Docking and 3-D QSAR Studies on Indolyl Aryl Sulfones. Binding Mode Exploration at the HIV-1 Reverse Transcriptase Non-Nucleoside Binding Site and Design of Highly Active *N*-(2-Hydroxyethyl)carboxamide and *N*-(2-Hydroxyethyl)carbohydrazide Derivatives

Rino Ragno,[†] Marino Artico,[‡] Gabriella De Martino,[‡] Giuseppe La Regina,[‡] Antonio Coluccia,[‡] Alessandra Di Pasquali,[‡] and Romano Silvestri*,[‡]

Istituto Pasteur−Fondazione Cenci Bolognetti, Dipartimento di Studi Farmaceutici, and Dipartimento di Studi di Chimica e Tecnologia delle Sostanze Biologicamente Attive, Università di Roma "La Sapienza", Piazzale Aldo Moro 5, I-00185 Roma, Italy
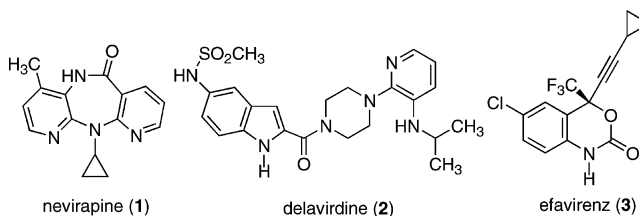
Three-dimensional quantitative structure−activity relationship (3-D QSAR) studies and docking simulations were developed on indolyl aryl sulfones (IASs), a class of novel HIV-1 non-nucleoside reverse transcriptase (RT) inhibitors (Silvestri, et al. *J. Med. Chem.* **2003**, *46*, 2482−2493) highly active against wild type and some clinically relevant resistant strains (Y181C, the double mutant K103N−Y181C, and the K103R−V179D−P225H strain, highly resistant to efavirenz). Predictive 3-D QSAR models using the combination of GRID and GOLPE programs were obtained using a receptor-based alignment by means of docking IASs into the non-nucleoside binding site (NNBS) of RT. The derived 3-D QSAR models showed conventional correlation (r^2) and cross-validated (q^2) coefficients values ranging from 0.79 to 0.93 and from 0.59 to 0.84, respectively. All described models were validated by an external test set compiled from previously reported pyrryl aryl sulfones (Artico, et al. *J. Med. Chem.* **1996**, *39*, 522−530). The most predictive 3-D QSAR model was then used to predict the activity of novel untested IASs. The synthesis of six designed derivatives (prediction set) allowed disclosure of new IASs endowed with high anti-HIV-1 activities.

Introduction

Human immunodeficiency virus (HIV) is the causative agent of acquired immuno deficiency syndrome (AIDS)/infection characterized by loss of helper T lymphocytes and heavy damage of lymphatic tissues. Global estimates of WHO/UNAIDS showed 40 million people infected with HIV/AIDS at the end of 2003, with 5 million newly infected and 3 million deaths. Every day in 2003 an estimated 14000 people were newly infected with HIV.[1]

Anti-AIDS therapy is actually based on three classes of anti-HIV drugs, the nucleoside reverse transcriptase inhibitors (NRTIs), the non-nucleoside reverse transcriptase inhibitors (NNRTIs), and the protease inhibitors (PIs). More recently enfuvirtide, a 36-amino acid residue peptide acting as viral entry inhibitor, has been licensed for the treatment of HIV infection.[2] NRTIs, NNRTIs, and PIs are mixed in the highly active antiretroviral therapy (HAART), which dramatically reduces the incidence of AIDS infection and death. HAART regimens slow the viral replication to very low levels, but they are unable to eradicate the viral infection.[3] The needed long-term or permanent use of anti-AIDS drugs induces the selection of drug-resistant viral variants and the emergence of unwanted metabolic side effects.[3]

* Corresponding author. Phone +39 06 4991 3800. Fax +39 06 491 491; E-mail: romano.silvestri@uniroma1.it.
‡ Istituto Pasteur−Fondazione Cenci Bolognetti, Dipartimento di Studi Farmaceutici.
† Dipartimento di Studi di Chimica e Tecnologia delle Sostanze Biologicamente Attive.

Chart 1

nevirapine (**1**)   delavirdine (**2**)   efavirenz (**3**)

Moreover, similarities in chemical structures and/or mechanisms of action often lead to the emergence of cross-resistance among members of the same class.[4] Rapid development of drug resistance and toxicity problems make urgent the discovery of novel anti-HIV agents effective against resistant mutants and deprived of unpleasant side effects. Actually, the search for novel anti-HIV drugs is pursued either by improvement of the existing the drug classes (NRTIs, NNRTIs, and PIs) or by discovery of new agents targeting different mechanisms of action (integrase, Rnase-H, and viral entry inhibitors).[5]

The NNRTIs are characterized by different unrelated chemical structures. The marketed drugs nevirapine (**1**), delavirdine (**2**), and efavirenz (**3**) are significant examples of such inhibitors (Chart 1). At the molecular level NNRTIs act by a specific allosteric effect arising from noncompetitive binding to a hydrophobic pocket, the non-nucleoside binding site (NNBS), located near the catalytic site.[6,7] Despite their low toxicity, the use of first-generation NNRTIs in multidrugs anti-AIDS cocktails led to unsuccessful results, due to the rapid

Chart 2

L-737,126 (**4**)   **5a**   **5b**   **5c**

Chart 3

6 (PAS)   **30, 31, 38** (IASs)

30: R$_1$ = CH$_3$, R$_2$, R$_3$ = H;
31: R$_1$, R$_3$ =H, R$_2$ = CH$_3$;
38: R$_1$ = H, R$_2$, R$_3$ = CH$_3$.

Chart 4

L-737,126 (**4**)
highly activity against wt HIV-1
low activity against resistant mutants

chemical modulation

IAS (**38**)
high activity against Y181C,
K103N-Y181C and EFV^R

chemical modulation

5a: improved activity
against K103N RT

3D-QSAR Docking Studies

Y = HN⌒OH   **84-95**

Y = HN⌒N⌒OH   **96-107**

X = S, SO$_2$
R = H, alkyl, Cl, F, NO$_2$, NH$_2$

Figure 1. Graphical result of the dockings of reference compound **4** into 14 NNBSs. Conformations docked in the different NNBS are color coded; in the legend are reported the PDB entry codes of the used RTs. A sketch of the NNBS subpockets is also reported.
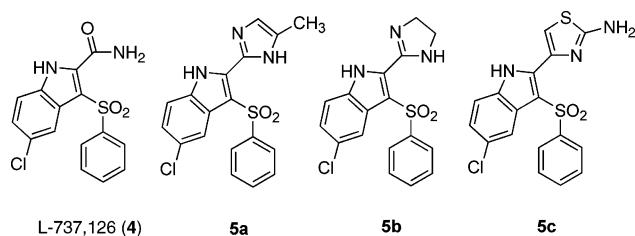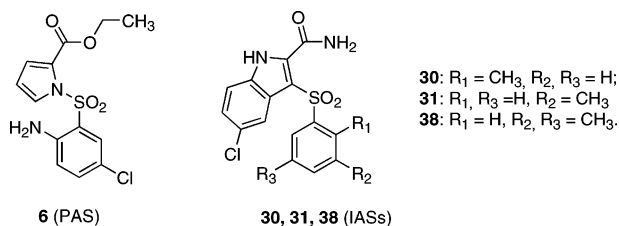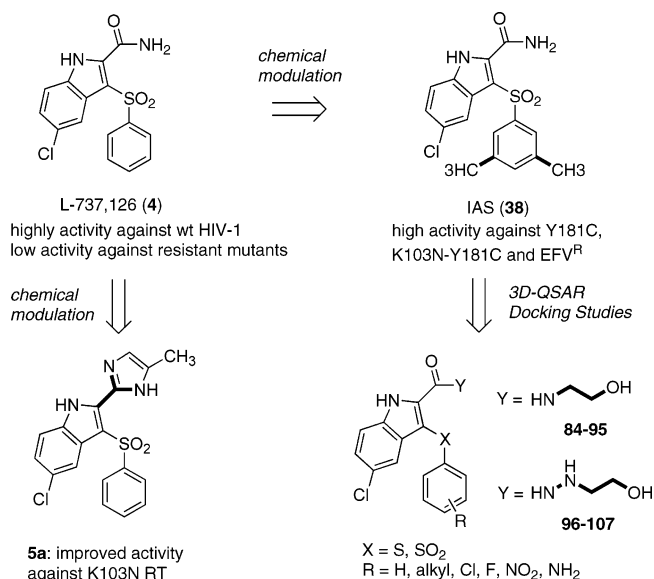
emergence of drug resistance.[8,9] As a consequence, second-generation NNRTIs having good activity against wild-type (wt) reverse transcriptase (RT) and the most prevalent mutant viral strains are nowadays needed. Continuous efforts in this field are documented by the wide number of NNRTIs described in the literature,[8-12] some of which are under clinical trials.[5]

Among them, L-737,126 (**4**), a benzenesulfonylindolcarboxamide endowed with potent antiviral activity and high selectivity, has been developed by Merck A. G. Despite its highest activity, L-737,126 was not suitable for clinical trials, because of its low bioavailability rising from its scarce solubility in water.[13] Merck Research Laboratories attempted to overcome this problem by replacing the 2-carboxamide function with some heterocycles (for example compounds **5a**−**5c**), but the new compounds increased only marginally the water solubility of **4**.[14] Nevertheless, some of them were found to be active against resistant mutants, with compound **5a** being 11 times more active than **4** against the K103N mutant (Chart 2). On the basis of these results, the search for L-737,126 analogues was abandoned by Merck.
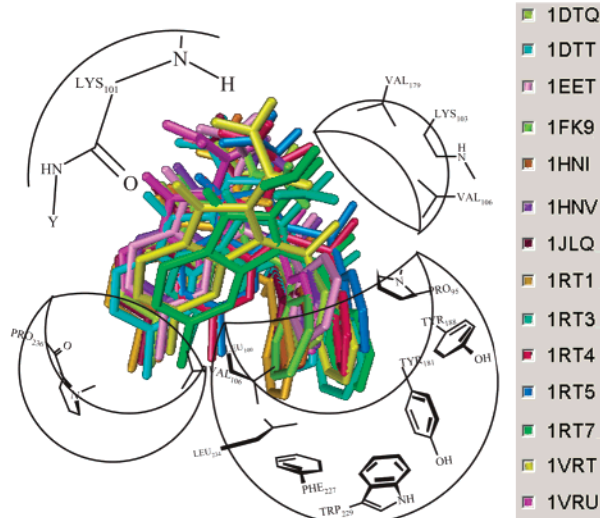
During extensive structure activity relationship (SAR) studies on diaryl sulfones, we first identified pyrryl sulfones (PASs, for example compound **6**)[15-17] as highly potent NNRTIs and then we extended our research to novel indolyl aryl sulfones (IASs compounds **30**, **31**, and **38**) related to L-737,126 (Chart 3). In particular, indole derivatives bearing 2-methylphenylsulfonyl (**30**) or 3-methylphenylsulfonyl (**31**) moieties were found to inhibit HIV-1 (EC$_{50}$s) at nanomolar concentrations. Furthermore, the introduction of a 3,5-dimethylphenylsulfonyl moiety led to compound **38**, displaying high activity and selectivity not only against the wt strain but also against the Y181C and K103N−Y181C viral variants and the efavirenz resistant (EFV^R) mutant K103R−V179D−P225H.[18]

Pursuing this research line, we have undertaken docking and 3-D QSAR studies on a training set (compounds **7**−**75**) of previously reported IASs.[18] To that end, Merck's findings[14] suggested that little chemical modification of the 2-carboxyamido moiety could lead to new IAS derivatives with improved antiviral activity against both wild type and mutant strains of HIV-1. In this paper we report the development of 3-D QSAR models by means of the programs Autodock, GRID, and GOLPE. The most predictive 3-D QSAR model was then used to design novel IAS derivatives containing a 2-hydroxyethylaminocarbonyl or a 2-hydroxyethylhydrazinocarbonyl moiety at position 2 of the indole ring (Chart 4). 2-Hydroxyethyl and 2-hydroxyethylamino moieties were chosen with the idea that introduction of chain elongators containing H-bond acceptors to the nitrogen of carboxamide group of indoles **4** and **38** would be beneficial for increasing the interaction between the new designed inhibitors and RT.

Docking and 3-D QSAR Studies

Docking Studies and Binding Mode Analysis of Lead Compound 4. The binding mode of IAS derivatives was investigated by means of docking studies into the HIV-1 NNBS using compound **4** as a representative member. To take into account the experimental NNBS

Table 1. Structures and Anti-HIV-1 Activities of Compounds **4** and **7−75**,[18] the Training Set Used for the 3-D QSARs
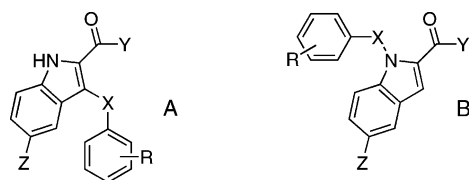
compd	structure	X	Y	Z	R	wt$_{IIIB}$ EC$_{50}$ (μM)
4	A	SO$_2$	NH$_2$	Cl	H	0.001
7	A	S	OEt	H	H	1.4
8	A	S	OEt	H	2-NH$_2$	>200
9	A	S	OEt	H	2-NH$_2$-5-Cl	≥200
10	A	S	OEt	Cl	2-NH$_2$	2.3
11	A	S	OEt	Cl	2-NH$_2$-5-Cl	2.5
12	A	SO$_2$	OEt	H	H	3.7
13	A	SO$_2$	OEt	Cl	2-NH$_2$	>200
14	A	SO$_2$	OEt	H	2-NH$_2$-5-Cl	2.5
15	A	SO$_2$	OEt	Cl	2-NH$_2$-5-Cl	1.9
16	A	S	NH$_2$	H	H	1.4
17	A	S	NH$_2$	H	2-NH$_2$-5-Cl	9
18	A	S	NH$_2$	Cl	H	0.02
19	A	S	NH$_2$	Cl	2-Me	0.3
20	A	S	NH$_2$	Cl	4-Me	>0.45
21	A	S	NH$_2$	Cl	4-F	1.4
22	A	S	NH$_2$	Cl	4-Cl	3.1
23	A	S	NH$_2$	Cl	4-iso-Pr	1.9
24	A	S	NH$_2$	Cl	4-tert-Bu	8
25	A	S	NH$_2$	Cl	3,5-Me2	0.006
26	A	S	NH$_2$	Cl	2,6-Cl2	1.2
27	A	S	NH$_2$	Cl	2-NH$_2$-5-Cl	1.6
28	A	SO$_2$	NH$_2$	H	H	0.18
29	A	SO$_2$	NH$_2$	H	2-NH$_2$-5-Cl	0.3
30	A	SO$_2$	NH$_2$	Cl	2-Me	0.001
31	A	SO$_2$	NH$_2$	Cl	3-Me	0.001
32	A	SO$_2$	NH$_2$	Cl	4-Me	0.003
33	A	SO$_2$	NH$_2$	Cl	4-F	0.014
34	A	SO$_2$	NH$_2$	Cl	4-Cl	0.011
35	A	SO$_2$	NH$_2$	Cl	4-iso-Pr	0.08
36	A	SO$_2$	NH$_2$	Cl	4-tert-Bu	0.13
37	A	SO$_2$	NH$_2$	Cl	2,4-Me$_2$	0.004
38	A	SO$_2$	NH$_2$	Cl	3,5-Me$_2$	0.004
39	A	SO$_2$	NH$_2$	Cl	2,6-Cl$_2$	0.1
40	A	SO$_2$	NH$_2$	Cl	2-NH$_2$-5-Cl	0.04
41	A	SO$_2$	NH$_2$	Br	3,5-Me$_2$	0.002
42	A	SO$_2$	NH$_2$	COMe	3,5-Me$_2$	0.015
43	A	SO$_2$	NH$_2$	CH(OH)Me	3,5-Me$_2$	0.025
44	A	S	NHNH$_2$	Cl	H	0.55
45	A	S	NHNH$_2$	Cl	4-Me	1.5
46	A	S	NHNH$_2$	Cl	4-F	5
47	A	S	NHNH$_2$	Cl	4-Cl	10
48	A	S	NHNH$_2$	Cl	2-NH$_2$-5-Cl	>13
49	A	SO$_2$	NHNH$_2$	H	H	0.53
50	A	SO$_2$	NHNH$_2$	Cl	H	0.01
51	A	SO$_2$	NHNH$_2$	Cl	4-Me	0.05
52	A	SO$_2$	NHNH$_2$	Cl	4-F	0.32
53	A	SO$_2$	NHNH$_2$	Cl	4-Cl	0.19
54	A	SO$_2$	NHNH$_2$	Cl	3,5-Me$_2$	0.13
55	A	SO$_2$	NHNH$_2$	Cl	2-NH$_2$-5-Cl	0.3
56	B	SO$_2$	H	H	H	>150
57	B	SO$_2$	COOEt	H	4-Cl	>100
58	B	SO$_2$	COOEt	H	2-NO$_2$	1.8
59	B	SO$_2$	COOEt	H	2-NO$_2$-5-Cl	>100
60	B	SO$_2$	COOEt	H	2-NH$_2$-5-Cl	1.8
61	B	SO$_2$	H	3-COOEt	H	>32
62	B	SO$_2$	H	3-COO-i-Pr	H	>54
63	B	SO$_2$	H	5-Cl	H	>75
64	B	SO$_2$	COOEt	5-Cl	H	>200
65	B	SO$_2$	COOEt	5-Cl	4-Me	>200
66	B	SO$_2$	COOEt	5-Cl	4-Cl	>200
67	B	SO$_2$	COOEt	5-Cl	2-NO$_2$-5-Cl	>31
68	B	SO$_2$	COOEt	5-Cl	2-NH$_2$-5-Cl	8.3
69	B	SO$_2$	H	5-Cl-3-COOEt	H	>42
70	B	SO$_2$	H	5-Cl-3-COO-i-Pr	H	>200
71	B	SO$_2$	CONH$_2$	H	H	15
72	B	SO$_2$	CONHNH$_2$	H	H	>200
73	B	SO$_2$	CONH$_2$	5-Cl	H	66.6
74	B	SO$_2$	H	3-CONH$_2$	H	>32
75	B	SO$_2$	H	5-Cl-3-CONH$_2$	H	≥200

^{a} Compound concentration (μM) required to achieve 50% protection of infected MT-4 cells from wt$_{IIIB}$-HIV-1-induced cytopathogenicity (MTT method). Data represent mean values for three separate experiments. Variation among triplicate samples was less than 15%.
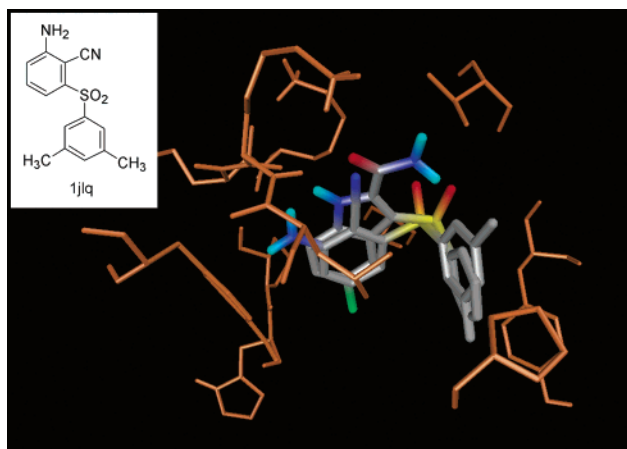
Figure 2. Structure of 739W94 (left). Superimposition of the experimental conformation found in the 739W94/RT complex with the docked conformation of **4** (right).

Figure 3. Training set molecules as in the best cluster alignment.

Table 2. Statistical Results of the 3D QSAR Models

model	alignment	grid probes	vars[a]	PC[b]	r^2	q^2	SDEP$_{CV}$
1	best cluster	C1=	665	3	0.88	0.72	0.88
2	best docked	C1=	683	3	0.93	0.84	0.69
3	best cluster	OH2	752	2	0.85	0.68	0.95
4	best docked	OH2	669	3	0.92	0.81	0.74
5	best cluster	C1= and OH2	1140	2	0.79	0.59	1.07
6	best docked	C1= and OH2	709	3	0.91	0.79	0.76

[a] Number of selected variables. [b] Number of principal components which showed the maximum q^2 value.

flexibility, the structural data of 14 RTs (pdb codes 1dtq,[19] 1dtt,[19] 1eet,[20] 1fk9,[21] 1hni,[22] 1hnv,[23] 1jlq,[24] 1rt1,[25] 1rt2,[25] 1rt3,[26] 1rt4,[27] 1rt5,[28] 1rt7,[28] 1vrt,[29] and 1vru[19,23]) were used to explore the binding mode of **4**. The program Autodock[30] was used for the docking of **4** into the 14 RTs. In all cases the docked conformations (either the first ranked or the most populated cluster representing conformations) were in good agreement with each other, identifying a unique binding mode of **4** for all the used RTs (Figure 1). The only chemical feature displaying some uncertainty was the carboxamide function in position 2 of the indole ring, which in some cases showed the amide carbonyl in the cis position with the indole NH, while in other cases the carboxamide moiety was rotated 180°. Apart from the carboxamide group, the other chemical features of **4** shared a common binding mode, namely, (i) the indole NH makes a hydrogen bond with the Lys101 carbonyl; (ii) the phenyl ring of the benzenesulfonyl moiety occupies a hydrophobic aromatic-rich pocket formed mainly by the side chains of Tyr181, Tyr188, Phe227, and Trp229; (iii) the sulfonyl group fits in a little hydrophobic pocket made by the side chains of Val106, Lys103 (only α and β CH₂), and Val179; and (iv) the 5-chlorine atom of indole makes favorable contacts with Pro236 (left side of Figure 1). To validate the docking procedure, parallel docking and cross-docking studies were conducted on all the complexes of the ligands cocrystallized within the NNBS of the 14 RTs. Autodock was able to reproduce with a minimal error the experimental binding mode of all the ligands.[31]

GRID/GOLPE 3-D QSAR of IASs. Training Set. A training set of 70 IASs derivatives[18] was used to develop the 3-D QSAR (Table 1). The structures of the training set were modeled starting from the closely structurally related diaryl sulfone 739W94[24] extracted from the corresponding complex with the HIV-1 RT (pdb code 1jlq) (Figure 2).

Alignment Rules. Receptor-based alignment rules[32,33] were obtained by means of docking experiments on the training set into the NNBS extracted from the 739W94/ RT complex. In several cases, different binding modes were found among the conformations suggested by the Autodock program.[34] Since the molecular alignment of the training set is a crucial step in a 3-D QSAR, two
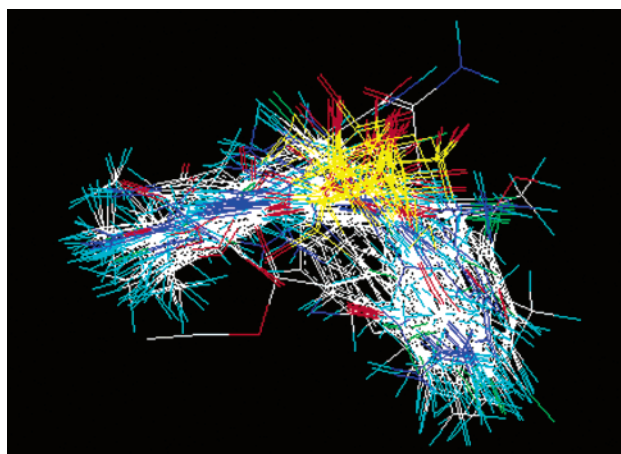
different alignment rules were used: the lowest energy docked conformations of the most populated clusters (named "best cluster alignment", Figure 3) and the first ranked docked conformations as obtained from the Autodock run (named "best docked alignment").

Calculation of Molecular Interaction Fields (MIFs). The MIFs for the 3-D QSAR studies were calculated by means of the GRID[35,36] program. The NNBS is characterized at least by two important spatial regions: (i) the hydrophobic aromatic-rich binding pocket formed mainly by residues Tyr181, Tyr188, Phe227, and Trp229 and (ii) the hydrogen-bonding site where most of the known RT inhibitors make at least one hydrogen bond with either the amide or carbonyl group of Lys101. To include these two binding site features into the 3-D QSAR studies, the aromatic carbon probe (grid format C1=) and the water probe (grid format OH2) were used to calculate the MIFs.
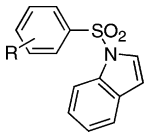
Definition of the 3-D QSAR Models. The program GOLPE[36,37] was used to define six 3-D QSAR models from two alignments and three combinations of MIFs (aromatic, water, and aromatic + water) (Table 2). The MIFs of the training set were imported in GOLPE along with the inverse of the logarithm of the experimental anti-HIV-1 activities (pEC₅₀s). It could be argued that cell-based activities would not be suitable for such a 3-D QSAR study, enzyme-based assays (IC₅₀s) being a more appropriate choice. Nevertheless, it has been proved that there is a good correlation between enzyme-based (IC₅₀) and cell-based (EC₅₀) assays,[38] and therefore, data derived from cell-based activities were expected not to affect the 3-D QSAR. Moreover, we aimed to obtain a 3-D QSAR model able to discriminate active versus inactive derivatives, the latter being a source of important structural information. Inactive compounds were

Chart 5. [a]

$$r^2 = 1 - \frac{\sum\limits_{i=1}^{n}(Y_i - \bar{Y})^2}{\sum\limits_{i=1}^{n}(Y_i - Y_{fit})^2} \qquad q^2 = 1 - \frac{\sum\limits_{i=1}^{n}(Y_i - \bar{Y})^2}{\sum\limits_{i=1}^{n}(Y_i - Y_{cv})^2} \qquad SDEP = \sqrt{\frac{\sum\limits_{i=1}^{n}(Y_i - Y_{cv})^2}{N}}$$

[a] Y_i = experimental value; Y_{fit} = recalculated value; $\bar{Y}$ = mean value; Y_{cv} = predicted value, and N = number of experiments.

Table 3. Test Set Predictions and SDEP Values Obtained with Models 1−6

| | | | pEC$_{50}$ | | | | | |
| | | | predicted[b] | | | | | |
compd	R	exp[a]	M1	M2	M3	M4	M5	M6
76	2-NO$_2$	5.30	5.37	5.87	4.42	4.59	4.77	4.08
77	2-NH$_2$	4.96	5.96	5.89	5.33	5.58	5.76	5.39
78	2-NO$_2$-5-Cl	5.40	5.11	3.26	4.42	3.68	4.33	4.32
79	2-NH$_2$-5-Cl	6.00	6.43	2.70	5.28	4.52	5.74	3.82
80	2-NO$_2$-4-Cl	4.80	3.43	5.58	4.19	5.18	3.71	4.61
81	2-NH$_2$-4-Cl	4.13	3.79	5.56	3.81	5.35	3.54	5.28
82	2-Cl-5-NO$_2$	5.22	5.19	5.35	4.68	5.50	5.12	4.53
83	2-Cl-5-NH$_2$	3.82	6.31	6.23	5.56	6.39	5.96	5.55
SDEP$_{Test-Set}$			1.09	1.31	0.88	1.29	1.08	1.25

[a] Experimental pEC$_{50}$ (pEC$_{50}$ = −log EC$_{50}$ (M)). [b] Predicted pEC$_{50}$: M1 = model 1, M2 = model 2, M3 = model 3, M4 = model 4, M5 = model 5, M6 = model 6.

thus included in the training set, to which was arbitrarily assigned an EC$_{50}$ value equal to 50% of the least active compound[39,40] (Table 1).

After data pretreatment (see Experimental Section), consecutive fractional factorial design (FFD) selections were conducted to refine the initial models. The FFD selections were continued until no further statistical improvements were observed. To measure the goodness of the model, the statistical indices r^2, q^2, and SDEP were employed. The six final 3-D QSAR models were characterized by correlation coefficient (r^2), predictive correlation coefficient (q^2), and cross-validated standard deviation of errors of prediction (SDEP$_{CV}$, Chart 5) values falling in the ranges 0.79−0.93, 0.59−0.84, and 0.69−1.07, respectively (Table 2). For both the FFD selections and the cross-validations, the group method was used, setting to five the number of groups to be used.

As the six 3-D QSAR models were endowed of similar statistical profiles, an external test set (compounds **76**−**83**) taken from a previously reported series of indolylaryl sulfones[16] was used to measure their predictive ability (Table 3). On the basis of the standard deviation error of prediction value on the test set (SDEP$_{Test-Set}$), the external validation proved the best cluster alignment/OH2 GRID probe combination (model 3, SDEP$_{Test-Set}$ = 0.88, Figure 4) as the most predictive. It was therefore selected for description of the 3-D QSAR maps and it was chosen as a tool for design and prediction of pEC$_{50}$s of new IAS derivatives. Apart from the optimal combination between the alignment and the probe choice, the higher predictive ability of model 3 can also be attributed to the fact that the model is defined with two principal components (PC) only. In fact, it

seems that the extraction of a further PC (models 1, 2, 4, and 5) enhances the self-consistency (cross-validation) of the model, but less predictive 3-D QSAR are obtained (Table 2).

Interpretation of Model 3. One important feature of 3-D QSARs is the graphical representation of the model, which makes its interpretation easier. In the GOLPE software, several options are available to display the final model. Among these, the PLS pseudo-coefficients and the activity contribution plots are very useful. The PLS pseudo-coefficients plot allows the visualization of the selected grid points at determined molecules/probe interaction energy level, indicating graphical information of the whole training set. However, in this plot the signs of the coefficients can induce errors, as coefficients have opposite meaning depending on the fact that the compound produces positive or negative field values in a specific area. On the other hand, the activity contribution plot gives the possibility to display spatial regions that are individually important for the selected molecule. The activity contribution plot, different for every molecule within the training set, results from multiplying the values of the coefficients by the actual values of the field for that molecule. In Figure 5 is reported the superimposition of the PLS coefficients plot with the activity contribution plots of the highly active compound **38** (EC$_{50}$ = 0.001 μM) and the poorly active **27** (EC$_{50}$ = 1.2 μM). Negative PLS coefficients, individualized in the red polyhedrons, indicate an increase or decrease of activity when favorable or unfavorable interaction occurs, respectively. On the contrary, blue polyhedrons (positive PLS coefficients) show areas with decreased or increased activity as a consequence of favorable or unfavorable interaction. Comparing the top view (highly active compound **38**) and the bottom view (low active compound **27**) of Figure 5, it can be seen that the two molecules show some differences in activity because of their dissimilar spatial arrangements of the substituted phenyl groups. The 3,5-dimethylphenyl moiety of **38** makes favorable interactions with either red (parallel to the benzene ring π cloud) or blue (positive interactions with methyl group) polyhedrons, while the 2-amino-5-chlorophenyl moiety of **27** lacks any favorable interaction. Another crucial difference is shown by the sulfone/sulfur bridge connecting indole with benzene. The sulfone group of **38** makes favorable interactions (blue polyhedrons), which are fully absent in the sulfur-bridged derivative **27**. Regarding the activity contribution plots, Figure 5 also shows for compound **38** a wider area of favorable interactions (yellow polyhedrons, top view) than for **27** (bottom view).

As a rule, the 3-D QSAR maps cannot be employed to extract information about ligand/receptor interactions, as the model has been created just with the training set structures. This implies any receptor model
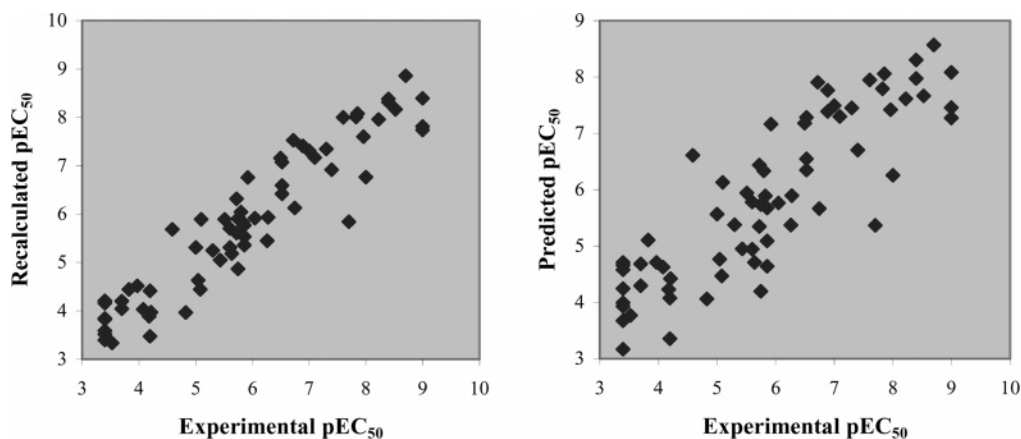
Figure 4. Fitting (left side) and cross-validation (right side) plots of the model 3 with two principal components.
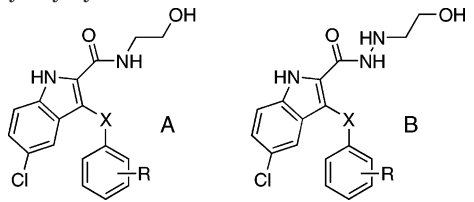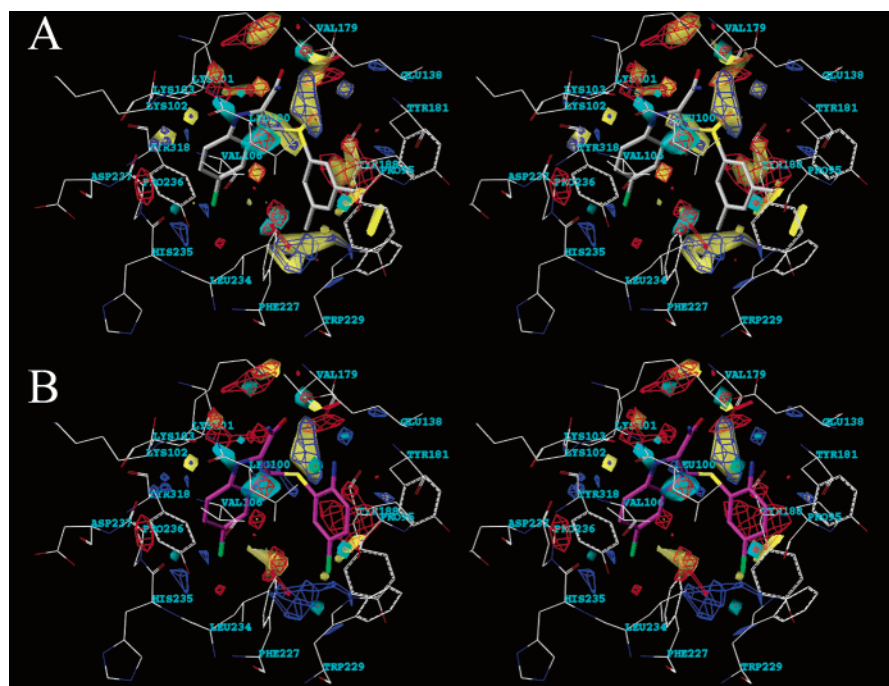

Figure 5. Stereoview of the PLS coefficients of the GRID plots. Polyhedrons in red are negative coefficients (data level = −0.008), polyhedrons in blue are positive coefficients (data level = +0.008). For interpretation are also reported activity contribution plots for derivatives **38** (Figure 5A, carbon atoms colored in white) and **27** (Figure 5B, carbon atoms colored in magenta). Positive contribution (data level = +0.035) to the activity is color coded as a yellow polyhedron. Negative contribution (data level = −0.035) to the activity is color coded as a cyan polyhedron. A 4 Å core of NNBS is also displayed for 3-D QSAR map/receptor matching. Hydrogen atoms are omitted for sake of clarity.

being extremely affected by lack in design, correlation in structural features, and alignments. Since the alignment of the IASs training set was based on docking studies, it would be useful to check for matching between the receptor pocket (NNBS) and the 3-D QSAR maps. In Figure 5 a 4 Å core of NNBS residues is overlapped with the 3-D PLS coefficients maps, and the close match between the NNBS residues and the PLS polyhedrons is evident. The biggest and thus most important polyhedrons are close to crucial residues of the NNBS: big red polyhedrons are situated near Lys101, Lys103, Val179, Tyr188, and Pro235 residues, likely identifying those NNBS regions more involved in electrostatic interactions with the bound ligand, while blue polyhedrons can be observed in the vicinity of Leu100, Val179, Tyr188, and Phe227, to which hydro-

phobic and steric ligand/receptor interactions could be mainly assessed.

Design of New IASs Derivatives, 3-D QSAR Predictions, and Binding Mode. On the basis of the structure−activity relationships of the training set, 24 new IAS derivatives bearing 2-hydroxyethylaminocarbonyl and 2-hydroxyethylhydrazinocarbonyl moieties at position 2 of the indole nucleus were modeled and receptor-based aligned. Application of the 3-D QSAR model 3 allowed the pEC_{50} prediction of the modeled derivatives (Table 4). Three compounds for each series (hydroxyethylamide and hydroxyethylhydrazide), for a total of six, were synthesized and tested for their anti-HIV-1 activity. Among predicted derivatives, the top predicted four (**94** and **95** for structure A and **104** and **107** for structure B) and the bottom predicted two (**84**

Table 4. pEC_{50} Predictions of 2-Hydroxyethylamide and 2-Hydroxyethylhydrazide Derivatives of IASs[a]

compd	structure	X	R	predicted pEC_{50}
84	A	S	H	5.93
85	A	S	2-Me	5.96
86	A	S	3-Me	6.00
87	A	S	4-Me	6.54
88	A	S	2,3-Me$_2$	5.97
89	A	S	3,5-Me$_2$	6.41
90	A	SO$_2$	H	6.76
91	A	SO$_2$	2-Me	6.32
92	A	SO$_2$	3-Me	7.05
93	A	SO$_2$	4-Me	7.11
94	A	SO$_2$	2,4-Me$_2$	7.13
95	A	SO$_2$	3,5-Me$_2$	7.53
96	B	S	H	6.49
97	B	S	2-Me	6.70
98	B	S	3-Me	6.12
99	B	S	4-Me	6.55
100	B	S	2,4-Me$_2$	6.16
101	B	S	3,5-Me$_2$	6.62
102	B	SO$_2$	H	7.06
103	B	SO$_2$	2-Me	7.10
104	B	SO$_2$	3-Me	7.36
105	B	SO$_2$	4-Me	6.98
106	B	SO$_2$	2,4-Me$_2$	6.65
107	B	SO$_2$	3,5-Me$_2$	7.79

[a] The four highest and the two least active predicted compounds are highlighted in bold.

for structure A and **100** for structure B) were chosen for the synthesis (Table 4, prediction set).

Chemistry. Amides **84**, **94**, and **95** were obtained by reacting the corresponding esters with ethanolamine at room temperature overnight. Similarly, hydrazides **100**, **104**, and **107** were prepared by reacting the esters with ethanolhydrazine under the same reaction conditions (Scheme 1). The required esters were prepared as previously reported.[18]

Results and Discussion

Once the molecules chosen from the prediction set were synthesized and tested against HIV-1 in cell-based assay, the experimental pEC_{50} values proved the 3-D QSAR model 3 to be predictive. Moreover, application of the other five 3-D QSAR models to the prediction set confirmed model 3 as the most predictive one. In fact, the standard deviation error of the prediction set (SDEP$_{Pred-Set}$) value, achieved from the application of model 3, was lower than those obtained from the other models (Table 5).

Inspection of the docked conformations of derivatives **84**, **94**, **95**, **100**, **104**, and **107** disclosed a different binding mode compared to that of the reference compound **4**. All six derivatives seem to arrange the indole ring in a different orientation (Figure 6), rotated by about 180°, while the benzene ring maintained almost overlapped with that of **4**. In this new orientation the indole ring still maintains the important hydrogen bond to Lys101 (not shown) as observed for **4**, allowing the

Scheme 1
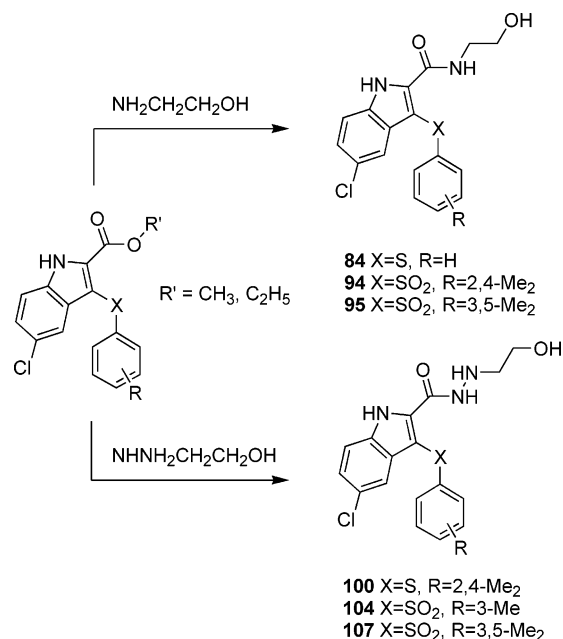
Table 5. Predicted pEC_{50} and SDEP$_{Pred-Set}$ Values Obtained with Models 1−6 after the Prediction Set

		pEC$_{50}$					
		predicted[b]					
compd	exp[a]	M1	M2	M3	M4	M5	M6
84	6.89	5.16	5.78	5.93	5.86	6.09	5.25
94	8.00	6.85	6.17	7.13	5.69	7.08	5.81
95	9.00	6.54	4.10	7.53	4.36	6.51	4.09
100	5.64	7.38	4.37	6.16	4.87	7.25	4.36
104	8.00	5.54	6.97	7.36	5.92	5.91	6.27
107	9.00	6.59	6.37	7.79	7.00	6.69	6.41
SDEP$_{Pred-Set}$		1.55	2.90	1.00	2.82	1.38	2.96

[a] Experimental pEC_{50} ($pEC_{50} = -\log EC_{50}$ (M)). [b] Predicted pEC_{50}: M1 = model 1, M2 = model 2, M3 = model 3, M4 = model 4, M5 = model 5, M6 = model 6.
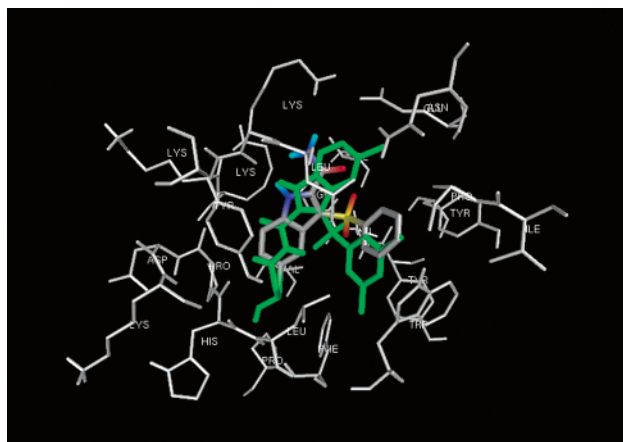
Figure 6. Superimposition of reference compound **5** and derivative **102** (colored in green). A 5 Å core of the NNBS is also included for comparison purposes. Nonpolar hydrogens are not displayed for sake of clarity.

hydroxyethylamide/hydroxyethylhydrazide moieties to fit into the NNBS entrance channel formed by both the main chain and the side chain of Val106, Pro225, Phe227, Leu234, His235, Pro236, and Tyr318. Furthermore, the highly active compounds **95** and **107** seem to make a hydrogen bond between the terminal OH of the

Table 6. Anti-HIV-1 Activity of Compounds **4**, **95**, and **107** against HIV-1 wt Strain and Some Clinically Relevant Resistant Mutants[a]

compd	wt$_{IIIB}$ EC$_{50}$[b]	wt$_{IIIB}$ EC$_{90}$[c]	Y181C EC$_{50}$[b]	K103N–Y181C EC$_{50}$[b]	EFVR EC$_{90}$[d]
95	0.001	0.005	0.01	0.5	0.1
107	0.001	0.006	0.03	10	0.2
4	0.001	0.0007	0.02	8	0.9
NVP^e (**1**)	0.37	0.12	>30	>30	>30
EFV^f (**3**)	0.004	0.008	0.025	0.15	1.8

[a] Data represent mean values for three separate experiments. [b] Compound concentration (μM) required to achieve 50% protection of MT-4 cells from the indicated strain HIV-1 induced cytopathogenicity as determined by the MTT method. [c] Compound concentration (μM) required to reduce the amount of p24 by 90% in wt$_{IIIB}$-infected C8166 cells. [d] Compound concentration (μM) required to reduce the amount of p24 by 90% in C8166 cells infected with an efavirenz-resistant strain EFV^R carrying mutations K103R, V179D, and P225H. [e] NVP, nevirapine. [f] EFV, efavirenz.

ethanolamide/ethanolhydrazide groups and the carbonyl of Leu234 (distance OH$_{95/107}$$\cdots$O=C$_{Leu234}$ = 1.9 Å). As a matter of fact, this change in the binding mode makes the 3-D QSAR model, while reproducing the exact activity trend, slightly underpredictive for five out of six derivatives (compare experimental and predicted activities for **84**, **94**, **100**, **104**, and **107** in Table 4). The problem of this underpredictivity would certainly be resolved by using a wider training set (i.e., including the newly synthesized IAS derivatives).

The above different binding mode of the new IASs suggested that a different activity profile against mutated forms of RT could occur. Thus, the most active derivatives **95** and **107** were also tested against HIV-1 mutant strains (Table 6) to see if the introduction of hydroxyethylamide/hydroxyethylhydrazide moieties would strengthen the antiretroviral activity of the novel IASs in comparison with **4** against resistant mutants. Indeed, derivative **95** showed against either wild type or mutant strains activities superior to those of **4**, while compound **107** was found less active than **4**, especially against the double mutant K103N–Y101C.

Conclusions

The need for novel NNRTIs with improved activity against resistant mutants was a stimulus for the design of novel indolyl aryl sulfones (IASs) related to Merck derivative L-737,126 (**4**). A receptor-based 3-D QSAR model with good predictive power was developed by means of the programs Autodock, GRID, and GOLPE, using a training set of 70 IAS derivatives recently described by us. The introduction of chain elongators containing H-bond acceptors to the nitrogen of 2-carboxamide function of IASs was expected to be beneficial for enhancing the interaction between the inhibitors and the RT at the NNBS. 2-Hydroxyethyl and 2-hydroxyethylamino moieties were chosen as activating elongators in order to examine our assumption and find more potent new IAS derivatives. Among six developed models the most predictive model 3 was used with success to prove its prediction power on a set of 24 new IAS derivatives bearing 2-hydroxyethylaminocarbonyl and 2-hydroxyethylhydrazinocarbonyl moieties. To confirm the goodness of the receptor-based 3-D QSAR model, six derivatives were selected and synthesized for acquiring experimental antiretroviral data to be com-

pared with data predicted from model 3. On the basis of the obtained results, the model proved to be of value in designing new IAS derivatives endowed with high activities and demonstrated the ability to discriminate between high and low activity compounds before their effective synthesis. Newly synthesized indole *N*-(2-hydroxyethyl)carboxamide **95** and *N*-(2-hydroxyethyl)-carbohydrazide **107**, as predicted by model 3, turned out highly active in cell-based assay against both HIV-1 RT and some resistant mutants of clinical interest.

Experimental Section

Chemistry. Melting points (mp) were determined on a Büchi 510 apparatus and are uncorrected. Infrared spectra (IR) were run on Perkin-Elmer 1310 and SpectrumOne spectrophotometers. Band position and absorption ranges are given in cm^{-1}. Proton nuclear magnetic resonance (^{1}H NMR) spectra were recorded on Bruker AM-200 (200 MHz) and Bruker Avance 400 MHz FT spectrometers in the indicated solvent. Chemical shifts are expressed in δ units (ppm) from tetramethylsilane. Column chromatographies were packed with alumina (Merck, 70−230 mesh) and silica gel (Merck, 70−230 mesh). Aluminum oxide TLC cards (Fluka, aluminum oxide precoated aluminum cards with fluorescent indicator at 254 nm) and silica gel TLC cards (Fluka, silica gel precoated aluminum cards with fluorescent indicator at 254 nm) were used for thin-layer chromatography (TLC). Developed plates were visualized by a Spectroline ENF 260C/F UV apparatus. Organic solutions were dried over anhydrous sodium sulfate. Concentration and evaporation of the solvent after reaction or extraction was carried out on a rotary evaporator (Büchi Rotavapor) operating at reduced pressure. Elemental analyses were found within ±0.4% of the theoretical values.

N-**(2-Hydroxyethyl)-3-(phenylthio)-5-chloro-1***H*-**indole-2-carboxamide (84).** A mixture of methyl 3-(phenylthio)-5-chloro-1*H*-indole-2-carboxylate[18] (0.48 g, 1.5 mmol), ethanolamine (7 mL), and ethanol (7 mL) was stirred at room temperature overnight. After dilution with water, the mixture was stirred for an additional 15 min and the outstanding solid was filtered, washed, and dried to afford pure **84** (0.43 g, 82%). Mp: 170−172 °C (from ethanol). ^{1}H NMR (DMSO-d_6): δ 3.40−3.53 (m, 4H), 4.77 (m, 1H, disappeared on treatment with D$_2$O), 7.08−7.36 (m, 6H), 7.45 (d, J = 1.9 Hz, 1H), 7.55 (d, J = 8.7 Hz, 1H), 8.43 (t, J = 4.9 Hz, 1H, disappeared on treatment with D$_2$O), 12.43 ppm (br s, 1H, disappeared on treatment with D$_2$O). IR (Nujol): ν 1630, 3210 and 3300 cm^{-1} Anal. Calcd (C$_{17}$H$_{15}$ClN$_2$O$_2$S (346.83)) C, H, N, Cl, S.

N-**(2-Hydroxyethyl)-3-[(2,4-dimethylphenyl)sulfonyl]-5-chloro-1***H*-**indole-2-carboxamide (94)** was prepared as **84** using methyl 3-[(2,4-dimethylphenyl)sulfonyl]-5-chloro-1*H*-indole-2-carboxylate.[18] Yield: 94%. Mp: 176−178 °C (from ethanol). ^{1}H NMR (DMSO-d_6): δ 2.29 (s, 3H), 2.31 (s, 3H), 3.34 (m, 2H), 3.55 (m, 2H), 4.83 (br s, 1H, disappeared on treatment with D$_2$O), 7.14 (s, 1H), 7.27 (d, J = 8.1 Hz, 1H), 7.35 (dd, J = 2.0 and 8.9 Hz), 7.58 (d, J = 8.9 Hz, 1H), 7.86 (d, J = 1.8 Hz, 1H), 8.02 (d, J = 8.1 Hz, 1H), 8.88 (t, J = 5.2 Hz, 1H, disappeared on treatment with D$_2$O), 13.03 ppm (br s, 1H, disappeared on treatment with D$_2$O). IR (Nujol): ν 1630, 3190, 3290, 3550 cm^{-1}. Anal. Calcd (C$_{19}$H$_{19}$ClN$_2$O$_4$S (406.88)) C, H, Cl, N, S.

N-**(2-Hydroxyethyl)-3-[(3,5-dimethylphenyl)sulfonyl]-5-chloro-1***H*-**indole-2-carboxamide (95)** was prepared as **84** using methyl 3-[(3,5-dimethylphenyl)sulfonyl]-5-chloro-1*H*-indole-2-carboxylate.[18] Yield: 92%. Mp: 272−274 °C (from ethanol/water). ^{1}H NMR (DMSO-d_6): δ 2.31 (s, 6H), 3.44 (m, 2H), 3.60 (t, J = 4.7 Hz, 2H), 4.81 (m, 1H, disappeared on treatment with D$_2$O), 7.27 (s, 1H), 7.34 (dd, J = 2.0 and 8.7 Hz, 1H), 7.54 (d, J = 8.7 Hz, 1H), 7.66 (s, 2H), 8.00 (d, J = 2.0 Hz, 1H), 9.05 (m, 1H, disappeared on treatment with D$_2$O), 12.96 ppm (br s, 1H, disappeared on treatment with D$_2$O). IR (Nujol): ν 1640, 3150, 3400−3510 cm^{-1}. Anal. Calcd (C$_{19}$H$_{19}$ClN$_2$O$_4$S (406.88)) C, H, Cl, N, S.

N′-(2-Hydroxyethyl)-3-[(2,4-dimethylphenyl)thio]-5-chloro-1H-indole-2-carbohydrazide (100) was prepared as **84** using ethanolhydrazine. Yield: 88%. Mp: 170−172 °C (from toluene/ligroin). ^{1}H NMR (DMSO-d_6): δ 2.17 (s, 3H), 2.41 (s, 3H), 2.82 (m, 2H), 3.43 (m, 2H), 4.54 (t, J = 5.2 Hz, 1H, disappeared on treatment with D$_2$O), 5.41 (br s, 1H, disappeared on treatment with D$_2$O), 6.44 (d, J = 8.0 Hz, 1H), 6.78 (d, J = 8.0 Hz, 1H), 7.04 (s, 1H), 7.21−7.34 (m, 2H), 7.54 (d, J = 8.6 Hz, 1H), 9.60 (br s, 1H, disappeared on treatment with D$_2$O), 12.43 ppm (br s, 1H, disappeared on treatment with D$_2$O). IR (Nujol): ν 1620, 3280 cm^{-1}. Anal. Calcd (C$_{19}$H$_{20}$ClN$_3$O$_2$S (389.90)) C, H, Cl, N, S.

N′-(2-Hydroxyethyl)-3-[(3-methylphenyl)sulfonyl]-5-chloro-1H-indole-2-carbohydrazide (104) was prepared as **84** using methyl 3-[(3-methylphenyl)sulfonyl]-5-chloro-1H-indole-2-carboxylate[18] and ethanolhydrazine. Yield: 100%. Mp: 210−215 °C (from ethanol). ^{1}H NMR (DMSO-d_6): δ 2.36 (s, 3H), 2.95 (m, 2H), 3.58 (m, 2H), 4.58 (br s, 1H, disappeared on treatment with D$_2$O), 5.46 (br s, 1H, disappeared on treatment with D$_2$O), 7.33 (dd, J = 1.7 and 8.3 Hz, 1H), 7.43−7.57 (m, 3H), 7.87−7.95 (m, 3H), 10.30 (br s, 1H, disappeared on treatment with D$_2$O), 13.02 ppm (very br s, 1H, disappeared on treatment with D$_2$O). IR (Nujol): ν 1620, 3250 cm^{-1}. Anal. Calcd (C$_{18}$H$_{18}$ClN$_3$O$_4$S (407.87)) C, H, Cl, N, S.

N′-(2-Hydroxyethyl)-3-[(3,5-dimethylphenyl)sulfonyl]-5-chloro-1H-indole-2-carbohydrazide (107) was prepared as **84** using methyl 3-[(3,5-dimethylphenyl)sulfonyl]-5-chloro-1H-indole-2-carboxylate[18] and ethanolhydrazine. Yield: 90%. Mp: 228−230 °C (from ethanol). ^{1}H NMR (CDCl$_3$): δ 1.75 (m, 2H), 2.32 (s, 6H), 2.69 (m, 2H), 4.62 and 4.64 (2 s, 2H), 7.16 (s, 1H), 7.31 (dd, J = 2.1 and 8.7 Hz, 1H), 7.45−7.57 (m, 3H), 8.18 (d, J = 2.1 Hz, 1H), 10.02 (br s, 1H, disappeared on treatment with D$_2$O), 11.2 ppm (very br s, 1H, disappeared on treatment with D$_2$O). IR (Nujol): ν 1635, 1770, 3200 cm^{-1}. Anal. (C$_{19}$H$_{20}$ClN$_3$O$_4$S (421.89)) C, H, N, Cl, S.

Antiviral Assay Procedures.[38] **Compounds.** Compounds were solubilized in DMSO at 200 mM and then diluted in culture medium.

Cells and Viruses. MT-4, C8166, and H9/IIIB cells were grown at 37 °C in a 5% CO$_2$ atmosphere in RPMI 1640 medium, supplemented with 10% fetal calf serum (FCS), 100 IU/mL penicillin G, and 100 μg/mL streptomycin. Cell cultures were checked periodically for the absence of mycoplasma contamination with a MycoTect Kit (Gibco). Human immunodeficiency viruses type-1 (HIV-1, IIIB strain) was obtained from supernatants of persistently infected H9/IIIB cells. The HIV-1 stock solutions had titers of 4.5×10^6 50% cell culture infectious dose (CCID$_{50}$)/mL. The K103R mutant (which also contains the mutations V179D and P225H) was derived from an IIIB strain passaged in C8166 cells in the presence of efavirenz (up to 2 μM). The Y181C mutant (NIH N119) derives from an AZT-sensitive clinical isolate passaged initially in CEM and then in MT-4 cells, in the presence of nevirapine (10 μM). The K103N−Y181C (NIH A17) derives from the IIIB strain passaged in H9 cells in the presence of BI-RG 587 (1 μM). K103R, Y181C, and K103N−Y181C stock solutions had titers of 3.0×10^5, 1.3×10^6, and 2.5×10^5 CCID$_{50}$/mL, respectively.

HIV Titration. Titration of HIV was performed in C8166 cells by the standard limiting dilution method (dilution 1:2, four replica wells per dilution) in 96-well plates. The infectious virus titer was determined by light microscope scoring of syncytia after 4 days of incubation. Virus titers were expressed as CCID$_{50}$/mL.

Anti-HIV Assays. The activity of test compounds against multiplication of wt HIV-1, Y181C, and K103N−Y181C in acutely infected cells was based on inhibition of virus-induced cytopathicity in MT-4 cells. The activity of the compounds against the K103R multiplication in acutely infected cells was based on inhibition of p24 antigen in C8166 cells. Briefly, 50 μL of culture medium containing 1×10^4 cells was added to each well of flat-bottom microtiter trays containing 50 μL of culture medium with or without various concentrations of test compounds. Then 20 μL of HIV suspensions (containing the

appropriate amount of CCID$_{50}$ to cause complete cytopathicity at day 4) was added. After incubation at 37°C, cell viability was determined by the 3-(4,5-dimethylthiazol-1-yl)-2,5-diphenyltetrazolium bromide (MTT) method.[39] Alternatively, p24 levels were determined by an immunoenzymatic kit (Abbott). The cytotoxicity of test compounds was evaluated in parallel with their antiviral activity and was based on the viability of mock-infected cells, as monitored by the MTT method.

Molecular Modeling. Docking Studies and Molecular Alignment. All molecular modeling calculations and manipulations were performed using the software packages Macromodel 7.1,[42] Autodock 3.0.5[30] running on IBM-compatible AMD Athlon 3.0 GHz workstations with the Linux operating system Mandrake 9.0. For the conformational analysis and for any minimization, the all-atom Amber[43] force field was adopted as implemented in the Macromodel package. The crystal structure of 739W94[24] extracted from the corresponding RT complex filed in the Brookhaven Protein Data Bank[44] (entry code 1jlq) was used as a template to model all the IASs derivatives. The NNBS was defined by selecting all the residues within 20 Å from the inhibitor.

The binding mode of **4** was analyzed by a docking procedure using the program Autodock. For the docking, a grid spacing of 0.375 Å and 60 × 80 × 60 points were used. The grid was centered on the mass center of the experimental bound 739W94 coordinates. The GA−LS method was adopted using the default settings. The Amber united atom was assigned to the protein using the program ADT (Auto Dock Tools). Autodock generated 100 possible binding conformations grouped in clusters. Similar setting were also adopted for the dockng of the IAS derivatives.

The starting conformations for all the docking studies were obtained using a molecular dynamic run with simulated annealing procedure as implemented in Macromodel version 7.1 and conducted as follows: each molecule was energy minimized to a low gradient. The nonbonded cutoff distances were set to 20 Å for both van der Waals and electrostatic interactions. An initial random velocity to all atoms corresponding to 300 K was applied. Three subsequent molecular dynamics runs were then performed. The first was carried out for 10 ps with a 1.5 fs time-step at a constant temperature of 300 K for equilibration purposes. The next molecular dynamic was carried out for 20 ps, during which time the system is coupled to a 150 K thermal bath with a time constant of 5 ps. The time constant represents approximately the half-life for equilibration with the bath; consequently, the second molecular dynamic command caused the molecule to slowly cool to approximately 150 K. The third and last dynamic cooled the molecule to 50 K over 20 ps. A final energy minimization was then carried out for 250 iterations using conjugate gradient. The minimizations and the molecular dynamics were in all cases performed in aqueous solution.

Upon docking based on the above setting, the IASs derivatives were receptor-based aligned using the Autodock program. To this each molecule contained in the three sets (training set, test set, and prediction set) was docked into the NNBS. The first ranked docked conformation (best docked conformation) and the lowest energy conformation of the most populated cluster (best cluster conformation) were selected to obtain two different alignment rules.

GRID Calculations. The interaction energies were calculated by using GRID (version 21)[35] running on IBM-compatible AMD Athlon 3.0 GHz workstations equipped with the Linux operating system SUSE 9.0. A grid spacing of 1 Å was used setting the grid dimensions (Å) to X_{min}/X_{max}, 6.0/28.0; Y_{min}/Y_{max}, −46.0/−14.0; and Z_{min}/Z_{max}, −12.0/10.0, centered on the experimental bound conformation of 739W94.

The aromatic C1= and water OH2 probes were used alone and in combination to evaluate the molecular interaction fields of the modeled molecules.

GOLPE PLS Analyses. PLS models were calculated with GOLPE 4.5.12[37] running on a O2 SGI workstation equipped with the operating system 6.5.11. Six models were derived from the two alignments and the three probe combinations

(C1=, OH2 and C1= + OH2). The resulting probe(s)−target interaction energies for each compound were unfolded to produce one-dimensional vector variables for each compound, which were assembled in the so-called **X** matrix. This matrix was pretreated by first using a cutoff of 5 kcal/mol to produce a more symmetrical distribution of energy values and then zeroing small (<0.01) variable values and finally by removing variables with small standard deviation, using appropriate cutoffs. In addition, variables taking only two distributions were also removed. The Smart Region Definition (SRD)[45] algorithm as implemented in the GOLPE program was also used. A number of seeds (1000) were selected using a *D*-optimal design criterion in the *weight space*. Structural differences between different molecules in the series will be reflected in groups of variables, and therefore, groups were generated around each seed in the 3D-space. Variables with a distance of no more than 2 Å to the seeds were included in the groups. If two neighboring groups (with a distance smaller than 10 Å) contained the same information, the groups were collapsed. The groups were used in the variable selection procedure replacing the original variables. The effect of the groups on the predictivity was evaluated, and groups instead of individual variables were removed from the data file. The effect of the grouped variables on the predictivity was evaluated using a fractional factorial design (FFD) procedure. A number of reduced models (twice the number of variables) were built by removing some of the variables according to the FFD design. The effect of dummy variables (20%) on the predictivity was calculated, and only if a variable had a positive effect on the predictivity larger than the effect of the average dummy variable was the variable included in the final model. The FFD selection was repeated until the r^2 and q^2 value did not increase significantly. In the FFD selection the cross-validation was conducted using five random groups for 20 times and a maximum of three principal components. The models were validated using random groups. Molecules were assigned in a random way to five groups of equal size. Reduced models were built by keeping out one group at a time. The formation of the groups was repeated 20 times and using a maximum model dimensionality of three components.

Supporting Information Available: Elemental analyses of new derivatives **84**, **94**, **95**, **100**, **104**, and **107**. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) WHO/UNAIDS AIDS Epidemic Update. December 2003.
(2) Williams, I. G. Enfuvirtide (Fuzeon): The first fusion inhibitor. *Int. J. Clin. Pract.* **2003**, *57*, 890−897.
(3) Vandamme, A. M.; Van Vaerenbergh, K.; De Clercq, E. Anti-human immunodeficiency virus drug combination strategies. *Antivir. Chem. Chemoth.* **1998**, *9*, 187−203.
(4) Condra, J. H.; Miller, M. D.; Hazuda, D. J.; Emini, E. A. Potential new therapies for the treatment of HIV-1 infection. *Annu. Rev. Med.* **2002**, *53*, 541−555.
(5) Young, S. D. Recent advances in the chemotherapy of HIV. *Annu. Rep. Med. Chem.* **2003**, *38*, 173−182.
(6) Tantillo, C.; Ding, J.; Jacobo-Molina, A.; Nanni, R. G.; Boyer, P. L.; Hughes, S. H.; Pauwels, R.; Andries, K.; Janssen, P. A.; Arnold, E. 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.
(7) Spence, R. A.; Kati, W. M.; Anderson, K. S.; Johnson, K. A. Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. *Science* **1995**, *267*, 988−993.
(8) Buckheit, R. W. Nonnucleoside reverse transcriptase inhibitors: Perspectives on novel therapeutic compounds and strategies for the treatment of HIV infection. *Exp. Opin. Invest. Drugs* **2001**, *10*, 1423−1442.
(9) Artico, M. Nonnucleoside anti-HIV-1 reverse transcriptase inhibitors (NNRTIs): A chemical survey from lead compounds to selected drugs for clinical trials. *Farmaco* **1996**, *51*, 305−331.
(10) Flynn, D. L. Abstracts 127−130 of 226th ACS National Meeting, New York, September 7−11, 2003.
(11) De Clercq, E. New anti-HIV agents and targets. *Med. Res. Rev.* **2002**, *22*, 531−565.
(12) De Clercq, E. Nonnucleoside reverse transcriptase inhibitors (NNRTIs): Past, present and future. *Chem. Biodiversity* **2004**, *1*, 44−64.
(13) Williams, T. M.; Ciccarone, T. M.; MacTough, S. C.; Rooney, C. S.; Balani, S. K.; Condra, J. H.; Emini, E. A.; Goldman, M. E.; Greenlee, W. J.; Kauffman, L. R.; et al. 5-chloro-3-(phenylsulfonyl)indole-2-carboxamide: A novel, non-nucleoside inhibitor of HIV-1 reverse transcriptase. *J. Med. Chem.* **1993**, *36*, 1291−1294.
(14) Young, S. D.; Amblard, M. C.; Brichter, S. F.; Grey, V. E.; Tran, L. O.; Lumma, W. C. J.; Huff, J. R.; Schleif, W. A.; Emini, E. E.; O'Brien, J. A.; Pettibone, D. J. 2-Heterocyclic indole-3-sulfones as inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 491−496.
(15) Artico, M.; Silvestri, R.; Stefancich, G.; Massa, S.; Pagnozzi, E.; Musu, D.; Scintu, F.; Pinna, E.; Tinti, E.; La Colla, P. Synthesis of pyrryl aryl sulfones targeted at the HIV-1 reverse transcriptase. *Arch. Pharm. (Weinheim)* **1995**, *328*, 223−229.
(16) Artico, M.; Silvestri, R.; Massa, S.; Loi, A. G.; Corrias, S.; Piras, G.; La Colla, P. 2-Sulfonyl-4-chloroanilino moiety: A potent pharmacophore for the anti-human immunodeficiency virus type 1 activity of pyrrolyl aryl sulfones. *J. Med. Chem.* **1996**, *39*, 522−530.
(17) Artico, M.; Silvestri, R.; Pagnozzi, E.; Bruno, B.; Novellino, E.; Greco, G.; Massa, S.; EtTorre, A.; Loi, A. G.; Scintu, F.; La Colla, P. Structure-based design, synthesis, and biological evaluation of novel pyrrolyl aryl sulfones: HIV-1 non-nucleoside reverse transcriptase inhibitors active at nanomolar concentrations. *J. Med. Chem.* **2000**, *43*, 1886−1891.
(18) Silvestri, R.; De Martino, G.; La Regina, G.; Artico, M.; Massa, S.; Vargiu, L.; Mura, M.; Loi, A. G.; Marceddu, T.; La Colla, P. Novel indolyl aryl sulfones active against HIV-1 carrying NNRTI resistance mutations: Synthesis and SAR studies. *J. Med. Chem.* **2003**, *46*, 2482−2493.
(19) Ren, J.; Diprose, J.; Warren, J.; Esnouf, R. M.; Bird, L. E.; Ikemizu, S.; Slater, M.; Milton, J.; Balzarini, J.; Stuart, D. I.; Stammers, D. K. Phenylethylthiazolylthiourea (PETT) nonnucleoside inhibitors of HIV-1 and HIV-2 reverse transcriptases. Structural and biochemical analyses. *J. Biol. Chem.* **2000**, *275*, 5633−5639.
(20) Hogberg M.; Sahlberg C.; Engelhardt P.; Noreen R.; Kangasmetsa J.; Johansson NG.; Oberg B.; Vrang L.; Zhang H.; BL, S.; Unge T.; Lovgren S.; Fridborg K.; K, B. Urea-PETT compounds as a new class of HIV-1 reverse transcriptase inhibitors. 3. Synthesis and further structure−activity relationship studies of PETT analogues. *J. Med. Chem.* **1999**, *42*, 4150−4160.
(21) Ren, J.; Milton, J.; Weaver, K. L.; Short, S. A.; Stuart, D. I.; Stammers, D. K. Structural basis for the resilience of efavirenz (DMP-266) to drug resistance mutations in HIV-1 reverse transcriptase. *Struct. Fold Des.* **2000**, *8*, 1089−1094.
(22) Pauwels, R.; Andries, K.; Debyser, Z.; Van Daele, P.; Schols, D.; Stoffels, P.; De Vreese, K.; Woestenborghs, R.; Vandamme, A. M.; Janssen, C. G.; et al. Potent and highly selective human immunodeficiency virus type 1 (HIV-1) inhibition by a series of alpha-anilinophenylacetamide derivatives targeted at HIV-1 reverse transcriptase. *Proc. Nat. Ac. Sci. U.S.A.* **1993**, *90*, 1711−1715.
(23) Ding, J.; Das, K.; Moereels, H.; Koymans, L.; Andries, K.; Janssen, P. A.; Hughes, S. H.; Arnold, E. Structure of HIV-1 RT/TIBO R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. *Nat. Struct. Biol.* **1995**, *2*, 407−415.
(24) Chan, J. H.; Hong, J. S.; Hunter, R. N. r.; Orr, G. F.; Cowan, J. R.; Sherman, D. B.; Sparks, S. M.; Reitter, B. E.; Andrews, C. W. r.; Hazen, R. J.; St Clair, M.; Boone, L. R.; Ferris, R. G.;

Creech, K. L.; Roberts, G. B.; Short, S. A.; Weaver, K.; Ott, R. J.; Ren, J.; Hopkins, A.; Stuart, D. I.; Stammers, D. K. 2-Amino-6-arylsulfonylbenzonitriles as non-nucleoside reverse transcriptase inhibitors of HIV-1. *J. Med. Chem.* **2001**, *44*, 1866−1882.

(25) Hopkins, A. L.; Ren, J.; Esnouf, R. M.; Willcox, B. E.; Jones, E. Y.; Ross, C.; Miyasaka, T.; Walker, R. T.; Tanaka, H.; Stammers, D. K.; Stuart, D. I. Complexes of HIV-1 reverse transcriptase with inhibitors of the HEPT series reveal conformational changes relevant to the design of potent non-nucleoside inhibitors. *J. Med. Chem.* **1996**, *39*, 1589−1600.

(26) Ren, J.; Esnouf, R. M.; Hopkins, A. L.; Jones, E. Y.; Kirby, I.; Keeling, J.; Ross, C. K.; Larder, B. A.; Stuart, D. I.; Stammers, D. K. 3′-Azido-3′-deoxythymidine drug resistance mutations in HIV-1 reverse transcriptase can induce long range conformational changes. *Proc. Nat. Ac. Sci. U.S.A.* **1998**, *95*, 9518−9523.

(27) Barnard, J.; Borkow, G.; Parniak, M. A. The thiocarboxanilide nonnucleoside UC781 is a tight-binding inhibitor of HIV-1 reverse transcriptase. *Biochemistry* **1997**, *36*, 7786−7792.

(28) Ren, J.; Esnouf, R. M.; Hopkins, A. L.; Warren, J.; Balzarini, J.; Stuart, D. I.; Stammers, D. K. Crystal structures of HIV-1 reverse transcriptase in complex with carboxanilide derivatives. *Biochemistry* **1998**, *37*, 14394−14403.

(29) Kohlstaedt, L. A.; Wang, J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **1992**, *256*, 1783−1790.

(30) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: Applications of AutoDock. *J. Mol. Recog.* **1996**, *9*, 1−5.

(31) Ragno, R. Cross-Docking: Applcation to HIV−NNRTI inhibitors. Manuscript in Preparation.

(32) Sippl, W.; Contreras, J. M.; Parrot, I.; Rival, Y. M.; Wermuth, C. G. Structure-based 3D QSAR and design of novel acetylcholinesterase inhibitors. *J. Comput. Aided Mol. Des.* **2001**, *15*, 395−410.

(33) Sippl, W. Receptor-based 3D QSAR analysis of estrogen receptor ligands−Merging the accuracy of receptor-based alignments with the computational efficiency of ligand-based methods. *J. Comput. Aided Mol. Des.* **2000**, *14*, 559−572.

(34) Lewis, P. J.; de Jonge, M.; Daeyaert, F.; Koymans, L.; Vinkers, M.; Heeres, J.; Janssen, P. A.; Arnold, E.; Das, K.; Clark, A. D., Jr.; Hughes, S. H.; Boyer, P. L.; de Bethune, M. P.; Pauwels, R.; Andries, K.; Kukla, M.; Ludovici, D.; De Corte, B.; Kavash, R.; Ho, C. On the detection of multiple-binding modes of ligands to proteins, from biological, structural, and modeling data. *J. Comput. Aided Mol. Des.* **2003**, *17*, 129−134.

(35) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(36) Pastor, M.; Cruciani, G.; Watson, K. A. A strategy for the incorporation of water molecules present in a ligand binding site into a three-dimensional quantitative structure−activity relationship analysis. *J. Med. Chem.* **1997**, *40*, 4089−4102.

(37) Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating optimal linear PLS estimations (GOLPE): An advanced chemometric tool for handling 3D-QSAR problems. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9−20.

(38) Cytotoxicities determinations and antiviral tests have been performed at Dipartimento di Scienze e Tecnologie Biomediche, Sezione di Microbiologia e Virologia Generale e Biotecnologie Microbiche, University of Cagliari (Italy).

(39) Pauwels, R.; Andries, K.; Debyser, Z.; Kukla, M. J.; Schols, D.; Breslin, H. J.; Woestenborghs, R.; Desmyter, J.; Janssen, M. A.; De Clercq, E.; et al. New tetrahydroimidazo[4,5,1-*jk*][1,4]-benzodiazepin-2(1H)-one and -thione derivatives are potent inhibitors of human immunodeficiency virus type 1 replication and are synergistic with 2′,3′-dideoxynucleoside analogues. *Antimicrob. Agents Chemother.* **1994**, *38*, 2863−2870.

(40) Tafi, A.; Anastassopoulou, J.; Theophanides, T.; Botta, M.; Corelli, F.; Massa, S.; Artico, M.; Costi, R.; Di Santo, R.; Ragno, R. Molecular modeling of azole antifungal agents active against *Candida albicans*. 1. A comparative molecular field analysis study. *J. Med. Chem.* **1996**, *39*, 1227−1235.

(41) Ragno, R.; Marshall, G. R.; Di Santo, R.; Costi, R.; Massa, S.; Rompei, R.; Artico, M. Antimycobacterial pyrroles: Synthesis, anti-*Mycobacterium tuberculosis* activity and QSAR studies. *Bioorg. Med. Chem.* **2000**, *8*, 1423−1432.

(42) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. MacroModel-An Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. *J. Comput. Chem.* **1990**, *11*, 440−467.

(43) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, I.; Thomas E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal-mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1−41.

(44) Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **2003**, *10*, 980.

(45) Pastor, M.; Cruciani, G.; Clementi, S. Smart region definition: A new way to improve the predictive ability and interpretability of three-dimensional quantitative structure−activity relationships. *J. Med. Chem.* **1997**, *40*, 1455−1464.