Grand Canonical Free-Energy Calculations of Protein−Ligand Binding

Matthew Clark,* Sia Meshkat, and Jeffrey S. Wiseman

Locus Pharmaceuticals, Four Valley Square, 512 Township Line Road, Blue Bell, Pennsylvania 19422

Received December 2, 2008

The principles behind the computation of protein−ligand binding free energies by Monte Carlo simulation in the Grand Canonical Ensemble are described in detail, and two variations of the calculation are presented. The computation can be performed by bathing a protein binding site either with ligand images that interact with each other or with ligand images that pass through each other. The second method is theoretically more rigorous, but we show that both methods lead to the same result, and there are distinct numeric advantages to using ligand images that interact with each other. The Grand Canonical simulation provides gas-phase binding free energies that can be converted to aqueous energies by generalized Born-surface area (GB/SA) solvation calculations to provide values that agree with experiment within ±1.5 kcal/mol. However, the accuracy of these simple solvation calculations is a major limiting factor in the accuracy of the overall binding free-energy computation. The Grand Canonical simulation has several characteristics beneficial to free-energy calculations. One is that the number of parameters that must be set for the simulation is small and can be determined objectively, making the outcome more deterministic, with respect to choice of input conditions, as compared to perturbation methods. Second, the simulation is free from assumptions about the starting pose or nature of the binding site. A final benefit is that binding free energies are a direct outcome of the simulation, and little processing is required to determine them.

INTRODUCTION

Discovering small molecules that bind a target protein tightly, with specificity, and that have desirable pharmacological properties is a difficult challenge. Fragment-based drug design is emerging as a powerful tool to address this challenge,[1] and the ability to accurately compute fragment-protein binding affinities should be a significant addition to the fragment-based design toolset. A computational method would not be limited by properties such as solubility and binding affinity, for example, that restrict the fragments that can be used in experimental methods such as nuclear magnetic resonance (NMR) and crystallography. The Monte Carlo Grand Canonical method described in this work promises to provide a simple method to quickly compute accurate ligand−binding free energies for the prediction of binding affinities. The method is very efficient in simulating the binding of small, rigid ligands to proteins and would be well-suited for quantitatively assessing fragment−protein binding affinities in fragment-based design.

Other rapid methods, such as docking, have shown great utility in quickly screening large libraries of molecules and scoring their ability to fit in a given binding site. While these methods can reproduce binding poses accurately, they have performed poorly at predicting binding affinities.[2] To be predictive, methods must focus on computing the binding free energy, which is directly related to the physically measurable K_d value; and significant effort has been devoted over the last decades to developing fast, robust methods to meet this goal.[3,4]

Binding free-energy calculations can be classified into two general approaches. Pathway approaches simulate the path of the molecule from the nonbound to the bound state. The potential of mean force method (PMF) is one pathway approach, which computes binding by integrating the work required to pull the bound ligand out of the binding site into bulk solvent. The free-energy perturbation (FEP) methods are a second example of the pathway class. These methods alchemically grow or remove ligands in the binding site and have been among the most used over the past decades.[5] FEP methods include the "double annihilation"[6] and "double decoupling"[7] methods.

The second family of methods, called "endpoint" methods, compute only properties of the bound and free states. The simplest of these is the linear interaction energy (LIE) model, which combines quantitative structure−activity relationships (QSAR) principles with molecular mechanics interaction energies to create a relationship that relates the average van der Waals and electrostatic energies to observed binding energies.[8] The molecular mechanics/generalized Born-surface area models (MM/GB/SA) or Poisson−Boltzmann surface-area models (MM/PB/SA) are more rigorous end-point methods. These methods use molecular dynamics (MD) to sample the binding poses available in the protein, while computing the solvation free energy using the GB/SA or PB/SA methods, as shown in eq 1.[9] For these methods, the gas-phase binding free energy ΔU_{gas} is computed using a molecular mechanics potential, while the ΔG_{solv} is computed using the GB-SA or PB-SA method. The $T\Delta S$ term is approximated either through a quasi-harmonic analysis or from MD simulations.

$$\Delta G_{binding} = \Delta U_{gas} + \Delta G_{solv} - T\Delta S \qquad (1)$$

All of these methods have demonstrated the ability to provide reasonable results for a variety of ligands and proteins. The

* To whom correspondence should be addressed. E-mail address: mclark@drmatthewclark.com.

10.1021/ci8004397 CCC: $40.75
Published on Web 03/23/2009

molecular mechanics model, for example, provides binding free energies with reported errors of ∼1 kcal/mol.[10,11] However, these methods are sensitive to the starting pose of the simulation. Each process begins with a starting pose defining the protein−ligand binding mode and computes the free energy of that pose. In some cases, the choices for the best starting pose are unambiguous; however, in other cases, ligands have a variety of possible binding modes. To fully study a set of proposed ligands, each ideally should be subjected to the free-energy calculation for a variety of starting points to sample many binding modes, which can become a prohibitively expensive calculation.

Because discovering novel drugs often involves exploring novel binding modes and interactions, there is a need for methods that can use free-energy calculations to efficiently scan the binding region for novel tight-binding modes. We have recently described a Grand Canonical free-energy method, which is a member of the end point family and provides this capability.[12,13] The Grand Canonical method was explored by Mezei and co-workers[14−17] and used to explore water binding to the DNA minor groove.[18] Resat has combined Grand Canonical sampling with the potential of mean force method of free-energy calculation.[19] The Grand Canonical ensemble has been used independently by several groups to explore protein−water interactions.[20−22] The Grand Canonical method computes the free energy of ligands binding to proteins by equilibrating concentrations between a reference state and a simulation cell that includes ligands bound to a protein surface. The Grand Canonical method is the inverse of other free-energy methods in that, instead of computing the binding free energy of each possible pose independently, it locates ensembles of poses consistent with a selected free-energy level. Thus, the results are not biased by the choice of a starting pose and require only a modest computational effort.

Grand Canonical ensembles have historically been used in chemistry to study systems such as the binding of gases to solid supports,[23−25] and excess chemical potential of electrolyte solutions.[26,27] Because Grand Canonical ensembles are relatively new in drug design applications, and because the protein binding problem differs in principle from historical applications of the method, the primary purpose of this manuscript is to describe how the results of a Grand Canonical simulation are interpreted in this new context. In addition, the specific method described herein represents an improvement over our original description of this method,[4] being both faster and theoretically more rigorous for studying ligand−protein interactions.

Bacteriophage T4 lysozyme has an important role in the hydrolysis of bacterial cell walls.[28] In this study, we use the artificial binding pocket of the T4 lysozyme mutant L99A to test the method.[29,30] The T4 binding pocket is a useful test case, because it is completely enclosed by the protein and admits only fragment-sized ligands with limited conformational flexibility. T4 lysozyme has the additional advantage of being reasonably rigid, with limited water molecules observed in the binding site in the presence of ligands.[31,32] There is a wealth of calorimetry and crystallographic data determined for T4 ligands that can be compared to the computed results; in addition, the binding of T4 ligands has been widely simulated by other groups, which allows comparison of a variety of free-energy computation methods.[33−37]

METHODS

T4-Lysozyme. A series of co-crystal structures and thermodynamic binding data are available for T4 lysozyme.[29,30] The 186 L co-crystal structure with the largest ligand, *n*-butyl benzene, was used for all simulations and was prepared by removing the ligand structure, adding H atoms and reorienting the protein to fit into the smallest rectangular box aligned along the Cartesian axes. The ligands were prepared by energy minimization with the AMBER* force field in MacroModel,[38,39] computing charges by the CHELPG method using Gaussian 98.[40] The solvation gas-to-aqueous transition free energies were computed for each fragment using the MacroModel GB/SA computation, and subtracted from the gas-phase binding free energy computed by the Monte Carlo simulation.

The Grand Canonical sampling, data collection, and analysis software was developed by Locus in the C++ language. The Lennard-Jones and electrostatic terms of the AMBER 2003 force field were implemented to compute the protein−ligand interaction energies.[41] Because the ligand is held rigid, only the nonbonded force field terms are used. To ensure high-quality random sampling when very long simulations are performed, a Mersenne Twister random number generator was implemented, providing a nominal random period of $2^{19937} - 1$.[42] The simulations were performed on a cluster of 64 computers with 2.6 GHz Intel Core2 Duo 6700 processors.

Free Energies Computed with the Grand Canonical Ensemble. The Grand Canonical method computes the concentration of ligand in a simulation cell, $[L_{\text{simulation cell}}]$, in equilibration with an infinite reference bath of ligand at a concentration $[L_{\text{ideal}}]$, which is treated as an ideal gas at a concentration set by the simulation parameters. The free energy of the system is then computed with eq 2. In the presence of a protein, the favorable interaction of the ligand with the protein results in a higher concentration of ligand in the simulation cell, compared to the ideal reference cell. The free energy is relative to the ideal gas concentration; if no protein is present, the concentration of the cell converges to a concentration identical to the ideal gas, leading to a binding free energy of zero. As discussed further below, if we do not allow ligand−ligand interactions in the simulation cell, then accumulation of the ligand in one binding pocket of the protein will be independent of accumulation elsewhere in the simulation cell.

$$\Delta G = -kT \ln\left(\frac{[L_{\text{simulation cell}}]}{[L_{\text{ideal}}]}\right) \tag{2}$$

The computational process simply simulates the ligand at various concentrations $[L_{\text{ideal}}]$ and measures the resulting concentration of the complex $[L_{\text{simulation cell}}]$. Computational details used to compute the simulation cell and ideal concentrations are presented in the next section.

Derivation of the Grand Canonical Method. The thermodynamics of the Grand Canonical ensemble, along with the canonical, and microcanonical ensembles, were first derived by J. Willard Gibbs in 1902.[43] The standard Grand

Canonical partition function is given by eq 3. $Q(N,V,T)$ is the standard canonical partition function given from statistical mechanics.

$$\Xi(\mu, T, V) = \sum_N \exp\left(\frac{\mu N}{kT}\right) Q(N, V, T) \tag{3}$$

$$Q(T, V, N) = \left(\frac{V^N}{\Lambda^{3N} N!}\right)\left[\frac{Z(T, V, N)}{V^N}\right] \tag{4}$$

The function $Z(T,V,N)$ is the configuration integral that describes the deviation from ideal gas behavior that is due to intermolecular interactions. The variable Λ is the thermal de Broglie wavelength, which is dependent on the temperature and mass of the molecule, as in eq 5:

$$\Lambda = \left(\frac{h^2}{2\pi mkT}\right)^{1/2} \tag{5}$$

Combining these equations provides the partition function in a convenient form:

$$\Xi(\mu, T, V) = \sum_N \left\{\frac{\exp[\mu N/(kT)]}{\Lambda^{3N} N!}\right\} Z(N, V, T) \tag{6}$$

Therefore, the probability of a finding a given configuration i in the (T,V,μ) ensemble is given by eq 7:

$$P_i = \left(\frac{1}{\Xi}\right)\left(\frac{1}{\Lambda^{3N_i}}\right) \exp\left(\frac{\mu N_i - E_i}{kT}\right) \tag{7}$$

The relative probabilities of the two states P_i and P_j is required to be equal to the probability of the transition P_i to P_j, which is denoted as p_{ij}. The chance of executing a insertion, or deletion, is the product of the probability of attempting the step q_{ij} and the probability of accepting the attempt α_{ij}. For the sake of example, herein we assume that the transition i to j is an insertion, and j to i is a deletion of a molecule.

$$p_{ij} = q_{ij}\alpha_{ij} \tag{8}$$

The probability of attempting an insertion into a particular volume is inversely proportional to the system volume. The probability of choosing any particular molecule for a delete attempt is inversely proportional to the number of molecules in the system. The use of a reference concentration N_0/V_0 gives the probability of the transition at a selected reference concentration (usually 1 M):

$$q_{ij} \propto \frac{V_0}{V} \tag{9}$$

$$q_{ij} \propto \frac{N_0}{N} \tag{10}$$

The requirement for microscopic reversibility demands that the proportion of the probabilities be the same as the transitions between them, as given in eq 14, which is derived from the probability of eq 7:

$$\frac{P_j}{P_i} = \frac{p_{ij}}{p_{ji}} \tag{11}$$

$$\frac{P_j}{P_i} = \frac{\alpha_{ij} q_{ij}}{\alpha_{ji} q_{ji}} \tag{12}$$

Substituting eqs 9 and 10 into eq 12 gives

$$\frac{P_j}{P_i} = \frac{\alpha_{ij} N_0}{\alpha_{ji} V_0} \tag{13}$$

This example assumes that the transition p_{ij} involves a molecule entering the system, $N_j = N_i + 1$, and, thus, the number of molecules N is eliminated from the exponent of eq 7. The factor of N in the exponent of the thermal de Broglie wavelength is removed in the same manner:

$$\frac{p_{ij}}{p_{ji}} = \frac{P_j}{P_i} = \frac{1}{\Lambda^3} \exp\left(\frac{\mu}{kT} - \frac{E_i - E_j}{kT}\right) \tag{14}$$

The excess chemical potential, or free energy, of the system is relative to the free energy of the ideal gas from statistical thermodynamics, which is given in eq 15. The ideal gas has no intermolecular interactions, and, therefore, no potential function is involved; the free energy is a general outcome of statistical thermodynamics of the concentration of the molecules.

$$\mu_{\text{ideal}} = -kT\left[\ln\left(\frac{V}{N}\right) + \left(\frac{3}{2}\right)\ln\left(\frac{2\pi mkT}{h^2}\right)\right] \tag{15}$$

The equation for the thermal de Broglie wavelength can be substituted, providing eq 16, which gives the ideal chemical potential for a specific N,V,T ensemble:

$$\mu_{\text{ideal}} = kT \ln\left(\frac{N\Lambda^3}{V}\right) \tag{16}$$

Substituting the excess chemical potential $(\mu - \mu_{\text{ideal}})$ from eq 16 removes the dependence on the de Broglie wavelength. This removes the mass of the molecule from the equation, thus allowing simpler comparisons of energies among different species.

$$\frac{p_{ij}}{p_{ji}} = \frac{N_0}{V_0}\frac{V \exp\{[(\mu - \mu_{\text{ideal}})/(kT)] + \ln(N_j) - [(E_j - E_i)/(kT)]\}}{N_j} \tag{17}$$

This is further simplified by introducing B, which is related to the excess chemical potential, as defined in eq 18.

$$B = \frac{\mu - \mu_{\text{ideal}}}{kT} + \ln(N_j) \tag{18}$$

Substituting eq 18 into eq 17, the transition probability, in terms of B, is given in eq 19:

$$\frac{p_{ij}}{p_{ji}} = \frac{N_0}{V_0}\frac{V \exp\{B - [(E_j - E_i)/(kT)]\}}{N_j} \tag{19}$$

The standard chemical potential (μ_0) is defined by the reference concentration N_0/V_0, and implicitly by the choice of the variable B. Generally. the reference concentration N_0/V_0 is chosen to be 1 M, which is 1 molecule/(1660 Å^3).

During the Monte Carlo simulation, molecules can translate and rotate as well as enter and leave the system. A random algorithm is used to attempt these moves and the

Table 1. Free Energies Computed for Benzene in the T4 Lysozyme L99A Binding Pocket with No Ligand−Ligand Interactions

log(concentration in reference box), B'	average population of noninteracting ligands in binding pocket[a]	ligand concentration in binding pocket[b] (M)	log(concentration in binding pocket)	computed free energy (kcal/mol)[c]
0	18497	40200	4.60	−6.3[d]
−2	14146	30000	4.48	−8.8[d]
−4	724	1540	3.19	−9.8
−6	8	20	1.31	−9.9
−7.14	0.5	1	0.059	−9.8
−8	0.05	0	−0.73	−9.9

[a] Simulations were run for at least 10^8 steps to reach equilibration, and ligand populations were subsequently determined by averaging over 10^7 steps. [b] The free volume of T4 L99A binding pocket is 782.5 Å^3. [c] Computed using eq 2. [d] The free energy did not converge after more than 4×10^8 steps.

probability of accepting them is given by eqs 20−22. In these equations, the term ΔE refers to the change in potential energy for the move being attempted. Equation 20 gives the probability of accepting the insertion of a new molecule into the system, moving from state j (N molecules) to state i ($N + 1$ molecules). Equation 21 gives the probability of a chosen molecule leaving the system state i ($N + 1$ molecules) to state j (N molecules). For insertion, the number of molecules changes from N to $N + 1$; for deletion, the molecules change from $N + 1$ to N for detailed balance. Equation 22 is the probability of accepting the translation or rotation of a molecule without changing the number N. This is the probability of the standard canonical step.

$$\alpha_{\text{insert}} = \min\left[1, \frac{N_0}{V_0}\exp\left(\frac{-\Delta E}{kT} + B\right)\frac{V}{N+1}\right] \quad (20)$$

$$\alpha_{\text{delete}} = \min\left[1, \frac{V_0}{N_0}\exp\left(\frac{-\Delta E}{kT} - B\right)\frac{N}{V}\right] \quad (21)$$

$$\alpha_{\text{move}} = \min\left[1, \exp\left(\frac{-\Delta E}{kT}\right)\right] \quad (22)$$

In the Monte Carlo/Metropolis simulation process, a new configuration is generated using one or more of these methods; a change in energy (ΔE) from the previous to the new transition is computed; a random number then is generated with a uniform distribution between 0 and 1. Finally, if the computed probability is greater than the random number, the new configuration is accepted.

After the Monte Carlo process has converged on a given concentration, the free energy can be computed from the resulting concentration, $[L_{\text{simulation cell}}]$, and the reference state, $[L_{\text{ideal}}]$, using eq 2. The concentration of the "ideal" system used in the denominator of the free energy equation given in eq 2 is given by $\exp(B)$, as in eq 23. This can be derived from eq 17 by setting the ΔE value to zero, as in an ideal gas, and noting that $p_{ij} = p_{ji}$ at equilibrium.

$$\frac{N_i}{V} = \frac{N_0}{V_0}\exp(B) \quad (23)$$

Equation 15 defines the ideal state to which the simulation chemical potential will be compared, based on the statistics of an ideal gas.

In the following examples, B is multiplied by 2.303 to convert from natural log to $\log_{10}$ concentration units. If N_0/V_0 is defined as 1 M in eq 23, the resulting B′ value can then be interpreted directly as concentration; a B′ value of −1 corresponds to a reference state of 0.1 M.

RESULTS AND DISCUSSION

Noninteracting Ligands. One advantage of Grand Canonical ensemble simulations is that they can be performed at high chemical potential to study weak interactions. However, the molar concentrations of a ligand are readily attained at these high potentials, so that ligand−ligand interactions are observed and ligands condense to form a solvent phase. Water will condense and form the expected hydrogen bonding networks, for example, and benzene will form stacked chains held together by $\pi-\pi$ interactions. This feature is desirable if we are simulating solvation by water;[44] however, if our goal is to simulate ligand binding, we are interested only in the interaction of each individual ligand with the protein, independent of interactions with any neighboring ligands. For example, $\pi-\pi$ interactions with neighboring benzene molecules would be irrelevant to our purpose. A straightforward way to achieve this effect is to turn off ligand−ligand interactions in the simulation (in other words, to allow ligands to pass through each other as ghosts).

Because the excess free energy of the system can be directly computed from the concentration in the simulation box, compared to the concentration in the reference box using eq 2, the free energy, in principle, could be computed from a single simulation. Moreover, the free energy in any specific region of the full simulation box can be computed via the ligand concentration in just that region. If the ligands do not interact with each other, for example, we could perform a single simulation encompassing the entire protein surface and then compute the binding affinity in any isolated binding pocket on the protein from the resulting data. Therefore, the current formalism is equivalent to the experimental formalism for measuring IC_{50} in which the protein is assumed to be at infinite dilution and the substrate is in large excess over the protein.

The principle that the computed free energy is independent of the B' value for the reference state is demonstrated for the binding of benzene to T4 lysozyme in Table 1, which is displayed to demonstrate how the free energy is computed from the results. Thus, from left to right in the table, the B' value is set to represent the desired concentration of the ideal gas reference state, the simulation is run to reach equilibrium (10^8 steps), using the acceptance and rejection criteria of eqs 20−22, and the population of benzene in the binding pocket is computed by averaging over an additional 10^7 steps. The average concentration of benzene in the binding pocket is

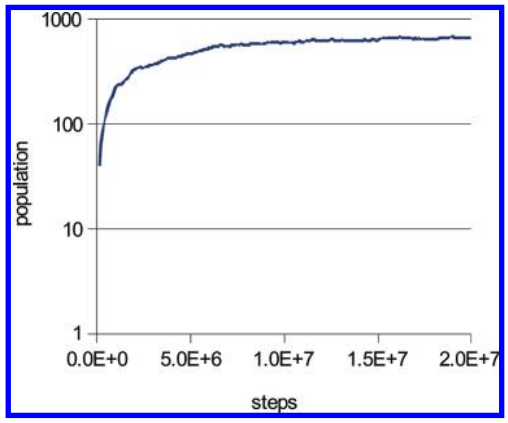

Figure 1. Population of benzene in T4 pocket versus the number of steps for simulation at $B' = -4$.

then computed, and the free energy is computed according to eq 2 from the difference between concentrations in the binding pocket and the reference state.

The data in Table 1 demonstrate that, in theory, the Grand Canonical ensemble could be set to a single arbitrary B' value for the calculation of binding free energy for any ligand, because, in each case that equilibrium was achieved ($B' \leq -4$), the simulation yielded the same computed binding free energy (between -9.8 and -9.9 kcal/mol).

The data in Table 1 also demonstrate that, in practice, the useful range of B' values is limited. The first limitation is seen at the higher free energies of $B' = -2$ and 0. In these cases, the system approaches the expected concentrations of 10^5 and 10^7 M (which correspond to a binding free energy of -9.9 kcal/mol) very slowly, because the small size of the binding pocket limits the orientations that can be accepted. Figure 1 shows that, at $B' = -4$, the system requires 2×10^7 steps to equilibrate asymptotically to ~700 ligands. At the rate of 5500 steps per second on a 2.6 GHz Intel CPU, this required ~1 h of simulation time. The slower rate of equilibrating higher populations arises from eq 20, where the probability that a ligand will be accepted into the bound population is inversely proportional to the number of ligands in the system. For example, because the number of ligands increases exponentially with the B' value, ~600 h would be required to equilibrate the system at $B' = -2$.

Simulations at lower B' levels equilibrate much more quickly, within $<10^7$ steps, making them more suitable for routine use; however, there is also a limit to how low a B' value can be. This limitation arises from the need to achieve a statistically significant value for the ligand population in the binding pocket. The simulation in Table 1 at $B' = -8$ is approaching that limit. The observed average population of 0.05 means that the total population observed over 10^7 steps was 5×10^5, with a Poisson standard deviation of ~0.1%.

Because the usable range of B' values is dependent on both the entropy and enthalpy of interaction (i.e., the size of the binding pocket, the size and flexibility of the ligand, and the protein−ligand interaction energy), the range will vary with each ligand and binding pocket that is simulated. In practice, it is easiest to identify an optimum B' value empirically by annealing B' from high to low (although the direction of annealing is not important) until we reach a range where the computed binding free energy is constant over several B' values. Other methods may be equally valid to estimate a usable B', perhaps by identifying the minimum binding energy pose and then selecting a corresponding B' value.

This method of noninteracting ligands is a correct thermodynamic model for the purpose of ligand-based drug design and works well for small binding pockets with a single, well-defined binding mode such as for the induced pocket on the T4 lysozyme L99A. With larger binding pockets with multiple potential ligand binding poses, or if the goal is to search for binding hot spots across the entire surface of a protein, then the method suffers from a practical drawback. For any system with multiple, noninteracting binding sites, the relative populations of bound ligand, $(EI)_i$, is determined simply by the relative equilibrium constants, Ki_i, at each site (see eq 24).

$$(EI)_1 \times Ki_1 = (EI)_2 \times Ki_2 = ...(EI)_n \times Ki_n \qquad (24)$$

The inhibition constants in eq 24 vary exponentially with binding free energy, so that the observed binding populations for different sites commonly vary by orders of magnitude at a given B' value. The practical implication of this result is that it is frequently not possible to identify a single B' value such that the simulation reaches equilibrium sufficiently rapidly and, at the same time, statistically significant population counts are achieved for all interesting ligand poses. This effect is amplified by the fact that solvation has a tendency to dampen the observed range of binding free energies so that the range of values observed in gas-phase simulations is greater than what is observed experimentally.

Interacting Ligands. While the noninteracting formalism is the correct model for simulating ligand−protein interactions, the exponential relationship previously described between population and free energy is a significant practical barrier to using this formalism. Methods to overcome this barrier must be directed at modulating the relationship in eq 24, so that the bound population is no longer directly proportional to the dissociation constant. One successful method to compress the dynamic range of populations uses grid-based energy offsets, for example, and is thermodynamically rigorous in terms of turning off ligand−ligand interactions.[45] Below, we describe another approach that does allow ligand−ligand repulsion (but not attraction) in a way that both compresses the dynamic population range as desired and, at the same time, returns protein−ligand binding affinities, independent of the ligand−ligand interaction energy.

When ligands under study do not interact with each other, the number of ligands observed in any given binding pose will increase without bound as the concentration of ligand in the reference state (B') is increased. If the ligands are allowed to exclude each other's volume, however, the population of the system is limited; and, in the T4 L99A example, the maximum population of benzene ligands observed in the binding site is limited to 1. We can use this simple result to limit the population in regions of high affinity so that more-extensive sampling may be accomplished in regions of low affinity, and we can compute the correct interaction free energies by limiting this calculation to regions where the average population for the interaction of interest is <1.

Because the optimum B' value for computing the binding free energy will vary depending on the binding pocket and

Table 2. Free Energies Computed for Benzene in the T4 Lysozyme L99A Binding Pocket with Ligand−Ligand Interactions Allowed

log(concentration in reference box), B'	average ligand population in binding pocket[a]	computed free energy (kcal/mol)
0	1.0	−0.45
−2	1.0	−3.2
−4	1.0	−5.9
−6	0.97	−8.4
−7	0.92	−9.0
−7.1	1.23	−9.8
−7.5	0.35	−9.7
−8	0.17	−9.9
−9	0.010	−9.5
−10	0.00016	−9.9
−11	0.00014	−9.8

[a] Ligand populations were determined by averaging over 5×10^6 steps without any prior equilibration. All ligands were cleared from the system before beginning the simulations at each new B' value.

ligand pose, the B' value is scanned to identify a range in which the average population is <1 but also high enough to be statistically significant. The binding free energy is then computed from the bound population at an appropriate B' level. The direction of annealing is not important; however, traditionally, we anneal from high B' to low B'.

Unlike the noninteracting ligand method, there is the potential for the interacting ligand method to introduce systematic errors. Thus, after a high-affinity pose is accepted into the simulation, the probability that subsequent removal attempts will be rejected is very low. Furthermore, because this pose will exclude any physically overlapping poses, there is a real possibility that alternative poses of equal or lower affinity will be excluded. As a result, the system does not reach a true equilibrium. This source of error can be minimized by clearing out all ligands from the system before each new B' value is simulated.

Results obtained by this process are shown Table 2. Although ligand−ligand interactions are nonphysical, there is a transition point at which the binding free energy the population becomes sparse, and the simulation is equivalent to the noninteracting model. At B' values below this transition point, both methods yield the same binding free energy. In this case, the transition from fully populated to sparse population occurs at $B' \approx -7.1$ (as expected from the results in Table 1) and the computed free-energy plateaus at −9.8 kcal/mol (which also is consistent with the results in Table 1). The statistical nature of the sampling results in an standard deviation of 0.15 kcal/mol for this value. The fact that the variability is greater than that for the results in Table 1 is a reflection of the need to clear all ligands from the system for each annealing step, as described above.

The method demonstrated in Table 2 will yield the correct gas-phase binding free energies within the variables of the protein state and the force field used. We compute a binding enthalpy of −16.4 kcal/mol and a free energy of −9.8 kcal/mol for the binding of benzene using the AMBER force field. In comparison, Hermans and Wang have studied the free energy of benzene in the T4 pocket with FEP methods.[37] They reported a gas-phase binding enthalpy of −16.4 kcal/mol, and a gas-phase free energy between −6 kcal/mol and −9 kcal/mol, depending on the protein structure used. To the extent that these two calculations are directly comparable, therefore, we conclude that the Grand Canonical ensemble and FEP methods will yield comparable results. The choice of protein structure is clearly a major variable in these calculations, but it is outside the scope of this paper.

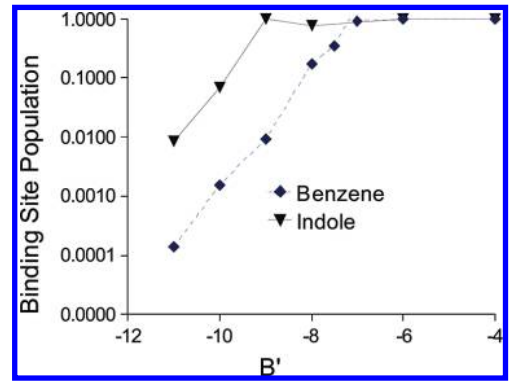

Figure 2. Free energy versus the B' value for the simulated binding of benzene and indole to T4 L99A with ligand−ligand repulsions allowed.

Allowing repulsive interactions between ligands requires sampling at many energy levels to identify the point of transition to sparse populations, but the overall result requires fewer steps than the model without ligand−ligand interactions. This is because the maximum number of ligands in the binding pockets is limited, and the rate at which equilibration is reached is dependent on the number of ligands in the system. For example, 10^8 steps were required to reach equilibrium at $B' = -4$ when ligands did not interact (Figure 1), whereas equilibration was reached in a small number of steps if repulsive interactions were allowed.

The value of allowing ligand−ligand repulsions is illustrated in Figure 2, which compares results for annealing benzene and indole. Indole binds more tightly, and the indole population exceeds benzene population by a factor of 100 at low B' values. If no ligand interactions were allowed, this result would hold at all B' values. If repulsions are allowed, the population curves reach a transition point to a maximum population; in this case, the transition points for the two ligands differ by 2 B' units, hence the obvious utility of an annealing approach. This result is more important for larger binding sites in which a single ligand may adopt multiple binding modes with widely varying binding affinities. In this case, multiple curves similar to those in Figure 2 would be observed for a single ligand, with each curve representing a different binding mode. In this case, it would be prohibitively time-consuming to sample at sufficiently high populations to quantify the low-energy modes if ligands were not allowed to interact. On the other hand, if ligand repulsions are allowed, the high-affinity modes will saturate at high B' values, and the low-affinity modes can be readily quantified in this region. For population curves similar to those in Figure 2, for example, the lower-affinity mode could be sampled efficiently at $B' = -8$.

Sampling Efficiency. If the volume available to a given binding mode, measured in the six dimensions of translation and rotation, is small, relative to the overall volume that is being sampled, the sampling procedure will not be able to find it. This principle is defined quantitatively as the probability of inserting a ligand into the required volume in eq 20 within a predefined number of steps. If the interaction energy is favorable compared to the B value, then the probability for insertion of the first ligand into the binding

site is given by eq 24, where N is the total number of attempts, V_0 the overall volume sampled, and V the volume of the target pose:

$$\rho = 1 - \left(1 - \frac{V}{V_0}\right)^N \tag{25}$$

The number of steps N required to provide a given chance ρ of accepting the fragment in the binding site, in turn, defines the sampling efficiency and is computed according to eq 25:

$$N = \frac{\log(1 - \rho)}{\log[1 - (V/V_0)]} \tag{26}$$

The volumes from eq 25 can be broken down into their translational and rotational components (V_{trans} and V_{rot}, respectively). The translational and rotational volumes of a single pose are not straightforward to define, because the probability of sampling a single pose *exactly* in both translational and rotational space is infinitesimally small. At the quantum limit, however, there is an uncertainty associated with this volume; therefore, we can relate the sampling efficiency to sampling within the resolution of this quantum uncertainty. For example, the computed van der Waals volume for benzene is 71 Å^3, and its thermal de Broglie wavelength is 0.1 Å. For a spherical molecule, which, in this case, would have a radius of 2.57 Å, this de Broglie wavelength would represent an uncertainty in volume of 8.6 Å^3. The translational component of the sampling efficiency is determined by this uncertainty in volume, relative to the entire sampled volume. In this case, the overall translational volume sampled was 782.5 Å^3, so that the translational component of the sampling efficiency is 1 in ~90. The rotational component is approximated as the volume of a cone, with a base equal to the de Broglie wavelength of 0.1 Å, relative to the spherical volume of the ligand. This is the larger component of the sampling efficiency, and, in this case, the volume ratio is 1 in ~11 000. This estimation agrees well with the measured rotational efficiency of 1 in 30 000 that has been measured previously for the binding of *n*-butyl benzene in this system.[12]

Therefore, the overall sampling efficiency for identifying any single binding pose of benzene on T4 lysozyme with 95% probability is computed to be 1 in 3×10^6, which is consistent with attaining equilibrium in 10^8 steps in the simulation of Figure 1. In our previous description of the Grand Canonical method,[12] the reference volume included the entire protein and was much larger, 193,116 Å^3. The sampling efficiency in this case was only 1 in 7×10^8. The difference demonstrates the much greater computational speed, ~1 CPU-hr, that can be achieved when the Grand Canonical method is applied solely to a single binding site. When the method is applied to map binding across an entire protein surface, the speed of the calculation can be increased by accepting a lower sampling resolution.

Solvation Correction. The thermodynamic cycle used to compute the experimentally measurable binding in solution is shown in Figure 3. The Grand Canonical simulation directly computes the binding free energy in line 1, performing the simulation in the gas phase. The solvation energies of lines 2, 3, and 4 can be computed by several methods,[4]

1.	AB_{gas}	$\rightarrow$	$A_{gas} + B_{gas}$
2.	AB_{sol}	$\rightarrow$	AB_{gas}
3.	A_{gas}	$\rightarrow$	A_{solv}
4.	B_{gas}	$\rightarrow$	B_{solv}
	AB_{solv}	$\rightarrow$	$A_{solv} + B_{solv}$

Figure 3. Complete thermodynamic cycle.

and in this case we use the generalized Born/solvent accessible surface model (GB/SA) implemented in Macromodel.[38,46]

Table 3 summarizes the computed values for several rigid ligands of T4 L99A, binding to a rigid protein model. Although the computed binding affinities do not take into account the free energy required to induce the protein binding pocket, the mean deviation between the computed and observed binding free energies is only −0.5 kcal/mol. Threfore, this result is an additional indication that the Grand Canonical approach will provide valid gas-phase binding affinities.

The accuracy of the solvation calculations, which sum three terms of different signs, can be a significant source of error in free-energy computations. For example, the standard deviation between the computed and observed free energies is 1.5 kcal/mol, and this large error is consistent with errors in the GB/SA solvation model. Table 4 compares the experimental and computed solvation free energies for five of the compounds for which experimental data are available. The average signed error is 0.9 kcal/mol, with a standard deviation of 1.0 kcal/mol. This error alone is large enough to account for the errors in Table 3, because solvation computations appear at three points in the cycle. If the observed solvation free energies are used instead of the GB/SA values for the compounds listed in Table 4, the average error drops from 1.46 kcal/mol to 0.69 kcal/mol. The importance of accurate solvation corrections is well-appreciated, and a thorough evaluation of alternative models is outside the scope of this paper. Therefore, we present the current results, simply as an indication that the gas-phase component of the free-energy calculations that are presented herein is inherently capable of yielding accurate free energies, with the uncertainty in the solvation correction potentially representing the largest source of error. Experimental studies of water in the T4 lysozyme L99A site show that, at 1 atm, an average of only 1.5 weakly bound water molecules may be present; therefore, a solvation correction for the protein may not be needed.[5] Resat has described an approach that used explicit water in this context as part of the potential of mean force calculations.[19] It may be possible to use that approach to improve the solvation treatment for more polar and open protein active sites.

CONCLUSIONS

The Grand Canonical simulation has several characteristics that are beneficial to free-energy calculations. One is that the number of parameters that must be set for the simulation is small and can be determined objectively, making the outcome more deterministic, with respect to choice of input conditions, compared to perturbation methods. The main variables are the number of steps of

Table 3. Solvation-Corrected Free Energies from Annealing of Free Energy in Simulations without Ligand−Ligand Interactions

	Free Energy (kcal/mol)				ΔG for $AB_{solv} \rightarrow A_{solv} + B_{solv}$ (kcal/mol)	
compound	$A_{gas} + B_{gas} \rightarrow AB_{gas}$	$AB_{gas} \rightarrow AB_{solv}$	$A_{solv} \rightarrow A_{gas}$	$B_{solv} \rightarrow B_{gas}$	computed	experimental
benzene	−9.8	−2302.6	2306.5	1.7	−4.2	−5.2
o-xylene	−9.9	−2304.0	2306.5	−0.5	−7.9	−4.6
m-xylene	−9.3	−2302.1	2306.5	−0.9	−5.8	−4.8
p-xylene	−8.5	−2303.1	2306.5	−0.9	−6.0	−4.7
benzofuran	−11.2	−2303.9	2306.5	2.2	−6.4	−5.5
indene	−10.5	−2301.3	2306.5	1.1	−4.2	−5.1
indole	−10.9	−2303.8	2306.5	3.0	−5.2	−4.9
thianaphthene	−10.7	−2301.5	2306.5	1.4	−4.3	−5.7
toluene	−10.9	−2302.1	2306.5	0.3	−6.5	−5.5

Table 4. Experimental and Computed Ligand Solvation Free Energies

	Free Energy for the Reaction $B_{gas} \rightarrow B_{solv}$ (kcal/mol)		error
compound	observed[a]	GB/SA	(kcal/mol)
benzene	−0.90	−1.70	0.80
o-xylene	−0.91	0.54	−1.45
m-xylene	−0.82	0.89	−1.71
p-xylene	−0.82	0.88	−1.70
toluene	−0.77	−0.32	−0.45

[a] From ref 47.

equilibration (which is dynamically based on equilibration of population and/or energy) and the number of snapshots taken for sampling the resulting poses (which can be set to give the desired accuracy in the computed binding population). While the computational model that does not allow ligand−ligand interactions is truest to the thermodynamics, it requires longer equilibration times and has limited dynamic range. We demonstrate that the model that allows ligand−ligand repulsions provides the same binding free energy with fewer equilibration steps. Therefore, the most practical way to compute free energies is to scan the energy values in a simulated annealing of chemical potential. System equilibration is efficient and the free energy can be easily derived from the population at the site of interest. The third variable, the free-energy levels at which to perform the annealing, is determined by the dynamic range of computed binding affinities.

The second benefit is that the simulation is free from assumptions about the starting pose or nature of the binding site. Free-energy perturbation methods compute the binding free energy of a preselected pose. The Grand Canonical simulation samples and computes the free energy of ligands without any assumptions other than proximity to the protein. This makes the method well-adapted to locating novel binding sites and poses. In addition, the accuracy of the free-energy computation is increased, because the sampling includes the entire binding pocket, or even the entire protein, and a large number of configurations, compared to perturbation calculations on a limited number of poses. Furthermore, perturbation methods require sophisticated constraints to be set up to maintain the desired pose during the perturbation, whereas the Monte Carlo Grand Canonical method does not require such constraints, which simplifies running the simulation for large numbers of ligands and eliminates the mathematics required to ensure that the constraints do not affect the resulting binding free energies.

A final benefit is that the binding free energies are a direct outcome of the simulation, and little processing is required to determine them. The chemical potential is set for the simulation, then the Monte Carlo process generates ensembles of poses that are consistent for that energy level, and the excess free energy over the ideal gas can be easily determined. No other integration methods are necessary to extract the free energy of the ligands.

Perhaps the overriding value of the Grand Canonical method is its speed. The rigid protein and fragment approximation allows fast estimation of binding but is also a potential weak point. The assumed rigidity does not represent the motion of the protein to accommodate ligands, which is generally the case in perturbation methods. In this study, the protein form with the largest binding pocket was used to allow all of the ligands to fit into the site. As long as the energy required to rearrange the protein for different ligands is small, or close to the quantity kT, the quality of the activity correlations obtained in the present case indicates that the approximation of rigid protein can be acceptable. However, there are certainly cases where this approximation will cause a loss of accuracy. In those situations, the speed of the Grand Canonical method is sufficient to permit comparison of several protein forms, to better understand the impact of protein flexibility.

The treatment of solvation energies is an important approximation to this method. The principle that the simple gas-phase electrostatic energy is incorrect in an absolute measure, but still related to binding energy, is suggested by the success of the linear interaction energy methods.[48−50] These methods derive a coefficient to the electrostatic potential that allows absolute prediction of the binding free energy, implying that the rank ordering of compound potencies is determined by the underlying gas-phase calculations (plus the solvation energy of the free ligand), which is consistent with results observed for the small, hydrophobic binding site modeled in this case.

Future work centers on methods of addressing more-flexible ligands by combining the results of single-fragment calculations and assembling more-complex molecules, so that the method can be applied to drug design.

ACKNOWLEDGMENT

The authors acknowledge Anthony Klon, Zenon Konteatis, Keith Milligan, George Talbot, and Jinming Zou for discussions and support of this work. We also thank the reviewers for their insightful comments.

REFERENCES AND NOTES

(1) Bembenek, S. D.; Tounge, B. A.; Reynolds, C. H. Ligand Efficiency and Fragment Based Drug Discovery. *Drug Discovery Today*Published Online December 30, 2008 (DOI: 10.1016/j.drudis.2008.11.007).

(2) Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.

(3) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical−Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72*, 1047–1069.

(4) Gilson, M. K.; Zhou, H.-X. Calculation of Protein−Ligand Binding Affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.

(5) Zwanzig, R. W. High-temperature Equation of State by a Perturbation Method. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

(6) Pranata, J.; Jorgensen, W. L. Monte-Carlo Simulations Yield Absolute Free Energies of Binding for Guanine−Cytosine and Adenine−Uracil Base Pairs in Chloroform. *Tetrahedron* **1991**, *47*, 2491–2501.

(7) Hamelberg, D.; McCammon, J. A. Standard Free Energy of Releasing a Localized Water Molecule from the Binding Pockets of Proteins: Double-Decoupling Method. *J. Am. Chem. Soc.* **2004**, *126*, 7683–7689.

(8) 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.

(9) Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA. *J. Am. Chem. Soc.* **2001**, *123*, 5221–5230.

(10) Weis, A.; Katebzadeh, K.; Soderhjelm, P.; Nilsson, I.; Ryde, U. Ligand Affinities Predicted with the MM/PBSA Method: Dependence on the Simulation Method and the Force Field. *J. Med. Chem.* **2006**, *49*, 6596–6606.

(11) Chang, C.-E.; Gilson, M. K. Free Energy, Entropy, and Induced Fit in Host-Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm. *J. Am. Chem. Soc.* **2004**, *126*, 13156–13164.

(12) Clark, M.; Guarnieri, F.; Shkurko, I.; Wiseman, J. Grand Canonical Monte Carlo Simulation of Ligand−Protein Binding. *J. Chem. Inf. Model.* **2006**, *46*, 231–242.

(13) Guarnieri, F. Computational Protein Probing to Identify Binding Sites, U.S. Patent 6,735,530, May 11, 2004.

(14) Mezei, M. Grand-Canonical Ensemble Monte Carlo Simulation of Dense Fluids: Lennard-Jones, Soft Spheres and Water. *Mol. Phys.* **1987**, *61*, 565−582 (and Erratum **1989**, *67*, 1207−1208).

(15) Resat, H.; Mezei, M.; McCammon, J. A. Use of the Grand Canonical Ensemble in Potential of Mean Force Calculations. *J. Phys. Chem.* **1996**, *100*, 1426–1432.

(16) Resat, H.; Mezei, M. Grand Canonical Ensemble Monte Carlo Simulation of the dCpG/Proflavine Crystal Hydrate. *Biophys. J.* **1996**, *71*, 1179–1190.

(17) Mezei, M. Comment on Molecular Dynamic Simulations in the Grand Canonical Ensemble: Formulation of a Bias Potential for Umbrella Sampling Comparison of the Sampling Efficiency in the Grand Canonical Ensemble Using Molecular Dynamics and Cavity-Biased Monte Carlo. *J. Chem. Phys.* **2000**, *112*, 1059–1060.

(18) Guarnieri, F.; Mezei, M. Simulated Annealing of Chemical Potential: A General Procedure for Locating Bound Waters. Application to the Study of the Differential Hydration Propensities of the Major and Minor Grooves of DNA. *J. Am. Chem. Soc.* **1996**, *118*, 8493–8494.

(19) Resat, H.; Marrone, T. J.; McCammon, J. A. Enzyme-Inhibitor Association Thermodynamics Explicit and Continuum Solvent Studies. *Biophys. J.* **1997**, *72*, 522–532.

(20) Kamberaj, H.; Helms, V. Monte-Carlo Simulation of Biomolecular Systems with BIOMCSCIM. *Comput. Phys. Commun.* **2001**, *3*, 375–402.

(21) Woo, H.-J.; Dinner, A. R.; Roux, B. Grand Canonical Monte Carlo Simulations of Water in Protein Environments. *J. Chem. Phys.* **2004**, *121*, 6392–6400.

(22) Marrone, T. J.; Resat, H.; Hodge, C. N.; Change, C. H.; McCammon, J. A. Solvation Studies of DMP323 and A76928 Bound to HIV Protease: Analysis of Water Sites Using Grand Canonical Monte Carlo Simulations. *Protein Sci.* **1998**, *7*, 573–579.

(23) Nagumo, R.; Takaba, H; Nakao, S.-I. Prediction of Ideal Permeability of Hydrocarbons Through an MFI-Type Zeolite Membrane by a Combined Method Using Molecular Simulation Techniques and Permeation Theory. *J. Phys. Chem. B* **2003**, *107*, 14422–14428.

(24) He, Y.; Seaton, N. A. Experimental and Computer Simulation Studies of the Adsorption of Ethane, Carbon Dioxide, and Their Binary Mixtures in MCM-41. *Langmuir* **2003**, *19*, 10132–10138.

(25) Meredith, J. C.; Johnston, K. P. Density Dependence of Homopolymer Adsorption and Colloidal Interaction Forces in a Supercritical Solvent: Monte Carlo Simulation. *Langmuir* **1999**, *15*, 8037–8044.

(26) Jayaram, B.; Beveridge, D. L. Grand Canonical Monte Carlo Simulations on Aqueous Solutions of NaCl and NaDNA: Excess Chemical Potentials and Sources of Nonideality in Electrolyte and Polyelectrolyte Solutions. *J. Phys. Chem.* **1991**, *95*, 2506–2516.

(27) Mills, P.; Anderson, C. F.; Record, M. T., Jr. Grand Canonical Monte Carlo Calculations of Thermodynamic Coefficients for a Primitive Model of DNA−Salt Solutions. *J. Phys. Chem.* **1986**, *90*, 6541–6548.

(28) Miller, E. S.; Kutter, E.; Mosig, G.; Arisaka, F.; Kunisawa, T.; Rüger, W. Bacteriophage T4 Genome. *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 86–156.

(29) Morton, A.; Baase, W. A.; Matthews, B. W. Energetic Origins of Specificity of Ligand Binding in an Interior Nonpolar Cavity of T4 Lysozyme. *Biochemistry* **1995**, *34*, 8564–8575.

(30) Morton, A.; Matthews, B. A. Specificity of Ligand Binding in a Buried Nonpolar Cavity of T4 Lysozyme: Linkage of Dynamics and Structural Plasticity. *Biochemistry* **1995**, *34*, 8576–8588.

(31) Collins, M. D.; Hummer, G.; Quillin, M. L.; Matthews, B. W.; Gruner, S. M. Cooperative Water Filling of a Nonpolar Protein Cavity Observed by High-Pressure Crystallography and Simulation. *Proc. Natl. Acad. Sci., USA* **2005**, *102*, 16668–16671.

(32) Liu, L.; Quillin, M. L.; Matthews, B. W. Use of Experimental Crystallographic Phases to Examine the Hydration of Polar and Nonpolar Cavities in T4 Lysozyme. *Proc. Natl. Acad. Sci., USA* **2008**, *105*, 14406–14411.

(33) Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. Prediction of Protein−Ligand Interactions. Docking and Scoring: Successes and Gaps. *J. Med. Chem.* **2006**, *49*, 5851–5855.

(34) Ferrari, A. M.; Wei, B. Q.; Constantino, L.; Shoichet, B. K. Soft Docking and Multiple Receptor Conformations in Virtual Screening. *J. Med. Chem.* **2004**, *47*, 5076–5084.

(35) Graves, A. P.; Brenk, R.; Shoichet, B. K. Decoys for Docking. *J. Med. Chem.* **2005**, *48*, 3714–3728.

(36) Deng, Y.; Roux, B. Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255–1273.

(37) Hermans, J.; Wang, L. Inclusion of Loss of Translational and Rotational Freedom in Theoretical Estimates of Free Energies of Binding. Application to a Complex of Benzene and Mutant T4 Lysozyme. *J. Am. Chem. Soc.* **1997**, *119*, 2707–2714.

(38) *Macromodel 8.6*; Schrodinger: New York, 2004.

(39) Mohamadi, F.; Richard, 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.

(40) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98 (Revision A.9)*; Gaussian, Inc.: Pittsburgh, PA, 1998.

(41) Cornell, W. D.; Cieplak, P. I.; Bayly, C. I.; Gould, I. R.; Merz, K. M. Jr; Ferguson, D.; 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.

(42) Matsumoto, M.; Nishimura, T. Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator. *ACM Trans. Modeling Comput. Sim.* **1998**, *8*, 3–30.

(43) Gibbs, J. W. *Elementary Principles in Statistical Mechanics*; Charles Scribners's Sons: New York, 1902.

(44) Guarnieri, F.; Mezei, M. Simulated Annealing of Chemical Potential: A General Procedure for Locating Bound Waters. Application to the Study of the Differential Hydration Propensities of the Major and Minor Grooves of DNA. *J. Am. Chem. Soc.* **1996**, *118*, 8493–8494.

(45) Brunner, S.; Karney, C. F. F. Method and Computer Program Product for Drug Discovery Using Weighted Grand Canonical Metropolis Monte Carlo Sampling, U.S. Pat. Appl. 2004-0267509, December 30, 2004.

(46) Qui, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.

(47) Viswanadhan, V. N.; Ghose, A. K.; Singh, U. C.; Wendoloski, J. J. Prediction of Solvation Free Energies of Small Organic Molecules: Additive−Constituitive Models Based on Molecular Fingerprints and Atomic Constants. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 405–412.

(48) Åqvist, J.; Marelius, J. The Linear Interaction Energy Method for Predicting Ligand Binding Free Energies. *Comb. Chem. High Throughput Screen.* **2001**, *4*, 613–626.

(49) Rizzo, R. C.; Udier-Blagovic, M.; Wang, D-P.; Watkins, E. K.; Kroeger Smith, M. B.; Smith, R. H., Jr.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of Binding Affinities for HEPT and Nevirapine Analogues with HIV-1 Reverse Transcriptase via Monte Carlo Simulations. *J. Med. Chem.* **2002**, *45*, 2970–2987.

(50) Pierce, A. C.; Jorgensen, W. L. Estimation of Binding Affinities for Selective Thrombin Inhibitors via Monte Carlo Simulations. *J. Med. Chem.* **2001**, *44*, 1043–1050.

CI8004397