A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.

(27) Wold, S.; Albano, C.; Dunn, W. J., III; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjostrom, M. *Multivariate Data Analysis in Chemistry. Chemometrics: Mathematics and Statistics in Chemistry*; Kowalski, B., Ed.; Reidel: Dordrecht, The Netherlands, 1984.

(28) (a) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E.; Frank, I. E. Bootstrapping, Crossvalidation and Partial Least Squares. *Quant. Struct.−Act. Relat.* **1988**, *7*, 18−25. (b) *Sybyl 6.6 QSAR Reference Manual*; Tripos, Inc.: St. Louis, MO.

(29) De Lucca, G. V.; Liang, J.; De Lucca, I. Stereospecific Synthesis, Structure−Activity Relationship, and Oral Bioavailability of Tetrahydropyrimidine-2-one HIV Protease Inhibitors. *J. Med. Chem.* **1999**, *42*, 135−152.

(30) De Lucca, G. V.; Liang, J.; Aldrich, P. E.; Calabrese, J.; Cordova, B.; Klabe, R. M.; Rayner, M. M.; Chang, C.-H. Design, Synthesis, and Evaluation of Tetrahydropyrimidinones as an Example of a General Approach to Nonpeptide HIV Protease Inhibitors. *J. Med. Chem.* **1997**, *40*, 1707−1719.

(31) *SYSTAT Statistical Package*, version 8.0; SPSS, Inc.: Chicago, IL.