Molecular Modeling Calculations of HIV-1 Reverse Transcriptase Nonnucleoside Inhibitors: Correlation of Binding Energy with Biological Activity for Novel 2-Aryl-Substituted Benzimidazole Analogues

Marilyn B. Kroeger Smith,*[,†,‡] Brian M. Hose,[‡] Arie Hawkins,[‡] James Lipchock,[‡] David W. Farnsworth,[†] Robert C. Rizzo,[§] Julian Tirado-Rives,[§] Edward Arnold,[||] Wanyi Zhang,[||] Stephen H. Hughes,[†] William L. Jorgensen,[§] Christopher J. Michejda,[†] and Richard H. Smith, Jr.[‡]

National Cancer Institute−Frederick, Frederick, Maryland 21702, Department of Chemistry, McDaniel College, Westminster, Maryland 21157, Department of Chemistry, Yale University, New Haven Connecticut 06520, and Center for Advanced Biotechnology and Medicine (CABM) and Department of Chemistry, Rutgers University, Piscataway, New Jersey 08854

Received June 25, 2002

The energies and physical descriptors for the binding of 20 novel 1-(2,6-difluorobenzyl)-2-(2,6-difluorophenyl)benzimidazole analogues (BPBIs) to HIV-1 reverse transcriptase (RT) have been determined using Monte Carlo (MC) simulations. The crystallographic structure of the lead compound, 1-(2,6-difluorobenzyl)-2-(2,6-difluorophenyl)-4-methylbenzimidazole, was used as a starting point to model the inhibitors in both the bound and the unbound states. The energy terms and physical descriptors obtained from the calculations were correlated with their respective experimental EC_{50} values, resulting in an r^2 value of 0.70 and a root-mean-square deviation (rms) of 0.53 kcal/mol. The terms in the correlation include the change in total Coulombic energy and solvent-accessible surface area. Structural analysis of the data files from the BPBI calculations reveals that all of the analogues with good biological activity show the formation of a hydrogen bond between the ligand and the backbone nitrogen atom of lysine 103. By use of the structural results, two novel BPBI inhibitors have been designed and calculations have been carried out. The results show the formation of the desired hydrogen bonds, and the $\Delta G_{binding}$ values predict the compounds to be excellent RT inhibitors. Subsequent synthesis and biological activity testing of these analogues have shown the validity of the predictive calculations. If the BPBIs are modeled in a site constructed from the crystal coordinates of a member of another class of nonnucleoside inhibitors (the 4,5,6,7-tetrahydroimidazo[4,5,1-*jk*][1,4]benzodiazepine-2(1*H*)-thione and -one (TIBO) compounds), the correlation with the same terms drops slightly, giving an r^2 value of 0.61 with an associated root-mean-square value of 0.53 kcal/mol. Conversely, if the TIBO compounds are modeled in a site constructed from the BPBI complex crystal coordinates, a correlation can be obtained using the drug−protein interaction energy and change in the total number of hydrogen bonds, giving an r^2 value of 0.63. These are the same descriptors that were used for the TIBO compounds modeled in their own sites, where the r^2 value was 0.72. These data suggest that it may be possible, in some cases, to design novel inhibitors utilizing structural data from related, but not identical, inhibitors.

Introduction

In the past few years, there has been a major effort to design inhibitors that bind to and interfere with the function of one of the three key enzymes of the HIV virus: protease, integrase, and reverse transcriptase (RT). The last enzyme is an excellent target for drug design because it is essential for HIV replication but is not required for normal cell replication. Nonnucleoside inhibitors of this enzyme (NNRTIs) are especially attractive drug candidates because they do not function as chain terminators and do not bind at the dNTP site,[1,2] making them less likely to interfere with the normal function of other DNA polymerases and therefore less toxic than nucleoside inhibitors (NRTIs) such as AZT.

While many good inhibitors of the NNRTI class have been reported and three, including nevirapine, are used clinically, the discovery of new, more efficacious inhibitors is becoming increasingly important in light of the emergence of HIV strains that are resistant to the current drugs. Presently, the only way to circumvent this problem is to administer a drug cocktail, usually consisting of both NRTIs and NNRTIs together with protease inhibitors, to patients. It has thus become obligatory to design compounds that are active against key variants that emerge upon treatment with the inhibitors, in addition to wild-type. The end result of these point mutations is the decreased activity of the inhibitors, most likely due to decreased binding affinity.[3]

One possible way to reduce the development time for clinically useful NNRTIs is to utilize computational

* To whom correspondence should be addressed. Phone: (410)-857-2496. Fax: (410) 857-2497. E-mail: msmith@mcdaniel.edu.
† National Cancer Institutes−Frederick.
‡ McDaniel College.
§ Yale University.
|| Rutgers University.

10.1021/jm020271f CCC: $25.00
Published on Web 04/11/2003

methodology as a way to predict free energies of binding ($\Delta G_{binding}$) of candidate inhibitors prior to their synthesis. The computational approach presented here provides a relatively rapid way to estimate the activity of drug candidates prior to synthesis.

The present work uses Monte Carlo (MC) simulations, in combination with a linear response (LR) method, to estimate the free energy of binding of the inhibitor to the RT enzyme. Recent results suggest that MC calculations of this type may be particularly efficient for conformational sampling of protein structure.[4,5] The LR approach is much faster than the more traditional free energy perturbation (FEP) or thermodynamic integration (TI) methods for the conversion of the lead substrate into another analogue.[6] While inherently successful, these latter methods are computationally taxing because of the necessity of carrying out calculations for numerous intermediate stages of the transformation. The linear response approach, on the other hand, requires only one simulation each for the bound and the unbound states of each new inhibitor to produce a Boltzmann distribution of conformations. The resulting set of structures allows for assessment of changes in the orientation of the inhibitor following modifications in either the drug's or the protein's structure. Such information is invaluable in the design of new, more potent inhibitors.

As introduced by Åqvist,[6] the original LR method to estimate $\Delta G_{binding}$ used empirically derived parameters obtained from calculated van der Waals and Coulombic components of the energy of interaction of the solute with its environment. In some cases, a solvent-accessible surface area (SASA) term is also added to give an equation of the form[7,8]

$$\Delta G_{binding} = \alpha(\Delta E_{vdw}) + \beta(\Delta E_{coul}) + \gamma(\Delta SASA) \quad (1)$$

Simulations using the MCPRO[9] software, which incorporates this methodology, have previously been applied to a series of sulfonamide inhibitors of human thrombin[10] and neurotrophic inhibitors of FK506 binding protein.[11] In these LR calculations, the differences in the interactions between the ligand in the bound (complexed) and unbound states were successfully correlated with their biological activity. The fact that comparable results were obtained for FEP[11] and LR[5] calculations for a series of FKBP inhibitors showed the utility of the LR for lead optimization of drugs. LR calculations have also been able to reproduce the experimental binding energies of nonnucleoside 4,5,6,7-tetrahydroimidazo[4,5,1-*jk*][1,4]benzodiazepine-2(1*H*)-thione and -one (TIBO) inhibitors[12] of HIV-1 RT with excellent accuracy. Computationally derived free energies of binding for 12 TIBO analogues complexed with wild-type RT produced an excellent correlation with experimentally derived EC_{50} values (root-mean square (rms) error of 0.87 kcal/mol).[12] In addition, we have probed with good quantitative success using both LR and FEP methods,[13] the response of 8-Cl TIBO to the L100I and Y181C mutations commonly observed in patients following treatment. Most recently, the LR method has been extended to correlate the biological activity with descriptive LR terms for two[14] or more[15] series of NNRTIs.

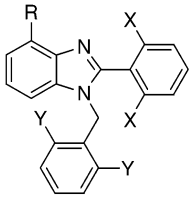
In the present study, the calculations have been extended to a new class of inhibitors, the 2-aryl-substituted benzimidazoles[16,17] (BPBIs). Preliminary crystallographic data[18] for the lead compound, 1-(2,6-difluorobenzyl)-2-(2,6-difluorophenyl)-4-methylbenzimidazole (analogue **1**, Table 1) with HIV-1 RT, allowed us to construct a model of the binding pocket. MC simulations were conducted on wild-type RT complexed with 20 previously synthesized 4-substituted BPBIs (Table 1). The aim of this work is to use computer simulations to obtain a correlation of binding energies for these compounds with their respective measure of biological activity, the EC_{50} values. A correlation of this type can then be used to predict the activity of proposed inhibitors that can be synthesized and tested for biological potency. We also wished to determine if the same correlation for the BPBIs could be made between their descriptors and the EC_{50} values in a site constructed from another inhibitor's crystal structure coordinate data. If this is possible, it might eliminate the potentially long wait for crystal structure data for novel classes of compounds.

Experimental Section

Protein and Inhibitor Parameters and Z-Matrix Generation. The potential energy functions used in these calculations have been previously described.[19] The OPLS (Optimized Potentials for Liquid Simulations) all-atom force field[19] for the protein was used in this study. The starting conformation for the lead inhibitor was taken from the crystal structure coordinates of the bound complex. The analogues used in this study have previously been reported[16] or are unpublished.[17] The structure of the inhibitor was modified as needed using the Spartan program[20] and was then subjected to an AM1 geometry optimization. The Z-matrix connectivity was graphically assigned, and the results were saved as a PEPZ[21] database and a BOSS[22] input file. The AM1-CM1A partial charges were then determined using BOSS. Atom types were determined, and all parameters for the inhibitors other than partial charges were assigned from the OPLS all-atom force field. Last, a fully flexible Z-matrix was then made for each inhibitor using the PEPZ program. Both bond stretching and angle bending parameters were assigned automatically using the atom-type combinations. The PEPZ program automatically assigned the corresponding parameters drawing from a dihedral database file, and any missing torsions were assigned by analogy to existing torsions in the file. Each inhibitor has its own nonbonded parameter file, the PEPZ database file, and is treated as a unique residue.

Construction of the Site. Representative models of the binding site of the protein were constructed from the crystal structure coordinates of the 4-Me-BPBI[18] or 8-Cl-TIBO[23] inhibitor complexes with wild-type HIV-1 RT by including all residues within approximately 15 Å of the inhibitor. To avoid excessive fragmentation of the protein chains, a few residues outside the original 15 Å cutoff were included in the model site. Hydrogen atoms were then added to the site, and the chains were capped with ACE and NME groups where appropriate so that all protein termini were neutral. The final system size, including capping groups, was 123 protein residues plus the inhibitor. To maintain charge neutrality for the entire system, all nonmoving (e.g., not varied during the simulation) Asp, Lys, Glu, and Arg residues were made neutral with the exception of p66 residue Asp 192, which remained charged. The proper tautomeric states of His residues were assigned by visual inspection. Both a fully flexible and a restricted protein−inhibitor Z-matrix file were made with PEPZ using the Z-matrix file for the inhibitor and the crystal coordinate file for the protein. The modified BPBI analogues were then superimposed over the lead compound, 4-methyl-BPBI, and subjected to manual docking to minimize any bad

Table 1. Structures and Experimental and Calculated ΔG Values for 1-(2.6-Difluorobenzyl)-2-(2,6-difluorophenyl)benzimidazole (BPBI) Inhibitors[a]



BPBI inhibitor	R substituent	Y substituent	X substituent	EC_{50} (μM)	ΔG, exptl (kcal/mol)	ΔG, calcd (kcal/mol)
1	CH_3	F, F	F, F	0.44	−9.03	−9.01
2	OCH_3	F, F	F, F	0.08	−10.08	−8.99
3	Cl	F, F	F, F	0.4	−9.09	−8.49
4	$NH-CH_3$	F, F	F, F	2.4	−7.99	−8.53
5	NO_2	F,F	F,F	0.45	−9.01	−9.31
6	NH_2	F, F	F, F	0.53	−8.92	−8.95
7	NH_2	F, F	2-OCH_3, 5-F	2.71	−7.91	−8.79
8	CH_3	H, H	F, F	0.75	−8.7	−9.43
9	H	F, F	F, F	1.7	−8.19	−7.9
10	CH_2-OH	F, F	F, F	0.05	−10.4	−9.64
11	CH_2-Cl	F, F	F, F	0.16	−9.65	−9.82
12	CH_2-NH_2	F, F	F, F	1.36	−8.33	−7.98
13	CH_2-CH_3	F, F	F, F	0.82	−8.64	−10.08
14	$C(O)NH_2$	F, F	F, F	0.37	−9.14	−8.46
15	$CH_2OC(O)CH_3$	F, F	F, F	0.06	−10.24	−10.29
16	C(O)H	F, F	F, F	0.1	−9.94	−9.44
17	$NHC(O)CH_3$	F, F	F, F	1.65	−8.23	−10.02
18	$N(CH_3)C(O)CH_3$	F, F	F, F	0.04	−10.51	−10.25
19	Br	F, F	F, F	0.37	−9.13	−9.34
20	H	F, F	H, H	6.06	−7.41	−7.94
A	CH_2-CN	F, F	F, F	0.06	−10.24	−9.48
B	CH_2-N_3	F, F	F, F	0.045	−10.43	−10.6

[a] Activity (cell-based assay) is in μM at 37 °C.

steric interactions before the protein−inhibitor Z-matrix was constructed. For the calculations of BPBI inhibitors in the 8-Cl-TIBO site, the various inhibitors were overlaid onto 8-Cl-TIBO in the correct orientation and 8-Cl-TIBO was deleted from the complex. The reverse of this procedure was followed for the TIBO derivatives in the 4-Me-BPBI site. For the MC simulations, residues outside a 10 Å sphere of the center of the inhibitor were kept rigid and included p66 residues 91−94, 109−110, 116−178, 184−185, 192−197, 199−205, 222−224, 230−232, 240−242, 316−317, 320−321, 343−349, and 381−383 and p51 residues 134−135, 137, and 140. Residues whose conformations were allowed to vary included p66 residues 95−108, 179−183, 186−191, 198, 225−229, 233−239, and 318−319 and p51 residues 136 and 138. The inhibitor was also free to move during the simulation. A 9 Å solvent−solvent, solvent−solute, and intrasolute nonbonded cutoff was used for all simulations.

Conjugate Gradient Energy Minimization. The fully flexible version of the Z-matrix for each protein−inhibitor complex was minimized briefly using 50 steps of conjugate gradient energy minimization with a distance-dependent dielectric constant of 4 ($\epsilon = 4r$) prior to the MC simulations to relax the crystal structure. The resulting coordinates were then used as the starting point for the MC simulations of the protein−inhibitor complexes (bound) or inhibitor (unbound).

Linear Response Calculations. All MC calculations were carried out with the MCPRO[9] (Monte Carlo simulation of proteins) software. For these simulations, a 22 Å water cap was used consisting of 851 TIP4P water molecules for the bound state and 1485 molecules for the unbound inhibitor. A 1.5 kcal/(mol Å^2) half-harmonic restraining force was applied to waters at the surface of the sphere to prevent possible evaporation. All protein side chains with any atom within ca. 10 Å of the cap atom, which was placed near the center of the binding site, were sampled, while all backbone atoms remained fixed; however, each inhibitor was fully variable. Bond lengths for the protein remained fixed after the initial minimization. A fixed protein residue−inhibitor list was specified for each simulation and determined for each complex during the initial solvent equilibration; the list was not updated during the simulation. The frequency of attempted moves for protein residue side chains was every 10 configurations, while an attempted move for each inhibitor molecule was made every 56 configurations; all remaining moves were for the solvent. Solvent−solvent neighbor lists that were periodically updated were also used during the simulations, while the maximum number of variables to be sampled for a given attempted move was set to 10.

Each MC simulation for a bound inhibitor consisted of 1 million configurations of solvent-only equilibration, 10 million configurations of full equilibration, and 10 million configurations of energy averaging. Separate running averages were obtained for the structural descriptors and the Lennard-Jones and Coulombic energy components for both inhibitor−solvent and inhibitor−protein nonbonded interactions. The standard deviation for each of these components was ≤1.5% of the mean. Unbound inhibitor simulations were carried out using an annealing protocol[14] after it was determined that the Coulumbic energies were not well converged when the same procedure was used as for the bound inhibitor. Each unbound MC simulation consisted of 1 million configurations of solvent-only equilibration at the experimental temperature of 37 °C. This was followed by 5 million configurations of equilibration in which the solvent and only the dihedral angles of the inhibitor were sampled. The MC acceptance rate for the inhibitor was increased through a local heating option in MCPRO for the inhibitor at the specified temperature of 727 °C (1000 K). This was then followed by an additional 5 million configurations of equilibration at the normal temperature (37 °C) and sampling schemes followed by 10 million configurations of averaging. This entire process was repeated from the heating phase for five cycles and resulted in well-converged energies for the inhibitor. As above, separate running averages were maintained for the various structural descriptors and the Lennard-Jones (U_{vdw}) and Coulombic (U_{elec}) energy components of the nonbonded interaction energies throughout the simulation.

Inhibitor Activity. The EC_{50} values for the model compounds in this study have previously been reported[16] or are

unpublished.[17,24] These values were converted to estimated experimental free energy values using the equation $\Delta G_{binding} = -RT \ln K_{form}$ at 37 °C. Here, K_{form}, the equilibrium constant for the formation of the complex, is taken to be the reciprocal of the EC_{50} value. This estimate was required because K_i data are not available for BPBI complexes with RT. Given the very close similarity in structure of these variants, however, it is reasonable to expect that EC_{50} and K_i, while not equivalent, should be linearly related.[25]

The difference in the structural descriptors and the energy terms between the bound and unbound inhibitors, along with a term for the change in the solvent-accessible surface area (SASA), was then calculated. The coefficients of the various descriptors and energy terms were then determined using a generalized linear response equation,

$$\Delta G_{binding} = \sum c_n \zeta_n + \text{constant} \quad (2)$$

where c_n represents an optimized coefficient for the associated descriptor ζ_n that is determined by a multiregression fit using the JMP program[26] to the experimental observable, the free energy of binding. A partial listing of the various descriptors that were evaluated includes terms such as EXX-LJ and EXX-coul, which are the van der Waals and electrostatic inhibitor−protein interaction energies, ΔFOSA or ΔPISA, which are the changes in the hydrophobic or aromatic SASAs, respectively, the number of hydrogen bonds donated by the ligand to the protein or to water, and the change in dipole moment of the compounds.

The linear regression analyses were carried out, and the descriptor sets were chosen to maximize the correlation coefficient, r^2, and to minimize the rms error with as few descriptors as possible. As a part of the standard fitting protocol, approximately 10% of the compounds could be eliminated as outliers. The statistical significance of the chosen descriptors was confirmed by analysis of the variance using the F ratios (regression model mean square/error mean square) and making sure that the probability of a greater F value occurring by chance (Prob > F) is less than 0.005.

The coordinate files from the simulations were analyzed using the Insight II module (MSI). Hydrogen bonds were determined using the preset distance measurement within this module (3.04 Å).

Synthesis and Activity Determination of New BPBIs. The synthesis of novel BPBI analogues was accomplished through slight modifications of known procedures.[16,17] The activity of the analogues was tested in an HIV-1-based vector pNLNgoMI/R⁻E assay, as has been previously described.[27] Briefly, the vector expresses the murine cells surface marker gene hsa from the nef reading frame to allow for the identification of cells infected with the vector. The env gene is inactivated in this vector. The vector is packaged into virus by cotransfecting 293 cells with plasmid encoding the vector and with the plasmid pHCMV-g, which expresses the VSV-g envelope, limiting the vector to a single cycle of retroviral replication.

Results and Discussion

Correlation of Experimental Binding Energies with MC Descriptors. The difference between the bound and unbound states for the various energetic and structural terms collected during the MCPRO runs were fit to the experimental observable, the free energy of binding. Recent work has shown that additional descriptors other than the purely energy-based terms may be useful in these correlations, especially where substituent changes occur in the more flexible portions of the drug, and protein interactions in these areas of the inhibitor remained relatively constant across the drug series. The basis for that notion stemmed from a study of the correlation of energetic terms and structural descriptors of ~200 organic molecules that were fit to free energies of solvation with excellent results.[28] Only four descriptors were needed to yield a correlation with an r^2 value of 0.91 and an rms error of 0.53.

Given the application of octanol/water partioning to protein/water systems and its effect on protein/ligand binding, the EC_{50} and IC_{50} values for 40 HIV-1 nonnucleoside inhibitors from the HEPT and nevirapine classes of compounds were recently correlated with the differences between the bound and unbound states for their respective energy terms and structural parameters,[14] according to the equation

$$\Delta G_{calc} = -0.94(\Delta\text{Hbtotal}) + 0.30(\text{EXX-LJ}) + 0.0085(\Delta\text{PHOBarea}) - 2.8(2° \text{ amide}) + 4.6 \quad (3)$$

The descriptors that emerged as best for correlating the combined data sets were (1) the protein−inhibitor Lennard-Jones energy term (EXX-LJ), (2) the change in the total number of hydrogen bonds for the ligand (ΔHBtotal), (3) the change in hydrophobic solvent-accessible surface area for the ligand (ΔPHOBarea), and (4) a correction term for the presence of 2° amide substituents in the ligand. The correlation coefficient, r^2, for this study was 0.75 with an rms deviation of 0.93 kcal/mol. The 20 HEPT analogues alone were found to correlate with the same terms (without the correction for amide substitutents because none were present in this class of compounds) with an r^2 value of 0.83. For the nevirapine analogues, however, only the ΔHBtotal and the 2° amide descriptors were significant, yielding an r^2 value of 0.58. A somewhat narrower range of available activities may account for the relatively low correlation for this latter set of compounds.

By use of this same approach, 20 BPBI inhibitors and their corresponding EC_{50} values were selected for analysis (Table 1). Following MCPRO simulations on the 4-substituted BPBI inhibitors, neither the standard ELR descriptors (ΔE_{LJ}, ΔE_{Coul}, and ΔSASA) nor the descriptors for the combined HEPT and nevirapine compounds yielded a correlation (r^2 = 0.25 or 0.11, respectively). However, a modified set of terms was found for these analogues that gave a reasonable correlation for 16 of the BPBI drugs. The calculated data for individual analogues are given in Table 2. The equation describing the fit is shown below:

$$\Delta G_{calc} = 0.101\langle\Delta\text{Coul}_{Total}\rangle + 0.027\langle\Delta\text{SASA}\rangle + 3.52 \quad (4)$$

This equation gave a correlation coefficient with an r^2 value of 0.70, an rms deviation of 0.53 kcal/mol, and an unsigned error of 0.46 kcal/mol when the experimental values were plotted against the calculated $\Delta G_{binding}$ values. The probability to F ratios (regression model mean square error divided by the error mean square) are all less than 0.005, implying that the probability of a greater F value occurring by chance is not great. The experimental and calculated ΔG values obtained from the simulations are given in Table 1 and a plot of these data is shown in Figure 1. Exclusion of any of the above terms from the fit significantly lowered the correlation coefficient (data not shown).

We decided to exclude compound **12** from the correlation in Figure 1. This compound is a primary amine and is probably protonated in aqueous solution. Its state of

Table 2. Change in Energy Values (kcal/mol) and Descriptors between Bound and Unbound States for BPBI Inhibitors

BPBI analogue	ΔESX-LJ	ΔESX-Coul	ΔEXX-LJ	ΔEXX-Coul	ΔLJ total	$\Delta Coul_{Total}$	ΔSASA	ΔPHOBarea	ΔPHILarea	ΔAROMarea	ΔHBondtot
1	26.2	28.37	−53.9	−8.29	−27.7	20.09	−541.82	−99.37	−12.71	−325.67	−1.12
2	28.39	24.19	−51.05	−8.41	−22.65	15.79	−526.36	−108.13	−12.65	−306.87	−0.76
3	25.41	24.98	−54.13	−10.23	−28.72	14.75	−502.92	−24.79	−14.72	−297.48	−0.95
4	23.84	39.76	−53.75	−6.31	−29.91	33.44	−570.05	−104.52	−36.29	−326.09	−1.73
5	26.12	30.45	−55.28	−16.6	−29.16	13.85	−531.91	−20.83	−107.74	−301.7	−0.88
6	23.54	31.98	−51.69	−17.97	−28.16	14.02	−518.47	−22.67	−73.15	−322.39	−1.26
7	26.38	33.04	−51.87	−16.7	−25.48	16.34	−520.35	−77.16	−72.85	−297.26	−0.76
8	25.52	27.57	−43.17	−9.4	−17.65	18.17	−552.03	−97.16	−11.53	−378.51	−1.28
9	25.75	21.86	−48.16	−10.66	−22.41	11.21	−467.19	−19.59	−22.41	−320.49	−0.96
10	24.97	33.01	−48.51	−13.65	−23.54	19.36	−564.43	−69.17	−62.36	−325.94	−2.7
11	27.42	24.63	−55.33	−9.5	−27.92	15.14	−556.61	−67.05	−12.2	−308.93	−0.95
12	21.62	51.78	−54.83	−21.19	−33.21	30.59	−538.39	−63.87	−70.73	−301.25	−1.07
13	28.9	27.41	−53.5	−7.73	−24.6	19.69	−582.75	−143.78	−11.62	−324.42	−0.93
14	24.55	43.05	−51.88	−17.32	−27.33	25.73	−540.08	−21.05	−116.86	−296.66	−2.67
15	29.74	58.3	−53.49	−28.29	−23.75	30.01	−627.47	−135.9	−72.29	−314.56	−3.68
16	25.98	22.79	−55.39	−12.6	−29.41	10.18	−524.32	−37	−87.8	−299.83	−0.89
17	31.21	29.43	−45.54	−15.79	−14.33	13.64	−568.35	−156.87	−52.1	−262.79	−1.54
18	28.35	27.39	−53.95	−3.82	−25.6	23.56	−565.67	−161.71	−8.36	−296.59	−1.02
19	25.7	23.73	−53.95	−9.5	−28.26	14.23	−534.45	−20.5	−15.58	−324.05	−0.94
20	25.49	23.77	−46.67	−11.83	−21.18	11.95	−471.37	−23.83	−23.33	−380.83	−1.04

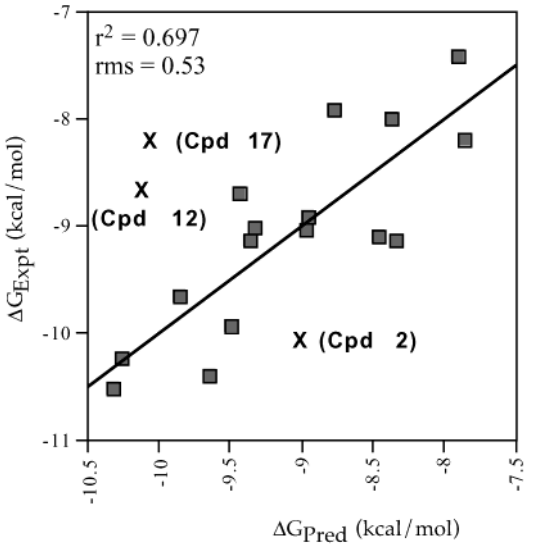


Figure 1. Predicted binding affinities (ΔG_{calc}) computed using eq 4 vs experimental activities (ΔG_{expt}) for 17 BPBI analogues with HIVRT.

protonation when bound to RT is uncertain, although, given the hydrophobic character of the pocket, it could be assumed to be unprotonated in the unbound form. Our earlier studies[12] on TIBO derivatives also contained some amine ligands, and we found that the unprotonated amino groups gave us an excellent fit. Nevertheless, the uncertainty about the state of protonation of **12** argues for its removal from the correlation, even though the results were not affected ($r^2 = 0.697$ without compound **12**; $r^2 = 0.695$ with **12**).

Three compounds, **2**, **13**, and **17**, fell outside the 2σ limit of the correlation. While it is difficult to identify the precise reasons for this, it is significant that the calculation predicts much higher activity for 13 and 17 than is actually observed. Both of these compounds can adopt several conformations, and it is tempting to speculate that the minimization procedure chooses the incorrect conformation. Structural studies currently in progress will help to resolve this issue. This is especially significant for compound **2**, which is a potent inhibitor of RT and whose activity is underestimated in the computational procedure.

The two terms involved in the correlation of the BPBI analogues include a term for the total change in Coulombic energy between the bound and unbound states $\langle \Delta Coul_{Total} \rangle$ and a term for the change in solvent-accessible surface area between the bound and unbound states $\langle \Delta SASA \rangle$. The energy term implies a dominant role for increased intersolute electrostatic interactions between the inhibitor and the protein, which is, of course, desirable upon binding. The SASA term indicates that the burying of the ligand, which is exposed to solvent in the unbound state, is favorable during complexation.

The descriptors that correlated with the experimental ΔG values for the BPBIs were not the same as those for the HEPT and nevirapine analogues mentioned above, although they are related. In fact, a recent study of NNRTIs using eight different NNRTI cores[15] found that each core had its unique descriptor set, which is not unresonable given the structural differences between classes of drugs. Nonetheless, almost all of the classes of NNRTIs had a term related to increasing the energy of interaction upon complexation. In addition, the burying of aromatic or hydrophobic areas was also seen to be highly significant, as was hydrogen bonding between the ligand and protein.

Thermodynamic and Structural Analysis for Substituted BPBIs. Compilation of energy values and descriptive structural terms during the MC calculations provides a mechanism for assessing the contribution of each of these parameters during the binding process. In addition, individual snapshots were saved every 500 000 steps during the MC simulation, providing structural information on the final inhibitor position, including any hydrogen bonds formed to the protein.

Analysis of the structure files reveals that the most of the inhibitors show a geometry repositioning during the Monte Carlo simulation that decreases the distances to two key protein side chain residues, K101 and K103 (Table 3). However, no correlation between activity and the change in this distance could be drawn. Clearly, geometry repositioning alone is not responsible for the activity of the compounds. Not much movement is seen for any of the drugs relative to the Y181, Y188, or W229

Table 3. Average Distance Measurements between Key Residues[a] of Selected BPBIs in the 4-Me-BPBI Wild-Type Crystal Structure before and after MCPRO Calculations

drug	K101−O (Å)	K103−CB (Å)	Y181−CE2 (Å)	Y188−CE2 (Å)	W229−CH_2 (Å)	ΔG_{calc} (kcal/mol)
1	−0.35	−0.6	0.2	0.13	0.18	−9.01
2	−0.52	−0.49	0.19	0.4	0.12	−8.99
3	−0.77	−0.52	0.28	0.29	0.28	−8.49
5	−0.33	−0.76	0.07	0.08	0.24	−9.31
8	−1.35	−1.07	0.34	0.07	0.36	−9.43
10	−0.9	−0.49	0.01	0.25	0.1	−9.64
11	0.57	−0.44	−0.12	0.48	0.11	−9.82
12	0.28	−0.84	0.23	0.28	0.25	−7.98
14	−1.2	−1.28	0.12	0.56	0.24	−8.46
16	−0.55	−0.41	−0.22	0.5	0.14	−9.44

[a] Distances are measured to the closest heavy atom in the inhibitor molecule. A positive number indicates movement further away from the respective amino acid residue. A negative number indicates movement closer to the amino acid residue.

residue. These aromatic residues have all been implicated in π-stacking interactions,[13,23] which add further stability to the bound inhibitors.

In addition to these general observations, the structure files reveal that all of the analogues that have an experimental $\Delta G_{binding}$ of better than −9.0 kcal/mol with an oxygen atom in the 4-substituent show the formation of a hydrogen bond between this O atom and the H−N of K103. One other member of the series, compound **12**, also shows a hydrogen bond to K103; however, its experimental ΔG value is only −8.33 kcal/mol. The substituent in this case is an amino methyl group; the acceptor N in the 4-position is less effective than an O. Analogue **11**, with 4-chloromethyl as a substituent and whose ΔG value is −9.65 kcal.mol, also has a substitutent that can function as a hydrogen bond acceptor, although it is generally not as effective as O or N. Analogue **6**, with a 4-NH_2 group, forms a hydrogen bond from the N atom to the backbone carbonyl oxygen atom of K101, as opposed to K103. The activity of this analogue was good but not excellent. Thus, all of the most efficacious inhibitors that have hydrogen-bonding acceptor atoms in the 4-position show hydrogen-bonding capability, which presumably adds stability to the drug's position in the binding pocket. An example of this capability is shown in Figure 2 for compound **16**. The information obtained from this type of analysis is extremely valuable in the design of proposed members of the BPBI series.

Prediction of Novel BPBI Inhibitors. Armed with a correlation of BPBI energy values and descriptors with EC_{50} values, a preliminary attempt was made to predict the activity of novel BPBI analogues against wild-type RT. To this end, two new analogues, 1-(2,6-difluorobenzyl)-2-(2,6-difluorophenyl)-4-cyanomethylbenzimidazole (compound **A**, Table 1) and 1-(2,6-difluorobenzyl)-2-(2,6-difluorophenyl)-4-azidomethylbenzimidazole (compound **B**, Table 1), were designed. The rationale for the selection of these compounds was based on the premise that these groups would enhance contacts with the K103 residue. They were placed into the binding site, and their binding energies were estimated using eq 4 above. Following the simulations, the $\Delta G_{binding}$ values of these analogues were estimated to be −9.48 and −10.6 kcal/mol, respectively, which corresponds to EC_{50} values of 0.075 and 0.035 μM. Subsequent synthesis, purification, and testing of RT activity showed that the EC_{50} values of the analogues were 0.06 and 0.045 μM, respectively. Therefore, the method for estimation of the efficacy of new analogues is quite accurate, and thus, the simulation of additional new analogues is currently in progress.

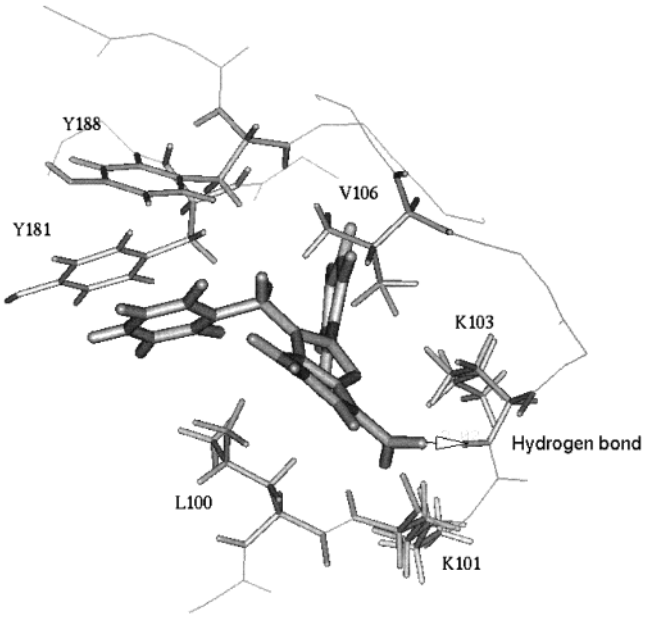
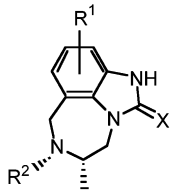


Figure 2. Representative example of hydrogen bonds between ligand and protein in the HIVRT binding pocket.

Prediction of Activities for BPBI Inhibitors in the 8-Cl-TIBO Binding Pocket or TIBO Inhibitors in the 4-Me-BPBI Binding Pocket. To determine if the differences in crystal structures of the respective HIV-1 RT complexes of the lead inhibitors play a role in predictive ability, MC simulations were carried out on the 8-Cl-TIBO complex in its native site so that the results could be compared to those of the 4-Me-BPBI simulation. Following these calculations, backbone superposition of the protein complexes constructed from the crystal structure coordinates of these inhibitors showed an rms deviation of only 0.87 Å. On the basis of this close structural similarity, the 20 known BPBI inhibitors were then placed into the RT binding pocket in place of 8-Cl-TIBO and, conversely, 22 TIBO derivatives (Table 4) were placed in the RT binding pocket in place of 4-methyl-BPBI. Following the MC calculations on all of these complexes, fitting of the resulting energy terms and descriptors to the respective experimental activities was again evaluated.

For the BPBI analogues placed in the TIBO binding pocket, using eq 4 gave a reasonable correlation (r^2 = 0.61; rms = 0.56 kcal/mol) for experimental vs calculated $\Delta G_{binding}$ if one additional analogue (BPBI 20) was

Table 4. Structures and Experimental Activities[29] for 4,5,6,7-Tetrahydroimidazo[4,5,1-*jk*][1,4]benzodiazepine-2(1*H*)-thione and -one (TIBO) Analogues



compd	R^1	R^2	X	activity[a] (μM)	ΔG_{expt} (kcal/mol)
T01	8-Br	dimethylallyl	S	0.003	−12.09
T02	8-Cl	dimethylallyl	S	0.0043	−11.87
T03	8-F	dimethylallyl	S	0.0058	−11.69
T04	8-Me	dimethylallyl	S	0.0136	−11.16
T05	9-F	dimethylallyl	S	0.025	−10.79
T06	9,10-di-Cl	dimethylallyl	S	0.0255	−10.76
T07	8-C≡CH	dimethylallyl	S	0.0296	−10.68
T08	9-Cl	dimethylallyl	S	0.034	−10.6
T09	H	dimethylallyl	S	0.044	−10.44
T10	8-Br	dimethylallyl	O	0.0473	−10.39
T11	8-CN	dimethylallyl	S	0.0563	−10.29
T12	8-COH	dimethylallyl	S	0.188	−9.54
T13	9-Me	diethylallyl	O	0.3142	−9.23
T14	8-C≡CH	dimethylallyl	O	0.4376	−9.02
T15	9-CF_3	dimethylallyl	S	0.485	−8.96
T16	8-Me	dimethylallyl	O	0.989	−8.52
T17	10-Br	dimethylallyl	S	1.075	−8.47
T18	8-CN	dimethylallyl	O	1.1396	−8.43
T19	9-NO_2	methyl-c-Pr	S	2.45	−7.96
T20	H	dimethylallyl	O	3.155	−7.81
T21	9-CF_3	dimethylallyl	O	5.919	−7.42
T22	9-NO_2	methyl-c-Pr	O	33.43	−6.35

[a] Activity (cell-based assay) at 37 °C.

dropped from the fit. In general, the final distances to K101 and K103 decreased for the good inhibitors in the TIBO site (see Table 5), while the poorer inhibitors moved closer to Y188 and W229. While the final position of the inhibitors in the binding pocket was not identical to that in the natural complex, this did not greatly affect the ability for correlation with the biological activity of these compounds.

For the TIBO analogues placed in the BPBI binding pocket, a reasonable correlation ($r^2 = 0.63$; rms = 1.00 kcal/mol) could be obtained for the calculated energy terms and descriptors used in the fit for 17 of the TIBO compounds in their own sites, according to eq 5, which uses the same descriptors as those determined previously:[15]

$$\Delta G_{calc} = 0.40\langle EXX_{LJ}\rangle - 1.41\langle \Delta HB_{total}\rangle + 4.13 \quad (5)$$

However, an alternative fit using 19 compounds could be obtained (eq 6, $r^2 = 0.75$, rms = 0.80 kcal/mol) if different descriptors were instead used:

$$\Delta G_{calc} = -0.03\langle \Delta PHIL_{area}\rangle - 0.74\langle \Delta HB_{tDtoWTRl}\rangle - 12.30 \quad (6)$$

In this equation, the useful terms are the change in hydrophilic area, which has a negative coefficient indicating that exposure of these areas is desirable upon binding, and the change in the number of hydrogen bonds donated to water, each of which adds 0.75 kcal/mol of energy upon complexation. The coefficient of this latter term is negative, signifying that water is the best hydrogen-bonding medium, and the loss of hydrogen bonds is thus unfavorable. Comparison of the locations of the compounds before and after calculations revealed that in general the best drugs moved closer to residue Y181 while moving further away from K101 and K103 (see Table 6).

The ability to predict the activity for members of a new class of inhibitors using existing crystal structure information is desirable in that it would greatly accelerate the drug design timetable. From the results pre-

Table 5. Average Distance Measurements between Key Residues[a] of BPBIs in the 8-Cl-TIBO Crystal Structure before and after MCPRO Calculation

drug	K101−O (Å)	K103−CB (Å)	Y181−CE2 (Å)	Y188−CE2 (Å)	W229−CH_2 (Å)	ΔG_{calc} (kcal/mol)
1	0.53	−0.43	0.52	−0.26	0.16	−9.01
2	−0.8	−0.82	0.02	0.52	−0.05	−8.99
3	−0.33	−0.36	1.1	1.01	1.76	−8.49
5	0.22	−0.42	−0.37	−0.59	−0.95	−9.31
8	−0.03	−0.2	−0.73	0.22	−0.54	−9.43
10	−1.14	−0.64	−0.23	−0.07	−0.29	−9.64
11	0.14	−0.7	0.62	0.48	1.63	−9.82
12	0.19	0.07	−0.57	−0.78	−0.69	−7.98
14	−1.07	1.22	0.54	2.78	1.63	−8.46
16	−1.18	−0.96	−0.39	3.59	−1.24	−7.94

[a] Distances are measured to the closest heavy atom in the inhibitor molecule. A positive number indicates movement further away from the respective amino acid residue. A negative number indicates movement closer to the amino acid residue.

Table 6. Average Distance Measurements between Key Residues[a] of Selected TIBOs in the 4-Me-BPBI Wild-Type Crystal Structure before and after MCPRO Calculations

drug	K101−O (Å)	K103−CB (Å)	Y181−CE2 (Å)	Y188−CE2 (Å)	W229−CH_2 (Å)	ΔG_{calc} (kcal/mol)
1	1.7	2.53	−0.48	0.3	−0.72	−9.01
2	1.02	0.87	0.13	−0.15	−0.09	−8.99
5	1.56	1.64	−1.19	−0.05	−0.08	−9.31
8	1.89	2.65	0.78	−1.78	−0.09	−9.43
10	0.38	2.09	0.3	0.06	−0.23	−9.64
11	0.55	0.51	0.72	−0.44	−0.31	−9.82
12	2.48	1.2	−0.13	0.47	−0.37	−7.98
14	−0.05	−0.49	0.87	−0.42	−0.7	−8.46
16	−0.51	−0.56	0.96	−0.47	−0.07	−9.44

[a] Distances are measured to the closest heavy atom in the inhibitor molecule. A positive number indicates movement further away from the respective amino acid residue. A negative number indicates movement closer to the amino acid residue.

sented above, it is apparent that it is possible to obtain reasonable correlations after placing the drugs in a site that is not constructed from crystal structure data of the lead member of a specific class. In addition, the information gleaned from the distance measurements allows for the formulation of new members of a class that might adjust for the gap in distance.

Conclusions

A good correlation was obtained between a measure of biological activity (EC_{50} values) and the various energy terms and descriptors for a training set of 20 known analogues of the new BPBI class against wild-type RT. These relationships may be useful for the prediction of the activity of additional novel analogues. On the basis of the structural analysis of the simulations, valuable information was obtained that will be useful in the design of new members of this class. Clearly, the ability of the BPBI inhibitors to form a hydrogen bond to the backbone nitrogen of lysine 103 in the protein contributes to drug potency and can be used in the design of novel analogues. Hydrogen bonding to backbone atoms in the protein has also been observed with the TIBO inhibitors[23] and with the HEPT series of drugs,[14] suggesting it may be a universal feature for good drug potency. By use of this information, two new compounds have been synthesized and tested for biological activity with excellent results.

In addition, the ability to use the RT−TIBO complex to predict the activity of BPBI analogues also lends support for this idea. This could circumvent the long wait for crystal structures of lead compounds for novel NNRTIs, a major problem in the past for drug prediction research. Proposed compounds could be simulated on the computer using available structural information and existing biological database information to obtain a correlation. The calculated data could then be used to predict the activity of compounds not only for wild-type but also for mutant RTs prior to synthesis of the inhibitors. This type of approach would certainly be beneficial in shortening the path to more universally potent AIDS drugs.

Acknowledgment. M.K.S. acknowledges NSF-PRF Grant 35097-B4 for financial support. E.A. thanks NIH for Grant NIGMS P01 56690 and NCI for Contract No. 75-1741 for financial support.

References

(1) Kopp, E. B.; Miglietla, J. J.; Shrutkowski, A. G.; Shih, C. K.; Grob, P. M.; Skoog, M. T. Steady state kinetics and inhibition of HIV-1 reverse transcriptase by a non-nucleoside dipyridodiazepinone, BI-RG-587, using a heteropolymeric template. *Nucleic Acids Res.* **1991**, *19*, 3035−3039.

(2) Romero, D. L.; Bussa, M.; Tan, C. K.; Reusser, F.; Palmer, J. R.; Poppe, S. M.; Aristoff, P. A.; Downery, K. M.; So, A. G.; Resnick, L.; Tarpley, W. G. Nonnucleoside Reverse Transcriptase Inhibitors That Potently and Specifically Block Human Immunodeficiency Virus Type 1 Replication. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 8806−8810.

(3) Kroeger Smith, M. B.; Michejda, C. J.; Hughes, S. H.; Boyer, P. L.; Janssen, P. A. J.; Andries, K.; Buckheit, R. W., Jr.; Smith, R. H., Jr. Molecular modeling of HIV-1 reverse transcriptase drug-resistant mutant strains: implications for the mechanism of polymerase action. *Protein Eng.* **1997**, *10*, 1379−1383.

(4) Essex, J. W.; Severance, D. L.; Tirado-Rives, J.; Jorgensen, W. L. Monte Carlo Simulations for Proteins: Binding Affinities for Trypsin−Benzamidine Complexes via Free-Energy Pertubations. *J. Phys. Chem. B* **1997**, *36*, 9663−9669.

(5) Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of binding affinity for FKBP12 inhibitors by a linear response method. *Bioorg. Med. Chem.* **1999**, *7*, 851−860.

(6) Åqvist, J.; Medina, C.; Samuelsson, J.-E. A new method for predicting binding affinity in comput-aided drug design. *Protein Eng.* **1994**, *7*, 385−391.

(7) Carlson, H. A.; Jorgensen, W. L. An extended linear response method for determining free energies of hydration. *J. Phys. Chem.* **1995**, *99*, 10667−10673.

(8) 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.

(9) Jorgensen, W. L. *MCPRO*, version 1.6.5; Yale University: New Haven, CT, 1998.

(10) Hertoz-Jones, 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.

(11) Lamb, M. L.; Jorgensen, W. L. Investigations of neurotropic inhibitors of FK506 binding protein via Monte Carlo simulations. *J. Med. Chem.* **1998**, *41*, 3928−3939.

(12) Smith, R. H., Jr.; Jorgensen, W. L.; Tirado-Rives, J.; Lamb, M. L.; Janssen, P. A. J.; Michejda, C. J.; Kroeger Smith, M. B. 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* (26), 5272−5286.

(13) Kroeger 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.

(14) Rizzo, R. C.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of binding affinities for HEPT and nevirapine analogs with HIV-1 reverse transcriptase via Monte Carlo simulations. *J. Med Chem.* **2001**, *44*, 145−154.

(15) Rizzo, R. C.; Blagovic, M. U.; Wang, D.-P.; Watkins, E. K.; Smith, M. B.; K.; Smith, R. H., Jr.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of Activity for Non-Nucleoside Inhibitors with HIV Reverse Transcriptase Based on Monte Carlo Simulations. *J. Med. Chem.* **2002**, *45*, 2970−2987.

(16) Roth, T.; Morningstar, M. L.; Boyer, P. L.; Hughes, S. H.; Buckheit, R. W., Jr.; Michejda, C. J. Synthesis and biological activity of novel nonnucleoside inhibitors of HIV-1 reverse transcriptase. 2-Aryl-substituted benzimidazoles. *J. Med. Chem.* **1997**, *40*, 4199−4207.

(17) Morningstar, M. L.; Roth, T.; Kroeger Smith, M. B.; Zajac, M.; Watson, K.; Bujacz, G.; Buckheit, R. W., Jr.; Michejda, C. J. Synthesis and biological activity of potent non-nucleoside inhibitors of HIV-1 reverse transcriptase that retain activity against mutant forms of the enzyme. Manuscript in preparation.

(18) Arnold, E.; Zhang, W. To be published.

(19) 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.

(20) *Spartan*, version 5.0.3; Wavefunction, Inc.: Irvine, CA, 1997.

(21) Tirado-Rives, J. *PEPZ*, version 1.0; Yale University: New Haven, CT, 1999.

(22) Jorgensen, W. L. *BOSS*, version 4.2; Yale University: New Haven, CT, 1998.

(23) Ding, J.; Das, K.; Moereels, H.; Koymans, L.; Andries, K.; Janssen, P. A. J.; Hughes, S. H.; Arnold, E. Structure of HIV-1 RT/TIBO R86183 complex reveals remarkable similarity in the binding of diverse nonnucleoside inhibitors. *Nat. Struct. Biol.* **1995**, *2*, 407−415.

(24) Buckheit, R. W., Jr. Personal communication.

(25) Cheng, Y.; Prusoff, W. H. Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50% inhibition (IC_{50}) of an enzymatic reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099−3108.

(26) *JMP*, version 3; SAS Institute, Inc.: Cary, NC, 1995.

(27) Julias, J. G.; Ferris, A. L.; Boyer, P. L.; Hughes, S. H. Replication of Phenotypically Mixed Human Immunodeficiency Virus Type 1 Virions Containing Catalytically Active and Catalytically Inactive Reverse Transcriptase. *J. Virol.* **2001**, *75*, 6537−6546.

(28) 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.

(29) Rizzo, R. C.; Jorgensen, W. L. OPLS All-Atom Model for Amines: Resolution of the Amine Hydration Problem. *J. Am. Chem. Soc.* **1999**, *121*, 4827−4836.

JM020271F