

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/12158980>

# Estimation of Binding Affinities for HEPT and Nevirapine Analogues with HIV-1 Reverse Transcriptase via Monte Carlo Simulations

ARTICLE *in* JOURNAL OF MEDICINAL CHEMISTRY · FEBRUARY 2001

Impact Factor: 5.45 · DOI: 10.1021/jm000255n · Source: PubMed

---

CITATIONS

100

---

READS

21

3 AUTHORS, INCLUDING:



Julian Tirado-Rives

Yale University

86 PUBLICATIONS 13,711 CITATIONS

SEE PROFILE

## Estimation of Binding Affinities for HEPT and Nevirapine Analogues with HIV-1 Reverse Transcriptase via Monte Carlo Simulations

Robert C. Rizzo, Julian Tirado-Rives, and William L. Jorgensen\*

Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107

Received June 13, 2000

The interactions and energetics associated with the binding of 20 HEPT and 20 nevirapine nonnucleoside inhibitors of HIV-1 reverse transcriptase (RT) have been explored in an effort to establish simulation protocols and methods that can be used in the development of more effective anti-HIV drugs. Using crystallographic structures as starting points, all 40 inhibitors were modeled in the bound and unbound states via Monte Carlo (MC) statistical mechanics methods. Potentially useful descriptors of binding affinity were configurationally averaged for each inhibitor during the MC simulations, and correlations were sought with reported experimental activities. A viable regression equation was obtained using only four descriptors to correlate the 40 experimental activities with an  $r^2$  of 0.75 and cross-validated  $q^2$  of 0.69. The computed activities show a rmsd of 0.94 kcal/mol in comparison with experiment and an average unsigned error of 0.69 kcal/mol. The MC results reveal three physically reasonable parameters that control the binding affinities: (1) loss of hydrogen bonds with the inhibitor is unfavorable, (2) burial of hydrophobic surface area is favorable, and (3) a good geometrical fit without steric clashes is needed for the protein–inhibitor complex. It is gratifying that the corresponding descriptors are statistically the most important quantities for determining the anti-HIVRT activity for the 40 compounds. Representative examples are also given in which structural and thermodynamic information from the MC simulations is used to help understand binding differences for related compounds. A key  $\pi$ -type hydrogen bond has been identified between secondary-amide nevirapine analogues and Tyr188A of HIVRT that explains their otherwise surprising activity and the ineffectiveness of nevirapine against the Y188C mutant.

### Background

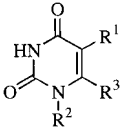
The human immunodeficiency virus (HIV), which has been identified as the causative agent of acquired immune-deficiency syndrome (AIDS),<sup>1</sup> infected about 15 000 people each day in 1999.<sup>2</sup> The World Health Organization and the Joint United Nations Programme on HIV/AIDS estimate that 16.3 million persons have died from the disease, 33.6 million people are currently infected with HIV, and over 95% of new infections are in developing countries.<sup>2</sup> The need for potent, safe, and inexpensive chemotherapeutics is clear, and the therapies must also be effective against mutant strains of HIV which arise from and circumvent existing anti-HIV treatments.<sup>3</sup>

One of the key enzymes packaged within the HIV virion capsid is a reverse transcriptase (RT) that plays an essential role in the replication of the virus.<sup>1,4,5</sup> Consequently, HIVRT has emerged as a prime target for the development of drugs for HIV/AIDS therapy.<sup>1,3</sup> The HIVRT protein has both RNA-dependent DNA polymerase and RNaseH activities that are required for the conversion of genomic viral RNA to DNA; this viral DNA is subsequently incorporated into the host cell genome.<sup>1,4,6</sup> Inhibitors of HIVRT fall into two main classes:<sup>3,6</sup> (1) Nucleoside inhibitors (NRTIs) are compounds that mimic normal nucleoside substrates but lack the 3'-OH group required for DNA chain elongation. NRTIs compete with native nucleosides and effectively

stall polymerase activity by becoming incorporated into the growing DNA strand thereby causing premature chain termination.<sup>3,6</sup> (2) Nonnucleoside inhibitors (NNRTIs) are molecules that bind to a region of HIVRT located near the polymerase catalytic site.<sup>6</sup> The binding event alters the conformation of critical residues and thereby inhibits the ability of the enzyme to perform normal RT functions.<sup>3</sup>

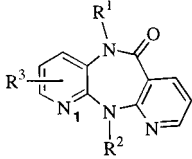
Although both NRTIs and NNRTIs dramatically decrease viral load in most infected persons on initiation of antiviral therapy, resistance to the chemotherapeutics invariably develops.<sup>3</sup> After the onset of infection, the virus replicates quickly within the host and a genetically related swarm (quasispecies) of virions is soon established.<sup>7,8</sup> This viral pool of variants arises rapidly mainly due to the low fidelity of HIVRT, which has been estimated to yield from 5–10 errors per HIV genome per round of replication.<sup>9,10</sup> Since as many as  $10^9$  virions are produced each day,<sup>11</sup> resistance to both nucleoside and nonnucleoside drugs quickly develops.<sup>3</sup> Since resistance arises in response to the chemotherapy, structurally unique inhibitors are needed that can challenge the swarm of virions in different ways. The use of combinations of NRTIs, NNRTIs, and HIV protease inhibitors is currently the best method for controlling HIV infection.<sup>12,13</sup> However, it is also desirable to have multiple inhibitors within a class since their unique modes of binding can lead to different resistance profiles. The present study has used computer simulations in an effort to develop protocols and methods that can

\* To whom correspondence should be addressed. Tel: 203-432-6278. Fax: 203-432-6299. E-mail: bill@adrik.chem.yale.edu.

**Table 1.** Inhibition of HIV-1 RT by HEPT Analogues


no.	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	EC <sub>50</sub>	ca. Δ <i>G</i> <sub>exptl</sub>
<b>H01</b>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	SPh	7.0 <sup>a</sup>	−7.32
<b>H02</b>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	SPh	3.6 <sup>a</sup>	−7.73
<b>H03</b>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>3</sub>	SPh	0.33 <sup>a</sup>	−9.20
<b>H04</b>	Me	CH <sub>2</sub> OCH <sub>3</sub>	SPh	2.1 <sup>a</sup>	−8.06
<b>H05</b>	Me	CH <sub>2</sub> OCH <sub>2</sub> Ph	SPh	0.088 <sup>a</sup>	−10.01
<b>H06</b>	<i>i</i> -Pr	CH <sub>2</sub> OCH <sub>2</sub> Ph	SPh	0.0027 <sup>a</sup>	−12.16
<b>H07</b>	Me	Et	SPh	2.2 <sup>a</sup>	−8.03
<b>H08</b>	Me	Me	SPh	>150.0 <sup>a</sup>	> −5.43
<b>H09</b>	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>3</sub>	SPh	0.019 <sup>a</sup>	−10.96
<b>H10</b>	<i>i</i> -Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>3</sub>	SPh	0.012 <sup>a</sup>	−11.24
<b>H11</b>	<i>i</i> -Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>3</sub>	CH <sub>2</sub> Ph	0.004 <sup>b</sup>	−11.89
<b>H12</b>	<i>c</i> -Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>3</sub>	SPh	0.1 <sup>a</sup>	−9.93
<b>H13</b>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	CH <sub>2</sub> Ph	23.0 <sup>c</sup>	−6.52
<b>H14</b>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	OPh	85.0 <sup>c</sup>	−5.78
<b>H15</b>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	SPh-3,5-diMe	0.26 <sup>d</sup>	−9.35
<b>H16</b>	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	SPh-3,5-diMe	0.013 <sup>d</sup>	−11.19
<b>H17</b>	<i>i</i> -Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	SPh-3,5-diMe	0.0027 <sup>d</sup>	−12.16
<b>H18</b>	Et	CH <sub>2</sub> OCH <sub>2</sub> Ph	SPh	0.0059 <sup>a</sup>	−11.68
<b>H19</b>	Me	H	SPh	>250.0 <sup>a</sup>	> −5.11
<b>H20</b>	Bu	Bu	SPh	1.2 <sup>a</sup>	−8.40

<sup>a</sup> Ref 14. <sup>b</sup> Ref 15. <sup>c</sup> Ref 16. <sup>d</sup> Ref 17. **H01** is parent HEPT; **H11** is MKC-442. EC<sub>50</sub> in μM at 37 °C. Estimated experimental binding energies Δ*G*<sub>exptl</sub> ≈ *RT* ln(EC<sub>50</sub>) in kcal/mol.

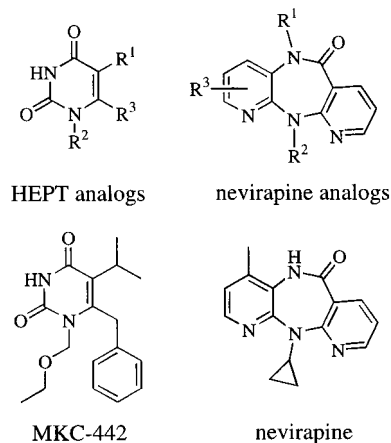
**Table 2.** Inhibition of HIV-1 RT by Nevirapine Analogues


no.	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	IC <sub>50</sub> <sup>a</sup>	ca. Δ <i>G</i> <sub>exptl</sub>
<b>N01</b>	Me	Et	H	0.125	−9.42
<b>N02</b>	Me	Et	2-Me	0.17	−9.24
<b>N03</b>	Me	Et	2-Cl	0.15	−9.31
<b>N04</b>	Me	Et	3-Me	0.76	−8.35
<b>N05</b>	Me	Et	3-Cl	>1.0	> −8.19
<b>N06</b>	Me	Et	4-Me	1.9	−7.81
<b>N07</b>	H	Et	H	0.44	−8.67
<b>N08</b>	H	Et	4-Me	0.035	−10.17
<b>N09</b>	H	Et	4-Cl	0.095	−9.58
<b>N10</b>	H	<i>c</i> -Pr	4-Me	0.084	−9.65
<b>N11</b>	Me	<i>c</i> -Pr	4-Me	>1.0	> −8.19
<b>N12</b>	Me	Pr	H	0.45	−8.66
<b>N13</b>	Me	<i>t</i> -Bu	H	11.0	−6.77
<b>N14</b>	Me	COCH <sub>3</sub>	H	15.3	−6.57
<b>N15</b>	Me	Et	4-Et	0.11	−9.49
<b>N16</b>	Me	CH <sub>2</sub> SCH <sub>3</sub>	H	0.85	−8.28
<b>N17</b>	H	<i>c</i> -Pr	4-CH <sub>2</sub> OH	3.0	−7.54
<b>N18</b>	H	<i>c</i> -Pr	4-CN	1.25	−8.05
<b>N19</b>	Me	CH <sub>2</sub> CH <sub>2</sub> F	H	2.9	−7.56
<b>N20</b>	H	<i>c</i> -Pr	H	0.45	−8.66

<sup>a</sup> Ref 18. **N10** is nevirapine. IC<sub>50</sub> in μM at 25 °C. Estimated experimental binding energies Δ*G*<sub>exptl</sub> ≈ *RT* ln(IC<sub>50</sub>) in kcal/mol.

be used in the design of improved anti-HIV drugs. In particular, computations have been carried out for the binding affinities of 40 analogues of the NNRTIs, HEPT, and nevirapine (Tables 1 and 2). Nevirapine was the first FDA-approved NNRTI, and the HEPT analogue MKC-442 is in clinical trials.

Specific goals of the research are twofold: (1) estimation of binding affinities in the context of structure-based drug design using available experimental data and (2) understanding the variations in binding affini-



ties through interpretation of energetic and structural results from simulations.

## Theoretical Methods

The most rigorous computational approaches used for the calculation of binding affinities (Δ*G*<sub>b</sub>) are the free energy perturbation (FEP) and thermodynamic integration (TI) methods.<sup>19–22</sup> These methods typically employ molecular dynamics (MD) or Monte Carlo (MC) simulations and have yielded impressive results for a number of protein–ligand systems, as reviewed elsewhere.<sup>19–22</sup> A more approximate method for the estimation of Δ*G*<sub>b</sub> is based on linear response (LR) theory, as introduced by Åqvist and co-workers (eq 1).<sup>23</sup> This approach is considerably faster than the FEP or TI alternatives because no intermediate transformation process is required to calculate the binding affinity.<sup>23</sup>

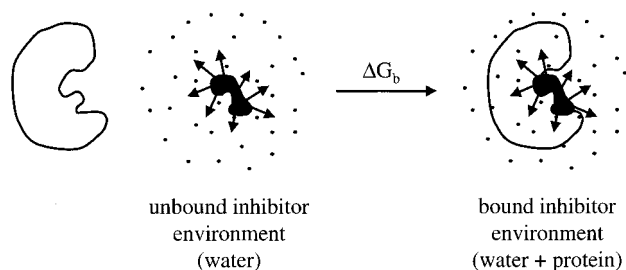
$$\Delta G_b = \alpha \langle \Delta E_{\text{vdw}} \rangle + \beta \langle \Delta E_{\text{Coul}} \rangle \quad (1)$$

Here,  $\langle \rangle$  signifies an ensemble average of the difference in interaction energies (Δ*E*) of the inhibitor–solvent plus inhibitor–protein interaction energies in the bound state and of the inhibitor–solvent interaction energies in the unbound state (Figure 1).<sup>23</sup> 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 β = 0.50, while the coefficient α was determined by fitting the simulation results to known experimental binding affinities.<sup>23</sup>

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.<sup>24,25</sup> 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 for solute cavity formation in water.<sup>24,25</sup>

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

Encouraged by prior MD/LR<sup>23,26–30</sup> and MC/LR<sup>31–33</sup> binding studies, we endeavored to treat larger data sets to see if good correlations with experimental data could



**Figure 1.** Schematic representation of a binding event showing different environments for HIVRT inhibitors. Small arrows depict potential interactions of a drug with water (unbound state) or water and protein (bound state).

still be obtained with a higher ratio of data points to parameters. Simultaneously, Duffy and Jorgensen have correlated results from aqueous MC simulations with solvation properties for more than 200 diverse organic compounds.<sup>34</sup> 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.

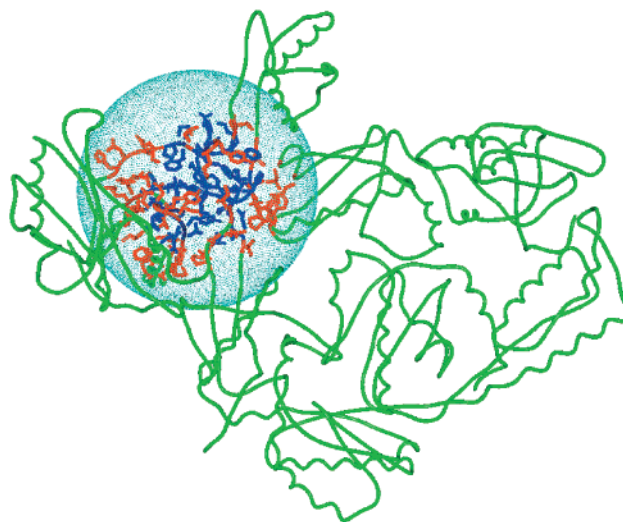
$$\Delta G_b = \sum_n c_n \xi_n + \text{constant} \quad (3)$$

Here,  $c_n$  represents an optimizable coefficient for the associated descriptor  $\xi_n$ . In principle, any physically reasonable quantity could be considered 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 an  $r^2$  of 0.91 and a rms error of 0.53.<sup>34</sup> Given the potential parallel between solute octanol–water partitioning and ligand protein–water partitioning, we sought to consider alternative descriptors too for protein–ligand binding with the data set for the 40 NNRTIs.

It should be emphasized that the ELR method relies 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 is needed in order to make activity predictions for novel compounds. Simulations for the bound and unbound states are all that is needed to make activity predictions for any new compound. Ideally, a universal regression equation (scoring function) may emerge through additional studies.

### Computational Details

**System Setup.** Given the large size of HIVRT, simulations of the entire protein–ligand complex are currently impractical. Therefore, a model of the NNRTI binding site was constructed which incorporated only nearby residues (Figure 2). Using the initial crystal structure coordinates for MKC-442 bound to HIVRT (PDB entry 1rt1),<sup>35</sup> a representative model was constructed by including only those residues within ca. 15 Å of atom C6 of the HEPT uracil core. To avoid excessive fragmentation of the protein backbone, a few additional amino acids were included. Hydrogen atoms were added, and clipped residues were then capped with acetyl and methylamine groups. Residues with all atoms outside a 10 Å sphere from C6 were kept rigid during the MC simulations. The final



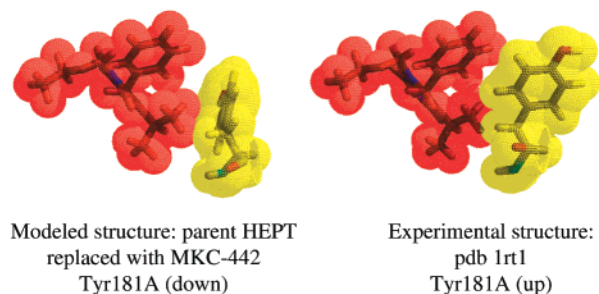
**Figure 2.** HIVRT binding site model surrounded by a 22 Å cap of water. Blue residues sampled in the MC simulations, red residues rigid, green residues not used. Crystal structure coordinates, PDB entry 1rt1, from ref 35.

system size was 123 protein residues plus the inhibitor. Specifically, the rigid residues are 91–94A, 109–110A, 116–178A, 84–185A, 192–197A, 199–205A, 222–224A, 230–232A, 240–242A, 316–317A, 320–321A, 343–349A, 381–383A, 134–135B, 137B, and 140B. The flexible residues are 95–108A, 179–183A, 186–191A, 198A, 225–229A, 233–239A, 318–319A, 136B, and 138B. To impose overall charge neutrality for the system,<sup>26</sup> all but one of the rigid Asp, Lys, Glu, and Arg residues were made neutral. The tautomeric states of His residues in the binding site were assigned by visual inspection. A residue-based cutoff at 9 Å was used for the solute–solvent and intrasolute nonbonded interactions. The water–water cutoff was also at 9 Å, based on the O–O separation. The nevirapine analogues were treated similarly starting from the coordinates of the X-ray structure of nevirapine bound to HIVRT (PDB entry 1vrt).<sup>36</sup>

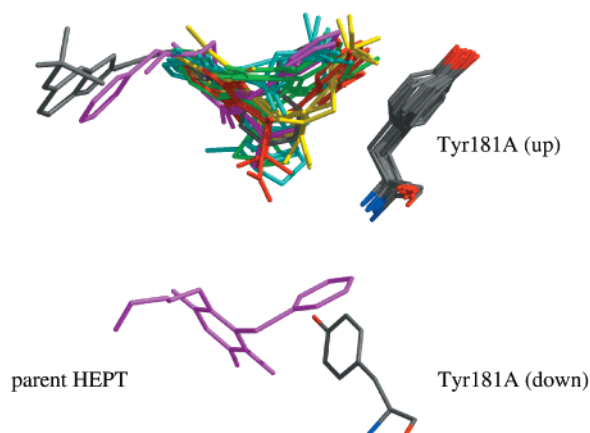
The initial Cartesian coordinates for each HEPT or nevirapine analogue were generated by analogy to the conformations in the crystal structures of HIVRT with MKC-442<sup>35</sup> and nevirapine<sup>36</sup> using the XChemEdit program.<sup>37,38</sup> The Z-matrix connectivity was then graphically assigned, and the results saved both as a PEPZ database<sup>39</sup> and as a Gaussian95 input file.<sup>40</sup> The OPLS-AA force field<sup>41,42</sup> was used for the systems except the partial charges for the inhibitors were determined using the ChelpG procedure at the HF/6-31G\* level.<sup>40</sup> Any missing OPLS-AA torsional parameters were assigned by analogy to existing ones with the exception of two new torsions, which were fit to results of dihedral angle energy scans at the HF/6-31G\* level for the model compounds methyl benzyl ether and thioanisole, as previously described.<sup>41</sup> The OPLS-AA parameters have been developed to reproduce accurately molecular geometries, torsional energetics, free energies of hydration, enthalpies of vaporization, and liquid densities for a wide range of model compounds.<sup>41</sup>

**Crystal Structure Choice.** **H01** (the parent HEPT) and analogue **H11** (MKC-442) differ in potency by about 4.6 kcal/mol (Table 1); a possible explanation has been suggested by Hopkins et al. based on an interpretation of crystallographic evidence.<sup>35</sup> A difference of ca. 100° in the  $\chi_1$  dihedral angle for Tyr181A was found between the structures for **H01** (1rt1) and **H11** (1rt1).<sup>35</sup> It was suggested that **H11** ( $R^1 = i\text{-Pr}$ ) is a more potent compound than **H01** ( $R^1 = \text{Me}$ ) because the larger group at  $R^1$  sterically forces Tyr181A “up”.<sup>35</sup> A favorable aromatic  $\pi$ -stacking interaction can then occur between Tyr181A and the phenyl ring in the  $R^3$  substituent (Table 1).<sup>35</sup> We believe that this interpretation is flawed for the following two reasons: (1) No steric clashes are visually apparent when **H11** is docked into the **H01** crystal structure (Figure 3), and conjugate





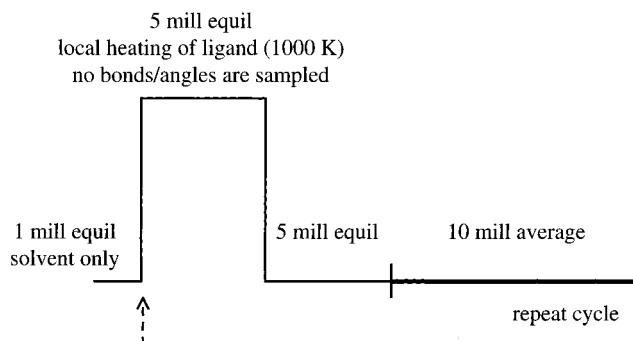
**Figure 3.** No steric clash is observed between HIVRT side chain Tyr181A and the *i*-Pr group of MKC-442 in the modeled structure using the “down” conformation, which is only reported for the parent HEPT.



**Figure 4.** Experimental conformation of Tyr181A for 16 HIVRT nonnucleoside inhibitor complexes: nevirapine (green), HEPT (magenta), BHAP (gray),  $\alpha$ -APA (red), TIBO (yellow), and carboxanilide (cyan) analogues. The complexes were aligned by minimizing the rmsd between C $\alpha$  carbons at residues Leu100A, Lys103A, Tyr181A, and Val106A. See text for PDB references.

gradient energy minimizations reveal no energetically unfavorable steric interactions between the *i*-Pr group of **H11** when Tyr181A is “down” as in the parent HEPT (**H01**) structure. This suggests that unfavorable steric interaction are not responsible for the “up” conformational preference observed in the MKC-442 crystal structure. (2) More importantly, an overlay of 16 experimental HIVRT/NNRTI crystal structures all show Tyr181A to be in the same “up” conformation with the lone exception of the structure for the parent HEPT (Figure 4). The 16 experimental structures include six different inhibitor cores. In fact, the NNRTIs based on nevirapine (green; 3hvt, 1rth, 1vrt),<sup>36,43</sup> HEPT (magenta; 1rt1, 1rt2, 1rti),<sup>35</sup>  $\alpha$ -APA (red; 1hni, 1vru),<sup>36,44</sup> TIBO (yellow; 1hmv, 1tvr, 1rev),<sup>45,46</sup> BHAP (gray; 1klm),<sup>47</sup> and carboxanilide (cyan; 1rt4, 1rt5, 1rt6, 1rt7)<sup>48</sup> could potentially allow Tyr181A to adopt the “down” conformation given that no steric clashes would result, yet this is not reported. A change in  $\chi_1$  for Tyr181A of ca. 100° would be expected to be a rare event in computer simulations of the present lengths and was not observed. Therefore, given the consistency in which Tyr181A is observed to be in the “up” conformation (Figure 4), PDB entry 1rt1<sup>35</sup> was chosen as the starting point for all simulations of HEPT analogues.

**MC Simulations.** Each protein–inhibitor complex was subjected to 50 steps of conjugate gradient energy minimization, using a distant-dependent dielectric constant of 4 ( $\epsilon = 4r$ ) to relax the crystal structure with the force field prior to the MC simulations. For the MC simulations, a 22 Å water cap was used containing 851 (bound) and 1485 (unbound) TIP4P water molecules.<sup>49</sup> All HIVRT side chains with an atom within ca. 10 Å from the defined center of the water cap were sampled, the protein backbone was fixed, and each inhibitor

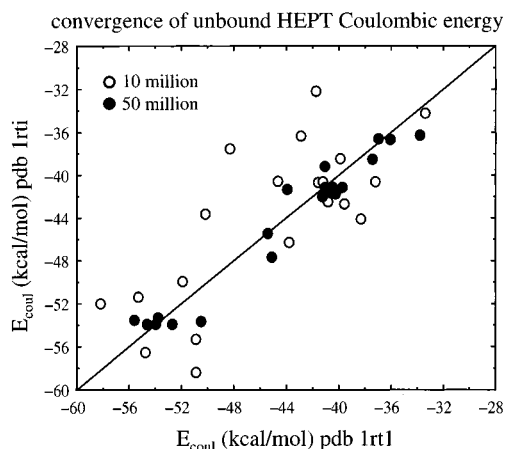


**Figure 5.** Annealing protocol showing heating, equilibration, and averaging portions used in the MC simulations for the unbound inhibitors.

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 attempted 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 minimizations were performed with the MCPRO program.<sup>50</sup> The computations were executed on a PC cluster with ca. 70 processors running Linux. The complete processing of one inhibitor (bound and unbound) requires 2.5 days using one 800 MHz PentiumIII processor. Thus, ca. 300 inhibitors could be processed in 1 week on a PC cluster with 100 top-end processors.

**Bound Simulations.** Each MC simulation for a protein–inhibitor complex consisted of 1 million configurations of solvent-only equilibration, 10 million configurations of full equilibration, and 10 million configurations of averaging. In general, convergence of the results for complexes is less problematic than for the simulations of the inhibitors alone in water. This probably results from the facts that in the simulations of the complexes the ligands are more conformationally restricted than in pure water and about one-half as many water molecules are used for the complexes than for the unbound inhibitors.

**Unbound Simulations Using an Annealing Protocol.** Initial results for the inhibitors alone in water revealed that the solute–water Coulombic interaction energy showed the slowest convergence among the descriptors and that it was not well-converged with MC simulations of the same length as for the complexes. Surprisingly, the same average energies were not obtained when the simulations were initiated from two similar yet distinct geometries. After additional testing, an annealing protocol (Figure 5) was developed to enhance the convergence. Each unbound MC simulation consisted of 1 million configurations of solvent-only equilibration at the experimental temperature of 37 or 25 °C. Then, 5 million configurations of equilibration ensued in which only the water and the dihedral angles of the inhibitor were sampled. The MC acceptance rate for the inhibitor was also increased through a local heating option in MCPRO with the temperature specified to be 727 °C (1000 K) for the attempted moves of the inhibitor. This was followed by an additional 5 million configurations of full equilibration at the normal temperature, followed by 10 million configurations of averaging. The latter three processes were then repeated for a total of five cycles (Figure 5). The local heating is applied only to the inhibitor and no bonds or angles are sampled during this stage. The focus is on increased conformational sampling for the inhibitor. Since local heating is only specified for the inhibitor, the bulk water structure is largely unaffected during the heating phase, and the dihedral-only sampling ensures that bond lengths and angles do not have to be cooled upon reequilibration.



**Figure 6.** Convergence of the inhibitor–water Coulombic energy for the HEPT data set after 10 million (1 cycle) and 50 million (5 cycles) configurations of averaging using the annealing protocol. Each inhibitor was simulated twice starting from two different conformations. Initial geometries of the inhibitors were derived from the 1rt1 and 1rti crystal structures.

Convergence of the solute–solvent  $E_{\text{Coul}}$  was greatly improved, as illustrated in Figure 6, using the new protocol. For Figure 6, simulations were initiated from two alternative geometries of all 20 HEPT analogues; one was based on the 1rt1 structure and the other on the 1rti structure.<sup>35</sup> The annealing results for the HEPT compounds clearly show that in five cycles acceptable convergence is obtained independent of small differences in the starting geometry of the unbound inhibitors. It may be noted that the annealing protocol formally corresponds to averaging the MC results from five independent simulations of 10 million configurations each. The importance of well-converged results cannot be overestimated for the LR or ELR equations to have good predictive value. For example, 5 kcal/mol of noise in the unbound  $E_{\text{Coul}}$  value can easily translate to 1–3 kcal/mol of noise in the predicted  $\Delta G_b$  with usual values for  $\beta$  in eqs 1 or 2.

**FEP.** To help in interpreting the results for nevirapine analogues, a FEP calculation was performed to determine the difference in free energy of hydration ( $\Delta\Delta G_{\text{hyd}}$ ) between model tertiary and secondary amides. Specifically, *N,N*-dimethylacetamide (DMA) was converted to *N*-methylacetamide (NMA) using well-established methods.<sup>19</sup> No internal degrees of freedom were sampled, so  $\Delta\Delta G_{\text{hyd}}$  could be computed simply by performing one mutation in water.<sup>19,51</sup> The FEP calculations were performed for the solute in a periodic cube containing 500 TIP4P water molecules. Both solute–solvent and solvent–solvent cutoffs were at 10 Å based on the separations of amide nitrogens and water oxygens. Each of the 10 windows consisted of 6 million configurations of equilibration, followed by an additional 5 million configurations of averaging. The potential functions for the amides were the same as for the HEPT and nevirapine inhibitors, OPLS-AA with HF/6-31G\* ChelpG atomic charges.

**Experimental Activities.** The experimental  $\text{EC}_{50}$  activities at 37 °C reported for the HEPT series<sup>14–17</sup> and the  $\text{IC}_{50}$  values at 25 °C for nevirapine analogues<sup>18</sup> were converted into approximate free energies of binding ( $\Delta G_{\text{exptl}}$ ) by eq 4 as listed in Tables 1 and 2. Although not formally equivalent, relative activities should correspond to relative free energies of binding for closely related series of inhibitors.<sup>52</sup>

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

To correlate both data sets simultaneously, an offset might be necessary, though it turned out not to be needed. Measured activities from the same laboratory<sup>53</sup> indicate that nevirapine (**N10**) is more potent than the parent HEPT (**H01**) by ca. 2.8 kcal/mol in general agreement with the difference of 2.3 kcal/

mol from the data in Tables 1 and 2. In another study,<sup>54</sup> MKC-442 (**H11**) was reported to be more potent than nevirapine (**N10**) by about 1.0 kcal/mol, while the data in Tables 1 and 2 imply 2.2 kcal/mol. The experimental HEPT activities span a range of 7.1 kcal/mol, which is twice as large as the range for the nevirapine analogues (Tables 1 and 2). Uncertainties were not reported for the experimental data, but they are typically at least 0.5 kcal/mol.

## Results and Discussion

**Regression Equations.** Correlations were derived using the statistical software package JMP.<sup>55</sup> Equation 5 shows the best four-descriptor equation obtained by fitting the experimental activities of the 40 compounds using the generic regression, eq 3.

$$\Delta G_{\text{calcd}} = -0.94\langle\Delta HB_{\text{total}}\rangle + 0.30\langle EXX-LJ \rangle + 0.0085\langle\Delta PHOB_{\text{area}}\rangle - 2.8(\text{secondary amide}) + 4.6 \quad (5)$$

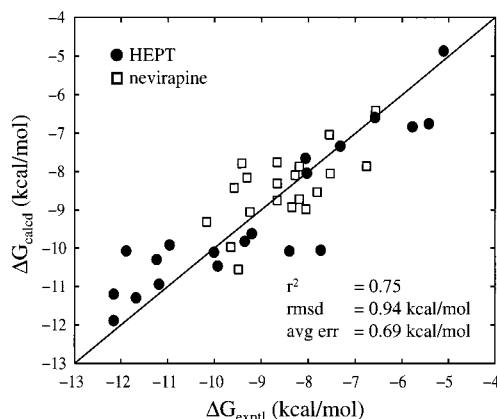
$\Delta HB_{\text{total}}$  is the change in the total number of hydrogen bonds for the inhibitor; a hydrogen bond is defined here by a distance of less than 2.5 Å between an N, O, or S atom and a hydrogen attached to a heteroatom.<sup>34</sup>  $EXX-LJ$  is the ligand–protein Lennard-Jones interaction energy,  $\Delta PHOB_{\text{area}}$  is the change in hydrophobic SASA upon binding, and secondary amide is an indicator variable (1 or 0) for compounds with or without a secondary-amide functional group. The contributions for each term are recorded in Table 3. For the 40 compounds, the correlation coefficient  $r^2$  of 0.75 reflects good accord between theory and experiment (Figure 7). Cross-validation by the leave-one-out procedure yields a  $q^2$  of 0.69 and implies reasonable predictive power for compounds not included in the original data set. The computed activities show a rmsd of 0.94 kcal/mol in comparison with experiment and an average unsigned error of only 0.69 kcal/mol. The uncertainties in the experimental data and in the convergence of the MC results are estimated to be at this level. All of the descriptors in eq 5 are significant. Probability >  $F$  ratios (regression model mean square/error mean square) are small:  $\Delta HB_{\text{total}}$  (0.0005),  $EXX-LJ$  (<0.0001),  $\Delta PHOB_{\text{area}}$  (0.0037), secondary amide (<0.0001). This implies that the probability of a greater  $F$  value occurring by chance is low. No systematic deviation in the predicted  $\Delta G_{\text{calcd}}$  values was found; the computed residuals show random scatter.

The four descriptors in eq 5 make physical sense: (1)  $\Delta HB_{\text{total}}$  is always negative; water is the best hydrogen-bonding medium, so there is an inevitable loss in number of hydrogen bonds for an inhibitor upon binding. The coefficient implies that the loss of each hydrogen bond costs 0.94 kcal/mol in free energy of binding. (2) The  $EXX-LJ$  term implies that a good geometrical fit between the ligand and the protein is also important. Favorable packing contributions to binding are contained in this term as well as any unfavorable steric interactions. The change in ligand–water Lennard-Jones energy ( $\Delta ESX-LJ$ ) is highly correlated with  $EXX-LJ$  (greater loss in  $\Delta ESX-LJ$  corresponds with greater gain in  $EXX-LJ$ ), so its inclusion does not improve the regression. (3) The  $\Delta PHOB_{\text{area}}$  term is also negative; SASA for a ligand is always lost upon binding. The associated coefficient is positive so that the removal of hydrophobic surface area upon binding is favorable for

**Table 3.** Individual Contributions to the Total Computed Free Energies of Binding for HEPT and Nevirapine Analogues with HIV-1 RT

no.	$\Delta G_{\text{exptl}}$ total	total	$\Delta HB_{\text{total}}$	$\Delta G_{\text{calcd}}$ EXX-LJ	$\Delta PHOB_{\text{area}}$	secondary amide
H01	-7.32 <sup>a</sup>	-7.33	3.24	-13.70	-1.47	0.00
H02	-7.73 <sup>a</sup>	-10.05	2.24	-14.93	-1.96	0.00
H03	-9.20 <sup>a</sup>	-9.62	2.05	-14.39	-1.87	0.00
H04	-8.06 <sup>a</sup>	-7.65	2.29	-13.04	-1.48	0.00
H05	-10.01 <sup>a</sup>	-10.11	1.67	-15.41	-0.96	0.00
H06	-12.16 <sup>a</sup>	-11.89	1.88	-16.79	-1.56	0.00
H07	-8.03 <sup>a</sup>	-8.04	1.46	-12.69	-1.39	0.00
H08	-5.43 <sup>a</sup>	-6.75	1.64	-11.89	-1.09	0.00
H09	-10.96 <sup>a</sup>	-9.92	2.23	-14.59	-2.15	0.00
H10	-11.24 <sup>a</sup>	-10.30	2.18	-14.68	-2.39	0.00
H11	-11.89 <sup>b</sup>	-10.07	2.34	-14.61	-2.39	0.00
H12	-9.93 <sup>a</sup>	-10.47	1.99	-14.87	-2.18	0.00
H13	-6.58 <sup>c</sup>	-6.59	3.17	-12.77	-1.58	0.00
H14	-5.78 <sup>c</sup>	-6.83	3.48	-13.32	-1.58	0.00
H15	-9.35 <sup>d</sup>	-9.82	3.24	-14.79	-2.86	0.00
H16	-11.19 <sup>d</sup>	-10.94	2.96	-15.29	-3.19	0.00
H17	-12.16 <sup>d</sup>	-11.20	3.10	-15.60	-3.28	0.00
H18	-11.68 <sup>a</sup>	-11.30	1.46	-16.05	-1.29	0.00
H19	-5.11 <sup>a</sup>	-4.86	1.68	-10.56	-0.58	0.00
H20	-8.40 <sup>a</sup>	-10.07	1.62	-14.35	-1.93	0.00
N01	-9.42 <sup>e</sup>	-7.78	2.08	-12.85	-1.60	0.00
N02	-9.24 <sup>e</sup>	-9.05	2.15	-13.47	-2.32	0.00
N03	-9.31 <sup>e</sup>	-8.16	2.44	-13.62	-1.57	0.00
N04	-8.35 <sup>e</sup>	-8.93	2.08	-13.32	-2.28	0.00
N05	-8.19 <sup>e</sup>	-7.87	2.54	-13.46	-1.54	0.00
N06	-7.81 <sup>e</sup>	-8.53	2.23	-13.28	-2.07	0.00
N07	-8.67 <sup>e</sup>	-7.76	3.58	-12.19	-0.91	-2.82
N08	-10.17 <sup>e</sup>	-9.32	3.71	-13.26	-1.54	-2.82
N09	-9.58 <sup>e</sup>	-8.42	3.92	-13.15	-0.97	-2.82
N10	-9.65 <sup>e</sup>	-9.98	3.60	-13.62	-1.73	-2.82
N11	-8.19 <sup>e</sup>	-8.72	2.67	-13.87	-2.10	0.00
N12	-8.66 <sup>e</sup>	-8.30	2.24	-13.37	-1.77	0.00
N13	-6.77 <sup>e</sup>	-7.86	2.77	-13.38	-1.84	0.00
N14	-6.57 <sup>e</sup>	-6.41	3.36	-13.21	-1.15	0.00
N15	-9.49 <sup>e</sup>	-10.56	3.22	-13.67	-1.88	-2.82
N16	-8.28 <sup>e</sup>	-8.09	2.57	-13.74	-1.52	0.00
N17	-7.54 <sup>e</sup>	-8.05	5.53	-13.85	-1.50	-2.82
N18	-8.05 <sup>e</sup>	-8.99	4.32	-13.81	-1.27	-2.82
N19	-7.55 <sup>e</sup>	-7.04	2.85	-13.17	-1.31	0.00
N20	-8.66 <sup>e</sup>	-8.75	3.42	-12.95	-0.99	-2.82

<sup>a</sup> Ref 14. <sup>b</sup> Ref 15. <sup>c</sup> Ref 16. <sup>d</sup> Ref 17. <sup>e</sup> Ref 18.  $\Delta G_{\text{exptl}} \approx RT \ln(\text{activity})$ .  $\Delta G_{\text{calcd}}$  obtained from eq 5. Energies in kcal/mol.

**Figure 7.** Predicted binding affinities ( $\Delta G_{\text{calcd}}$ ) using eq 5 vs experimental activities ( $\Delta G_{\text{exptl}}$ ) for 20 HEPT (●) and 20 nevirapine (□) analogues with HIVRT.

the free energy, which simply reflects the hydrophobic effect. (4) Finally, as described in the next section, a secondary-amide indicator is needed to account for deficiencies in the partial charges.

The separate data sets yield modified optimal fits. For the HEPT analogues alone, an  $r^2$  of 0.83 is obtained with eq 6:

$$\Delta G_{\text{calcd}} = -1.00\langle\Delta HB_{\text{total}}\rangle + 0.31\langle EXX-LJ\rangle + 0.0112\langle\Delta PHOB_{\text{area}}\rangle + 5.6 \quad (6)$$

All descriptors in eq 5 are still significant except no secondary amides are present for HEPT analogues, so this descriptor is eliminated. For the nevirapine data set alone, however, only the  $\Delta HB_{\text{total}}$  and secondary-amide descriptors are significant. A fit with these two descriptors plus a constant yields an  $r^2$  of 0.58 (eq 7). In this case, the lower  $r^2$  may reflect challenges associated with the compressed range of the experimental activities in comparison with the data for the HEPT series.

$$\Delta G_{\text{calcd}} = -1.10\langle\Delta HB_{\text{total}}\rangle - 2.44(\text{secondary amide}) - 11.2 \quad (7)$$

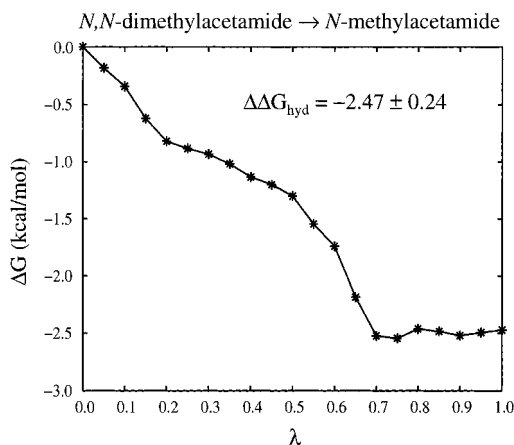
Binding affinity fits with the traditional ELR approach (eq 2) were also made for comparison. A reasonable  $r^2$  of 0.56 and rmsd of 1.24 kcal/mol are obtained with eq 2 augmented by the secondary-amide indicator and a constant. Nevertheless, eq 5 is clearly superior with the same number of descriptors. It may be noted that eq 5 does not include a term that obviously reflects differences in flexibility for the inhibitors. A rotatable-bond count was considered but was not found to be statistically significant. For more diverse sets of ligands, it is likely that such a term may be needed to reflect the entropic penalty for loss of conformational freedom upon binding.<sup>56</sup>

**Secondary-Amide Indicator.** During the fitting, it was discovered that acceptable correlations could not be obtained for the nevirapine analogues unless an indicator variable was included for secondary amides. Suspecting that the use of the 6-31G\* ChelpG charges was overestimating hydration differences in the unbound state, the FEP calculation was performed for the model tertiary-amide  $\rightarrow$  secondary-amide conversion of DMA  $\rightarrow$  NMA in water (Figure 8). The computed  $\Delta\Delta G_{\text{hyd}}$  of  $-2.47 \pm 0.24$  kcal/mol is too negative by 1 kcal/mol in comparison with the experimental value of  $-1.53$  kcal/mol.<sup>57</sup> By analogy, nevirapine analogues with secondary amides would be expected to be too well-hydrated in the unbound state and thus pay an artificially high desolvation penalty for binding. Thus the indicator coefficient of  $-2.8$  in eq 5 has the correct sign, though the magnitude is larger than from the simple model, as clarified below.

It should be noted that obtaining correct relative free energies of hydration for amines and amides has been a long-standing problem in the computational community.<sup>58–62</sup> Successful parameters for primary, secondary, and tertiary aliphatic, cyclic, and aromatic amines have now been reported,<sup>62</sup> and parallel improvements for amides have recently been achieved.<sup>63</sup>

**Analysis of Binding Trends – HEPT Series.** Equation 5 presents a straightforward framework for understanding the trends in the observed activities. For the HEPT analogues in Table 3, the ranges for the contributions from the hydrogen-bond loss, protein-inhibitor Lennard-Jones energy, and burial of hydrophobic surface area are 1.9, 6.2, and 2.7 kcal/mol, respectively. There is variation in the R<sup>2</sup> side chain (Table 1), and the side chains with no oxygen atoms

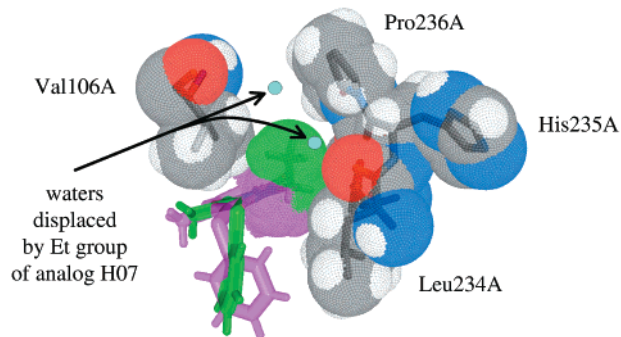




**Figure 8.** Plot of  $\Delta G$  (kcal/mol) vs  $\lambda$  for the perturbation of *N,N*-dimethylacetamide to *N*-methylacetamide. The nonbonded parameters and geometries were scaled using the coupling coordinate  $\lambda$ .

(**H07**, **H08**, **H19**, **H20**) show smaller desolvation penalties (Table 3). In the simulations of the complexes,  $R^2$  is in a channel that contains some water and the terminal hydroxyl in, for example, **H01** is involved in hydrogen bonds with water or the backbone carbonyl of L234A. So, the range of desolvation penalties is not as great as might have been expected. The larger analogues then benefit from more favorable Lennard-Jones interactions (**H05**, **H06**, **H16**, **H17**, **H18**), which is the dominant discriminator. The HEPT derivatives with  $R^3$  as 3,5-dimethylthiophenyl (**H15**, **H16**, **H17**) or with isopropyl groups at  $R^1$  (**H06**, **H10**, **H11**, **H17**) get an additional boost for burial of more hydrophobic surface area than their less substituted analogues. The factors combine such that **H06**, **H17**, and **H18** are observed and predicted to be in the most active group. Some comments can also be made on specific pairs of inhibitors with small structural but large activity differences.

**H08 vs H07.** The HEPT analogues **H08** ( $R^2$  = Me) and **H07** ( $R^2$  = Et) differ only by a Me group yet have an experimental activity difference ( $\Delta\Delta G_{\text{exptl}}$ ) of more than 2.6 kcal/mol. The computed relative free energy of binding ( $\Delta\Delta G_{\text{calcd}}$ ) is 1.3 kcal/mol, in qualitative agreement with experiment. In Table 3, the computed free energy penalties for lost hydrogen bonds ( $\Delta HB_{\text{total}}$ ) are similar: 1.46 kcal/mol for **H07** and 1.64 kcal/mol for **H08**. However, the larger Et group of **H07** improves the hydrocarbon packing in the binding pocket (Figure 9) and yields a more favorable *EXX-LJ* contribution by about 0.8 kcal/mol over **H08**. Additional benefit for **H07** comes from the burial of more hydrophobic surface area (−1.4 kcal/mol) than for **H08** (−1.1 kcal/mol). Given the structure in Figure 9, these results for the descriptors are reasonable. In addition, two water molecules are displaced from the binding pocket upon expansion of the methyl group in **H08** to ethyl in **H07** (Figure 9). In general, this should be an entropically favorable process since the bound water molecules likely gain translational and rotational freedom upon transfer into the bulk solvent<sup>56,64</sup> In contrast, homologation of **H03** to **H02**, though reported to diminish activity (Table 1), is also predicted to enhance activity (Table 3). A steric problem is not found for **H02** here, which is consistent



**Figure 9.** Two water molecules (cyan) are displaced by compound **H07** (green; Et analogue) that are observed in simulations of compound **H08** (magenta; Me analogue) with HIVRT.

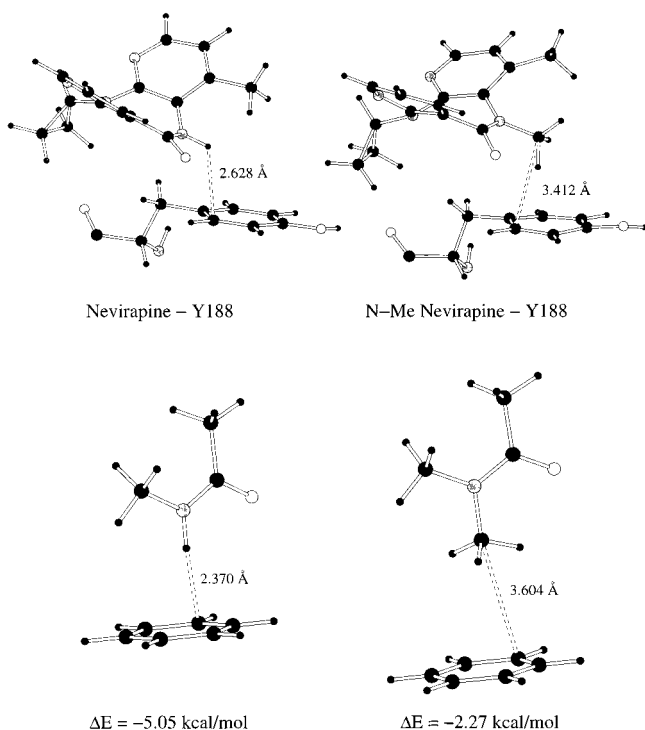
with the observed accommodation of the even larger benzyloxy group for **H05**.

**H14 vs H01.** **H14** ( $R^3$  = OPh) and the parent HEPT **H01** ( $R^3$  = SPh) only differ in the atom linking the phenyl ring to the uracil core, yet **H01** is more potent than **H14** by 1.5 kcal/mol (Table 1). The computed results are again qualitative accord with the difference diminished to 0.5 kcal/mol. Examination of the components in Table 3 shows that **H14** has a computed  $\Delta G(\Delta HB_{\text{total}})$  of 3.48 kcal/mol compared to 3.24 for **H01**. Though an ether oxygen is expected to be better hydrated in the unbound state than a thioether sulfur, these atoms are hindered in the bisaryl analogues **H01** and **H14**, so there is only a small differential. However, the more favorable  $\Delta G(\text{EXX-LJ})$  contribution for **H01** (−13.70) compared to **H14** (−13.32) makes sense given that sulfur is more polarizable than oxygen and has a larger Lennard-Jones  $\epsilon$  (0.25 vs 0.14 kcal/mol).<sup>41</sup> Finally, the  $\Delta G(\Delta PHOB_{\text{area}})$  values are essentially the same for **H01** (−1.47 kcal/mol) and **H14** (−1.58 kcal/mol) reflecting similar burial of hydrophobic surface area. Thus, the greater activity of the sulfur analogue **H01** is predicted to come primarily from better van der Waals interactions with some help from a smaller desolvation penalty.

**Analysis of Binding Trends – Nevirapine Series.** For the nevirapine series, the ranges are 3.5, 1.7, and 1.4 kcal/mol for the desolvation penalty, protein–inhibitor Lennard-Jones interactions, and burial of hydrophobic surface area (Table 3). The compressed ranges are consistent with the smaller variation in activities (Figure 7). The small ranges for the latter two effects are also consistent with the diminished differences in total size and hydrophobic surface area; i.e., the ranges of SASA and FOSA values are 468–530 and 115–275 Å<sup>2</sup> for the nevirapines and 448–648 and 66–409 Å<sup>2</sup> for the HEPT analogues. The dominant term then becomes desolvation and the secondary-amide indication. Thus, the inhibitors with more polar side chains are less active, i.e., **N14**, **N17**, and **N18** (Table 2).

However, the experimental results for the secondary-amide vs tertiary-amide analogues are interestingly mixed: **N08** and **N10** are experimentally more active than their tertiary homologues; **N06** and **N11**, by more than a factor of 10 (Table 2), while the secondary **N07** is reported to be less active than its tertiary derivative **N01** by a factor of 3.5, and another pair with  $R^2$  = Et

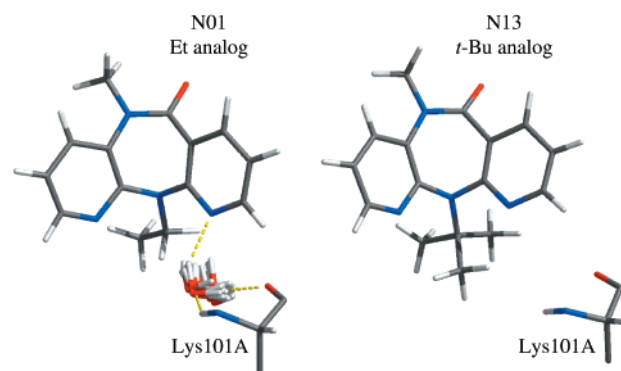




**Figure 10.** (Top) Computed snapshots of nevirapine (**N10**) and *N*-methylnevirapine (**N11**) with Y188 from the MC simulations. (Bottom) Optimized structures of model secondary and tertiary amides, *N*-methylacetamide and *N,N*-dimethylacetamide, with benzene. The net interaction energy is shown along with the shortest distances to aromatic carbons.

and  $R^3 = 2,3$ -dimethyl is reported to have the tertiary compound more active than the secondary by a factor of 2.<sup>18</sup> In the crystal structure for nevirapine (**N10**) with HIVRT,<sup>36</sup> there are water molecules hydrogen-bonded to both pyridine nitrogens and the amide carbonyl, though there is no hydrogen bond for the amide NH. The simulation results typically have one water molecule hydrogen-bonded to a pyridine nitrogen, but there is no water molecule within hydrogen-bonding range of the amide carbonyl. Thus, the secondary-amide fragment is not well-accommodated in any event, and in the absence of another factor, the secondary amides should not be so competitive with the tertiary analogues.

The missing factor appears to be a favorable NH-aryl  $\pi$ -type hydrogen bond for the secondary amides with the phenyl ring of Y188A. Though this has not been specifically noted in the crystallographic studies, it is illustrated in Figure 10.<sup>36,43</sup> The structures are shown for **N10** and **N11** with Y188 from the last configuration of both MC runs, which is representative. For comparison, the optimal structures and interaction energies using the OPLS-AA force field for the complexes of the model amides, *cis*-NMA and DMA, with benzene are also shown. The shortest distance between the amide NH and an aromatic carbon is only 0.26 Å longer for **N10** than *cis*-NMA and a comparably attractive interaction is indicated. The longer distance is reasonable since the optimized NMA-benzene structure is effectively at a temperature of 0 K; it may also be noted that the interaction energy for *trans*-NMA with benzene is somewhat more attractive, -5.55 kcal/mol. For DMA and benzene, the  $\pi$ -type hydrogen bond is lost and the attraction drops nearly 3 kcal/mol. The shortest distance between the *N*-methyl carbon and a ring carbon of Y188



**Figure 11.** A water-mediated hydrogen bond is consistently observed between **N01** (Et analogue) and Lys101A that is not observed in the MC simulations of **N13** (*t*-Bu analogue) with HIVRT.

is now 3.4 Å, which is 0.2 Å shorter than in the optimal DMA-benzene complex. In the crystal structure for **N10** with HIVRT,<sup>36</sup> the shortest distance between the amide N (coordinates are not given for the H) and the Y188 ring carbons is 3.54 Å for CD2, while the corresponding distance for the computed structure in Figure 10 is 3.24 Å. Thus, we propose that the binding of the secondary amides in the nevirapine series benefits significantly from a  $\pi$ -type hydrogen bond with Y188. This factor coupled with the overestimate of the desolvation energy of secondary amides is responsible for the magnitude and significance of the secondary-amide indicator in eq 5. Such strong  $\pi$ -type hydrogen bonds could be included in the hydrogen-bond counts in the future. In support of this analysis, it is known that the Y188C mutant of HIVRT is 100–1000-fold less sensitive to nevirapine (**N10**) than the wild-type protein.<sup>65</sup> The decrease in activity for tertiary analogues such as **N11** should not be as severe, but this has not been studied to our knowledge.

One final point for the nevirapines is that **N13** ( $R^2 = t$ -Bu) is observed to have low activity (Table 2). This is the only compound in this series with a tertiary substituent at  $R^2$ , and not surprisingly, the hydration of the proximal pyridine nitrogens is affected. The effect is actually not great for **N13** unbound in water; it is computed to accept an average of 3.0 hydrogen bonds from water, which is just a little less than the 3.2–3.6 for tertiary amides **N01**–**N06**. However, in the complex with HIVRT, the bulkier *tert*-butyl group displaces the water molecule from the pocket near the pyridine nitrogens. Both the hydration of a pyridine nitrogen and the backbone of K101A are adversely affected. This is illustrated in Figure 11 by contrasting representative configurations from the MC simulations of the complexes for **N13** and **N01**. The energetic penalty for the net loss of the hydrogen bonding with the pyridine nitrogen is about 0.7 kcal/mol in comparing **N13** with **N01** in Table 3. Equation 5 does not obviously reflect the penalty for the poorer solvation of K101A, which may account for **N13** being predicted to be too active by 1 kcal/mol.

## Conclusion

The results of the MC simulations presented here revealed three physically reasonable parameters that control binding for two series of inhibitors with

HIVRT: loss of hydrogen bonds with the inhibitor upon binding is unfavorable, burying hydrophobic surface area of the inhibitor is favorable, and a good geometrical match between the inhibitor and the protein is important. The best regression equation that was generated (eq 5) reveals a strong correlation with experimental activities (Figure 7,  $r^2 = 0.75$ ), and the cross-validated  $q^2$  of 0.69 implies reasonable predictive power for compounds not included in the original data set. Given the comparatively large size of the data set (40 compounds), the results provide strong support for the utility of the ELR method. On the technical side, convergence of the results for the unbound inhibitors in water was carefully investigated and led to the development of an effective annealing method. Further efforts on improving the efficiency and convergence of both the unbound and bound simulations are ongoing.

The structural details from the MC simulations are also valuable in interpreting trends in the binding and activity data. In particular, a key  $\pi$ -type hydrogen bond between the secondary-amide fragment of nevirapine analogues and the aryl ring of Tyr188A of HIVRT was identified that explains the otherwise surprising activity of the secondary amides and the poor activity of nevirapine against the Y188C mutant. Detailed knowledge of the hydration of the inhibitor and the protein by specific water molecules is also repeatedly found to be relevant in interpreting binding/activity data.

Finally, the development of improved, low-cost anti-HIV drugs is critical.<sup>2</sup> The present study has been successful in advancing the potential for computational methods to participate in achieving this goal. It has been demonstrated that computer simulations can be used to make predictions of binding affinities for sizable data sets in a reasonable time frame. And, the examination of the associated energetic and structural results can provide bases for understanding activity differences and for rational drug design.

**Acknowledgment.** Gratitude is expressed to Dr. Dongchul Lim for the XChemEdit program that facilitated inhibitor Z-matrix generation, to Dr. Marilyn B. Kroeger Smith, Prof. Richard H. Smith, and Mark A. Wilson for helpful discussions, and to the National Institute of Allergy and Infectious Diseases (Grant AI44616) for support.

## References

- Mitsuya, H.; Yarchoan, R.; Broder, S. Molecular Targets For Aids Therapy. *Science* **1990**, *249*, 1533–1544.
- AIDS epidemic update: December 1999, Joint United Nations Programme on HIV/AIDS (UNAIDS) and The World Health Organization (WHO). Available from <http://www.unaids.org>.
- De Clercq, E. HIV Resistance to Reverse Transcriptase Inhibitors. *Biochem. Pharmacol.* **1994**, *47*, 155–169.
- Katz, R. A.; Skalka, A. M. The Retroviral Enzymes. *Annu. Rev. Biochem.* **1994**, *63*, 133–173.
- Turner, B. G.; Summers, M. F. Structural Biology of HIV. *J. Mol. Biol.* **1999**, *285*, 1–32.
- 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.
- Goodenow, M.; Huet, T.; Saurin, W.; Kwok, S.; Sninsky, J.; Wainhobson, S. HIV-1 Isolates Are Rapidly Evolving Quasispecies: Evidence For Viral Mixtures and Preferred Nucleotide Substitutions. *J. Acquired Immune Defic. Syndr. Hum. Retrovirol.* **1989**, *2*, 344–352.
- Eigen, M. On the nature of virus quasispecies. *Trends Microbiol.* **1996**, *4*, 216–218.
- Preston, B. D.; Poiesz, B. J.; Loeb, L. A. Fidelity of HIV-1 Reverse Transcriptase. *Science* **1988**, *242*, 1168–1171.
- Roberts, J. D.; Bebenek, K.; Kunkel, T. A. The Accuracy of Reverse Transcriptase From HIV-1. *Science* **1988**, *242*, 1171–1173.
- Perelson, A. S.; Neumann, A. U.; Markowitz, M.; Leonard, J. M.; Ho, D. D. HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time. *Science* **1996**, *271*, 1582–1586.
- Wilson, E. K. AIDS Conference Highlights Hope of Drug Cocktails, Chemokine Research. *Chem. Eng. News* **1996**, *74*, 42–46.
- Cohen, J. AIDS Therapies: The Daunting Challenge of Keeping HIV Suppressed. *Science* **1997**, *277*, 32–33.
- 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.
- 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.
- 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.
- 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.
- 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 Dipyrroldiazepinones. *J. Med. Chem.* **1991**, *34*, 2231–2241.
- 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.
- Lamb, M. L.; Jorgensen, W. L. Computational approaches to molecular recognition. *Curr. Opin. Chem. Biol.* **1997**, *1*, 449–457.
- Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
- Jorgensen, W. L. Free Energy Calculations: A Breakthrough For Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, *22*, 184–189.
- 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.
- Carlson, H. A.; Jorgensen, W. L. An Extended Linear Response Method For Determining Free Energies of Hydration. *J. Phys. Chem.* **1995**, *99*, 10667–10673.
- 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.
- Hansson, T.; Åqvist, J. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng.* **1995**, *8*, 1137–1144.
- Paulsen, M. D.; Ornstein, R. L. Binding free energy calculations for P450cam-substrate complexes. *Protein Eng.* **1996**, *9*, 567–571.
- 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.
- Hansson, T.; Marelus, J.; Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27–35.

- (30) 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.
- (31) Jones-Hertzog, D. K.; Jorgensen, W. L. Binding affinities for Sulfonamide Inhibitors with human Thrombin Using Monte Carlo Simulations with a Linear Response Methodology. *J. Med. Chem.* **1997**, *40*, 1539–1549.
- (32) 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 Methodology. *J. Med. Chem.* **1998**, *41*, 5272–5286.
- (33) 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.
- (34) 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.
- (35) 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.
- (36) 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.
- (37) Lim, D. *XChemEdit Version 0.82*; Yale University: New Haven, CT, 1999.
- (38) Lim, D.; Jorgensen, W. L. ChemEdit. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; Wiley: New York, 1998; Vol. 5, pp 3295–3302.
- (39) Tirado-Rives, J. *PEPZ Version 1.0*; Yale University: New Haven, CT, 1997.
- (40) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Strain, M. C.; Burant, J. C.; Stratman, R. E.; Petersson, G. A.; Montgomery, J. A.; Zakrzewski, V. G.; Raghavachari, K.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Stewart, J. J. P.; Gonzalez, C.; Head-Gordon, M.; Gill, P. M. W.; Johnson, B. G.; Pople, J. A. *Gaussian 95, Development Version (Revision E.1)*; Gaussian Inc.: Pittsburgh, PA, 1996.
- (41) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, 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.
- (42) Jorgensen, W. L. *BOSS Version 4.1*; Yale University: New Haven, CT, 2000.
- (43) Smerdon, S. J.; Jager, J.; Wang, J.; Kohlstaedt, L. A.; Chirino, A. J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. Structure of the Binding Site for Nonnucleoside Inhibitors of the Reverse Transcriptase of Human Immunodeficiency Virus Type 1. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3911–3915.
- (44) Ding, J.; Das, K.; Tantillo, C.; Zhang, W.; Clark, A. D., Jr.; Jessen, S.; Lu, X.; Hsiou, Y.; Jacobo-Molina, A.; Andries, K.; et al. Structure of HIV-1 reverse transcriptase in a complex with the nonnucleoside inhibitor  $\alpha$ -APA R 95845 at 2.8 Å resolution. *Structure* **1995**, *3*, 365–79.
- (45) Das, K.; Ding, J. P.; Hsiou, Y.; Clark, A. D.; Mooreels, H.; Koymans, L.; Andries, K.; Pauwels, R.; Janssen, P. A. J.; Boyer, P. L.; Clark, P.; Smith, R. H.; Smith, M. B. K.; Michejda, C. J.; Hughes, S. H.; Arnold, E. Crystal structures of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181Cys HIV-1 RT drug-resistant mutant. *J. Mol. Biol.* **1996**, *264*, 1085–1100.
- (46) Ren, J.; Esnouf, R.; Hopkins, A.; Ross, C.; Jones, Y.; Stammers, D.; Stuart, D. The structure of HIV-1 reverse transcriptase complexed with 9-chloro-TIBO: lessons for inhibitor design. *Structure* **1995**, *3*, 915–26.
- (47) Esnouf, R. M.; Ren, J. S.; Hopkins, A. L.; Ross, C. K.; Jones, E. Y.; Stammers, D. K.; Stuart, D. I. Unique features in the structure of the complex between HIV-1 reverse transcriptase and the bis(heteroaryl)piperazine (BHAP) U-90152 explain resistance mutations for this nonnucleoside inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3984–3989.
- (48) 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.
- (49) 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.
- (50) Jorgensen, W. L. *MCPRO Version 1.65*; Yale University: New Haven, CT, 2000.
- (51) Jorgensen, W. L.; Ravimohan, C. Monte Carlo Simulation of Differences in Free Energies of Hydration. *J. Chem. Phys.* **1985**, *83*, 3050–3054.
- (52) Cheng, Y.; Prusoff, W. H. Relationship Between Inhibition Constant (K<sub>i</sub>) and Concentration of Inhibitor Which Causes 50 Per Cent Inhibition (I<sub>50</sub>) of an Enzymatic Reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099–3108.
- (53) Balzarini, J.; Karlsson, A.; Sardana, V. V.; Emini, E. A.; Camarasa, M. J.; Declercq, E. Human Immunodeficiency Virus 1 (HIV-1)-Specific Reverse- Transcriptase (RT) Inhibitors May Suppress the Replication of Specific Drug-Resistant (E138K)-RT HIV-1 Mutants or Select For Highly Resistant (Y181 → C181I) RT HIV-1 Mutants. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 6599–6603.
- (54) Baba, M.; Shigeta, S.; Yuasa, S.; Takashima, H.; Sekiya, K.; Ubasawa, M.; Tanaka, H.; Miyasaka, T.; Walker, R. T.; Declercq, E. Preclinical Evaluation of MKC-442, a Highly Potent and Specific Inhibitor of Human-Immunodeficiency Virus Type 1 In Vitro. *Antimicrob. Agents Chemother.* **1994**, *38*, 688–692.
- (55) *JMP Version 3*; SAS Institute Inc.: Cary, NC, 1995.
- (56) 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.
- (57) Wolfenden, R. Interaction of the Peptide Bond With Solvent Water: A Vapor Phase Analysis. *Biochemistry* **1978**, *17*, 201–204.
- (58) Morgantini, P. Y.; Kollman, P. A. Solvation Free Energies of Amides and Amines: Disagreement Between Free Energy Calculations and Experiment. *J. Am. Chem. Soc.* **1995**, *117*, 6057–6063.
- (59) Ding, Y. B.; Bernardo, D. N.; Kroghjerspersen, K.; Levy, R. M. Solvation Free Energies of Small Amides and Amines From Molecular-Dynamics Free Energy Perturbation Simulations Using Pairwise Additive and Many-Body Polarizable Potentials. *J. Phys. Chem.* **1995**, *99*, 11575–11583.
- (60) Meng, E. C.; Caldwell, J. W.; Kollman, P. A. Investigating the anomalous solvation free energies of amines with a polarizable potential. *J. Phys. Chem.* **1996**, *100*, 2367–2371.
- (61) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. New model for calculation of solvation free energies: Correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bonding effects. *J. Phys. Chem.* **1996**, *100*, 11775–11788.
- (62) 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.
- (63) Pearlman, S.; Jorgensen, W. L. Manuscript submitted for publication.
- (64) Dunitz, J. D. The Entropic Cost of Bound Water in Crystals and Biomolecules. *Science* **1994**, *264*, 670.
- (65) Buckheit, R. W.; Fliakaboltz, V.; Yeagybargo, S.; Weislow, O.; Mayers, D. L.; Boyer, P. L.; Hughes, S. H.; Pan, B. C.; Chu, S. H.; Bader, J. P. Resistance to 1-[(2-Hydroxyethoxy)methyl]-6-(Phenylthio)Thymine Derivatives Is Generated By Mutations At Multiple Sites in the HIV-1 Reverse-Transcriptase. *Virology* **1995**, *210*, 186–193.

JM000255N