

# CoMFA and CoMSIA 3D QSAR and Docking Studies on Conformationally-Restrained Cinnamoyl HIV-1 Integrase Inhibitors: Exploration of a Binding Mode at the Active Site

John K. Buolamwini\* and Haregewein Assefa

Department of Pharmaceutical Sciences, College of Pharmacy, University of Tennessee Health Science Center, 847 Monroe Avenue Suite 327, Memphis, Tennessee 38163

Received August 22, 2001

Anti-HIV (human immunodeficiency virus) drug discovery has been increasingly focusing on HIV integrase (IN) as a potential therapeutic target. This enzyme is required for the integration of reverse transcribed proviral DNA into the host cell's genome and is essential for the propagation of the HIV life cycle. Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) three-dimensional quantitative structure–activity relationship (3D QSAR) studies and docking simulations were conducted on a series of potent conformationally restrained cinnamoyl inhibitors of HIV-1 IN (Artico; et al. *J. Med. Chem.* **1998**, *41*, 3948–3960). Predictive 3D QSAR models were established using SYBYL multifit molecular alignment rule, which had conventional  $r^2$  and cross-validated coefficient ( $q^2$ ) values up to 0.981 and 0.721 for CoMFA and 0.975 and 0.804 for CoMSIA, respectively. These models were validated by an external test set (Burke; et al. *J. Med. Chem.* **1995**, *38*, 4171–4178). CoMFA and CoMSIA 3D QSAR models were also derived using a molecular alignment obtained by docking the compounds into the active site of HIV IN. These latter models were comparable to multifit-derived models in terms of relative descriptor field contributions and the partial least squares (PLS) contour maps. The CoMSIA 3D QSAR models performed better than the CoMFA models. The superior performance of CoMSIA was attributed to the large contribution of hydrogen-bonding interactions to the inhibitory activity differences among the compounds. This was supported by FlexX binding energy scores that correlated well with the inhibitory activity differences between hydroxylated compounds and their corresponding methoxy or deoxy counterparts. The CoMFA and CoMSIA PLS contour maps and MOLCAD-generated active site electrostatic, lipophilicity, and hydrogen-bonding potential surface maps, as well as the docking results, were integrated to propose a binding mode for the cinnamoyl inhibitors at the active site of HIV-1 IN.

## Introduction

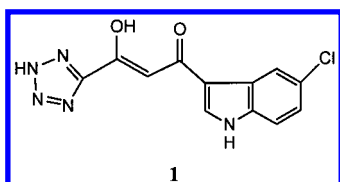
Acquired immunodeficiency syndrome (AIDS) is a formidable pandemic that is still wreaking havoc worldwide. Targeting the human immunodeficiency virus (HIV), a retrovirus that is implicated in the etiology of AIDS, has been useful in AIDS chemotherapy. Upon infection, HIV enters CD4 positive T lymphocytes by binding to the CD4 receptor via its coat protein, p120. Following the infection, this retrovirus uses three key enzymes to propagate its life cycle. These enzymes are (i) reverse transcriptase (RT), an RNA-dependent DNA polymerase that is used by the virus to transcribe the viral genomic RNA into proviral DNA for incorporation into the host cell DNA; (ii) integrase (IN), which is the enzyme responsible for insertion of the proviral DNA into the host cell genome; and (iii) HIV protease (PR), the enzyme necessary for the processing and packaging of new virulent viral particles for exiting the host cell.<sup>1,2</sup> Clinical agents that target RT and PR have been successfully developed and are used in combination to keep the HIV virus under control for many AIDS patients and improve the lives of AIDS patients.<sup>3</sup> However, targeting RT and PR still does not eliminate

the virus from patients,<sup>4,5</sup> making it necessary to explore other targets. Attention has recently been focused on the HIV IN enzyme, which is an attractive potential target for anti-HIV drug development because it is essential for viral replication and productive infection<sup>6,7</sup> and does not appear to have a counterpart in human cells. Furthermore, compounds that unambiguously inhibit HIV replication in cell culture by targeting IN have been identified,<sup>8</sup> and an oligonucleotide<sup>9</sup> purported to inhibit HIV IN, Zentivir, has been undergoing phase I/II clinical trials in AIDS patients.<sup>10</sup> Much effort is being expended in the discovery of small molecule IN inhibitors<sup>11</sup> for therapeutic use, which would be better drug candidates than the oligonucleotides in regards to bioavailability and metabolic stability considerations. Various approaches have been adopted in the search for small molecule IN inhibitors. These include compound library screening of both synthetic compounds and natural products<sup>12</sup> and computer-aided drug design methods such as three-dimensional (3D) database searching<sup>13,14</sup> and molecular docking into the IN catalytic core.<sup>15</sup> A diverse array of molecules have been identified as IN inhibitors, with polyhydroxylated aromatic compounds being quite well-represented.<sup>11,16</sup> No drugs targeted to IN are yet on the market due to the fact

\* To whom correspondence should be addressed. Tel: (901)448-7533. Fax: (901)448-6828. E-mail: jbuolamwini@utmem.edu.

that the inhibitors identified to date lack selectivity,<sup>2</sup> creating the need for the discovery of new inhibitors. This requires insights into the factors that influence inhibitor potency and selectivity toward HIV IN to guide drug design.

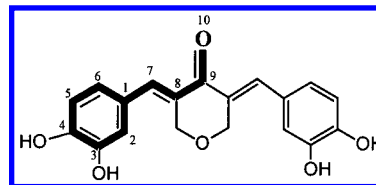
Three-dimensional quantitative structure–activity relationship (3D QSAR) methods, such as comparative molecular field analysis (CoMFA),<sup>17</sup> have been successfully applied in many instances to guide the design of new bioactive molecules.<sup>18</sup> The application of 3D QSAR methodology to the design of HIV IN inhibitors has received relatively little attention.<sup>19</sup> The present study has extended CoMFA and the more recently introduced comparative molecular similarity indices analysis (CoMSIA)<sup>20,21</sup> 3D QSAR methodologies to potent conformationally restrained cinnamoyl HIV IN inhibitors<sup>21</sup> to gain insights into how steric, electrostatic, hydrophobic, and hydrogen-bonding interactions influence their activity, and to derive predictive 3D QSAR models for designing and forecasting the activity of IN inhibitors of this class. It has been previously surmised that subtle steric and electrostatic effects come into play regarding the bioactivity of these compounds, the nature of which requires additional investigation.<sup>22</sup> In addition to the 3D QSAR analyses, docking simulations were performed using the only published X-ray crystallographic structure of the human IN catalytic core in complex with an inhibitor,<sup>23</sup> to explore the binding modes of these compounds at the IN active site. Interest for this study also stems from the presence of the 1,3-keto–enol moiety in many of the compounds, a common structural feature found in potent and potentially useful HIV IN inhibitors<sup>8,23</sup> such as 1-(5-chloroindol-3-yl)-3-hydroxy-3-(2*H*-tetrazo-5-yl)-propenone (CITEP, **1**), the ligand in the published X-ray structure of the human IN catalytic core–ligand complex.<sup>23</sup>



## Materials and Methods

**Data Sets.** The training set used comprises a series of conformationally restrained cinnamoyl derivatives, which have been shown to be potent specific inhibitors of IN relative to other nucleic acid processing enzymes, namely, HIV RT and RNA polymerase.<sup>22</sup> An attractive feature of the compounds in the training set is their relative conformational rigidity, which makes them more amenable to meaningful CoMFA and CoMSIA analyses than flexible molecules. Furthermore, X-ray crystallographic structural representatives are available that provide templates for modeling the compounds.<sup>22</sup> The IC<sub>50</sub> (concentration causing 50% inhibitory effect for inhibiting 3'-processing by IN) values for the training set<sup>22</sup> were converted to pIC<sub>50</sub> (–log IC<sub>50</sub>) values and used as dependent variables in the CoMFA and CoMSIA QSAR analyses (see Table 1). The test set was composed of 23 compounds, and their experimental HIV-1 IN inhibitory data were taken from a report by Burke et al.<sup>24</sup> (see Table 2).

**Molecular Modeling and Alignment.** Three-dimensional structure building and all modeling were performed using the SYBYL program package, version 6.7<sup>25</sup> on a Silicon Graphics Octane (R1200) workstation with the IRIX 6.5 operating system. Conformations of compounds in the training set were generated using torsion angles ( $\tau$  values) indicated for the



**Figure 1.** Compound used as template for molecular alignment showing; in bold face, and numbered 1–10, are shown the reference atoms used in multifit alignment protocol of the SYBYL program.

same or similar compounds in the Cambridge Crystallographic Structural Database as provided by Artico et al.<sup>22</sup> Energy minimizations were performed using the Tripos force field<sup>26</sup> with a distance-dependent dielectric and the Powell conjugate gradient algorithm with a convergence criterion of 0.01 kcal/(mol Å). Partial atomic charges were calculated using the semiempirical program MOPAC 6.0 and applying the PM3 Hamiltonian.<sup>27</sup>

CoMFA and CoMSIA studies require that the 3D structures of the molecules to be analyzed be aligned according to a suitable conformational template, which is assumed to be a “bioactive” conformation.<sup>17</sup> The molecular alignments used for the studies were obtained by means of the SYBYL “multifit” alignment function or FlexX-docked alignment. For the multifit alignment, the default SYBYL spring constant value of 20 was used, and compound **2**, one of the most active and conformationally constrained analogues, was used as the alignment template. The reference atoms in compound **2** used for alignment were as follows: (i) the six aromatic carbon atoms of the common catechol or its methylated or deoxy counterparts and (ii) heavy atoms of the adjoining  $\alpha,\beta$ -unsaturated carbonyl moiety (see atoms numbered 1–10 in Figure 1). For the FlexX-docked alignment, conformations were selected from the “east–west” binding orientation at the active site of IN, taking into consideration the FlexX ranking, consensus scores, and molecular geometry.

**CoMFA and CoMSIA 3D QSAR Models.** In deriving the CoMFA and CoMSIA descriptor fields, a 3D cubic lattice with grid spacing of 2 or 1 Å and extending 4 Å units beyond the aligned molecules in all directions was created to encompass the aligned molecules. CoMFA descriptors were calculated using an sp<sup>3</sup> carbon probe atom with a van der Waals radius of 1.52 Å and a charge of +1.0 to generate steric (Lennard–Jones 6-12 potential) field energies and electrostatic (Coulombic potential) fields with a distance-dependent dielectric at each lattice point. The SYBYL default energy cutoff of 30 kcal/mol was used as well as cutoffs of 25, 20, and 15 kcal/mol. The CoMFA steric and electrostatic fields generated were scaled by the CoMFA–STD method in SYBYL.

CoMSIA similarity indices descriptors were derived according to Klebe et al.<sup>20</sup> with the same lattice box as was used for the CoMFA calculations, with a grid spacing of 2 or 1 Å employing a C<sup>1+</sup> probe atom with a radius of 1.0 Å as implemented in SYBYL. CoMSIA similarity indices ( $A_F$ ) for a molecule  $j$  with atoms  $i$  at a grid point  $q$  are calculated by eq 1 as follows:

$$A_{F,k}^q(j) = -\sum \omega_{\text{probe},k} \omega_{ik} e^{-\alpha r} \quad (1)$$

Five physicochemical properties  $k$  (steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor) were evaluated using the probe atom. A Gaussian type distance dependence was used between the grid point  $q$  and each atom  $i$  of the molecule. The default value of 0.3 was used as the attenuation factor ( $\alpha$ ). In CoMSIA, the steric indices are related to the third power of the atomic radii, the electrostatic descriptors are derived from atomic partial charges, the hydrophobic fields are derived from atom-based parameters developed by Viswanadhan et al.,<sup>28</sup> and the hydrogen bond donor and acceptor indices are obtained by a rule-based method derived from experimental values.<sup>29</sup>

**Table 1.** Structures, HIV-1 IN Inhibitory Activity, and FlexX Energy Scores of Compounds in the Training Set

Compd	Structure	pIC <sub>50</sub> <sup>a</sup>	FlexX <sup>b</sup> Energy	Compd	Structure	pIC <sub>50</sub> <sup>a</sup>	FlexX <sup>b</sup> Energy
2		6.70	-18.50	14		6.15	-18.90
3		6.70	-18.70	15		4.52	-14.90
4		6.70	-15.50	16		4.00	-13.80
5		6.05	-15.40	17		4.00	-12.60
6		4.00	-14.90	18		4.00	-16.80
7		6.00	-12.70	19		4.70	-17.60
8		6.22	-16.62	20		5.89	-14.80
9		4.43	-14.40	21		4.00	-13.20
10		4.00	-14.40	22		5.64	-13.10
11		4.00	-12.40	23		5.52	-15.60
12		6.10	-17.10	24		4.07	-12.10
13		4.52	-12.80	25		4.00	-12.80

<sup>a</sup> pIC<sub>50</sub> (−log IC<sub>50</sub>) values were derived from biological activity data reported by Artico et al.<sup>22</sup> <sup>b</sup> FlexX total energy score in kcal/mol.

The CoMFA and CoMSIA descriptors were used as independent variables, and pIC<sub>50</sub> values were used as dependent variables in partial least squares (PLS) regression analyses to derive 3D QSAR models using the standard implementation in the SYBYL package. The predictive value of the models was evaluated first by leave-one-out (LOO) cross-validation. The cross-validated coefficient,  $q^2$ , was calculated using eq 2

$$q^2 = 1 - \frac{\sum (Y_{\text{predicted}} - Y_{\text{observed}})^2}{\sum (Y_{\text{observed}} - Y_{\text{mean}})^2} \quad (2)$$

where  $Y_{\text{pred}}$ ,  $Y_{\text{actual}}$ , and  $Y_{\text{mean}}$  are predicted, actual, and mean values of the target property (pIC<sub>50</sub>), respectively.  $\sum (Y_{\text{pred}} - Y_{\text{actual}})^2$  is the predictive sum of squares (PRESS). To maintain the optimum number of PLS components and minimize the tendency to over fit the data, the number of components corresponding to the lowest PRESS value was used for deriving the final PLS regression models.<sup>30</sup> In addition to the  $q^2$  and the corresponding PRESS and number of components, the conventional correlation coefficient  $r^2$  and its standard error  $s$  were also computed. CoMFA and CoMSIA coefficient maps were generated by interpolation of the pairwise products between the PLS coefficients and the standard deviations of the corresponding CoMFA or CoMSIA descriptor values.

**Molecular Docking.** The FlexX program<sup>31</sup> version 1.8 interfaced with SYBYL 6.7 was used to dock the compounds to the active site of IN. FlexX is a fast automated docking

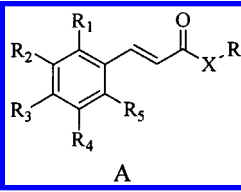
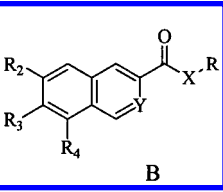
program that considers ligand conformational flexibility by an incremental fragment placing technique.<sup>31,32</sup> We used the program to dock the training set molecules into the active site of monomeric unit "A" of the crystal structure of the HIV-1 IN catalytic core. The 3D coordinates of the IN active site were taken from the recently reported X-ray crystal structure of the HIV-1 IN catalytic core in complex with an inhibitor deposited in the Brookhaven Protein Databank (PDB code: 1QS4).<sup>23</sup> The missing residues at positions 141–144 were inserted by means of the loop search algorithm in the SYBYL BIOPOLYMER module. The active site for docking was defined as all amino acids within 6.5 Å proximity of the cocrystallized ligand (compound 1).

**Molecular Surface Physicochemical Properties.** Surface physicochemical property maps, i.e., electrostatic potential, hydrophobicity (lipophilicity) potential, and hydrogen bonding (donor/acceptor) potential maps of the IN active site were generated on the solvent assessable (connolly) surface using the MOLCAD program in SYBYL. These surface property maps were examined for the determination of complementarity with CoMFA and CoMSIA PLS contour maps.

## Results and Discussion

**3D QSAR Modeling.** PLS, the statistical method used in deriving the 3D QSAR models, is a variation of principal component regression<sup>33</sup> in which the original variables are replaced by a small set of linear combina-

**Table 2.** Structures and HIV-1 IN Inhibitory Activity of Compounds in the Test Set

<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>A</p> </div> <div style="text-align: center;">  <p>B</p> </div> </div>										
compd	type	X	Y	R	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	pIC <sub>50</sub> <sup>a</sup>
26	A	O		PhCH <sub>2</sub> CH <sub>2</sub>	H	OH	OH	H	H	5.15
27	A	O		PhCH <sub>2</sub> CH <sub>2</sub>	H	OH	OCH <sub>3</sub>	H	H	4.22
28	A	O		PhCH <sub>2</sub> CH <sub>2</sub>	H	OCH <sub>3</sub>	OH	H	H	4.00
29	A	O		PhCH <sub>2</sub> CH <sub>2</sub>	H	OCH <sub>3</sub>	OCH <sub>3</sub>	H	H	4.00
30	A	O		PhCH <sub>2</sub> CH <sub>2</sub>	H	F	F	H	H	4.00
31	A	O		PhCH <sub>2</sub> CH <sub>2</sub>	OH	H	H	OH	H	4.00
32	A	O		PhCH <sub>2</sub> CH <sub>2</sub>	OH	OH	OH	H	H	5.70
33	A	O		PhCH <sub>2</sub> CH <sub>2</sub>	H	OH	OH	OH	H	5.70
34	A	O		PhCH <sub>2</sub> CH <sub>2</sub>	H	OH	OH	H	OH	4.26
35	A	O		CH <sub>3</sub>	H	OH	OH	H	H	4.00
36	A	O		CH <sub>3</sub> CH <sub>2</sub>	H	OH	OH	H	H	4.00
37	A	O		PhCH <sub>2</sub>	H	OH	OH	H	H	5.05
38	A	O		PhCH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub>	H	OH	OH	H	H	5.10
39	A	O		2-naphthylethyl	H	OH	OH	H	H	4.36
40	A	O		1-naphthylethyl	H	OH	OH	H	H	4.82
41	B	O	C	PhCH <sub>2</sub> CH <sub>2</sub>		OH	OH	H		4.17
42	B	O	C	PhCH <sub>2</sub> CH <sub>2</sub>		H	OH	OH		4.74
43	B	O	C	OCH <sub>3</sub>		H	OH	OH		5.10
44	B	O	N	OCH <sub>3</sub>		OH	OH	H		4.00
45	B	O	N	OCH <sub>3</sub>		H	OH	OH		4.85
46	B	NH	N	PhCH <sub>2</sub>		OH	OH	H		4.00
47	B	NH	N	PhCH <sub>2</sub>		H	OH	OH		4.12
48	B	NH	N	PhCH <sub>2</sub> CH <sub>2</sub>		H	OH	OH		4.40

<sup>a</sup> pIC<sub>50</sub> (−log IC<sub>50</sub>) values were derived from biological activity data reported by Burke et al.<sup>24</sup>

tions thereof. The latent variables so generated are used for multivariate regression, maximizing the communality of predictor and response variable blocks. Several attractive features of PLS need mention, such as (i) the ability to handle multivariate regression analysis in cases where the number of independent variables is greater than the number of samples (compounds) as found in CoMFA and CoMSIA 3D QSAR analyses; (ii) the ability to perform well even when interdescriptor correlations exist that would pose a problem for traditional multiple linear regression;<sup>34</sup> (iii) the reduction of the risk of chance correlations.<sup>35</sup>

Initial LOO cross-validated PLS analyses were used to determine the optimum number of components to be used in the final QSAR models. PLS results are summarized in Table 3. Figures 2 and 3 show the prediction curves obtained with final CoMFA and CoMSIA 3D QSAR models. Table 4 gives the residual values from the prediction of the activity values of the test set.

More rigorous statistical testing was performed by group cross-validation and scrambling of the biological data (randomization control). The average  $q^2$  values and standard deviations obtained in the group cross-validation and biological data randomization exercises are presented in Table 5. These results indicate stability and robustness in the CoMFA and CoMSIA QSAR models.

**Docking of Cinnamoyl Inhibitors into the Active Site of HIV-1 IN.** We used the FlexX program<sup>32,33</sup> to dock the database of molecules in our training set into the active site of a monomeric unit of the catalytic core of HIV-1 IN. The monomeric unit A was used although the crystal structure contains a symmetry-related neighboring ligand causing a juxtaposition of two inhibitors each bound to the active site of a different monomer but

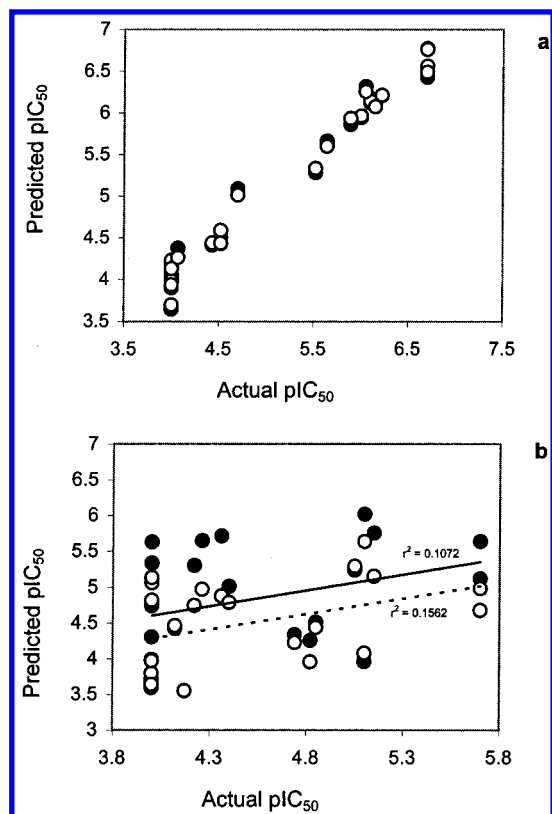
**Table 3.** PLS Statistics of CoMFA and CoMSIA 3D QSAR Models

step size	PLS statistics	24 compds model		22 compds model	
		CoMFA	CoMSIA	CoMFA	CoMSIA
2 Å	$q^2$	0.423	0.685	0.696	0.804
	PRESS	0.937	0.673	0.689	0.535
	$r^2$	0.973	0.969	0.976	0.975
	$s$	0.204	0.210	0.195	0.192
	$F$	101	114	100	125
	PLS components	6	5	6	5
	field contribution				
	steric	0.469	0.066	0.459	0.040
	electrostatic	0.531	0.251	0.541	0.173
	hydrophobic		0.144		0.086
1 Å	donor		0.343		0.450
	acceptor		0.196		0.250
	$q^2$	0.544	0.660	0.721	0.795
	PRESS	0.832	0.699	0.660	0.517
	$r^2$	0.981	0.951	0.980	0.942
	$s$	0.171	0.264	0.176	0.275
	$F$	145	71	124	98
	PLS components	6	5	6	3
	field contribution				
	steric	0.482	0.096	0.482	0.041
	electrostatic	0.518	0.192	0.518	0.138
	hydrophobic		0.141		0.114
	donor		0.329		0.425
	acceptor		0.242		0.282

interacting with each other.<sup>23</sup> This crystal structure representation has been characterized as being significantly influenced by crystal packing effects and therefore not unequivocal.<sup>36</sup>

Graphical representations of the docking results are presented in Figure 4. To assess the capability of the FlexX program, we tested how it would perform in docking the bound ligand, compound 1 (5CITEP) in the active site of the crystal structure. FlexX docked this ligand in the same pocket as the crystal structure and

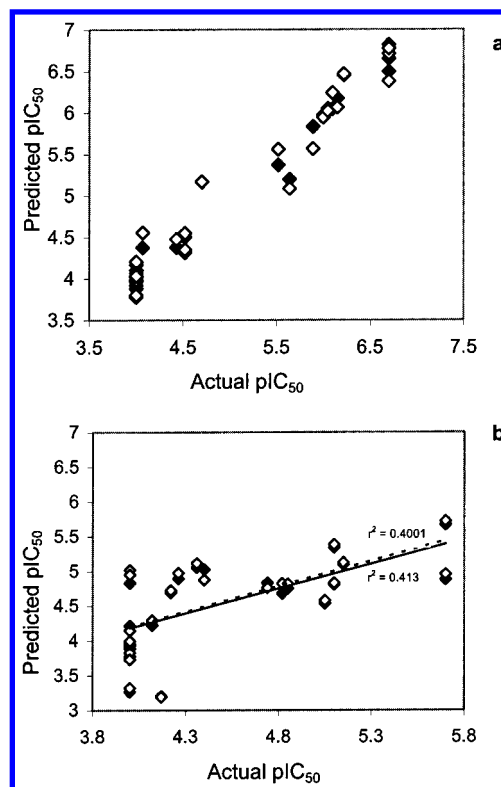




**Figure 2.** CoMFA predicted vs experimental  $pIC_{50}$  values. (a) Fitted predictions for the training set; filled circles represent predictions by the 2 Å model, while open circles represent predictions by the 1 Å model. (b) Predictions for the test set; filled circles and the solid line represent predictions by the 2 Å model, while open circles and the broken line represent predictions by the 1 Å model.

aligned the docked structures in the same horizontal plane within the active site similar to the bound ligand in the crystal structure. The second-ranked FlexX-docked structure among 30 structures, according to FlexX binding energy scoring, was the closest to the bound ligand with a root-mean-square deviation of 3.43 Å. As illustrated in Figure 4a, the chloroindole ring of this FlexX-docked ligand was in the same general vicinity as the chloroindole ring of the bound ligand. However, it was turned upside down, whereby the chlorine became buried in the cleft between Gln148 and Asp116 while the keto-enol moiety interacted with Asn155, Cys65, and Thr66. The tetrazole ring also is in a similar orientation as that of the bound ligand but was placed in the cleft between Lys159 and His67, with which it interacted extensively. This orientation appears somewhat similar to the alternative (energetically degenerate) orientation of the bound ligand recently proposed by Sotrifer et al.<sup>36,37</sup> Figure 4a shows the superimposition of the FlexX-docked structure (ball-and-stick rendering) over the crystal structure location (capped-sticks rendering) of the bound ligand in the IN active site.

The docking of the cinnamoyl compounds produced structures that were divided into two clusters occupying two separate but overlapping binding regions. The first cluster, which is aligned horizontally across the binding site as shown in Figure 4b, occupies what we term the east-west binding region, which encompasses the same region that FlexX docked the bound ligand. The second



**Figure 3.** CoMSIA predicted vs experimental  $pIC_{50}$  values. (a) Fitted predictions for the training set; filled diamonds represent predictions by the 2 Å model, while open diamonds represent predictions by the 1 Å model. (b) Predictions for the test set; filled diamonds and the solid line represent predictions by the 2 Å model, while open diamonds and the broken line represent predictions by the 1 Å model.

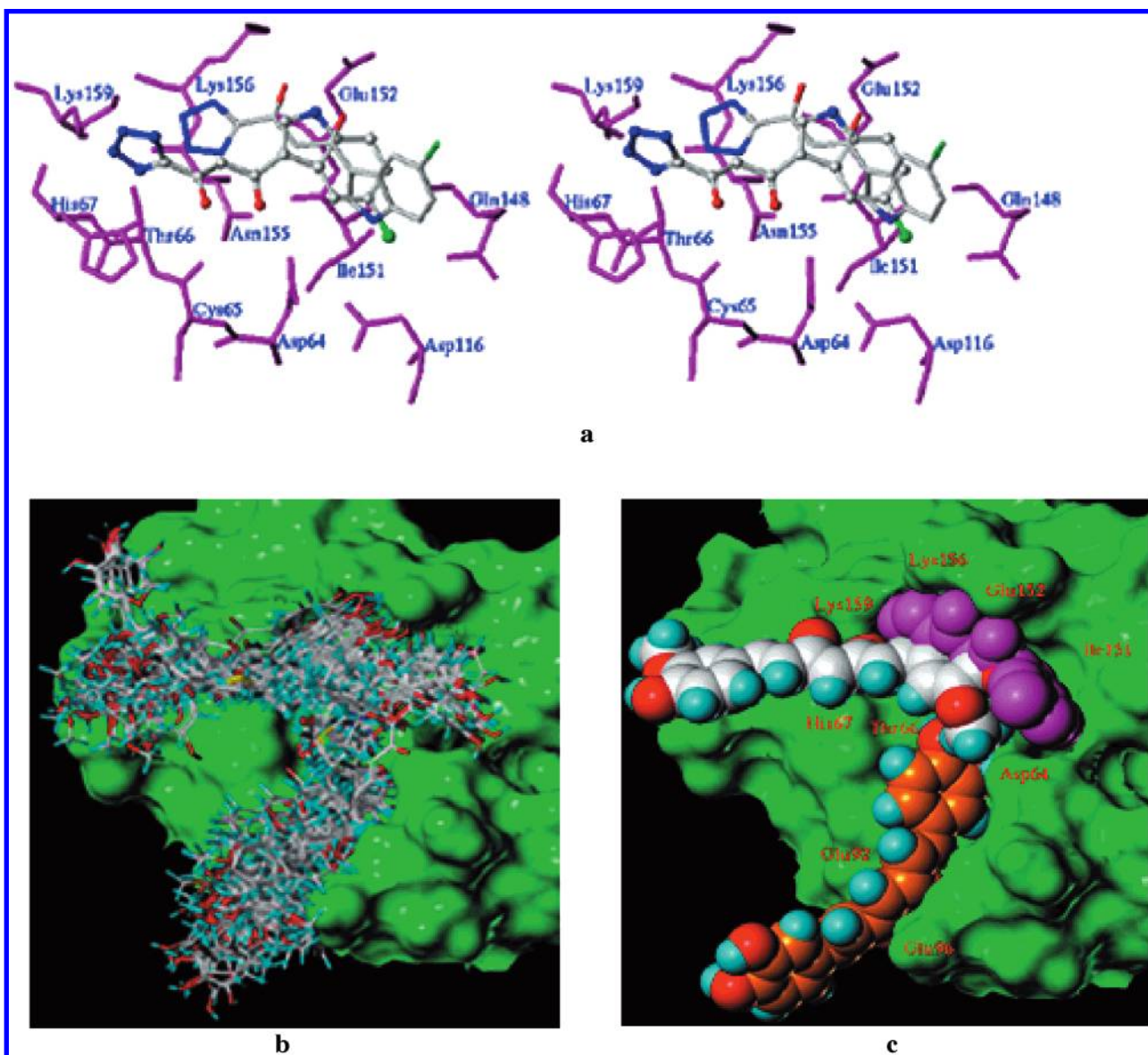
**Table 4.** Residuals of the Predictions of the Test Set by the CoMFA and CoMSIA Models

compd	$pIC_{50}$	residuals			
		CoMFA 2 Å	CoMFA 1 Å	CoMSIA 2 Å	CoMSIA 1 Å
26	5.15	0.606	0.002	-0.056	-0.022
27	4.22	1.081	0.526	0.470	0.505
28	4.00	0.744	-0.024	-0.069	-0.004
29	4.00	0.308	-0.402	-0.112	-0.044
30	4.00	1.627	1.136	-0.113	-0.269
31	4.00	-0.005	-0.354	0.213	0.147
32	5.70	-0.580	-1.021	-0.029	0.022
33	5.70	-0.065	-0.716	-0.822	-0.741
34	4.26	1.386	0.713	0.636	0.717
35	4.00	1.126	0.822	0.833	0.953
36	4.00	1.336	1.062	1.019	1.020
37	5.05	0.188	0.242	-0.514	-0.474
38	5.10	0.917	0.536	0.243	0.286
39	4.36	1.353	0.518	0.696	0.755
40	4.82	-0.559	-0.858	-0.143	0.005
41	4.17	-0.619	-0.615	-0.965	-0.977
42	4.74	-0.398	-0.510	0.090	0.020
43	5.10	-1.136	-1.019	-0.291	-0.268
44	4.00	-0.281	-0.332	-0.222	-0.172
45	4.85	-0.332	-0.408	-0.096	-0.035
46	4.00	-0.298	-0.199	-0.735	-0.679
47	4.12	0.307	0.349	0.104	0.171
48	4.40	0.614	0.392	0.629	0.476

cluster is aligned nearly vertically within the active site and is termed the north-south binding region. Figure 4c shows space-filling models of two different cinnamoyl molecules, one colored by the atom type (C, white; H, cyan; O, red, in the east-west binding region) and the other colored by the atom type (C, orange; H, cyan; O,

**Table 5.** Results of Group Cross-Validation and Randomization Exercises

exercise	average PLS statistics	24 compds		22 compds	
		CoMFA	CoMSIA	CoMFA	CoMSIA
group validation	$q^2$ <sup>a</sup>	0.419 (0.082)	0.685 (0.029)	0.653 (0.052)	0.795 (0.019)
randomization	$q^2$	-0.278	-0.245	-0.295	-0.245

<sup>a</sup> Value in parentheses is the standard deviation.

**Figure 4.** (a) Stereoview of FlexX docking of the ligand in complex (ball-and-stick rendering) to the active site of the monomer structure of HIV-1 IN; comparison with the experimental position of the bound ligand (capped-sticks rendering). The ligands are colored by atom type (C, gray; N, blue; O, red; Cl, green), and important active site amino acids are shown in magenta. (b) FlexX docking of cinnomoyl IN inhibitors into the IN active site, showing the two clusters obtained: (i) east-west (horizontal) and (ii) north-south (vertical), on the Connolly surface of the active site. (c) Space-filling rendering showing ligands (colored by atom type) docked in the two binding locations as depicted in panel b; the experimentally bound ligand is colored magenta, the carbon atoms of the east-west-oriented docked ligand are colored gray, whereas those of the north-south-bound ligand are colored orange.

red, in the north-south binding region) as well as the cocrystallized inhibitor (compound **1**, colored magenta) in the active site. Most compounds in the training set are much larger than the bound inhibitor; therefore, they needed to extend much beyond the bound inhibitor pocket. The main active site amino acid residues that interacted with most of the docked cinnomoyl compounds in our study were Asp64, Cys65, Thr66, His67, Glu92, Glu96, Glu152, Lys156, and Lys159 as labeled

in Figure 4c. Structural information regarding inhibitor binding to the HIV-1 catalytic site is very limited. Recent docking studies have suggested binding modes for several different IN inhibitors at the active site, and invariably, different inhibitors adopt different binding orientations with some overlapping interactions.<sup>37</sup> The HIV-1 IN active site is quite large and relatively shallow and therefore able to accommodate large variations in ligand size and shape and different binding modes.

The top-ranking members of the docked structures were mostly found in the east–west binding region. This also was the binding region that complemented the CoMFA and CoMSIA contour maps from the 3D QSAR models as discussed in the respective sections below. Consequently, we have proposed this orientation to be probably the best binding mode for the cinnamoyl inhibitors at the IN active site. Interestingly, the docking energy scores calculated by FlexX correlated quite well with the biological activity ( $\text{pIC}_{50}$ ) in a general sense; compounds with the highest activity showed the lowest FlexX energy scores (see Table 1). This correlation was more pronounced for molecules that differed only by the presence of phenolic hydroxyl groups on one hand and their absence or methylation on the other hand. Thus, compound **5**, which is more potent than the corresponding dimethylated derivative compound **6**, is ranked higher than the latter in predicted binding affinity as shown by the lower FlexX energy scores (Table 1). Likewise, compound **8** is ranked higher than compound **9**, which ranks similar to compound **10**, and both **9** and **10** rank higher than compound **11** in which all four phenolic OH groups have been methylated. Furthermore, compound **12** is more potent than and was predicted to bind more tightly than compound **13**, which has two less hydroxyl groups. The same trend holds true among compounds **14**–**17** as well as between compounds **19** and **21** and compounds **20** and **22**. This trend indicates the importance of hydrogen-bonding interactions in determining binding at the active site. The carbonyl groups in the indanedione and indanone derivatives, compounds **23** and **24**, respectively, are also indicated to contribute to the binding by the FlexX docking results. The only group of compounds that did not show the trend for the presence of OH groups is compounds **19** and **20**, where one would have expected **19** to be more potent than **20** but the experimental results are the opposite. Unlike most of the compounds that show the OH trend, compounds **20** and **21** like compounds **19**–**25** are not symmetrical. It appears that the presence of the OH groups on the phenyl ring of the cinnamoyl portion of these molecules is more important for inhibitory activity than their presence on the phenyl ring of the benzoyl portion in compounds **19**–**25**. It also appears that the presence of phenolic hydroxyl groups adjacent to each (ortho relationship) is more important than just the mere presence of these groups.

Some exceptions to the general trends have been addressed in the following discussion. The reason for the high binding energy score of compound **18** is not apparent since it lacks the phenolic hydroxyl groups and has a low potency. This may be an artifactual occurrence in which compensation of energy gain by hydrophobic interaction may offset the absence of the interaction with phenolic hydroxyl groups. The other discrepancy is the less than expected binding energy score for the cyclopentyl derivative, compound **7**, as compared to its cyclohexyl counterpart, compound **6**. This may result from how their different ring systems fit into the cleft between Lys159 and His67 (see Figure 4c).

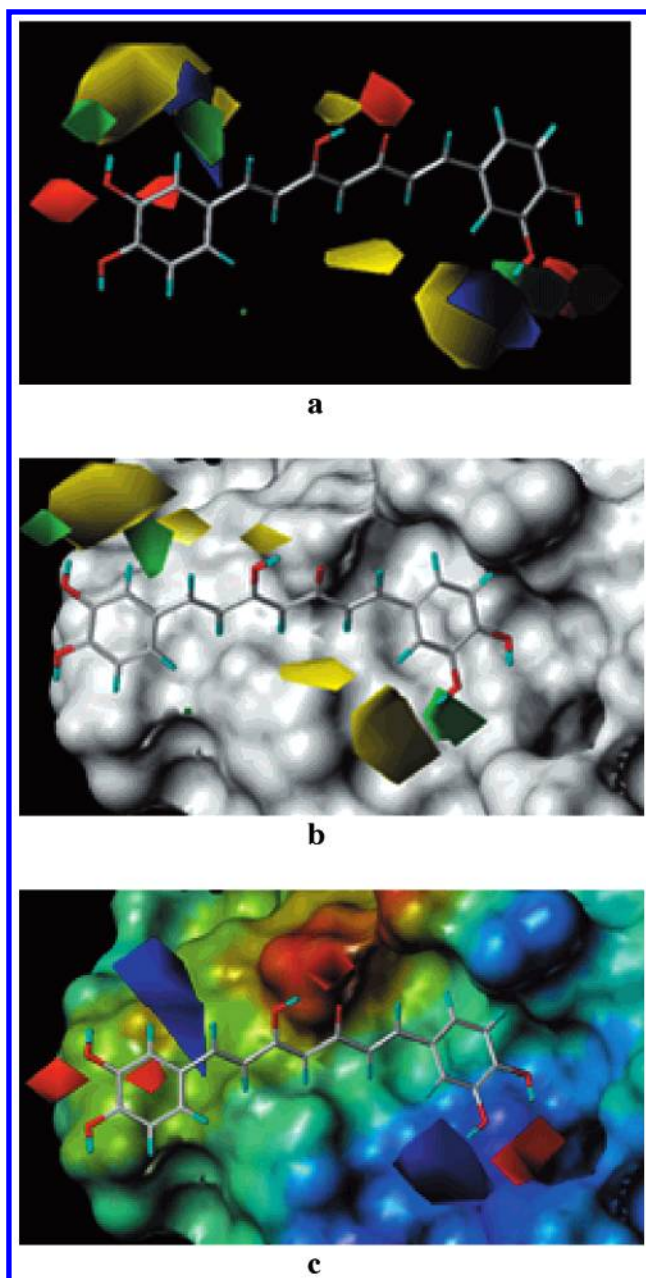
**CoMFA 3D QSAR Models.** PLS analysis on the multifit alignment of all of the compounds in the training set using default parameters resulted in a

CoMFA QSAR model with a modest  $q^2$  value of 0.423. Examination of the residuals from the LOO cross-validated predictions indicated that some compounds might be outliers. Omission of two compounds, **24** and **25**, resulted in an increase in the  $q^2$  value to 0.696 for the remaining 22 compounds. Several factors may contribute to the outlier status of compounds **24** and **25**, including low biological activity and structural uniqueness. Compound **24** appears to have a lower activity than expected when compared to its counterpart, compound **23**. The outlier status of compound **25** could stem from its structural uniqueness. The reduction of an  $\alpha,\beta$ -unsaturated double bond in compound **25** introduces much conformational flexibility that is absent in its unsaturated counterparts, compounds **8**–**11**. One would expect a higher potency for compound **25** than it exhibited, based upon the presence of the catechol moiety, which appears to confer high potency on all of the other compounds in which it occurs.

Reducing the lattice step size from 2 to 1 Å resulted in a significant change in the  $q^2$  value from 0.423 to 0.544 for the model with all 24 compounds and from 0.696 to 0.721 for the model with 22 compounds. The predictive performance of the CoMFA models is shown in Figure 2, which indicates that the 1 Å model performed better at predicting the activities of the test set than the 2 Å model (Figure 2b). The CoMFA model obtained using all compounds in the training set performed better in predicting in the biological activities of the compounds in the test set than the CoMFA model derived from a training set of 22 compounds (data not shown), even though the latter had a higher  $q^2$  value. This shows that a higher  $q^2$  value does not necessarily make a model more predictive with reference to an external test set. In this case, it appears that structural diversity dictated the predictive performance of the models with regard to the test set. The outliers had structural features that resembled those in many of the compounds in the test set; therefore, their removal caused the QSAR models to be less accommodating to all of the compounds in the test set.

Lower energy cutoff values, namely, 25, 20, and 15 kcal/mol, were investigated, but all led to a decrease in the  $q^2$  value (data not shown). In a previous CoMFA 3D QSAR study of flavone HIV-1 IN inhibitors, a 10 kcal/mol cutoff was found to be optimal.<sup>37</sup> The PLS stdev\* coefficient contour maps for the CoMFA model are shown in Figure 5a. Green regions indicate areas where steric bulk is predicted to enhance biological activity, whereas yellow contours indicate regions where steric bulk is predicted to be detrimental to biological activity. Blue-colored regions indicate areas where electropositive groups are predicted to enhance biological activity, while red regions represent areas where electronegative groups are predicted to favor activity. There are green contours close to the catechol moieties especially the one on the lower right corner. These are immediately flanked by yellow contours suggesting that limited bulk right close to the molecules will be favorable but cannot be extended beyond that. There are two yellow regions flanking the carbonyl or keto–enol moiety of the compounds suggesting steric restriction in this region, possibly a narrow cleft at the binding site. This is supported by the FlexX docking of the molecules into





**Figure 5.** (a) CoMFA stdev\*coeff contour plots; green contours indicate regions where bulky groups increase activity, whereas yellow contours indicate regions where bulky groups decrease activity. Blue contours indicate regions where positive groups increase activity, whereas red contours indicate regions where negative charge increases activity. (b) CoMFA steric contours projected over the solvent accessible (connolly) topological surface of the IN active site. (c) CoMFA electrostatic contours projected over the electrostatic potential surface (blue, negative potential; red/brown, positive potential) of the IN active site.

the IN active site, as shown in Figure 4c. Projecting the steric CoMFA contour maps onto the docked-inhibitor orientation on the MOLCAD-generated solvent accessible (connolly) surface of the IN active site, as depicted in Figure 5b, shows that these yellow contours demarcate the narrow cleft between Lys159 and His67 residues (see amino acid labeling in Figure 4c), which accommodates the carbonyl or keto–enol moiety.

There are prominent red contours close to the oxygen atoms of the catechol moieties suggesting the preference for electronegative groups at those locations, which is consistent with the observation that polyhydroxylation

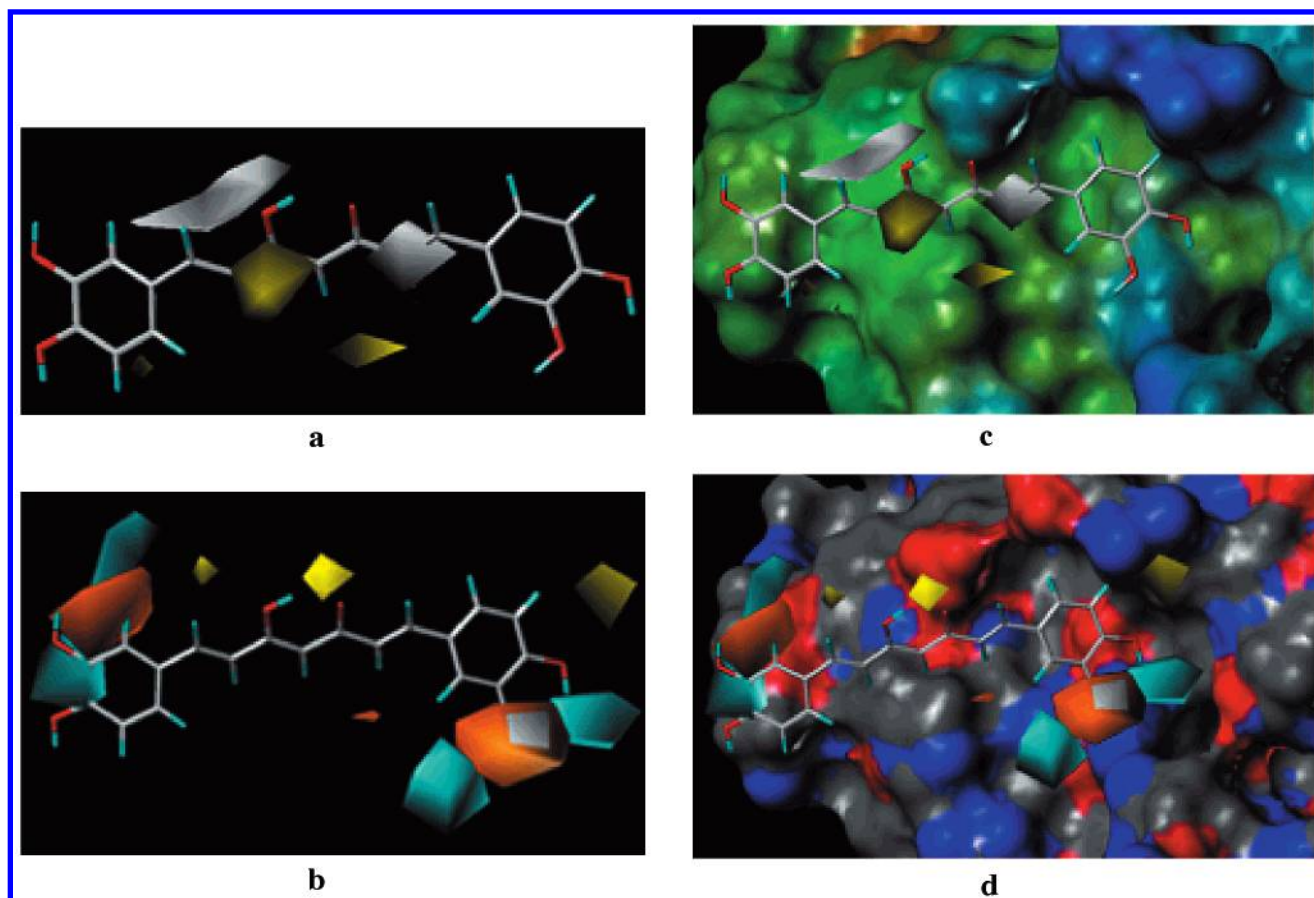
generally enhances HIV IN inhibitory activity.<sup>16</sup> Blue contours reside around the upper left-hand and lower right-hand corners of the molecule, the latter just bracing the hydrogen atoms of the catechol moiety, indicating that electropositive groups are predicted to increase activity in those areas. This is also a possible indication that H-bond donor groups will be favored in those locations.

The projection of the CoMFA electrostatic contour map onto the electrostatic potential surface map of the binding site generated by the MOLCAD shows a general complementarity as depicted in Figure 5c. Red regions on the binding site surface map represent positive electrostatic potential, whereas blue regions on the surface represent negative electrostatic potential. Therefore, the red CoMFA contour around the carbonyl group (electronegative) matches well with the red, highly electropositive surface of the binding site provided by amino acid residue Lys159. The large blue CoMFA contour around the hydroxyl hydrogen (electropositive) at the right lower corner complements the dark blue (highly electronegative) surface of the active site provided by Cys65 and Asp64. The other red and blue CoMFA contours appear to be in intermediate electrostatic potential regions or pointing out into solvent. Of the two possible binding domains that were identified by FlexX docking (Figure 4b,c), the east–west domain complements the CoMFA PLS contour maps better than the north–south domain, supporting our proposed binding mode. It should be kept in mind, however, that the CoMFA and CoMSIA coefficient maps show relative differences in that “more positive” or “more negative” may also mean that all compared values are still in the negative range or positive range, respectively. In that case, mapping the actual electrostatic potential onto the surface could show a mismatch.

**CoMSIA 3D QSAR Models.** CoMSIA is a relatively new alternative molecular field analysis method to CoMFA. It is touted to be less affected by changes in molecular alignment and to provide more smooth and interpretable contour maps as a result of employing Gaussian type distance dependence with the molecular similarity indices it uses.<sup>20</sup> Furthermore, in addition to the steric and electrostatic fields, CoMSIA defines explicit hydrophobic and hydrogen bond donor and acceptor descriptor fields, which are not available with standard CoMFA. The use of CoMSIA along with CoMFA shows that in most instances it performs similarly to CoMFA in terms of predictive ability, sometimes slightly better, and other times slightly worse than CoMFA. Overall, it is a good addition to the 3D QSAR tool kit and is gaining popularity.

In the present study, we obtained better PLS statistics and predictive performance on the test set with the CoMSIA 3D QSAR models than with the CoMFA models (Tables 3 and 4). Using default SYBYL CoMSIA parameters and all compounds in the training set with the multifit molecular alignment resulted in a PLS model with a  $q^2$  value of 0.685 (as compared to a  $q^2$  value of 0.423 for CoMFA) and an  $r^2$  value of 0.969 for 5 PLS components. Elimination of the same two outliers, compounds **23** and **24**, as was done in the case of CoMFA, brought the  $q^2$  value up to 0.804 (as compared to a  $q^2$  value of 0.696 for CoMFA) and the  $r^2$  value to





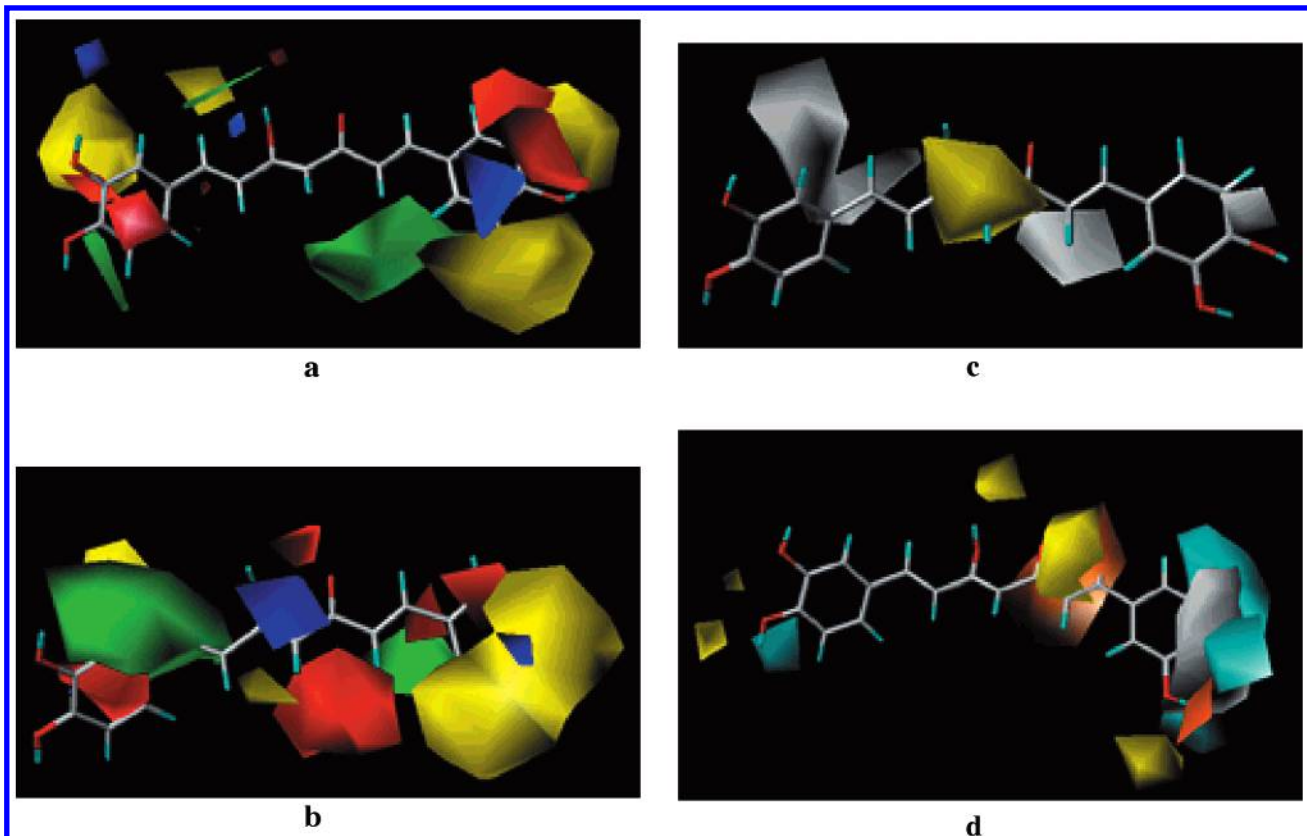
**Figure 6.** (a) CoMSIA stdev\*coeff hydrophobic contour plots; white contours indicate regions where hydrophobic (lipophilic) groups increase activity, whereas yellow contours indicate regions where hydrophobic (lipophilic) groups decrease activity. (b) CoMSIA stdev\*coeff hydrogen bond donor and hydrogen bond acceptor contour plots; cyan contours indicate regions where hydrogen bond donor groups increase activity, whereas yellow contours indicate regions where a hydrogen bond donor decreases activity. White contours indicate regions where hydrogen bond acceptor groups increase activity, whereas orange contours indicate regions where a hydrogen bond acceptor decreases activity. (c) CoMSIA hydrophobic contours projected over the lipophilicity potential surface map (brown, hydrophobic; blue, hydrophilic) of the IN active site. (d) CoMSIA H-bond donor and acceptor contour maps projected on the H-bond potential surface map of the IN active site (red, H-bond donor; blue, H-bond acceptor).

0.975. Unlike CoMFA, reducing the grid spacing from 2 to 1 Å resulted in CoMSIA PLS models with slightly lower  $q^2$  and  $r^2$  values (Table 3). However, it provided a slightly better performance in predicting the activities of the test set (see Figure 3b). CoMSIA 3D QSAR models performed well in the prediction of the IN inhibitory activities of the test compounds and much better than the CoMFA 3D QSAR models. This can be seen from the prediction curves for the test compounds (see Figures 2b and 3b) and the residual values in Table 4. The superior performance of CoMSIA relative to CoMFA, with this data set, may be attributed mainly to the higher contributions from the hydrophobic and hydrogen bond donor and acceptor fields to the CoMSIA QSAR models (Table 3). CoMFA, unlike CoMSIA, does not have explicit hydrophobic and hydrogen-bonding descriptors, which are assumed to be implicitly treated in the CoMFA steric and electrostatic fields, respectively. Overall, the electrostatic field components made higher contributions to the models (Table 3).

The CoMSIA steric and electrostatic PLS contours (not shown) were similarly placed as those of the CoMFA model. The additional hydrophobic, hydrogen bond donor, and hydrogen bond acceptor contours of CoMSIA are displayed in Figure 6a,b, respectively. The hydrophobic fields (white, hydrophobic group favored;

yellow, hydrophobic disfavored) and the H-bond donor (cyan, favored; yellow, disfavored) and H-bond acceptor (white, favored; orange, disfavored) fields indicate areas around the molecules where changes increased or decreased activity. The H-bond donor fields made the highest contribution to the CoMSIA QSAR models (Table 3), which suggests that among the descriptors considered, the hydrogen bond donor is the most important factor influencing the IN inhibitory activity of the compounds in the training set. This was strongly supported by the FlexX docking results as discussed above, which indicated that the presence of free hydroxyl groups was important for binding to the active site.

Furthermore, as depicted in Figure 6c,d, the hydrophobic, hydrogen bond donor, and hydrogen bond acceptor contour maps of the CoMSIA 3D QSAR models complement the lipophilicity (Figure 6c) and hydrogen bond (Figure 6d) potential surface maps of the binding site. With regard to the lipophilicity potential surface, brown is hydrophobic whereas blue is hydrophilic. As Figure 6c indicates, the regions of the CoMSIA hydrophobic map where hydrophobicity is favored (white) occur over the receptor surface of moderate hydrophobic potential. The regions of the CoMSIA hydrophobic map where hydrophobicity is disfavored are in proximity to



**Figure 7.** (a) CoMFA stdev\*coeff contour plots; green contours indicate regions where bulky groups increase activity, whereas yellow contours indicate regions where bulky groups decrease activity. Blue contours indicate regions where positive groups increase activity, whereas red contours indicate regions where negative charge increases activity. (b) CoMSIA stdev\*coeff steric and electrostatic contour plots; the region colors have the same interpretation as those in panel a. (c) CoMSIA stdev\*coeff hydrophobic contour plots; white contours indicate regions where hydrophobic (lipophilic) groups increase activity, whereas yellow contours indicate regions where hydrophobic (lipophilic) groups decrease activity. (d) CoMSIA stdev\*coeff hydrogen bond donor and hydrogen bond acceptor contour plots; cyan contours indicate regions where hydrogen bond donor groups increase activity, whereas yellow contours indicate regions where a hydrogen bond donor decreases activity. White contours indicate regions where hydrogen bond acceptor groups increase activity, whereas orange contours indicate regions where a hydrogen bond acceptor decreases activity.

the polar keto–enol moiety, which interacts with hydrophilic amino acid residues. The complementarity of the CoMSIA hydrogen bond donor and acceptor contour maps with the hydrogen bond potential map of the binding site surface is even more dramatic. As shown in Figure 6d, the regions of the CoMSIA hydrogen bond donor map where donor is favored (cyan) occur over the receptor surface where there is high hydrogen bond acceptor potential (blue), whereas a prominent H-bond donor disfavored contour (yellow) occurs right over the site surface with high H-bond donor potential (red), in proximity to the Lys159. This is consistent because a strong H-bond donor group at the binding site will interact with a H-bond acceptor group on the ligand and not a H-bond donor group. The CoMSIA hydrogen bond acceptor contours also show general complementarity with this binding site surface map as well, albeit not as strongly as the CoMSIA H-bond donor fields. This may be a reflection of the lower contribution of the H-bond acceptor fields relative to the H-bond donor fields in the CoMSIA QSAR models (Table 3).

The high complementarity between the CoMFA and the CoMSIA contour maps and the surface property maps (electrostatic potential, lipophilicity potential, and hydrogen-bonding potential) of the east–west binding site leads us to propose that site as the likely binding site and mode of these cinnamoyl inhibitors at the HIV-1 IN active site.

**3D QSAR Studies with Docked Alignment.** In addition to the 3D QSAR studies using the multifit alignment described above, we also conducted CoMFA and CoMSIA studies with molecules aligned by docking at the active site in the horizontal binding region. The docked alignment was quite different as the molecules were more staggered, and compounds **12**, **13**, **22**, and **25** were outliers from the start as they were docked differently from the rest of the compounds. A CoMFA 3D QSAR model with a positive, albeit low,  $q^2$  value (0.259 with 4 PLS components, see Table 6) was obtained with 17 of the compounds in the docked alignment, leaving out compounds **2**, **23**, and **24**. It is interesting to note that compounds **24** and **25** were indicated outliers in the CoMFA model using the multifit alignment (vide supra). Compound **8** was indicated as an outlier in this CoMFA model as well, and its elimination resulted in a substantial increase in the  $q^2$  value of the model to 0.477 with 4 PLS components (Table 6). As shown also in Table 6, the final QSAR model with the 16 compounds had an  $r^2$  value of 0.988. The steric and electrostatic field contributions to this model were 0.365 and 0.635, respectively. In the multifit alignment CoMFA, the corresponding field contributions were 0.469 and 0.531, respectively (Table 3), indicating that the electrostatic fields dominated in both CoMFA models. The CoMFA PLS steric and electrostatic field coefficient map obtained for the docked



**Table 6.** PLS Statistics of CoMFA and CoMSIA 3D QSAR Models Obtained from Dock Alignment

step size	PLS statistics	CoMFA <sup>a</sup>	CoMFA <sup>b</sup>	CoMSIA <sup>c</sup>
2 Å	$q^2$	0.259	0.477	0.425
	PRESS	1.087	0.930	0.951
	$r^2$	0.982	0.998	0.922
	$s$	0.171	0.064	0.112
	$F$	160	1103	349
	PLS components	4	4	5
	field contribution			
	steric	0.365	0.365	0.08
	electrostatic	0.635	0.635	0.252
	hydrophobic			0.135
	H-bond donor			0.353
	H-bond acceptor			0.117

<sup>a</sup> Model was obtained using 17 out of the 20 compounds docked within a similar binding region. <sup>b</sup> Model was obtained using 16 out of the 20 compounds docked within a similar binding region. <sup>c</sup> Model was obtained using all 20 compounds docked within a similar binding region.

alignment is shown in Figure 7a, which shows similarity to the map obtained for the multifit alignment CoMFA (Figure 5a). This indicates a general agreement between the two CoMFA models.

Similarly, CoMSIA 3D QSAR study was conducted on the molecules in the docked alignment. CoMSIA PLS analysis afforded a  $q^2$  value of 0.429 with 5 PLS components for the 20 compounds in the alignment (see Table 6). CoMSIA provided a better QSAR model than CoMFA with this alignment as indicated by the  $q^2$  values and the number of compounds that could be accommodated. The superiority of CoMSIA over CoMFA found with the multifit alignment QSAR models was magnified with the docked alignment, indicating that the CoMSIA model was less affected by alignment heterogeneity. The statistics of the nonvalidated CoMSIA model for the docked alignment are provided in Table 6. The relative order of importance of the various CoMSIA descriptor fields, i.e., steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor is the same as the order obtained from the multifit alignment CoMSIA (see Table 3). Here again, the hydrogen bond donor field is shown to be the most important. The CoMSIA PLS field coefficient contour maps obtained with the docked alignment are shown in Figure 7b–d. The steric and electrostatic contour map is similar to the CoMFA map (Figure 7a) while the hydrophobic and hydrogen bond field contour maps are very similar to the corresponding maps of the multifit alignment CoMSIA (Figure 6a,b), further showing that there is an agreement between the two CoMSIA models.

## Conclusion

We have established predictive CoMFA and CoMSIA 3D QSAR models for conformationally constrained cinnamoyl HIV IN inhibitors, with CoMSIA 3D QSAR modeling performing better than CoMFA. Molecular docking and molecular surface property (electrostatic, lipophilicity, and hydrogen-bonding potential) mapping have been integrated with CoMFA and CoMSIA 3D QSAR, to identify a potential binding site and mode for these cinnamoyl inhibitors at the HIV-1 IN active site. This study should provide further insights to support structure-based design of anti-HIV compounds as potential AIDS drugs. If confirmed by the solution of the

structures of cinnamoyl inhibitors in complex with the HIV-1 IN catalytic core, this study would be a demonstration of the power of combining 3D QSAR and docking to explore the binding sites and modes for ligands where these are unknown.

**Acknowledgment.** Support from Start-up funds and resources provided by the College of Pharmacy and Department of Pharmaceutical Sciences is gratefully acknowledged.

## References

- (1) Johnston, M. I.; Hoth, D. F. Present Status and Future Prospects for HIV Therapies. *Science* **1993**, *260*, 1286–1293.
- (2) Craigie, R. HIV Integrase, a Brief Overview from Chemistry to Therapeutics. *J. Biol. Chem.* **2001**, *276*, 23213–23216.
- (3) De Clercq, E. From Anti-HIV Agents to Anti-AIDS Chemotherapy: A Critical Appraisal. In *Anti-AIDS Drug Development: Challenges, Strategies and Prospects*; Mohan, P., Baba, M., Eds.; Harwood Academic Publishers: Chur, Switzerland, 1995; pp 1–37.
- (4) De Clercq, E. Towards Improved Anti-HIV Chemotherapy: Therapeutic Strategies for Intervention with HIV Infections. *J. Med. Chem.* **1995**, *38*, 2491–2517.
- (5) Appelt, K.; Clare, M.; Pettit, S.; Michael, S. F.; Swantsrom, R.; Moore, M. C.; Dreyer, G. B.; Erickson, J. W. Therapeutic Approaches to HIV. *Perspect. Drug Discovery Des.* **1993**, *1*, 23–128.
- (6) La Femina, R. L.; Schnider, C. L.; Robbins, H. L.; Callahan, P. L.; LeGrow, K.; Roth, E.; Schleif, W. A.; Emini, E. A. Requirement of Active Human Immunodeficiency Virus Type 1 Integrase Enzyme for Productive Infection of Human T-lymphoid Cells. *J. Virol.* **1992**, *66*, 7414–7419.
- (7) Sakai, H.; Kawamura, M.; Sakuragi, J.; Sakuragi, S.; Shibata, R.; Ishimoto, A.; Ono, N.; Ueda, S.; Adachi, A. Integration is Essential for Efficient Gene Expression of Human Immunodeficiency Virus Type 1. *J. Virol.* **1993**, *67*, 1169–1174.
- (8) Hazuda, D. J.; Felock, P.; Witmer, M.; Wolfe, A.; Stillmock, K.; Grobler, J. A.; Espeseth, A.; Gabryelski, L.; Schleif, W.; Blau, C.; Miller, M. D. Inhibitors of Strand Transfer that Prevent Integration and Inhibit HIV-1 Replication in Cells. *Science* **2000**, *287*, 646–650.
- (9) Mazumder, A.; Neamati, N.; Ojwang, J. O.; Sunder, S.; Rando, R. F.; Pommier, Y. Inhibition of Human Immunodeficiency Virus Type 1 Integrase by Guanidine Quartet Structures. *Biochemistry* **1996**, *35*, 13762–13771.
- (10) Aronex Pharmaceuticals, Woodland, TX and New York Hospital.
- (11) Neamati, N.; Sander, S.; Pommier, Y. Design and Discovery of HIV-1 Integrase Inhibitors. *Drug Discovery Today* **1997**, *11*, 487–498.
- (12) Eich, E.; Pertz, H.; Kaloga, M.; Schulz, J.; Fesen, M. R.; Mazumder, A.; Pommier, Y. (–)-Arctigenin as a Lead Structure for Inhibitors of Human Immunodeficiency Virus Type-1 Integrase. *J. Med. Chem.* **1996**, *39*, 86–95.
- (13) Nicklaus, M. C.; Neamati, N.; Hong, H.; Mazumder, A.; Sunder, S.; Chen, J.; Milne, G. W. A.; Pommier, Y. HIV-1 Integrase Pharmacophore: Discovery of Inhibitors Through Three-Dimensional Database Searching. *J. Med. Chem.* **1997**, *40*, 920–929.
- (14) Hong, H.; Neamati, N.; Wang, S.; Nicklaus, M. C.; Mazumder, A.; Zhao, H.; Burke, T. R., Jr.; Pommier, Y.; Milne, G. W. Discovery of HIV-1 Integrase Inhibitors by Pharmacophore Searching. *J. Med. Chem.* **1997**, *40*, 930–936.
- (15) Robinson, W. E., Jr.; Cordeiro, M.; Abdel-Malek, S.; Jia, Q.; Chow, S. A.; Reinecke, M. G.; Mitchell, W. M. Dicafeoylquinic Acid Inhibitors of Human Immunodeficiency Virus Integrase: Inhibition of the Core Catalytic Domain of Human Immunodeficiency Virus Integrase. *Mol. Pharmacol.* **1996**, *50*, 846–855.
- (16) Fesen, M. R.; Kohn, K. W.; Leteurtre, F.; Pommier, Y. Inhibitors of Human Immunodeficiency Virus Integrase. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2399–2403.
- (17) 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.
- (18) *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM Science Publishers: Leiden, 1993.
- (19) Raghavan, K.; Buolamwini, J. K.; Fesen, M. R.; Pommier, Y.; Kohn, K. W.; Weinstein, J. N. Three-Dimensional Quantitative Structure–Activity Relationship (QSAR) of HIV Integrase Inhibitors: A Comparative Molecular Field Analysis (CoMFA) Study. *J. Med. Chem.* **1995**, *38*, 890–897.
- (20) Kliebe, G.; Abraham, U.; Mietzner, T. Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict their Biological activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.



- (21) Bohm, M.; Stutzerbercher, J.; Kliebe, G. Three-Dimensional Quantitative Structure–Activity Relationship Analyses Using Comparative Molecular Field Analysis and Comparative Molecular Similarity Indices Analysis to Elucidate Selectivity Differences of Inhibitors Binding to Trypsin, Thrombin, and Factor Xa. *J. Med. Chem.* **1999**, *42*, 458–477.
- (22) Artico, M.; Di Santo, R.; Costi, R.; Novellino, E.; Greco, G.; Massa, S.; Tramontano, E.; Marongiu, M. E.; De Montis, A.; Colla, P. L. Geometrically and Conformationally Restrained Cinnamoyl Compounds as Inhibitors of HIV-1 Integrase: Synthesis, Biological Evaluation, and Molecular Modeling. *J. Med. Chem.* **1998**, *41*, 3948–3960.
- (23) Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Mural, H.; Davies, D. R. Structure of the HIV-1 Integrase Catalytic Domain Complexed with an Inhibitor: A Platform for Antiviral Drug Design. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13040–13043.
- (24) Burke, T. R., Jr.; Fesen, M. R.; Mazumder, A.; Wang, J.; Carothers, A. M.; Grunberger, D.; Driscoll, J.; Kohn, K.; Pommier, Y. Hydroxylated Aromatic Inhibitors of HIV Integrase. *J. Med. Chem.* **1995**, *38*, 4171–4178.
- (25) SYBYL, version 6.7; Tripos Associates: St. Louis, MO, 2000.
- (26) Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. The Tripos Force Field. *J. Comput. Chem.* **1989**, *10*, 982–1012.
- (27) Stewart, J. J. MOPAC. *J. Comput.-Aided Mol. Des.* **1990**, *4* (1).
- (28) Viswanadhan, V. N.; Ghose, A. K.; Revenkar, G. R.; Robins, R. Atomic and Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure–Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally Occurring Antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.
- (29) Klebe, G. The Use of Composite Crystal-Field Environments in Molecular Recognition and the *De Novo* Design of Protein Ligands. *J. Mol. Biol.* **1994**, *237*, 212–235.
- (30) Buolamwini, J. K.; Raghavan, K.; Fesen, M. R.; Pommier, Y.; Kohn, K. W.; Weinstein, J. N. Application of the Electrotopological State Index to QSAR Analysis of Flavone Derivatives as HIV-1 Integrase Inhibitors. *Pharm. Res.* **1996**, *13*, 1891–1894.
- (31) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
- (32) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX Incremental Construction Algorithm for Protein–Ligand Docking. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 228–241.
- (33) Wold, S.; Albano, C.; Dunn, W. J., III; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johanson, E.; Lindberg, W.; Sjostrom, M. Multivariate Data Analysis in Chemistry. *NATO Sci. Ser., Ser. C* **1984**, *138* (Chemometrics), 17–95.
- (34) Wold, S.; Rhue, A.; Wold, H.; Dunn, W. J., III; The Covariance Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SAIM J. Sci. Stat. Comput.* **1994**, *5*, 735–743.
- (35) Clark, M.; Cramer, R. D., III. The Probability of Chance Correlation Using Partial Least Squares (PLS). *Quant. Struct.-Act. Relat.* **1993**, *12*, 137–145.
- (36) Sotriffer, C. A.; Ni, H.; McCammon, J. A. HIV-1 Integrase Inhibitor Interactions at the Active Site: Prediction of Binding Modes Unaffected by Crystal Packing. *J. Am. Chem. Soc.* **2000**, *122*, 6136–6137.
- (37) Sotriffer, C. A.; Ni, H.; McCammon, J. A. Active Site Binding Modes of HIV-1 Integrase Inhibitors. *J. Med. Chem.* **2000**, *43*, 4109–4117.

JM010399H