

CoMFA 3D-QSAR Analysis of HIV-1 RT Nonnucleoside Inhibitors, TIBO Derivatives Based on Docking Conformation and Alignment

Zhigang Zhou and Jeffry D. Madura*

Department of Chemistry and Biochemistry, Duquesne University, Pittsburgh, Pennsylvania 15282

Received March 28, 2004

HIV-1 RT is one of the key enzymes in the duplication of HIV-1. Inhibitors of HIV-1 RT are classified as nonnucleoside RT inhibitors (NNRTIs) and nucleoside analogues. NNRTIs bind in a region not associated with the active site of the enzyme. Within the NNRTI category, there is a set of inhibitors commonly referred to as TIBO inhibitors. Fifty TIBO inhibitors were used in the work to build 3-D QSAR models. The two known crystal structures of complexes are used to investigate and validate the docking protocol. The results show that the docking simulations reproduce the crystal complexes very well with RMSDs of ~ 1 Å and ~ 0.6 Å for 1REV and 1COU, respectively. The alignment of molecules and “active” conformation selection are the key to a successful 3D-QSAR model by CoMFA. The flexible docking (Autodock3) was used on determination of “active” conformation and molecular alignment, and CoMFA and CoMSIA were used to develop 3D-QSAR models of 50 TIBOs in the work. The 3D-QSAR models demonstrate a good ability to predict the activity of studied compounds ($r^2 = 0.972, 0.944, q^2 = 0.704, 0.776$). It is shown that the steric and electrostatic properties predicted by CoMFA contours can be related to the binding structure of the complex. The results demonstrate that the combination of ligand-based and receptor-based modeling is a powerful approach to build 3D-QSAR models.

INTRODUCTION

The Reverse Transcriptase of Human Immunodeficiency Virus type 1 (HIV-1 RT) function is to transcribe a single-stranded viral RNA genome into a double-stranded DNA, and it plays a vital role in the replication of HIV-1.^{1–4} Several drugs that target this enzyme have been approved to treat Acquired Immune Deficiency Syndrome (AIDS). There are two types of RT inhibitors. One type of RT inhibitor is commonly referred to as a nucleoside inhibitor. This inhibitor inserts as a nucleoside analogue into DNA and acts as a chain-terminating agent, therefore, terminating viral synthesis. The other type is called the nonnucleoside inhibitor (NNRTI).^{5–12} NNRTIs bind in a nonnucleoside binding pocket (NNBP) to inhibit the activity of RT. TIBO and its derivatives are a class of NNRTIs that have demonstrated good activity toward RT inhibition. One (Tivirapine) of them has moved onto the clinical development cycle.⁸ The crystal structures of several TIBOs/RT complexes are currently available.^{13,14} These complexes provide some insight into the binding and interactions of TIBOs in RT. However the inhibition model of TIBOs still needs to be elucidated in order to find and design new and more potent inhibitors that remain effective for HIV-1 RT mutants due to the presence of NNRTIs.

Docking is one method in which the binding of an inhibitor to a receptor can be explored.^{15–23} Comparative molecular field analysis (CoMFA)²⁴ and comparative molecular similarity indices analysis (CoMSIA)^{25,26} are powerful and versatile tools to build and design an activity model (QSAR) for a given set of molecules in rational drug design and

related applications.^{27–36} Recently, we use an Autodock3 to successfully dock a set of NNRTIs into RT. The calculated binding energies, based on the docked structures, agree well with the experimental activities.³⁷ QSAR models of 46 TIBOs were studied by Hannongbua et al.³⁸ using CoMFA and by Huuskonen³⁹ based on the atom level E-state indices and calculated molecular properties ($\log P$, MR). Also the correlation between activities and $\log P$ of several sets of TIBOs were explored by Garg et al.⁴⁰ Some works show that the binding affinity calculated by Monte Carlo and the Linear Response equation has a good correlation with the activity of TIBOs.^{41,42}

In CoMFA or other 3D-QSAR studies, the molecule alignment and conformation determination are so important that they affect the success of a model. In most cases a bound TIBO/RT complex is not available, and therefore a computation method has to be deployed to determine conformations and alignment of a set of molecules so that 3D-QSAR work can be carried out. Several strategies have been used to determine conformation and align molecules. Of them, docking is an attractive way to align molecules for CoMFA. Several applications of docking alignment with CoMFA have been reported.^{43–45}

In this paper, determination of the “active” conformation of each molecule and the molecular alignment are done using the flexible docking program, Autodock3.⁴⁶ The molecular alignment is done according to the electrostatic and structural properties of the active site of RT. Then 3D-QSAR models based on the active conformation and the aligned cluster are constructed using CoMFA and CoMSIA. The strategy of combining conformations and alignment obtained from the Autodock3 with the CoMFA produces a natural and reasonable elucidation of activation from a 3D-QSAR calculation.

* Corresponding author phone: (412)396-6341; fax: (412)396-5683; e-mail: Madura@duq.edu.

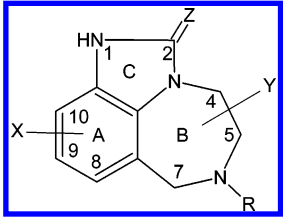
METHODS

Data Set and Molecule Preparation. The construction and preparation of molecular coordinates of all molecules were done using Molecular Operating Environment (MOE) program (Chemical Computing Group, Montreal, Canada). The starting coordinates of the HIV-1 RT/TIBO complex (pdb code, 1REV)¹⁴ and the HIV-1 RT/BM5 (pdb code, 1COU)⁵⁶ were taken from the Protein Data Bank.¹⁴ After hydrogen atoms were added using MOE, the substrate (9CI-TIBO) and the protein (RT) were saved separately. Partial charges for the protein were assigned from the AMBER94 force field.⁴⁷ The protein was minimized holding all non-hydrogen atoms fixed. All other inhibitors were built using the 9CI-TIBO as a template. The PEOE charge set⁴⁸ was used on the ligands, and full optimization was performed to minimize each structure. The structure and experimental activity (pIC₅₀) for the inhibitors used in this work are listed in Table 1.^{49–53}

Docking Simulation. Docking simulations of TIBO inhibitors into the RT NNRTI binding pocket were performed using Autodock3⁴⁶ in this study. All single bonds of a substrate were treated as flexible by allowing them to rotate freely. The Lamarckian Genetic Algorithm (LGA)⁵⁴ in Autodock3 was used to explore the energy landscape. The hybrid search technique consists of a global optimizer⁵⁵ modified from a genetic algorithm with 2-point crossover, random mutation, and a local optimizer with a Solis and Wets algorithm. A docking box of 60 × 60 × 60 points with a grid spacing of 0.375 Å was used in the calculations. Random conditions were used in the settings of seed, initial quaternion, coordinates, and torsions. A 0.2 Å step was used for translation, and a 25-degree was used for quaternion and torsion. The maximum number of energy evaluation was 250 000, and the maximum number of generations was 27 000. The rate of gene mutation was 0.02, and the rate of crossover was 0.8. The number of cycles was set to 10. So a total of 10 docking configurations were determined in each docking calculation. A “preferable” docking configuration was chosen based on the lowest empirical binding free energy and the most frequent cluster.³⁷ This configuration was treated as the “bioactive” binding conformation and used in two alignment schemes.

Alignment. The program SYBYL (version 6.8) was used in the development of the 3D-QSAR models. CoMFA and CoMSIA studies require the coordinates of molecules to be aligned according to reasonable bioactive conformations. In this case we used the conformation obtained from our docking calculations as the “bioactive” conformation needed in the alignment step. Two alignment computational designs were used to build the 3D-QSAR models. The first computational design is that the relative binding positions of all molecules obtained from the docking calculations were used. In other words, the alignment was done using flexible docking based on the steric and electrostatic properties of the binding pocket of the receptor (RT). The second computational design is to use the *Atom Fit* method in SYBYL. The 9CI-TIBO was used as a template to align the remaining inhibitor molecules. The core structure used for the alignment is shown in Figure 1. The reference atoms, marked black in the figure, are all in one plane and were used to align all the other molecules.

Table 1. Structures and HIV-1 RT Inhibitory Activity of Compounds Used in the Work



compd	X	Z	R	Y	pIC ₅₀
1	H	S	DMA ^a	5-Me(S)	7.36
2	9-Cl	S	DMA	5-Me(S)	7.47
3	8-Cl	S	DMA	5-Me(S)	8.37
4	8-F	S	DMA	5-Me(S)	8.24
5	8-SMe	S	DMA	5-Me(S)	8.30
6	8-OMe	S	DMA	5-Me(S)	7.47
7	8-OC ₂ H ₅	S	DMA	5-Me(S)	7.02
8	8-CN	S	DMA	5-Me(S)	7.25
9	8-CHO	S	DMA	5-Me(S)	6.73
10	8-CONH ₂	O	DMA	5-Me(S)	5.20
11	8-Br	O	DMA	5-Me(S)	7.33
12	8-Br	S	DMA	5-Me(S)	8.52
13	8-I	O	DMA	5-Me(S)	7.06
14	8-C = -CH	S	DMA	5-Me(S)	7.53
15	8-Me	O	DMA	5-Me(S)	6.00
16	8-Me	S	DMA	5-Me(S)	7.87
17	8-NMe ₂	O	CPM ^b	5-Me(S)	5.18
18	9-NH ₂	O	CPM	5-Me(S)	4.22
19	9-NMe ₂	O	CPM	5-Me(S)	5.18
20	9-NHCOMe	O	CPM	5-Me(S)	3.80
21	9-NO ₂	S	CPM	5-Me(S)	5.61
22	9-F	S	DMA	5-Me(S)	7.60
23	9-CF ₃	O	DMA	5-Me(S)	5.23
24	9-CF ₃	S	DMA	5-Me(S)	6.31
25	10-OMe	O	DMA	5-Me(S)	5.18
26	10-OMe	S	DMA	5-Me(S)	5.33
27	10-Br	S	DMA	5-Me(S)	5.97
28	H	O	CH ₂ CH=CH ₂	5-Me(S)	4.15
29	H	O	2-MA	5-Me(S)	4.33
30	H	O	CH ₂ CO ₂ Me	5-Me(S)	3.04
31	H	O	CH ₂ -2-furanyl	5-Me(S)	3.97
32	H	O	CH ₂ CH ₂ CH=CH ₂	5-Me(S)	4.30
33	H	O	CH ₂ CH ₂ CH ₃	5-Me(S)	4.05
34	H	O	CPM	5-Me(S)	4.36
35	H	O	CH ₂ CH=CHMe(E)	5-Me(S)	4.24
36	H	O	CH ₂ CH=CHMe(Z)	5-Me(S)	4.46
37	H	O	CH ₂ CH ₂ CH ₂ Me	5-Me(S)	4.00
38	H	O	DMA	5-Me(S)	4.90
39	H	O	CH ₂ C(Br)=CH ₂	5-Me(S)	4.21
40	H	O	CH ₂ C(Me)=CHMe(E)	5-Me(S)	4.54
41	H	O	CH ₂ C(C ₂ H ₅)=CH ₂	5-Me(S)	4.43
42	H	O	CH ₂ CH=CHC ₆ H ₅ (Z)	5-Me(S)	3.91
43	H	O	CH ₂ C(CH=CH ₂)=CH ₂	5-Me(S)	4.15
44	8-Cl	S	DMA	H	7.34
45	9-Cl	S	DMA	H	6.80
46	9-Cl	S	CPM	4-Me(R)	5.66
47	9-Cl	O	DMA	5-Me(S)	6.74
48	9-Cl	S	CPM	5-Me(S)	7.47
49	H	S	CPM	5-Me(S)	7.22
50	H	O	DMA	5-Me(S)	5.48

^a 3,3-Dimethylallyl. ^b Cyclopropylmethyl. ^c 2-Methylallyl.

CoMFA and CoMSIA 3D-QSAR Models. CoMFA²⁴ and CoMSIA^{25,26} descriptors were calculated using the following parameters. A 3D grid spacing of 2 Å in x, y, and z directions and an extension of 4 Å beyond the aligned molecules in all directions are used. An sp³ carbon probe atom with a charge of +1.0 and a vdW radius of 1.52 Å was used to calculate CoMFA steric and electrostatic field descriptors. The distant-dependent dielectric constant was used for treating electro-

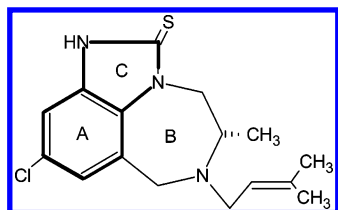


Figure 1. The molecule used as a template for molecule alignment. The bold part is the core for the alignment.

static term. A default cutoff of 30 kcal/mol was used to truncate the steric field and electrostatic field energies. The CoMFA standard method was used for scaling.

CoMSIA^{25,26} calculates the similarity descriptors by way of a grid lattice. For a molecule j with atoms i at the grid point q , the CoMSIA similarity indices A_F are calculated by the equation as follows

$$A_{F,k}^q(j) = -\sum \omega_{\text{probe},k} \omega_{jk} e^{-\alpha r_{iq}^2}$$

where ω_{jk} is the actual value of the physicochemical property k of atom i ; $\omega_{\text{probe},k}$ is the property of the probe atom with a preset charge (+1 in this case), a radius (1.53 Å), and a hydrophobicity of 1; and r_{iq} is the mutual distance between the probe atom at grid point q and atom i of the molecule. In the CoMSIA calculations, five physicochemical properties (steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor) were determined for all of the molecules. The same parameters used in the CoMFA calculations were used here.

PLS Analysis. After all of the CoMFA and CoMSIA descriptors were calculated, Partial Least-Squares analysis (PLS) was performed to obtain a 3D-QSAR model. The PLS method has been used in numerous applications in correlating the activity with various physicochemical properties. The PLS regression tries to build a relationship between a dependent variable (normally a activity) and several independent variables (property descriptors). The CoMFA standard scaling and column filtering of 2.0 were used in PLS analysis.

Cross-validations in PLS were done by the leave-one-out (LOO) procedure to find out the optimal number of components in building the regression models and to check statistic significance of models. The leave-one-out technique provides a good way to quantitatively evaluate the internal predictive ability of a model by removing one compound out at a time and then building the QSAR model and calculating the activity of the compound using the newly model constructed from the remaining compounds in the data set. The quality of a model is expressed as the cross-validated correlation coefficient q^2 .

The optimal number of components is the smallest cross-validated standard error of estimate SE_{press} (or the number giving the largest value of q^2 , as they are consistent most of the time).

The optimal number of components obtained is then used to derive the final QSAR model using all of the compounds (without cross-validation). The conventional correlation coefficient (r^2) is used to measure the quality of the model.

RESULTS AND DISCUSSIONS

Docking Reproduction of Crystal Complex Structures.

To validate the ability of the docking protocol to reproduce

Table 2. RMSD and Docking Energies from Docking Simulation of 9Cl-TIBO into Its Original Crystal Binding Pocket of RT (1REV) (kcal/mol)

configuration	RMSD ^a (Å)	$\Delta G_{\text{Binding}}$	ΔE_{Docked}	ΔE_{inter}	ΔE_{tors}^b
1	1.076	-11.16	-13.45	-13.03	-0.42
2	1.046	-11.16	-13.45	-13.03	-0.42
3	0.917	-11.15	-13.44	-13.02	-0.42
4	1.008	-11.19	-13.44	-13.06	-0.38
5	0.969	-11.08	-13.42	-12.94	-0.48
6	0.938	-11.13	-13.42	-13.00	-0.42
7	1.020	-10.74	-13.15	-12.61	-0.55
8	6.644	-7.18	-9.49	-9.05	-0.44
9	8.725	-5.89	-8.25	-7.76	-0.50
10	8.343	-5.75	-8.08	-7.62	-0.46

^a RMSD, root-mean-square deviation of coordinates between the configuration and initial position (from crystal structure). ^b Internal energy of ligand.

a TIBO binding structure of the crystal complex, two docking simulations were conducted against the crystal complexes. First the crystal complex of TIBO/RT (1REV) was used to validate the docking protocol. The 9Cl-TIBO was docked back into the NNRTI binding pocket of RT using a protocol described in the Method section. The RMSD of coordinates between the docked pose and the crystal complex and related energies are listed in Table 2. Out of 10 poses, seven poses have similar binding positions and the same orientation with crystal structures with RMSD ~ 1 Å. The other three binding poses are outside of the NNRTI binding pocket. The seven poses binding inside have nearly the same binding energy. They bind in the same pattern. The three poses binding outside have higher binding energy ($-7.18 \sim -5.75$ kcal/mol) than the inside poses ($-11.19 \sim -10.74$ kcal/mol). The torsional energy of these poses is small ($-0.55 \sim -0.38$ kcal/mol, $<4\%$ of internal energy), and the variance (≤ 0.17 kcal/mol) on the energy among these poses is very small. The energy is not a major factor on the binding. The reason is that 9-Cl-TIBO is a molecule with a triring and a short chain and is relative rigid.

To further investigate the docking protocol, a similar ligand/RT complex was selected to perform docking against the crystal complex. The ligand, (R)-(+)-5(9bH)-oxo-9b-phenyl-2,3-dihydrothiazolo[2,3-hetam a]isoindol-3-carboxylic acid methyl ester (BM5), consists of a triple fused ring and has a dimension similar to TIBO. The docking result (Table 3) of the complex shows that out of 10 docked poses, 9 have very similar binding positions and the same orientation as the crystal complex. One pose binds outside of the NNRTI binding pocket. Similar to the 9Cl-TIBO docking, the energy ($-12.67 \sim -12.49$ kcal/mol) of poses inside is lower than that (-6.07 kcal/mol) of the pose outside. In the two cases, the ligands favor binding inside. As BM5 has two single chain linked groups, it is more flexible than 9Cl-TIBO. The torsional energy of BM5 (1.48–1.72 kcal/mol, 12% of internal energy) is higher than 9Cl-TIBO. The torsional energy of the docking indicates that BM5 adopts an unfavorable conformation compared to the free state with a positive torsional energy. The two docking simulations demonstrate that the docking protocol can reproduce crystal structures of NNRTI/RT complexes well. The preferable docked structures are very similar to the corresponding crystal complex. This result is consistent with our previous applications of docking simulation, in which the similar

Table 3. RMSD and Docking Energies from Docking Simulation of BM5 ((R)-(+)-5(9bH)-Oxo-9b-phenyl-2,3-dihydrothiazolo-[2,3-hetham alisoindol-3-carboxylic Acid Methyl Ester) into Its Original Crystal Binding Pocket of RT (1COU) (kcal/mol)

configuration	RMSD ^a (Å)	$\Delta G_{\text{Binding}}$	ΔE_{Docked}	ΔE_{inter}	ΔE_{tors}^b
1	0.647	-12.5	-12.27	-13.75	1.48
2	0.643	-12.52	-12.26	-13.77	1.51
3	0.636	-12.51	-12.25	-13.75	1.5
4	0.685	-12.51	-12.24	-13.75	1.51
5	0.693	-12.66	-12.2	-13.91	1.71
6	0.679	-12.65	-12.2	-13.89	1.69
7	0.668	-12.52	-12.2	-13.76	1.56
8	0.629	-12.67	-12.19	-13.91	1.72
9	0.63	-12.49	-12.19	-13.74	1.55
10	7.409	-6.07	-5.44	-7.32	1.88

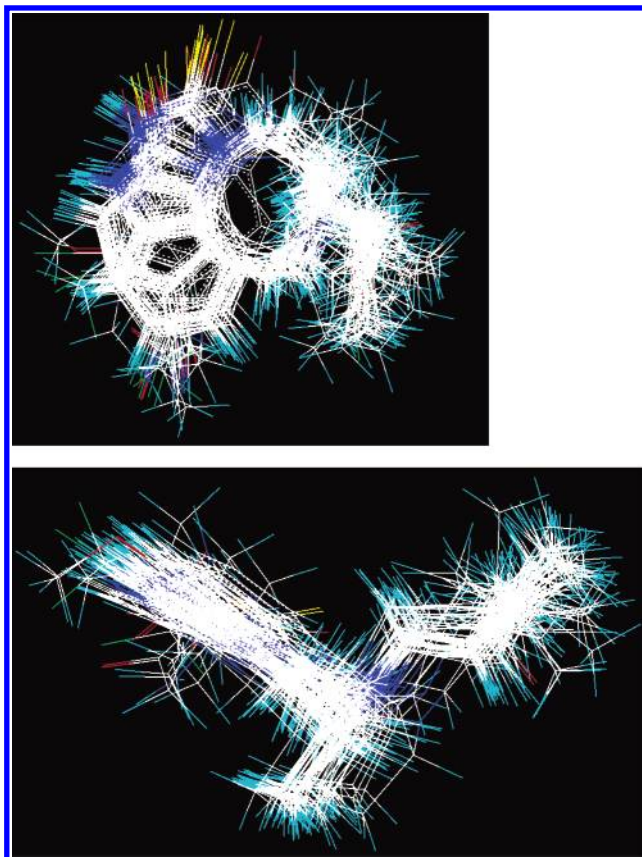
^a RMSD, root-mean-square deviation of coordinates between the configuration and initial position (from crystal structure). ^b Internal energy of ligand.

docking protocol has been successful in reproducing several RT crystal complexes (1VRT, 3HVT, and 1REV) and applying a series of ligands into RT.³⁷ This docking protocol was used to dock the series of TIBOs into the NNRTI binding pocket of RT to get their binding poses.

Docking of TIBOs. In every TIBO docking simulation, there is a dominating cluster that has lower binding free energy than the pose docked outside. The poses in the cluster for all TIBOs have similar binding positions and the same orientation with each other and with the 9CI-TIBO in the crystal complex. As these TIBOs have a common core and are similar to each other, they are assumed to have a similar binding pattern in RT. One pose was picked up from the cluster for each molecule as a “preferable” binding pose and was superposed together (Figure 2). It is seen that all these binding poses can be clustered into two. One group shifts a little from the other. After checking which molecules are in each cluster, we found that all molecules with a non-hydrogen 8-substituent fall into one cluster, and all others with an H at 8 fall into other cluster. The substituents on the 8-position will interact with the Tyr188, Phe227, and Leu234 residues of the P66 subunit. The position difference of the two clusters is quantitatively measured in terms of RMSD of the common core (ring A, B, and C on Figure 1) between each pose and compound 9 in Table 4. The RMSD within a cluster is smaller than 1 Å and that between clusters is a little larger than 1 Å.

From the energies in Table 4, we can see that all these molecules have a good binding with a negative estimated free energy of binding (FEB, <-10.11 kcal/mol) and a negative intermolecular potential energy (<11.35 kcal/mol). These molecules can fit into the binding pocket of RT without clash contact. The torsional energy of each molecule is very small and most are negative. These molecules can adopt a reasonable conformation within the active site. The calculated free energy of binding (FEB) (Table 4) by Autodock has some similar trend with actual activity (pIC₅₀). But the linear correlation between the calculated FEB is very low with a square of correlation constant $R^2 < 0.1$.

From the structural and energetic information, we can say that this docking protocol is suitable for docking this series of TIBOs in a single rigid binding pocket of RT and found a reasonable bound conformation for these molecules. Currently, the docking simulation can just treat a receptor

**Figure 2.** Superposition of all TIBOs aligned by the docking. All molecules are aligned according to the bound position in the nonnucleoside binding pocket of RT using flexible docking (Autodock3).

as rigid. It is a risk to dock a large molecule into a binding pocket which has no right pocket to accommodate the substrate, in a case where side chain conformations or pocket shape need to be changed to make the substrate binding possible. In other words, the docking method is not suitable for a case where an induced-fitting occurs. In our case, the docking method produced convincing results, and this method can be used in other studies.

Conformation Determination and Molecule Alignment.

In previous docking simulations, a preferable binding pose was chosen for each molecule. This conformation is treated as any “active” conformation when a ligand binds in the NNRT binding pocket of RT, which was determined by docking simulation according to the characteristics of the NNRT binding pocket. Meanwhile the whole set of molecules is aligned together. To compare with the alignment, the *atom fit* alignment in the Sybyl is used to align the whole set of molecules. In the later alignment, the docking-determined conformation is still used for each molecule. The two alignments (*Atom fit* and flexible docking) were used to explore the effect of molecular alignment on the CoMFA and CoMSIA analysis. The aligned molecules by docking and atom fit are shown in Figures 2 and 3, respectively. In Figure 3, one observes that the overall overlap of the tripping part is nearly the same. Side chains of molecules are located in different locations (especially the R group). On the other hand, the docking aligned cluster (Figure 2) has a different picture from the atom fit aligned cluster. It can be seen that the docking alignment produced a cluster with a more

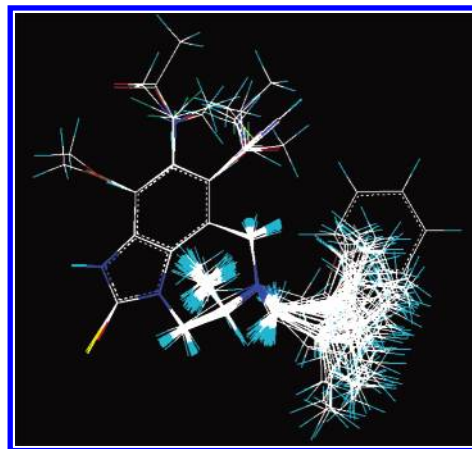
Table 4. RMSD and Docking Energies from the “Preferable” Docking Pose from Simulation of Each TIBO to 9Cl-TIBO to the NNRTI Binding Pocket of RT (1REV) (kcal/mol)

compd	RMSD ^a (Å)	$\Delta G_{\text{Binding}}$	ΔE_{Docked}	ΔE_{inter}	ΔE_{tors}^b
1	1.106	-10.91	-13.35	-12.78	-0.57
2	1.205	-11.16	-13.45	-13.03	-0.42
3	0.0	-11.51	-13.81	-13.38	-0.43
4	0.363	-11.46	-13.85	-13.33	-0.52
5	0.264	-11.6	-14.72	-14.09	-0.63
6	0.373	-11.18	-13.76	-13.67	-0.08
7	0.372	-11.95	-14.88	-14.75	-0.14
8	0.259	-12.03	-14.88	-14.2	-0.68
9	0.186	-11.61	-14.48	-13.79	-0.7
10	0.373	-12.27	-15.13	-14.44	-0.68
11	0.233	-11.6	-13.99	-13.47	-0.52
12	0.299	-11.6	-14.01	-13.47	-0.54
13	0.139	-11.47	-13.84	-13.34	-0.51
14	0.276	-11.95	-15.12	-14.44	-0.68
15	0.224	-11.47	-14.28	-13.65	-0.62
16	0.212	-11.35	-14.13	-13.53	-0.61
17	0.291	-11.2	-13.05	-12.45	-0.6
18	1.160	-11.39	-13.12	-12.63	-0.49
19	1.688	-11.76	-13.98	-13.62	-0.36
20	1.582	-11.48	-13.89	-13.66	-0.23
21	1.201	-11.04	-12.86	-12.29	-0.58
22	1.352	-11.22	-13.66	-13.09	-0.57
23	1.707	-11.71	-13.78	-13.89	0.11
24	1.585	-12.13	-14.19	-14.3	0.11
25	1.236	-11.31	-13.75	-13.8	0.05
26	1.014	-11.22	-13.74	-13.71	-0.03
27	1.211	-11.31	-13.71	-13.18	-0.53
28	1.083	-10.78	-13.17	-12.65	-0.52
29	1.214	-10.16	-11.78	-11.4	-0.37
30	0.982	-10.35	-12.44	-11.91	-0.53
31	1.245	-10.26	-12.45	-11.81	-0.63
32	1.154	-12.25	-13.74	-13.19	-0.55
33	1.193	-10.63	-12.66	-12.18	-0.48
34	1.227	-10.11	-11.77	-11.35	-0.42
35	1.226	-11.15	-12.59	-12.09	-0.51
36	1.214	-10.64	-12.64	-12.19	-0.44
37	1.202	-10.64	-12.64	-12.19	-0.44
38	1.221	-10.6	-12.67	-12.15	-0.52
39	1.172	-11.06	-13.51	-12.93	-0.59
40	1.254	-10.57	-12.29	-11.81	-0.48
41	1.319	-10.66	-13	-12.52	-0.47
42	1.261	-10.68	-13.03	-12.55	-0.48
43	1.409	-13.95	-15.84	-15.51	-0.34
44	0.466	-11.07	-12.99	-12.63	-0.37
45	1.292	-10.63	-12.5	-12.19	-0.31
46	1.442	-11.27	-12.58	-12.2	-0.38
47	1.369	-11.2	-13.61	-13.07	-0.54
48	1.352	-11.18	-12.61	-12.11	-0.49
49	1.174	-10.96	-12.4	-11.9	-0.5
50	1.183	-11.05	-13.5	-12.92	-0.58

^a RMSD, root-mean-square deviation of common core (ring A, B, and C) between the “preferable” docking poses of the compound and compound 3 (8-Cl-TIBO or tivrapipe). ^b Internal energy of ligand.

uniform surface contour. It is natural, as docking aligns molecules according to the feature of the binding site.

The position of each atom is important to CoMFA and CoMSIA, because the descriptors are calculated based on the 3-D space grid. So the method to determine conformation of each molecule and the way to align molecules together are keys to get a reasonable model. In the study, the same conformation for each molecule is used in the two schemes. Just the alignment (rigid supposition) difference is compared in the work. In the atom fit alignment (Figure 3) the atoms of the triring part have nearly the same coordinates for all molecules. The contributions of these atoms to CoMFA and CoMSIA descriptors are nearly the same. They will be

**Figure 3.** Superposition of all TIBOs aligned by atom fit, in which all molecules are aligned according to the core atoms in the ring A and ring C (see Figure 1 for ring A and ring C).**Table 5.** Comparison of PLS Statistics Results of the 3D-QSAR Models of CoMFA and CoMSIA

	docking alignment		atom fit alignment	
	CoMFA	CoMSIA	CoMFA	CoMSIA
PCs	4	5	6	5
r^2	0.972	0.944	0.959	0.889
q^2	0.704	0.776	0.661	0.682
SE_{press}	0.326	0.381	0.335	0.482
	fraction			
steric	0.41	0.077	0.50	0.070
electrostatic	0.59	0.224	0.50	0.246
hydrophobic		0.483		0.453
H-acceptor		0.216		0.231

canceled when comparing the difference among these molecules. So in the atom fit alignment, nearly all of the difference of the final calculated activity is contributed by the atoms not from the triring part. So in the alignment CoMFA (or CoMSIA) explores the activity difference by focusing on these side chains. It seems to be reasonable for ligand-based drug design. In docking alignment each molecule tried to fit the binding pocket of RT the best in terms of steric and electronic interactions with the protein. The centers of Ring A and C are at different points between the two clusters (Figure 2). The core atoms likely have a contribution to the difference of CoMFA and CoMSIA descriptors for different molecules. So the difference on the final calculated activity comes from side chain atoms as well as core atoms. The model encodes the feature of the binding site and will be more reasonable for the study activity than the ligand-based model.

QSAR Models. The CoMFA and CoMSIA descriptors were calculated for the two models. QSAR models were analyzed and optimized using the PLS method based on these descriptors. The statistical results of CoMFA and CoMSIA studied are summarized in Table 5. These analyses were based on the clusters of molecules that were aligned by the two methods. The regression coefficient (r^2) and the cross-validation coefficient (q^2) of the QSAR model constructed by CoMFA based on the docking alignment are 0.972 and 0.704, respectively. The two coefficients for the CoMSIA model are 0.944 and 0.776, respectively. Based on the coefficient values, the CoMFA model (without validation) appears slightly better than CoMSIA models. CoMSIA seems to have better predictability than CoMFA. Both models

Table 6. Predicted Activities and Residues for the CoMFA and CoMSIA Models of Docking Alignment

compd	obsd pIC ₅₀	CoMFA				CoMSIA			
		leave-one-out		no-validation		leave-one-out		no-validation	
		activity	residue	activity	residue	activity	residue	activity	residue
1	7.36	7.461	-0.1	7.633	-0.27	6.812	0.548	6.966	0.394
2	7.47	6.563	1.037	7.202	0.398	7.273	0.327	7.42	0.18
3	8.37	8.097	0.273	8.089	0.281	8.378	-0.01	8.439	-0.07
4	8.24	7.99	0.25	7.956	0.284	7.555	0.685	7.719	0.521
5	8.30	7.197	1.103	8.093	0.207	7.355	0.945	7.768	0.532
6	7.47	6.977	0.493	7.221	0.249	6.898	0.572	7.156	0.314
7	7.02	6.818	0.202	7.024	-0	7.609	-0.59	7.567	-0.55
8	7.25	7.741	-0.49	7.482	-0.23	6.824	0.426	7.324	-0.07
9	6.73	7.656	-0.93	7.052	-0.32	7.986	-1.26	6.673	0.057
10	5.20	7.564	-2.36	5.423	-0.22	5.031	0.169	4.681	0.519
11	7.33	6.499	0.831	7.306	0.024	6.413	0.917	6.783	0.547
12	8.52	7.179	1.341	8.446	0.074	7.165	1.355	7.833	0.687
13	7.06	7.551	-0.49	7.6	-0.54	7.458	-0.4	7.514	-0.45
14	7.53	7.336	0.194	7.548	-0.02	6.791	0.739	7.405	0.125
15	6.00	7.897	-1.9	6.168	-0.17	7.052	-1.05	6.311	-0.31
16	7.87	7.178	0.692	7.516	0.354	7.725	0.145	7.794	0.076
17	5.18	5.233	-0.05	5.366	-0.19	5.504	-0.32	5.362	-0.18
18	4.22	6.059	-1.84	4.563	-0.34	3.948	0.272	3.984	0.236
19	5.18	4.171	1.009	5.207	-0.03	3.918	1.262	5.072	0.108
20	3.80	4.645	-0.85	4.016	-0.22	4.916	-1.12	3.88	-0.08
21	5.61	6.149	-0.54	5.969	-0.36	5.343	0.267	5.46	0.15
22	7.60	7.79	-0.19	7.845	-0.25	7.218	0.382	7.315	0.285
23	5.23	4.316	0.914	5.031	0.199	5.203	0.027	5.091	0.139
24	6.31	6.078	0.232	6.323	-0.01	6.609	-0.3	6.473	-0.16
25	5.18	4.344	0.836	4.983	0.197	4.307	0.873	4.566	0.614
26	5.33	6.09	-0.76	5.205	0.125	7.173	-1.84	6.369	-1.04
27	5.97	6.798	-0.83	6.124	-0.15	7.152	-1.18	6.397	-0.43
28	4.15	3.034	1.116	3.533	0.617	4.119	0.031	4.112	0.038
29	4.33	4.377	-0.05	4.336	-0.01	4.751	-0.42	4.581	-0.25
30	3.04	3.148	-0.08	2.977	0.093	3.265	-0.2	3.035	0.035
31	3.97	4.741	-0.77	4.284	-0.31	4.594	-0.62	4.157	-0.19
32	4.30	4.087	0.213	4.058	0.242	4.8	-0.5	4.667	-0.37
33	4.05	4.043	0.007	3.952	0.098	4.673	-0.62	4.557	-0.51
34	4.36	4.992	-0.63	4.598	-0.24	4.607	-0.25	4.56	-0.2
35	4.24	4.038	0.202	4.142	0.098	4.478	-0.24	4.416	-0.18
36	4.46	4.066	0.394	4.153	0.307	4.215	0.245	4.247	0.213
37	4.00	4.551	-0.55	4.209	-0.21	4.623	-0.62	4.495	-0.5
38	4.90	5.691	-0.79	5.448	-0.55	4.971	-0.07	4.927	-0.03
39	4.21	5.033	-0.82	4.404	-0.19	5.392	-1.18	4.526	-0.32
40	4.54	5.128	-0.59	4.628	-0.09	4.792	-0.25	4.701	-0.16
41	4.43	4.273	0.157	4.255	0.175	4.433	-0	4.402	0.028
42	3.91	5.131	-1.22	3.906	0.004	4.078	-0.17	3.777	0.133
43	4.15	4.018	0.132	4.093	0.057	4.148	0.002	4.056	0.094
44	7.34	7.394	-0.05	7.156	0.184	8.341	-1	8.111	-0.77
45	6.80	7.585	-0.79	7.027	-0.23	7.847	-1.05	7.608	-0.81
46	5.66	6.571	-0.91	5.741	-0.08	6.485	-0.83	5.645	0.015
47	6.74	5.967	0.773	6.345	0.395	5.757	0.983	6.098	0.642
48	7.47	7.118	0.352	7.47	3E-04	7.62	-0.15	7.722	-0.25
49	7.22	6.091	1.129	6.804	0.416	6.341	0.879	6.582	0.638
50	5.48	5.166	0.314	5.328	0.152	4.864	0.616	4.934	0.546

exhibit good predictive capabilities as shown by the cross-validation method. The standard errors of estimate for the two models are 0.326 and 0.381, respectively.

In the QSAR models based on the atom fit alignment, the regression coefficient (r^2) and the cross-validation correlation coefficient (q^2) for the CoMFA models are 0.959 and 0.661, respectively. The two coefficients of the CoMSIA model are 0.889 and 0.682, respectively. Based on the coefficients, the CoMFA model is a little better than the CoMSIA model. They both demonstrate good predictive capabilities. The standard errors of the two models are not high (0.355 and 0.482).

Comparing the docking models and atom fit models, it is seen that the correlation coefficients of docking models are all slightly higher than the corresponding values of atom fit

models. The docking alignment produced better results than the traditional atom fit alignment. Compared to others work based on 42 TIBOs,³⁹ the docking models of 52 TIBOs are slightly better. There is a need to mention that the atom fit models also use the conformation determined by docking for each molecule. So the atom fit models still encode information from docking. The docking simulation demonstrates a good ability to predict a binding conformation and alignment for CoMFA to build a QSAR model.

The CoMFA and CoMSIA calculated electrostatic and steric properties (descriptors) are based on the grid built around these molecules. As discussed earlier, in the atom fit system, the largest difference between these calculated properties for these molecules derives from the side chain groups rather than from the three rings (A, B, and C). On

Table 7. Predicted Activities and Residues for the CoMFA and CoMSIA Models of *Atom Fit* Alignment

compd	obsd pIC ₅₀	CoMFA				CoMSIA			
		leave-one-out		no-validation		leave-one-out		no-validation	
		activity	residue	activity	residue	activity	residue	activity	residue
1	7.36	7.289	0.071	7.431	-0.07	6.66	0.7	6.737	0.623
2	7.47	5.862	1.738	7.019	0.581	7.053	0.547	7.217	0.383
3	8.37	8.133	0.237	8.401	-0.03	8.442	-0.07	8.506	-0.14
4	8.24	7.716	0.524	8.08	0.16	7.567	0.673	7.644	0.596
5	8.30	7.471	0.829	8.128	0.172	7.542	0.758	7.718	0.582
6	7.47	6.733	0.737	7.328	0.142	6.461	1.009	6.567	0.903
7	7.02	8.293	-1.27	7.291	-0.27	7.235	-0.22	7.171	-0.15
8	7.25	7.097	0.153	7.209	0.041	6.627	0.623	7.525	-0.28
9	6.73	7.418	-0.69	6.7	0.03	7.909	-1.18	7.234	-0.5
10	5.20	6.53	-1.33	5.043	0.157	4.692	0.508	4.214	0.986
11	7.33	7.047	0.283	7.21	0.12	6.533	0.797	6.781	0.549
12	8.52	6.64	1.88	8.505	0.015	6.236	2.284	7.358	1.162
13	7.06	7.122	-0.06	7.261	-0.2	7.292	-0.23	7.524	-0.46
14	7.53	6.549	0.981	7.095	0.435	6.192	1.338	6.864	0.666
15	6.00	6.534	-0.53	5.852	0.148	7.186	-1.19	6.292	-0.29
16	7.87	7.813	0.057	7.959	-0.09	7.615	0.255	7.651	0.219
17	5.18	5.72	-0.54	5.229	-0.05	5.652	-0.47	5.413	-0.23
18	4.22	4.855	-0.64	4.159	0.061	4.5	-0.28	4.332	-0.11
19	5.18	4.452	0.728	5.077	0.103	4.548	0.632	4.636	0.544
20	3.80	5.227	-1.43	3.826	-0.03	4.994	-1.19	4.683	-0.88
21	5.61	5.043	0.567	5.436	0.174	5.351	0.259	5.433	0.177
22	7.60	7.922	-0.32	8.021	-0.42	6.827	0.773	6.911	0.689
23	5.23	4.681	0.549	4.823	0.407	5.071	0.159	5.132	0.098
24	6.31	8.722	-2.41	6.299	0.011	7.883	-1.57	7.642	-1.33
25	5.18	4.09	1.09	4.894	0.286	4.379	0.801	4.503	0.677
26	5.33	6.13	-0.8	5.414	-0.08	6.739	-1.41	6.531	-1.2
27	5.97	7.221	-1.25	6.481	-0.51	7.178	-1.21	6.903	-0.93
28	4.15	3.972	0.178	3.986	0.164	4.434	-0.28	4.36	-0.21
29	4.33	4.743	-0.41	4.579	-0.25	4.638	-0.31	4.583	-0.25
30	3.04	3.417	-0.35	2.829	0.241	4.363	-1.29	3.003	0.067
31	3.97	3.84	0.13	3.76	0.21	4.402	-0.43	4.169	-0.2
32	4.30	4.839	-0.54	4.68	-0.38	4.622	-0.32	4.538	-0.24
33	4.05	4.453	-0.4	4.279	-0.23	4.575	-0.53	4.502	-0.45
34	4.36	5.082	-0.72	4.919	-0.56	4.496	-0.14	4.474	-0.11
35	4.24	4.35	-0.11	4.277	-0.04	4.481	-0.24	4.444	-0.2
36	4.46	4.302	0.158	4.338	0.122	4.477	-0.02	4.477	-0.02
37	4.00	4.718	-0.72	4.479	-0.48	4.502	-0.5	4.453	-0.45
38	4.90	5.193	-0.29	5.15	-0.25	4.728	0.172	4.736	0.164
39	4.21	4.926	-0.72	4.589	-0.38	4.715	-0.51	4.53	-0.32
40	4.54	5.363	-0.82	4.729	-0.19	5.052	-0.51	4.948	-0.41
41	4.43	4.233	0.197	4.257	0.173	4.652	-0.22	4.6	-0.17
42	3.91	5.003	-1.09	4.04	-0.13	4.686	-0.78	4.39	-0.48
43	4.15	4.252	-0.1	4.128	0.022	4.753	-0.6	4.566	-0.42
44	7.34	7.749	-0.41	7.597	-0.26	8.395	-1.06	8.291	-0.95
45	6.80	7.477	-0.68	7.306	-0.51	7.101	-0.3	7.088	-0.29
46	5.66	6.9	-1.24	5.831	-0.17	7.372	-1.71	6.529	-0.87
47	6.74	6.325	0.415	6.687	0.053	5.025	1.715	5.289	1.451
48	7.47	6.69	0.78	7.196	0.274	6.833	0.637	6.958	0.512
49	7.22	5.808	1.412	6.431	0.789	6.421	0.799	6.499	0.721
50	5.48	4.816	0.664	5.003	0.477	4.626	0.854	4.693	0.787

the other hand, in the docking alignment system, the difference between these calculated properties derives from all the atoms.

Also it is observed that with the CoMFA results, the steric and electrostatic contributions are similar (0.41 vs 0.59 in docking alignment, 0.50 vs 0.50 in atom fit alignment, in Table 5). The electrostatic contributions in the CoMSIA models are nearly three times larger than the steric contributions. It is also observed that the hydrophobic contributions are the largest part in the CoMSIA models. This is consistent with the concept that the NNRTI active site of RT is hydrophobic. The H-acceptor effect has a similar contribution as does the electrostatic effect. The steric effect seems to be the least important among the four terms. In our previous docking study,³⁷ it was recognized that the hydrophobic effect and water bridge network around the entrance of the active

site plays a role on the binding. It is believed that this water network helps stabilize the binding of NNRTI in RT. These results support the idea that the hydrophobic and hydrophilic properties of NNRTIs are important in the design of NNRTIs.

The predicted activities and residues (between the predicted value and the experimental value) by validation (LOO) and no-validation methods of the four models are listed in Tables 6 and 7. In the docking models (Table 6), the numbers of residues which are larger than 1 in the LOO predictions for both CoMFA and CoMSIA models are 10 (20% out of all). The number of residues which are larger than 0.5 in no-validations for CoMFA and CoMSIA models are 3 (6%) and 10 (20%), respectively. It shows that the CoMFA predictability is slightly better than CoMSIA. In the atom fit models (Table 7), the numbers of residues which are larger than 1 in LOO predictions for CoMFA and CoMSIA models

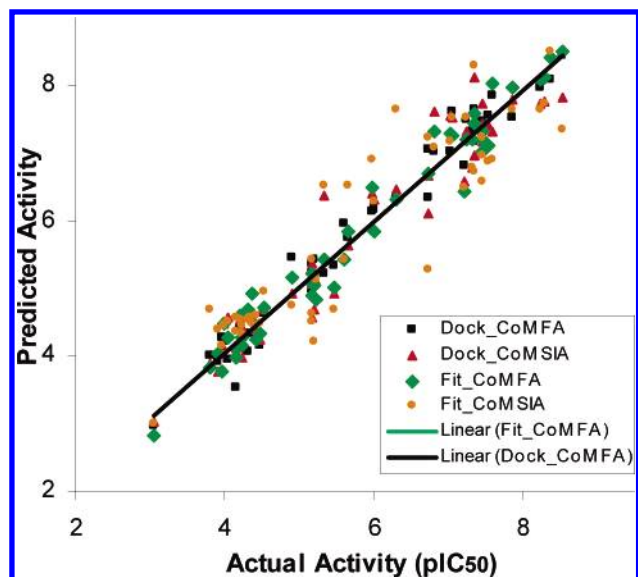


Figure 4. Experimental activity (pIC_{50}) vs calculated activity values of 3D-QSAR models. The black squares and red triangles are CoMFA and CoMSIA results based on docking alignment and conformation determination. The green diamonds and yellow cycles are CoMFA and CoMSIA results based on atom fit alignment and docking conformation determination. The black line is the trend line of the CoMFA model of docking, and the green line is the trend line of the CoMFA model of atom fit. The two set data nearly have the same trend lines.

are 11 (22%) and 13 (26%), respectively. The number of residues which are larger than 0.5 in no-validations for CoMFA and CoMSIA are 5 (10%) and 19 (38%), respectively. This shows again that CoMFA predictability seems to be better than CoMSIA.

Comparing the number of larger residues between docking models and atom fit models (10 vs 11 in CoMFA-LOO, 3 vs 5 in CoMFA-no-validation, 10 vs 13 in CoMSIA-LOO, and 10 vs 19 in CoMSIA-no-validation), it is clearly seen that the docking alignment produced better results than the atom fit. It reensures the conclusion that we came to in the previous section.

To visualize the predicting results of the models, the final calculated activity from no-validation calculation vs experimental activity of each compound using the atom fit and docking models is shown in Figure 4. The black squares and red triangles are CoMFA and CoMSIA results based on docking alignment and conformation determination. The green diamonds and yellow cycles are CoMFA and CoMSIA results based on atom fit alignment and docking conformation determination. The black line is the trend line of the CoMFA model of docking, and the green line is the trend line of the CoMFA model of atom fit. It is seen that the trend lines of CoMFA models based on the docking and atom fit are nearly identical. The fits are nearly perfect with a slope of 1. It indicates that the CoMFA and CoMSIA models do not have a systematic deviation. The prediction residuals for both models are shown in Figure 5. It is observed that the CoMFA model (black and green) has fewer long-bars than the CoMSIA model (red and yellow). This means that the CoMFA model yields a little better prediction than the CoMSIA model. By comparing the residue bars of docking models and atom fit models, it is noticed that more black bars (Docking) are shorter than green bars (atom fit) and most red bars (Docking) are shorter than yellow bars (atom

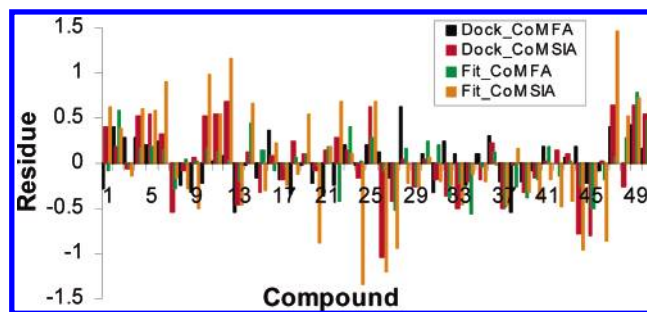


Figure 5. The residuals between experimental activities and predicted activities from the four QSAR models. The black and red bars are CoMFA and CoMSIA results based on docking alignment and conformation determination. The green and yellow bars are CoMFA and CoMSIA results based on atom fit alignment and docking conformation determination.

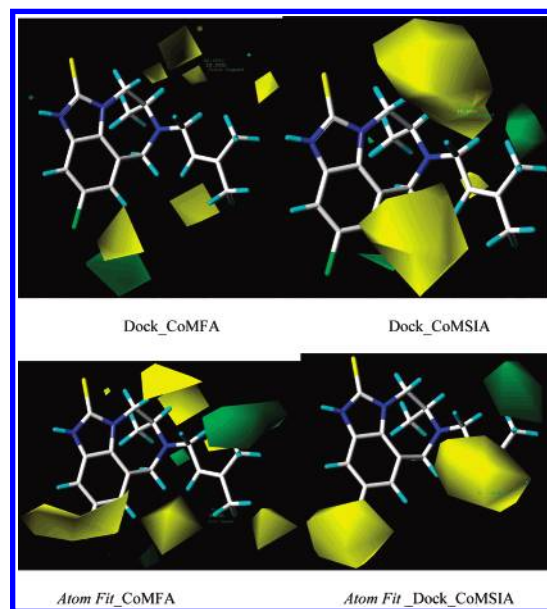


Figure 6. Steric contour maps of the CoMFA and CoMSIA models. 9-Cl-TIBO is used to demonstrate the corresponding areas where a change in a molecule may affect its activity. Green contours indicate the regions where the addition of bulky groups may increase activity. Yellow contours indicate the regions where the addition of bulky groups may decrease activity.

fit). So docking models produce better predictions to activity of these TIBOs than atom fit models. It is also seen that for most compounds, the Dock_CoMFA (black) and Fit_CoMFA (green) have the same signs (above for + or below for -), and Dock_CoMSIA (red) and Fit_CoMSIA (yellow) also have the same signs. The alignments give similar deviation on prediction of activity.

Graphical Interpretation of the Results. To further explore the hypothetical interaction feature of a ligand with its receptor, the steric and electrostatic contour maps of the four models are shown in Figures 6 and 7. The compound 9-Cl-TIBO is used in the figures for analysis. Considering the steric contour first, it is seen that two regions (green) have been recognized as the regions where an addition of bulky groups may increase activity. The first steric-favorable region is in the direction of the R group, which is recognized by Dock_CoMSIA, atom fit_CoMFA, and CoMSIA. It indicates that changing to a larger group from the methyl group could increase the activity of the ligand. The second steric-favorable region is near the 8 position of ring A above

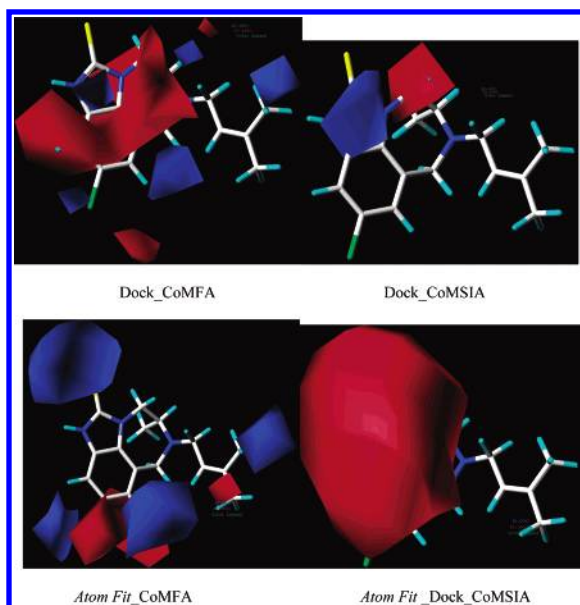


Figure 7. Electrostatic contour of the CoMFA and CoMSIA models from docking alignment. 9-Cl-TIBO is used to demonstrate the corresponding areas where a change on molecule may affect its activity. Blue contours indicate regions where positive groups may increase activity. Red contours indicate regions where negative groups may increase activity.

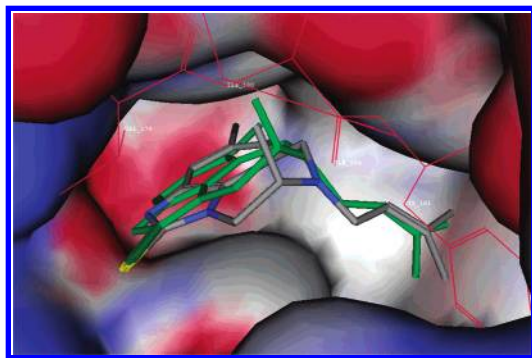


Figure 8. The actual binding structure of compound 2 (colored by element, 9-Cl-TIBO) and docked structure of compound 3 (green, 8-Cl-TIBO or Tivirapine) from crystal structure. The blue is a positive area and red is a negative area for the active site surface of RT.

the ring plane, which is recognized by Dock_CoMFA, Dock_CoMSIA, and atom fit_CoMFA. The 4- and 5-positions of ring B are recognized as nonfavorable regions by Dock_CoMFA, Dock_CoMSIA, and atomfit_CoMFA. The 9-position on ring A is also recognized as a nonfavorable region. The four models give some common features but not all are consistent with one another. These features are based on ligands for further improving the need to be conformed by the binding site feature of the receptor, because a ligand activity results from its binding and interaction with the receptor. Examining the bound complex of RT/9-Cl-TIBO (1REV) and RT/8-Cl-TIBO shown in Figure 8, it is seen that there is some but not a large amount of room near the region of the methyl group of R, which corresponds to the green region in the CoMFA steric contour. The group at the position should be a proper one which can fit into the binding site, not a large one which could penetrate into protein after the ligand shift toward another direction as far as possible. But there is no room for a large substituent at the 8-position. There is more room at the 9-position than at

the 8-position. So all docked structures with a non-hydrogen at the 8-position shift a little bit from those without a non-hydrogen substituent at the 8-position, which was discussed previously (for Table 2). It is also seen that the ligand in the complex makes good contact with the protein near the linking area of the R group of ring B (unfavorable region colored in yellow). Some of the predicted results from the steric CoMFA contour are consistent with the complex. This suggests that to combine the ligand-based CoMFA and CoMSIA QSAR develops a model with protein-based docking simulation which will provide a clearer direction and can avoid some blind images.

On the electrostatic contour map (Figure 7), is it seen that different models give different indications on electrostatic property for further improvement. CoMFA models indicate more improvement regions than CoMSIA models. In the Dock_CoMFA model, the 8, 9 and one methyl group of the R substituent are marked as positive-favorable regions (blue). The ring area is marked as the negative-favorable region. The four models do not give the same pattern. Compared with the pocket electrostatic property (Figure 8), the negative-favorable area matches the protein electrostatic characteristics, except for the 10-position of ring A, which is indicated as a negative-favorable region and located in a negative area of the binding pocket. The positive-favorable 9-position and the methyl group direct to a weak negative area of the pocket. But the positive-favorable 8-position seems not to match the pocket property. The most obvious difference between Dock_CoMFA and atom fit_CoMFA is that atom fit_CoMFA indicates that the 2-position of ring C is positive-favorable, which does not match the pocket property. The CoMSIA models do not provide much useful information on electrostatic prediction in the study.

In docking simulation, electrostatic is one of the major forces to determine the binding structure for a ligand. If studied ligands have similar electrostatic patterns, they can bind to the same receptor in a similar position and the same orientation. If one ligand has a different electrostatic pattern, for example, one negative substituent goes to a position where a positive should go. If the electrostatic dissimilarity affects the binding enough, the ligand could flip away from another one in a docked pose in docking simulation. In this case, the ligand should be treated as an outlier and excluded from the 3-D CoMFA (CoMSIA) QSAR model. This indicates that to check the cluster of docked poses (as shown in Figures 2 and 3) is important before QSAR development.

To determine "active" conformation and position-orientation for each molecule is a prerequisite for CoMFA (CoMSIA) 3-D QSAR. In a conventional template alignment, all other molecules are aligned along one known molecule or some common features based on steric and (or) electrostatic, etc. properties. Generally speaking the position-orientation alignment is the easier part. The conformation determination is a challenging task for some molecules. In some cases, there is no right way to determine the conformation of some molecules. As there is not a receptor present in the alignment, some aligned molecules may not fit into their binding site in the aligned conformation. Some molecules could penetrate into protein. Thus, there is no guarantee that the newly designed leads based on such a model can bind receptor well and interact effectively with the receptor, even if the QSAR model is very good in terms of good correlation

with experimental activity. So such an approach does not always lead in the right direction. On the other hand, docking alignment introduces the binding site feature into the QSAR model. This ensures that all calculated variables (property descriptors) which will be used to define the model are based on real poses which can bind to the receptor. By combining receptor-based docking simulation with ligand-based CoMFA (CoMSIA) QSAR, the enhanced approach will have more chances to find right leads on computer-aided drug design.

CONCLUSIONS

The validation results of two docking simulations have shown that the docking protocol can reproduce the two crystal structures of two complexes (IREV and 1COU) well with RMSDs of ~ 1 Å and ~ 0.6 Å for IREV and 1COU, respectively. The docking results for 50 TIBO derivatives show that they bind in the RT in the same direction and similar positions. They can be grouped into two as the docked structures distinguished by the 8-substituent, if it is hydrogen or not.

The satisfactory 3D-QSAR models of 50 TIBO derivatives have been constructed using the CoMFA and CoMSIA methods based on the docking conformation determination and two molecular alignments. They demonstrate that flexible docking is a good method to determine the "active" conformation of molecules and align them according to the active site feature for 3D-QSAR analyses.

The CoMFA QSAR models show that the steric part and the electrostatic part contribute equally to the activity. The CoMSIA QSAR models show that the largest contribution comes from the hydrophobic part. This was supported by our previous docking results that hydrophobic and hydrophilic interactions are important for NNRTI binding in RT active site.

The steric and electrostatic contours from CoMFA (CoMSIA) provide some useful insight into the design of inhibitors. By comparing these predictions from the CoMFA (CoMSIA) contours with the bound complex of RT/8-Cl-TIBO, we can see that part of the prediction is consistent with the characteristics of the inhibitor and RT. This suggests that to verify the ligand-based contours consistent with the binding site feature of their receptor is important. The verified information about electrostatic and steric preferable areas will serve better with further ligand development. It increases the probability of getting a right direction by getting more useful information and eliminating the false information which could result in a ligand which cannot bind and interact well with its receptor.

A ligand-based approach is used in rational drug design to build activity models, which provide important information on possible improvements in ligand structure to increase activity. Many studies have demonstrated that the methods can produce successful results. As it does not need receptor information and does not have a sense about the receptor feature, a conventional CoMFA (CoMSIA) QSAR cannot guarantee that its findings can bind well to the receptor and will create successful ligands. Meanwhile receptor-based modeling provides an insight into the interaction model of a ligand in its receptor.

Docking is based on the active site feature of a receptor by evaluating the interaction between ligands and receptor.

This interaction model is essential to understand the ligand interaction with its receptor and to help to find and improve successful candidates in rational drug design. If a correct binding model is gotten from the docking simulation and its result is integrated into ligand-based drug design, it will enhance the drug design approach and give a better chance to find leads and improve activity. Although both approaches provide a powerful means in drug design, the combination of both will be a better approach in rational drug design. The combination of flexible docking with CoMFA is an attractive way to construct 3D-QSAR models, because this approach integrates the receptor feature into ligand-based CoMFA (CoMSIA) QSAR by finding conformation and aligning molecules according to the binding site feature. This combination decreases the possibility to lead candidates which are not suitable for binding into the receptor, which very possibly occurs in a conventional template-alignment approach. The study intends to explore a way and demonstrates a successful case to combine a receptor-based drug design method—docking with ligand-based drug design methods—CoMFA (CoMSIA) QSAR.

ACKNOWLEDGMENT

The authors would like to thank Dr. Aleem Gangjee for allowing access to the computational resources in his group and Mr. Xing Lin for his help. This research was supported in part by a grant from the National Computational Science Alliance (MCB990008Nr00).

REFERENCES AND NOTES

- (1) Jacobo-Molina, A.; Arnold, E. HIV reverse transcriptase structure—function relationships. *Biochemistry* **1991**, *30*, 6351–6361.
- (2) Le Grice, S. F. J. Human immunodeficiency virus reverse transcriptase. *Reverse transcriptase*; Cold Spring Harbor Laboratory Press: Plainview, NY, 1993; pp 163–191.
- (3) Goff, P. Retroviral reverse transcriptase: Synthesis, structure and function. *JAIDS, J. Acquired Immune Defic. Syndr.* **1990**, *3*, 817–831.
- (4) Whitcomb, J. M.; Hughes, S. H. Retroviral reverse transcription and integration: Progress and problems. *Annu. Rev. Cell Biol.* **1992**, *8*, 275–306.
- (5) Arnold, E.; Das, K.; Ding, J.; Yadav, P.; Hsiou, Y. et al. Targeting HIV reverse transcriptase for anti-AIDS drug design: structural and biological considerations for chemotherapeutic strategies. *Drug Design Discov.* **1996**, *13*, 29–47.
- (6) Larder, B. A. Inhibitors of HIV reverse transcriptase as antiviral agents and drug resistance. *Reverse Transcriptase*; Cold Spring Harbor Laboratory Press: Plainview, NY, 1993; pp 163–191.
- (7) Tantillo, C.; Ding, J.; Jacobo-Molina, A.; Nanni, R. G.; Boyer, P. L. et al. Locations of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase: implications for mechanisms of drug inhibition and resistance. *J. Mol. Biol.* **1994**, *243*, 369–387.
- (8) De Clercq, E. Antiviral therapy for human immunodeficiency virus infections. *Clin. Microbiol. Rev.* **1995**, *8*, 200–239.
- (9) Ding, J.; Das, K.; Hsiou, Y.; Zhang, W.; Arnold, E. et al. Structural Studies of HIV-1 Reverse Transcriptase and Implications for Drug Design. *Structure-Based Drug Des.* **1997**, 41–82.
- (10) Pedersen, O. S.; Pedersen, E. B. nonnucleosid reverse transcriptase inhibitors. *Antiviral Chem. Chemother.* **1999**, *10*, 285–314.
- (11) Koup, R. A.; Merluzzi, V. J.; Hargrave, K. D.; Adams, J.; Grozinger, K. et al. Inhibition of human immunodeficiency virus type 1 replication by the dipyrrocliazepinone BI-RG-587. *J. Infect. Dis.* **1991**, *163*, 966–970.
- (12) Richman, D.; Rosenthal, A. S.; Shoog, M.; Eckner, R. J.; Chou, T. C. et al. BI-RG-587 in active against zidovudine-pesant human immunodeficiency virus type 1 and synergistic with zidovudine. *Antimicrob. Agents Chemother.* **1991**, *35*, 305–308.
- (13) Das, K.; Ding, J.; Hsiou, Y.; Clark, A. J.; Moereels, H. et al. Crystal structures of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1

- RT and 8-Cl TIBO complexed with the Tyr181Cys HIV-1 RT drug-resistant mutant. *J. Mol. Biol.* **1996**, *264*, 1085–1100.
- (14) Ren, J.; Esnouf, R.; Hopkins, A.; Ross, C.; Jones, Y. et al. The structure of HIV-1 reverse transcriptase complexed with 9-chloro-TIBO: lessons for inhibitor design. *Structure* **1995**, *3*, 915–926.
 - (15) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737.
 - (16) Jain, A. N. Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. *J. Med. Chem.* **2003**, *46*, 499–511.
 - (17) Johnson, M. A.; Hoeoeg, C.; Pinto, B. M. A Novel Modeling Protocol for Protein Receptors Guided by Bound-Ligand Conformation. *Biochemistry* **2003**, *42*, 1842–1853.
 - (18) Zhou, Z.; Fisher, D.; Spidel, J.; Greenfield, J.; Patson, B. et al. Kinetic and Docking Studies of the Interaction of Quinones with the Quinone Reductase Active Site. *Biochemistry* **2003**, *42*, 1985–1994.
 - (19) Wu, X.; Milne, J. L. S.; Borgnia, M. J.; Rostapshov, A. V.; Subramaniam, S. et al. A core-weighted fitting method for docking atomic structures into low-resolution maps: Application to cryo-electron microscopy. *J. Struct. Biol.* **2003**, *141*, 63–76.
 - (20) Todorov, N. P.; Mancera, R. L.; Monthoux, P. H. A new quantum stochastic tunnelling optimisation method for protein–ligand docking. *Chem. Phys. Lett.* **2003**, *369*, 257–263.
 - (21) Wang, L.; Merz, A. J.; Collins, K. M.; Wickner, W. Hierarchy of protein assembly at the vertex ring domain for yeast vacuole docking and fusion. *J. Cell Biol.* **2003**, *160*, 365–374.
 - (22) Vicker, N.; Ho, Y.; Robinson, J.; Woo, L. L. W.; Purohit, A. et al. Docking studies of sulphamate inhibitors of estrone sulphatase in human carbonic anhydrase II. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 863–865.
 - (23) Sabnis, Y. A.; Desai, P. V.; Rosenthal, P. J.; Avery, M. A. Probing the structure of falcipain-3, a cysteine protease from *Plasmodium falciparum*: Comparative protein modeling and docking studies. *Protein Sci.* **2003**, *12*, 501–509.
 - (24) Cramer, R. D. I.; 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.
 - (25) Klebe, 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.
 - (26) TBohm, M.; Sturzebecher, J.; Klebe, 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.
 - (27) Buolamwini, J. K.; Assefa, H. 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. *J. Med. Chem.* **2002**, *45*, 841–852.
 - (28) Nair, A. C.; Jayatilke, P.; Wang, X.; Miertus, S.; Welsh, W. J. 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. *J. Med. Chem.* **2002**, *45*, 973–983.
 - (29) Pungpo, P.; Hannongbua, S. Three-dimensional quantitative structure–activity relationships study on HIV-1 reverse transcriptase inhibitors in the class of dipyrindodiazepinone derivatives, using comparative molecular field analysis. *J. Mol. Graphics Modell.* **2000**, *18*, 581–590.
 - (30) Jayatilke, 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 CoMFA Models. *J. Med. Chem.* **2000**, *43*, 4446–4451.
 - (31) Barreca, M. L.; Carotti, A.; Carrieri, A.; Chimirri, A.; Monforte, A. M. et al. Comparative molecular field analysis (CoMFA) and docking studies of nonnucleoside HIV-1 RT inhibitors (NNIs). *Bioorg. Med. Chem.* **1999**, *7*, 2283–2292.
 - (32) 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.
 - (33) 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.
 - (34) Raghavan, K.; Buolamwini, J. K.; Fesen, M. R.; Pommier, Y.; Kohn, K. W. et al. 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.
 - (35) Oprea, T. I.; Waller, C. L.; Marshall, G. R. 3D-QSAR of human immunodeficiency virus (H) protease inhibitors. III. Interpretation of CoMFA results. *Drug Des. Discovery* **1994**, *12*, 29–51.
 - (36) Debnath, A. K.; Jiang, S.; Strick, N.; Lin, K.; Haberfield, P. et al. Three-Dimensional Structure–Activity Analysis of a Series of Porphyrin Derivatives with Anti-HIV-1 Activity Targeted to the V3 Loop of the gp120 Envelope Glycoprotein of the Human Immunodeficiency Virus Type 1. *J. Med. Chem.* **1994**, *37*, 1099–1108.
 - (37) Zhou, Z.; Madrid, M.; Madura, J. D. New Nonnucleoside Inhibitors in HIV-1 RT Docking Calculations. *Proteins: Struct., Funct., Genet.* **2002**, *49*, 529–542.
 - (38) Hannongbua, S.; Pungpo, P.; Limtrakul, J.; Wolschann, P. Quantitative structure–activity relationships and comparative molecular field analysis of TIBO derivatised HIV-1 reverse transcriptase inhibitors. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 563–577.
 - (39) Huuskonen, J. QSAR Modeling with the Electrotopological State: TIBO Derivatives. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 425–429.
 - (40) Garg, R.; Gupta, S. P.; Gao, H.; Babu, M. S.; Debnath, A. K. et al. Comparative Quantitative Structure–Activity Relationship Studies on Anti-HIV Drugs. *Chem. Rev.* **1999**, *99*, 3525–3602.
 - (41) Smith, M. B. K.; Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L.; Michejda, C. J. et al. Monte Carlo calculations on HIV-1 reverse transcriptase complexed with the nonnucleoside inhibitor 8-Cl TIBO: contribution of the L100I and Y181C variants to protein stability and biological activity. *Protein Eng.* **2000**, *13*, 413–421.
 - (42) Smith, R. H., Jr.; Jorgensen, W. L.; Tirado-Rives, J.; Lamb, M. L.; Janssen, P. A. J. et al. Prediction of Binding Affinities for TIBO Inhibitors of HIV-1 Reverse Transcriptase Using Monte Carlo Simulations in a Linear Response Method. *J. Med. Chem.* **1998**, *41*, 5272–5286.
 - (43) Cui, M.; Huang, X.; Luo, X.; Briggs, J. M.; Ji, R. et al. Molecular Docking and 3D-QSAR Studies on Gag Peptide Analogue Inhibitors Interacting with Human Cyclophilin A. *J. Med. Chem.* **2002**, *45*, 5249–5259.
 - (44) Buolamwini, J. K.; Assefa, H. 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. *J. Med. Chem.* **2002**, *45*, 841–852.
 - (45) Muegge, I.; Podlogar, B. L. 3D-quantitative structure activity relationships of biphenyl carboxylic acid MMP-3 inhibitors: exploring automated docking as alignment method. *Quant. Struct.–Act. Relat.* **2001**, *20*, 215–222.
 - (46) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
 - (47) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C. et al. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765.
 - (48) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity – A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219.
 - (49) Kukla, M. J.; Breslin, H. J.; Pauwels, R.; Fedde, C. L.; Miranda, M. et al. Synthesis and anti-HIV-1 activity of 4,5,6,7-tetrahydro-5-methylimidazo[4,5,1-jk][1,4]benzodiazepin-2(1H)-one (TIBO) derivatives. *J. Med. Chem.* **1991**, *34*, 746–751.
 - (50) Kukla, M. J.; Breslin, H. J.; Diamond, C. J.; Grous, P. P.; Ho, C. Y. et al. Synthesis and anti-HIV-1 activity of 4,5,6,7-tetrahydro-5-methylimidazo[4,5,1-jk][1,4]benzodiazepin-2(1H)-one (TIBO) derivatives. 2. *J. Med. Chem.* **1991**, *34*, 3187–3197.
 - (51) Breslin, H. J.; Kukla, M. J.; Ludovici, D. W.; Mohrbacher, R.; Ho, W. et al. Synthesis and Anti-HIV-1 Activity of 4,5,6,7-Tetrahydro-5-methylimidazo[4,5,1-jk][1,4]benzodiazepin-2(1H)-one (TIBO) Derivatives. 3. *J. Med. Chem.* **1995**, *38*, 771–793.
 - (52) Ho, W.; Kukla, M. J.; Breslin, H. J.; Ludovici, D. W.; Grous, P. P. et al. Synthesis and Anti-HIV-1 Activity of 4,5,6,7-Tetrahydro-5-methylimidazo[4,5,1-jk][1,4]benzodiazepin-2(1H)-one (TIBO) Derivatives. 4. *J. Med. Chem.* **1995**, *38*, 794–802.
 - (53) Breslin, H. J.; Kukla, M. J.; Kromis, T.; Cullis, H.; De Knaep, F. et al. Synthesis and anti-HIV activity of 1,3,4,5-tetrahydro-2H-1,4-benzodiazepin-2-one (TBO) derivatives. Truncated 4,5,6,7-tetrahydro-5-methylimidazo[4,5,1-jk][1,4]benzodiazepin-2(1H)-ones (TIBO) analogues. *Bioorg. Med. Chem.* **1999**, *7*, 2427–2436.

- (54) Solis, F. J.; EWets, R. J.-B. Minimization by random search techniques. *Math. Oper. Res.* **1981**, 6, 19–30.
- (55) L'Ecuyer, P.; Cote, S. Implementing a random number package with splitting facilities. *ACM Trans. Math. Software* **1991**, 17, 98–111.
- (56) Ren, J.; Esnouf, R. M.; Hopkins, A. L.; Stuart, D. I.; Stammers, D. K. Crystallographic analysis of the binding modes of Thiazoloisoindolindolinone nonnucleoside inhibitors of HIV-1 Reverse Transcriptase and comparison with modeling studies. *J. Med. Chem.* **1999**, 42, 3845–3851.

dolindolinone nonnucleoside inhibitors of HIV-1 Reverse Transcriptase and comparison with modeling studies. *J. Med. Chem.* **1999**, 42, 3845–3851.

CI049893V