

Prediction of Activity for Nonnucleoside Inhibitors with HIV-1 Reverse Transcriptase Based on Monte Carlo Simulations

Robert C. Rizzo,[†] Marina Udier-Blagović,[†] De-Ping Wang, Edward K. Watkins, Marilyn B. Kroeger Smith,[‡] Richard H. Smith, Jr.,[‡] Julian Tirado-Rives, and William L. Jorgensen*

Department of Chemistry, Western Maryland College, and the Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107

Received December 21, 2001

Results of Monte Carlo (MC) simulations for more than 200 nonnucleoside inhibitors of HIV-1 reverse transcriptase (NNRTIs) representing eight diverse chemotypes have been correlated with their anti-HIV activities in an effort to establish simulation protocols and methods that can be used in the development of more effective drugs. Each inhibitor was modeled in a complex with the protein and by itself in water, and potentially useful descriptors of binding affinity were collected during the MC simulations. A viable regression equation was obtained for each data set using an extended linear response approach, which yielded r^2 values between 0.54 and 0.85 and an average unsigned error of only 0.50 kcal/mol. The most common descriptors confirm that a good geometrical match between the inhibitor and the protein is important and that the net loss of hydrogen bonds with the inhibitor upon binding is unfavorable. Other physically reasonable descriptors of binding are needed on a chemotype case-by-case basis. By including descriptors in common from the individual fits, combination regressions that include multiple data sets were also developed. This procedure led to a refined “master” regression for 210 NNRTIs with an r^2 of 0.60 and a cross-validated q^2 of 0.55. The computed activities show an rms error of 0.86 kcal/mol in comparison with experiment and an average unsigned error of 0.69 kcal/mol. Encouraging results were obtained for the predictions of 27 NNRTIs, representing a new chemotype not included in the development of the regression model. Predictions for this test set using the master regression yielded a q^2 value of 0.51 and an average unsigned error of 0.67 kcal/mol. Finally, additional regression analysis reveals that use of ligand-only descriptors leads to models with much diminished predictive ability.

Background

Inhibitors of HIV-1 reverse transcriptase (HIVRT) reduce the ability of the enzyme to perform RNA-dependent DNA polymerase and RNaseH activities required for the conversion of viral RNA to DNA.^{1–3} Because viral DNA copies are necessary for subsequent replication steps, inhibition of HIVRT has evolved to become an important part of anti-HIV therapies.^{1,4} Two broad inhibitor classes toward HIVRT have been discovered.^{3,4} The first, termed nucleoside reverse transcriptase inhibitors (NRTIs), mimic normal active site substrates of HIVRT but lack the 3'-OH group required for DNA chain elongation, which causes premature termination of the growing viral DNA strand (Figure 1).^{3,4} The second class, nonnucleoside reverse transcriptase inhibitors (NNRTIs), binds to a hydrophobic pocket on HIVRT ca. 10–15 Å from the NRTI site (Figure 1).³ This binding event alters the conformation of active site residues hampering normal enzymatic activity.^{4,5}

Clinical trials have determined that the most effective way to suppress HIV replication is through a combination of NRTIs, NNRTIs, and HIV protease inhibitors (HIVPR), a treatment regimen termed highly active antiretroviral therapy (HAART). The success of HAART

Nucleoside binding site (NRTI) Non-nucleoside binding site (NNRTI)

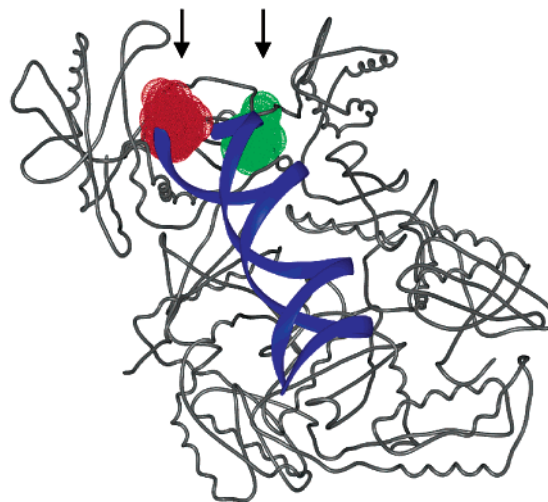


Figure 1. Ribbon diagram showing HIVRT (gray) complexed with DNA (blue) and the location of the NRTI (red) and NNRTI (green) binding sites. Crystal structure coordinates (PDB entry 1rtf) are from refs 32 and 63.

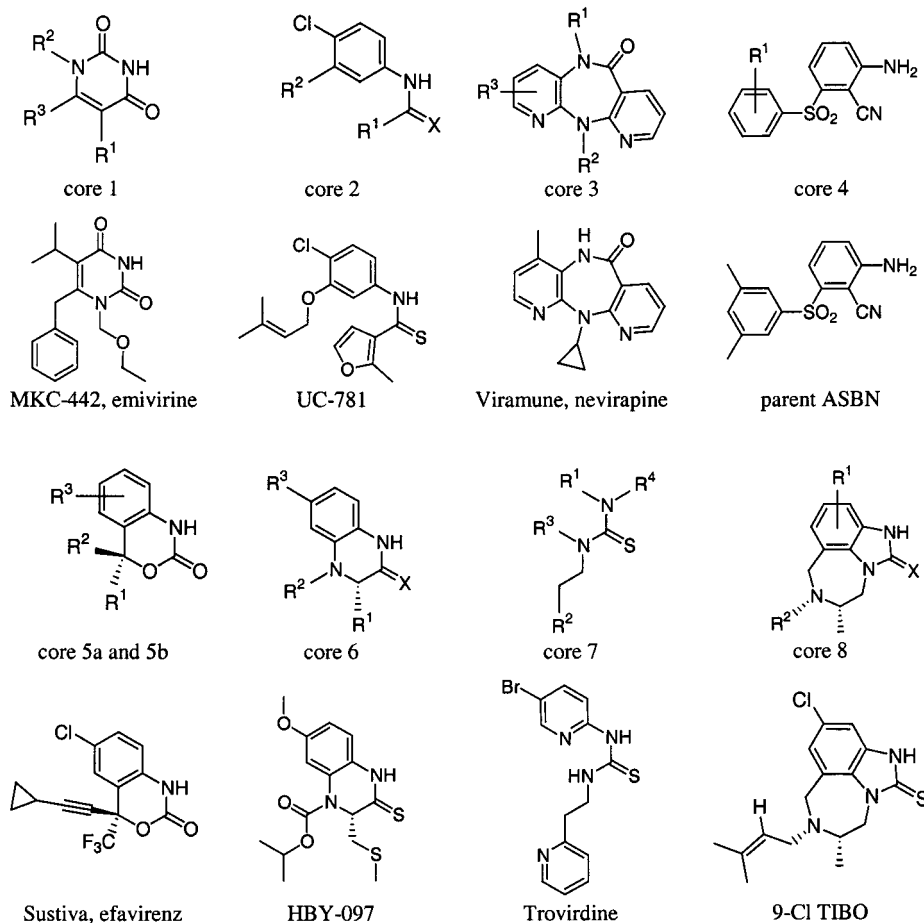
in delaying the onset of acquired immunodeficiency syndrome (AIDS) is attributed in part to the suppressed emergence of mutant virus strains selected for in the presence of single inhibitors⁶ and by the fact that some inhibitors have been shown to act synergistically toward

* To whom correspondence should be addressed. Tel.: 203-432-6278. Fax: 203-432-6299. E-mail: William.Jorgensen@yale.edu.

[†] Authors contributed equally to the work.

[‡] Western Maryland College.

Scheme 1



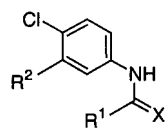
inhibition of HIV replication.⁷ Because mutations in HIVRT affect various NNRTIs differently,⁴ the need for diversity among structurally unique NNRTI cores and within a given core is important. The present work is an extension of our ongoing efforts toward the development of protocols and methods to be used in the design of improved anti-HIV drugs.⁸⁻¹² Computer simulations have been carried out for more than 200 compounds based on the eight different NNRTI cores shown in Scheme 1 and in Tables 1-8 for which anti-HIV activities have been reported in the literature. Viramune (nevirapine, core 3) and Sustiva (efavirenz, cores 5a,b) have been approved by the FDA, while second generation NNRTIs such as carboxanilide (core 2), quinoxaline (core 6), and phenethylthiazolethiourea (PETT, core 7) analogues are in preclinical or early clinical development.^{13,14}

The goal of this study is to develop a physically reasonable computational model for the prediction of NNRTI activity by correlating results of atomic level computer simulations including solvent with experimentally determined anti-HIV activities. If the computational model (scoring function) for each individual core can be incorporated into one regression equation, a general scoring function may emerge that could be used to predict the anti-HIV activity for new NNRTIs. Elucidating the physical reasons for the variations in binding affinities through interpretation of the simulation results is an additional benefit that can contribute to the development of improved procedures for inhibitor design.

Table 1. HEPT Analogues, Core 1

no.	R ¹	R ²	R ³	activity	ca. ΔG_{exptl}
H01	Me	CH ₂ OCH ₂ CH ₂ OH	SPh	7.0 ^a	-7.32
H02	Me	CH ₂ OCH ₂ CH ₂ CH ₃	SPh	3.6 ^a	-7.73
H03	Me	CH ₂ OCH ₂ CH ₃	SPh	0.33 ^a	-9.20
H04	Me	CH ₂ OCH ₃	SPh	2.1 ^a	-8.06
H05	Me	CH ₂ OCH ₂ Ph	SPh	0.088 ^a	-10.01
H06	<i>i</i> -Pr	CH ₂ OCH ₂ Ph	SPh	0.0027 ^a	-12.16
H07	Me	Et	SPh	2.2 ^a	-8.03
H09	Et	CH ₂ OCH ₂ CH ₃	SPh	0.019 ^a	-10.96
H10	<i>i</i> -Pr	CH ₂ OCH ₂ CH ₃	SPh	0.012 ^a	-11.24
H11	<i>i</i> -Pr	CH ₂ OCH ₂ CH ₃	CH ₂ Ph	0.004 ^b	-11.89
H12	<i>c</i> -Pr	CH ₂ OCH ₂ CH ₃	SPh	0.1 ^a	-9.93
H13	Me	CH ₂ OCH ₂ CH ₂ OH	CH ₂ Ph	23.0 ^c	-6.52
H14	Me	CH ₂ OCH ₂ CH ₂ OH	OPh	85.0 ^c	-5.78
H15	Me	CH ₂ OCH ₂ CH ₂ OH	SPh-3,5 di-Me	0.26 ^d	-9.35
H16	Et	CH ₂ OCH ₂ CH ₂ OH	SPh-3,5 di-Me	0.013 ^d	-11.19
H17	<i>i</i> -Pr	CH ₂ OCH ₂ CH ₂ OH	SPh-3,5 di-Me	0.0027 ^d	-12.16
H18	Et	CH ₂ OCH ₂ Ph	SPh	0.0059 ^a	-11.68
H20	Me	Bu	SPh	1.2 ^a	-8.40
MKC-442				0.004 ^b	-11.89

^a Ref 47. ^b Ref 48. ^c Ref 49. ^d Ref 50. H01 is parent HEPT, and H11 is MKC-442. H17 was excluded as an outlier from the individual regression (eq 5), from the build-up model (eqs 16-20), and from the final regression model (eq 21). Activity (cell-based assay) is in micromolar at 37 °C. Estimated experimental binding energies $\Delta G_{\text{exptl}} \approx RT \ln(\text{activity})$ are in kcal/mol. ΔG_{exptl} range = 6.38, $N = 18$.

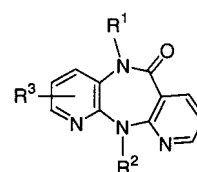
Table 2. Carboxanilide Analogues, Core 2

no.	R ¹	X	R ²	activity	ca. ΔG_{exptl}
U02	2-Ome-Phe	O	COO- <i>c</i> -pentyl	1.2 ^a	-8.40
U03	OCH(CH ₃) ₂	S	COO- <i>i</i> -Pr	0.029 ^b	-10.72
U04	3-(2-Me-Fur)	S	OCH ₂ CH=C(CH ₃) ₂	0.009 ^a	-11.42
U05	2,3-dihydro(1,4)-oxathiine	S	COO- <i>i</i> -Pr	0.019 ^b	-10.96
U06	3-(2-Me-Fur)	S	COO-CH ₂ CH(CH ₃) ₂	0.028 ^a	-10.72
U07	3-(2-Me-Fur)	S	COO-CH ₂ - <i>c</i> -Pr	0.043 ^a	-10.45
U08	3-(2-Me-Fur)	S	OCH ₂ CH=CHCH ₃ (trans)	0.018 ^a	-10.99
U09	3-(2-Me-Fur)	S	COO-CH ₂ CF ₃	0.018 ^a	-10.99
U10	3-(2-Me-Fur)	S	COO- <i>c</i> -hexyl	0.011 ^a	-11.30
U11	2-F-Phe	S	COO- <i>c</i> -hexyl	0.102 ^a	-9.92
U12	3-(2-Me-Fur)	S	CH=N-OCH ₃	0.070 ^a	-10.15
U13	3-(2-Me-Fur)	O	CH ₂ OC(CH ₃) ₃	2.2 ^a	-8.03
U14	3-(2-Me-Fur)	S	COO- <i>c</i> -pentyl	0.013 ^a	-11.19
U18	3-(2-Me-Fur)	S	CH ₂ OC(CH ₃) ₃	0.087 ^a	-10.02
U19	3-(2-Me-Fur)	S	CH=N-OC(CH ₃) ₃	0.142 ^a	-9.72
U20	3-(2-Me-Fur)	S	OCH ₂ C(CH ₃) ₃	0.45 ^a	-9.01
U22	Phe	S	CH=N-OC(CH ₃) ₃	0.086 ^a	-10.03
U23	3-(2-Me-Fur)	S	COO-CH-(<i>i</i> -Pr) ₂	0.076 ^a	-10.10
U24	OCH(CH ₃) ₂	S	CH=N-OCH(CH ₃) ₂	0.032 ^b	-10.64
U25	2-(3-Me-Thie)	S	OCH ₂ CH=CH ₂	0.108 ^a	-9.89
U26	3-(2-Me-Fur)	S	CH=N-OCH ₂ -CH=CH ₂	0.021 ^a	-10.90
U27	2-(3-Me-Thie)	S	CH=N-OCH(CH ₃) ₂	0.022 ^a	-10.87
U28	3-(2-Me-Fur)	S	COO-CH(CH ₃)-(<i>c</i> -Pr)	0.025 ^a	-10.81
U29	OCH(CH ₃) ₂	S	COO-CH-(<i>i</i> -Pr) ₂	0.188 ^b	-9.54
U30	3-Thie	S	COO- <i>c</i> -hexyl	0.078 ^a	-10.09
U31	3-(2-Me-Fur)	S	OCH ₂ COOC(CH ₃) ₃	0.757 ^a	-8.69
U32	2-(3-Me-Thie)	S	COO-CH-(<i>i</i> -Pr) ₂	0.110 ^a	-9.88
U33	OCH(CH ₃) ₂	S	OCH ₂ C(CH ₃)=CH ₂	0.3 ^b	-9.26
U34	3-(2-Me-Fur)	S	CH=N-O- <i>c</i> -pentyl	0.063 ^a	-10.22
U35	2-Ome-Phe	S	COO- <i>i</i> -Pr	0.513 ^a	-8.93
U36	2-(3-Me-Thie)	S	CH=N-OC(CH ₃) ₃	0.027 ^a	-10.74
U37	3-(2-Me-Fur)	S	OCH ₂ CH ₂ CH(CH ₃) ₂	0.021 ^a	-10.90
U38	3-(2-Me-Fur)	S	COO- <i>i</i> -Pr	0.021 ^a	-10.90
U39	3-(2-Me-Fur)	O	COO- <i>i</i> -Pr	1.4 ^a	-8.31
U40	2-(3-Me-Thie)	O	CH=N-OC(CH ₃) ₃	0.644 ^a	-8.79
U41	2-Cl-Phe	S	COO- <i>c</i> -hexyl	1.2 ^a	-8.40
U43	3-(2-Me-Fur)	S	SCH ₂ COOC(CH ₃) ₃	0.55 ^a	-8.88
U44	OCH(CH ₃) ₂	O	COO- <i>i</i> -Pr	1.8 ^b	-8.14
U45	OCH(CH ₃) ₂	S	OCH ₂ CH=CHCH ₃ (trans)	0.133 ^b	-9.76
U47	OCH(CH ₃) ₂	S	COO-Et	0.099 ^b	-9.94
U48	Phe	S	COO- <i>i</i> -Pr	0.15 ^a	-9.68
U49	Phe	O	COO- <i>i</i> -Pr	1.3 ^a	-8.35
U50	2-F-Phe	O	COO- <i>c</i> -hexyl	1.6 ^a	-8.23
U51	OCH(CH ₃) ₂	S	COOCH ₂ - <i>i</i> -Pr	0.03 ^b	-10.67
U52	OCH(CH ₃) ₂	S	COOCH ₂ - <i>c</i> -Pr	0.1 ^b	-9.81
U53	OCH(CH ₃) ₂	S	COO- <i>c</i> -pentyl	0.06 ^b	-10.26
MKC-442				0.01 ^a	-11.35
nevirapine				0.03 ^b	-10.76
9-Cl TIBO				0.05 ^b	-10.37

^a Ref 51. ^b Ref 52. U03 is UC-38 (PDB entry 1rt6), U04 is UC-781 (PDB entry 1rt4), and U19 is UC-10 (PDB entry 1rt5). PDB entries are from ref 32. U10, U26, U38, U41, and U50 are excluded as outliers from the individual regression (eq 6) and from the build-up model (eqs 17–20). U02, U41, U44, and U50 are excluded as outliers from the final regression model (eq 21). Activity (cell-based assay) is in micromolar at 37 °C. Estimated experimental binding energies $\Delta G_{\text{exptl}} \approx RT \ln(\text{activity})$ are in kcal/mol. ΔG_{exptl} range = 3.39, $N = 46$.

Theoretical Methods

The most rigorous computational approaches used for the calculation of binding affinities (ΔG_b) are the free energy perturbation (FEP) and thermodynamic integration (TI) methods.^{15–18} These methods typically employ molecular dynamics (MD) or Monte Carlo (MC) simulations and have yielded impressive results for a number

Table 3. Nevirapine Analogues, Core 3

no.	R ¹	R ²	R ³	activity ^a	ca. ΔG_{exptl}
N01	Me	Et	H	0.125	-9.42
N02	Me	Et	2-Me	0.17	-9.24
N03	Me	Et	2-Cl	0.15	-9.31
N04	Me	Et	3-Me	0.76	-8.35
N06	Me	Et	4-Me	1.9	-7.81
N07	H	Et	H	0.44	-8.67
N08	H	Et	4-Me	0.035	-10.17
N09	H	Et	4-Cl	0.095	-9.58
N10	H	<i>c</i> -Pr	4-Me	0.084	-9.65
N12	Me	Pr	H	0.45	-8.66
N13	Me	<i>t</i> -Bu	H	11.0	-6.77
N14	Me	COCH ₃	H	15.3	-6.57
N15	H	Et	4-Et	0.11	-9.49
N16	Me	CH ₂ SCH ₃	H	0.85	-8.28
N17	H	<i>c</i> -Pr	4-CH ₂ OH	3.0	-7.54
N18	H	<i>c</i> -Pr	4-CN	1.25	-8.05
N19	Me	CH ₂ CH ₂ F	H	2.9	-7.56
N20	H	<i>c</i> -Pr	H	0.45	-8.66
nevirapine				0.084	-9.65

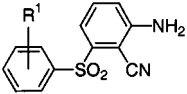
^a Ref 53. N10 is nevirapine. N08 and N13 are excluded as outliers from the individual regression (eq 7) and from the build-up model (eqs 19–20). N13 is excluded as an outlier from the final regression model (eq 21). Activity (enzyme-based assay) is in micromolar at 25 °C. Estimated experimental binding energies $\Delta G_{\text{exptl}} \approx RT \ln(\text{activity})$ are in kcal/mol. ΔG_{exptl} range = 3.60, $N = 18$.

of protein–ligand systems, as reviewed elsewhere.^{15–18} A more approximate method for the estimation of ΔG_b is based on linear response (LR) theory, as introduced by Åqvist and co-workers (eq 1).¹⁹ This approach is considerably faster than the FEP or TI alternatives because no intermediate transformation process is required to calculate the binding affinity.¹⁹

$$\Delta G_b = \alpha \langle \Delta E_{\text{vdw}} \rangle + \beta \langle \Delta E_{\text{Coul}} \rangle \quad (1)$$

Here, $\langle \rangle$ signifies an ensemble average of the difference of the inhibitor–solvent plus inhibitor–protein interaction energies (ΔE) in the bound state and of the inhibitor–solvent interaction energies in the unbound state.¹⁹ The two energy terms represent the differences in average van der Waals (Lennard–Jones) and electrostatic (Coulombic) contributions, respectively, which are normally calculated using a molecular mechanics force field and either MD or MC simulations. The Coulombic energy differences were originally scaled by a factor $\beta = 0.50$, while the coefficient α was determined by fitting the simulation results to known experimental binding affinities.¹⁹

Jorgensen et al. modified the LR approach for the calculations of free energies of solvation, which corresponds to eq 2 for computing free energies of binding.^{20,21} In this approach, both coefficients α and β are allowed to vary, and a third term representing the solvent accessible surface area (SASA) of the solute is included and scaled by a coefficient γ . The rationale for the SASA term is that it provides a means to account for possible positive free energies of hydration caused by the penalty

Table 4. ASBN Analogues, Core 4


no.	R ¹	activity ^a	ca. ΔG _{exptl}
A01	H	2.0	-8.09
A02	2-OMe	0.6	-8.83
A03	3-OMe	0.9	-8.58
A04	4-OMe	25	-6.53
A05	2-Me	2.3	-8.00
A06	3-Me	0.4	-9.08
A07	4-Me	9.5	-7.13
A08	2-Cl	4.1	-7.65
A09	3-Cl	0.59	-8.84
A10	4-Cl	3	-7.84
A11	2-Br	5.0	-7.52
A12	3-Br	0.54	-8.89
A14	2-F	3.0	-7.84
A15	3-F	3.0	-7.84
A16	2-CN	5.4	-7.48
A17	3-CN	2.4	-7.98
A19	3-CF ₃	3.5	-7.74
A20	2,5-di-Cl	0.3	-9.26
A21	3,5-di-Cl	0.07	-10.15
A22	3,5-di-Me	0.01	-11.35
A23	3-Br, 5-Me	0.02	-10.93
A24	3-Cl, 5-Me	0.03	-10.68
A25	3-OMe, 5-Me	0.05	-10.36
A26	3-OMe, 5-CF ₃	0.09	-10.00
A27	3-OH, 5-Me	0.43	-9.04
A28	3-OEt, 5-Me	0.06	-10.25
A29	3-OPr, 5-Me	0.06	-10.25
A30	3-OBu, 5-Me	0.6	-8.83
nevirapine		0.089	-10.01
Sustiva		0.0009	-12.84

^a Ref 54. A02 and A07 are excluded as outliers from the individual regression (eq 8) and from the build-up model (eqs 14–20). A19 and A22 are excluded as outliers from the final regression model (eq 21). Activity (cell-based assay) is in micromolar at 37 °C. Estimated experimental binding energies ΔG_{exptl} ≈ RT ln(activity) are in kcal/mol. ΔG_{exptl} range = 4.82, N = 28.

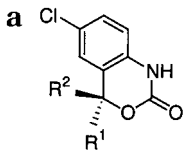
for solute cavity formation in water.^{20,21}

$$\Delta G_b = \alpha(\Delta E_{\text{vdw}}) + \beta(\Delta E_{\text{Coul}}) + \gamma(\Delta \text{SASA}) \quad (2)$$

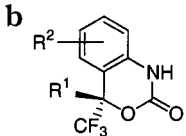
Encouraged by prior MD/LR^{19,22–26} and MC/LR^{8,27,28} binding studies, we endeavored to treat larger data sets to see if good correlations with experimental data could still be obtained with a higher ratio of data points to parameters. Simultaneously, Duffy and Jorgensen correlated results from aqueous MC simulations with solvation properties for more than 200 diverse organic compounds.²⁹ The descriptors were expanded from those in eq 2 to include, for example, hydrogen bond counts and the hydrophobic, hydrophilic, and aromatic components of the SASA. A multivariate fitting approach was used, which corresponds to eq 3 for computing binding affinities.

$$\Delta G_b = \sum_n c_n \xi_n + \text{constant} \quad (3)$$

Here, c_n represents an optimizable coefficient for the associated descriptor ξ_n . In principle, any physically reasonable quantity could be used as a descriptor in this extended linear response (ELR) approach. Specifically relevant to protein–ligand binding was the success in predictions of log P (octanol/water) for 200 solutes. Only four descriptors were needed to yield a correlation with

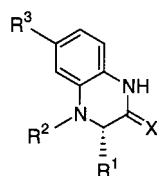
Table 5. Sustiva Analogues, Core 5


Section A				
no.	R ¹	R ²	activity ^a	ca. ΔG _{exptl}
Xs25	CF ₃	ethynyl- <i>c</i> -Pr	0.002	-12.35
Xs26.2	CF ₃	ethynyl-Ph	0.0086	-11.45
Xs27.1	CF ₃	Et	2.3	-8.43
Xs28.1	CF ₃	Ph	0.145	-10.13
Xs29.1	CF ₃	SPh	0.860	-9.04
Xs31.1	CF ₃	allyl	0.22	-9.88
Xs32.1	<i>c</i> -Pr	Pr	0.307	-9.67
Xs33.1	Pr	Ph	5.4	-7.90
Xs34.1	Et	Ph	0.3	-9.68
Xs35	Et	Et	16.5	-6.79
Xs36.1	ethynyl	Ph	0.65	-9.21
Xs37.1	Et	Pr	0.052	-10.76
Xs38	CF ₂ Cl	ethynyl-Ph	0.012	-11.67
Xs41.1	CF ₃	ethynyl-CH ₂ -OMe	0.015	-11.53
Xs43	CF ₃	ethynyl-CH ₂ -OH	0.55	-9.31
Xs44.2	ethynyl	Pr	1.9	-8.55
Xs45	<i>c</i> -Pr	ethynyl-CH ₂ -OMe	0.41	-9.49
Xs46.1	Pr	<i>t</i> -Bu	5.3	-7.91
Sustiva			0.002	-12.35



Section B				
no.	R ¹	R ²	activity	ca. ΔG _{exptl}
s01	ethynyl- <i>c</i> -Pr	H	0.478 ^b	-9.40
s02	ethynyl- <i>c</i> -Pr	6-F	0.19 ^b	-9.97
s03	ethynyl- <i>c</i> -Pr	6- <i>i</i> -Pr	1.958 ^b	-8.53
s04	ethynyl- <i>c</i> -Pr	6-N(CH ₃) ₂	0.816 ^b	-9.07
s05	ethynyl- <i>c</i> -Pr	6-O-CF ₃	1.249 ^b	-8.81
s06	ethynyl- <i>c</i> -Pr	5,6-di-F	0.084 ^b	-10.47
s07	ethynyl- <i>c</i> -Pr	5,8-di-F	0.796 ^b	-9.08
s08	ethynyl- <i>c</i> -Pr	5,6,8-tri-F	0.8 ^b	-9.08
s09	ethynyl- <i>c</i> -Pr	5,6,7-tri-F	0.442 ^b	-9.45
s10	ethynyl- <i>c</i> -Pr	6-OMe	0.131 ^b	-10.20
s11	ethynyl- <i>c</i> -Pr	6-Me	0.133 ^b	-10.19
s12	ethynyl- <i>c</i> -Pr	5-F	0.078 ^c	-10.52
s13	ethynyl-Et	5-F	0.127 ^c	-10.21
s14	ethynyl-Pr	5-F	0.156 ^c	-10.09
s15	ethynyl- <i>i</i> -Pr	5-F	0.102 ^c	-10.35
s16	ethynyl- <i>c</i> -Pr	6-NO ₂	0.209 ^c	-9.91
s17	ethynyl-Et	6-NO ₂	0.276 ^c	-9.74
s18	ethynyl-Pr	6-NO ₂	0.304 ^c	-9.68
s19	ethynyl- <i>i</i> -Pr	6-NO ₂	0.199 ^c	-9.94
s20	ethynyl- <i>c</i> -Pr	6-NH ₂	0.802 ^c	-9.08
s21	ethynyl-Et	6-NH ₂	1.894 ^c	-8.55
s22	ethynyl-Pr	6-NH ₂	1.506 ^c	-8.69
s23	ethynyl- <i>i</i> -Pr	6-NH ₂	0.896 ^c	-9.01
s24	ethynyl- <i>c</i> -Pr	6-NHCH ₃	0.608 ^c	-9.25
s25	ethynyl- <i>i</i> -Pr	6-NHCH ₃	0.473 ^c	-9.40
Sustiva			0.048 ^{b,c}	-10.40

^a Ref 55. Xs33.1 and Xs37.1 are excluded as outliers from the individual regression (eq 9) and from the build-up model (eqs 18–20). Xs25, Xs26.2, Xs35, Xs37.1, and Xs38 are excluded as outliers from the final regression model (eq 21). Activity (enzyme-based assay) is in micromolar at 37 °C. The activities of analogues Xs25 (Sustiva) and Xs26.2 reflect values determined for single enantiomers. For all other compounds, assays were performed using racemic mixtures. Therefore, the anti-HIV activities were divided by two, since the biological evaluation of both enantiomers of Xs26.2 revealed that only one enantiomer is active. Estimated experimental binding energies ΔG_{exptl} ≈ RT ln(activity) are in kcal/mol. ΔG_{exptl} range = 5.56, N = 18. ^b Ref 56. ^c Ref 57. s01, s03, and s05 are excluded as outliers from the individual regression (eq 10) and from the build-up model (eq 20). s03, s04, s05, and s12 are excluded as outliers from the final regression model (eq 21). Activity (enzyme-based assay) is in micromolar at 37 °C. ΔG_{exptl} range = 1.99, N = 25.

Table 6. Quinoxaline Analogues, Core 6

no.	R ¹	R ²	R ³	X	activity	ca. ΔG_{exptl}
Q01	CH ₂ SCH ₃	COO- <i>i</i> -Pr	OCH ₃	S	0.006 ^{a,b}	-11.68
Q02	Et	COO- <i>i</i> -Pr	H	S	0.004 ^a	-11.98
Q03	Pr	COO- <i>i</i> -Pr	H	S	0.014 ^a	-11.16
Q04	Et	COO- <i>i</i> -Pr	Cl	S	0.006 ^a	-11.63
Q05	CH ₂ SCH ₃	COO- <i>i</i> -Pr	F	S	0.009 ^a	-11.41
Q06	Et	COO- <i>i</i> -Pr	OCH ₃	S	0.012 ^a	-11.22
Q07	CH ₂ OCH ₃	COO- <i>i</i> -Pr	OCH ₃	S	0.025 ^a	-10.80
Q08	Et	COOC(CH ₃)=CH ₂	Cl	S	0.003 ^a	-12.19
Q09	Et	COO-allyl	Cl	S	0.005 ^a	-11.76
Q10	CH ₂ SCH ₃	COOC(CH ₃)=CH ₂	Cl	S	0.005 ^a	-11.82
Q11	CH ₂ SCH ₃	COO- <i>i</i> -Pr	OH	S	0.123 ^a	-9.81
Q12	Me	COO-allyl	Cl	S	0.135 ^a	-9.75
Q13	Me	COOC(CH ₃)=CH ₂	Cl	S	0.034 ^a	-10.60
Q14	CH ₂ SCH ₃	COOC(CH ₃)C ₂ H ₅	OCH ₃	S	0.023 ^a	-10.85
Q15	CH ₂ SCH ₃	COOCH ₂ CH(CH ₃) ₂	OCH ₃	S	0.113 ^a	-9.86
Q16	Et	COOCH ₃	F	S	0.037 ^a	-10.54
Q17	Et	COO- <i>n</i> -Bu	F	S	0.005 ^a	-10.63
Q18	Et	COOCH ₂ CH(CH ₃) ₂	F	S	0.032 ^a	-10.63
Q19	Et	COOCH(CH ₃)C ₂ H ₅	F	S	0.026 ^a	-10.77
Q20	CH ₂ OCH ₃	COO- <i>i</i> -Pr	F	S	0.026 ^a	-10.77
Q21	Et	COO- <i>i</i> -Pr	F	O	0.029 ^a	-10.71
Q22	CH ₂ SCH ₃	COOC(CH ₃)=CH ₂	Cl	O	0.061 ^a	-10.24
Q23	Et	COOC(CH ₃)=CH ₂	F	O	0.287 ^a	-9.28
Q24	CH ₂ SCH ₃	COOC(CH ₃)=CH ₂	OCH ₃	O	0.025 ^a	-10.79
Q25	Et	COO- <i>n</i> -Bu	F	O	0.272 ^a	-9.32
nevirapine					0.208 ^b	-9.48

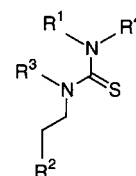
^a Ref 58. ^b Ref 59. Q02 and Q12 are excluded as outliers from the individual regression (eq 11) and from the build-up model (eqs 14–20). Q02 is excluded as an outlier from the final regression model (eq 21). Activity (cell-based assay) is in micromolar at 37 °C. Estimated experimental binding energies $\Delta G_{\text{exptl}} \approx RT \ln(\text{activity})$ are in kcal/mol. ΔG_{exptl} range = 2.91, $N = 25$.

an r^2 of 0.91 and an rms error of 0.53.²⁹ Given the potential parallel between solute octanol/water partitioning and ligand protein/water partitioning, we also considered alternative descriptors for protein–ligand binding for 40 NNRTIs, representing 20 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT, core 1) and 20 nevirapine analogues (core 3).¹⁰ The present study tests the viability of this approach on a much larger and more diverse collection of NNRTIs in search of a general method for prediction of anti-HIV activity. ELR regressions have also been reported for 20 inhibitors of thrombin.³⁰

It should be emphasized that LR and ELR methods rely on using experimental data, in conjunction with a set of descriptors obtained via computer simulations to derive a regression equation. However, once a reasonable, cross-validated regression equation is derived, no additional experimental data are needed in order to make activity predictions for novel compounds. Only simulations for the bound and unbound states are required to make activity predictions for any new compound.¹⁰ Given the potential utility of the ELR method for ranking hundreds of compounds in a structure-based drug design scenario, additional investigation, validation, and refinement of the method are warranted.

Computational Details

System Setups. Given the large size of HIVRT, a reduced model of the NNRTI binding site is utilized, as

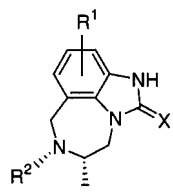
Table 7. PETT Analogues, Core 7

no.	R ¹	R ²	R ³	R ⁴	activity	ca. ΔG_{exptl}
PT301	2-thiazole	Ph	H	H	1.3 ^a	-8.35
PT302	2-thiazole	4-F-Ph	H	H	3.3 ^a	-7.78
PT304	2-thiazole	3-F-Ph	H	H	0.25 ^a	-9.37
PT306	2-thiazole	2-F-Ph	H	H	0.1 ^a	-9.93
PT307	2-thiazole	2,6-di-F-Ph	H	H	0.02 ^a	-10.93
PT308	2-thiazole	2-Cl-Ph	H	H	0.4 ^a	-9.08
PT309	2-thiazole	2-Cl, 6-F-Ph	H	H	0.05 ^a	-10.36
PT310	2-thiazole	2-OH-Ph	H	H	4.0 ^a	-7.66
PT311	2-thiazole	2-Ome-Ph	H	H	0.4 ^a	-9.08
PT312	5-Cl-2-thiazole	Ph	H	H	2.7 ^a	-7.90
PT313	4-Me-2-thiazole	Ph	H	H	0.4 ^a	-9.08
PT314	4-CF ₃ -2-thiazole	Ph	H	H	0.5 ^a	-8.94
PT315	4-Et-2-thiazole	Ph	H	H	0.7 ^a	-8.74
PT316	4-Pr-2-thiazole	Ph	H	H	1.6 ^a	-8.23
PT317	4- <i>i</i> -Pr-2-thiazole	Ph	H	H	1.3 ^a	-8.35
PT318	<i>t</i> -Bu	Ph	H	H	1.3 ^a	-8.35
PT319	2-thiazole	3-Ome-Ph	H	H	0.6 ^a	-8.83
PT320	2-thiazole	4-Ome-Ph	H	H	5.5 ^a	-7.46
PT322	2-thiazole	Ph	H	Me	10.0 ^a	-7.10
PT323	2-thiazole	Ph	Me	H	2.3 ^a	-8.00
PT101	2-pyridyl	Ph	H	H	0.2 ^b	-9.51
PT102	5-Br-2-pyridyl	Ph	H	H	0.05 ^b	-10.36
PT103	5-Me-2-pyridyl	Ph	H	H	0.15 ^b	-9.68
PT104	5-Br-2-pyridyl	2,6-di-F-Ph	H	H	0.013 ^b	-11.19
PT106	2-pyrazine	Ph	H	H	10.0 ^b	-7.10
PT107	5-Br-2-pyridyl	2-pyridyl	H	H	0.02 ^b	-10.93
PT108	5-Me-2-pyridyl	2,6-di-F-Ph	H	H	0.01 ^b	-11.35
PT109	5-Cl-2-pyridyl	2,4,6-tri-F-Ph	H	H	0.01 ^b	-11.35
PT110	5-Cl-2-pyridyl	3-CN, 2,6-di-F-Ph	H	H	0.006 ^b	-11.67
PT111	5-Br-2-pyridyl	2-CN, 2-Ome, 6-F-Ph	H	H	0.003 ^b	-12.10
PT112	5-Br-2-pyridyl	2-Cl, 3-OEt, 6-F-Ph	H	H	0.007 ^b	-11.57
PT113	5-Br-2-pyridyl	3-CONHCH ₃ , 2,6-di-F-Ph	H	H	0.08 ^b	-10.07
PT114	5-Cl-2-pyridyl	4-N(CH ₃) ₂ , 2,6-di-F-Ph	H	H	0.018 ^b	-10.99
PT115	5-Br-2-pyridyl	3-Ome, 2,6-di-F-Ph	H	H	0.025 ^b	-10.79
nevirapine					0.15 ^{a,b}	-9.68
9-Cl TIBO					0.25 ^{a,b}	-9.37

^a Ref 60. ^b Ref 61. PT309, PT107, and PT111 are excluded as outliers from the individual regression (eq 12) and from the build-up model (eqs 14–20). PT314, PT318, and PT107 are excluded as outliers from the final regression model (eq 21). Activity (cell-based assay) is in micromolar at 37 °C. Estimated experimental binding energies $\Delta G_{\text{exptl}} \approx RT \ln(\text{activity})$ are in kcal/mol. ΔG_{exptl} range = 5.00, $N = 34$.

previously described in detail.¹⁰ Briefly, the model includes each inhibitor plus the nearest 123 protein residues that line the NNRTI site (Figure 2). Forty-two residues are actively sampled in the MC simulations (95–108A, 179–183A, 186–191A, 198A, 225–229A, 233–239A, 318–319A, 136B, and 138B). The remaining residues are held fixed (91–94A, 109–110A, 161–178A, 184–185A, 192–197A, 199–205A, 222–224A, 230–232A, 240–242A, 316–317A, 320–321A, 343–349A, 381–383A, 134–135B, 137B, and 140B) after an initial full energy minimization of each complex.

The simulations for each protein–ligand complex were initiated from the X-ray structure coordinates of HIVRT with MKC-442³¹ (core 1), UC-781³² (core 2), nevirapine³³ (core 3), Sustiva¹¹ (core 5), HBV-097³⁴ (core 6), PETT³⁵ (core 7), and 9-Cl 4,5,6,7-tetrahydroimidazo-

Table 8. TIBO Analogues, Core 8


no.	R ¹	R ²	X	activity	ca. ΔG_{exptl}
B01	8-Br	dimethyl-allyl	S	0.003 ^a	-12.09
B02	8-Cl	dimethyl-allyl	S	0.0043 ^a	-11.87
B03	8-F	dimethyl-allyl	S	0.0058 ^a	-11.69
B04	8-Me	dimethyl-allyl	S	0.0136 ^a	-11.16
B05	9-F	dimethyl-allyl	S	0.025 ^a	-10.79
B06	9,10-di-Cl	dimethyl-allyl	S	0.0255 ^a	-10.76
B07	8-C≡CH	dimethyl-allyl	S	0.0296 ^a	-10.68
B08	9-Cl	dimethyl-allyl	S	0.034 ^a	-10.60
B09	H	dimethyl-allyl	S	0.044 ^a	-10.44
B10	8-Br	dimethyl-allyl	O	0.0473 ^a	-10.39
B11	8-CN	dimethyl-allyl	S	0.0563 ^a	-10.29
B12	8-COH	dimethyl-allyl	S	0.188 ^a	-9.54
B13	9-Me	dimethyl-allyl	O	0.3142 ^a	-9.23
B14	8-C≡CH	dimethyl-allyl	O	0.4376 ^a	-9.02
B15	9-CF ₃	dimethyl-allyl	S	0.485 ^a	-8.96
B16	8-Me	dimethyl-allyl	O	0.989 ^a	-8.52
B17	10-Br	dimethyl-allyl	S	1.075 ^a	-8.47
B18	8-CN	dimethyl-allyl	O	1.1396 ^a	-8.43
B19	9-NO ₂	methyl- <i>c</i> -Pr	S	2.45 ^a	-7.96
B20	H	dimethyl-allyl	O	3.155 ^b	-7.81
B21	9-CF ₃	dimethyl-allyl	O	5.919 ^a	-7.42
B22	9-NO ₂	methyl- <i>c</i> -Pr	O	33.43 ^a	-6.35
9-Cl TIBO				0.034 ^a	-10.60

^a Ref 62. ^b Ref 8. B17 and B20 are excluded as outliers from the individual regression (eq 13) and from the build-up model (eqs 15–20). B15, B17, and B21 are excluded as outliers from the final regression model (eq 21). Activity (cell-based assay) is in micromolar at 37 °C. Estimated experimental binding energies $\Delta G_{\text{exptl}} \approx RT \ln(\text{activity})$ are in kcal/mol. ΔG_{exptl} range = 5.74, $N = 22$.

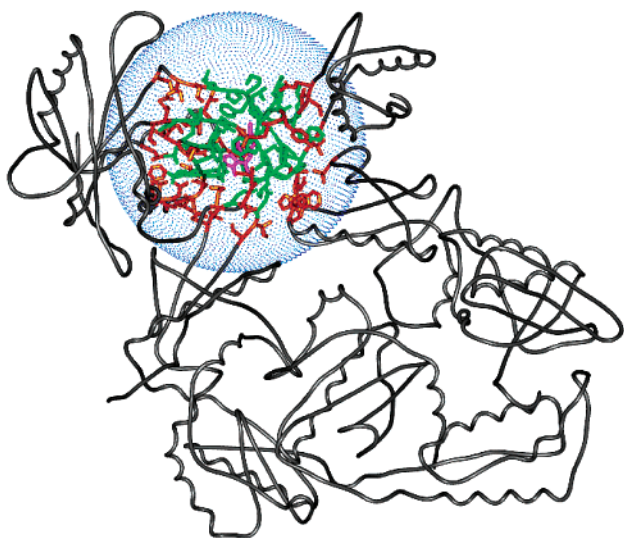


Figure 2. HIVRT binding site model solvated by a 22 Å cap of water. Green residues sampled in the MC simulations, orange residues are rigid, and gray residues are not used. Compound UC-781 (core 2) in magenta shows the location of the NNRTI binding site. Crystal structure coordinates (PDB entry 1rt4) are from ref 32.

[4,5,1-*jk*][1,4]benzodiazepine-2(1*H*)thione and -one (TIBO)³⁶ (core 8). The 2-amino-6-arylsulfonylbenzonitrile (ASBN, core 4) simulations were initiated from a model constructed by docking in the same manner as previously described for Sustiva.¹¹ The simulations for the

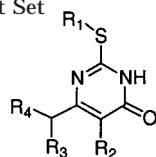
2-alkylthio-6-[1-(2,6-difluorophenyl)alkyl]-3,4-dihydro-5-alkylpyrimidin-4-(3*H*)one (DABO) analogues (test set) were initiated from the X-ray structure coordinates of HIVRT with MKC-442³¹ by manual docking. The initial conformations for bound ligands were generated using manual docking techniques or automatically using the GenMol program.³⁷ GenMol performs a conformational search and optimization for a ligand in a binding site; it then saves the lowest energy complexes for subsequent MC calculations or other analyses.

MC Simulations. Each protein–inhibitor complex was subjected to 50 steps of conjugate gradient (CG) energy minimization, using a distant-dependent dielectric constant of 4 ($\epsilon = 4r$), to relax the crystal structure prior to the MC simulations. For the MC simulations, a 22 Å water cap was used containing 851 (bound) and 1485 (unbound) TIP4P³⁸ water molecules centered on the ligand. All HIVRT side chains with an atom within ca. 10 Å from the center of the ligand were sampled, the protein backbone was fixed after the CG minimization, and each inhibitor was fully flexible. Bond lengths for the protein remained fixed after the initial energy minimizations. A protein residue–inhibitor list, which was kept constant during the entire simulation, was determined for each complex during the initial solvent equilibration stage of the simulation. A MC move for a side chain was randomly chosen every 10 configurations, while a move for the inhibitor was attempted every 56 configurations. All remaining moves were for solvent molecules. Solvent–solvent neighbor lists were also used, and the maximum number of internal coordinates to be varied for an attempted move was limited to 30. All MC simulations and energy calculations were performed with the MCPRO³⁹ program utilizing CM1A charges scaled by 1.08 with the OPLS-AA force field⁴⁰ for the ligand. Residue-based cutoffs at 9 Å were used for the solvent–solvent, solute–solvent, and intrasolute nonbonded interactions.

Bound MC simulations for protein–inhibitor complexes consisted of 10 million configurations of solvent-only equilibration, 10 million configurations of full equilibration, and 10 million configurations of averaging. Unbound MC simulations were subjected to an annealing protocol in order to reduce the effect of starting conditions on simulation results, as described in the earlier work.¹⁰ Briefly, 10 million configurations of solvent-only equilibration are performed followed by 5 million configurations in which only the water and the dihedral angles for each inhibitor are sampled. The MC acceptance rates for free inhibitors are enhanced at this stage by increasing the temperature to 727 °C (1000 K) for each attempted inhibitor move. The system is then subjected to an additional 5 million configurations of full equilibration at the normal temperature (25 or 37 °C), followed by 10 million configurations of averaging. The latter three processes are repeated for a total of five cycles.¹⁰

Experimental Activities. The experimental activities reported for each analogue series were converted into approximate free energies of binding (ΔG_{exptl}) using eq 4, which should correspond to relative free energies of binding for a closely related series of inhibitors.⁴¹

$$\Delta G_{\text{exptl}} \approx RT \ln(\text{activity}) \quad (4)$$

Table 9. S-DABO, Test Set

no.	R ¹	R ²	R ³	R ⁴	activity	ca. ΔG_{exptl}
D01	<i>i</i> -Pr	H	Me	Ph	0.9 ^a	-8.58
D02	Me	H	Me	2,6-di-F-Ph	0.18 ^a	-9.57
D03	<i>i</i> -Pr	H	Me	2,6-di-F-Ph	0.05 ^a	-10.35
D04	<i>n</i> -Bu	H	Me	2,6-di-F-Ph	0.07 ^a	-10.15
D05	Me	Me	Me	2,6-di-F-Ph	0.04 ^a	-10.50
D06	<i>i</i> -Pr	Me	Me	2,6-di-F-Ph	0.007 ^a	-11.57
D07	<i>n</i> -Bu	Me	Me	2,6-di-F-Ph	0.008 ^a	-11.49
D08	<i>i</i> -Pr	H	Et	Ph	0.8 ^a	-8.65
D09	<i>i</i> -Pr	H	Et	2,6-di-F-Ph	0.08 ^a	-10.07
D10	<i>i</i> -Pr	H	H	2,6-di-F-Ph	0.05 ^a	-10.36
D11	<i>i</i> -Pr	Me	H	2,6-di-F-Ph	0.05 ^a	-10.36
D12	<i>c</i> -hexyl	H	Me	Ph	0.6 ^a	-8.83
D13	<i>c</i> -hexyl	H	Me	2,6-di-F-Ph	0.16 ^a	-9.64
D14	<i>c</i> -hexyl	Me	Me	2,6-di-F-Ph	0.018 ^a	-10.99
D15	<i>c</i> -hexyl	H	Et	Ph	1.3 ^a	-8.35
D16	<i>c</i> -hexyl	H	Et	2,6-di-F-Ph	0.05 ^a	-10.36
D17	<i>c</i> -pentyl	H	Me	Ph	0.6 ^a	-8.83
D18	<i>c</i> -pentyl	H	Me	2,6-di-F-Ph	0.03 ^a	-10.68
D19	<i>c</i> -pentyl	Me	Me	2,6-di-F-Ph	0.006 ^a	-11.67
D20	<i>c</i> -pentyl	H	Et	Ph	1.0 ^b	-8.51
D21	<i>c</i> -pentyl	H	Et	2,6-di-F-Ph	0.15 ^a	-9.68
D22	<i>c</i> -pentyl	H	H	2,6-di-F-Ph	0.08 ^a	-10.07
D23	<i>c</i> -pentyl	Me	H	2,6-di-F-Ph	0.08 ^a	-10.07
D24	<i>s</i> -Bu	H	H	4-F-Ph	8.7 ^b	-7.18
D25	<i>s</i> -Bu	Me	H	4-F-Ph	11 ^b	-7.04
D26	<i>s</i> -Bu	Me	H	4-NO ₂ -Ph	2 ^b	-8.09
D27	<i>s</i> -Bu	H	H	4-NO ₂ -Ph	1.5 ^b	-8.27
MKC-442					0.03 ^{a,b}	-10.68
nevirapine					0.3 ^a	-9.26

^a Ref 45. ^b Ref 46. All compounds are included in predictions. Activity (cell-based assay) is in micromolar at 37 °C. Estimated experimental binding energies $\Delta G_{\text{exptl}} \approx RT \ln(\text{activity})$ are in kcal/mol. ΔG_{exptl} range = 4.63, $N = 27$.

For the NNRTIs, both cell-based (cores **1**, **2**, **4**, and **6–8**) and enzyme-based (cores **3** and **5a,b**) activities are reported (Tables 1–8). For three series (cores **4**, **6**, and **7**), both types of activity measurements were reported. Plotting cell-based vs enzyme-based activities for these three series yields correlation coefficients of $r^2 = 0.92$, 0.69, and 0.84, respectively, which implies reasonable agreement between the two types of independent measurements. Experimental binding free energies were then approximated for each data set (Tables 1–8) using the anti-HIV activities obtained from the cell-based assays, if available (cores **1**, **2**, **4**, and **6–8**). Otherwise, enzyme-based data was employed (cores **3** and **5a,b**).

Table 10. Estimated Experimental Binding Energies $\Delta G_{\text{exptl}} \approx RT \ln(\text{Activity})$ in kcal/mol for Compounds in Common (Anchors) among the Eight NNRTI Cores from Different Literature Sources^a

core (name)	literature ref	A	B	C	D
		ca. ΔG_{exptl} nevirapine	ca. ΔG_{exptl} 9-Cl TIBO	ca. ΔG_{exptl} Sustiva	ca. ΔG_{exptl} MKC-442
1 (HEPT)	48				-11.92
2 (carboxanilide)	51, 52	-10.76	-10.37		-11.35
3 (nevirapine)	53	-9.65			
4 (ASBN)	54	-10.01			
5a (Sustiva)	55			-12.84	
5b (Sustiva)	56, 57			-12.35	
6 (quinoxaline)	59	-9.48		-10.40	
7 (PETT)	60, 61	-9.68			
8 (TIBO)	62		-9.37		
test set (DABO)	45, 46	-9.26	-10.60		

^a Experimental values from Tables 1–9. The compounds in common give some indication as to how much a given data set should be shifted to bring the experimental values for the same compound from different literature sources in reasonable agreement. For example, the activities for nevirapine (column A) reported in the nevirapine, ASBN, quinoxaline, and PETT series (cores **3**, **4**, **6**, and **7**) agree within 0.5 kcal/mol. However, the corresponding value within the carboxanilide series (core **2**, column A) is ca. 1 kcal/mol lower, indicating that an offset of ca. +1 kcal/mol is needed. A similar difference in experimentally determined anti-HIV activities is observed with 9-Cl TIBO (column B, cores **2** and **7**). The 9-Cl TIBO activity reported in the TIBO series (core **8**, column B) is very close to the value reported in the carboxanilide series (core **2**, column B), suggesting a similar offset of +1 kcal/mol for the TIBO analogues. Similarly, the difference between the activities for Sustiva in cores **4** and **5a,b** (column C) indicates an offset of approximately -2 kcal/mol for core **5b**. Finally, the difference in activity of ca. +0.5 kcal/mol for MKC-442 between the HEPT and carboxanilide series (column D, cores **1** and **2**) adds to the carboxanilide difference suggested above, indicating an offset of ca. +1.5 kcal/mol for the HEPT series.

In some cases, the activity reported in different sources for a given compound varies by more than an acceptable amount (ca. 0.5 kcal/mol), which may be due to differences in assay conditions, protocols, and/or methods. To minimize the effects of these differences and allow regressions to include multiple data sets simultaneously, indicator variables (which can adopt values of 1 or 0) were employed where needed, with adjustable coefficients to shift the computed ΔG values. The experimental anti-HIV activities of compounds common to the different assays (Table 10) were used as a qualitative measure of which data sets need to be corrected and by how much. Combination fits that included HEPT (core **1**), carboxanilide (core **2**), Sustiva (core **5b**), and TIBO (core **8**) series employed adjustable shifting parameters for the indicator variables H_{corr} , U_{corr} , S_{corr} , and B_{corr} , respectively.

Results and Discussion

Individual Regression Equations. Correlations to each individual set of anti-HIV activities were pursued with the statistical software package JMP⁴² using the generic ELR regression, eq 3. Only statistically significant descriptors were included in the linear regressions. This condition was enforced by requiring that the probability greater than F ratios (regression model mean square/error mean square) is small (typically less than 0.005). Consequently, the probability of a greater F value occurring by chance ($\text{Prob} > F$) is low. It became apparent quickly that various descriptor combinations yield similar correlation coefficients (r^2 values) and that only one or two outliers in the data can adversely affect the significance of important descriptors. For this reason, as part of the standard fitting protocol, physically reasonable regression equations were sought that contained descriptors in common among the nine different data sets, subject to the constraint that up to 10% of the outliers would be eliminated. On average, the cross-correlation coefficient between any two of the descriptors used was less than 0.5. Outliers were selected and excluded based on the largest residual values, provided that the number of descriptors in common was maximized. Outlier compounds excluded from the individual regression equations (eqs 5–13), combined regressions (eqs 14–20), and final regression

Table 11. Significant Descriptors from Individual Fits of Each NNRTI Core (Eqs 5–13)^a

core	name	r^2	EXX_{LJ}	ΔHB_{tot}	water-bridges	qp_#rotor	$\Delta FOSA$	$\Delta WPSA$	$\Delta PISA$	$\Delta dipole$	qp_ ΔG_{hyd}	DtoProPi
1	HEPT	0.85	+	-			+					
2	carboxanilide	0.63	+					+		+		
3	nevirapine	0.54	+	-								-
4	ASBN	0.74	+		-	+						
5a	Sustiva	0.67	+					+				
5b	Sustiva	0.62							+			
6	quinoxaline	0.82	+	-		+						
7	PETT	0.66	+	-	-							
8	TIBO	0.79	+	-								

^a Descriptors in common share a + or - in the same column, which also indicates the sign of the fitted coefficient.

model (eq 21) are noted in Tables 1–8. Following this protocol, only 10 different descriptors are needed to describe the binding of the different NNRTI cores, as presented in regression eqs 5–13.

$$\Delta G_{calcd}(\text{core } \mathbf{1}) = 0.29\langle EXX_{LJ} \rangle - 1.3\langle \Delta HB_{total} \rangle + 0.013\langle \Delta FOSA \rangle + 4.7 \quad (5)$$

$$\Delta G_{calcd}(\text{core } \mathbf{2}) = 0.11\langle EXX_{LJ} \rangle + 0.015\langle \Delta WPSA \rangle + 0.32\langle \Delta dipole \rangle - 0.24\langle qp_{\Delta G_{hyd}} \rangle - 5.9 \quad (6)$$

$$\Delta G_{calcd}(\text{core } \mathbf{3}) = 0.24\langle EXX_{LJ} \rangle - 0.85\langle \Delta HB_{total} \rangle - 1.1\langle DtoProPi \rangle - 0.015 \quad (7)$$

$$\Delta G_{calcd}(\text{core } \mathbf{4}) = 0.40\langle EXX_{LJ} \rangle - 1.1\langle water-bridges \rangle + 0.81\langle qp_{\#rotor} \rangle + 6.7 \quad (8)$$

$$\Delta G_{calcd}(\text{core } \mathbf{5a}) = 0.35\langle EXX_{LJ} \rangle + 0.015\langle \Delta WPSA \rangle + 6.6 \quad (9)$$

$$\Delta G_{calcd}(\text{core } \mathbf{5b}) = 0.012\langle \Delta PISA \rangle - 0.11\langle qp_{\Delta G_{hyd}} \rangle - 11 \quad (10)$$

$$\Delta G_{calcd}(\text{core } \mathbf{6}) = 0.17\langle EXX_{LJ} \rangle - 0.79\langle \Delta HB_{total} \rangle + 0.31\langle qp_{\#rotor} \rangle - 5.6 \quad (11)$$

$$\Delta G_{calcd}(\text{core } \mathbf{7}) = 0.26\langle EXX_{LJ} \rangle - 0.91\langle \Delta HB_{total} \rangle - 1.7\langle water-bridges \rangle + 3.1 \quad (12)$$

$$\Delta G_{calcd}(\text{core } \mathbf{8}) = 0.53\langle EXX_{LJ} \rangle - 0.91\langle \Delta HB_{total} \rangle + 12 \quad (13)$$

Given the diversity among the NNRTI cores, it is unlikely that all of the descriptors would be significant for all of the regressions, although on the basis of our earlier ELR study of NNRTIs,¹⁰ we anticipated that the ligand–protein Lennard–Jones interaction energy (EXX_{LJ}) and the change in the total number of hydrogen bonds for the inhibitor upon binding (ΔHB_{total}) would correlate well with anti-HIV activities. These two terms consistently emerge as important factors that control binding (eqs 5–13, Table 11). Other significant descriptors were $\Delta FOSA$, the change in the hydrophobic SASA; $\Delta WPSA$, the change in the weakly polar (halogens, P, and S) SASA; $\Delta PISA$, the change in the aromatic SASA; $\Delta dipole$, the change in dipole moment of the inhibitor; $qp_{\Delta G_{hyd}}$, an estimate of the free energy of hydration for the inhibitor obtained using the QikProp program;⁴³ $DtoProPi$, the number of hydrogen bonds donated by the ligand to a protein π system;¹⁰ $water-bridges$, the number of bridging water molecules that mediate

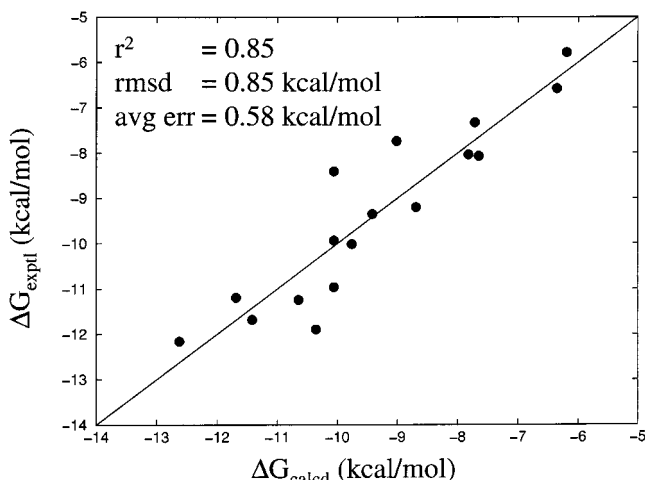


Figure 3. Predicted binding affinities (ΔG_{calcd}) computed using eq 5 vs experimental activities (ΔG_{exptl}) for HEPT analogues (core 1, Table 1) with HIVRT.

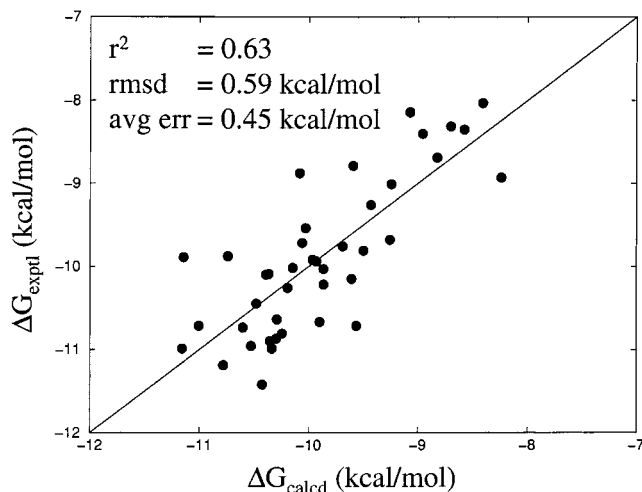


Figure 4. Predicted binding affinities (ΔG_{calcd}) computed using eq 6 vs experimental activities (ΔG_{exptl}) for carboxanilide analogues (core 2, Table 2) with HIVRT.

hydrogen bonding between ligand and protein; and $qp_{\#rotor}$, the number of rotatable bonds in the ligand. Correlation plots from the nine fits (eqs 5–13) of ΔG_{calcd} vs ΔG_{exptl} are presented in Figures 3–11 and show good accord between experiment and theory. An average correlation coefficient r^2 of 0.70, an average rms error of 0.67 kcal/mol, and an average unsigned error of only 0.50 kcal/mol were obtained across the NNRTI series. Thus, good fits are readily found for individual chemotypes using a small number of descriptors.

Descriptor Significance. In most cases, the descriptors used in the regressions make obvious physical

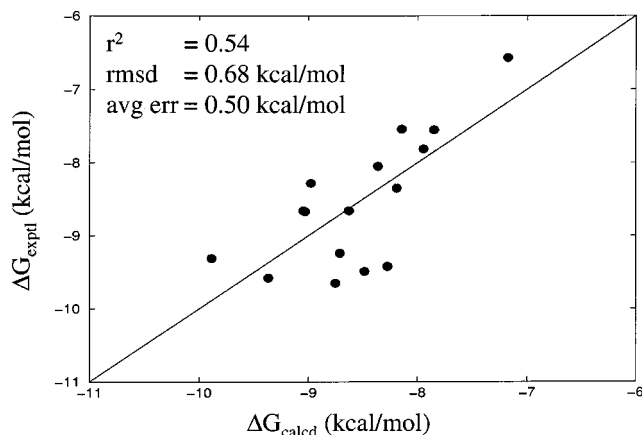


Figure 5. Predicted binding affinities (ΔG_{calcd}) computed using eq 7 vs experimental activities (ΔG_{exptl}) for nevirapine analogues (core 3, Table 3) with HIVRT.

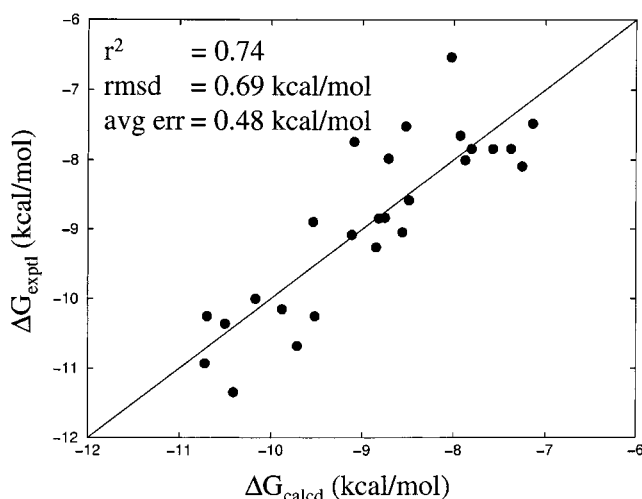


Figure 6. Predicted binding affinities (ΔG_{calcd}) computed using eq 8 vs experimental activities (ΔG_{exptl}) for ASBN analogues (core 4, Table 4) with HIVRT.

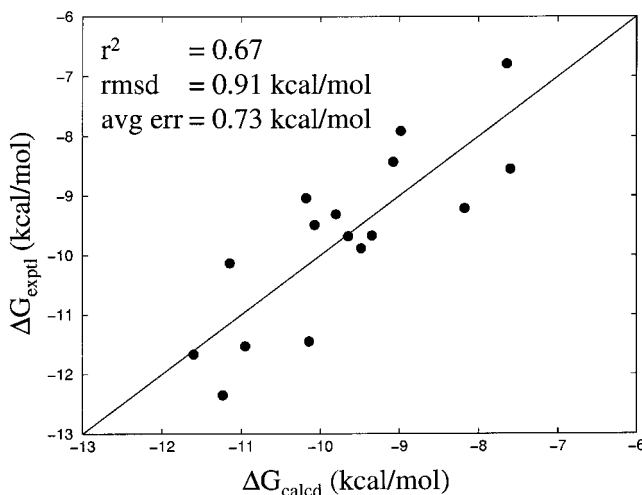


Figure 7. Predicted binding affinities (ΔG_{calcd}) computed using eq 9 vs experimental activities (ΔG_{exptl}) for Sustiva analogues (core 5a, Table 5a) with HIVRT.

sense. It should be noted that the descriptors used for the HEPT (core 1, eq 5) and nevirapine (core 3, eq 7) analogues are consistent with those reported in the earlier ELR study¹⁰ despite the fact that a different

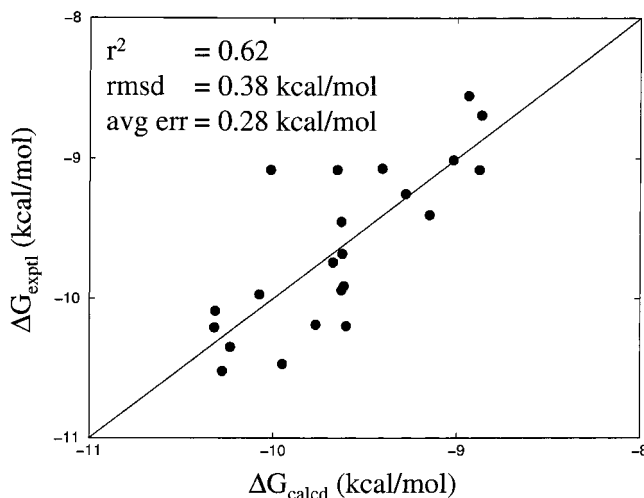


Figure 8. Predicted binding affinities (ΔG_{calcd}) computed using eq 10 vs experimental activities (ΔG_{exptl}) for Sustiva analogues (core 5b, Table 5b) with HIVRT.

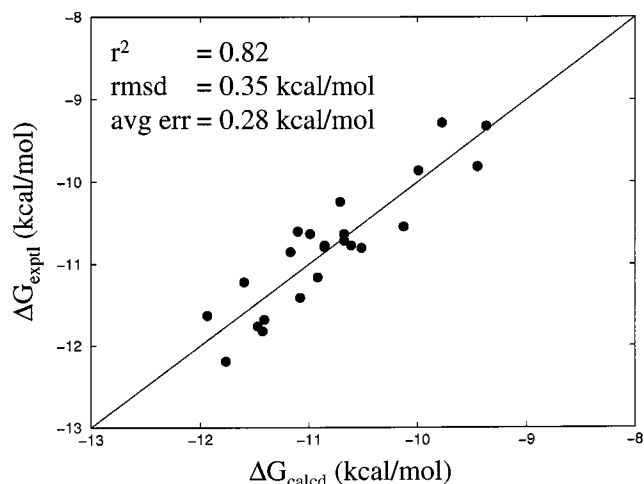


Figure 9. Predicted binding affinities (ΔG_{calcd}) computed using eq 11 vs experimental activities (ΔG_{exptl}) for quinoxaline analogues (core 6, Table 6) with HIVRT.

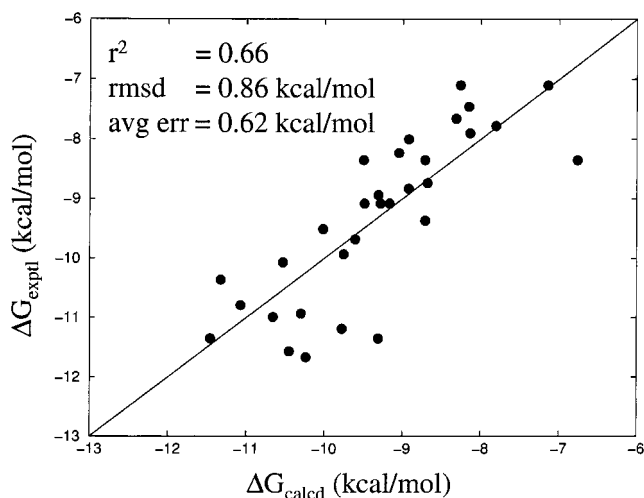


Figure 10. Predicted binding affinities (ΔG_{calcd}) computed using eq 12 vs experimental activities (ΔG_{exptl}) for PETT analogues (core 7, Table 7) with HIVRT.

charge model was used. Also, the *DtoProPi* descriptor in eq 7 replaces the secondary amide indicator used in the previous study to account for π type hydrogen bonds,

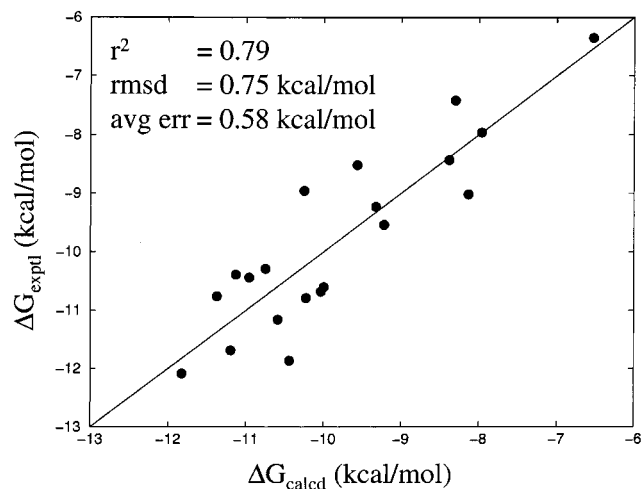


Figure 11. Predicted binding affinities (ΔG_{calcd}) computed using eq 13 vs experimental activities (ΔG_{expt}) for TIBO analogues (core **8**, Table 8) with HIVRT.

which were identified as important for the binding of nevirapines.¹⁰

Eight out of the nine regressions shown in eqs 5–13 contain the EXX_{LJ} descriptor. Because the EXX_{LJ} values are always negative, the associated positive coefficient implies that a good geometrical fit between the ligand and the protein is important. Favorable packing contributions to binding are contained in this term as well as any unfavorable steric interactions.

The next most common descriptor in eqs 5–13 is ΔHB_{total} (cores **1**, **3**, and **6–8**). In general, the change in the total number of hydrogen bonds for the inhibitor upon binding is negative given that water provides a better hydrogen-bonding environment than the protein and the fitted coefficients for ΔHB_{total} are negative. For these regressions, there is an energetic penalty between 0.79 and 1.3 kcal/mol for each net hydrogen bond lost in the binding event.

Two regressions contain the *water-bridges* descriptor (cores **4** and **7**). Here, the negative coefficient implies a gain in binding free energy of between -1.1 and -1.7 kcal/mol for each water-mediated protein–ligand hydrogen bond. This is consistent with the magnitudes of the coefficients for the other hydrogen-bonding descriptors, ΔHB_{total} and *DtoProPi*. Examples for the *DtoProPi* and *water-bridges* descriptors are presented in Figures 12 and 13. Note that the position of bridging water molecules observed in the MC simulations is consistent with those from the HEPT³¹ and nevirapine³³ X-ray structures despite the fact that crystallographic waters were not used.

Two regressions contain the *qp_#rotor* descriptor (cores **4** and **6**). This descriptor is expected to provide some measure of the entropy loss due to conformational restrictions imposed on the ligands upon binding. Here, a positive coefficient implies an energetic penalty of between 0.31 and 0.81 kcal/mol for each rotatable bond. Böhm et al.⁴⁴ reported a rotatable bond penalty of 0.34 kcal/mol based on a study of 45 different protein–ligand systems, and a penalty of 0.44 kcal/mol was obtained in the ELR study of thrombin inhibitors.³⁰ Inclusion of a logical descriptor like this for other series does not necessarily lead to a poorer regression; it simply does not emerge as statistically significant.

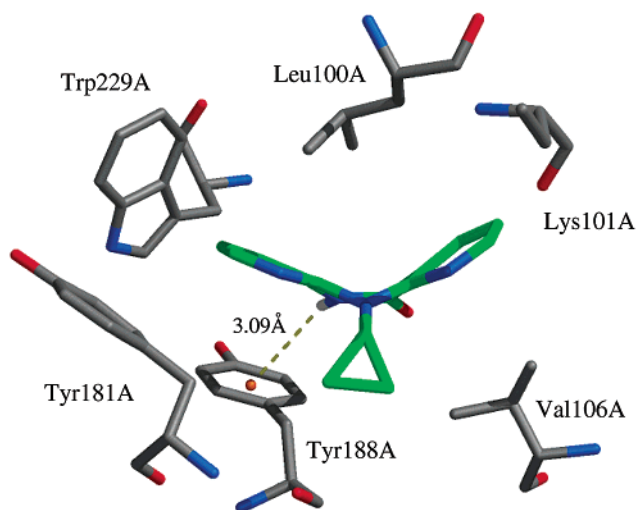


Figure 12. Example of a hydrogen bond donated by a ligand (nevirapine in green) to an aromatic ring in the NNRTI binding pocket (*DtoProPi* descriptor).

The burial of hydrophobic SASA ($\Delta FOSA$) found to be important here and in the previous ELR study for the HEPT series¹⁰ is easily interpreted as reflecting the hydrophobic effect given the nonpolar nature of the NNRTI binding site. Burial of aromatic SASA ($\Delta PISA$), an important descriptor for the Sustiva analogues from core **5b** (Table 5b), also makes physical sense given the large number of aromatic residues in the NNRTI site. The burial of weakly polar SASA ($\Delta WPSA$; halogens, sulfur, and phosphorus) emerges as important in fits for carboxanilides (eq 6) and Sustiva analogue **5a** (eq 9). For the halogens, this may also be reflective of their relatively hydrophobic nature.

The $\Delta dipole$ descriptor is significant only for the carboxanilide analogues (core **2**). The crystal structure complexes of carboxanilide analogues with HIVRT³² reveal a *cis* configuration around the amide/thioamide bond, although this bond is expected to be and was treated here as *trans* in the unbound state. The $\Delta dipole$ descriptor reflects the conformational change in the molecules between bound and unbound states, which ranges from 0.64 to 7.59 D for the carboxanilide series. The conformation of the thioamide bond in PETT analogues also adopts a *cis* configuration in the bound state. For this core, however, conformational analysis shows that the most likely configuration for unbound states is also *cis* due to the formation of an internal hydrogen bond between the pyridine or the thiazole ring nitrogen and a thiourea hydrogen (Scheme 1, Table 7). Other NNRTI cores are not expected to undergo such large changes in conformation between the two states.

Finally, the $qp_{\Delta G_{\text{hyd}}}$ descriptor emerges as important only for carboxanilide (core **2**) and Sustiva (core **5b**) compounds. This descriptor provides an estimate of the free energy of transfer from the gas phase into water and is always negative. Because the fitted coefficients in eqs 6 and 10 are negative, the most hydrophilic compounds from cores **2** and **5b** pay the largest desolvation penalty toward the computed binding free energy, as would be expected.

Combination Regression Equations. Correlations that included multiple data sets were pursued using a

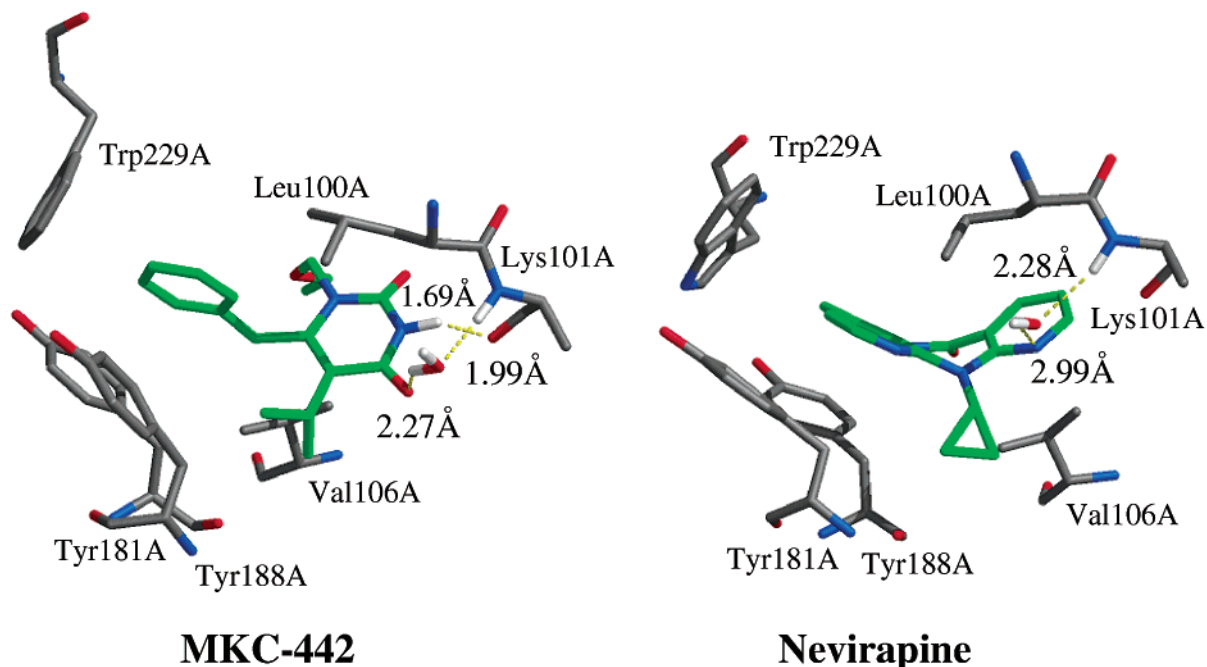


Figure 13. Two representative examples of water-mediated hydrogen bonds (*water-bridges* descriptor) between ligand and protein. Ligands are shown in green, MKC-442 is on the left, and nevirapine is on the right.

build-up approach in which the descriptors and signs of their coefficients were examined and the variation and type (cell or enzyme-based) of experimental data were considered. Table 11 lists the descriptors and signs of the coefficient for each descriptor in regression eqs 5–13. Table 10 lists the anti-HIV activities for the compounds in common (anchors) among the eight NNRTI cores from different literature sources.

The cell-based anti-HIV activities for nevirapine are in agreement to within ca. 0.5 kcal/mol from the data sets for the ASBN (core 4), quinoxaline (core 6), and PETT series (core 7), as shown in Table 10. Fitting these series together and incorporating the four descriptors common to the individual fits (eqs 8, 11, 12, Table 11) yields eq 14 with an r^2 of 0.66 for the 80 compounds. The same outlier compounds excluded from the individual fits were removed here during the build-up procedure. Other series were systematically added to this fit as described below.

$$\Delta G_{\text{calcd}} (\text{cores } 4, 6, 7) = 0.33\langle EXX_{LJ} \rangle - 0.44\langle \Delta HB_{\text{total}} \rangle - 0.69\langle \text{water-bridges} \rangle + 0.67\langle qp_#\text{rotor} \rangle + 1.9 \quad (14)$$

The individual fit for TIBO analogues (eq 13, Table 11) contains only two descriptors (EXX_{LJ} and ΔHB_{total}), both being included in eq 14. Therefore, no additional descriptors are required although an indicator variable (B_{corr}) was added to correct for the experimental variation, as reflected in Table 10 and described in the Experimental Activities section above. The resultant regression (eq 15) for the 100 compounds has an r^2 value of 0.66 and represents cores 4 and 6–8. The coefficient for the indicator variable in eq 15 has the expected sign and magnitude (+0.70 kcal/mol) for correcting the TIBO

data to the ASBN, quinoxaline, and PETT activities.

$$\Delta G_{\text{calcd}} (\text{cores } 4 \text{ and } 6-8) = 0.34\langle EXX_{LJ} \rangle - 0.48\langle \Delta HB_{\text{total}} \rangle - 0.72\langle \text{water-bridges} \rangle + 0.67\langle qp_#\text{rotor} \rangle + 0.70\langle B_{\text{corr}} \rangle + 2.4 \quad (15)$$

Incorporating HEPT (core 1) data into the fit required the addition of $\Delta FOSA$, which was found to be significant in the individual HEPT regression (eq 5, Table 11). An indicator variable was also added (H_{corr}) as suggested by the differences between MKC-442 activities and the other experimental data (Table 10). The resultant regression (eq 16) includes 117 compounds representing cores 1, 4, and 6–8 and has an r^2 value of 0.68. Here, the coefficient for the indicator variable H_{corr} has the correct sign but is too large based on the experimental data in Table 10, which suggests a value of 1.5 instead of the fitted 3.1 kcal/mol.

$$\Delta G_{\text{calcd}} (\text{cores } 1, 4, 6-8) = 0.31\langle EXX_{LJ} \rangle - 0.49\langle \Delta HB_{\text{total}} \rangle - 0.87\langle \text{water-bridges} \rangle + 0.58\langle qp_#\text{rotor} \rangle + 0.0033\langle \Delta FOSA \rangle + 0.84\langle B_{\text{corr}} \rangle + 3.1\langle H_{\text{corr}} \rangle + 2.0 \quad (16)$$

The carboxanilide data set (core 2) was incorporated into the combination fit by the addition of $\Delta WPSA$, $\Delta dipole$, and qp_G_{hyd} descriptors used for the individual fit (eq 6, Table 11) and the indicator variable U_{corr} . The resultant regression (eq 17) representing cores 1, 2, 4, and 6–8 now includes 156 compounds although the r^2 value degrades to 0.60. The coefficients for the indicator variables in eq 17 have the correct sign, although the magnitudes for H_{corr} and U_{corr} are 0.5–

0.75 kcal/mol larger than expected.

$$\Delta G_{\text{calcd}}(\text{cores } \mathbf{1}, \mathbf{2}, \mathbf{4}, \mathbf{6-8}) = 0.21\langle EXX_{LJ} \rangle - 0.12\langle \Delta HB_{\text{total}} \rangle - 0.78\langle \text{water-bridges} \rangle + 0.35\langle qp_ \#rotor \rangle + 0.0046\langle \Delta FOSA \rangle + 0.0089\langle \Delta WPSA \rangle + 0.18\langle \Delta dipole \rangle - 0.070\langle qp_ \Delta G_{\text{hyd}} \rangle + 0.65(B_{\text{corr}}) + 2.3(H_{\text{corr}}) + 1.5(U_{\text{corr}}) - 1.1 \quad (17)$$

The inclusion of Sustiva (cores **5a,b**) and nevirapine (core **3**) analogues into the combination fit was delayed in order to first derive correlations to anti-HIV activities obtained using the more common cell-based (ASBN, quinoxaline, PETT, TIBO, HEPT, and carboxanilide) rather than enzyme-based (Sustiva and nevirapine) assays. To include the Sustiva analogues from Table 5a, no indicator variable or additional descriptors should be necessary since eq 17 already incorporates all of the descriptors used in the individual fit (eq 9, Table 11). Furthermore, Table 10 indicates that the experimental anti-HIV activities for Sustiva from the ASBN and Sustiva study (Table 5a) are similar. The combination regression (cores **1**, **2**, **4**, **5a**, and **6-8**) shown in eq 18 has 172 compounds with an r^2 value of 0.56. The coefficients for the indicator variables continue to have the correct sign although the expected values for H_{corr} and U_{corr} are still about 0.5 kcal/mol too large.

$$\Delta G_{\text{calcd}}(\text{cores } \mathbf{1}, \mathbf{2}, \mathbf{4}, \mathbf{5a}, \mathbf{6-8}) = 0.22\langle EXX_{LJ} \rangle - 0.10\langle \Delta HB_{\text{total}} \rangle - 0.69\langle \text{water-bridges} \rangle + 0.36\langle qp_ \#rotor \rangle + 0.0038\langle \Delta FOSA \rangle + 0.011\langle \Delta WPSA \rangle + 0.16\langle \Delta dipole \rangle - 0.052\langle qp_ \Delta G_{\text{hyd}} \rangle + 0.59(B_{\text{corr}}) + 2.0(H_{\text{corr}}) + 1.6(U_{\text{corr}}) - 0.48 \quad (18)$$

Nevirapine analogues (core **3**) were then incorporated into the combination fit by adding in the *DtoProPi* descriptor. No indicator variable is required to correct the nevirapine data in Table 10. This combination fit (eq 19) represents eight cores **1-4**, **5a**, and **6-8** and yields an r^2 of 0.56 for 188 compounds. The coefficients for the indicator variables continue to have the correct sign although B_{corr} is now too small and no longer significant while U_{corr} is still 0.5 kcal/mol too large. H_{corr} is now close to the expected value of ca. 1.5 kcal/mol.

$$\Delta G_{\text{calcd}}(\text{cores } \mathbf{1-4}, \mathbf{5a}, \mathbf{6-8}) = 0.19\langle EXX_{LJ} \rangle - 0.15\langle \Delta HB_{\text{total}} \rangle - 0.70\langle \text{water-bridges} \rangle + 0.19\langle qp_ \#rotor \rangle + 0.0045\langle \Delta FOSA \rangle + 0.014\langle \Delta WPSA \rangle + 0.19\langle \Delta dipole \rangle - 0.079\langle qp_ \Delta G_{\text{hyd}} \rangle - 0.70\langle DtoProPi \rangle + 0.18(B_{\text{corr}}) + 1.7(H_{\text{corr}}) + 1.6(U_{\text{corr}}) - 1.1 \quad (19)$$

Finally, the last data set representing the second set of Sustiva analogues shown in Table 5b was added to the combination fit by incorporating one additional descriptor $\Delta PISA$, and indicator variable, S_{corr} . The resultant combination regression, eq 20, contains 210 compounds, has an r^2 value of 0.53, and has an rms error of 0.95 kcal/mol. As before, all coefficients for indicator variables continue to have the correct sign. The magnitude for S_{corr} to correct the Sustivas from

Table 5b is too small at -0.11 kcal/mol as compared with the expected value of ca. -2.0 kcal/mol (Table 10). For other offsets, B_{corr} is still too small, U_{corr} is still 0.5 kcal/mol too large, and H_{corr} has the expected value of ca. 1.5 kcal/mol.

$$\Delta G_{\text{calcd}}(\text{all cores}) = 0.18\langle EXX_{LJ} \rangle - 0.14\langle \Delta HB_{\text{total}} \rangle - 0.69\langle \text{water-bridges} \rangle + 0.18\langle qp_ \#rotor \rangle + 0.0041\langle \Delta FOSA \rangle + 0.013\langle \Delta WPSA \rangle + 0.18\langle \Delta dipole \rangle - 0.082\langle qp_ \Delta G_{\text{hyd}} \rangle - 0.55\langle DtoProPi \rangle - 0.00041\langle \Delta PISA \rangle + 0.17(B_{\text{corr}}) + 1.6(H_{\text{corr}}) + 1.5(U_{\text{corr}}) - 0.11(S_{\text{corr}}) - 1.8 \quad (20)$$

The regression results obtained from this build-up approach are condensed into Table 12. Here, the signs of each fitted coefficient in the combination regression remain constant, but as more data sets are included, the significance of the descriptors changes. Although the results are encouraging given the large number of compounds, refinements to the last regression (eq 20) were pursued in an attempt to increase the generality and robustness of the model. Relaxing the restrictions imposed in the build-up procedure on the number and type of descriptors yielded a new regression equation (eq 21, Table 12) with an improvement in r^2 from 0.53 to 0.60. The correlation with experiment for this fit is presented in Figure 14. Here, the rms error is 0.86 kcal/mol and the average unsigned error is only 0.69 kcal/mol for 210 compounds.

$$\Delta G_{\text{calcd}}(\text{all cores}) = 0.15\langle EXX_{LJ} \rangle - 0.22\langle \Delta HB_{\text{tot}} \rangle - 0.56\langle \text{water-bridges} \rangle + 0.24\langle qp_ \#rotor \rangle + 0.0061\langle \Delta FOSA \rangle + 0.015\langle \Delta WPSA \rangle + 0.14\langle \Delta dipole \rangle - 0.73\langle DtoProPi \rangle + 0.036\langle \Delta \text{eintra} \rangle + 1.2(H_{\text{corr}}) + 0.94(U_{\text{corr}}) - 1.3 \quad (21)$$

It should be noted that 10% of the total of 234 compounds were still removed as outliers (Tables 1-8); 42% of the outliers excluded from eq 21 are the same as from the individual fits. The coefficients for H_{corr} and U_{corr} indicator variables are now near the target values of 1.5 and 1 kcal/mol, respectively, and the regression intercept is small. Furthermore, the B_{corr} and S_{corr} offsets, as well as the $\Delta PISA$ and $qp_ \Delta G_{\text{hyd}}$ descriptors, are no longer significant. The two descriptors were significant only in the individual regressions for the carboxanilide (core **2**, eq 6) and Sustiva (core **5b**, eq 10) series. Given the small data range of experimental anti-HIV activities for these two data sets (Tables 2 and 5b), reasonable individual regressions could not be obtained using more common descriptors, such as EXX_{LJ} and ΔHB_{tot} (Table 11). Consequently, the $\Delta PISA$ and $qp_ \Delta G_{\text{hyd}}$ descriptors became insignificant when all eight series were considered in deriving the final regression model. A new significant descriptor (Δeintra) emerged, which represents the change in intramolecular strain energy of the ligand upon binding. Δeintra was also found to be useful in the previous thrombin studies.³⁰ Cross-correlation of descriptors in the final model was reexamined, and it was determined that the average value of all the pairwise correlation coefficients between the

Table 12. Regression Equations that Incorporate Multiple NNRTI Cores^a

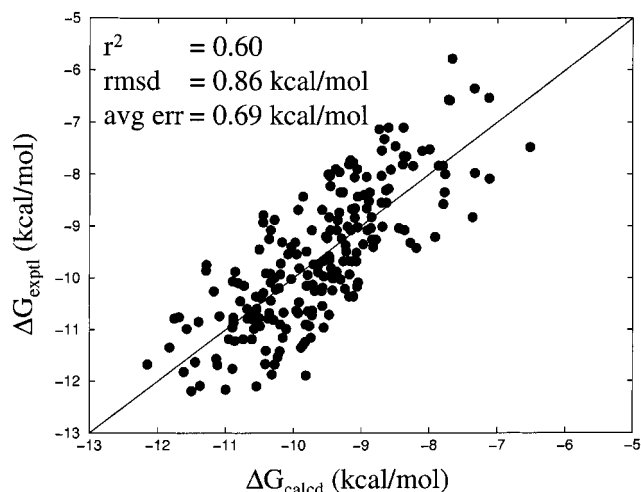
eq	cores	no.	r ²	EXX _{LJ}	ΔHB _{tot}	water-bridges	qp-#rotor	ΔFOSA	ΔWPSA	ΔPISA	Δdipole	ΔG _{hyd} ^{qp-}	DtoProPi	B	H	U	S	Δ _{intra}	
14	4, 6, 7	80	0.66	+	-	-	+												
15	4, 6-8	100	0.66	+	-	-	+								+				
16	1, 4, 6-8	117	0.68	+	-	-	+	+							+	+			
17	1, 2, 4, 6-8	156	0.60	+	-	-	+	+	+		+	-			+	+	+		
18	1, 2, 4, 5a, 6-8	172	0.56	+	-	-	+	+	+		+	-			+	+	+		
19	1-4, 5a, 6-8	188	0.56	+	-	-	+	+	+		+	-			+	+	+		
20	1-8	210	0.53	+	-	-	+	+	+	-	+	-			+	+	+	-	
21	all (refined)	210	0.60	+	-	-	+	+	+		+	-			+	+		+	

^a Descriptors in common share a + or - in the same column, which also indicates the sign of the fitted coefficient.

Table 13. Absolute and Relative^a Errors from Cross-Validated Anti-HIV Activity Predictions Made Using Eq 21

core	name	no. in training set	no. predicted	predicted absolute error (kcal/mol)	predicted relative error (kcal/mol)	predicted q ²	range of ΔG _{exptl}
1	HEPT	193	18	1.42	1.24	0.66	6.38
8	TIBO	191	22	0.89	0.87	0.48	5.74
test set	DABO	210	27	0.67	0.67	0.51	4.63

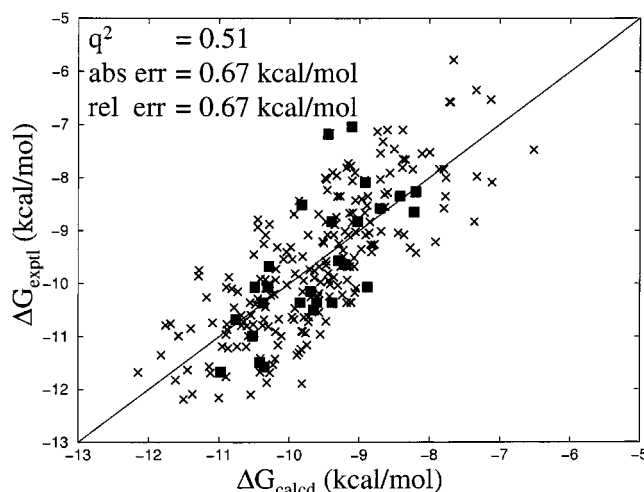
^a The relative errors are obtained by adding the difference of the mean between the predicted and the experimental values to the predictions for the current core.

**Figure 14.** Predicted binding affinities (ΔG_{calcd}) computed using eq 21 vs experimental activities (ΔG_{exptl}) for multiple NNRTIs (cores 1-8, Tables 1-8) with HIVRT.

nine descriptors and the two offsets was 0.17. Equation 21 represents the best regression model found that correlates well with all nine data sets using physically reasonable descriptors of binding. The absolute and relative errors are low given the size and diversity of the training set and the fact that only nine descriptors and two offsets are used (eq 21, Table 12). Smaller errors would be unrealistic given the fluctuations in the MC simulations, the uncertainties in the individual experimental activities, and the merging of the experimental data sets (Table 10).

Cross-Validated Anti-HIV Activity Predictions.

Cross-validation was pursued to gauge the predictive ability of the model for compounds not included in the training set. In the previous ELR studies, cross-validated correlation coefficients (q^2) were reported as obtained by the leave-one-out (LOO) procedure.^{10,28,30} In LOO, a series of fits to the activity data are generated leaving out a single inhibitor. The equation for each fit is then used to predict the activity for the compound left out of the fit, and the correlation of these predictions with the experimental values is computed and reported

**Figure 15.** Predicted binding affinities (ΔG_{calcd}) computed using eq 21 vs experimental activities (ΔG_{exptl}) for the DABO analogues (■, Table 9) with HIVRT shown together with predictions for the training set (×).

as q^2 . As in the prior studies, the q^2 value determined by the LOO procedure shows little degradation from the r^2 results. For the 210 NNRTIs, the cross-validated q^2 was 0.55, with an rms error of 0.89 kcal/mol and an average unsigned error of 0.73 kcal/mol as compared to experiment.

To better assess the quality and utility of the regression model developed here, the LOO idea was further extended by leaving out one of the nine NNRTI data sets, refitting the model for the remaining compounds, and then predicting the activities for the excluded data set. This is clearly a more difficult challenge. These calculations were carried out for the two cores with the largest activity ranges, HEPT (core 1) and TIBO (core 8). As summarized in Table 13, in both cases, correlations are obtained that would be useful in library design to filter more active from less active compounds. The q^2 values are 0.66 and 0.48 for the HEPT and TIBO series, respectively. It should be emphasized that the former outliers were not excluded from the data set being predicted for these tests. However, this represents

Table 14. Significant Descriptors from Individual QP Fits of Each NNRTI Core (Eqs 22–30) and from the Final QP Model (Eq 31)^a

core	name	r^2	qp_FISA	qp_volume	$qp_#rotor$	qp_PISA	qp_acctHB	qp_WPSA	$qp_ΔG_{hyd}$	U	B
1	HEPT	0.81	+	–							
2	carboxanilide	0.49	+		+	+					
3	nevirapine	0.66				+	+				
4	ASBN	0.76				+	+				
5a	Sustiva	0.52		–					–		
5b	Sustiva	0.72						+		–	
6	quinoxaline	0.52	+								
7	PETT	0.63		–					–		
8	TIBO	0.47		–	+						
all (refined)		0.42	+	–		+			–		+

^a Descriptors in common share a + or – in the same column, which also indicates the sign of the fitted coefficient.

a realistic application of the ELR method given that outlier compounds would not be known a priori.

Finally, the master regression model (eq 21) was used to predict the inhibitory activities for a test set of 27 DABO analogues (Table 9). This set was not considered in the development of the regression model. For 19 out of 27 compounds used in the test set, both cell-based and enzyme-based anti-HIV activities were reported.^{45,46} Because a cell- vs enzyme-based activities plot yields a correlation coefficient of $r^2 = 0.95$, activities determined by the more common cell-based assay were used to derive the approximate experimental binding free energies via eq 4. The plot of predicted vs experimental activities for the DABO compounds is shown in Figure 15 with a q^2 of 0.51 and an average unsigned error of only 0.67 kcal/mol. Residual values for most of the compounds are close to the uncertainties in the experimental measurements (ca. 0.5 kcal/mol), with the exception of analogues D24 and D25. The application of the master regression to successfully predict the activities of a new set of NNRTIs not included in the training set is encouraging.

QikProp Regressions. EXX_{LJ} consistently emerges as a significant descriptor of binding because it reflects the complementarity of the binding site of the protein and the structure of the ligands. To test whether such information is crucial in deriving scoring functions, additional calculations on the free ligands were performed using the QikProp (QP) program.⁴³ QikProp utilizes only the unbound ligand to calculate 35 molecular descriptors and properties, which were used here as regression descriptors. QikProp was executed on the ligand's structure from the last accepted move in the unbound MC simulations. The QP results lack structural information about the binding site of the protein and do not include statistical sampling. Following the same protocol described earlier in the Results and Discussion section, correlations to each individual set of anti-HIV activities were pursued first, utilizing the same constraint that up to 10% outliers would be eliminated. Individual regressions (eqs 22–30) did not contain descriptors in common for most of the data sets. However, reasonable correlations for each core were obtained using only seven QP descriptors with coefficients sharing the same sign among different data sets (with the exception of core **5b**, Table 14) with an average r^2 value of 0.62, an rms error of 0.76 kcal/mol, and an average unsigned error of 0.59 kcal/mol. The corre-

sponding results for eqs 5–13 were 0.70, 0.67, and 0.50 kcal/mol, respectively.

$$\Delta G_{\text{calcd}}(\text{core } \mathbf{1}) = 0.033(qp_FISA) - 0.014(qp_volume) + 0.65 \quad (22)$$

$$\Delta G_{\text{calcd}}(\text{core } \mathbf{2}) = 0.037(qp_FISA) + 0.0045(qp_PISA) + 0.51(qp_rotor) - 15 \quad (23)$$

$$\Delta G_{\text{calcd}}(\text{core } \mathbf{3}) = 0.0079(qp_PISA) + 0.94(qp_acctHB) - 16 \quad (24)$$

$$\Delta G_{\text{calcd}}(\text{core } \mathbf{4}) = 0.028(qp_PISA) + 0.55(qp_acctHB) - 19 \quad (25)$$

$$\Delta G_{\text{calcd}}(\text{core } \mathbf{5a}) = -0.01(qp_WPSA) - 0.0083(qp_volume) - 0.49 \quad (26)$$

$$\Delta G_{\text{calcd}}(\text{core } \mathbf{5b}) = 0.01(qp_WPSA) - 0.16(qp_ΔG_{\text{hyd}}) - 14 \quad (27)$$

$$\Delta G_{\text{calcd}}(\text{core } \mathbf{6}) = 0.026(qp_FISA) - 13 \quad (28)$$

$$\Delta G_{\text{calcd}}(\text{core } \mathbf{7}) = -0.022(qp_WPSA) - 0.0036(qp_volume) - 3.2 \quad (29)$$

$$\Delta G_{\text{calcd}}(\text{core } \mathbf{8}) = -0.028(qp_volume) + 1.8(qp_#rotor) + 13 \quad (30)$$

QP descriptors used in individual regressions include qp_FISA , qp_PISA , and qp_WPSA , representing the hydrophilic, aromatic, and weakly polar portions of the SASA, respectively; qp_volume , molecular volume; $qp_#rotor$, number of rotatable bonds for the ligand; qp_acctHB , number of hydrogen bonds accepted by the ligand; and $qp_ΔG_{\text{hyd}}$, estimated free energy of hydration. The logical build-up approach proven successful in deriving the master NNRTI regression (eq 21) was initially applied to the QP analysis but did not yield satisfactory results, primarily because only a few cores contain descriptors in common across the nine diverse data sets (Table 14). Therefore, a master QP regression (eq 31) was derived by including all of the significant descriptors from individual equations and the offsets to

Table 15. Absolute and Relative^a Errors from Cross-Validated Anti-HIV QP Activity Predictions Made Using Eq 31

core	name	no. in training set	no. predicted	predicted absolute error (kcal/mol)	predicted relative error (kcal/mol)	predicted q^2	range of ΔG_{expt}
1	HEPT	192	18	1.48	1.46	0.61	6.38
8	TIBO	193	22	1.09	1.05	0.30	5.74
test set	DABO	210	27	1.00	0.95	0.15	4.63

^a The relative errors are obtained by adding the difference of the mean between the predicted and the experimental values to the predictions for the current core.

take into account the experimental variations (Table 10).

$$\Delta G_{\text{calcd}}(\text{all cores}) = 0.0071(qp_FISA) + 0.0027(qp_PISA) - 0.0057(qp_WPSA) - 0.0059(qp_volume) + 1.2(U_{\text{corr}}) + 0.54(B_{\text{corr}}) - 4.9 \quad (31)$$

An r^2 of 0.42, an rms error of 0.98 kcal/mol, and an average unsigned error of 0.79 kcal/mol were obtained using only four QP descriptors and two offsets. The value of the intercept for this fit is significantly larger than the corresponding value in the master regression (eq 21), which can be attributed to the absence of parameters that include important protein–ligand interactions. Generally, large values of an intercept might indicate factors not accounted for in a regression model. Nevertheless, the magnitudes and signs of the descriptors are reasonable and physically intuitive. Ligands with large hydrophilic and aromatic SASA pay a higher desolvation penalty upon binding, and contributions from those terms to the overall ΔG_{bind} should be unfavorable. Weakly polar SASA emerges as a significant descriptor in the fit and, as before, might reflect the relatively hydrophobic nature of the halogens. The molecular volume descriptor indicates that the binding process is favored by many contacts between the ligand and the protein. However, in the absence of information about the nature of the binding site, such an interpretation is not straightforward. Interestingly, 46% of the outliers in both master regressions (eqs 21 and 31) are the same, and the two offsets in eq 31, U_{corr} and B_{corr} , are close to the expected values of 1 kcal/mol for cores **2** and **8** (Table 10). The correlation between experimental and predicted values for 210 NNRTIs using only QP descriptors is presented in Figure 16.

Cross-validation of the model was pursued following the same LOO protocol established previously. As expected, these LOO predictions yielded significantly degraded q^2 values and higher absolute and relative errors (Table 15). The q^2 obtained for the 210 NNRTIs using eq 31 is 0.37 with an rms of 1.00 kcal/mol and an average unsigned error of 0.82 kcal/mol. The values for the cross-validated correlation coefficients obtained by the extended LOO protocol for the HEPT and TIBO series are 0.61 and 0.30, respectively, with relatively high absolute and relative errors that differ negligibly (Table 15). Predictions were then made for the DABO series and yielded a q^2 value of only 0.15. These findings unequivocally support the value of the information on the protein–ligand complexes that is obtained in the ELR approach. In the absence of such information, the predictive ability of regression models is seriously compromised.

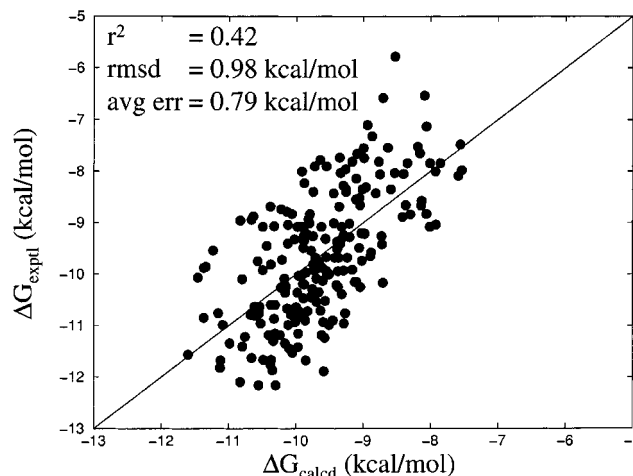


Figure 16. Predicted binding affinities (ΔG_{calcd}) computed using eq 31 vs experimental activities (ΔG_{expt}) for 210 NNRTIs (cores **1–8**, Tables 1–8) with HIVRT using QikProp descriptors.

Conclusion

Individual regression equations (eqs 5–13) have been developed for the estimation of binding affinities for eight series of NNRTIs by fitting results from MC simulations for each compound in the bound and unbound states to experimental anti-HIV activities. Good correlations with the experimental activities (Table 11, Figures 3–11) were obtained in all cases with an average r^2 value of 0.70, an rms error of 0.67 kcal/mol, and an average unsigned error of 0.50 kcal/mol, using only 10 unique descriptors (Table 11). In each case, no more than 10% of the compounds were removed as outliers (Tables 1–8). Thus, good fits can be obtained for individual series, as previously found.^{8,10,28,30} The most common descriptors that were significant for the individual regressions (Table 11) are the EXX_{LJ} and ΔHB_{tot} terms. These physically sensible parameters show that a good geometrical match between the inhibitor and the protein is important and that loss of hydrogen bonds with the inhibitor upon binding is unfavorable. Other quantities that were found to be important on a case-by-case basis are also physically reasonable and include the number of water bridges between inhibitor and protein (*water-bridges*), the number of rotatable bonds in the inhibitor ($qp_#rotor$), changes in SASA components ($\Delta FOSA$, $\Delta WPSA$, and $\Delta PISA$), changes in dipole moment ($\Delta dipole$), an estimate of the free energy of hydration of the inhibitor ($qp_ \Delta G_{\text{hyd}}$), and the number of H-bonds between the inhibitor and an aromatic ring in the protein (*DtoProPi*).

The regression equations developed for each data set were then combined (Table 12) by systematic incorporation of the important descriptors from each fit and by inclusion of an offset factor, if necessary, to account for

the uncertainties in merging the experimental data from different literature sources (Table 10). The last combination regression (eq 20) was refined in an attempt to generate a more robust equation that could be used to predict the activity of any NNRTI. Descriptors that were no longer significant ($\Delta PISA$ and $qp_{\Delta G_{hyd}}$) were removed, one additional descriptor ($\Delta eintra$) was added, and the outliers were redetermined. The resultant master regression model (eq 21) yielded a reasonable correlation ($r^2 = 0.60$, Figure 14) with the experimental activities for 210 NNRTIs using nine physically understandable descriptors and two offsets.

The ability of the refined model to predict anti-HIV activities for NNRTI series not included in the training set was investigated using a LOO cross-validation procedure. A reasonable cross-validated correlation coefficient ($q^2 = 0.55$) was obtained when each of the 210 NNRTIs was excluded from the training set, the master regression was refit, and the predictions were made for the compound left out. An extended LOO procedure was investigated in which an entire data set was excluded from the training set, the master regression was refit, and the predictions were made for the data set left out. Encouraging results were obtained for the HEPT ($q^2 = 0.66$) and TIBO ($q^2 = 0.48$) data sets, which represent the series with the largest activity ranges. Problematic issues for the analyses of all of the series include the fact that the experimental anti-HIV activities were obtained from multiple sources, the activity ranges for many data sets are small, and a few outliers in the data can remove the statistical significance of an important descriptor. These issues are expected to remain problematic in the absence of a large, accurate data set from a single experimental source. Encouragingly, the regression model can successfully predict anti-HIV activities for compounds not considered in the development of the master equation, as shown in Figure 15 for 27 DABO analogues. The reasonable q^2 of 0.51 and a low unsigned error of only 0.67 kcal/mol indicate the potential for application of the ELR method in library design to discriminate between active (potential leads) and less active/inactive compounds. The importance of the EXX_{LJ} and hydrogen-bonding terms was enforced through additional regression analysis using only ligand-based descriptors from QikProp. This approach revealed only moderate correlations ($r^2 = 0.42$, Figure 16) with the experimentally determined anti-HIV activities and poor predictive ability (Table 15) in the absence of structural information about the protein binding site. The lack of common descriptors (Table 14) among the nine data sets introduced additional difficulties in the development of the master regression (eq 31).

The present study has further explored the potential for computational methods to participate in inhibitor design with specific application to anti-HIV compounds. It has been demonstrated that results from computer simulations for NNRTIs and their complexes with HIVRT can yield physically reasonable regression-based models, which can be used to make useful predictions of binding affinities and activities. Extension of a current series is the most reliable, while jumping to a new series is also successful with minor degradation in the predictions.

Acknowledgment. Gratitude is expressed to Matthew P. Repasky for programming assistance, Dennis Ostrovsky for assistance in the simulation setup, Dr. Jayaraman Chandrasekhar and Dr. Mark A. Wilson for helpful discussions, and the National Institute of Allergy and Infectious Diseases (AI44616) for financial support.

References

- (1) Mitsuya, H.; Yarchoan, R.; Broder, S. Molecular Targets For Aids Therapy. *Science* **1990**, *249*, 1533–1544.
- (2) Katz, R. A.; Skalka, A. M. The Retroviral Enzymes. *Annu. Rev. Biochem.* **1994**, *63*, 133–173.
- (3) Tantillo, C.; Ding, J. P.; Jacobomolina, A.; Nanni, R. G.; Boyer, P. L.; Hughes, S. H.; Pauwels, R.; Andries, K.; Janssen, P. A. J.; Arnold, E. Locations of Anti-Aids Drug Binding Sites and Resistance Mutations in the 3-Dimensional Structure of HIV-1 Reverse Transcriptase: Implications For Mechanisms of Drug Inhibition and Resistance. *J. Mol. Biol.* **1994**, *243*, 369–387.
- (4) De Clercq, E. HIV Resistance to Reverse Transcriptase Inhibitors. *Biochem. Pharmacol.* **1994**, *47*, 155–169.
- (5) Smith, M. B. K.; Rouzer, C. A.; Taneyhill, L. A.; Smith, N. A.; Hughes, S. H.; Boyer, P. L.; Janssen, P. A. J.; Moereels, H.; Koymans, L.; Arnold, E.; Ding, J. P.; Das, K.; Zhang, W. Y.; Michejda, C. J.; Smith, R. H. Molecular Modeling Studies of HIV-1 Reverse-Transcriptase Nonnucleoside Inhibitors – Total-Energy of Complexation As a Predictor of Drug Placement and Activity. *Protein Sci.* **1995**, *4*, 2203–2222.
- (6) De Clercq, E. The role of nonnucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection. *Antiviral Res.* **1998**, *38*, 153–179.
- (7) Fletcher, R. S.; Arion, D.; Borkow, G.; Wainberg, M. A.; Dmitrienko, G. I.; Parniak, M. A. Synergistic inhibition of HIV-1 reverse transcriptase DNA polymerase activity and virus replication in vitro by combinations of carboxanilide nonnucleoside compounds. *Biochemistry* **1995**, *34*, 10106–10112.
- (8) Smith, R. H.; Jorgensen, W. L.; Tirado-Rives, J.; Lamb, M. L.; Janssen, P. A. J.; Michejda, C. J.; Smith, M. B. K. 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.
- (9) Smith, M. B.; Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L.; Michejda, C. J.; Ruby, S. K.; Smith, R. H., Jr. 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.
- (10) Rizzo, R. C.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of Binding Affinities for HEPT and Nevirapine Analogues with HIV-1 Reverse Transcriptase via Monte Carlo Simulations. *J. Med. Chem.* **2001**, *44*, 145–154.
- (11) Rizzo, R. C.; Wang, D.; Tirado-Rives, J.; Jorgensen, W. L. Validation of a Model for the Complex of HIV-1 Reverse Transcriptase with Sustiva through Computation of Resistance Profiles. *J. Am. Chem. Soc.* **2000**, *122*, 12898–12900.
- (12) Wang, D.; Rizzo, R. C.; Tirado-Rives, J.; Jorgensen, W. L. Antiviral Drug Design: Computational Analyses of the Effects of the L100I Mutation for HIV-RT on the Binding of NNRTIs. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2799–2802.
- (13) De Clercq, E. New developments in anti-HIV chemotherapy. *Farmacol.* **2001**, *56*, 3–12.
- (14) De Clercq, E. Novel compounds in preclinical/early clinical development for the treatment of HIV infections. *Rev. Med. Virol.* **2000**, *10*, 255–277.
- (15) Jorgensen, W. L. Free Energy Changes in Solution. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; Wiley: New York, 1998; Vol. 2, pp 1061–1070.
- (16) Lamb, M. L.; Jorgensen, W. L. Computational approaches to molecular recognition. *Curr. Opin. Chem. Biol.* **1997**, *1*, 449–457.
- (17) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (18) Jorgensen, W. L. Free Energy Calculations: A Breakthrough For Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, *22*, 184–189.
- (19) Åqvist, J.; Medina, C.; Samuelsson, J.-E. A New Method For Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng.* **1994**, *7*, 385–391.
- (20) Carlson, H. A.; Jorgensen, W. L. An Extended Linear Response Method For Determining Free Energies of Hydration. *J. Phys. Chem.* **1995**, *99*, 10667–10673.
- (21) McDonald, N. A.; Carlson, H. A.; Jorgensen, W. L. Free energies of solvation in chloroform and water from a linear response approach. *J. Phys. Org. Chem.* **1997**, *10*, 563–576.
- (22) Hansson, T.; Åqvist, J. Estimation of binding free energies for HIV protease inhibitors by molecular dynamics simulations. *Protein Eng.* **1995**, *8*, 1137–1144.

- (23) Paulsen, M. D.; Ornstein, R. L. Binding free energy calculations for P450cam-substrate complexes. *Protein Eng.* **1996**, *9*, 567–571.
- (24) Hulten, J.; Bonham, N. M.; Nillroth, U.; Hansson, T.; Zuccarello, G.; Bouzide, A.; Åqvist, J.; Classon, B.; Danielson, U. H.; Karlen, A.; Kvarnstrom, I.; Samuelsson, B.; Hallberg, A. Cyclic HIV-1 Protease Inhibitors Derived from Mannitol: Synthesis, Inhibitory Potencies, and Computational Predictions of Binding Affinities. *J. Med. Chem.* **1997**, *40*, 885–897.
- (25) Hansson, T.; Marelius, J.; Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27–35.
- (26) Wang, W.; Wang, J.; Kollman, P. A. What Determines the van der Waals Coefficient beta in the LIE (Linear Interaction Energy) Method to Estimate Binding Free Energies Using Molecular Dynamics Simulations? *Proteins* **1999**, *34*, 395–402.
- (27) Jones-Hertzog, D. K.; Jorgensen, W. L. Binding affinities for Sulfonamide Inhibitors with human Thrombin Using Monte Carlo Simulations with a Linear Response Method. *J. Med. Chem.* **1997**, *40*, 1539–1549.
- (28) Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of the binding affinities of FKBP12 inhibitors using a linear response method. *Bioorg. Med. Chem.* **1999**, *7*, 851–860.
- (29) Duffy, E. M.; Jorgensen, W. L. Prediction of Properties from Simulations: Free Energies of Solvation in hexadecane, Octanol and Water. *J. Am. Chem. Soc.* **2000**, *122*, 2878–2888.
- (30) Pierce, A. C.; Jorgensen, W. L. Estimation of Binding Affinities for Selective Thrombin Inhibitors via Monte Carlo Simulations. *J. Med. Chem.* **2001**, *44*, 1043–1050.
- (31) Hopkins, A. L.; Ren, J. S.; 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 nonnucleoside inhibitors. *J. Med. Chem.* **1996**, *39*, 1589–1600.
- (32) 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.
- (33) Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.; et al. High-resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat. Struct. Biol.* **1995**, *2*, 293–302.
- (34) Hsiou, Y.; Das, K.; Ding, J.; Clark, A. D., Jr.; Kleim, J. P.; Rosner, M.; Winkler, I.; Riess, G.; Hughes, S. H.; Arnold, E. Structures of Tyr188Leu mutant and wild-type HIV-1 reverse transcriptase complexed with the nonnucleoside inhibitor HBY 097: inhibitor flexibility is a useful design feature for reducing drug resistance. *J. Mol. Biol.* **1998**, *284*, 313–323.
- (35) 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) non-nucleoside inhibitors of HIV-1 and HIV-2 reverse transcriptases. Structural and biochemical analyses. *J. Biol. Chem.* **2000**, *275*, 5633–5639.
- (36) Ding, J. P.; Das, K.; Moereels, H.; Koymans, L.; Andries, K.; Janssen, P. A. J.; 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.
- (37) Jorgensen, W. L. *GenMol, version 1.5*; Yale University: New Haven, CT, 2001.
- (38) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions For Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (39) Jorgensen, W. L. *MCPRO, version 1.67*; Yale University: New Haven, CT, 2001.
- (40) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (41) Cheng, Y.; Prusoff, W. H. Relationship Between Inhibition Constant (K_i) and Concentration of Inhibitor Which Causes 50 Per Cent Inhibition (I_{50}) of an Enzymatic Reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099–3108.
- (42) Sall, J. *JMP, version 4*; SAS Institute Inc.: Cary, NC, 2000.
- (43) *QikProp, version 1.67*; Schrodinger, Inc.: New York, 2001.
- (44) Böhm, H.-J.; Klebe, G. What Can We Learn from Molecular Recognition in Protein–Ligand Complexes for the Design of New Drugs? *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 2588–2614.
- (45) Mai, A.; Sbardella, G.; Artico, M.; Ragno, R.; Massa, S.; Novellino, E.; Greco, G.; Lavecchia, A.; Musiu, C.; La Colla, M.; Murgioni, C.; La Colla, P.; Loddo, R. Structure-based design, synthesis, and biological evaluation of conformationally restricted novel 2-alkylthio-6-[1-(2,6-difluorophenyl)alkyl]-3,4-dihydro-5-alkylpyrimidin-4(3H)-ones as nonnucleoside inhibitors of HIV-1 reverse transcriptase. *J. Med. Chem.* **2001**, *44*, 2544–2554.
- (46) Mai, A.; Artico, M.; Sbardella, G.; Massa, S.; Novellino, E.; Greco, G.; Loi, A. G.; Tramontano, E.; Marongiu, M. E.; La Colla, P. 5-Alkyl-2-(alkylthio)-6-(2,6-dihaloophenylmethyl)-3,4-dihydropyrimidin-4(3H)-ones: novel potent and selective dihydro-alkoxybenzyl-oxypyrimidine derivatives. *J. Med. Chem.* **1999**, *42*, 619–627.
- (47) Tanaka, H.; Takashima, H.; Ubasawa, M.; Sekiya, K.; Nitta, I.; Baba, M.; Shigeta, S.; Walker, R. T.; Declercq, E.; Miyasaka, T. Synthesis and Antiviral Activity of Deoxy Analogues of 1-[(2-Hydroxyethoxy)methyl]-6-(Phenylthio)Thymine (HEPT) As Potent and Selective Anti-HIV-1 Agents. *J. Med. Chem.* **1992**, *35*, 4713–4719.
- (48) Tanaka, H.; Takashima, H.; Ubasawa, M.; Sekiya, K.; Inouye, N.; Baba, M.; Shigeta, S.; Walker, R. T.; Declercq, E.; Miyasaka, T. Synthesis and Antiviral Activity of 6-Benzyl Analogues of 1-[(2-Hydroxyethoxy)methyl]-6-(Phenylthio)Thymine (HEPT) As Potent and Selective Anti-HIV-1 Agents. *J. Med. Chem.* **1995**, *38*, 2860–2865.
- (49) Tanaka, H.; Baba, M.; Hayakawa, H.; Sakamaki, T.; Miyasaka, T.; Ubasawa, M.; Takashima, H.; Sekiya, K.; Nitta, I.; Shigeta, S.; Walker, R. T.; Balzarini, J.; Declercq, E. A New Class of HIV-1-Specific 6-Substituted Acyclouridine Derivatives: Synthesis and Anti-HIV-1 Activity of 5-Substituted or 6-Substituted Analogues of 1-[(2-Hydroxyethoxy)methyl]-6-(Phenylthio)Thymine (HEPT). *J. Med. Chem.* **1991**, *34*, 349–357.
- (50) Tanaka, H.; Takashima, H.; Ubasawa, M.; Sekiya, K.; Nitta, I.; Baba, M.; Shigeta, S.; Walker, R. T.; Declercq, E.; Miyasaka, T. Structure–Activity–Relationships of 1-[(2-Hydroxyethoxy)methyl]-6-(Phenylthio)Thymine Analogues: Effect of Substitutions at the C-6 Phenyl Ring and At the C-5 Position On Anti-HIV-1 Activity. *J. Med. Chem.* **1992**, *35*, 337–345.
- (51) Balzarini, J.; Brouwer, W. G.; Dao, D. C.; Osika, E. M.; De Clercq, E. Identification of novel thiocarboxanilide derivatives that suppress a variety of drug-resistant mutant human immunodeficiency virus type 1 strains at a potency similar to that for wild-type virus. *Antimicrob. Agents Chemother.* **1996**, *40*, 1454–1466.
- (52) Balzarini, J.; Brouwer, W. G.; Felauer, E. E.; De Clercq, E.; Karlsson, A. Activity of various thiocarboxanilide derivatives against wild-type and several mutant human immunodeficiency virus type 1 strains. *Antiviral Res.* **1995**, *27*, 219–236.
- (53) Hargrave, K. D.; Proudfoot, J. R.; Grozinger, K. G.; Cullen, E.; Kapadia, S. R.; Patel, U. R.; Fuchs, V. U.; Mauldin, S. C.; Vitous, J.; Behnke, M. L.; Klunder, J. M.; Pal, K.; Skiles, J. W.; McNeil, D. W.; Rose, J. M.; Chow, G. C.; Skoog, M. T.; Wu, J. C.; Schmidt, G.; Engel, W. W.; Eberlein, W. G.; Saboe, T. D.; Campbell, S. J.; Rosenthal, A. S.; Adams, J. Novel Nonnucleoside Inhibitors of HIV-1 Reverse-Transcriptase. 1. Tricyclic Pyridobenzodiazepinones and Dipyridodiazepinones. *J. Med. Chem.* **1991**, *34*, 2231–2241.
- (54) Chan, J. H.; Hong, J. S.; Hunter, R. N., III; Orr, G. F.; Cowan, J. R.; Sherman, D. B.; Sparks, S. M.; Reitter, B. E.; Andrews, C. W., III; 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-arylsulfonylbenzotriazoles as nonnucleoside reverse transcriptase inhibitors of HIV-1. *J. Med. Chem.* **2001**, *44*, 1866–1882.
- (55) Young, S. D.; et al. Benzoxazinones as Inhibitors of HIV Reverse Transcriptase. U.S. Patent 5,811,423, Sept. 22, 1998.
- (56) Patel, M.; Ko, S. S.; McHugh, R. J., Jr.; Markwalder, J. A.; Srivastava, A. S.; Cordova, B. C.; Klabe, R. M.; Erickson-Viitanen, S.; Trainor, G. L.; Seitz, S. P. Synthesis and evaluation of analogues of Efavirenz (Sustiva) as HIV-1 reverse transcriptase inhibitors. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2805–2810.
- (57) Patel, M.; McHugh, R. J., Jr.; Cordova, B. C.; Klabe, R. M.; Erickson-Viitanen, S.; Trainor, G. L.; Ko, S. S. Synthesis and evaluation of benzoxazinones as HIV-1 reverse transcriptase inhibitors. Analogues of Efavirenz (SUSTIVA). *Bioorg. Med. Chem. Lett.* **1999**, *9*, 3221–3224.
- (58) Rösner, M.; Billhardt-Troughton, U.-M.; Kirsch, R. K., J.-P.; Meichsner, C.; Riess, G. I. W. Quinoxalines, A Process for their Preparation and their Use. U.S. Patent 5,723,461, March 3, 1998.
- (59) Kleim, J. P.; Bender, R.; Kirsch, R.; Meichsner, C.; Paessens, A.; Rosner, M.; Rubsam-Waigmann, H.; Kaiser, R.; Wichers, M.; Schneweis, K. E.; et al. Preclinical evaluation of HBY 097, a new nonnucleoside reverse transcriptase inhibitor of human immunodeficiency virus type 1 replication. *Antimicrob. Agents Chemother.* **1995**, *39*, 2253–2257.
- (60) Bell, F. W.; Cantrell, A. S.; Hogberg, M.; Jaskunas, S. R.; Johansson, N. G.; Jordan, C. L.; Kinnick, M. D.; Lind, P.; Morin, J. M., Jr.; Noreen, R.; et al. Phenethylthiazolethiourea (PETT)

- compounds, a new class of HIV-1 reverse transcriptase inhibitors. 1. Synthesis and basic structure–activity relationship studies of PETT analogues. *J. Med. Chem.* **1995**, *38*, 4929–4936.
- (61) Cantrell, A. S.; Engelhardt, P.; Hogberg, M.; Jaskunas, S. R.; Johansson, N. G.; Jordan, C. L.; Kangasmetsa, J.; Kinnick, M. D.; Lind, P.; Morin, J. M., Jr.; Muesing, M. A.; Noreen, R.; Oberg, B.; Pranc, P.; Sahlberg, C.; Ternansky, R. J.; Vasileff, R. T.; Vrang, L.; West, S. J.; Zhang, H. Phenethylthiazolythiourea (PETT) compounds as a new class of HIV-1 reverse transcriptase inhibitors. 2. Synthesis and further structure–activity relationship studies of PETT analogues. *J. Med. Chem.* **1996**, *39*, 4261–4274.
- (62) Ho, W.; Kukla, M. J.; Breslin, H. J.; Ludovici, D. W.; Grous, P. P.; Diamond, C. J.; Miranda, M.; Rodgers, J. D.; Ho, C. Y.; Declercq, E.; Pauwels, R.; Andries, K.; Janssen, M. A. C.; Janssen, P. A. J. 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.* **1995**, *38*, 794–802.
- (63) Huang, H.; Chopra, R.; Verdine, G. L.; Harrison, S. C. Structure of a Covalently Trapped Catalytic Complex of HIV-1 Reverse Transcriptase: Implications for Drug Resistance. *Science* **1998**, *282*, 1669–1674.

JM010580Q