Prediction of the Binding Free Energies of New TIBO-like HIV-1 Reverse Transcriptase Inhibitors Using a Combination of PROFEC, PB/SA, CMC/MD, and Free Energy Calculations

Mats A. L. Eriksson,[†,§] Jed Pitera,[‡] and Peter A. Kollman*,[†]

Department of Pharmaceutical Chemistry and Graduate Group in Biophysics, University of California at San Francisco, San Francisco, California 94143-0446

We have ranked 13 different TIBO derivatives with respect to their relative free energies of binding using two approximate computational methods: adaptive chemical Monte Carlo/molecular dynamics (CMC/MD) and Poisson–Boltzmann/solvent accessibility (PB/SA) calculations. Eight of these derivatives have experimentally determined binding affinities. The remaining new derivatives were constructed based on contour maps around R86183 (8Cl-TIBO), generated with the program PROFEC (pictorial representation of free energy changes). The rank order among the derivatives with known binding affinity was in good agreement with experimental results for both methods, with average errors in the binding free energies of 1.0 kcal/mol for CMC/MD and 1.3 kcal/mol for the PB/SA method. With both methods, we found that one of the new derivatives was predicted to bind 1–2 kcal/mol better than R86183, which is the hitherto most tightly binding derivative. This result was subsequently supported by the most rigorous free energy computational methods: free energy perturbation (FEP) and thermodynamic integration (TI). The strategy we have used here should be generally useful in structure-based drug optimization. An initial ligand is derivatized based on PROFEC suggestions, and the derivatives are ranked with CMC/MD and PB/SA to identify promising compounds. Since these two methods rely on different sets of approximations, they serve as a good complement to each other. Predictions of the improved affinity can be reinforced with FEP or TI and the best compounds synthesized and tested. Such a computational strategy would allow many different derivatives to be tested in a reasonable time, focusing synthetic efforts on the most promising modifications.

Introduction

Inhibitors to HIV-1 RT are one of the cornerstones in the treatment of AIDS patients, preventing the progression of HIV infection. The enzyme is an attractive target for drug therapy not only because it is essential for HIV replication but also since it is not required for normal host cell replication. HIV-1 RT is a multifunctional enzyme that copies the RNA genome of HIV-1 into DNA which is subsequently integrated into the host cell's genome. The enzyme is a heterodimer composed of the two subunits p66 and p51, and its unliganded structure has been determined at 2.35-Å resolution.[1] The active site (or the dNTP site), which contains the catalytically essential amino acids (primarily a triad of aspartic acids), is located in the p66 palm subdomain with the 3′-OH of the primer terminus near the active site. The types of inhibitors currently discovered can be divided into two classes: nucleoside inhibitors [NIs, for example, AZT, ddI, and ddC (for reviews, see refs 2–5)] and nonnucleoside inhibitors [NNIs, for example, TIBO, HEPT, α-APA, and nevirapine (reviewed in refs 4, 6–8)]. The NIs cause termination of the growing DNA chain because elongation is blocked due to the lacking 3′-OH

functional group, which is essential for incorporation of additional nucleosides (reviewed in ref 3). However, the NIs can also be incorporated into cellular DNA by the host DNA polymerases and therefore cause serious side effects. Unlike the NIs, NNIs are HIV-1 RT-specific and do not inhibit host cell polymerases. The binding site of the NNIs is located in the p66 palm subdomain near, but distinct from the dNTP-binding site. Recently, the two similar crystal structures of HIV-1 RT in complex with the TIBO NNIs R86183 (8-Cl TIBO)[9] and R82913 (9-Cl TIBO)[10] have been solved at 3.0-Å resolution. The NNI binding pocket constitutes mainly hydrophobic and aromatic residues (green residues in Figure 1). Comparisons of structures of HIV-1 RT complexed with different NNIs[1] reveal that there is a significant rearrangement of a three-stranded β-sheet in the p66 subunit (containing the catalytic triad of aspartic acids), with respect to the rest of the polymerase site. This suggests that NNIs inhibit HIV-1 RT by locking the polymerase active site in an inactive conformation, similar to the conformation observed in the inactive p51 subunit.[1] In addition, the NNIs have low cytotoxicity and produce few side effects.[11] A serious problem with the NNI HIV-1 RT inhibitors is the emergence of viral strains that have point mutations in the region encoding HIV-1 RT which prevent these drugs from inhibiting RT.

There is a considerable interest in developing computational methods that are sufficiently efficient to

---

* To whom correspondence should be addressed: tel, (415) 476-4637; fax, (415) 476-0688; e-mail: pak@cgl.ucsf.edu.
  † Department of Pharmaceutical Chemistry.
  ‡ Graduate Group in Biophysics.
  § Present address: Department of Biochemistry, University of Stockholm, SE-10691 Stockholm, Sweden.
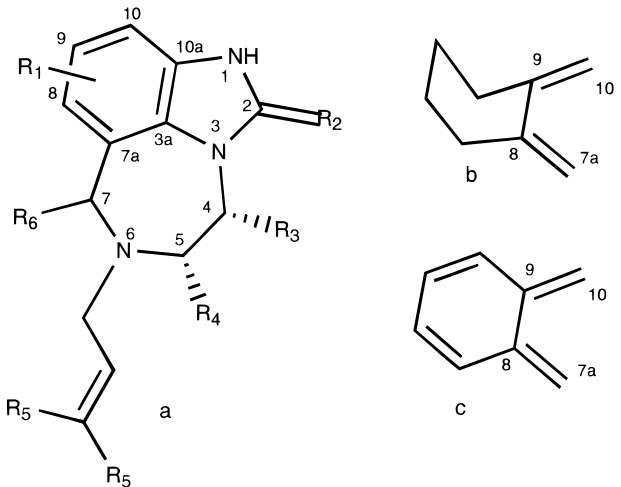
Figure 1. 8Cl-TIBO (R86183, red) in HIV-1 RT. The nonpolar (green) and polar (white) residues shown have any atom closer than 3.5 Å from any TIBO atom. This is a snapshot from the 500-ps MD simulation, and some close water molecules are also shown as blue spheres. This picture and the other molecular graphics images in this paper have been created with the program MidasPlus.[43]

allow ranking of several (10−100) inhibitors with respect to their binding free energy to a common receptor. This stems from the fact that the most rigorous computational methods—free energy perturbation (FEP) and thermodynamic integration (TI) calculations (for reviews, see, for example, refs 12, 13)—are both too slow for practical use in drug optimization. These two methods typically give good (<1 kcal/mol[14]) estimates of the relative binding energies. However, only one pair of inhibitors can be compared in a single FEP/TI run, which may require from days to weeks to complete due to their computationally intensive nature. Since lead optimization requires the comparison of *many* possible modifications to the lead compound, there is a need for more rapid methods. One such method, denoted chemical Monte Carlo/molecular dynamics (CMC/MD), has recently been developed by Pitera and Kollman.[15] CMC/MD combines the Monte Carlo method for sampling the chemical space of a system and the MD method for generating a set of coordinates for a distinct chemical system. The chemical space can be typically 5−10 different derivatives of an inhibitor. During the course of a CMC/MD run, the probabilities of each derivative are generated, which can be related to their relative binding free energies. The CMC/MD method has successfully been applied to estimate relative solvation free energies for small organic molecules and to study the strength of small ligands binding to an organic host.[15] In CMC/MD the solvent is typically described with explicit water molecules, which is a computational bottleneck since a great portion of computer time is spent calculating forces on the solvent molecules. The problem is circumvented in the Poisson−Boltzmann/

solvent accessibility (PB/SA) method,[16,17] where the solvent is treated implicitly as a dielectric continuum. The protein and inhibitor are modeled as low-dielectric cavities containing fixed partial charges. By solving the Poisson−Boltzmann equation the electrostatic contribution to the binding free energy is calculated. The nonpolar contribution to the binding free energy is estimated, assuming an empirical linear dependence on the solvent accessibility areas.[18,19] This method has been applied to a number of protein−ligand complexes,[20−25] for estimation of absolute and relative ligand binding free energies. A third method is the linear interaction energy (LIE) method, developed by Åqvist.[26] In this method, the binding free energy is approximated to be linearly dependent on the ligands' interaction energies in the protein or in solution. The LIE method (sometimes with slight modifications) has also been applied to a variety of ligand−protein complexes,[26−30] where absolute as well as relative binding free energies have been estimated.

To generate suggestions for modifications on a lead inhibitor that would improve its binding, Radmer and Kollman have developed PROFEC[31] (pictorial representation of free energy changes). This approach uses MD trajectories to estimate the average cost of adding a particle around the inhibitor in the protein and in solution, respectively. The difference cost can then be visualized as contour maps around the inhibitor. Favorable (negative) regions of the contour maps indicate positions where modifications to the inhibitor should improve its binding free energy.

In this study, we have used the PROFEC method to suggest modifications on R86183, which is so far the

Table 1. Selected Set of TIBO Derivatives

compd	R_1	R_2	R_3	R_4	R_5	R_6	EC_{50} (nM)[a]
R86183	8-Cl	S	H	CH_3	H	H	4.6
R82913	9-Cl	S	H	CH_3	H	H	33
R82150	H	S	H	CH_3	H	H	44
R80902	H	O	H	CH_3	H	H	4200
R84674	8-CH_3	S	H	CH_3	H	H	14
R84963	H	S	H	CH_3	H	-CH_3 (*trans*)[b]	39
R84914	H	S	H	CH_3	H	-CH_3 (*cis*)[b]	790
R87027	8-Cl	S	H	CH_3	CH_3	H	5.1
45MeT	8-Cl	S	CH_3	CH_3	H	H	
4MeT	8-Cl	S	CH_3	H	H	H	
4ClT	8-Cl	S	Cl	H	H	H	
HET	b	S	H	CH_3	H	H	
BET	c	S	H	CH_3	H	H	

[a] Reference 32. [b] Relative stereochemistry of the methyl groups at positions 5 and 7.

tightest binding inhibitor (see Table 1). Five new derivatives was constructed with the help of the contour maps from PROFEC. These new inhibitors were ranked together with eight derivatives with experimentally[32] known binding affinities, using both the CMC/MD and the PB/SA methods. The purposes of the study are (1) to test these relatively new methods (especially CMC/MD) against experimental results and against each other; (2) to develop a feasible computational strategy for structure-based lead optimization; and (3) to generate a better binding TIBO derivative than that previously known. We show that both methods work surprisingly well, given the approximations involved, and rank the inhibitors in good agreement with the experiment. Since the two methods are based on different sets of approximations, they serve as good complements to each other and a consistent result between them increases its validity. Both methods predict that one of the new PROFEC derivatives should bind HIV-1 RT about 1−2 kcal/mol stronger than R86183. Subsequently, we performed "full" free energy calculations (FEP and TI) on this best binding inhibitor. The full free energy calculations also suggest that this new inhibitor binds better than R86183. The strategy we have employed herein (outlined in Figure 2) could generally be used as one of the final stages in a structure-based lead optimization using computational methods.

Computational Methods

Force-Field Parameters for the TIBO Derivatives. The van der Waals (vdW), bond, angle, dihedral, and improper
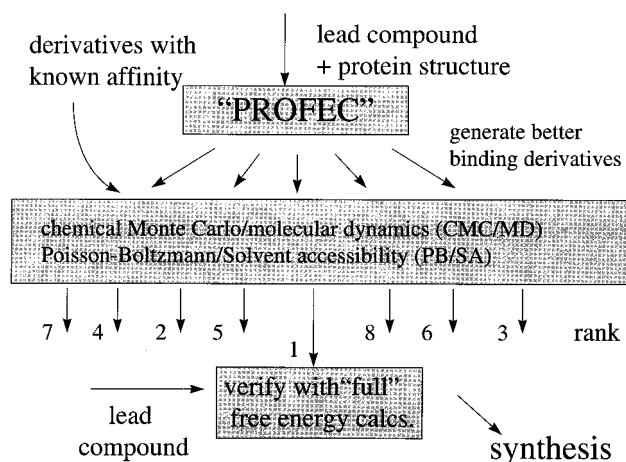

Figure 2. General outline of a procedure that can be applied as one of the final steps in computational structure-based lead optimization. We have used this strategy for the TIBO derivatives in this work.

dihedral parameters for the sulfur atom were adopted from a parametrization of thiobiotin,[33] and vdW parameters of the chlorine atom were taken from parameters used for chloroform.[34] To estimate the partial atomic charges of 8Cl-TIBO (R86183; see Table 1), 9Cl-TIBO (R82913), and unchlorinated TIBO (R82150), we used both the conformation of 8Cl-TIBO in complex with HIV-1 RT[9] as well as the A-form of the crystal structure of 9Cl-TIBO (R82913).[35] The two respective conformers were geometry-optimized using Gaussian94[36] at the STO-3G level. The electrostatic potential around the TIBO derivatives was then calculated with the 6-31G* basis set, and atomic partial charges were fitted to the electrostatic potentials around the two structures using the RESP method.[37] Since the partial charges evaluated from the two conformers individually were very similar, we evaluated the partial charges of the remaining TIBO derivatives (Table 1) using only the conformation of 8Cl-TIBO in HIV-1 RT.

Setup and Equilibration of HIV-1 RT in Complex with 8Cl-TIBO. The simulations were carried out with the AMBER 4.1[38] program SANDER using the Cornell et al. force field.[39] Starting with the 3.0-Å resolution crystal structure of 8Cl-TIBO in HIV-1 RT,[9] we added unresolved residues, modeled as alanines in the crystal structure, as well as hydrogen atoms. The hydrogen atoms were then minimized for 200 steps (steepest descent) in vacuo. To let the protein relax in an aqueous environment, the complex was immersed in a 55-Å radius sphere of TIP3P−water.[40] The solvent sphere together with the protein−inhibitor complex were minimized with a gradual decrease in the position restraints of the protein atoms. Thereafter, all water molecules beyond the first hydration shell (i.e., at a distance > 3.5 Å from any protein atom) were removed, and to achieve electroneutrality 11 chloride ions were added, using the program module "CION" within Amber 4.1. Protein residues with any atom closer than 12 Å from 8Cl-TIBO were chosen to be flexible in the simulations. All protein residues, water molecules, and counterions further away than 15 Å from any flexible residue were then deleted, due to the size of HIV-1 RT. We then centered a 20-Å radius spherical cap of TIP3P−water around TIBO, including the hydrating water molecules within the sphere from the previous step. The water cap was equilibrated for 50 ps at 300 K, keeping the protein, 8Cl-TIBO, and hydrating water molecules outside the water cap rigid. Thereafter, the flexible residues (as defined above) and 8Cl-TIBO together with the cap of water molecules were heated (50 ps) and equilibrated for 300 ps at 300 K. A time step of 2 fs was used, with the nonbonded list updated every 20 time steps, and all bonds were constrained with the SHAKE algorithm.[41] We applied a dual cutoff of 9 and 13 Å, respectively, where energies and forces due to interactions between 9 and 13 Å were updated every 20 time steps. The temperature was maintained using the Berendsen method,[42]

with separate couplings of the solute and solvent to the heat. This system was then run for 500 ps for a subsequent analysis with the PROFEC program (see below).

Setup and Equilibration of 8Cl-TIBO in Solution. For 8Cl-TIBO in solution we started with the A-form of the crystal structure of 9Cl-TIBO,[35] with a substitution of the atoms at positions 8 and 9. 8Cl-TIBO was then immersed in a box of TIP3P–water with dimensions $34 \times 33 \times 29$ Å^3. The water molecules were equilibrated at constant pressure for 100 ps, keeping the inhibitor rigid. We then released the TIBO atoms, and the system was equilibrated for 200 ps, using the same dual cutoff and time step as for 8Cl-TIBO in HIV-1 RT. Also here, we performed an additional 500-ps MD simulation for the PROFEC analysis.

Pictorial Representation of Free Energy Changes (PROFEC). The contour maps that are generated from the program PROFEC[31] can be used as guides to where atoms/groups should be added or deleted on the inhibitor in order to improve its binding free energy to a protein. The maps are generated from trajectories of two MD simulations: one of the protein–inhibitor complex and the other of the inhibitor in solution. The insertion free energy of a test particle (ΔG_{ins}) at various grid points close to the inhibitor is calculated according to:

$$\Delta G_{ins}(i,j,k) = -RT \ln \langle \exp(-\Delta V(i,j,k)/RT \rangle_0 \qquad (1)$$

where i, j, and k are the coordinates of a grid point, $\Delta V(i,j,k)$ is the interaction energy between the test particle and the surrounding atoms, and $\langle ... \rangle_0$ is an average over the trajectories. To generate the coordinate system of the grid points, three atoms in the inhibitor determine a coordinate plane and the third axis is formed as a vector product of two axes in that plane. Since the coordinate system is molecule-fixed, corresponding grid points are comparable for the inhibitor in solution and in the protein, respectively. The difference, $\Delta\Delta G_{ins}$, of particles in the inhibitor–protein complex and in the inhibitor in solution, respectively, is formed for each grid point, and contour maps of $\Delta\Delta G_{ins}$ can be constructed and displayed. The electrostatic properties of the added test particles can also be estimated by calculating the derivative of $\Delta\Delta G_{ins}(i,j,k)$ with respect to charge at each grid point. This derivative is then displayed by coloring the contour map (at, for example, $\Delta\Delta G_{ins} = 0$) and might thus suggest how the charge distribution should be changed for an improved binding.

The PROFEC contour maps were calculated from the two 500-ps MD trajectories of 8Cl-TIBO in HIV-1 RT and in solution, respectively. In each PROFEC calculation we chose the grid size to be 7.5 Å with a grid spacing of 0.5 Å. We selected different atoms of 8Cl-TIBO in each calculation to obtain detailed contour maps centered on various regions of the ligand. Through a special delegate program written by R. J. Radmer (UCSF), the contour maps were visualized with UCSF MidasPlus.[43]

Chemical Monte Carlo (CMC)/Molecular Dynamics (MD). The CMC/MD method (described in detail in ref 15) has recently been developed for determination of relative free energies of a series of ligands binding to a common receptor. The method employs MD to generate a set of coordinates for one distinct chemical system and MC to sample the *chemical* space of the system, which can be 5–10 different derivatives of an inhibitor in a protein–inhibitor system. The derivatives are all present in the protein binding pocket during the simulation, but they do not exert forces on each other. In addition, the protein only feels the presence of one ("real") inhibitor at a given time. An MC step consists of choosing an inhibitor 'i' at random, and this inhibitor will be accepted as the new "real" ligand, according to the Metropolis[44] criteria:

$$\text{if } \Delta E_i \leq 0 \Rightarrow P_i = 1$$

$$\text{if } \Delta E_i > 0 \Rightarrow P_i = \exp(-\Delta E_i/RT) \qquad (2)$$

where ΔE_i is the difference in protein–inhibitor interaction

energy between a derivative 'i' and the old derivative and P_i is the acceptance probability. We use the protein–ligand interaction energy instead of the total system energy in our Monte Carlo step since the only thing that changes in the MC move is which ligand is "real", i.e., interacting with the protein. During the course of the MC/MD run, the probability of each derivative being the "real" ligand P_i is accumulated, resulting in a probability distribution that mirrors the relative free energies of the bound state of the derivatives. To better determine the relative free energies of unfavorable states, the "Boltzmann" probabilities of each inhibitor 'i' can also be calculated prior to each MC step according to:

$$P_i = \exp(-\Delta E_i/RT)/\sum_i \exp(-\Delta E_i/RT) \qquad (3)$$

If an infinite number of MC steps were performed on a *single* Cartesian conformation, the resulting probability distribution $\{P_i\}$ would be exactly the same as that calculated with eq 3. We used the averaged P_i's from eq 3 herein, since they also allow for estimations of the relative free energies of poorly sampled derivatives. The resulting probability distribution is then related to the relative free energy of the bound state (ΔG_{bound}) for derivatives 'j' and 'i' according to:

$$\Delta G_{bound,j} - \Delta G_{bound,i} = -RT \ln(P_j/P_i) \qquad (4)$$

We found that CMC/MD, as outlined above, converged very slowly when applied to the TIBO derivatives in HIV-1 RT. In order to increase the convergence rate, a variant of this method—herein called the "adaptive CMC/MD" method (J. Pitera, unpublished)—was developed. The goal with adaptive CMC/MD is to sample the chemical space *evenly* instead of sampling this space according to the relative free energies of the derivatives. This can be achieved by introducing biasing offsets, $\Delta G_{offs,i}$, that for each inhibitor 'i' reflect its relative free energy in the bound state. These biasing offsets are introduced by umbrella sampling, as previously described.[15,45] MC sampling by testing the acceptance against $\Delta E_i - \Delta G_{offs,i}$, rather than ΔE_i as in eq 2, would then result in an even sampling of all inhibitors, since all $\Delta E_i - \Delta G_{offs,i}$ would equal zero, on average. Starting with all $\Delta G_{offs,i} = 0$, the offsets are solved for iteratively and the probabilities of each inhibitor are calculated according to eq 3, averaged over a certain number of MC/MD cycles (a CMC/MD run). A first set of $\Delta G_{offs,i}$'s, relative to some arbitrarily chosen ligand, is estimated from eq 4, and these offsets are used in the next CMC/MD run. The offsets are then adjusted iteratively after each CMC/MD run by averaging the P_i's from eq 3 and using the $\Delta G_{bound,i}$ obtained from eq 4 to adjust $\Delta G_{offs,i}$. Upon convergence, all P_i's are roughly equal and the relative free energies of the bound state ($\Delta G_{bound,i}$) are equal to $-\Delta G_{offs,i}$. Finally, the relative free energies of binding ($\Delta\Delta G_{bind}$) are calculated by subtracting the solvation free energies (ΔG_{solv}) from ΔG_{bound}. This adaptive procedure is effectively the same as the WHAM procedure[46] for calculating conformational free energy differences. However, chemical MC/MD allows us to use it for the calculation of chemical free energy differences.

The adaptive CMC/MD method was applied to eight different TIBO derivatives with experimentally known[32] relative binding affinities (Table 1) together with five new derivatives that were suggested by visualization of the PROFEC contour maps. The derivatives were created by substituting and/or deleting atoms of the HIV-1 RT conformer of 8Cl-TIBO and positioned in the equilibrated HIV-1 RT–8Cl-TIBO complex (see above). The inhibitors were then allowed to relax in the binding pocket by individually minimizing them, keeping everything but the inhibitors rigid. In the subsequent adaptive CMC/MD calculations, the MD time step was reduced to 1.5 fs due to problems with the SHAKE algorithm and 1 MC step was performed every 20 MD time steps. We applied the adaptive CMC/MD method for two sets of inhibitors, each containing 10 derivatives. The $\Delta G_{offs,i}$'s were iteratively adjusted every 500 MC steps for set 1. For set 2 we shortened that interval to every 125 MC steps. The values of $\Delta G_{offs,i}$ were

graphed and monitored until they appeared converged by visual inspection. This required 450 ps (30 iterations) for set 1 and 560 ps (150 iterations) for set 2.

The solvation free energies of the TIBO derivatives were estimated from generalized Born/solvent accessibility (GB/SA)[47] calculations, using the program MacroModel/BatchMin, version 4.5,[48] with our RESP-derived charges on the derivatives. The derivatives were substitutions from the A-form of the 9Cl-TIBO crystal structure[35] that were minimized in vacuo prior to the calculations. While this approach does not include the relative internal entropies of each compound in solution, we expect those contributions to be small among our family of highly similar and relatively rigid compounds.

Poisson–Boltzmann/Solvent Accessibility (PB/SA) Calculations. In the PB/SA calculations, which were carried out with the latest Delphi package,[49,50] the solvent is represented by a continuum with a dielectric constant $\epsilon = 80$, with or without implicit ions. In this work we added implicit ions to an ionic strength of 0.13 M. The protein and the TIBO derivatives are represented by a cavity with a dielectric constant $\epsilon = 2$, containing fixed partial charges on their atoms. The relative free energy of solvation ($\Delta\Delta G_{solv}$) for two TIBO derivatives (L_1 and L_2) was estimated according to:

$$\Delta\Delta G = \Delta G_{react}^{g\to aq}(L_2) - \Delta G_{react}^{g\to aq}(L_1) + \Delta\Delta G_{nonpol} \quad (5)$$

where $\Delta G_{react}^{g\to aq}$ is the reaction field energy when transferring the derivative from vacuum ($\epsilon = 1$) to aqueous solution ($\epsilon = 80$). The nonpolar contribution ($\Delta\Delta G_{nonpol}$) can be estimated according to the following empirical relation, which correlates the solvation free energies of nonpolar solutes to their solvent-accessible surface area:[18,19]

$$\Delta\Delta G_{nonpol} = \sigma\Delta A \quad (6)$$

where ΔA is the difference in solvent-accessible area between L_1 and L_2. ΔA was calculated with Connolly's MS program[51] using vdW radii from the Cornell et al. force field.[39] σ is the empirical solvation parameter, and we used a value of 5 cal mol^{-1} A^{-2} in this work. The same structures as for the GB/SA calculations (above) were used for estimations of ΔG_{solv}.

A corresponding calculation of the relative free energies in the bound state ($\Delta\Delta G_{bound}$) involves an estimation of the difference in solvation free energies between L_1P and L_2P. This is very difficult in practice since these energies are large numbers and subtracting them might result in large errors. Therefore, as a first approximation, we estimated the polar contribution to $\Delta\Delta G_{bound}$ simply as the difference in reaction field energy on the two ligands (L_1 and L_2) in the protein [$\Delta G_{react,L_2}^{g\to aq}(L_2P) - \Delta G_{react,L_1}^{g\to aq}(L_1P)$] plus the difference in ligand–protein electrostatic energy ($\Delta\Delta G_{lig-prot,elec}$). A more rigorous PB calculation would also include the intramolecular energies of the protein–ligand complexes and of the free ligands. However, the intramolecular energies of the proteins in the protein–ligand complexes are large numbers and might add large errors when subtracting them. We assume therefore to a first approximation that these energies cancel then forming the difference between the complexes. Moreover, the intramolecular energies of the free ligands are sensitive to their conformation, and a correct inclusion of these energies would thus require multiple conformations of the ligands, generated, for example, by using MD simulations. This would make the PB method relatively slow and tedious, and since we are interested in rapid, approximate methods that are able to rank the derivatives, we only use one conformation in solution and assume to a first approximation that the relative intramolecular energies of the ligands cancel in the protein and in solution, respectively. Finally, since the TIBO derivatives are completely buried in the hydrophobic binding pocket (i.e., their accessible surface areas are zero), the nonpolar contribution can simply be estimated as the difference in ligand–protein vdW energy ($\Delta\Delta G_{lig-prot,vdW}$). The resulting simplified expression for $\Delta\Delta G_{bound}$ is thus:

$$\Delta\Delta G_{bound} = \Delta G_{react,L_2}^{g\to aq}(L_2P) - \Delta G_{react,L_1}^{g\to aq}(L_1P) + \Delta\Delta G_{lig-prot,elec} + \Delta\Delta G_{lig-prot,vdW} \quad (7)$$

The resulting relative free energies of binding ($\Delta\Delta G_{bind}$) are then estimated from the difference $\Delta\Delta G_{bound} - \Delta\Delta G_{solv}$. The 13 different TIBO–HIV-1 RT complexes were further minimized, now with flexible residues, water molecules, and counterions as in the MD simulations (see above). Prior to the PB/SA calculations, all water molecules and counterions were removed.

Free Energy Calculations (TI and FEP). To estimate the relative binding free energies ($\Delta\Delta G_{bind}$) of two TIBO derivatives (L_1 and L_2) to HIV-1 RT (P), we made use of the following thermodynamic cycle:

$$L_1 + P \xrightarrow{\Delta G_1} L_1P$$
$$\downarrow \Delta G_s \qquad\qquad \downarrow \Delta G_b$$
$$L_2 + P \xrightarrow{\Delta G_2} L_2P \qquad (8)$$
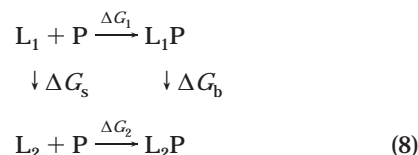
where ΔG_1 and ΔG_2 are the experimentally determined[32] binding affinities. Since G is a state function the following applies:

$$\Delta\Delta G_{bind} = \Delta G_2 - \Delta G_1 = \Delta G_b - \Delta G_s \quad (9)$$

where ΔG_b and ΔG_s are calculated with the TI or FEP methods (see below).

Except for one set of calculations (described below) we used the TI algorithm to estimate ΔG_b and ΔG_s (eq 8). In this method a coupling parameter λ is introduced, which varies from $\lambda = 0$ (for L_1P) to $\lambda = 1$ (for L_2P). The free energy change is then evaluated according to:

$$\Delta G = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (10)$$

where $H(\lambda)$ is the potential energy of the system as a function of the coupling parameter λ and $\langle\ \rangle_\lambda$ is an ensemble average at λ. The integral is evaluated numerically from a number of evenly spaced windows (spacing $= \Delta\lambda$) with λ values ranging from 0 to 1. $\langle\partial H(\lambda)/\partial\lambda\rangle_\lambda$ is calculated by averaging over molecular dynamics trajectories run at a certain number of steps in each window. The calculations were run with the AMBER 4.1 program GIBBS, and we applied the same parameters and protocol as for the MD simulations (above). Starting with the equilibrated systems of 8Cl-TIBO in HIV-1 RT and in solution, respectively, the 8-chloro atom was perturbed into a hydrogen (R82150). We continued with a perturbation where the position of the chlorine was changed from 8 to 9 (i.e., R86183 to R82913, see Table 1). A window size ($\Delta\lambda$) of 0.02 was used, i.e., 51 windows in the λ interval [0,1], and for TIBO in solution each window was equilibrated for 2 ps prior to a data collection time of 5 ps per window. The corresponding equilibration/data collection times for TIBO in HIV-1 RT were 3 and 8 ps, respectively. In the final, full free energy calculation of derivatives with experimentally known binding affinities, we perturbed the sulfur of R82150 into an oxygen (R80902, Table 1). In the first set (1b and 1s, respectively, see Table 2), we used the same parameters and protocol as for the perturbations described above. This yields a $\Delta\Delta G_{bind}$(R82150 − R80902) = +1.7 kcal/mol, which is far from the experimental value of −2.69 kcal/mol. This perturbation involves much larger changes in the electrostatics than the two previous, and the free energy change might thus be more sensitive to the treatment of long-ranged electrostatics as well as the local counterion configuration. We therefore changed the protocol as follows (sets 2b and 2s, respectively): for TIBO in HIV-1 RT we added counterions (Cl$^-$) at salt bridges that were truncated during the setup of the TIBO–HIV-1 RT system (see above) to obtain a net electroneutral system. One of the counterions was constrained to be 3.8 Å from the sulfur/
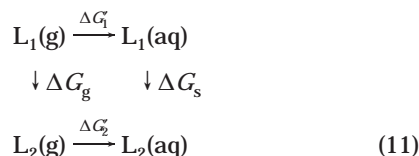
Table 2. Free Energy (kcal/mol) Perturbation (TI), R80902 → R82150, Using Various Protocols and Parameters

	in HIV-1 RT				in solution			
	set 1b[a]	set 2b	set 3b	set 4b	set 1s[a]	set 2s	set 3s	set 4s
ΔG_{tot}	27.92 ± 0.01	27.00	24.54	23.03	26.20 ± 0.05	26.29	24.91	24.03
$\Delta G_{el.stat}$	23.82 ± 0.03	22.61	23.07	22.92	23.14 ± 0.03	23.06	22.95	22.86
ΔG_{vdW}	3.03 ± 0.01	3.58	1.13	0.24	3.01 ± 0.02	3.08	1.89	1.08
ΔG_{pmf}	1.07 ± 0.04	0.81	0.35	−0.13	0.05 ± 0.04	0.015	0.076	0.089

[a] Set 1b: dual cutoff (9/13 Å), $\delta\lambda = 0.02$, $t_{eq} = 3$ ps, $t_{sampl} = 8$ ps. Set 2b: cutoff 10 Å, cutoff for TIBO 100 Å, electroneutral system, one Cl$^-$ constrained to be 3.8 Å from sulfur/oxygen atom in TIBO, $\delta\lambda = 0.02$, $t_{eq} = 2$ ps, $t_{sampl} = 5$ ps. Set 3b: as set 2, but with R^* changed from 2.0 to 1.9 and ϵ from 0.2 to 0.381. Set 4b: as set 2, but with $R^* = 1.844$ and $\epsilon = 0.55$. Set 1s: as set 1b, but in a water box of water and with $t_{eq} = 2$ ps, $t_{sampl} = 5$ ps. Set 2s: as set 1s, but in a water sphere with radius 20 Å and cutoff 10 Å, cutoff for TIBO 100 Å. Set 3s: as set 2s, but with R^* changed from 2.0 to 1.9 and ϵ from 0.2 to 0.381. Set 4s: as set 2s, but with $R^* = 1.844$ and $\epsilon = 0.55$.

oxygen, since three lysine residues are very close to this part of TIBO (two of them are shown in Figure 1). The nonbonded cutoff was increased to 10 Å, and all interactions with TIBO closer than 100 Å were included (i.e., all atoms in the system). Since the perturbation of set 1 was well converged we decreased the equilibration and data collection times to 2 and 5 ps per window, respectively, and we ran the remaining perturbations only in one direction. For TIBO in solution (set 2b), we eliminated possible discrepancies between free energy estimates from a periodic box of water and the "cap" protein simulation, by instead simulating TIBO in a 20-Å radius sphere of TIP3P−water. Also here, we applied the 10/100-Å cutoff as described above. Set 2 gives a slightly lower free energy difference (0.7 kcal/mol) than set 1 but is still relatively far from the experimental value. Next, we suspected that the nonbonded parameters of the sulfur atom, which have been developed for sp³-sulfur,[33] might contribute to the erroneous value that we obtained. Therefore, R^* was changed from 2.0 to 1.9 and ϵ from 0.2 to 0.381, which results in a slightly smaller sulfur atom but with an unchanged repulsive contribution to the vdW energy (sets 3b and 3s, respectively). With these parameters (and with the same protocol as in set 2) we obtained a relative binding free energy of −0.4 kcal/mol, which is considerably closer to, but still relatively far away from, the experimental value. The vdW contribution to the free energy (ΔG_{vdW}) is very sensitive to the choice of nonbonded parameters, as is seen from Table 2. A further reduction of R^* to 1.844, which with $\epsilon = 0.55$ gives an unchanged repulsive contribution to the vdW energy (set 4b and 4s, respectively), results in a relative binding free energy of −1.0 kcal/mol, which is closer to the experimental value. It does not make physical sense to reduce R^* further.

Similarly, differences in free energy of solvation ($\Delta\Delta G_{solv}$) for two TIBO derivatives can be estimated from the following thermodynamic cycle of the derivatives in gas phase (g) and in solution (aq), respectively:

$$L_1(g) \xrightarrow{\Delta G'_1} L_1(aq)$$
$$\downarrow \Delta G_g \qquad \downarrow \Delta G_s$$
$$L_2(g) \xrightarrow{\Delta G'_2} L_2(aq) \qquad (11)$$

and $\Delta\Delta G_{solv}$ is obtained from the relation

$$\Delta\Delta G_{solv} = \Delta G'_2 - \Delta G'_1 = \Delta G_s - \Delta G_g \qquad (12)$$

ΔG_g was calculated by perturbing the TIBO derivatives in vacuo with the same $\Delta\lambda$, equilibration, and data collection times as when calculating ΔG_s above.

Finally, we performed perturbations on the next best binding TIBO derivative—HET—according to both the adaptive CMC/MD and the PB calculations. Starting from a 500-ps MD equilibration of the HIV-1 RT−HET complex (same parameters and protocol as for 8Cl-TIBO in HIV-1 RT), HET was perturbed into 8Cl-TIBO in two steps (see Figure 3). In the first step the cyclohexyl ring was perturbed into dummy atoms, and to avoid the singularity when these atoms disappear at $\lambda = 1$, we used a soft-core nonbonded potential energy func-
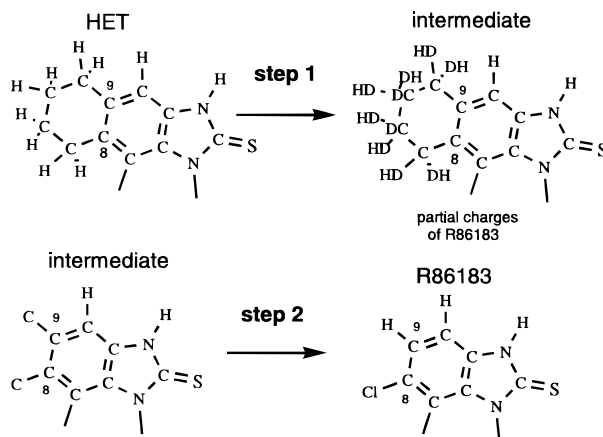

Figure 3. Two-step perturbation of HET → R86183 (only the aromatic part of TIBO is shown). DH and DC are dummy hydrogen and carbon atoms, respectively. These atoms have the same masses are the original atoms, but no interactions with the surrounding.

tion.[52,53] For atoms that disappear at $\lambda = 1$, this function has the form:

$$V_{nb} = (1 - \lambda)\sum_{i<j}\left\{\left[\frac{A_{ij}}{(\alpha_{LJ}\sigma_{ij}\lambda^2 + r_{ij}^6)^2} - \frac{B_{ij}}{(\alpha_{LJ}\sigma_{ij}\lambda^2 + r_{ij}^6)}\right] + \frac{q_i q_j}{4\pi\epsilon(\alpha_c\lambda^2 + r_{ij}^2)^{1/2}}\right\} \quad (13)$$

where α_{LJ} and α_c are the soft-core parameters for the Lennard−Jones and the electrostatic terms, respectively. We used values of $\alpha_{LJ} = 0.5$ Å² and $\alpha_{LJ} = 15.0$ Å², that previously[54] had been found suitable. This function is identical to the regular AMBER nonbonded potential function at the perturbation endpoints ($\lambda = 0$ and $\lambda = 1$), and they should thus give identical results. The function has the advantage of smoothing the interactions at short interatomic distances, which results in a well-behaved $\Delta G(\lambda)$ function. We used the slightly different free energy perturbation (FEP) scheme for this step, since the soft-core potential energy function currently is implemented only for that method. The FEP method relies on the following master equation instead of eq 10 for TI:

$$\Delta G = \sum_{\lambda=0}^{1} -RT \ln \langle e^{-[H(\lambda+\delta\lambda)-H(\lambda)]/RT}\rangle_\lambda \quad (14)$$

For this perturbation we used the same $\Delta\lambda$ and equilibration/data collection times in solution and protein, respectively, as above. The partial charges of HET were also perturbed into those of 8Cl-TIBO in this step. Bond potential of mean force (pmf) calculations cannot be performed when changing bond lengths of systems belonging to closed rings (as here), when using the FEP method. We therefore kept all bond lengths at their initial values by keeping the atoms bound to positions 8

Table 3. RESP-Derived[37] Partial Charges for Benzene and Chlorobenzene

atom	q (benzene)	q (chlorobenzene)
1C	−0.146	−0.033
1H/Cl	0.146	−0.115
2C	−0.146	−0.053
2H	0.146	0.133
3C	−0.146	−0.156
3H	0.146	0.147
4C	−0.146	−0.145
4H	0.146	0.150

and 9 in HET (see Figure 3) as carbons (atom type CT). In the second step, the two carbon atoms at positions 8 and 9 were perturbed into 8Cl and 9H, respectively, using the TI method. $\Delta\lambda$ in step 2 was increased to 0.05, and the equilibration/data collection times in solution were chosen as 3 and 8 ps, respectively. The corresponding times in the protein were prolonged to 4 and 10 ps, respectively. Instead of running the perturbations forward and reverse, as in the previous perturbations, we performed two forward (i.e., HET → R86183) perturbations (runs 1 and 2), which differ by an equilibration of 100 ps of HET in HIV-1 RT and in solution, respectively.

Test of Chlorine Parameters. To test whether the vdW parameters for chlorine, which were adopted from chloroform,[34] also could be used for the TIBO derivatives, we estimated the relative free energy of solvation ($\Delta\Delta G_{solv}$) between benzene and chlorobenzene. Partial charges of these two compounds (Table 3) were obtained from RESP fits[37] of the 6-31G* electrostatic potentials, calculated with the program Gaussian94.[36] Benzene was perturbed into chlorobenzene using the TI method. We used a window size ($\Delta\lambda$) of 0.01 with equilibration and data collection times of 1 ps in each window for both the perturbation in solution and in vacuo. Nonbonded interactions were cut off at 9 Å, and a time step of 2 fs was used. For the perturbation in solution benzene was immersed in a box of TIP3P−water[40] of dimensions $25 \times 27 \times 21$ A^3 and equilibrated for 50 ps.

Results

PROFEC Contour Maps. We were able to extract meaningful information from the PROFEC contour maps centered around C4 and Cl8 (Figure 4). Figure 4 (top) shows the zero level (i.e., $\Delta\Delta G_{ins} = 0$) PROFEC contour map centered around atom C4. The cavity around C4 suggests that addition of an atom/group would improve the binding of the inhibitor. Therefore, we added a methyl group in a cis position relative to the methyl group at position 5, and we denoted this derivative 45MeT (see Table 1). The contour map also partly overlaps the methyl group at position 5, and since it is unfavorable to have groups outside the 'cage' formed by the contour map (see Figure 4, top), a removal of this methyl group was predicted to improve binding. Thus, for the following two compounds we removed the C5 methyl group, and in one of these, we kept the methyl at the C4 position (4MeT). In the other, we added a chlorine in the C4 position (4ClT), since $\partial\Delta\Delta G_{ins}/\partial q$ (see Computational Methods) is positive in the map around the C4 cavity (blue in Figure 4), suggesting that the added group/atom should be electronegative. There is also a cavity around the atoms at positions 8 and 9, as seen from the contour map centered around Cl8 (Figure 4, bottom), and the multicolored plot suggests that the added group should be electroneutral. We added two new substituents at this position—a cyclohexyl group and a benzene ring condensed at positions 8 and 9—and these compounds are denoted "HET" and "BET", respectively (see Table 1).

Adaptive CMC/MD. Adaptive CMC/MD were run for two sets of derivatives, and the first set consisted of R86183, R82913, R84963, R80150, R84194, R80902, HET, BET, 4MeT, and 4ClT (see Table 4). The free energies of solvation (ΔG_{solv}) were subtracted from the energy "offsets" (see Computational Methods) that were obtained from the 450-ps simulation, and the relative free energies of binding are shown in Table 4. The values of ΔG_{solv}, which were estimated from GB/SA calculations, are shown in Table 5. From the first adaptive CMC/MD run we note that R80902 is clearly ranked as being the poorest inhibitor, in agreement with experiments, and we therefore discarded this compound in the next set. Moreover, HET and the similar BET were shown to be tight binding inhibitors (HET is about 2 kcal/mol better than R86183), so we also discarded these two derivatives in the next run. In set 2, the three discarded compounds were replaced with two derivatives with experimentally known binding affinities, R87027 and R84674, and we also included one new compound (45MeT, see Table 1). The second set was run for 560 ps, and for derivatives present in both sets we averaged the two estimates of the relative binding free energy. The rank order of the eight different TIBO derivatives with known experimental binding affinity according to adaptive CMC/MD is in good agreement with experiments (Table 4 and Figure 5, top) with an unsigned average error of the relative binding free energy of 1.0 kcal/mol. In this context, we should however point out that discrepancies between computed and experimental values can also be due to an imperfect agreement between HIV-1 RT activity and binding affinity caused by differences in cell-penetrating ability and metabolic stability between the TIBO derivatives. The three best binding derivatives, according to experimental results, were also ranked as the three best among the inhibitors with known binding affinity (bold numbers in Table 4). The binding free energy of R87027 and R84674 have both been underestimated with the adaptive CMC/MD method, which erroneously ranks then as better binders than R86183. The next three derivatives have almost the same experimental binding affinity, and they are also ranked between 4 and 6 among the derivatives with known experimental binding affinities. Finally the two derivatives with poorest experimental binding free energy, R84194 and R80902, have also been ranked as the worst binders among the derivatives with known binding affinity according to the adaptive CMC/MD method. Among the new derivatives, HET is ranked as being the best with a binding free energy of 2 kcal/mol better than that of R86183. BET is also one of the better inhibitors, whereas the other three PROFEC compounds were found to be poor binders.

PB/SA Calculations. ΔG_{solv}, as estimated with the PB/SA calculations (see Computational Methods), are in reasonable agreement with those obtained from the GB/SA method (Table 5). The relative solvation free energy ($\Delta\Delta G_{solv}$), which is the property of interest in this comparative study, has an average (unsigned) error of 0.4 kcal/mol between the two methods. Also from the PB/SA calculations, we obtain a relatively good agreement with the experimental rank order of binding to HIV-1 RT (Table 4 and Figure 5, bottom), and the
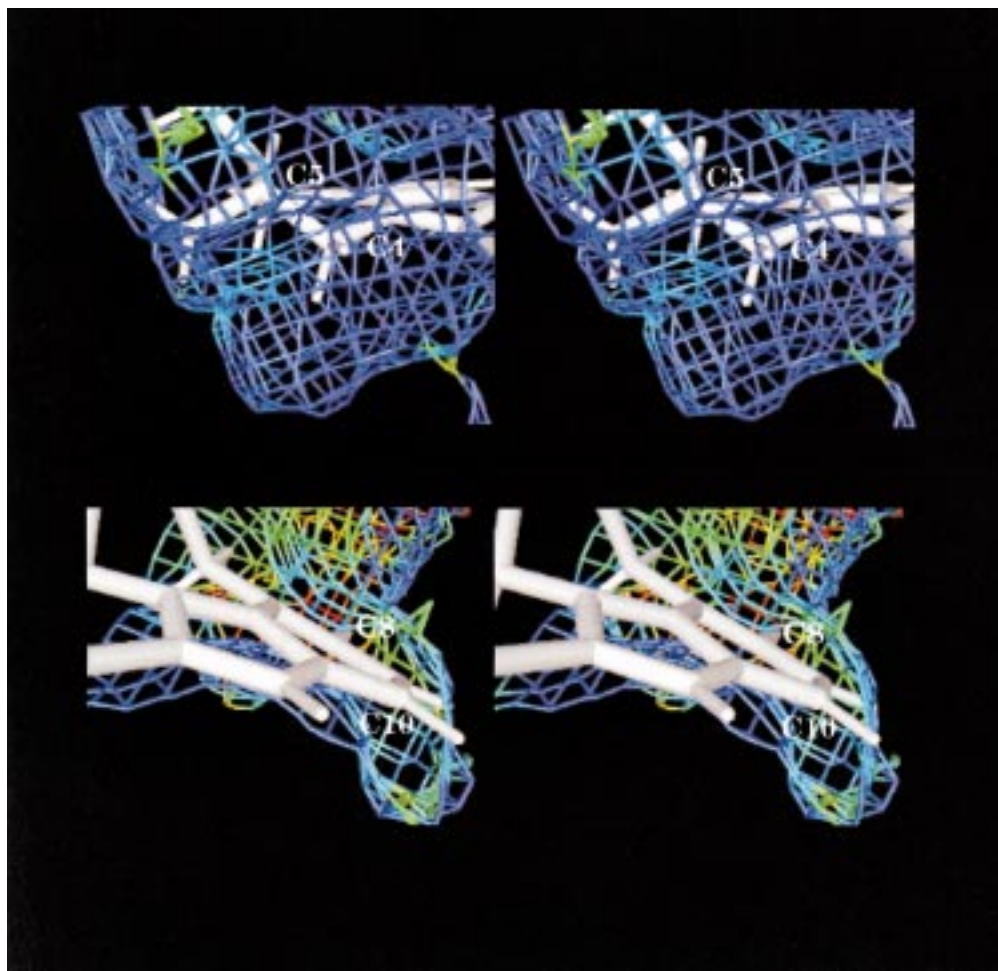
Figure 4. (Stereoviews) PROFEC[31] contour maps (contour level, $\Delta\Delta G_{ins} = 0$) around R86183, centered around atoms C4 (top) and 8Cl (bottom). Outside the cages that are formed from the contour maps, it is unfavorable to add a particle ($\Delta\Delta G_{ins} > 0$), whereas inside, an addition of a particle improves the binding free energy ($\Delta\Delta G_{ins} < 0$). The color of the map ranges from blue when $\partial\Delta\Delta G_{ins}/\partial q > 0$ to red when $\partial\Delta\Delta G_{ins}/\partial q < 0$. Green and yellow thus correspond to areas with $\partial\Delta\Delta G_{ins}/\partial q$ closer to 0 (see Computational methods).

Table 4. Binding Free Energies (relative to R86183, kcal/mol) and Rank Order of Binding to HIV-1 RT According to Adaptive CMC/MD, PB Calculations, and Experimental[32] Values

derivative	adaptive CMC/MD		PB		experimental	
	$\Delta\Delta G_{bind}$	rank	$\Delta\Delta G_{bind}$	rank	$\Delta\Delta G_{bind}$	rank
R86183	$(0^a+0^b)/2 = 0$	4 (**3**)	0	3 (**1**)	0	1
R87027	-2.56^b	1 (**1**)	1.87	7 (**4**)	0.06	2
R84674	-0.74^b	3 (**2**)	0.21	5 (**2**)	0.66	3
R82913	$(0.69^a+1.70^b)/2 = 1.19$	8 (**6**)	2.24	8 (**5**)	1.17	4
R84963	$(-0.12^a+1.28^b)/2 = 0.58$	6 (**4**)	2.31	9 (**6**)	1.27	5
R82150	$(0.67^a+1.15^b)/2 = 0.91$	7 (**5**)	2.67	11 (**7**)	1.34	6
R84914	$(0.78^a+2.01^b)/2 = 1.39$	9 (**7**)	0.86	6 (**3**)	3.05	7
R80902	3.71^a	12 (**8**)	5.19	13 (**8**)	4.04	8
HET	-1.94^a	2	-1.28	2		
45MeT	1.80^b	11	-1.47	1		
BET	0.50^a	5	0.09	4		
4MeT	$(1.11^a+1.74^b)/2 = 1.42$	10	2.36	10		
4ClT	$(2.02^a+1.58^b)/2 = 1.80$	11	4.19	12		

^{a} Set 1, 450-ps adaptive CMC/MD. ^{b} Set 2, 560-ps adaptive CMC/MD.

unsigned average error is 1.3 kcal/mol. The binding free energy of the derivative R84914 has been underestimated, and it thus has a too favorable ranking. When omitting this derivative, the rank order among the derivatives with known binding affinity coincides with that of the experiment except for R87027 and R84674, where the rank order is reversed. The derivative 45MeT is ranked as being the best binding derivative, closely followed by HET, which according to this method is

ranked as number 2. BET is also a good binder with this method, whereas 4MeT and 4ClT are both poor binding inhibitors as also was found in the adaptive CMC/MD method. We will discuss possible reasons for the large discrepancy between the two methods for 45MeT below. Dissecting the terms of ΔG_{bound} (see Computational Methods) we find that the vdW energy between the derivative and HIV-1 RT (Table 6) is the most favorable of all derivatives for HET closely followed

Table 5. Estimated ΔG_{solv} (kcal/mol) of the TIBO Derivatives, Using GB/SA, PB/SA, and TI Calculations

	GB/SA		PB		TI
TIBO derivative	ΔG_{solv}	$\Delta G_{solv} - \Delta G_{solv}$ (R86183)	ΔG_{solv}	$\Delta G_{solv} - \Delta G_{solv}$ (R86183)	$\Delta G_{solv} - \Delta G_{solv}$ (R86183)
R86183	−5.14	0	−3.92	0	0
R82150	−5.50	−0.36	−4.18	−0.26	−0.73
R82913	−5.21	−0.07	−4.18	−0.26	−0.60
R80902	−7.12	−1.98	−5.00	−1.08	−3.04
R84674	−4.48	0.66	−3.99	−0.07	
R84963	−4.38	0.76	−4.30	−0.38	
R84914	−5.20	−0.06	−3.89	0.03	
R87027	−4.99	0.15	−3.69	0.23	
4ClT	−6.41	−1.27	−4.86	−0.94	
4MeT	−4.96	0.18	−3.87	0.05	
45MeT	−4.45	0.69	−3.59	0.33	
HET	−4.19	0.95	−3.43	0.49	
BET	−5.28	−0.14	−4.27	−0.35	

Table 6. Energy Quantities (kcal/mol, eqs 5 and 7) for Calculation of Binding Free Energies, According to the PB/SA Method

derivative	$\Delta G_{react,L}^{g \to ag}$(LP)	$\Delta G_{lig-prot,vdW}$	$\Delta G_{lig-prot,elec}$	ΔG_{solv}	ΔG_{bind}
R86183	0.175	−52.71	−6.45	−3.92	−55.06
R87027	0.209	−50.64	−6.47	−3.69	−53.19
R84674	0.505	−52.05	−7.30	−3.99	−54.86
R82913	0.555	−51.00	−6.56	−4.18	−52.82
R84963	0.460	−50.12	−7.40	−4.30	−52.76
R82150	0.520	−49.84	−7.26	−4.18	−52.40
R84914	0.447	−51.50	−7.04	−3.89	−54.20
R80902	0.378	−48.95	−6.30	−5.00	−49.87
HET	0.520	−53.00	−7.29	−3.43	−56.34
45MeT	0.307	−52.87	−7.56	−3.59	−56.53
BET	0.505	−52.76	−6.99	−4.27	−54.97
4MeT	0.294	−49.52	−7.35	−3.87	−52.71
4ClT	0.196	−48.52	−7.41	−4.86	−50.87
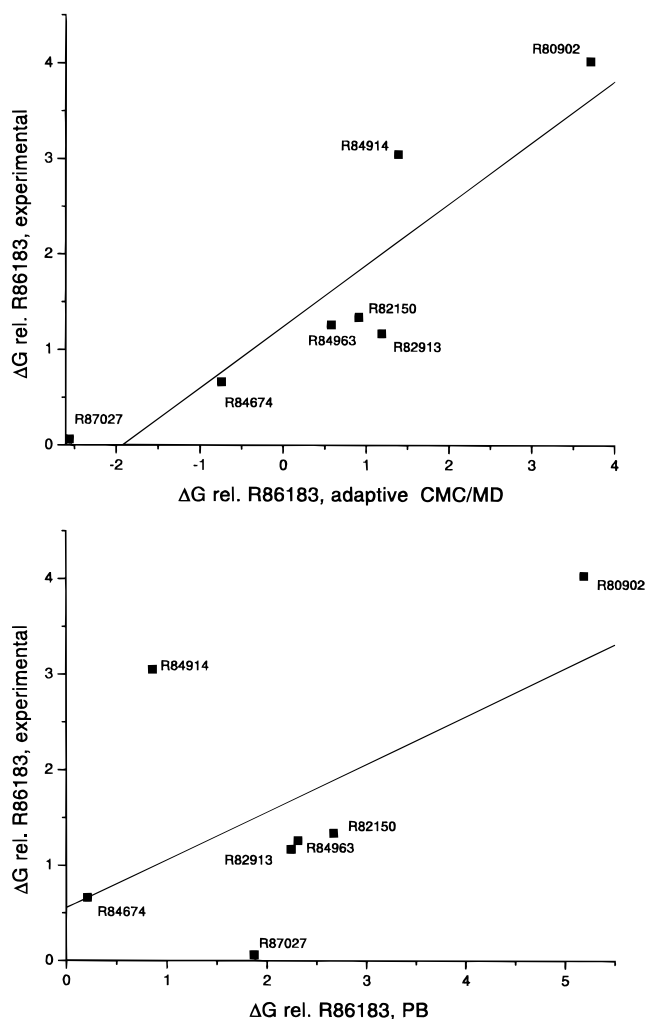
Figure 5. Estimated binding free energies (kcal/mol) relative to R86183 according to CMC/MD (top) and PB/SA (bottom) calculations, plotted against experimental values.[32] The least-squares linear fits are also shown in the plots.

by 45MeT and BET. From this table it is also apparent that the vdW energy is the most important term, determining almost solely the strength of binding the derivatives to HIV-1 RT. This is expected since most of the variation between the derivatives consists of modifications to hydrophobic groups.

Free Energy Calculations (TI and FEP). The relative free energies of solvation ($\Delta\Delta G_{solv}$) for R86183, R82150, and R80902 as estimated with the TI method (Table 7) are all well converged and in qualitative agreement with those estimated from the GB/SA and PB/SA calculations (Table 5). A much poorer convergence is found for the perturbations in HIV-1 RT despite the prolonged equilibration and data collection times (Table 8). In both the R82150 → R86183 and R82913 → R86183 perturbations, the vdW interaction is almost solely responsible for the hysteresis between the forward and reverse runs. We also note that differences in vdW interactions are responsible for the difference in binding free energy between R86183 and R82150. This is consistent with these interactions dominating the differences in binding free energies as found from the PB calculations (Table 6). Considering that TIBO is bound in a pocket, with predominantly hydrophobic and aromatic residues (see Figure 1), it is not surprising that differences in binding strength are governed by vdW interactions. The relative binding free energy of −1.9 ± 0.5 kcal/mol that we obtain is in close agreement with experimental results[32] (−1.34 kcal/mol). In the R82913 → R86183 perturbations, the contributions to the differences in binding affinities are shared between vdW and pmf contributions, whereas the electrostatic contribution to the difference in binding affinity is negligible. Here, we get a $\Delta\Delta G_{bind}$ of −3.2 ± 0.5 kcal/mol, in qualitative agreement with experiments (−1.17 kcal/mol).

The relative free energy of perturbing HET to R86183, via the intermediate (see Figure 3), is estimated to −0.8 ± 0.7 kcal/mol (Table 9). This agrees with the other two methods, supporting the prediction that HET should have improved affinity for HIV-1 RT compared to the parent compound R86183. A comparison of the individual contributions to the stability, summing over both steps of the perturbation, shows that HET is a tighter binder mainly because of a stronger vdW interaction (Table 9). This is also consistent with the PB/SA calculations, where the inhibitor−protein vdW interaction energy was strongest for HET.

Test of Chlorine Parameters. The results from the TI calculations are well converged, with very close values for the forward and reverse runs (Table 10). We obtain a $\Delta\Delta G_{solv}$(benzene − chlorobenzene) of −0.19 ± 0.06 kcal/mol which is in reasonable agreement with the experimental result[55] of 0.12 kcal/mol.

Table 7. Relative Solvation Free Energies (kcal/mol) of a Selected Set of TIBO Derivatives Estimated with the TI Method

	in solution			in vacuo		
	fwd	rev	av	fwd	rev	av
			R82150 → R86183			
ΔG_{tot}	1.531	1.691	1.611 ± 0.080	0.858	0.905	0.882 ± 0.023
$\Delta G_{el.stat}$	2.522	2.697	2.609 ± 0.072	1.843	1.848	1.846 ± 0.002
ΔG_{vdW}	0.699	0.703	0.701 ± 0.002	1.319	1.202	1.261 ± 0.059
ΔG_{pmf}	−1.690	−1.709	−1.700 ± 0.010	−2.303	−2.145	−2.224 ± 0.080
$\Delta\Delta G_{solv}$(R86183 − R82150) = 0.73 ± 0.08 kcal/mol						
			R82913 → R86183			
ΔG_{tot}	1.658	1.787	1.722 ± 0.064	1.113	1.081	1.118 ± 0.006
$\Delta G_{el.stat}$	1.341	1.377	1.35 ± 0.018	1.134	1.140	1.137 ± 0.003
ΔG_{vdw}	1.677	1.750	1.714 ± 0.036	1.595	1.605	1.600 ± 0.005
ΔG_{pmf}	−1.358	−1.341	−1.350 ± 0.009	−1.616	−1.621	−1.618 ± 0.003
$\Delta\Delta G_{solv}$(R86183 − R82913) = 0.60 ± 0.06 kcal/mol						
			R80902 → R82150			
ΔG_{tot}	26.148	26.250	26.199 ± 0.051	23.87	23.903	23.886 ± 0.016
$\Delta G_{el.stat}$	23.111	23.174	23.142 ± 0.032	22.052	21.964	22.008 ± 0.044
ΔG_{vdw}	3.024	2.986	3.005 ± 0.019	0.549	0.456	0.502 ± 0.046
ΔG_{pmf}	0.013	0.090	0.052 ± 0.038	1.269	1.480	1.374 ± 0.106
$\Delta\Delta G_{solv}$(R82150 − R80902) = 2.31 ± 0.05 kcal/mol						

Table 8. Relative Free Energies (kcal/mol) of Binding to HIV-1 RT for a Selected Set of TIBO Derivatives Estimated with the TI Method

	in HIV-1 RT			in solution		
	fwd	rev	av	fwd	rev	av
			R82150 → R86183			
ΔG_{tot}	−0.777	0.127	−0.325 ± 0.452	1.531	1.691	1.611 ± 0.080
$\Delta G_{el.stat}$	2.707	2.202	2.454 ± 0.253	2.522	2.697	2.609 ± 0.072
ΔG_{vdW}	−1.653	−0.271	−0.962 ± 0.691	0.699	0.703	0.701 ± 0.002
ΔG_{pmf}	−1.832	−1.727	−1.780 ± 0.105	−1.690	−1.709	−1.700 ± 0.010
$\Delta\Delta G_{bind}$(R86183 − R82150) = −1.9 ± 0.5 kcal/mol, exptl value:[32] −1.34 kcal/mol						
			R82913 → R86183			
ΔG_{tot}	−2.109	−0.784	−1.447 ± 0.662	1.658	1.787	1.722 ± 0.064
$\Delta G_{el.stat}$	1.131	1.372	1.25 ± 0.12	1.341	1.377	1.35 ± 0.018
ΔG_{vdW}	−0.658	0.091	−0.28 ± 0.37	1.677	1.750	1.714 ± 0.036
ΔG_{pmf}	−2.582	−2.247	−2.41 ± 0.17	−1.358	−1.341	−1.350 ± 0.009
$\Delta\Delta G_{bind}$(R86183 − R82913) = −3.2 ± 0.7 kcal/mol, exptl value:[32] −1.17 kcal/mol						

Table 9

Two-Step Free Energy Calculation (kcal/mol) of R86183 → HET

	in HIV-1 RT			in solution		
	run 1[a]	run 2	av	run 1	run 2	av
			Step 1 (FEP)			
ΔG_{tot}	−4.043	−5.333	−4.668 ± 0.645	−2.830	−2.232	−2.531 ± 0.299
$\Delta G_{el.stat}$	−3.475	−4.308	−3.892 ± 0.417	−3.094	−3.178	−3.136 ± 0.042
ΔG_{vdW}	−1.723	−2.403	−2.063 ± 0.340	−0.657	0.268	−0.194 ± 0.462
ΔG_{14vdW}	2.854	2.818	2.836 ± 0.018	3.020	3.021	3.020 ± 0.001
$\Delta G_{14elstat}$	−1.655	−1.410	−1.532 ± 0.122	−2.094	−2.088	−2.091 ± 0.003
			Step 2 (TI)			
ΔG_{tot}	1.963	2.324	2.144 ± 0.180	0.838	0.806	0.822 ± 0.016
ΔG_{vdW}	0.513	0.719	0.616 ± 0.103	−0.088	−0.133	−0.110 ± 0.022
ΔG_{14vdW}	1.048	1.036	1.042 ± 0.006	0.860	0.897	0.878 ± 0.018
ΔG_{badh}[b]	0.027	0.021	0.024 ± 0.003	0.020	0.035	0.028 ± 0.008
ΔG_{pmf}	0.389	0.556	0.472 ± 0.083	0.050	0.010	0.075 ± 0.020
			Summary Table from the Two Perturbation Steps			

	ΔG_{tot}(in HIV-1 RT)	ΔG_{tot}(in solution)	ΔG_{tot}(in HIV-1 RT) − ΔG_{tot}(in solution)
step 1	−4.668 ± 0.645	−2.531 ± 0.299	−2.137 ± 0.711
step 2	2.144 ± 0.180	0.822 ± 0.016	1.322 ± 0.181
step 1 + step 2	−2.524 ± 0.670	−1.709 ± 0.299	−0.815 ± 0.733
$\Delta\Delta G_{bind}$(HET − R86183) = −0.82 ± 0.73 kcal/mol			

[a] See Computational Methods. [b] Free energy contribution from bonds, angles, and dihedrals.

Discussion

The "full" free energy calculations on TIBO derivatives with known experimental binding affinities were performed as an initial check whether it was possible to reproduce the relative binding affinities with the most rigorous method prior to applying the more approximate approaches to this system. Despite the relatively poor convergence for the R82150 → R86183 and R89193 →

Table 10. Thermodynamic Integration (kcal/mol), Chlorobenzene → Benzene

	in solution			in vacuo		
	fwd	rev	av	fwd	rev	av
ΔG_{tot}	1.980	1.887	1.93 ± 0.05	2.072	2.161	2.12 ± 0.04
$\Delta G_{el.stat}$	0.764	0.600	0.68 ± 0.08	1.485	1.486	1.485 ± 0.001
ΔG_{vdW}	0.905	0.931	0.92 ± 0.01	0.150	0.151	0.15 ± 0.001
ΔG_{pmf}	0.311	0.356	0.33 ± 0.02	0.437	0.525	0.48 ± 0.04
$\Delta\Delta G_{solv}$(benzene − chlorobenzene) = -0.19 ± 0.06 kcal/mol, exptl value:[55] 0.12 kcal/mol						

R86183 perturbations, they both give reasonable estimates for the relative free energy. However, we were only able to qualitatively reproduce the relative binding free energy of R80902 versus R82150, with significant changes in the nonbonded parameters of sulfur (described in Computational Methods). These calculations show that the nonbonded interactions of the sulfur atom are strongly contributing to the erroneous result we obtained for our initial set of parameters (Table 2). Since our initial nonbonded parameters for sulfur were the same as those used in cysteine and methionine, it is not unreasonable that the nonbonded parameters for sp^2-sulfur (as in TIBO) have a smaller R^* and a larger ϵ than found for an sp^3 single-bonded sulfur. However, even with the modified sulfur parameters there is a significant quantitative difference (1.7 kcal/mol) between calculated and experimental $\Delta\Delta G_{bind}$ for R80902 → R82150 (C=O → C=S). Interestingly, the PB/SA and CMC/MD methods both estimate this $\Delta\Delta G_{bind}$ in close agreement with the experimental value (Table 4) even with the initial set of sulfur nonbonded parameters. We do not understand why the "most rigorous" approach is less accurate in this regard, but this is further support for the use of multiple methods to make binding free energy predictions.

The derivatives 4MeT and 4ClT are both estimated as being poor binders although we would expect them to bind better than R86183 from the PROFEC contour maps. However, it is not obvious that the cavity that was found around atom C4 is present when the C5 methyl group is also removed. Since the contour maps are based on simulations of a single parent compound (R86183), they give no information about what would happen with the cavity if other changes are made to the inhibitor. Comparing the minimized structures of 45MeT, 4MeT, and 4ClT from the PB calculations, we also observe small tendencies for the two latter derivatives to be pushed away from the original cavity at position C4. This is probably also reflected when comparing the inhibitor−protein vdW energies (Table 6). Both 4MeT and 4ClT have unfavorable vdW energies compared to 45MeT which is the major reason for their poor binding according to the PB calculations. The high solubility of 4ClT (Table 5) makes this inhibitor an even weaker binder. The very tight binding of 45MeT that was predicted by the PB method is not consistent with the results from the adaptive MC/MD runs, where this derivative instead is estimated as being a poor binder (Table 6). This discrepancy illustrates one of the most severe problems with the adaptive CMC/MD as currently implemented. Within the limited time of sampling, certain derivatives might be over- or undersampled, due to the fact that the binding mode of one derivative might not be favorable for another derivative.

Therefore, during the course of the CMC/MD run, a certain derivative may never (or rarely) interact optimally with the protein, especially if such an interaction would require rather large structural changes of the protein. This might be the case for 45MeT, in that rearranging nearby protein residues for an optimal interaction is a slower process than can be caught within our sampled time of 560 ps. The oversampling of R87027 (leading to a too favorable relative binding free energy) is most probably an example of the same problem, but reversed. The surrounding side chains of the protein have to rearrange in order to accommodate the relatively long "tail" of R87027 that results from the two added methyl groups (see Table 1). When performing an MC step with R87027 as the sampled inhibitor, it might be difficult for the other derivatives to find an optimal conformation of the protein, resulting in rejections of these trial steps (see Computational Methods). The algorithm might therefore get temporarily "stuck" sampling R87027 due to incompatible binding modes, leading to an overestimation of its binding free energy. The "adaptive" CMC/MD version was partly constructed in order to overcome this problem, and it is an improvement over the nonadaptive protocol, where the convergence was extremely slow for this system. However, convergence of the calculated free energies is still hampered by the problem described above. This difficulty might further be reduced by adding multiple conformers (rotamers) of some critical side chains of the protein and also permitting them to participate in the CMC sampling. Work is currently in progress (J. Pitera) to implement multiple copies of protein side chains, to improve the convergence rate. Another observation, comparing sets 1 and 2, is that the relative binding free energies are consequently higher in set 2 (except for 4ClT). The stems from the fact that the "reference derivative", R86183, is being estimated as a relatively better binder in set 2 than in set 1, which thus shifts the relative free energies of all the other compounds.

Most encouraging is that we find HET to be a much better binder than R86183 with both the adaptive CMC/MD and the PB/SA methods. "Full" free energy calculations also support this prediction (albeit less clearly due to the large error estimate), lending support to the conclusions of the other two methods. The physical picture of HET binding to HIV-1 RT also suggests that this affinity is plausible. The hydrophobic cyclohexyl moiety of TIBO fits very well into a pocket composed of many hydrophobic side chains: Val106, Tyr188, Phe227, and Leu234 (Figure 6). This is also consistent with the calculation that HET has the highest vdW interaction with the protein of all compared derivatives. Coupled with a more unfavorable solvation free energy, this yields a much greater affinity for HIV-1 RT than the parent compound, R86183.

From this study is it not possible to judge which of the two approximate methods (adaptive CMC/MD and PB/SA) performs best in ranking the derivatives with respect to the binding free energies since both methods have their strengths and shortcomings. In CMC/MD, on one hand, the solvent with counterions is modeled explicitly and a large number of conformations are sampled, but as now implemented it suffers from the slow convergence when the sampled inhibitors have
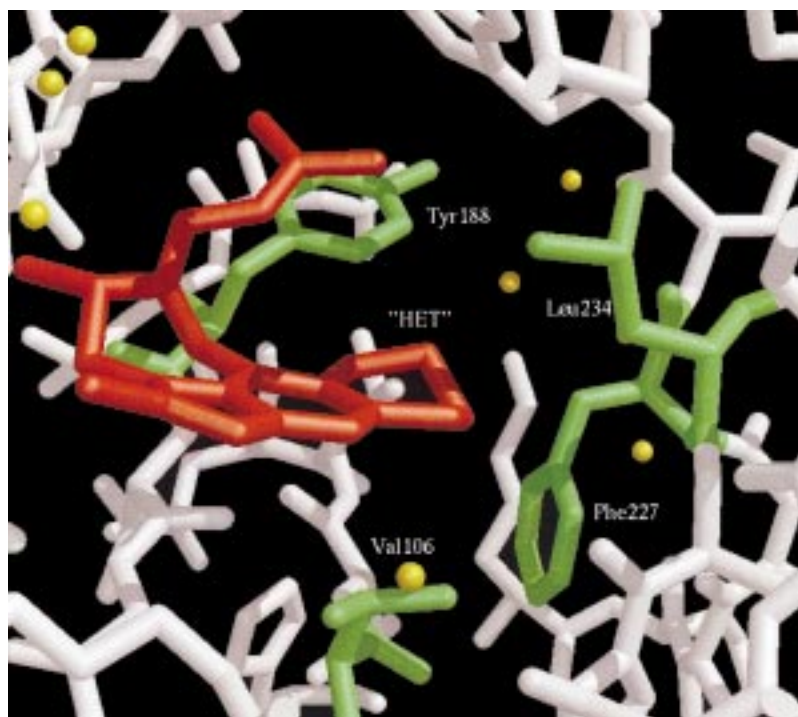
Figure 6. Snapshot from the MD simulation of HET (red) in HIV-1 RT (white). Hydrophobic residues with any atom < 3.5 Å from the cyclohexyl moiety atoms are shown in green, and water molecules are shown as yellow spheres.

differing binding modes. The PB/SA method, on the other hand, does not suffer from that problem, but the solvent is instead approximately described as a continuum and we only consider one conformation of the ligand and the complex, respectively. Since the two methods have different sets of approximations, we feel that they complement each other and that a consistent result between them increases its validity.

Conclusions

In this study we have used two methods in order to rank 13 different TIBO derivatives: the adaptive CMC/MD method and PB/SA calculations. Five of these derivatives were new modifications that were made from suggestions generated by PROFEC contour maps. The rest of the TIBO derivatives have experimentally determined[32] binding affinities. Both methods work surprisingly well, yielding rank orders in good agreement with experimental results. Since the two methods are quite different in nature, each with their own sets of approximations, they serve as a good complement to each other. That is, if both methods predict the same rank order, the reliability of this prediction will significantly increase. The methods are also relatively rapid—a total of 0.8-ns simulation time was needed in order to rank the 13 TIBO derivatives with the adaptive CMC/MD method. To put this in perspective, considerable simulation times (1.1 ns or more) were required in order to obtain an estimate of the relative binding free energy between only one pair of derivatives with a FEP/TI calculation. We found that one of the new modifications (HET), as suggested from PROFEC, was binding about 1−2 kcal/mol better than R86183 according to both methods. This result was confirmed with subsequent FEP/TI calculations. The hydrophobic cyclohexyl moiety that was added to TIBO fits well into a cavity in HIV-1 RT that consists of many nonpolar residues. This is also

consistent with the observation that HET has the most favorable vdW interaction with HIV-1 RT among the TIBO derivatives studied.

The protocol we have used in this paper (outlined in Figure 2) is a general strategy for computational structure-based lead optimization. While we have not considered crucial pharmacological issues such as bioavailability and toxicity, our approach appears to be useful for optimization of affinity. Starting from a parent lead compound (or a family of compounds), PROFEC can be used to suggest where modifications of the lead should be made to improve the binding affinity of the lead compound. PB/SA and adaptive CMC/MC can then be applied for ranking of the PROFEC derivatives, preferably together with derivatives of known binding affinity if such are available. Thereafter FEP/TI can be used to study particularly interesting derivatives and to confirm results from the more approximate methods, followed by synthesis and in vitro testing of the best binding derivative(s). However, as one reviewer noted, the expense of FEP/TI calculations suggests a revised strategy, where the compounds selected by CMC/MD and PB/SA are synthesized and tested, without carrying out a FEP/TI calculation. We agree that one can simply use any predictions from PROFEC, CMC/MD, or PB/SA directly, but if significant synthetic efforts are required, it is certainly worth a confirmatory calculation with FEP/TI to see if such efforts are justified. Regardless of whether FEP/TI calculations are used, we feel that the strategy used in this paper provides an excellent blueprint for lead optimization in structure-based drug design.

Note Added in Proof. While this work was in press, a paper by Smith et al. appeared (*J. Med. Chem.* **1998**, *41*, 5272–5286) which studied some of the same inhibitors with RT using a method close to those described in refs 26–30.

Supporting Information Available: Partial charges of the 13 TIBO derivatives. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stammers, D.; Stuart, D. Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. *Struct. Biol.* **1995**, *2*, 303–308.

(2) Larder, B. A. *Inhibitors of HIV reverse transcriptase as antiviral agents and drug resistance*. Cold Spring Harbor Lab. Press: Cold Spring Harbor, NY, 1993; pp 205–222.

(3) Schinazi, R. F. Competitive inhibitors of human immunodeficiency virus reverse transcriptase. *Perspect. Drug Discov. Des.* **1993**, *1*, 151–180.

(4) De Clercq, E. Antiviral therapy of human immunodeficiency virus infections. *Clin. Microbiol. Rev.* **1995**, *8*, 200–239.

(5) DeClercq, E. HIV resistance to reverse transcriptase inhibitors. *Biochem. Pharmacol.* **1994**, *47*, 155–169.

(6) Young, S. D. Nonnucleoside inhibitors of HIV-1 reverse transcriptase. *Perspect. Drug Discov. Des.* **1993**, *1*, 181–192.

(7) De Clercq, E. Toward improved anti-HIV chemotherapy: therapeutic strategies for intervention with HIV infections. *J. Med. Chem.* **1995**, *38*, 2491–2517.

(8) Arnold, E.; Das, K.; Ding, J.; Yadav, P. N. S.; Hsiou, Y.; Boyer, P. L.; Hughes, S. H. Targeting HIV reverse transcriptase for anti-AIDS drug design. *Drug Des. Discov.* **1996**, *13*, 29–47.

(9) 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 R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. *Struct. Biol.* **1995**, *2*, 407–415.

(10) Das, K.; Ding, J.; Hsiou, Y.; Clark, A. D. J.; Moereels, H.; Koymans, L.; Andries, K.; Pauwels, R.; Janssen, P. A. J.; Boyer, P. L.; Clark, P.; Smith, R. H. J.; Kroeger-Smith, M. B.; Michedja, 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.

(11) Tantillo, C.; Ding, J.; Jacobo-Molina, 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 three-dimensional structure of HIV-1 reverse transcriptase. *J. Mol. Biol.* **1994**, *243*, 369–387.

(12) Beveridge, D. L.; DiCapua, F. M. Free energy via molecular simulation: Application of chemical and biochemical systems. *Annu. Rev. Biophys. Biophys. Chem.* **1983**, *18*, 431–492.

(13) Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.

(14) Rao, B. G.; Kim, E. E.; Murcko, M. A. Calculation of solvation and binding free energy differences between VX-478 and its analogues by free energy perturbation and AMSOL methods. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 23–30.

(15) Pitera, J.; Kollman, P. Designing an optimum guest for a host using multimolecule free energy calculations: Predicting the best ligand for Rebek's "Tennis Ball". *J. Am. Chem. Soc.* **1998**, *120*, 7557–7567.

(16) Gilson, M. K.; Honig, B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies and conformational analysis. *Proteins: Struct. Funct., Genet.* **1988**, *4*, 7–18.

(17) Sharp, K. A.; Honig, B. Electrostatic interactions in macromolecules: theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301–332.

(18) Nozaki, Y.; Tanford, C. The solubility of amino acids and two glycine peptides in aqueous solution and dixonane solutions. Establishment of a hydrophobicity scale. *J. Biol. Chem.* **1971**, *246*, 2211–2217.

(19) Hermann, R. B. Theory of hydrophobic binding. II. The correlation of hydrocarbon solubility in water with solvent cavity surface area. *J. Phys. Chem.* **1971**, *76*, 2754–2759.

(20) Wendoloski, J. J.; Shen, J.; Oliva, M. T.; Weber, P. C. Biophysical tools for structure-based drug design. *Pharmacol. Ther.* **1993**, *66*, 169–183.

(21) Jackson, R. M.; Sternberg, M. J. E. A continuum model for protein–protein interactions: Application to the docking problem. *J. Mol. Biol.* **1995**, *250*, 258–275.

(22) Shen, J.; Quiocho, F. A. Calculation of binding energy differences receptor–ligand systems using the Poisson–Boltzmann method. *J. Comput. Chem.* **1995**, *16*, 445–448.

(23) Shen, J.; Wendoloski, J. Binding of phosphorus-containing inhibitors to thermolysin studied by the Poisson–Boltzmann method. *J. Comput. Chem.* **1996**, *17*, 350–357.

(24) Zhang, T.; Koshland, J. D. E. Computational method for relative binding energies of enzyme–substrate complexes. *Protein Sci.* **1996**, *5*, 348–356.

(25) Froloff, N.; Windemuth, A.; Honig, B. On the calculation of binding free energies using continuum methods: Application to MHC class I protein-peptide interactions. *Protein Sci.* **1997**, *6*, 1293–1301.

(26) 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.

(27) Åqvist, J.; Mowbray, S. L. Sugar recognition by a glucose/galactose receptor. *J. Biol. Chem.* **1995**, *270*, 9978–9981.

(28) Hansson, T.; Åqvist, J. Estimation of binding free energies for HIV-1 protease inhibitors by molecular dynamics simulations. *Protein Eng.* **1995**, *8*, 1137–1144.

(29) Paulsen, M. D.; Ornstein, R. L. Binding free energy calculations for P450cam-substrate complexes. *Protein Eng.* **1996**, *9*, 567–571.

(30) Åqvist, J. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *J. Comput. Chem.* **1996**, *17*, 1587–1597.

(31) Radmer, R. J.; Kollman, P. A. The application of three approximative free energy calculation methods to structure based ligand design: Trypsin and its complex inhibitors. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 215–227.

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

(33) Miyamoto, S.; Kollman, P. A. Absolute and relative binding free energy calculations of the interaction of biotin and its analogues with streptavidin using molecular dynamics/free energy perturbation approaches. *Proteins: Struct., Funct., Genet.* **1993**, *16*, 226–245.

(34) Fox, T.; Thomas, B. E.; McCarrick, M.; Kollman, P. A. Application of free energy perturbation calculations to the "tennis ball" dimer: Why is CF4 not encapsulated by this host? *J. Phys. Chem.* **1996**, *100*, 10779–10783.

(35) Liaw, Y. C.; Gao, Y. G.; Robinson, H.; Wang, A. H. J. Molecular structure of a potent HIV-1 inhibitor belonging to TIBO family. *J. Am. Chem. Soc.* **1991**, *113*, 1857–1859.

(36) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Jonhson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montegomery, J. A.; Rahavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. Gaussian 94, Revision B.3; Gaussian Inc.: Pittsburgh, PA.1995.

(37) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges − the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(38) Pearlman, D. A.; Case, D. A.; Caldwell, J. C.; Ross, W. S.; Cheatham, T. E., III; Ferguson, D. M.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. AMBER 4.1 (UCSF); University of California: San Francisco, 1994.

(39) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(40) Jorgensen, W. L.; Chandraseklar, 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.

(41) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(42) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(43) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. The Midas display system. *J. Mol. Graph.* **1988**, *6*, 13–27.

(44) Metropolis, N. R.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087−1092.

(45) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.

(46) Kumar, S.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The weighted histogram analysis for free energy calculations on biomolecules. 1. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(47) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(48) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. MacroModel − An integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* **1990**, *11*, 440−467.

(49) Gilson, M. K.; Sharp, K. A.; Honig, B. Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comput. Chem.* **1988,** *9*, 327−335.

(50) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995,** *268*, 1144−1149.

(51) Connolly, M. L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983,** *221*, 709−713.

(52) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-shifted scaling, a new method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* **1994,** *100*, 9025−9031.

(53) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994,** *222*, 529−539.

(54) Simmerling, C.; Fox, T.; Kollman, P. Use of locally enhanced sampling in free energy calculations: Testing and application to the $\alpha \rightarrow \beta$ anomerization of glucose. *J. Am. Chem. Soc.* **1998,** *120*, 5771−5782.

(55) Hine, J.; Mookerjee, P. K. The intrinsic character of organic compounds. Correlations in terms of structural contributions. *J. Org. Chem.* **1975,** *40*, 292−298.