JCTC Journal of Chemical Theory and Computation

Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant

Yuqing Deng[†] and Benoît Roux*[,†,‡]

Biosciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Building 221, Argonne, Illinois 60439, and Institute of Molecular Pediatric Sciences, Gordon Center for Integrative Sciences, University of Chicago, 929 East 57th Street, Room W323B, Chicago, Illinois 60637

Received February 1, 2006

Abstract: Calculations of the binding free energy of various nonpolar aromatic ligands with the L99A mutant of T4 lysozyme using molecular dynamics (MD) simulation are presented. To ensure better convergence, biasing potentials are used to restrain the ligand orientation and center-of-mass movement relative to the binding site when the ligand is decoupled from its environment in the binding pocket. The bias introduced by the restraint potentials is removed once the ligand fully interacts with the rest of the system and the calculated binding free energy is independent of the applied restraints. To decrease the computational cost, the simulations are generated with a reduced system in which protein and water atoms within a 15 Å-radius sphere around the ligand are included explicitly, while the rest of the system is treated with the generalized solvent boundary potential (GSBP). For all the ligands, the precision of the calculated free energy is less than 0.5 kcal/mol. For small nonpolar ligands such as benzene, toluene, and ethylbenzene, the calculated binding free energies are within 1.1 kcal/mol of the experimental values. For larger ligands, the computed binding free energies are slightly more favorable than the experimental values. The nonbinding polar molecule, phenol, has a calculated binding free energy of −0.88 kcal/mol. The simulation protocol presented here provides a way to calculate the binding free energy of small molecules to receptors at moderate computational cost.

Introduction

"*Corpora non agunt nisi fixata*"[1]—A substance is not (biologically) active unless it is fixed (bound) by a receptor. Nearly a century ago, this Latin maxim from Paul Ehrlich (1913) epitomized the central role of molecular recognition in chemistry and biology. Such a view is currently more pertinent then ever, as the rational design of small ligands binding to macromolecules with high affinity and specificity has become one of the cornerstones in the development of new drugs.[2−6] Theoretical and computational tools, if they were systematically able to predict binding free energies with quantitative accuracy, would have a big impact on drug design. Developing and improving these methods is, thus, an important goal of computational chemistry and biophysics.

Calculations of binding free energies employ a wide variety of methods, ranging from simple empirical scoring functions to thermodynamic free energy simulations with explicit solvent and full atomic details.[7] While it may be expected that fast and simple scoring schemes will remain the method of choice to sieve through a large database comprising millions of compounds, more accurate treatments are clearly necessary in order to accurately rank-order similar and potentially active ligands. With the growth in computational resources, advances in simulations methodologies, and improvement of all-atom force fields, approaches to calculate standard (absolute) binding free energies based on rigorous statistical mechanical treatments become increasingly practical.[8−11] It is of interest to take advantage of these recent advances in order to make rigorous calculations of binding free energies tracktable computationally to the point that they could be utilized routinely.

* Corresponding author e-mail: roux@uchicago.edu.
[†] Argonne National Laboratory.
[‡] University of Chicago.

10.1021/ct060037v CCC: $33.50
Published on Web 07/20/2006

Fundamentally, the binding free energy can be expressed from the reversible thermodynamic work to separate the ligand from its receptor in solution.[12−20] Alchemical free energy perturbation (FEP) simulations,[21,22] in which the ligand is annihilated (decoupled from its surrounding), offer a practical route to carry out such computations.[23] Alternatively, it is possible to integrate the Boltzmann factor of the potential of mean force (PMF) as a function of the ligand−receptor distance.[12,18,20] While the latter may be advantageous in the case of highly charged ligands bound to the surface of a receptor (e.g., phosphotyrosine peptide pYEEI binding to an SH2 domain),[20] the former is particularly useful in the case of a ligand bound deeply in the interior of a macromolecule. The main advantage of alchemical FEP is that a realistic physical path for the association/dissociation of the ligand from the binding pocket is not required. In other words, the calculations can proceed by annihilating the ligand via unphysical intermediate states as along as the final endpoint states are correctly constructed. A straightforward double-annihilation scheme[23] is, however, impractical because the decoupled (noninteracting) ligand can drift away from the binding site and wander anywhere in the simulated system, leading to an ill-defined reference endpoint state (see the discussion in ref 19). These problems are solved in the double-decoupling method (DDM), in which biasing restraint potentials are applied to the ligand translational degrees of freedom relative to the receptor. Handling these restraints properly enables a rigorous treatment of the endpoint decoupled state.[13−17,19,24]

In the context of this theoretical framework, it is possible to carry detailed FEP molecular dynamics (MD) simulations to calculate the standard (absolute) binding free energy of a ligand to a receptor. Conducting all-atom FEP/MD free energy simulations for an entire protein−solvent system with periodic boundary conditions is, however, often prohibitive, and it is important to seek ways to reduce the computational cost of the calculations. In the present calculations, all the simulations are performed on reduced atomic systems using the spherical solvent boundary potential (SSBP)[25] for bulk solvent and the generalized solvent boundary potential (GSBP) for the binding site.[26] In these treatments, all the atomic details near the ligand are simulated explicitly, while the influence of the rest of the system is incorporated implicity via a mean-field continuum electrostatic approximation. Free energy calculations using both SSBP[27] and GSBP[28] have been shown to yield accurate results at a much reduced computational cost. With these approximations, the number of atoms that are simulated explicitly typically ranges from 1500 to about 3000, which reduces significantly the computational effort for FEP simulations.

Additional considerations may also help to construct a robust and efficient method for calculating binding free energies. For example, intermolecular forces are typically dominated by short-range harsh repulsive interactions and more slowly varying van der Waals attraction and electrostatic interactions; it is consequentially advantageous to stage the decoupling alchemical FEP simulations accordingly with a proper choice of intermediate states by exploiting the drastically different characters in spatial range and magnitude of these various contributions. A staged decoupling FEP scheme based on the Weeks−Chandler−Andersen (WCA)[29] separation of the Lennard-Jones (LJ) 12−6 potential has been shown to provide a clear separation of the competing repulsive and dispersive forces in the solvation process and to greatly enhance the efficiency of solvation free energy computations.[27] The WCA separation leads to the interesting observation that, while the total contribution to the solvation free energy from nonpolar interactions is relatively small, it typically results from the cancellation of much larger repulsive and dispersive contributions. It may be expected that a similar separation of the binding free energy will help clarify important aspects of the binding process which might otherwise remain hidden.

Our goal with this paper is to present a detailed examination of theoretical and computational issues related to the calculation of the absolute binding free energy using FEP simulations. As a test system, we chose to study the binding of aromatic ligands to the engineered cavity inside the L99A T4 lysozyme mutant, for which there are many available crystallographic structures.[30,31] The availability of accurate structure of the complexes allows us to avoid the challenging step of determining the ligand position and configuration (*docking*) in its binding site. From this point of view, the present effort is primarily concerned with the evaluation of the absolute binding free energy associated with a given ligand position and configuration (*scoring*) using a rigorous but computationally inexpensive FEP simulation protocol. The paper is organized as follows. The theoretical formulation of the binding free energy calculation forms the next section. What follows is all computational details. The results and discussion are then presented, and the paper is concluded with a brief summary of the most important points.

II. Theory

A. Equilibrium Binding Constant. For the sake of simplicity, we consider a single receptor molecule (R) in thermodynamic equilibrium with a dilute solution containing flexible ligand molecules (L). The equilibrium constant, K_b, of the binding reaction $L + R \rightleftharpoons LR$ is defined as $K_b = [LR]/([L][R])$, where [L], [R], and [LR], are the equilibrium concentrations of the ligand, receptor, and complex, respectively. It is assumed that the *binding site* of the receptor and the *bulk* can be distinguished clearly as two distinct spatial regions. This implies that the binding site of the receptor molecule can either be empty or occupied by a ligand. A rigorous expression for the equilibrium binding constant can be derived directly from population configurational ensemble averages.[14,18,20] This derivation, which can be traced back to Bjerrum,[32] is arguably clearer and more direct than the more traditional treatment that consists of equating the chemical potentials of the three species, L, R, and LR.[15,19] Let the probability to find the receptor with no ligand bound be $\mathscr{P}_0$, and the probability to find the receptor with one ligand bound be $\mathscr{P}_1$;[14] by normalization, $\mathscr{P}_0 + \mathscr{P}_1 = 1$. The ratio of these occupancy probabilities, $\mathscr{P}_1/\mathscr{P}_0 = K_b[L]$, is related to the thermodynamic reversible work to take one ligand molecule in the bulk solution and insert it into the binding site.[14,33,20] It follows that K_b can be expressed in terms of a ratio of

configurational integrals as

$$K_b = \frac{1}{[L]} \frac{N\int_{site} d(\mathbf{1}) \int_{bulk} d(\mathbf{2}) \cdots \int_{bulk} d(\mathbf{N}) \int d\mathbf{X}\, e^{-\beta U}}{\int_{bulk} d(\mathbf{1}) \int_{bulk} d(\mathbf{2}) \cdots \int_{bulk} d(\mathbf{N}) \int d\mathbf{X}\, e^{-\beta U}} \quad (1)$$

where U is the total potential energy of the system; **(1)**, **(2)**, ..., (N) and **X** are the coordinates of the N ligand molecules and the remaining atoms (solvent and receptor), respectively; $\beta \equiv 1/k_B T$, where k_B is the Boltzmann constant, and T is the temperature. The factor of N accounts for the fact that any of the N identical ligands could be chosen to occupy the binding site. Because the bulk region is homogeneous and invariant by translation, this expression can be rewritten as

$$K_b = \frac{1}{[L]} \times \frac{N\int_{site} d(\mathbf{1}) \int_{bulk} d(\mathbf{2}) \cdots \int_{bulk} d(\mathbf{N}) \int d\mathbf{X}\, e^{-\beta U}}{V_{bulk} \int_{bulk} d(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int_{bulk} d(\mathbf{2}) \cdots \int_{bulk} d(\mathbf{N}) \int d\mathbf{X}\, e^{-\beta U}} \quad (2)$$

where $\boldsymbol{r}_1$ is the position of the center of mass (COM) of ligand 1, and $\boldsymbol{r}^*$ is some arbitrary location in the bulk region. In the limit of an excess ligand concentration, the binding constant is

$$K_b = \frac{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta U}}{\int_{bulk} d(\mathbf{1}) \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta U}} \quad (3)$$

where the integrals over the ($N - 1$) remaining ligand in the bulk have been omitted for the sake of clarity. Eq 3 is our starting point for designing effective binding free energy simulation strategies.[14] It should be noted that the delta function in the denominator of eq 3 implies that K_b has dimension of volume.

The general idea to compute the equilibrium binding constant from eq 3 can be expressed by inserting unity as the ratio of configurational integrals, Z_i/Z_i, $i = 1...n$, corresponding to suitably chosen intermediate states

$$K_b = \frac{\int_{site} d\mathbf{1} \int d\mathbf{X}\, e^{-\beta U}}{Z_n} \times \frac{Z_n}{Z_{n-1}} \times \cdots \frac{Z_3}{Z_2} \times \frac{Z_2}{Z_1} \times \frac{Z_1}{\int_{bulk} d\mathbf{1}\, \delta(\mathbf{r}_1 - \mathbf{r}_1^*) \int d\mathbf{X}\, e^{-\beta U}} \quad (4)$$

Here, the numerator of the leftmost ratio and the denominator of the rightmost ratio may be considered as *initial* and *final* states of the binding process: the ligand bound to the receptor and the ligand held with its COM at $\mathbf{r}_1^*$, far away from the receptor. The remaining configurational integrals, Z_i, represent intermediate states chosen for the sake of computational effectiveness and convenience. Wise choices of those intermediates are the key to the design of a practical scheme— the contribution of each of the intermediates should be amenable to efficient computer simulations. One possible strategy is to insert intermediates with different fixed distances between the ligand and the receptor, so that the binding constant is expressed as an integration over a potential of mean force (PMF)[12,18,33,34,20] along the chosen pathway. An alternative strategy is to annihilate the ligand in the binding site and re-create it in the bulk solution.[13−19] This strategy corresponds to the double-annihilation method,[23] where a single intermediate configurational integral Z_1 is introduced in which the ligand is completely decoupled from its surroundings (the protein receptor or bulk solvent). A straightforward application of the double-annihilation scheme is, however, impractical because the noninteracting ligand can drift away from the binding site and wander anywhere in the simulation box (see the discussion in ref 19). In fact, a simulation has to be exceedingly (impossibly) long to sample the decoupling of the ligand from the receptor in a true double-annihilation computation. Such difficulties are avoided very simply by introducing one additional configurational integral Z_2, an intermediate state of a decoupled ligand confined in the binding site by an extra potential, u_t, which restrains the position of the ligand relative to the receptor.[13,14,16,17,19] Similarly, additional restraining potentials can also be introduced to bias the orientation[16,19] and conformation[20] of the ligand to accelerate the convergence further. To facilitate further development, it is useful to define U_1 as the total potential energy of the fully interacting system and U_0 as the total potential energy of a fictitious system in which the ligand does not interact with the rest of the system. The binding constant can then be expressed as

$$\begin{aligned} K_b = {} & \frac{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta U_1}}{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c]}} \times \\ & \frac{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c]}}{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c+u_t]}} \times \\ & \frac{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c+u_t]}}{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c+u_t+u_r]}} \times \\ & \frac{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c+u_t+u_r]}}{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_c+u_t+u_r]}} \times \\ & \frac{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_c+u_t+u_r]}}{\int_{bulk} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_c+u_t]}} \times \\ & \frac{\int_{bulk} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_c+u_t]}}{\int_{bulk} d(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_0+u_c]}} \times \\ & \frac{\int_{bulk} d(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_0+u_c]}}{\int_{bulk} d(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_1+u_c]}} \times \\ & \frac{\int_{bulk} d(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_1+u_c]}}{\int_{bulk} d(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta U_1}} \end{aligned} \quad (5)$$

where u_c is a restraint potential acting on the internal configuration of the ligand, u_t is a restraint potential acting on the position of the COM (relative to the receptor) of the ligand, and u_r is a restraint potential acting on the rotational orientation (relative to the receptor) of the ligand. Typically,

the restraint potentials are chosen as harmonic potentials centered around some reference values. The configurational restraint potential, u_c, for example, is a harmonic potential with respect to the root-mean-square deviation (RMSD) of the ligand relative to a reference bound configuration.[20] The formulation of the binding constant developed here is quite similar to the PMF formulation in ref 20. The difference is that in the PMF formulation, the interaction between the ligand and the receptor is never turned off, and the binding is achieved by the change in the distance between the ligand and the protein. In terms of free energies, the binding constant can be written as

$$K_b = e^{+\beta\Delta G_c^{site}} \times e^{+\beta\Delta G_t^{site}} \times e^{+\beta\Delta G_r^{site}} \times e^{-\beta\Delta G_{int}^{site}} \times F_r \times F_t \times e^{+\beta\Delta G_{int}^{bulk}} \times e^{-\beta\Delta G_c^{bulk}} \quad (6)$$

where the various free energies are defined as what follows

$$e^{-\beta\Delta G_c^{site}} = \frac{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c]}}{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta U_1}} = \langle e^{-\beta u_c} \rangle_{(site;\, U_1)} \quad (7)$$

$$e^{-\beta\Delta G_t^{site}} = \frac{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c+u_t]}}{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c]}} = \langle e^{-\beta u_t} \rangle_{(site;U_1+u_c)} \quad (8)$$

$$e^{-\beta\Delta G_r^{site}} = \frac{\sigma\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c+u_t+u_r]}}{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c+u_t]}} = \langle e^{-\beta u_r} \rangle_{(site;U_1+u_{c+}u_t)} \quad (9)$$

$$e^{-\beta\Delta G_{int}^{site}} = \frac{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c+u_t+u_r]}}{\int_{site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_c+u_t+u_r]}} = \langle e^{-\beta[U_1-U_0]} \rangle_{(site;U_0+u_c+u_t+u_r)} \quad (10)$$

$$e^{-\beta\Delta G_{int}^{bulk}} = \frac{\int_{bulk} d(\mathbf{1})\, \delta(\mathbf{r}_1-\mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_1+u_c]}}{\int_{bulk} d(\mathbf{1})\, \delta(\mathbf{r}_1-\mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_0+u_c]}} = \langle e^{-\beta[U_1-U_0]} \rangle_{(bulk;U_0+u_c)} \quad (11)$$

and

$$e^{-\beta\Delta G_c^{bulk}} = \frac{\int_{bulk} d(\mathbf{1})\, \delta(\mathbf{r}_1-\mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_1+u_c]}}{\int_{bulk} d(\mathbf{1})\, \delta(\mathbf{r}_1-\mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta U_1}} = \langle e^{-\beta u_c} \rangle_{(bulk;U_1)} \quad (12)$$

where $\mathbf{r}^*$ is the fixed position of the ligand COM in the bulk region. Note that the symmetry number of the ligand, σ, has been included in eq 9 because the configurational integral for ΔG_r^{site} is restricted to a single distinct orientation (σ is the number of rotational operations in the point group of the ligand). All those ΔG's can be calculated from separate free energy simulations according to a standard methodology.[21,22,35]

In contrast to those contributions above, the translational factor F_t

$$F_t = \frac{\int d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_c+u_t]}}{\int d(\mathbf{1})\, \delta(\mathbf{r}_1-\mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_0+u_c]}} \quad (13)$$

and the rotational factor F_r

$$F_r = \frac{\int d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_c+u_t+u_r]}}{\int d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_c+u_t]}} \quad (14)$$

can be evaluated directly with numerical integration schemes. The position of the ligand with respect to the receptor is restrained with the potential $u_t(\mathbf{r}_1)$, where $\mathbf{r}_1$ is the relative vector between the COM of the ligand and the receptor. Since the ligand molecule does not interact with its surroundings in the uncoupled state with potential U_0, the translational factor F_t can be expressed as

$$F_t = \int d\mathbf{r}_1\, e^{-\beta u_t(\mathbf{r}_1)} \quad (15)$$

(all other terms cancel out). Likewise, the orientation of the ligand molecule (relative to the receptor) is restrained via the potential $u_r(\Omega_1)$ where Ω_1 is the set of three angles for the rigid body rotation

$$F_r = \frac{\int d\Omega_1\, e^{-\beta u_r(\Omega_1)}}{\int d\Omega_1} \quad (16)$$

Similar treatment of restraining potentials has been discussed previously.[13−20] The simplification of the translational and rotational factors lies in the fact that there are six translational and rotational degrees of freedom of the bound complex. As long as the restraint potentials are placed on coordinates relative to the receptor, it is always possible to separate the ligand degrees of freedom from that of the proteins, thereby reducing the factors to products of one-dimensional integrals. It should be noted that the translational factor F_t (and hence the binding constant K_b) has a dimension of volume (with a natural unit of Å^3 in atomistic simulations). This dimensionality of the binding constant gives rise to the standard free energy, defined relative to the standard concentration 1 mol/L

$$\Delta G_b^\circ \equiv -k_B T \ln[K_b C^\circ] \quad (17)$$

where $C^\circ = 1$ mol·L^{-1} (=1/1660 Å^{-3}). This the statistical mechanical basis for the definition of the thermodynamic standard state in the binding free energy.[14,16,19]

The standard binding constant (K_bC°) can be expressed as $K_b^\circ \equiv \exp[-\beta\Delta G_b^\circ]$, where the standard binding free energy is

$$\Delta G_b^\circ = \Delta\Delta G_{int} + \Delta\Delta G_t^\circ + \Delta\Delta G_r + \Delta\Delta G_c \quad (18)$$

where $\Delta\Delta G_{int} = [\Delta G_{int}^{site} - \Delta G_{int}^{bulk}]$, $\Delta\Delta G_c = [\Delta G_c^{bulk} - \Delta G_c^{site}]$, $\Delta\Delta G_r = [-\Delta G_r^{site} - k_B T \ln(F_r)]$, and $\Delta\Delta G_t^\circ = [-\Delta G_t^{site} - k_B T \ln(F_t C^\circ)]$. Each contribution has a well-

defined meaning: ΔG_{int}^{bulk} corresponds to the free energy to introduce the interaction between the ligand (restrained in its bound conformation by the potentials u_c) and the surrounding bulk solution; ΔG_{int}^{site} corresponds to the free energy to introduce the interaction of the ligand with its binding pocket (restrained in its conformation, translation, and rotation by the potentials u_c, u_t, and u_r, respectively); $\Delta\Delta G_{int}$ corresponds to the free energy difference associated with removing the ligand from the bulk and inserting it in the binding site; $\Delta\Delta G_t$ and $\Delta\Delta G_r$ correspond to the free energy cost associated with application and removal of the translational and rotational restraints on the ligand; $\Delta\Delta G_c$ corresponds to the free energy cost associated with restricting the conformational freedom of the ligand upon binding. The free energies associated with the introduction of the conformational restraint potential can be computed, for example, by using the RMSD relative to the conformation of the bound ligand[36]

$$\Delta G_c^{site} = -k_B T \ln \frac{\int d\zeta\, \rho_{site}(\zeta) e^{-\beta u_c}}{\int d\zeta\, \rho_{site}(\zeta)} \tag{19}$$

$$\Delta G_c^{bulk} = -k_B T \ln \frac{\int d\zeta\, \rho_{bulk}(\zeta) e^{-\beta u_c}}{\int d\zeta\, \rho_{bulk}(\zeta)} \tag{20}$$

where $\rho_{site}(\zeta)$ and $\rho_{bulk}(\zeta)$ are the unbiased distributions of RMSD. One may note that the contribution from the conformational restraint, $-\beta\Delta\Delta G_c$, is necessarily unfavorable because it is related to the reduction in the number of accessible conformations of the ligand in the binding site relative to the bulk, i.e., this may be formally represented as $\mathcal{N}_c^{site}/\mathcal{N}_c^{bulk}$. Equation 18 is the basis for all of the computations presented in this work. Apart from the additional conformational restraint u_c, the present framework is equivalent to the scheme of Boresch et al.,[19] though it is derived using a different route.

B. Translational and Rotational Restriction in the Bound State. With appropriate choices of the translational and orientational restraint potentials, it is possible to relate $\Delta\Delta G_t^\circ$ and $\Delta\Delta G_r$ directly to the reduction in the translational and rotational freedom of the bound ligand. The free energy associated with the translation of the ligand is

$$e^{-\beta\Delta\Delta G_t^\circ} = C^\circ \times \frac{\int_{site} d(\mathbf{1}) \int d\boldsymbol{X}\, e^{-\beta[U_1+u_c]}}{\int_{site} d(\mathbf{1}) \int d\boldsymbol{X}\, e^{-\beta[U_1+u_c+u_t]}} \times \int d\mathbf{r}_1\, e^{-\beta u_t(\mathbf{r}_1)}$$

$$= C^\circ \times \frac{\int_{site} d\mathbf{r}_1\, P^{site}(\mathbf{r}_1)}{\int_{site} d\mathbf{r}_1\, P_t^{site}(\mathbf{r}_1) e^{-\beta u_t(\mathbf{r}_1)}} \times \int d\,\mathbf{r}_1\, e^{-\beta u_t(\mathbf{r}_1)} \tag{21}$$

where P_t^{site} is the probability distribution of the ligand position in the binding site (this distribution depends implicitly on u_c). In the limit of strong restraint potential, $u_t(\mathbf{r}_1)$, the translational restraint acts essentially as a delta function

$$\frac{e^{-\beta u_t(\mathbf{r}_1)}}{\int_{site} d\,\mathbf{r}_1\, e^{-\beta u_t(\mathbf{r}_1)}} \approx \delta(\mathbf{r}_1 - \mathbf{r}_m) \tag{22}$$

where $\mathbf{r}_m$ is the energy minimum in the translational restraint. Assuming that $\mathbf{r}_m$ is set to match the most probable position of the COM of the ligand in the binding site (i.e., the maximum of P_t^{site}), the translational contribution becomes

$$e^{-\beta\Delta\Delta G_t^\circ} \approx C^\circ \times \int_{site} d\,\mathbf{r}_1 \frac{P_t^{site}(\mathbf{r}_1)}{P_t^{site}(\mathbf{r}_m)} = C^\circ \Delta V \tag{23}$$

where ΔV is the effective accessible volume of the COM of the ligand in the binding site. Similarly, the free energy associated with the orientation of the ligand is

$$e^{-\beta\Delta\Delta G_r} = \frac{\int_{site} d(\mathbf{1}) \int d\boldsymbol{X}\, e^{-\beta[U_1+u_c+u_t]}}{\int_{site} d(\mathbf{1}) \int d\boldsymbol{X}\, e^{-\beta[U_1+u_c+u_t+u_r]}} \times \frac{\int d\Omega_1\, e^{-\beta u_r(\Omega_1)}}{\int d\Omega_1} = \frac{\int d\Omega_1\, P_r^{site}(\Omega_1)}{\int d\Omega_1\, P_r^{site}(\Omega_1) e^{-\beta u_r(\Omega_1)}} \times \frac{\int d\Omega_1\, e^{-\beta u_r(\Omega_1)}}{\int d\Omega_1} \tag{24}$$

where P_r^{site} is the distribution of the orientation angles (this distribution depends implicitly on u_c and u_t). Here we drop the symmetry factor for the moment and will add it in the final binding constant formulation. In the limit of strong $u_r(\Omega)$, the orientational restraint acts essentially as a delta function

$$\frac{e^{-\beta u_r(\Omega_1)}}{\int d\Omega_1\, e^{-\beta u_r(\Omega_1)}} \approx \delta(\Omega_1 - \Omega_m) \tag{25}$$

which is sharply peaked at Ω_m, chosen to match the most probable orientation of the ligand in the binding site (i.e., the maximum of P_r^{site}). For a nonlinear ligand, it follows that

$$e^{-\beta\Delta\Delta G_r} \approx \frac{1}{\int_{site} d\Omega_1} \int_{site} d\Omega_1 \frac{P_r^{site}(\Omega_1)}{P_r^{site}(\Omega_m)} = \frac{\Delta\Omega}{8\pi^2} \tag{26}$$

Because the factor $\Delta\Omega/8\pi^2$ is necessarily smaller than (or equal to) 1, the change in orientational degrees of freedom makes an unfavorable contribution to binding free energy. For both the orientational and translational probabilities, a Gaussian approximation can be made to relate the accessible orientation and volume to fluctuations in the angles and position.[18,37] In the following, we will use the limit of strong restraints to compute the accessible volume and orientation from FEP simulations.

With the definition of the accessible volume and orientation of the ligand in the binding site, the standard binding constant K_b° can be expressed in the familiar form

$$K_b^\circ = C^\circ \Delta V \left(\frac{\sigma\Delta\Omega}{8\pi^2}\right)\left(\frac{\mathcal{N}_c^{site}}{\mathcal{N}_c^{bulk}}\right) e^{-\beta[\Delta G_{int}^{site} - \Delta G_{int}^{bulk}]} \tag{27}$$

In eq 27 the conventional interpretation of the binding process becomes self-evident: a ligand in a single conformation is desolvated from the bulk solution and inserted to the binding site [$\Delta G_{\rm int}^{\rm site} - \Delta G_{\rm int}^{\rm bulk}$], and the contribution from conformation change as well as reduction in translational and orientational freedom are accounted for by $\mathcal{N}_{\rm c}^{\rm site}/\mathcal{N}_{\rm c}^{\rm bulk}$, $C^{\circ}\Delta V$ (with the proper standard state) and $\Delta\Omega/8\pi^2$, respectively. Conceptually, this offers a useful decomposition of the binding constant in terms of various contributions. Although the statistical mechanical definition of all the terms is unambiguous, each term depends (conditionally) on the details of how the restraint potentials are implemented. In comparison with other methods, only a quantitative comparison of the final binding constant values will be meaningful in general.

III. Computational Details

A. Free Energy Perturbation. The computation of $\Delta G_{\rm int}^{\rm site}$ and $\Delta G_{\rm int}^{\rm bulk}$ is done with a step-by-step series of alchemical transformations. The interaction free energies ($\Delta G_{\rm int}^{\rm site}$ and $\Delta G_{\rm int}^{\rm bulk}$) are decomposed into repulsive, dispersive, and electrostatic contributions[27]

$$\Delta G_{\rm int}^{\rm a} = \Delta G_{\rm rep}^{\rm a} + \Delta G_{\rm dis}^{\rm a} + \Delta G_{\rm elec}^{\rm a} \tag{28}$$

where a = site, bulk. The free energy components are computed with free energy perturbation (FEP)[21,22,35] techniques. To this end, an auxiliary simulation potential energy is constructed with couplings parameters controlling the interactions between the ligand and its environment: $\lambda_{\rm rep}$, and $\lambda_{\rm dis}$ for the repulsive and dispersive part of the Lennard-Jones potential, and $\lambda_{\rm elec}$, for electrostatics. In addition, when the ligand is in the binding site, translational and rotational restraints are applied. Their magnitude is controlled by the coupling parameters $\lambda_{\rm t}$ and $\lambda_{\rm r}$, respectively (when $\lambda_{\rm t}$ and $\lambda_{\rm r}$ go to zero, all translational and orientational restraints are removed). The total auxiliary potential energy is

$$U(\lambda_{\rm rep},\lambda_{\rm dis},\lambda_{\rm elec},\lambda_{\rm t},\lambda_{\rm r}) = U_0 + U^{\rm rep}(\lambda_{\rm rep}) + \lambda_{\rm dis}U^{\rm dis} + \lambda_{\rm elec}U^{\rm elec} + \lambda_{\rm t}u_{\rm t} + \lambda_{\rm r}u_{\rm r} \tag{29}$$

where U_0 is the potential of the system with the noninteracting ligand; $\lambda_{\rm rep}$, $\lambda_{\rm dis}$, $\lambda_{\rm elec}$, $\lambda_{\rm t}$,$\lambda_{\rm r}$ $\in$[0, 1]; $U^{\rm rep}(\lambda_{\rm rep})$ is the shifted Weeks−Chandler−Andersen (WCA)[29] repulsive interaction[27] between the ligand and its environment (when $\lambda_{\rm rep} = 0$ the repulsive interaction is completely turned off, when $\lambda_{\rm rep} = 1$, it is fully switched on); $U^{\rm dis}$ and $U^{\rm elec}$ are the WCA dispersive interaction and electrostatic interaction between the ligand and its environment, respectively. The total repulsive interaction is the sum over the shifted repulsive LJ-WCA pair potential[27]

$$u_{ij}^{\rm rep}(r;\lambda_{\rm rep}) = \begin{cases} \epsilon_{ij}\left\{\dfrac{(R_{ij}^*)^{12}}{[r^2+(1-\lambda_{\rm rep})^2(R_{ij}^*)^2]^6} - 2\dfrac{(R_{ij}^*)^6}{[r^2+(1-\lambda_{\rm rep})^2(R_{ij}^*)^2]^3} + 1\right\} & r \le R_{ij}^*\sqrt{1-(1-\lambda_{\rm rep})^2} \\ 0 & r > R_{ij}^*\sqrt{1-(1-\lambda_{\rm rep})^2} \end{cases} \tag{30}$$

where r and R_{ij}^* are the distance and minimal energy distance between atom type i and j, respectively. The dispersion interaction is the sum over the attractive LJ-WCA pair potential

$$u_{ij}^{\rm dis}(r) = \begin{cases} -\epsilon_{ij} & r \le R_{ij}^* \\ \epsilon_{ij}\left[\left(\dfrac{R_{ij}^*}{r}\right)^{12} - 2\left(\dfrac{R_{ij}^*}{r}\right)^6\right] & r \le R_{ij}^* \end{cases} \tag{31}$$

The insertion of the ligand into the binding pocket ($\Delta G_{\rm int}^{\rm site}$) is done in three steps, with the help of the coupling parameters controlling the nonbonded interaction ($\lambda_{\rm rep}$, $\lambda_{\rm dis}$, and $\lambda_{\rm elec}$). Two additional steps are used to remove the translational and orientational restraints. The repulsive contribution $\Delta G_{\rm rep}^{\rm site}$ corresponds to the process

$$U(\lambda_{\rm rep}=0,\lambda_{\rm dis}=0,\lambda_{\rm elec}=0,\lambda_{\rm t}=1,\lambda_{\rm r}=1) \rightarrow U(\lambda_{\rm rep}=1,\lambda_{\rm dis}=0,\lambda_{\rm elec}=0,\lambda_{\rm t}=1,\lambda_{\rm r}=1) \tag{32}$$

the dispersive contribution $\Delta G_{\rm dis}^{\rm site}$ corresponds to the process

$$U(\lambda_{\rm rep}=1,\lambda_{\rm dis}=0,\lambda_{\rm elec}=0,\lambda_{\rm t}=1,\lambda_{\rm r}=1) \rightarrow U(\lambda_{\rm rep}=1,\lambda_{\rm dis}=1,\lambda_{\rm elec}=0,\lambda_{\rm t}=1,\lambda_{\rm r}=1) \tag{33}$$

and the electrostatic contribution $\Delta G_{\rm elec}^{\rm site}$ corresponds to the process

$$U(\lambda_{\rm rep}=1,\lambda_{\rm dis}=1,\lambda_{\rm elec}=0,\lambda_{\rm t}=1,\lambda_{\rm r}=1) \rightarrow U(\lambda_{\rm rep}=1,\lambda_{\rm dis}=1,\lambda_{\rm elec}=1,\lambda_{\rm t}=1,\lambda_{\rm r}=1) \tag{34}$$

The free energies, $\Delta G_{\rm r}^{\rm site}$ and $\Delta G_{\rm t}^{\rm site}$, correspond respectively to the process

$$U(\lambda_{\rm rep}=1,\lambda_{\rm dis}=1,\lambda_{\rm elec}=1,\lambda_{\rm t}=1,\lambda_{\rm r}=1) \rightarrow U(\lambda_{\rm rep}=1,\lambda_{\rm dis}=1,\lambda_{\rm elec}=1,\lambda_{\rm t}=1,\lambda_{\rm r}=0) \tag{35}$$

and

$$U(\lambda_{\rm rep}=1,\lambda_{\rm dis}=1,\lambda_{\rm elec}=1,\lambda_{\rm t}=1,\lambda_{\rm r}=0) \rightarrow U(\lambda_{\rm rep}=1,\lambda_{\rm dis}=1,\lambda_{\rm elec}=1,\lambda_{\rm t}=0,\lambda_{\rm r}=0) \tag{36}$$

The insertion of the ligand into the bulk sovent is calculated according to the same protocol but without translational and orientation restraints ($\lambda_{\rm t} = 0$ and $\lambda_{\rm r} = 0$). It should be noted that the values of $\Delta G_{\rm rep}^{\rm site}$, $\Delta G_{\rm dis}^{\rm site}$, and $\Delta G_{\rm elec}^{\rm site}$ depend on the translational and orientational restraints (thus, they are *conditional* free energies), though a comparison with the values in the bulk, $\Delta G_{\rm rep}^{\rm bulk}$, $\Delta G_{\rm dis}^{\rm bulk}$, and $\Delta G_{\rm elec}^{\rm bulk}$, is nonetheless expected to be informative. Table 1 lists all the values of the coupling parameters used in the step-by-step alchemical free energy calculations.

B. MD Simulations. All the molecular dynamics calculations were carried out with the CHARMM program.[38] For the translational/orientational restraint potentials (with the

Table 1: Values of the Coupling Parameters Used in the Free Energy Calculations

λ_{rep}	0.0	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0					
λ_{dis}	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0[a]				
λ_{elec}	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0				
λ_t	0.0	0.0025	0.005	0.0075	0.01	0.02	0.04	0.06	0.08	0.1	0.2	0.4	0.6	0.8	1.0
λ_r	0.0	0.0025	0.005	0.0075	0.01	0.02	0.04	0.06	0.08	0.1	0.2	0.4	0.6	0.8	1.0

[a] See Results and Discussion section and ref 27 for the choice of the optimal number of windows in the dispersive free energy calculation.

ligand in the binding site), two trajectories of 40 ps were generated starting from different initial velocities with values of the coupling parameters λ_t and λ_r equal to [0.0, 0.0025, 0.0050, 0.0075, 0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0]. The last 20 ps were collected and processed with the weighted histogram analysis method (WHAM).[39−41] For the repulsion contribution, two MD trajectories of 110 ps were run for values of the coupling parameter λ_{rep} equal to [0.0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]. The last 100 ps were processed with WHAM for each pair of adjacent coupling constants. For the dispersion contribution, values of the coupling parameter λ_{dis}, equal to [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0], were used in the simulation (see also Results and Discussion). For the electrostatic contribution, two MD trajectories of 110 ps with values of the coupling parameter λ_{elec} equal to [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0] were generated, and the last 100 ps were processed with WHAM.

All the simulations were generated on reduced atomic systems using the generalized solvent boundary potential (GSBP)[26] for the bound ligand and using the spherical solvent boundary potential (SSBP)[25,27] for free ligand inbulk solvent. For the bound ligand simulations, only the atoms within 15 Å from the center of the binding pocket were allowed to move, while the influence of the rest was incorporated with GSBP. A basis set of 16 spherical harmonic functions was used to approximate generalized reaction field. The electrostatic interactions between explicit atoms were not truncated within the simulation sphere, but those beyond 12 Å were represented using an extended electrostatic method.[42] The ligand in solution was simulated using SSBP[25,27] with 400 explicit water molecules. The coupling parameters λ_{rep}, λ_{dis}, and λ_{elec} were the same as for the bound ligand simulations. Those nonbonded coupling parameters affect also the internal interactions of the ligand, thus an additional calculation in vacuo was performed for each ligand which was subtracted from simulations of the solvated and bound ligand to enable the evaluation of solvation free energy. The length of each window simulation was 20 ps equilibration followed by 40 ps of sampling. Langevin dynamics with a friction constant of 5 ps^{-1} was used to ensure thermalization and keep the temperature constant at 300 K. All bonds involving hydrogen atoms were fixed with the SHAKE algorithm.[43] The integration time-step of the Langevin dynamics was 2 fs. For comparison, the free energy calculations with the ligand in bulk solution were repeated with periodic boundary conditions (PBC) at constant pressure using particle mesh Ewald[44] summation for the electrostatics. The PBC simulations comprise the ligand and 125 TIP3P water molecules placed in a cubic box. The Ewald coefficient was set to 0.34 $Å^{-1}$, and a sixth-order B-spline interpolation was used with a grid of $32 \times 32 \times 32$. The pressure was kept constant at 1.0 atm by using the Langevin piston[45] method with a pressure piston mass of 400.0 amu and a piston collision frequency of 20 ps^{-1}. The temperature was held at 300 K with a mass 3000.0 kcal·ps^2 Nose-Hoover thermostat.[46] The integration time-step was 2 fs. Two 30 ps trajectories were run for each value of the coupling parameters, and the last 20 ps were used in WHAM analysis to compute the free energies. For both SSBP and PBC calculations, the COM of the ligand was weakly restrained to the center of the simulation region using a harmonic potential with a force constant 2.0 kcal/mol/$Å^2$ to prevent the ligand from drifting.

The binding free energy of benzene, toluene, *o*-xylene, *p*-xylene, ethylbenzene, benzofuran, indene, indole, isobutylbenzene, and *n*-butylbenzene is computed without the conformational restraint. The binding free energy of a polar molecule phenol that does not bind experimentally[47] was also computed for comparison. All of the binding free energy calculations start from the bound structure (with the exception that the benzene bound structure were used for phenol and toluene). The computing time for each ligand binding free energy comprises 156 independent simulations, for a total of about 130 CPU h on a 2.4 GHz Xeon processor.

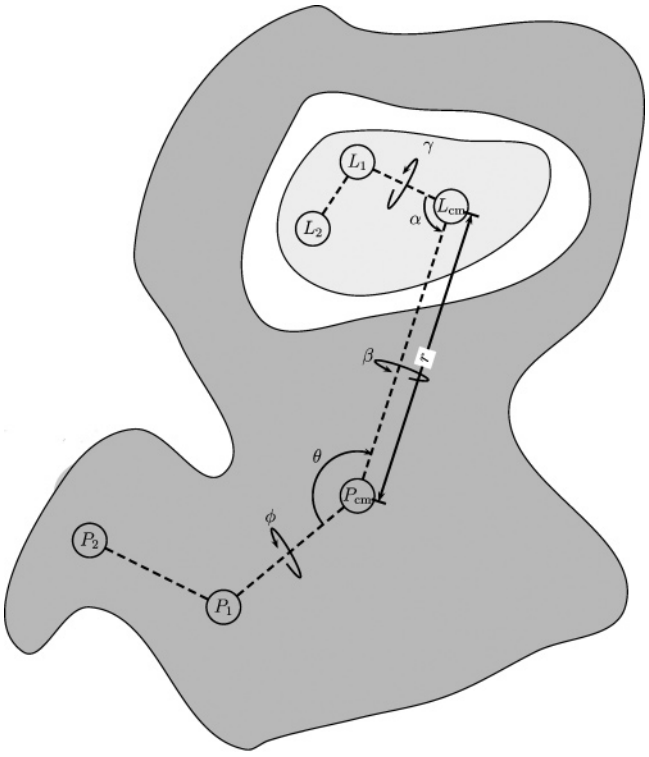

Figure 1. Restraint on the ligand. P_{cm} and L_{cm} are the COM of the protein and ligand, respectively. L_1 and L_2 are ligand atoms. P_1 and P_2 are protein atoms.

C. Restraint Factors. In the present computations, the translational restraint potential is implemented in polar coordinates

$$u_t(r,\theta,\phi) = \frac{1}{2}k_{dist}(r - r_0)^2 + \frac{1}{2}k_{ang}(\theta - \theta_0)^2 + \frac{1}{2}k_{ang}(\phi - \phi_0)^2 \quad (37)$$

where r is the distance between the COM of the ligand and protein; θ is the angle $\angle P_1P_{cm}L_{cm}$; and ϕ is the dihedral angle between plane $P_2P_1P_{cm}$ and $P_1P_{cm}L_{cm}$ (cf. Figure 1). The translational factor F_t is thus

$$F_t = \int_0^\infty dr\, r^2 \int_0^\pi d\theta \sin\theta \int_{-\pi}^{\pi} d\phi\, e^{-\beta u_t(r,\theta,\phi)} \approx r^2{}_0 \sin\theta_0 \sqrt{\frac{(2\pi k_B T)^3}{k_{dist} k_{ang}{}^2}} \quad (38)$$

where the rigid rotor approximation[19] has been used to derive the last equation. The rotational restraint potential has the following form (cf. Figure 1)

$$u_r = \frac{1}{2}k_r(\alpha - \alpha_0)^2 + \frac{1}{2}k_r(\beta - \beta_0)^2 + \frac{1}{2}k_r(\gamma - \gamma_0)^2 \quad (39)$$

where k_r is the force constant; α is the angle $\angle P_{cm}L_{cm}L_1$; β is the dihedral angle between plane $L_1L_{cm}P_{cm}$ and $P_1P_{cm}L_{cm}$; γ is the dihedral angle between plane $L_2L_1L_{cm}$ and $L_1L_{cm}P_{cm}$. The rotational factor is

$$F_r = \frac{1}{8\pi^2}\int_0^\pi d\alpha \int_{-\pi}^{\pi} d\beta \int_{-\pi}^{\pi} d\gamma \sin(\alpha) e^{-\beta u_r(\alpha,\beta,\gamma)} \approx \frac{1}{8\pi^2}\left(\frac{2\pi k_B T}{k_r}\right)^{3/2} \quad (40)$$

where the last equation assumed a rigid rotor approximation. All these integrals can also be computed numerically with high accuracy using Simpson's rule. In the present applications, the rigid rotor approximation for F_t and F_r yields essentially identical results to direct numerical integration.

D. Preparation of Protein Structure. Crystal structures of the complexes in the Protein Data Bank were used to generate the all-atom structures. The explicit simulation spheres were henceforth constructed from those all-atom structures. Take the benzene-T4 lysozyme complex, for example. The 15 Å explicit sphere was defined around the center of the binding pocket. Grand canonical Monte Carlo simulation (GCMC)[36] was run to put 107 water molecules in the sphere. Almost all the water molecules were distributed on the surface of the protein. Even though four water molecules could be inserted in binding pocket, they all drifted out within several picoseconds of a subsequent MD run. This is consistent with the observation that the hydrophobic engineered binding pocket in the L99A mutant does not contain water.[16,17] The anchor atoms in the ligand and the protein were assigned automatically by selecting the first non-hydrogen atom in the ligand or protein and then another non-hydrogen atom randomly. For example, for the benzene-bound lysozyme, P_1 was the backbone nitrogen atom of MET1; P_2 was the backbone carbon atom of TYR161. L_1 and L_2 were the two carbon atoms in the benzene ring in meta-position to each other. The corresponding initial values were $r_0 = 5.71568$ Å, $\theta_0 = 120.785°$, $\phi_0 = 151.906°$, $\alpha_0 = 74.9721°$, $\beta_0 = -93.3057°$, and $\gamma_0 = 145.364°$. For the toluene and phenol calculations, the protein coordinates were taken from the crystal structure of the benzene-T4 lysozyme complex (PDB 181L). For the rest of the ligands (*o*-xylene, *p*-xylene, ethylbenzene, isobutylbenzene, *n*-butylbenzene, benzofuran, indole, and indene), the initial simulation spheres were constructed from the correspond crystal structures (PDB 188L, 187L, 1NHB, 184L, 186L, 182L, 185L, and 183L, respectively).

E. Ligand Force Field. Apart for benzene, the binding free energies of phenol, toluene, *o*-xylene, *p*-xylene, ethylbenzene, benzofuran, indene, indole, isobutylbenzene, and *n*-butylbenzene were computed. The partial charges of the ligands were generated from the CHARMM PARAM22 models,[48] except for benzofuran and indene. The partial charges of *o*-xylene and *p*-xylene were created by replacing a hydrogen with a methyl group on the aromatic ring of toluene and setting the partial charge on the carbon attached to the methyl group to zero. Similarly, ethylbenzene, isobutylbenzene, and *n*-butylbenzene models were created by replacing a hydrogen on the benzene ring with ethyl and butyl groups and setting the charge on the corresponding carbon to zero. The model of phenol was constructed by replacing the methyl group of *p*-cresol with a hydrogen of partial charge $0.115e$ and setting the corresponding carbon partial charge to $-0.115e$. For benzofuran and indene, the partial charges were computed by the CHELPG[49] method implemented in Gaussian 98.[50] The rest of the ligand force field parameters were taken from the PARAM22 force field of CHARMM.[48]

IV. Results and Discussion

Table 2 lists the computed binding free energies for nine compounds with available crystal structures in the engineered lysozyme binding pocket. For small ligands such as benzene, toluene, and ethylbenzene, the agreement with experimental results is excellent. For larger ligands, the agreement is reasonable, although the binding free energies appear to be overestimated by a few kilocalories for the xylenes and butylbenzenes. In contrast, the computed free energy is several kilocalories less favorable in the case of indene. In the rest of this section, we examine the details of the binding free energy calculations, analyze the origin of the discrepancies, and conclude with the discussion of the issues that need to be addressed in order to improve the methodology

A. Independence from the Restraining Potentials. According to eq 5, the calculated binding free energy must be formally independent of the strength of the translational and orientational restraining potentials and of the choice of protein anchor atoms used to define those restraints (see also ref 14 or 19 for discussion). The binding of benzene was used as a test system to illustrate the independence of the free energy calculation on the restraints. First, a series of free energy calculations with different translational restraint constants while the rotational restraint is held constant is considered. The results in Table 3 show that the final binding

Table 2: Values of the Binding Free Energy of Aromatic Compounds with Lysozyme[a]

ligand	ΔG_{int}^{site}	ΔG_{int}^{bulk}	$\Delta\Delta G_r + \Delta\Delta G_t^\circ$	ΔV	$\Delta\Omega/8\pi^2$	σ	ΔG_b°	exp[b]
benzene	-11.52 ± 0.09	-0.14 ± 0.12	5.42 ± 0.10	0.87	0.018	12	-5.96 ± 0.19	-5.19 ± 0.16
toluene	-11.72 ± 0.29	-0.25 ± 0.08	7.01 ± 0.06	1.57	3.1×10^{-3}	2	-4.46 ± 0.31	-5.52 ± 0.04
o-xylene	-15.74 ± 0.10	-0.20 ± 0.10	7.95 ± 0.12	0.95	1.4×10^{-3}	2	-7.59 ± 0.19	-4.60 ± 0.06
p-xylene	-16.04 ± 0.02	0.14 ± 0.20	7.12 ± 0.06	0.85	2.9×10^{-3}	4	-9.06 ± 0.21	-4.67 ± 0.06
ethylbenzene	-13.49 ± 0.24	-0.12 ± 0.22	8.33 ± 0.09	0.80	1.8×10^{-3}	1	-5.04 ± 0.34	-5.76 ± 0.07
benzofuran	-16.25 ± 0.02	-1.54 ± 0.20	9.09 ± 0.02	0.90	4.3×10^{-4}	1	-5.62 ± 0.20	-5.46 ± 0.03
indene	-11.02 ± 0.16	-0.28 ± 0.17	8.27 ± 0.03	0.60	2.0×10^{-3}	1	-2.47 ± 0.24	-5.46 ± 0.03
indole	-17.68 ± 0.13	-3.96 ± 0.11	9.48 ± 0.01	1.11	1.6×10^{-4}	1	-4.24 ± 0.17	-4.89 ± 0.06
isobutylbenzene	-18.45 ± 0.18	0.43 ± 0.33	9.21 ± 0.03	0.52	6.2×10^{-4}	1	-9.67 ± 0.38	-6.51 ± 0.06
n-butylbenzene	-18.31 ± 0.31	-1.08 ± 0.18	8.48 ± 0.02	0.57	1.9×10^{-3}	1	-8.75 ± 0.36	-6.70 ± 0.02
phenol	-13.11 ± 0.07	-4.68 ± 0.07	7.55 ± 0.12	1.35	1.9×10^{-3}	2	-0.88 ± 0.13	decoy

[a] The unit of the free energies is kcal/mol. The translational restraint force constants are 10 kcal·mol^{-1}·Å^{-2} for the distant and 200 kcal·mol^{-1}·rad^{-2} for the orientation. The rotational restraint force constant is 200 kcal·mol^{-1}·rad^{-2}. The unit of ΔV is Å^3. The bulk free energies are computed with PBC. The errors are the standard deviation of three simulations from different initial velocities. [b] Taken from ref 31.

Table 3: Values of the Binding Free Energy of Benzene with Different Translational Restraint[a]

	1:50	5:100	10:200	20:400	40:800	100:2000	200:4000
ΔG_{rep}^{site}	9.47 ± 0.11	8.73 ± 0.14	8.49 ± 0.10	8.30 ± 0.08	8.04 ± 0.09	8.18 ± 0.10	8.17 ± 0.16
ΔG_{dis}^{site}	-18.78 ± 0.05	-18.79 ± 0.04	-18.83 ± 0.02	-18.86 ± 0.02	-18.77 ± 0.03	-18.78 ± 0.02	-18.79 ± 0.05
ΔG_{elec}^{site}	-1.19 ± 0.01	-1.18 ± 0.00	-1.18 ± 0.02	-1.18 ± 0.01	-1.19 ± 0.01	-1.2 ± 0.00	-1.19 ± 0.01
ΔG_{int}^{site}	-10.50 ± 0.10	-11.24 ± 0.10	-11.52 ± 0.09	-11.73 ± 0.08	-11.91 ± 0.10	-11.79 ± 0.10	-11.81 ± 0.20
$\Delta\Delta G_r + \Delta\Delta G_t^\circ$	3.96 ± 0.28	4.95 ± 0.11	5.42 ± 0.10	5.65 ± 0.16	5.70 ± 0.25	6.19 ± 0.13	5.98 ± 0.16
$-k_BT \ln(F_tC^\circ)$	3.58	4.48	5.10	5.72	6.34	7.16	7.78
$-k_BT \ln(F_tC^\circ)$ RRA	3.58	4.48	5.10	5.72	6.34	7.15	7.78
ΔV	5.53	1.65	0.87	0.52	0.54	0.30	0.28
$\Delta\Omega/8\pi^2$	0.033	0.021	0.018	0.020	0.025	0.014	0.022
ΔG_b°	-6.40 ± 0.30	-6.15 ± 0.15	-5.96 ± 0.13	-5.94 ± 0.18	-6.07 ± 0.27	-5.46 ± 0.16	-5.69 ± 0.26

[a] The unit of the free energies is kcal/mol. The units of the force constants k_{dist} and k_{ang} are kcal·mol^{-1}·Å^{-2} and kcal·mol^{-1}·rad^{-2}, respectively. The translational restraints are written as k_{dist}: k_{ang}. The rotational restraint of benzene is kept fixed at 200 kcal·mol^{-1}·rad^{-2}. Total is defined as a sum of ΔG_{int}^{site} and $\Delta\Delta G_r + \Delta\Delta G_t^\circ$ (the symmetry factor 12 is included in $\Delta\Delta G_r$). The unit of ΔV is Å^3. The solvation free energy ΔG_{bulk} is -0.14 kcal·mol^{-1} and is used to compute ΔG_b°. The errors are the standard deviations from five simulations with different initial velocity. The value of $-k_BT \ln F_t$ computed with numerical integration and its rigid rotor approximation (RRA) are also listed.

free energy is indeed independent of the restraint. In Figure 2, the dependence of ΔG_t^{site},ΔG_r^{site}, ΔG_{int}^{site}, and $-k_BT \ln(F_tC^\circ)$ with respect to the strength of the translational restraint is plotted. Despite significant variations in the different components, the sum of $-k_BT \ln(F_tC^\circ)$, $-\Delta G_t^{site}$, and ΔG_{int}^{site} remains nearly invariant with respect to the change of the restraint. At weak restraint, there are small deviations, but the sum reaches a stable value quickly at moderate and strong restraints. This suggest that sampling is less efficient if the restraints are not sufficiently strong. To avoid adding strong artificial forces in the simulation system, we chose the translational restraint force constants to be 10 kcal·mol^{-1}·Å^{-2} for the distance and 200 kcal·mol^{-1}·rad^{-2} for the angles.

Second, a series of free energy calculations with different rotational restraint constants with constant transational restraint are considered. Again, the results in Table 4 show that the final binding free energy is indeed independent of the restraint. In Figure 3, the free energies with respect to the change in rotational restraint are plotted. It is observed that the variations in $-\Delta G_r^{site}$ and $-k_BT \ln(F_r)$ nearly cancel each other, and the resulting binding free energy is independent of the restraint. For the remaining calculations, we chose the rotational restraint force constant to be 200 kcal·mol^{-1}·rad^{-2} for all the rotational angles. The present results do not display the dependence on the restraint force constants encountered by Hamelberg and McCammon in binding free energy computations of water in the interior of proteins.[10] Most likely, the convergence of the free energy results is sensitive to various sampling issues, which can be system-dependent.

The information in Tables 3 and 4 can also be exploited to estimate the accessible volume and orientational freedom of the bound ligand acccording to eqs 23 and 26. The accessible volume, ΔV, of benzene is about 0.5 Å^3; the values at low restraint strength should be discarded because eq 23 holds only in the strong restraint limit. The orientational freedom, $\Delta\Omega/8\pi^2$, is about 1%. These values confirm that benzene does not have much freedom to translate or rotate in the binding pocket. In a previous study, Hermans and Wang[16] also estimated the free energies corresponding to the loss of translational and rotational freedom of benzene bound to T4 lysozyme to be around 5.1 and 4.8 kcal/mol, respectively. However, an exact match is difficult because the definition of $\Delta\Delta G_t^\circ$and $\Delta\Delta G_r$ depends on the applied restraining potentials. In particular, although the final results for the binding free energy do not depend on the applied restraints, a meaningful identification of the contributions associated with the loss of translation or rotation evaluated via eqs 22−26 requires that the applied restraints be centered on the most probable position and orientation of the ligand in the binding pocket. Such estimates may also be related to the results obtained from a quasi-harmonic approximation

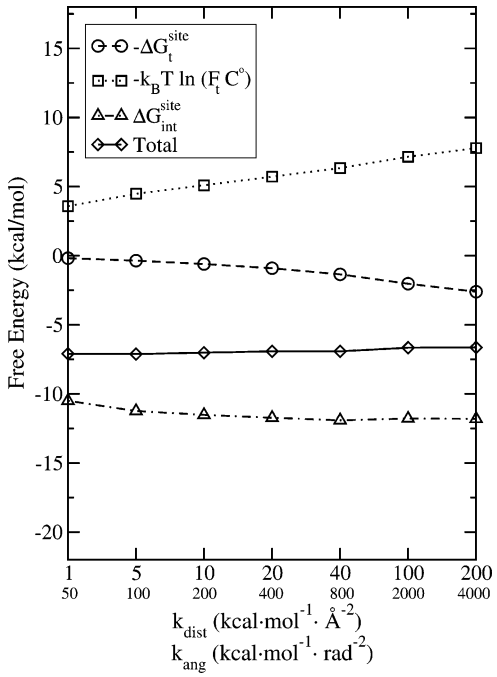

Figure 2. Independence of the binding free energy of benzene on the strength of the translational restraining potential. Total is the sum of $\Delta G_{\text{int}}^{\text{site}}$, $-\Delta G_{\text{t}}^{\text{site}}$, and $-k_{\text{B}}T \ln(F_{\text{t}}C^{\circ})$. The top row tick labels of the *x*-axis show the distance force constants k_{dist}. The lower row tick labels show the angular force constants k_{ang}. The restraint on the rotation is fixed at 200 kcal·mol^{-1}·rad^{-2}.

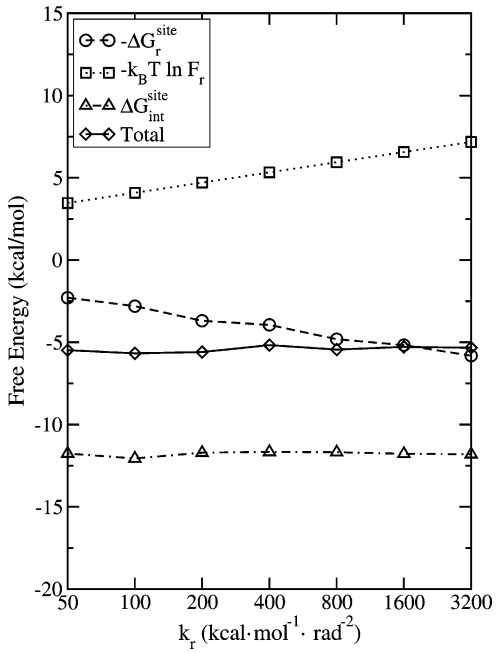

Figure 3. Independence of the binding free energy of benzene on the strength of the orientational restraining potential. Total is the sum of $\Delta G_{\text{int}}^{\text{site}}$, $-\Delta G_{\text{r}}^{\text{site}}$, and $-k_{\text{B}}T \ln F_{\text{r}}$. The translational restraint is held fixed at 10 kcal·mol^{-1}·Å^{-2} for the distance and 200 kcal·mol^{-1}·rad^{-2} for the angles.

by assuming that the position and orientation distribution of the bound ligand is Gaussian. Generally, the accessible volumes for different ligands in the binding pocket have similar values, though there can be considerable reduction of orientational freedom for the largest ligands.

The computed standard binding free energy of benzene is very close to the −5.19 kcal/mol experimental value for all of the restraint strength in both Tables 3 and 4 except the weakest restraint. The final computed standard free energy is comparable to the former calculated binding free energy of −5.14 kcal/mol (the experimental solvation free energy of benzene was used in that calculation).[16]

B. Decomposition of Binding Free Energy. The present free energy calculations distinctively use the WCA separation[27] to separate the LJ potential contributions into strict repulsive and dispersive contributions. Since the free energy computed in the binding site contains the effect from the restraint potentials, it should be noted that the resulting values for $\Delta G_{\text{rep}}^{\text{site}}$, $\Delta G_{\text{dis}}^{\text{site}}$, and $\Delta G_{\text{elec}}^{\text{site}}$ are *conditional* free energies. Nonetheless, the WCA decomposition is useful in clarifying the origin of the free energy change occurring upon ligand binding. In the following, we examine the importance of the different interaction components in the case of benzene.

In Figure 4, the progression of the repulsive free energy is shown as a function of the coupling parameter λ_{rep}. It is clear that the variation in the free energy with respect to λ_{rep} is quite different in the binding site and in bulk water. In general, the free energy contribution from the repulsive part of the LJ potential is smoother in the binding site than in the solvent. In particular, a cusp at $\lambda_{\text{rep}} = 0.2$ in the solvent environment is not observed in the binding site environment. Clearly, the L99A T4 lysozyme mutant has a preformed

Table 4: Values of the Binding Free Energy Values of Benzene with Rotational Restraint[a]

	50	100	200	400	800	1600	3200
$\Delta G_{\text{rep}}^{\text{site}}$	8.3	8.01	8.32	8.43	8.40	8.31	8.34
$\Delta G_{\text{dis}}^{\text{site}}$	−18.92	−18.89	−18.87	−18.89	−18.88	−18.93	−18.98
$\Delta G_{\text{elec}}^{\text{site}}$	−1.14	−1.18	−1.16	−1.20	−1.20	−1.16	−1.16
$\Delta G_{\text{int}}^{\text{site}}$	−11.76	−12.06	−11.71	−11.66	−11.68	−11.78	−11.8
$\Delta\Delta G_{\text{r}} + \Delta\Delta G_{\text{t}}^{\circ}$	5.68	5.79	5.42	5.89	5.64	5.90	5.87
$-k_{\text{B}}T \ln(F_{\text{r}})$	3.47	4.09	4.71	5.33	5.95	6.57	7.18
$-k_{\text{B}}T \ln(F_{\text{r}})$ RRA	3.46	4.08	4.70	5.32	5.94	6.56	7.18
ΔV	0.87	0.87	0.87	0.87	0.87	0.87	0.87
$\Delta\Omega/8\pi^2$	0.012	0.0096	0.018	0.0081	0.012	0.0072	0.0084
$\Delta G_{\text{b}}^{\circ}$	−5.94	−6.13	−6.15	−5.63	−5.90	−5.74	−5.79

[a] The unit of the free energies is kcal/mol. The unit of the rotational restraint force constant is kcal·mol^{-1}·rad^{-2}. All three rotational degrees of freedom are restrained with the same force constant. The translational restraint force constants are fixed at 10 kcal·mol^{-1}·Å^{-2} and 200 kcal·mol^{-1}·rad^{-2}. Total is defined as a sum of $\Delta G_{\text{int}}^{\text{site}}$ and $\Delta\Delta G_{\text{r}} + \Delta\Delta G_{\text{t}}^{\circ}$ (the symmetry factor 12 is included in $\Delta\Delta G_{\text{r}}$). The unit of ΔV is Å^3. The solvation free energy ΔG_{bulk} is −0.14 kcal·mol^{-1}. The value of $-k_{\text{B}}T \ln(F_{\text{r}})$ computed with numerical integration and its rigid rotor approximation (RRA) are also listed.

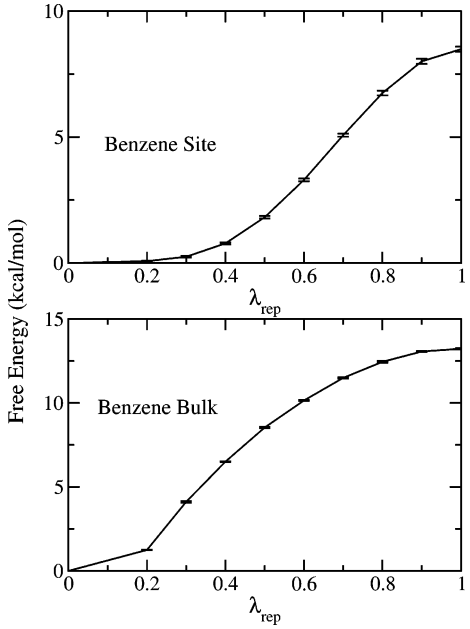

Figure 4. Progression of ΔG_{rep} with respect to the coupling parameter λ_{rep} for benzene in the binding site (top) and in bulk water (bottom).

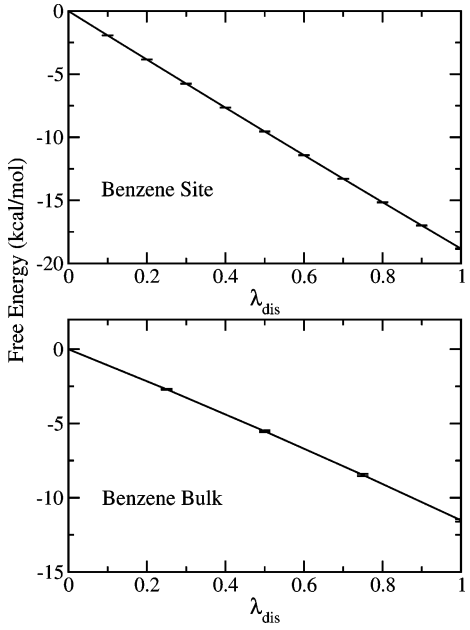

Figure 5. Progression of ΔG_{dis} with respect to the coupling parameter λ_{dis} for benzene in the binding site (top) and in bulk water (bottom).

cavity of the appropriate size for benzene, hence the early stages of repulsion are introduced without significant free energy cost. As a consequence, the free energy contribution from the repulsive part of the LJ potential favors ligand binding. The overall gain is about 5 kcal/mol in favor of ligand binding.

In Figure 5, the progression of the dispersive free energy is plotted as a function of the coupling parameter λ_{dis}. In both bulk solvent and in the binding site, the dispersive free energy increases linearly with λ_{dis}, though the slope is different in the two environments. The origin of the difference can be directly traced back to the number density of

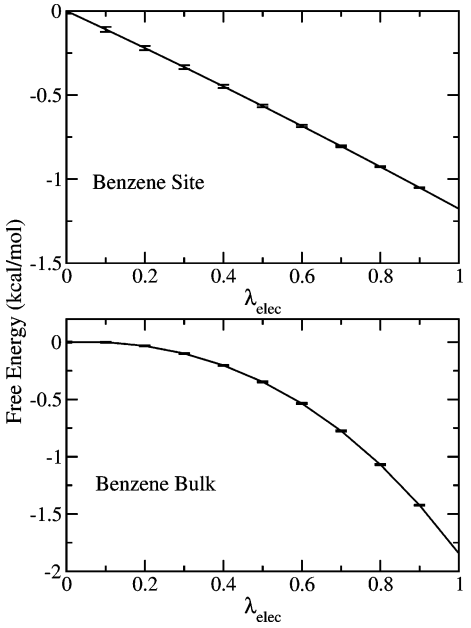

Figure 6. Progression of ΔG_{elec} with respect to the coupling parameter λ_{elec} for benzene in the binding site (top) and in bulk water (bottom).

van der Waals interaction centers per unit volume surrounding the ligand. This number is invariably larger in a protein compared with bulk water, which is a fairly *loose* and extended liquid. A similar observation was made by Levy and co-workers in developing implicit solvent models.[51] An important consequence is that the free energy contribution from the dispersive van der Waals interaction favors ligand binding. This is also related to trends previously observed with the linear interaction energy (LIE) approximation.[52] The linear progression of the dispersive free energy with λ_{dis} reflects the fact that there is no significant structural change as this interaction is switched-on during the simulation. That also leads to a computational advantage that this contribution converges very easily with a reduced number of simulation windows.[27] For the remaining calculations with ligands in the binding site, values of [0, 0.25, 0.5, 0.75, 1.0] were used for the coupling parameter λ_{dis} with trajectories of 100 ps.

In Figure 6, the progression of the electrostatic free energy of benzene is plotted with respect to the coupling parameter, λ_{elec}. The free energy varies linearly with the coupling parameter for benzene in the binding site, whereas it displays a typical quadratic behavior for benzene in bulk solvent. This indicates that the protein environment does not respond (electrostatically) to the charging of benzene. The ligand−protein interaction is dominated by a constant static field. This can be compared with the progression of the electrostatic free energy contribution of two polar ligands. The variation of the electrostatic free energy of indole and phenol as a function of λ_{elec} is shown in Figures 7 and 8, respectively. In the case of indole (a binder), the electrostatic free energy in the protein site is linear with respect to the coupling parameter λ_{elec}, whereas for phenol (a nonbinder) the progression of the electrostatic free energy in the binding site is not linear anymore. The difference results from the stronger polarity of phenol, which induces an electrostatic response in the protein. The simple progression of the

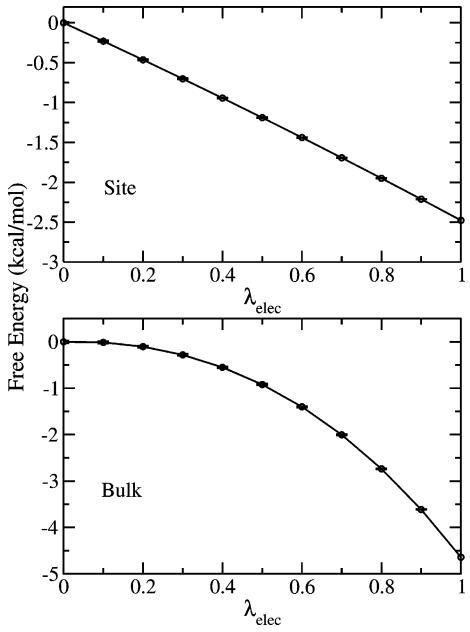

Figure 7. Progression of ΔG_{elec} with respect to the coupling parameter λ_{elec} for indole in the binding site (top) and bulk water (bottom).

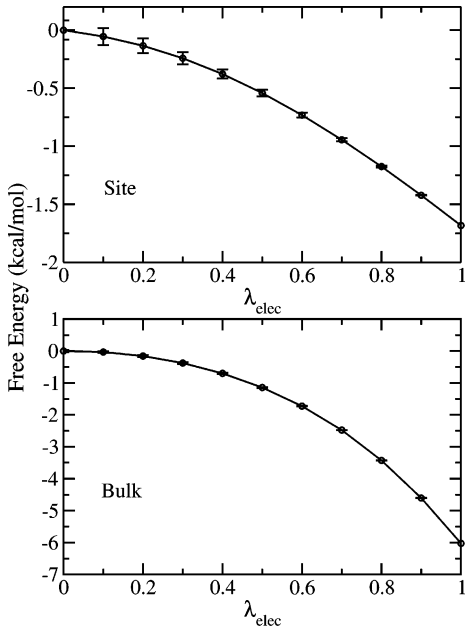

Figure 8. Progression of ΔG_{elec} with respect to the coupling parameter λ_{elec} for phenol in the binding site (top) and in bulk water (bottom).

electrostatic free energy in the binding site also suggests that it might be possible to reduce the number of sampling windows considerably without loss of precision. These observation suggest that the average electrostatic energy used in the linear interaction energy approximation (LIE)[52] may yield accurate results in the case of nonpolar ligands.

In Table 6, the free energy contributions in the binding site and bulk water are listed. The decomposition of solvation free energies computed with SSBP and PBC is also shown. As previously noted,[27] calculations with SSBP tend to yield slightly more unfavorable solvation free energy for the nonpolar molecules compared with PBC computations. For the final results of the present study, the values from PBC are used. Generally, both the repulsive and dispersive contributions favor the binding process, whereas the electrostatic contribution does not. For all of the ligands except indole, electrostatics contributes about 1 kcal/mol of unfavorable free energy. For the more polar indole, the electrostatic contribution to binding is 2.3 kcal/mol. The favorable repulsive contribution varies from molecule to molecule. The dispersive free energy has a clear dependence on the ligand size. As their size increases, the dispersive contribution becomes more and more favorable for binding. Ligands of similar size have similar dispersive contributions. This does seem to support the empirical observation that the *addition* of atoms improves the binding affinity (as long as they do not cause steric clashes).[53] For all of the ligands, most of the binding free energy comes from the nonpolar contribution. Even for the nonbinding molecule, phenol, the nonpolar contribution is comparable to that of toluene; it is the 4.4 kcal/mol of unfavorable electrostatic contribution that causes the unfavorable binding free energy for phenol. It is worth emphasizing that nonpolar contributions always exist independent of electrostatic contributions. Therefore, given the structural arrangement for the formation of the binding pocket buried in a protein, it is likely that nonpolar contributions always favor binding, for the molecules that do bind. The results of water binding free energy in different receptors[10] indeed show favorable nonpolar contributions, although the electrostatic contribution dominates.

C. Force Field and Free Energy Calculation. Accurate free energy calculations require highly refined force fields.[48,54] The atomic model affects the binding free energy in two ways: the solvation free energy in the bulk solvent and the free energy in the binding site. Consequently, the atomic model must be accurate for the ligand, protein, and water. The quality of the models in water solution can be evaluated by comparing computed solvation free energy with experimental values. In the present results, the calculated solvation free energies are about 1 kcal/mol more unfavorable. This gives the possibility of more favorable computed binding free energy, although, to arrive at this kind of conclusion, the deviation of the site free energy is ignored. The compatibility with the protein is more difficult to test directly. Here we resort to a less rigorous test of the ligand models by investigating how much the binding free energy changes when the ligand partial charges are changed.

First, the charges of the compounds are perturbed from CHARMM charges to AM1-CM2 AMSOL[55] charges. The simulation windows and sampling length are the same as the site and bulk electrostatic free energy calculation, respectively. The results are shown in Table 7. For all compounds except phenol, AMSOL charges give more favorable solvation and site free energies. The AMSOL electrostatic free energy of phenol is slightly more positive than that of the CHARMM charges. The AMSOL charges for indole and isobutylbenzene give 3 kcal/mol more favorable solvation free energies. In general, the site and solvation free energy change in the same direction when the ligand charges are changed. Therefore, for most of the compounds, the difference in the binding free energy with

Table 5: Values of the Solvation Free Energy of the Aromatic Ligand Compounds Computed with SSBP and PBC[a]

ligand		ΔG_{rep}^{bulk}	ΔG_{dis}^{bulk}	ΔG_{elec}^{bulk}	ΔG_{int}^{bulk}	$\Delta G^{exp\ b}$
benzene	SSBP	15.19	−13.30	−1.86	0.03	−0.87
	PBC	13.22 ± 0.04	−11.49 ± 0.11	−1.87 ± 0.01	−0.14 ± 0.12	
toluene	SSBP	17.27	−15.27	−1.73	0.27	−0.76
	PBC	14.59 ± 0.23	−13.17 ± 0.13	−1.67 ± 0.02	−0.25 ± 0.08	
o-xylene	SSBP	19.48	−17.05	−1.64	0.78	−0.90
	PBC	16.14 ± 0.10	−14.72 ± 0.05	−1.62 ± 0.01	−0.20 ± 0.10	
p-xylene	SSBP	19.64	−17.06	−1.56	1.02	−0.81
	PBC	16.33 ± 0.02	−14.64 ± 0.20	−1.55 ± 0.03	0.14 ± 0.20	
ethylbenzene	SSBP	19.05	−16.98	−1.86	0.22	−0.80
	PBC	16.38 ± 0.18	−14.65 ± 0.08	−1.84 ± 0.04	−0.12 ± 0.22	
benzofuran	SSBP	18.40	−17.31	−2.51	−1.42	
	PBC	15.46 ± 0.13	−14.81 ± 0.11	−2.19 ± 0.02	−1.54 ± 0.20	
indene	SSBP	18.99	−16.97	−1.92	0.10	
	PBC	16.16 ± 0.19	−14.57 ± 0.12	−1.87 ± 0.01	−0.28 ± 0.17	
indole	SSBP	18.41	−17.72	−4.68	−3.99	
	PBC	15.88 ± 0.15	−15.19 ± 0.12	−4.64 ± 0.02	−3.96 ± 0.11	
isobutylbenzene	SSBP	23.12	−20.67	−1.78	0.67	−0.44
	PBC	19.51 ± 0.14	−17.37 ± 0.20	−1.70 ± 0.03	0.44 ± 0.33	
n-butylbenzene	SSBP	22.20	−20.64	−1.98	−0.42	−0.40
	PBC	18.37 ± 0.14	−17.39 ± 0.15	−2.07 ± 0.02	−1.08 ± 0.18	
phenol	SSBP	15.92	−14.37	−6.18	−4.63	−6.62
	PBC	13.76 ± 0.04	−12.42 ± 0.05	−6.03 ± 0.03	−4.68 ± 0.07	

[a] The unit of free energies is kcal/mol. ΔG_{int}^{bulk} is defined as a sum of all the free energy components (repulsion, dispersion, and electrostatics). The errors are the standard deviation of three simulations from different initial velocities. [b] Taken from ref 65.

Table 6: Decomposition of the Interaction Free Energy in Bulk Solution and in the Binding Site for the Various Ligand Compounds[a]

ligand		ΔG_{rep}	ΔG_{dis}	ΔG_{elec}	ΔG_{int}	ΔG^{expb}
benzene	site	8.49 ± 0.10	−18.83 ± 0.02	−1.18 ± 0.02	−11.52 ± 0.09	
	bulk	13.22 ± 0.04	−11.49 ± 0.11	−1.87 ± 0.01	−0.14 ± 0.12	−0.87
toluene	site	11.40 ± 0.17	−22.19 ± 0.11	−0.93 ± 0.01	−11.72 ± 0.29	
	bulk	14.59 ± 0.23	−13.17 ± 0.13	−1.67 ± 0.02	−0.25 ± 0.08	−0.76
o-xylene	site	10.60 ± 0.13	−25.65 ± 0.03	−0.69 ± 0.01	−15.74 ± 0.10	
	bulk	16.14 ± 0.10	−14.72 ± 0.05	−1.62 ± 0.01	−0.20 ± 0.10	−0.90
p-xylene	site	9.11 ± 0.03	−24.46 ± 0.04	−0.69 ± 0.01	−16.04 ± 0.02	
	bulk	16.33 ± 0.02	−14.64 ± 0.20	−1.55 ± 0.03	0.14 ± 0.20	−0.81
ethylbenzene	site	13.53 ± 0.23	−26.02 ± 0.09	−0.99 ± 0.00	−13.49 ± 0.24	
	bulk	16.38 ± 0.18	−14.65 ± 0.08	−1.84 ± 0.04	−0.12 ± 0.22	−0.80
benzofuran	site	11.87 ± 0.02	−25.95 ± 0.03	−2.17 ± 0.00	−16.25 ± 0.02	
	bulk	15.46 ± 0.13	−14.81 ± 0.11	−2.19 ± 0.02	−1.54 ± 0.20	
indene	site	16.27 ± 0.13	−25.80 ± 0.03	−1.48 ± 0.01	−11.02 ± 0.16	
	bulk	16.16 ± 0.19	−14.57 ± 0.12	−1.87 ± 0.01	−0.28 ± 0.17	
indole	site	11.69 ± 0.10	−26.90 ± 0.03	−2.48 ± 0.01	−17.68 ± 0.13	
	bulk	15.88 ± 0.15	−15.19 ± 0.12	−4.64 ± 0.02	−3.96 ± 0.11	
isobutylbenzene	site	14.49 ± 0.22	−32.04 ± 0.04	−0.90 ± 0.00	−18.45 ± 0.18	
	bulk	19.51 ± 0.14	−17.37 ± 0.20	−1.70 ± 0.03	0.44 ± 0.33	−0.44
n-butylbenzene	site	15.15 ± 0.22	−32.45 ± 0.13	−1.00 ± 0.04	−18.31 ± 0.31	
	bulk	18.37 ± 0.14	−17.39 ± 0.15	−2.07 ± 0.02	−1.08 ± 0.18	−0.40
phenol	site	9.18 ± 0.06	−20.61 ± 0.05	−1.68 ± 0.08	−13.11 ± 0.07	
	bulk	13.76 ± 0.04	−12.42 ± 0.05	−6.03 ± 0.03	−4.68 ± 0.07	−6.62

[a] The solvation free energies are computed with PBC. The unit of the free energies is kcal/mol. The translation restraint force constants are 10 kcal·mol^{-1}·Å^{-2} for the distant and 200 kcal·mol^{-1}·rad^{-2} for the orientation. The rotational restraint force constant is 200 kcal·mol^{-1}·rad^{-2}. ΔG_{int} is defined as a sum of all the free energy components (repulsion, dispersion, and electrostatics). The errors are the standard deviation of three simulations from different initial velocities. [b] Taken from ref 65.

CHARMM and AMSOL charges are not very pronounced. For indole, the deviation in solvation free energy is expected because the CHARMM solvation free energy is about 2 kcal/mol less favorable (this is estimated with the experimental solvation free energy of 3-methylindole) than the experimental value. The indole model was recently revised to have a more realistic charge distribution and a more accurate solvation free energy in liquid water.[56] However, the binding free energy of −3.18 kcal/mol computed with the revised model is not as good in agreement with experiment than the

Table 7: Free Energy Differences from CHARMM Charges to AMSOL Charges[a]

ligand	$\Delta\Delta G_{elec}^{site}$	$\Delta\Delta G_{elec}^{bulk}$	$\Delta\Delta G_{b}^{\circ}$
benzene	−0.08	−0.36	0.28
ethylbenzene	−0.88	−0.89	0.01
isobutylbenzene	−1.27	−2.62	1.35
indole	−1.11	−3.23	2.12
n-butylbenzene	−0.64	−0.73	0.09
o-xylene	−1.14	−1.52	0.38
phenol	0.17	0.49	−0.32
p-xylene	−0.13	−1.53	1.40
toluene	−0.51	−0.91	0.40

[a] Free energies in kcal/mol.

Table 8: Free Energy Differences from CHARMM Charges to CHELPG Charges[a]

ligand	$\Delta\Delta G_{elec}^{site}$	$\Delta\Delta G_{elec}^{bulk}$	$\Delta\Delta G_{b}^{\circ}$
benzene	0.16	0.55	−0.39
ethylbenzene	0.26	0.43	−0.17
isobutylbenzene	0.74	−3.09	3.83
indole	0.22	0.23	−0.01
n-butylbenzene	0.48	−0.02	0.50
o-xylene	0.27	0.30	−0.03
phenol	−0.23	0.16	−0.39
p-xylene	−0.56	0.25	−0.81
toluene	0.58	0.39	0.19

[a] Free energies in kcal/mol.

original model. The decomposition of the free energies indicates that the difference results mostly from changes in the electrostatic contribution in the bulk solvent, which varied from −4.6 kcal/mol to nearly −5.8 kcal/mol, while the other terms have a similar magnitude. The interaction free energy in the binding site remain almost unchanged compared with the original model. The revised model is better solvated by improving the electrostatic interactions. The total nonpolar contribution to hydration for the revised model is 0.32 kcal/mol. This suggests that the balance of nonpolar (dispersion) and polar (electrostatic) contribution for the new model may not be optimal; even though the solvation free energy is more accurate, the binding free energy of the ligand inside a nonpolar protein cavity is not well described. This analysis illustrates (but does not resolve) the difficulties in designing a nonpolarizable force field with fixed partial charges that is simultaneously accurate in different environments such as liquid water and a nonpolar cavity in the interior of a protein.

Table 8 lists the changes of free energies from CHARMM to CHELPG charges. In this case, the solvation free energies computed with CHELPG charges are very close to those computed with CHARMM charges except for isobutylbenzene. The CHELPG charges overestimate the dipole moment for isobutylbenzene, yielding a large difference in solvation free energy. Table 9 lists the change of free energies from CHELPG to AMSOL charges for benzofuran and indene (there are no CHARMM partial charges for these molecules). The difference in solvation free energy is close to 3 kcal/mol. This large difference implies that the computed solvation free energy might not be accurate for benzofuran and indene, although we have found no experiment solvation free energy for these molecules. These tests with different compound force field models clearly indicate that binding free energy calculations place a very high demand on the force field, although, for many of the test compounds, the variations between different parameter sets are acceptable.

Table 9: Free Energy Differences from CHELPG Charges to AMSOL Charges[a]

ligand	$\Delta\Delta G_{elec}^{site}$	$\Delta\Delta G_{elec}^{bulk}$	$\Delta\Delta G_{b}^{\circ}$
benzofuran	−0.83	−3.07	2.24
indene	−0.38	−2.97	2.59

[a] Free energies in kcal/mol.

D. Flexibility and Conformational Sampling. In the present study, the alchemical FEP simulations were started from the experimentally determined structure of the protein−ligand complex. It obviously is advantageous to make use of any available high-resolution structural information about the protein−ligand complex, whenever this is possible. In the case of small ligands such as benzene, the change in the protein conformation is quite small. As the ligand vanishes in the alchemical simulation, the mobile atoms of the protein in the GSBP inner region relax to their average position in the apo structure, and the alchemical free energy simulation converges with no problems. The flexibility of a binding pocket makes an important contribution to the binding free energy.[57] In this context, the current FEP simulations protocol with the GSBP approximation is able to incorporate the influence of the local angstrom-scale thermal fluctuations of the protein and ligand as well as the loss of translational and orientational freedom into the calculated binding free energy. Nonetheless, it should noted that the finite GSBP simulation system gives rise to some artifacts. The average fluctuations of the mobile backbone atoms in the inner region of the GSBP simulation are slightly smaller than those from the fully solvated protein with PBC. There is an obvious compromise in choosing the size of the inner GSBP region, reducing the computational cost, and realistically simulating the dynamics of the atoms surrounding the ligand.

There can also be additional difficulties to sample ligands that are highly flexible in bulk solution but become restricted in the bound state.[20] This can be illustrated with the case of *n*-butylbenzene, as the movement of the butyl group becomes restricted upon binding. In this case, brute force sampling of the butyl group internal rotation might be inefficient. To address this issue, we introduce a conformational restraint on the butyl group and recomputed the free energies. The conformational restraint potential is $k_c\zeta^2$, where $k_c = 2$ kcal·mol·Å^{-2} and ζ is the RMSD from the bound conformation. The conformational factor (free energy) is computed by introducing a reference RMSD in the restraint potential, $k_c(\zeta - \zeta_0)^2$. The RMSD is sampled at each ζ_0 values of 0, 0.2, 0.4, 0.6, 0.8, ..., 1.8, 2.0 Å, for 60 ps, through which the unbiased distribution of RMSD at $\zeta_0 = 0$ Å, $\rho(\zeta)$, is computed.[58] The comparisons of the calculated free energies are listed in Table 10. In the particular case of *n*-butylbenzene, the forced sampling with the conformational restraint does not change significantly the resulting computed binding free energy. The system is dominated by a single bound conformation for the ligand, and only the local fluctuation

Table 10: Decomposition of the Binding Free Energy for *n*-Butylbenzene Binding Computed with Conformational Restraint[a]

ligand		ΔG_{rep}	ΔG_{dis}	ΔG_{elec}	ΔG_{int}	ΔG_c	$\Delta\Delta G_r + \Delta\Delta G_t^\circ$	ΔG_b°
n-butylbenzene with conf. restr.	site	14.95	−32.22	−0.95	−18.22	0.23	8.61	−7.99
	bulk	21.51	−20.90	−1.66	−1.05	0.80		
n-butylbenzene no conf. restr.	site	15.15	−32.45	−1.00	−18.31	0	8.48	−8.75
	bulk	18.37	−17.39	−2.07	−1.08	0		

[a] The unit of the free energies is kcal/mol. The translation restraint force constants are 10 kcal·mol^{-1}·Å^{-2} for the distant and 200 kcal·mol^{-1}·rad^{-2} for the orientation. The rotational restraint force constant is 200 kcal·mol^{-1}·rad^{-2}. The conformational restraint is $k_c\zeta^2$, where $k_c = 2$ kcal·mol·Å^{-2}, and ζ is the RMSD of *n*-butylbenzene from the bound conformation.

around it contributes to the binding free energy. Therefore, the conformational restraint is not essential in this case. However, for highly flexible ligands with many rotable groups the conformational restraint can help to achieve a better sampling.[59]

Of more significance are the difficulties encountered if the protein conformation in the complex differs significantly from that of the apo state, as in the case of the larger ligands. For example, there is considerable shift in the backbone position and large B-factor in helix F of the T4 lysozyme L99A mutant upon binding the largest ligands;[31] the RMSD per residue between the bound and apo structures for isobutylbenzene reaches a maximum of 2.5 Å, centered mostly around helix F (residues 109−113). An insufficiently sampled alchemical simulation started from the protein−ligand complex may underestimate the reversible work needed to distort and adapt the protein conformation to the ligand (the *strain* energy). The magnitude of the strain energy associated with distortion of the protein in the region of the binding site of T4 lysozyme has been estimated to be on the order of 2 kcal/mol by NMR experiments.[60] Because of the insufficient simulation time in the present calculations, the side chain atoms in helix F do not relax to their position in the apo structure and the calculated binding free energy is slightly too favorable for the largest ligands. This problem could be addressed by generating longer simulations, or alternatively, by generalizing the present theoretical formulation to introduce a conformational restriction potential u_c for biasing the conformation of the protein. Following this treatment, the influence of the bias could be rigorously removed and unbiased to obtain binding free energies without generating prohibitively long simulations. Further work along those lines is being pursued.

E. Importance of Accurate Starting Structure of the Complex. The calculated binding free energy of indene is several kcal/mol more unfavorable than the experimental value. Because the result for this molecule seems unaccountably inaccurate compared to the other ligands, it is particularly informative to understand the origin of this failure. According to the decomposition of the free energy, the difference arises mainly from the repulsive contribution in the binding site. This is indicative that, somehow, there is not enough space in the binding pocket for indene. The crystal structure shows that the binding of indene in the engineered lysozyme site induces increased disorder (as indicated by the temperature factors) and positional displacement of the helix F (residues 109−113).[31] Similar changes from the apo structure are observed for other ligands with a similar shape. But the other ligands (which are quite similar) do not display an unfavorable repulsive contribution as indene does. Because the displacement of helix F between the indene and *o*-xylene bound lysozyme is quite similar, we decided to compute the binding free energy of indene using the crystallographic structure of lysozyme with *o*-xylene bound (PDB 188L). As seen in Table 11, using this protein conformation, there is a significant reduction of the repulsive free energy (7 kcal/mol) upon binding. The resulting total binding free energy ends up being −8.10 kcal/mol (compared to −2.47 kcal/mol when indene is bound to its own crystal structure), which is 3 kcal/mol more favorable than the experimental value for indene. What could give rise to such discrepancy? Upon close examination, it is found that the determining factor is the χ_1 dihedral angle of Val111. When the side chain of Val111 in the indene bound crystal structure (PDB 183L) is set to the rotamer observed in the *o*-xylene bound crystal structure (PDB 188L), the computed free energy of indene becomes −8.06 kcal/mol. The unfavorable binding free energy of indene computed with crystal structure PDB 183L results directly from the rotamer conformation of Val111.

Table 11: Decomposition of the Binding Free Energy for Indene Computed from a Different Starting Structure of Lysozyme[a]

lysozyme structure	ΔG_{rep}^{site}	ΔG_{dis}^{site}	ΔG_{elec}^{site}	ΔG_{int}^{site}	$\Delta\Delta G_r + \Delta\Delta G_t^\circ$	ΔG_b°
PDB 183L	16.27	−25.80	−1.48	−11.02	8.27	−2.47
PDB 188L	9.04	−25.37	−0.82	−17.15	8.77	−8.10
PDB 183L*	9.12	−24.93	−0.95	−16.79	8.42	−8.06

[a] PDB 183L is the lysozyme crystal structure with indene bound. PDB 188L is the lysozyme crystal structure with *o*-xylene bound. PDB 183L* is the modified PDB 183L lysozyme crystal structure in which the χ_1 of Val111 changed to the value −45.48°, as in PDB 188L, a *o*-xylene bound structure. The unit of the free energies is kcal/mol.

In the *o*-xylene bound crystal structure, the backbone dihedral angles of Val111 are $\phi = -88.67°$ and $\psi = 1.14°$. Given the backbone dihedral angles, the rotamer library[61] predicts a single possible χ_1 of −60° (g^- configuration). The side chain dihedral angle in the crystal structure confirms the prediction with $\chi_1 = -45.58°$. Similarly, in the indene bound crystal structure, backbone dihedral angles of Val111 are $\phi = -82.22°$ and $\psi = 5.68°$, which leads to a predicted χ_1 angle of −60° in the same g^- configuration. However, the crystal structure has a χ_1 value of 171.42°. Analysis of all the other residues within 4 Å of all the nine complexes shows no such deviation from the rotamer library prediction in χ_1. In the indene bound crystal structure (PDB 183L), the backbone carbonyl oxygen of Val111 is only 2.74 Å away from the closest γ carbon atom. Examination of the electron density map of the indene bound structure[31] reveals that only one of the γ carbon atom of the valine side chain is clearly

resolved. In fact, both side chain dihedral χ_1 values would be compatible with the experimental density map. These structural observations support the idea that the Val111 side chain dihedral angle in the indene bound crystal structure (PDB 183L) might be incorrect.

The observation that the rotameric state of a single residue (in the crystal structure) in direct contact with the ligand can lead to several kcal/mol difference in the binding free energy is alarming but not unexpected. Clearly, this type of side chain barrier crossing, which takes place on a relatively slow time scale (typically, the barriers are several kcal/mol, e.g., see ref 62), could be sampled during very long FEP simulations.[63] However, having to generate prohibitively long trajectories and, in effect, trying to locate the correct conformation of the complex using brute force MD is obviously inefficient. An approach relying on a systematic search with a side chain rotamer library would be more efficient. Furthermore, it is antithetical to the spirit of the current methodological effort aimed at evaluating the binding free energy corresponding to a given state of the complex (scoring). These considerations highlight the great importance of an accurate starting structure (either from experiments or from docking followed by careful side chain optimization) for binding free energy calculation. They also show that much caution should be used to adjust empirical scoring functions on the basis of static X-ray protein−ligand structural complexes.

V. Conclusions

In this paper, we have presented a general statistical mechanical formulation of the standard binding constant and its application to the calculation of the binding free energy of aromatic ligands to an engineered nonpolar binding pocket in T4 lysozyme. The calculation of the free energy is done with several restraint potentials on the ligand translational and rotational (relative to the protein) degrees of freedom to improve the sampling. Those restraints are used to hold the uncoupled ligand in the binding pocket. We have shown that the final result (binding free energy) is independent of the choice of restraint potential. The definition of standard states appears naturally in the theoretical formulation. To decrease computational time, the MD simulations in the binding site were carried out with a reduced system including only atoms in a 15-Å sphere around the binding pocket, while the rest of the system is treated with a generalized solvent boundary potential (GSBP).[26] The solvation free energy calculation was done with both spherical solvent boundary potential (SSBP)[25] and periodic boundary potential. With GSBP the binding free energy calculation for a single ligand is achieved in 130 CPU h on a 2.4 GHz Xeon processor.

According to the different character of molecular interactions, the nonbonded interaction free energy was separated into repulsive, dispersive, and electrostatic contributions. Generally, the repulsive and dispersive interaction contribute significantly to the binding free energy, while the electrostatic interaction is slightly unfavorable. The repulsive free energy contribution is generally smaller in the binding site than in the bulk solvent, which is consistent with the existence of a preformed binding cavity in the protein. Furthermore, the dispersive van der Waals free energy contributes favorably to binding because the protein environment is more dense than bulk water. As expected, it is the large unfavorable contribution in the electrostatic free energy, insufficiently compensated by the nonpolar protein cavity, that that makes binding unfavorable in the case of the polar molecule phenol.

The agreement of calculated binding free energies with experimental values is generally satisfactory. For small ligands such as benzene, toluene, and ethylbenzene, the results are in excellent agreement with experiment, while for larger ligands, the agreement is reasonable though the computed free energies are somewhat more favorable than the experimental ones. Beyond specific successes and failures, the current study provides a good opportunity to highlight many of the key issues needed to predict accurate binding free energy using atomic models with realistic interactions and, in many ways, outlines the challenges that lie ahead. Those are as follows: (1) correct statistical mechanical formulation of the binding free energy, (2) accuracy and transferability of the atomic potential function, (3) access to accurate structure of the protein−ligand complex, and finally, (4) adequate conformational sampling of the molecular systems.

First and foremost, a valid statistical mechanical treatment of the binding process should establish the basis of any computational treatment. In particular, a translational restraint for the uncoupled ligand is an unavoidable ingredient of a sound double decoupling method (DDM) based on alchemical free energy. Trying to assess the convergence and precision of extensively long alchemical free energy simulations generated without translational restraint for the uncoupled state[64] only adds to the confusion because the significance of such calculations is unclear.

Calculations of absolute binding free energies place high demand on the accuracy of a force field. To examine how much the force field could change the computed results, we computed the change in binding free energies when partial charges of the ligands are changed. While the difference is not pronounced in general, for indene and benzofuran, the change in partial charges gives rise to differences up to 3 kcal/mol in the bulk electrostatic free energy. Even though the influence of the partial charge was certainly limited by the nonpolar nature of the ligands and binding pocket in the present case, such differences in binding free energies cannot generally be ignored.

Issues of protein conformation, flexibility, and thermal fluctuations are also particularly important. Having a good starting structure is essential for accurate binding free energy computations. This was dramatically illustrated in the case of indene, which yielded an incorrect free energy by several kcal/mol due to a single side chain misconfiguration in a high resolution crystallographic structure. Nonetheless, assuming that a good starting structure is available, the current binding free energy calculation method can be precise. The good news is that the small local angstrom-scale thermal fluctuations of the ligand and protein are sampled very well during relatively modest MD simulations. Simulation of a reduced system with GSBP may have limited the fluctuation of F-helix, which is somewhat displaced upon binding of

the larger ligands. The loss of conformational freedom of a flexible ligand upon binding requires special attention to get results of quantitative accuracy. A PMF approach based on the RMSD of the ligand relative to its bound conformation is used to obtain an efficient sampling of the ligand conformation restriction. This approach, which was illustrated with the case of *n*-butylbenzene, could be generalized to include biasing restraints on the conformation of the receptor as well as the ligand. Sampling the thermal fluctuations and motions in the system is absolutely essential; it simply cannot be avoided. It seems highly unlikely that simple effective scoring schemes based on a static (fixed) snapshot of the protein−ligand complex would ever be able to incorporate the effects of those fluctuations correctly.

The present effort shows that computationally tractable methods yielding first-principle absolute binding free energies are promising. With the development of more accurate force field and efficient sampling of the protein flexibility with conformational restraints, binding free energy predictions could soon become a useful tool in pharmaceutical studies.

Acknowledgment. We thank Brian Shoichet for helpful discussions and for providing AMSOL charges. We are grateful to Hyung-June Woo for providing the initial coordinates of the benzene-T4 lysozyme complex. We also thank Yanxiang Zhao for help with the analysis of the lysozyme crystal structure and Alexander D. MacKerell, Jr. for providing the new indole model. This work was supported by the National Science Foundation through Grant MCB-0415784.

References

(1) Ehrlich, P. Address in pathology on chemotherapeutics: scientific principles, methods, and results. *Lancet* **1913**, *182*, 445−451.

(2) Vindigni, A. Energetic dissection of specificity in serine proteases. *Comb. Chem. High Throughput Screening.* **1999**, *2*, 139−153.

(3) Cheng, A. C.; Calabro, V.; Frankel, A. D. Design of rna-binding proteins and ligands. *Curr. Opin. Struct. Biol.* **2001**, *11* (4), 478−484.

(4) Garvie, C. W.; Wolberger, C. Recognition of specific dna sequences. *Mol. Cell.* **2001**, *8* (5), 937−946.

(5) Smith, P. E.; Pettitt, B. M. Modeling solvent in biomolecular systems. *J. Phys. Chem.* **1994**, *98*, 9700−9711.

(6) Orozco, M.; Luque, F. J. Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem. Rev.* **2000**, *100*, 4187−4225.

(7) Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligand to macromolecular receptors. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644−2676.

(8) Pearlman, D. A.; Charifson, P. S. Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. *J. Med. Chem.* **2001**, *44*, 3417−3423.

(9) Dixit, S. B.; Chipot, C. Can absolute free energies of association be estimated from molecular mechanical simulations? The biotin-streptavidin system revisited. *J. Phys. Chem. A* **2001**, *105*, 9795−9799.

(10) 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.

(11) 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.

(12) Jorgensen, W. L. Interactions between amides in solution and the thermodynamics of weak binding. *J. Am. Chem. Soc.* **1989**, *111*, 3770−3771.

(13) Hermans, J.; Subramaniam, S. The free energy of xenon binding to myoglobin from molecular dynamics simulation. *Isr. J. Chem.* **1986**, *27*, 225−227.

(14) Roux, B.; Nina, M.; Pomès, R.; Smith, J. C. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: A molecular dynamics free energy perturbation study. *Biophys. J.* **1996**, *71*, 670−681.

(15) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermaldynamic basis for computation of binding affinities: A critical review. *Biophys. J.* **1997**, *72*, 1047−1069.

(16) 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.

(17) Mann, G.; Hermans, J. Modeling protein-small molecule interactions: Structure and thermodynamics of noble gases binding in a cavity in mutant phage T4 lysozyme L99A. *J. Mol. Biol.* **2000**, *302*, 979−989.

(18) Luo, H.; Sharp, K. On the calculation of absolute macromolecular binding free energies. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10399−10404.

(19) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535−9551.

(20) Woo, H.-J.; Roux, B. Calculation of absolute protein−ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6825−6830.

(21) Straatsma, T. P.; McCammon, J. A. Computational alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407−435.

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

(23) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. Efficient computation of absolute free energies of binding by computer simulations. application to the methane dimer in water. *J. Chem. Phys.* **1988**, *89*, 3742−3746.

(24) Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the use of orientational restraint and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* In press.

(25) Beglov, D.; Roux, B. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.* **1994**, *100*, 9050−9063.

(26) Im, W.; Bernèche, S.; Roux, B. Generalized solvent boundary potential for computer simulations. *J. Chem. Phys.* **2000**, *114*, 2924−2937.

(27) Deng, Y.; Roux, B. Hydration of amino acid side chains: Nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules. *J. Phys. Chem. B* **2004**, *108*, 16567−16576.

(28) Banavali, N. K.; Im, W.; Roux, B. Electrostatic free energy calculation using the generalized solvent boundary potential method. *J. Chem. Phys.* **2002**, *117*, 7381−7388.

(29) Weeks, J. D.; Chandler, D.; Andersen, H. C. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.* **1971**, *54*, 5237−5247.

(30) 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.

(31) Morton, A.; Matthews, B. W. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: Linkage of dynamics and structural plasticity. *Biochemistry* **1995**, *34*, 8576−8588.

(32) Bjerrum, N. Untersuchungen über Ionenassoziation. I. der Einfluss der Ionenassoziation auf die Akativität der Ionen bei Mittleren Assoziationsgraden. *K. Dan. Vidensk. Selsk., Mat.-Fys. Medd.* **1926**, *7*, 1−48.

(33) Mihailescu, M.; Gilson, M. K. On the theory of noncovalent binding. *Biophys. J.* **2004**, *87*, 23−36.

(34) Allen, T. W.; Andersen, O. S.; Roux, B. Energetics of ion conduction through the gramicidin channel. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 117−122.

(35) Zwanzig, R. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* **1954**, *22*, 1420−1426.

(36) Woo, H.-J.; Dinner, A. R.; Roux, B. Grand canonical Monte Carlo simulation of water in protein environments. *J. Chem. Phys.* **2004**, *121*, 6392−6400.

(37) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys. J.* **2004**, *86*, 67−74.

(38) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM − A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(39) Ferrenberg, A. M.; Swendsen, R. H. New monte carlo technique for studying phase transitions. *Phys. Rev. Lett.* **1988**, *61*, 2635−2638.

(40) Ferrenberg, A. M.; Swendsen, R. H. Optimized monte carlo data analysis. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.

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

(42) Stote, R. H.; States, D. J.; Karplus, M. On the treatment of electrostatic interactions in biomolecular simulation. *J. Chim. Phys. Phys.-Chim. Biol.* **1991**, *88*, 2419−2433.

(43) 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.

(44) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(45) Feller, S. E.; Zhang, Y.; Paster, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulations: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613−4621.

(46) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695−1697.

(47) Wei, B. Q.; Baase, W. A.; Weaver, L. H.; Matthews, B. W.; Shoichet, B. K. A model binding site for testing scoring functions in molecular docking. *J. Mol. Biol.* **2002**, *322*, 339−355.

(48) MacKerell, A. J.; Bashford, D.; Bellot, M.; Dunbrack, R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S., D. J.-M.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, W., III; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chems. B* **1998**, *102*, 3586−3616.

(49) Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361−373.

(50) 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.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98, Revision A.9*; Gaussian, Inc.: Pittsburgh, PA, 1998,

(51) Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. On the nonpolar hydration free energy of proteins: Surface area and continuum solvent models for the solute-solvent interaction energy. *J. Am. Chem. Soc.* **2003**, *125*, 9523−9530.

(52) Marelius, J.; Hansson, T.; Åqvist, J. Calculation of ligand binding free energies from molecular dynamics simulations. *Int. J. Quantum Chem.* **1998**, *69*, 77−88.

(53) Shaikh, S. A.; Ahmed, S. R.; Jayaram, B. A molecular thermodynamic view of DNA-drug interactions: a case study of 25 minor-groove binders. *Arch. Biochem. Biophys.* **2004**, *429*, 81−99.

(54) MacKerell, A. D., Jr. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **2004**, *25*, 1584−1604.

(55) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. New class IV charge model for extracting accurate partial charges from wave function. *J. Phys. Chem. A* **1998**, *102*, 1820−1831.

(56) Macias, A. T.; MacKerell, A. D., Jr. Ch/*pi* interactions involving aromatic amino acids: Refinement of the CHARMM tryptophan force field. *J. Comput. Chem.* **2005**, *26*, 1452−1463.

(57) Wei, B. Q.; Weaver, L. H.; Ferrari, A. M.; Matthews, B. W.; Shoichet, B. K. Testing a flexible-receptor docking algorithm in a model binding site. *J. Mol. Biol.* **2004**, *337*, 1161−1182.

(58) Souaille, M.; Roux, B. Extension to the weighted histogram analysis method: Combining umbrella sampling with free energy calculations. *Comput. Phys. Comm.* **2001**, *135*, 40−57.

(59) Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular dynamics simulations with restraint potentials. *Biophys. J.* In press.

(60) Mulder, F. A. A.; Mittermaier, A.; Hon, B.; Dahlquist, F. W.; Kay, L. E. Studying excited states of proteins by NMR spectroscopy. *Nat. Struct. Biol.* **2001**, *8*, 932−935.

(61) Dunbrack, R. L., Jr.; Karplus, M. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nat. Struct. Biol.* **1994**, *1*, 334−340.

(62) Petrella, R. J.; Karplus, M. The energetics of off-rotamer protein side-chain conformations. *J. Mol. Biol.* **2001**, *312*, 1161−1175.

(63) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* **2003**, *119*, 5740−5761.

(64) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. Direct calculation of the binding free energies of FKBP ligands. *J. Chem. Phys.* **2005**, *123*, 084108.

(65) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

CT060037V