CENTENNIAL FEATURE ARTICLE

Computations of Standard Binding Free Energies with Molecular Dynamics Simulations

Yuqing Deng[†,§] and Benoît Roux*[,†,‡]

Biosciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois, and Department of Biochemistry and Molecular Biology, Gordon Center for Integrative Sciences, University of Chicago, 929 East 57th Street, Chicago, Illinois

Received: August 29, 2008; Revised Manuscript Received: October 22, 2008

An increasing number of studies have reported computations of the standard (absolute) binding free energy of small ligands to proteins using molecular dynamics (MD) simulations and explicit solvent molecules that are in good agreement with experiments. This encouraging progress suggests that physics-based approaches hold the promise of making important contributions to the process of drug discovery and optimization in the near future. Two types of approaches are principally used to compute binding free energies with MD simulations. The most widely known is the alchemical double decoupling method, in which the interaction of the ligand with its surroundings are progressively switched off. It is also possible to use a potential of mean force (PMF) method, in which the ligand is physically separated from the protein receptor. For both of these computational approaches, restraining potentials may be activated and released during the simulation for sampling efficiently the changes in translational, rotational, and conformational freedom of the ligand and protein upon binding. Because such restraining potentials add bias to the simulations, it is important that their effects be rigorously removed to yield a binding free energy that is properly unbiased with respect to the standard state. A review of recent results is presented, and differences in computational methods are discussed. Examples of computations with T4-lysozyme mutants, FKBP12, SH2 domain, and cytochrome P450 are discussed and compared. Remaining difficulties and challenges are highlighted.

I. Introduction

In recent years, a number of studies have reported computations of standard (absolute) binding free energy of small ligands to proteins using molecular dynamics (MD) simulations with explicit solvent molecules. In many ways, the results are in good agreement with experiments. By this, one means that the calculated binding free energy is roughly within experimental variations, and thus a meaningful predicted quantity. The progress is encouraging and suggests that physics-based approaches hold the promise of becoming an important predictive tool for drug discovery and optimization in the near future. The goal of this article is to review the recent progress computations of standard binding free energies, contrast the different methodologies that are available, and highlight the remaining challenges for the future.

Issues of molecular recognition, involving the noncovalent association of small ligands to large macromolecules with high affinity and specificity, play a crucial role in biology and medicinal chemistry.[1−3] Computational studies can help elucidate the fundamental principles governing these issues at the molecular level. Moreover, improving our ability to screen large databases of compounds *in silico* to identify potential lead drug molecules with accurate prediction of binding affinities could have a great impact on structure-based drug design. So far, however, computational screening methods have had only a mixed success rate. In the language of molecular modeling, ligand screening can be separated into two loosely define steps, "docking" and "scoring".[4] The docking step aims at predicting the preferred orientation of the ligand molecule bound to the protein receptor (the ligand "pose"), and the scoring step aims at predicting the binding affinity of the ligand for a given ligand orientation. While docking can proceed successfully via heuristic simplifications, the worse shortcomings of ligand screening approaches stem from the approximate scoring functions.

The fundamental principles controlling ligand binding are relatively well understood, but scoring often relies on extremely simplified approximations in order to achieve the computational efficiency needed to handle large databases.[5−7] Nonetheless, to have any predictive and practical value, scoring must reflect the binding free energies with sufficient accuracy. Arguably, physics-based approaches such as free energy perturbation molecular dynamics (FEP/MD) simulations represent the most accurate approach to quantitatively characterize the binding free energy of small ligands with macromolecules. FEP/MD simulations can naturally handle the influence of solvent and dynamic flexibility,[6] and previous studies indicate that the method is more reliable than simpler scoring schemes.[8,9]

The calculation of free energies is among the most important applications of biomolecular simulations.[10−13] Initial applications of FEP/MD to biomolecular systems, such as, for example, the calculations of hydration free energies[14,15]

* Corresponding author.
† Argonne National Laboratory.
‡ University of Chicago.
§ Current address: Zymeworks Inc., 540-1385 West 8th Ave., Vancouver, BC V6H 3V9, Canada.

10.1021/jp807701h CCC: $40.75
Published on Web 01/15/2009

Yuqing Deng was born in Mianyang, Sichuan Province, in China in 1973. He received a B.Sc. in Applied Chemistry from the University of Electronic Science and Technology (Chengdu, China) in 1994 and obtained a M.Sc. under the guidance of Zhida Chen in theoretical physical chemistry from Peking University (Beijing, China) in 1996. He then moved on to Brown University where he obtained his Ph.D. in 2002 on the mechanism of polyatomic vibrational relaxation in simple liquids under the direction of Richard Stratt. During his postdoctoral work with Benoît Roux, he developed computational methodologies for free energy calculations. He currently works as a molecular simulation scientist at Zymeworks Inc. (Vancouver, BC).

Benoît Roux was born in the city of Montréal, Canada, in 1958. In 1981, he received a B.Sc. in Physics from the University of Montréal, followed by a M.Sc. in Biophysics in 1985 under the supervision of Rémy Sauvé. In 1990, he obtained a Ph.D. in Biophysics from Harvard University under the direction of Martin Karplus. In the last decade, he has held positions at the University of Montréal and the Weill Medical College of Cornell University. Since 2005, he is Professor in the Department of Biochemistry and Molecular Biology at the University of Chicago with a joint appointment as Senior Computational Biologist at Argonne National Laboratory.

and of binding free energies,[16−21] go back to the 1980s. The early work was obviously burdened by the limited sampling from short simulations, though the potential of FEP methods was recognized. Considerable progress has been made since then, in part due to the increased availability of powerful computers. Moreover, the theoretical framework for carrying out various free energy computations has been greatly clarified.[13] In recent years, free energy simulations have been used to characterize solvation properties of a wide range of molecules and have now become integral tools to test, validate, and refine biomolecular force fields.[22−25] Absolute hydration free energies were calculated for the amino acids side chains[26,27] and for a wide range of small molecules.[28−32] Water−cyclohexane transfer free energies of amino acids have been determined.[33] Decomposition of free energies into contributions by the core repulsion, van der Waals dispersion, and electrostatic interactions helps gain insight on the intermolecular forces underlying solvation properties.[27] While FEP/MD remains too demanding computationally for screening extremely large databases of compounds, the recent success and the increase in computational power suggest that these methods are going to play an important role in drug discovery and optimization in the near future.

In the following section, we review briefly the statistical thermodynamics formulation of molecular association, and describe the main approaches to compute standard binding free energies from MD simulations. There is the alchemical double decoupling method, in which the interaction of the ligand with its surroundings is progressively switched off, and there is the potential of mean force method, in which the ligand is physically separated from the receptor. For both of these computational strategies, restraining potentials may be activated and released during the simulation to sample more efficiently the changes in translational, rotational, and conformational freedom of the ligand and protein upon binding. A review of recent results is then presented, and differences in computational methods are discussed. Particular attention is given to studies of ligand binding to T4-lysozyme mutants[34−39] and FKBP12.[40−44] Our attention shall be devoted mainly to the recent advances in the computation of standard (absolute) binding free energies based on simulations with explicit solvent. The wide range of scoring methods that are available for virtual screening are not covered here, as they have been the object of recent reviews (e.g., see, refs 45−47). Similarly, the vast literature concerned with alchemical FEP calculations of *relative* free energies will not be reviewed extensively. Nevertheless, it is important to emphasize that familiar FEP techniques, when combined with systematic explorations of the effect of chemical modifications of lead compounds (e.g., heterocycle substitutions, methyl, chlorine, and hydroxyl scans, etc.), represent a powerful strategy in drug design (see refs 48−51 as recent examples). The review is concluded with a discussion highlighting the remaining challenges with an outlook to the future.

II. Theory and Methods

A. Statistical Mechanics and Equilibrium Binding Constant. The main quantity of interest is the *standard binding free energy*, ΔG_b°. One may note that the term *absolute* binding free energy has often been used in theoretical studies,[39,42,44,52−55] largely for the purpose of contrasting ΔG_b° with the *relative* free energies that are obtained via standard FEP simulations.[10−13] In the following, we shall use the term standard binding free energy.

To calculate ΔG_b° with computer simulations, a mathematical relation has to be established between macroscopic observables and microscopic variables. For the sake of concreteness, let us consider a protein macromolecule (P) in thermodynamic equilibrium with a dilute solution containing ligand molecules (L). The equilibrium constant K_b of the binding reaction L + P $\rightleftharpoons$ LP is defined as $K_b = [LP]/([L][P])$, where [L], [P], and [LP] are the equilibrium concentrations of the unbound ligand, unbound protein, and bound complex, respectively. The standard binding free energy is defined from the equilibrium constant by $\Delta G_b^\circ = -k_B T \ln(C^\circ K_b)$, where C° is a standard concentration, k_B is the Boltzmann constant, and T is the absolute temperature.

To obtain an expression for K_b, it is useful to consider the occupancy operator H equal to 1, when the ligand is in the binding site of the protein and 0 otherwise. The probability to find the protein with one ligand bound is $\mathscr{P}_1 = \langle H \rangle$, the probability to find the protein with no ligand bound is $\mathscr{P}_0 = 1 - \mathscr{P}_1$, and the ratio of these occupancy probabilities is directly related to the equilibrium constant, $\mathscr{P}_1/\mathscr{P}_0 = K_b[L]$. This ratio is independent of the concentration of the protein, as long as the solution is relatively dilute. Following classical statistical mechanics, it can be shown that the equilibrium constant can be expressed as[36,53,56,57]

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

where U is the total potential energy of the system, $\beta \equiv 1/k_B T$, $\mathbf{L}$ and $\mathbf{X}$ represent, respectively, the coordinates of the ligand and all remaining atoms (solvent and protein), $\mathbf{r}_L$ is the position of the center of mass of the ligand, and $\mathbf{r}^*$ is some arbitrary position far away in the bulk region.[36,55,56] The derivation, which can be traced back to Bjerrum,[58] is more direct than the traditional treatments based on chemical potentials.[53,57,59] One may note that K_b has the dimension of volume, hence the need to multiply by the standard concentration C° (1 mol/L = 1/1660 Å^3), to define the dimensionless quantity $K_b C^\circ$ and a meaningful binding free energy ΔG_b°.

On the basis of eq 1, one may picture the binding process as the ligand leaving the bulk region (in the denominator) and moving into the binding site (in the numerator). In principle, the binding free energy could be determined from the average population of the bound state $\langle H \rangle$, via an unbiased trajectory. However, while this has been done previously (e.g., see ref 60),

it is generally more advantageous to introduce a sequence of intermediate states between the bound and unbound "end-point" states in order to enable practical computations. The most popular approaches rely on alchemical free energy thermodynamic perturbation techniques in which the intermediates are chosen to progressively switch "off" the interactions of the ligand with its surroundings.[21,36,42,54,56,57,61] Alternatively, convenient intermediates may be chosen to control the physical separation between the ligand and the protein without alchemical decoupling. In this approach, the binding constant is computed via a protein−ligand potential of mean force (PMF).[43,55,62,63] One may refer to this PMF-based approach as the "pulling" method.[43] Such PMF-based approach can also be applied along a 4th spatial dimension,[64−66] which has some analogy with the soft-core potentials used in alchemical decoupling.[67] With these different methods, any number of restraining potentials may be activated and then released at different stages during the calculations to improve convergence. As long as the total reversible work between the end-points is properly calculated, the final result should be independent of the intermediate states.

B. Standard Binding Free Energy from Alchemical Perturbation. The alchemical method computes the reversible thermodynamic work for "decoupling" the ligand from its surrounding, that is, to switch off all the interaction of the ligand with the protein and/or bulk solvent. As an illustration, let us first consider the free energy associated with decoupling the ligand from the bulk solvent

$$e^{-\beta\Delta G_{int}^{bulk}} = \frac{\int d\mathbf{L} \int d\mathbf{X}\, e^{-\beta U}}{\int d\mathbf{L} \int d\mathbf{X}\, e^{-\beta U_0}} \qquad (2)$$

where U_0 represents the total potential energy of the system with a noninteracting (decoupled) ligand. The thermodynamic decoupling in eq 2, which corresponds to the transfer of the ligand to the gas phase, is often referred to as an "annihilation" process. This could, however, be misconstrued, since the ligand remains but is only decoupled from its environment. The solvation free energy in eq 2 can be computed using alchemical free energy perturbation (FEP) or thermodynamic integration (TI); by FEP and TI, we mean methods that rely on a finite or infinitesimal change in the Hamiltonian, respectively (see refs 11, 68, and 69 for reviews on methodology).

Applying directly the decoupling scheme of eq 2 for the ligand in the binding site can be attempted,[61] though this is often impractical. The noninteracting (decoupled) ligand should in principle drift away from the binding site and wander anywhere in the volume of the simulation box. This increases the difficulty to obtain a statistically converged and reversible free energy from standard methods (see the discussion in ref 54). A proper free energy calculation must be reversible; i.e., the change in free energy for introducing the ligand molecule into a system should be equal, but of opposite sign, to the free energy for removing the ligand. Nevertheless, if one assumes that the ligand has explored the entire simulation box and the free energy calculation is converged, then the equilibrium constant follows from eq 1

$$K_b = V_{pbc} \times e^{-\beta[\Delta G_{int}^{site} - \Delta G_{int}^{bulk}]} \qquad (3)$$

where V_{pbc} is the volume of the box for the periodic boundary conditions (PBC) used in the MD simulation to calculate ΔG_{int}^{site} and ΔG_{int}^{bulk} is the solvation free energy defined in eq 2. The relation to the standard state expressed in eq 3 presumes that the unrestrained alchemical decoupling state is fully converged, which would need to be monitored with attention (e.g., the ligand might get trapped in a metastable site and never return to the true binding site). Furthermore, using the above expression assumes implicitly that the ligand binds tightly to the protein and that the concentration of ligand in the MD box is large compared to $1/K_b$.

Such difficulties are easily avoided by introducing a restraining potential, u_t, at an intermediate step to control the translation of the ligand relative to the protein binding site. This strategy, introduced by Hermans and Subramaniam,[21] has been progressively enriched over the years by a number of additional developments and variants.[34−37,54,56] The restraining potential is introduced at one end-point to "confine" the uncoupled ligand within the binding site, and is then "released" at the other end-point, where this step can be carried out analytically.[21,54,56] Gilson et al.[57] called free energy calculations in which there is no translational restraint the "double annihilation method" (DAM) and calculations in which there is a translational restraint the "double decoupling method" (DDM).

In subsequent developments, restraining potentials have been introduced to control the ligand rotation,[34,36,37,54] conformation,[36,42,55] and protein side chain conformation.[38] The various potentials in FEP are sometimes called "virtual molecular tweezers"[34] or "virtual bonds".[54] The reversible work for the entire association/dissociation process is then carried out as a series of sequential steps, during which the interaction of the ligand with its surroundings (protein and solvent) as well as the various restraining potentials are switched "on" and "off".[42] The potentials restraining the translation, rotation, and conformation are denoted by u_t, u_r, and u_c, respectively. The standard binding free energy obtained from the step-by-step procedure can be expressed as[36,42]

$$\Delta G_b^\circ = \Delta\Delta G_{int} + \Delta\Delta G_t^\circ + \Delta\Delta G_r + \Delta\Delta G_c \qquad (4)$$

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 $\Delta\Delta G_t^\circ = [-\Delta G_t^{site} - k_B T \ln(F_t C^\circ)]$. ΔG_c^{site} and ΔG_c^{bulk} are the free energy cost associated with applying the conformational restraint on the ligand in the binding site and bulk solvent, F_t and F_r are the factors associated with applying restraints on ligand rotation and translation when it is decoupled, ΔG_t^{site} and ΔG_r^{site} are the free energy associated with applying restraints in the binding site, ΔG_{int}^{site} is the free energy associated with switching on ligand interaction in the binding site with all restraint potentials applied, ΔG_{int}^{bulk} is the free energy associated with switching on ligand interaction in bulk solvent with conformational restraint potential. The free energy associated with some of the restraints can be calculated analytically,[34,54,56] or via a numerical integral.[36,42,55] In practice, the potential u_c for restraining the conformation of the ligand has been written as a quadratic form $k\xi^2$, where ξ is the root mean square deviation (RMSD) of the ligand relative to its bound conformation.[36,42,55] Simple dihedral torsion potentials have also been used to confine the conformation of the protein by restraining the orientation of protein side chains.[38] Equation 4 provides a decomposition of the dominant factors affecting the binding free energy.[34,36,42] Normally, the interaction component, $\Delta\Delta G_{int}$, is highly favorable but is strongly opposed by unfavorable contributions from the ligand translation, $\Delta\Delta G_t^\circ$, rotation/orientation, $\Delta\Delta G_r$, and conformation, $\Delta\Delta G_c$. The

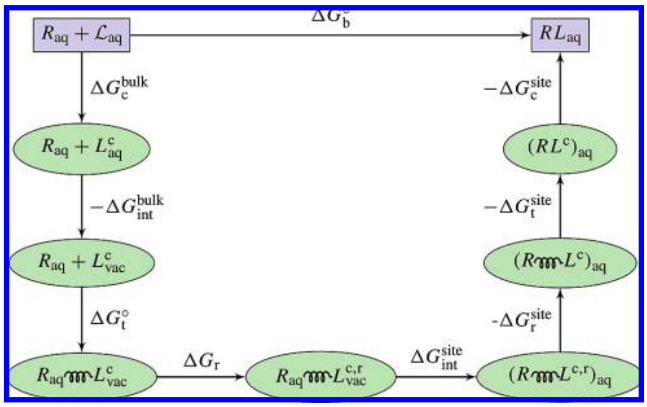

Figure 1. Schematic representation of the various steps to compute the standard binding free energy ΔG_b° according to the alchemical double decoupling method (DDM) with restraining potentials as expressed in eq 4. The complete "path" serves to link the two end-point states: the unbound state $R_{aq} + \mathcal{L}_{aq}$ with the receptor (R_{aq}) and the unrestricted ligand in bulk solvent ($\mathcal{L}_{aq}$) and the bound state RL_{aq}. In the figure, L_{aq}^c denotes the restricted ligand in bulk solvent with conformational restraint, L_{vac}^c denotes the noninteracting decoupled (vacuum) ligand with conformational restraint, and $L_{vac}^{c,r}$ denotes the noninteracting ligand with both conformational and rotational restraint potential (the spring represents the translational restraint). See ref 36 for a complete theoretical formulation of each term.

decomposition can also be used to shed some light on end-point approximations, such as the explicit/implicit solvent MM/PB-SA[13,70,71] and the linear interaction energy (LIE).[72,73] The value of the various contributions depends on the details of the restraint potential and the order in which specific terms are activated/deactivated, though the total standard binding free energy is independent of those factors. The complete DDM procedure is illustrated schematically in Figure 1.

The various restraining potentials serve as a "guide" to prevent large excursions of the configurations of the molecular system, thus helping to reduce the size of the configurational space that needs to be sampled between the end-points of a free energy calculation. However, there is sometimes the misperception that computations based on DAM yield more meaningful (unbiased) results because they do not involve restraining potentials. This is incorrect. Judiciously chosen restraining potentials actually enhance sampling efficiency.[37,38,42,55,74−76] Spatial confinement via biasing potentials is a general noise-reduction technique to aid the convergence of simulations (see ref 77). While the convergence and reversibility in standard DAM poses a real problem, restraining potentials may be activated and then released at different stages during a free energy calculation, as long as their free energy contributions are properly accounted for.

Any restraining potential that is activated (confine step) must be deactivated before the end state (release step) in order to yield properly unbiased results. The effects of switching any restraining potential on and off may be calculated via standard TI or FEP, or combined with umbrella sampling[78] to achieve a higher precision.[36,42,74] Typically, the free energies associated with u_t and u_r have been computed directly with FEP or TI.[34−37,42,55] In this case, unbiased simulations are assumed to be able to achieve sufficient conformational sampling. A more powerful approach is often needed to treat slow varying degrees of freedom involving the conformation of the ligand or the protein. For example, the free energy cost ΔG_c^{bulk} and ΔG_c^{site} associated with restraining the conformation of the ligand in the bulk or in the binding site by the potential u_c have been calculated as

$$\Delta G_c^{bulk} = -k_B T \ln\left(\frac{\int d\xi\, e^{-\beta W_c^{bulk}(\xi)} e^{-\beta u_c(\xi)}}{\int d\xi\, e^{-\beta W_c^{bulk}(\xi)}}\right) \quad (5)$$

and

$$\Delta G_c^{site} = -k_B T \ln\left(\frac{\int d\xi\, e^{-\beta W_c^{site}(\xi)} e^{-\beta u_c(\xi)}}{\int d\xi\, e^{-\beta W_c^{site}(\xi)}}\right) \quad (6)$$

where $W_c^{bulk}(\xi)$ and $W_c^{site}(\xi)$ are PMFs of the ligand in the bulk and in the binding site, respectively.[36,42,55] They must be calculated with biased simulations via umbrella sampling. A similar umbrella sampling technique has been utilized by Dill and co-workers to treat the effect of side chain configurations on ligand binding in T4-lysozyme.[38] The above procedure with eqs 5 and 6 constitutes the basis of what one might call a "deliberate" sampling strategy: a PMF along some clearly identified degree of freedom is first calculated with biased simulations, and then, unbiased averages are extracted via an explicit numerical integration of the probability distributions involving the Boltzmann factor of the unbiased PMF. The free energy difference $\Delta\Delta G_c = [\Delta G_c^{apo} - \Delta G_c^{holo}]$, corresponding to the reversible work for switching on a conformational restraint in one end-point state and switching it off in the other, has been referred to as a "confine-and-release cycle".[38]

C. Nonpolar and Electrostatics Interactions. A deeper insight into the microscopic factors driving ligand binding can be achieved by further dissecting the interaction components ΔG_{int}^{site} and ΔG_{int}^{bulk} according to the character of intermolecular forces. Intermolecular forces are dominated by short-range harsh repulsive interactions, arising from Pauli's exclusion principle, and long-range van der Waals attraction and electrostatic interactions, arising, respectively, from quantum mechanical dispersion and the nonuniform molecular charge distribution. In standard biomolecular potential functions, the nonpolar forces are modeled with Lennard-Jones (LJ) 6−12 potentials, while electrostatics forces are represented on the basis of Coulomb interactions between partial charges.[22−25] For example, the Lennard-Jones 6−12 potential can be separated into purely repulsive and attractive parts according to the Weeks−Chandler−Anderson (WCA) scheme.[27,79] Assuming that the repulsive, dispersive, and electrostatic components of the interactions are switched on sequentially, in a step-by-step process, the decoupling free energy may be written as[27]

$$\Delta\Delta G_{int} = \Delta\Delta G_{rep} + \Delta\Delta G_{dis} + \Delta\Delta G_{elec} \quad (7)$$

where $\Delta\Delta G_{rep} = [\Delta G_{rep}^{site} - \Delta G_{rep}^{bulk}]$, $\Delta\Delta G_{dis} = [\Delta G_{dis}^{site} - \Delta G_{dis}^{bulk}]$, and $\Delta\Delta G_{elec} = [\Delta G_{elec}^{site} - \Delta G_{elec}^{bulk}]$. The separation of intermolecular forces in eq 7 provides a useful framework for decomposing the interaction free energy of a molecular ligand with its surrounding into distinct contributions with a clear and well-defined physical meaning. The latter statement must be understood with caution because the various free energy terms are conditionally defined via the step-by-step reversible work in eqs 2 and 7. For example, the free energy ΔG_{dis} and ΔG_{elec} will not be the same if the electrostatic term is switched on before, or after, the van der Waals dispersion (though the value of ΔG_b° is invariant).

D. Binding Free Energy from a PMF. The equilibrium association constant and binding free energy may also be

calculated using a PMF,[80,81] without resorting to the alchemical decoupling steps as in DDM. In simple cases, a strategy based on the one-dimensional (1D) radial PMF may work[62]

$$K_b = \int_{site} 4\pi r^2 \, dr \, e^{-\beta[w(r)-w(r^*)]} \quad (8)$$

where r^* is a reference position far away in the bulk. However, the considerable complexity of biological systems often hinders the utilization of a straightforward approach. It may be hard to sample the spherical shell with a ligand at a distance r, required to utilize eq 8. Furthermore, the conformational freedom of the ligand may vary considerably between the bound and free states, which may be difficult to sample with unbiased simulations. To improve convergence in calculating the PMF, it may be advantageous to introduce various restraining potentials to limit the fluctuations of the conformation and orientation of the ligand, and to confine its translation along a well-defined axis "a" relative to the binding site. According to this restrained-PMF approach, the equilibrium binding constant is expressed in terms of a 1D integral[55]

$$K_b = S^* e^{-\beta[\Delta\Delta G_c + \Delta\Delta G_r - \Delta G_a^{site}]} \int_{site} dr \, e^{-\beta[w(r)-w(r^*)]} \quad (9)$$

where $w(r)$ is the 1D-PMF calculated in the presence of the configurational and orientational restraints. The quantity S^* corresponds to the effective cross-sectional area swept by the ligand restrained along the axis "a" when it is at a distance r^* from the protein. ΔG_a^{site} represents the free energy for restraining the bound ligand along the 1D axis. The additional factors $\Delta\Delta G_c$ and $\Delta\Delta G_r$ represent the loss of conformational and rotational freedom of the bound ligand relative to the unbound ligand and have the same meaning as in eq 4. As with the alchemical methods described above, restraining potentials of different shapes may be introduced to carry out a calculation, as long as their contributions are correctly accounted for in the final equilibrium binding constant.[82] It is illustrated schematically in Figure 2.

E. Molecular System and Solvation. It is possible to carry out detailed FEP molecular dynamics (MD) simulations to calculate the standard binding free energy of a ligand to a receptor with the theoretical framework presented above. For meaningful results, one must accurately simulate the thermal fluctuations and the environment-mediated interactions arising in diverse and complex systems (i.e., the bulk solution and the protein binding site). Computational approaches at different levels of complexity and sophistication have been used to describe the influence of solvent on biomolecular systems.[83] Those range from MD simulations based on all-atom models with PBC, in which the solvent is treated explicitly, to continuum electrostatic models in which the influence of the solvent is incorporated implicitly. Conducting all-atom FEP/MD simulations is, however, often prohibitive, and it is important to seek ways to reduce the computational cost of the calculations. An intermediate approach was developed, combining some aspects of both explicit and implicit solvent treatments.[84−86] It consists of simulating a nonperiodic system with a small number of explicit solvent molecules in the vicinity of a region of interest, while representing the influence of the surrounding solvent with an effective "solvent boundary potential". Many of our own free energy studies[27,36,42] based on reduced atomic systems were simulated using the spherical solvent boundary potential (SSBP)[85] for the bulk solvent and the generalized solvent boundary potential (GSBP) for the binding site.[86] The SSBP and GSBP are mixed explicit/implicit models: a detailed atomic model of the ligand and its nearest neighbors is simulated explicitly, while the influence of the rest of the system is incorporated implicitly via a mean-field continuum electrostatic approximation. Several mixed explicit/implicit schemes have been derived through ad hoc assumptions. In contrast, SSBP and GSBP have been formulated from a rigorous statistical mechanical reduction of the many-body configurational integrals.[85,86] SSBP includes the reaction field from the dielectric response of the solvent acting on the atoms of the simulation region, while GSBP includes the reaction field for a solvent region of arbitrary geometry, as well as the solvent-shielded static field from the distant atoms of the protein receptor. Simulations of a reduced system embedded into a continuum mean-field environment offer an attractive strategy to decrease the computational cost of MD/FEP simulations. Computationally, this can be very advantageous because binding specificity is often dominated by local interactions in the vicinity of the ligand while the remote regions of the receptor contribute only in an average manner. Nevertheless, one must be cautious with these approaches, particularly when the protein macromolecule undergoes a large conformational change upon ligand binding.

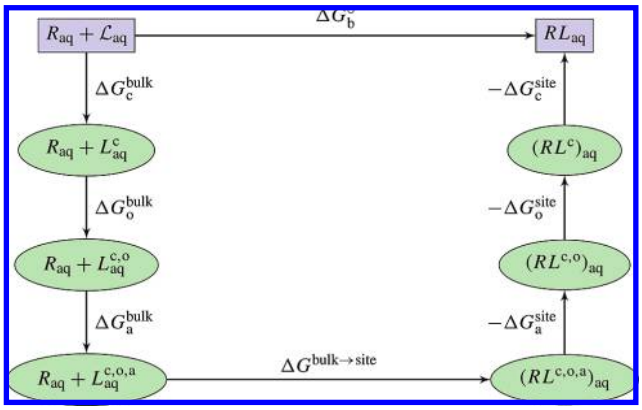

Figure 2. Schematic representation of the various steps to compute the standard binding free energy ΔG_b° according to a potential of mean force (PMF) method with restraining potentials as expressed in eq 9. The complete "path" serves to link the two end-point states: the unbound state $R_{aq} + \mathcal{L}_{aq}$ with the receptor (R_{aq}) and the unrestricted ligand in bulk solvent ($\mathcal{L}_{aq}$) and the bound state RL_{aq}. In the figure, L_{aq}^c denotes the configuration restricted ligand in bulk solvent, $L_{aq}^{c,o}$ denotes the ligand with both configuration and orientation restraints, and $L_{aq}^{c,o,a}$ denotes the ligand in the bulk with all configurational, orientational, and axial restraint potentials. ΔG_a^{bulk} and $\Delta G^{bulk\to site}$ are related to S^* and an integral over the ligand−receptor PMF in eq 9, respectively. See ref 55 for a complete theoretical formulation of each term.

III. Overview of Recent Results

A number of well-characterized ligand-binding protein systems have been the object of intense scrutiny, including mutants of T4-lysozyme (T4L),[87,88] the FK506 binding protein (FKBP12),[40−44] cytochrome P450,[52,89] and SH2 domains.[90−92] While current studies ultimately are aimed at clarifying the fundamental principles driving molecular association, they also serve in large part as "testing grounds" to highlight and address critical issues with the various computational methodologies.

Some of the most interesting and useful model systems to investigate noncovalent binding have been provided by mutants of T4-lysozyme (T4L) in which artificial cavities were engineered.[87,88] A hydrophobic cavity, created by the mutation

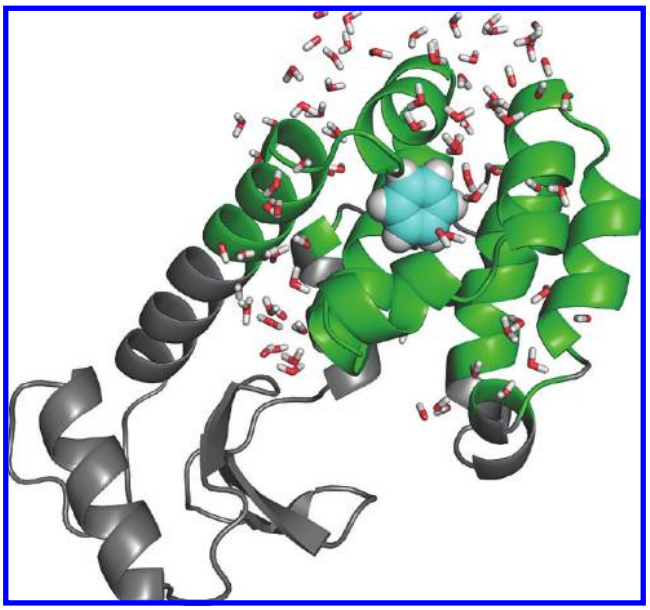

Figure 3. T4-lysozyme L99A mutant with benzene bound in the cavity. The gray parts are treated as a mean-field approximation with generalized solvent boundary potential.[86] See ref 36 for computational details.

L99A in T4L, has been found to bind nonpolar aromatic molecules such as benzene and indole.[87,93] Polar derivatives of benzene like phenol do not bind to this nonpolar pocket.[87] The double mutation M102Q-L99A provides a hydrogen bond partner for the ligand, thus making the cavity able to bind more polar ligands.[88] T4L/L99A with bound benzene is illustrated in Figure 3. The T4L systems have been the object of several computational studies.[34−39] By virtue of their simplicity, these systems are particularly attractive for computational studies. The binding sites are buried inside the protein and not directly accessible to the bulk solvent. In addition, the cavities are believed to be essentially "dry" in the holo state; i.e., no solvent molecules are displaced upon ligand binding.

Another system that has been extensively studied is the FK506 binding protein (FKBP12).[40−44] There is a vast collection of ligands which are real drug molecules displaying considerable flexibility. Furthermore, the binding site exposed to the bulk and solvent molecules must be displaced upon ligand binding. The FKBP12 system is also of extreme importance for immunosuppression because the binding of the drug FK506 to FKBP12 forms a complex that inhibits calcineurin and blocks the signal transduction pathway for T-cell activation.[94,95] One ligand included in a previous study[42] is shown in Figure 4. Although FKBP12 is a small rigid protein that does not undergo very large fluctuations, this system is considerably more complex than the simple pockets engineered in T4L.

A. Computational Methodologies. Relatively few studies have been based on PMF-based approaches. Those include the original study of amide association by Jorgensen,[62] the binding of K^+ into the gramicidin A channel,[63] the binding of the peptides to signaling modules,[90,92] and ligands of FKBP12.[43] The greatest advantage of a PMF approach is to avoid the alchemical decoupling (or annihilation) of the ligand with its surrounding. In that sense, the PMF method imitates roughly the bimolecular association/dissociation process whereby the ligand moves along a reaction path from the binding site to the bulk solution. This is particularly useful if a ligand is charged and its solvation free energy is very large.[55,63,92] The issue is well illustrated by the calculation of the binding free energy of

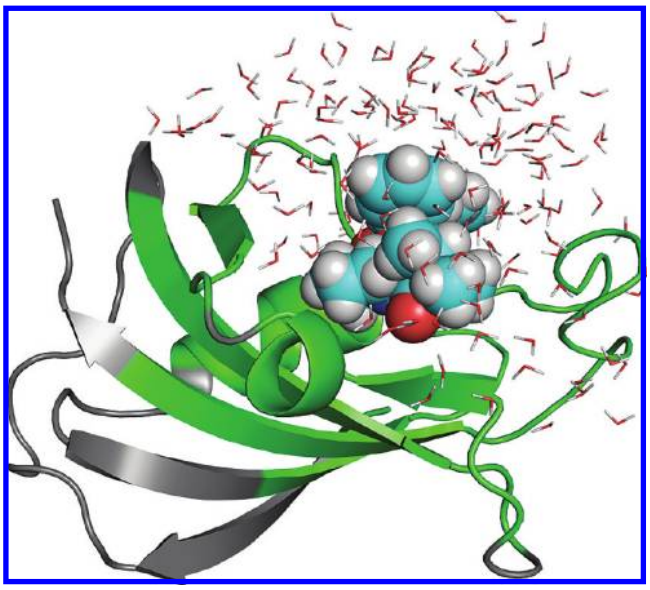

Figure 4. FKBP12 bound with ligand #8 studied previously.[40,42] The gray parts are treated as a mean-field approximation with generalized solvent boundary potential.[86] See ref 42 for computational details.

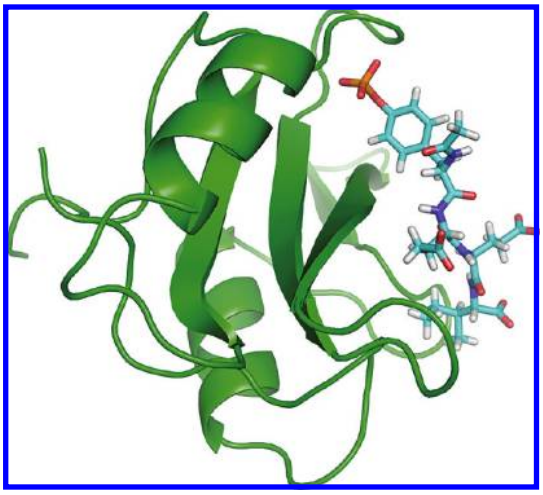

Figure 5. SH2 domain with bound peptide pYEEI. The system was simulated with PBC, and water is not shown for clarity. See ref 55 for computational details.

a phosphotyrosyl peptide, Ace-pYEEI, to the Src homology 2 (SH2) domain of Lck kinase.[55] Figure 5 shows the SH2 domain with bound phosphotyrosyl peptide. An alchemical FEP strategy is essentially impractical in this case because the solvation free energy of the ligand is on the order of −800 kcal/mol.[55] Even if the statistical uncertainty of a DDM calculation was only about 1% of the total solvation free energy, that would still translate into an error that is of the same order of magnitude as the quantity of interest itself. Using a restrained-PMF approach, the computed standard binding free energy is −8.8 kcal/mol, in good accord with the experimental value of −7.1 kcal/mol.[55]

PMF-based approaches become less practical if the binding site is deeply buried and a simple path for ligand association cannot be found. In this case, alchemical decoupling free energy techniques are more effective. For this reason, the majority of binding free energy calculations has been carried out using alchemical decoupling approaches. A large number of computations on a wide range of systems based on DAM have been reported. However, with a few exceptions, the significance of those results is unclear due to unresolved issues of convergence

TABLE 1: Binding Free Energy for the T4L Cavities

ligand	$\Delta\Delta G_{rep}$	$\Delta\Delta G_{dis}$	$\Delta\Delta G_{elec}$	$\Delta\Delta G^{\circ}_{t} + \Delta\Delta G_{r}$	ΔG°_{bind}	exptl
L99A Nonpolar Site						
benzene	−4.73	−7.43	0.69	5.42	−5.96	−5.19
phenol	−4.58	−8.19	4.35	7.55	−0.88	
L99A-M102Q Polar Site[a]						
benzene	−4.37	−8.53	0.51	6.78	−5.61	
phenol	−4.06	−9.74	−0.32	8.52	−5.64	−5.55

[a] Deng and Roux, unpublished result.

and reversibility. The study of FKPB12 ligands by Fujitani and co-workers based on DAM is of particular interest because extensive efforts were made to achieve convergence and reversibility in the calculations.[41] The connection to the standard state expressed in eq 3 was not specified, and the calculated binding free energies were empirically shifted by an offset constant for comparison with experiments. In a study of the same system, Pande and co-workers used an approach with some similarities to both DAM and DDM.[44] The simulation sampled a full decoupled state with the ligand wandering anywhere in the box, but only the configurations with the ligand within a small subvolume near the binding site were included in the free energy. This analysis made it possible to specify the connection with the standard state. Generally, however, DDM with translational restraints appears to be more advantageous than DAM in terms of its ability to yield well-converged and reversible results that can be related unambiguously to the standard state. Computations based on DDM have been carried out for a variety of systems. DDM has been used to characterize water occupancy in protein cavities,[56,96,97] and examine the binding of ligands to the T4L mutants,[34−39] FKBP12,[40−44] and cytochrome P450.[52,89] The latter system has been used to test a method combining DDM with a grand canonical Monte Carlo (GCMC) algorithm to account for the change in water occupancy in the buried cavity.[89]

B. Salient Contributions to the Binding Free Energy. While there is definitely room for improvement, it is fair to say that the results from many of the computational studies are in good agreement with experiments. It is clear that current computational approaches are sufficiently accurate to yield meaningful observations about the microscopic factors governing ligand binding. In the following, we review the most important conclusions drawn from the different computational studies.

1. Ligand−Receptor Interactions. One advantage of the alchemical approach based on DDM is the possibility to dissect the nonbonded interaction of the ligand with its surrounding (protein or solvent). Computing the reversible work in a step-by-step procedure makes it possible to separate the interaction free energy into repulsive, dispersive, and electrostatic components based on eq 7. The decomposition provides important insight into the microscopic forces driving ligand binding (keeping in mind the caveat that each term represents a conditional free energy contribution associated with the step-by-step reversible work to carry out the full process).

As an illustration, Tables 1 and 2 show the results for the T4L cavities and FKBP12, respectively, taken from our previous studies.[36,42] The free energy decomposition indicates that both the LJ-core repulsion and dispersion contribute favorably to the binding free energy. In the case of nonpolar ligand binding to T4L/L99A, the favorable contribution from the core repulsion is consistent with the idea that the protein provides an empty cavity that is preformed for ligand binding. Essentially, no reversible work is needed to insert the repulsive core of the ligand into the protein binding site. In the case of FKBP12, the repulsive component is still favorable but much smaller (Table 2). The main difference with T4L can be understood by the location of the binding site. The binding pocket of FKBP12 is at the protein surface, whereas the binding site in T4L mutants is located deep inside the protein core. For FKBP12, there is no preformed empty cavity to bind the ligand and the repulsive interaction does not make a strong contribution to the free energy.

TABLE 2: Binding Free Energy for Ligand 8 of FKBP12

$\Delta\Delta G_{rep}$	$\Delta\Delta G_{dis}$	$\Delta\Delta G_{elec}$	$\Delta\Delta G_{c}$	$\Delta\Delta G^{\circ}_{t}$	$\Delta\Delta G_{r}$	ΔG°_{bind}	exptl
−1.1	−21.1	−3.7	6.9	3.4	5.4	−10.2	−10.9

[a] Note that $\Delta\Delta G_{rep} + \Delta\Delta G_{dis} + \Delta\Delta G_{elec} = -25.9$ kcal/mol, and $\Delta\Delta G_{c} + \Delta\Delta G^{\circ}_{t} + \Delta\Delta G_{r} + 15.7$ kcal/mol.

Interestingly, van der Waals dispersive interaction makes a systematically favorable contribution to the binding free energy, both in the case of T4L and FKBP12; i.e., dispersion is more favorable when the ligand is in the binding site than when it is in the bulk solvent. The origin of the difference can be directly traced back to the number density of van der Waals interaction centers per unit volume surrounding the ligand. The density is invariably larger in a protein environment than in bulk water. A similar observation was made by Levy and co-workers in developing implicit solvent models.[98] The large dispersive contribution is correlated with the number of non-hydrogen atoms in the ligand. This observation agrees with the empirical observation that, barring large shape change, binding increased by adding heavy atoms to a ligand.[99] Furthermore, one may note that the contribution from the van der Waals dispersion to the binding free energy is almost equal to the average van der Waals dispersion of the ligand with its surrounding in the bound state minus the average van der Waals dispersion of the unbound ligand with the bulk solvent. This clarifies the circumstances for which the difference in simple end-point averages may be a useful indicator for assessing the relative affinity of series of ligands with a similar shape.

Ligand specificity is strongly affected by the polarity of the cavity.[39,88] This is well illustrated by contrasting the affinity of benzene and phenol for the T4L cavities engineered in the L99A and the M102Q-L99A mutants. As shown in Table 1, the free energy calculation is able to identify phenol as a binder of T4L/M102Q-L99A and as a nonbinder of T4L/L99A. In contrast, benzene binds to both the T4L/L99A and T4L/M102Q-L99A cavities. As the free energy decomposition shows, the nonpolar contribution for phenol is comparable to that of similar sized binding ligands, and it is only the high desolvation penalty from electrostatics that lowers its affinity for the L99A cavity. The M102Q-L99A double mutant provides a polar group that helps stabilize phenol electrostatically. This explains why phenol binds to the M102Q-L99A double mutant of T4L but not to the single mutant L99A.

2. Restriction of Ligand Motion. The loss of motional freedom of the ligand is intuitively associated with the concept of entropy. However, estimating changes in entropy accurately from simulations is very difficult,[28,100] and schemes to partition the entropy into particular motions involve further approximations.[101] Alternatively, the reversible work associated with the activation and release of restraining potentials can be utilized to estimate directly the loss of free energy associated with the loss of motional freedom upon binding. While such estimates depend on the specific pathway for decoupling the ligand, they can provide useful insight on the various contributions to the

binding free energy.[34,36] On the basis of an analysis with the restraining potential, the loss of translational freedom yields a free energy $\Delta\Delta G_t^\circ = -k_B T \ln(C^\circ \Delta V)$, where ΔV is an effective accessible volume for the center of mass of the ligand in the binding site. The microscopic volume ΔV is normally on the order of $\sim$1 Å^3, which yields the well-known standard state offset factor $-k_B T \ln(C^\circ \Delta V)$ of 4.4 kcal/mol. Thus, the reduction in translational freedom of the ligand almost invariably makes an unfavorable contribution to the binding free energy. On the basis of a similar analysis with the restraining potential, the loss of rotational freedom translates into a free energy $\Delta\Delta G_r = -k_B T \ln(\Delta\Omega/8\pi^2)$, where $\Delta\Omega$ is the magnitude of the orientational fluctuations of the ligand. Because the factor $\Delta\Omega/8\pi^2$ is typically much smaller than 1, the reduction in rotational freedom of the ligand always makes a considerable unfavorable contribution to binding free energy. Therefore, reduction in both translational and orientational freedom yields unfavorable contributions to the binding free energy. Interestingly, ΔV and $\Delta\Omega$ can be related to the dynamical fluctuations of the bound ligand. This can be exploited to clarify the significance of end-point approximations, in which the translational and orientational contributions are often estimated using a quasi-harmonic approximation.[57,70,71] For the T4L systems, roughly 5−9 kcal/mol arises from the loss of translational and rotational freedom of the ligand.[36] For the FKBP12 ligands, the loss of translation corresponds roughly to 3.5 kcal/mol and the loss of rotation corresponds roughly to 4.5−5.5 kcal/mol.[42]

Except for small and rigid ligands, one expects a reduction of internal ligand flexibility upon binding. The free energy associated with the loss of conformational freedom of the ligand, $\Delta\Delta G_c$, is the reversible work first to confine the ligand near its bound (reference) conformation with the restraining potential when it is in the bulk solvent (ΔG_c^{bulk}) and then to release it freely when it is in the binding site (ΔG_c^{site}).[36,42,55] The conformational restraining potential was written in terms of the rmsd of the ligand relative to its bound conformation, and the calculation proceeded via the PMF of the ligand in the bulk and in the binding site; see eqs 5 and 6. Upon binding to the SH2 domain, there is a loss of conformational freedom of the phosphotyrosyl-peptide ligand which gives rise to an unfavorable free energy of about 2 kcal/mol.[55] In the case of the flexible FK506-related ligands, the conformational free energy of the ligand varies from 1 to 7 kcal/mol;[42] it is 6.9 kcal/mol for ligand 8 (see Table 2). The utilization of a PMF to control the conformation of the ligand is useful to obtain more accurate estimates of the solvation free energy of the ligand, in cases when a direct decoupling scheme without conformational restraints would fail due to incomplete sampling.

In summary, the loss of motional freedom plays a key role in the resulting binding free energy. As a result, a large fraction of the total (favorable) contribution from the ligand−protein interaction is opposed by the (unfavorable) contribution arising from the loss of translational, orientational, and conformational freedom of the ligand. In the case of benzene shown in Table 1, about half of the total interaction free energy (−11.52 kcal/mol) is canceled by the loss of motional freedom upon binding (5.42 kcal/mol). In the case of the FK506-related ligand shown in Table 2, the interaction free energy of almost −26 kcal/mol is opposed by +15.7 kcal/mol of unfavorable contributions from the translational, orientational, and conformational loss, yielding a net binding free energy of only −10.2 kcal/mol. Even though FKBP12 itself is relatively rigid, about 60% of all the favorable interaction free energy of −26 kcal/mol is actually opposed by unfavorable contributions that are often ignored or discarded in simple scoring schemes. This has important implications for end-point methods[57,70,71] and for the parametrization of empirical scoring functions that are adjusted to estimate the binding free energy from one fixed configuration.[4]

3. Conformation of the Protein. Large conformational changes in the protein are expected to be an important component of the binding free energy. However, even seemingly minor conformation changes can have a considerable impact on the calculated binding free energy. A case in point is the orientation of the Val111 side chain in the T4L/L99A system. In the bound complexes with the small- and medium-size ligands (e.g., benzene, toluene), the protein structure is not strongly affected and the side chain of Val111 adopts a conformation similar to that in the apo structure. The calculated binding free energies for the small nonpolar ligands are generally in excellent agreement with experiments.[36,39] Difficulties arise in the case of larger ligands (e.g., indene, *n*-butylbenzene, isobutylbenzene, *o*-xylene, *p*-xylene). The side chain of Val111, which is in direct contact with the bound ligand, changes its rotameric states from a *t* conformation ($\chi_1 = 180°$) for the ligand-free or -bound state with small ligands, to a g^- conformation ($\chi_1 = -60°$) for the bound state with large ligands. The energy barrier around the χ_1 torsion is sufficient to prevent the Val111 side chain from rotating on the time scale of typical simulations, and the free energy for decoupling the ligand starting from the holo conformation is not correctly accounted for. As a result, calculated standard binding free energies are too favorable by 2−3 kcal/mol when the DDM calculations are initiated directly from the ligand-bound structure of the protein.[36] Additional difficulties were also noted in the case of indene, for which the side chain rotamer of Val111 in the crystallographic X-ray structure of the bound complex (PDB 183L) might be incorrect[36] (but see also refs 38 and 39). While the problems with Val111 can be resolved by adopting a deliberate sampling strategy involving the calculation of the PMF for the χ_1 dihedral angle of the side chain via umbrella sampling,[39] the quantitative impact of a single protein side chain on the calculated binding free energies is truly sobering. Whether a rotameric state is incorrect in the X-ray structure,[36] or insufficiently sampled in free energy simulations,[39] the consequences can amount to an error of several kilocalories per mole. This highlights the importance of having accurate structural models for the calculations, and the importance of sampling all of the relevant degrees of freedom.

4. Binding Site Hydration. While the solvent configurations are generated spontaneously by MD when the binding site is exposed to the bulk phase, sampling difficulties become particularly acute when a binding site is deeply buried and inaccessible. In this case, the exchange of water molecules with the bulk region may be very slow, and the accuracy of free energy perturbation (FEP) calculations based on unbiased MD trajectories is severely compromised. Several theoretical studies have specifically examined the thermodynamics stability of water molecules in buried protein cavities[56,96,97,102−104] and their impact on the thermodynamics of ligand binding.[105−107] Ligand binding often disrupts the hydration of the binding site.[108−110] A particularly challenging example is provided by cytochrome P450, a monooxygenase that oxidizes endogenous and xenobiotic substrates.[52,89] Simulation studies have shown that some conformational change is required to open up a channel, allowing exchange between the cavity and the bulk phase,[111−113] and about five water molecules must be expelled from the deeply buried cavity upon the binding of camphor. Such a change in hydration state must be captured in free energy calculations to

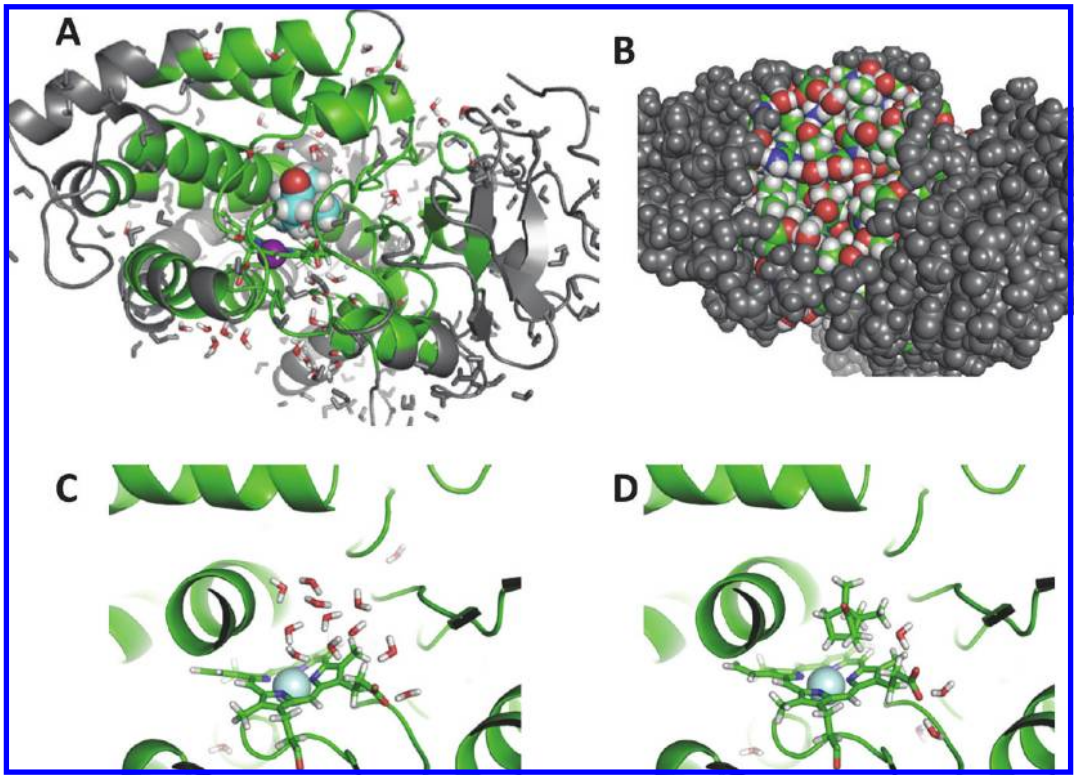

Figure 6. Binding site of cytochrome p450 with camphor. In part A is an overview of the simulation system with the buried binding site. In part B, the binding site is not visible with a space-filling representation. In part C, the site with no camphor is occupied by water, and in part D, camphor is bound. See ref 89 for theoretical formulation and computational details.

yield accurate and meaningful results. In a previous study, Wade and co-workers designed an alchemical transformation combining the annihilation of five water molecules together with insertion of the ligand in the cavity.[52] While the strategy can produce accurate results, it requires prior knowledge of the number of water molecules to annihilate in the cavity. Figure 6 shows the buried binding site in p450, and the impact on water occupancy upon camphor binding. The binding of camphor to cytochrome P450 was studied using a free energy method combining MD with a grand canonical Monte Carlo (GCMC) algorithm[90,114,115] to account for the change in water occupancy in the buried cavity during the alchemical free energy calculation.[89] In this context, the reversible free energy work to alchemically decouple the bound ligand can be expressed in terms of standard TI[89]

$$\Delta G = \int_0^1 \mathrm{d}\lambda \sum_n \mathscr{P}_n(\lambda)\left\langle \frac{\partial W}{\partial \lambda}\right\rangle_{\lambda,n} = \int_0^1 \mathrm{d}\lambda \left\langle\left\langle \frac{\partial W}{\partial \lambda}\right\rangle\right\rangle_{\lambda} \quad (10)$$

where $\mathscr{P}_n(\lambda)$ denotes the probability of n solvent molecules occupying the binding pocket, the bracket $\langle\cdots\rangle_{\lambda,n}$ means a constrained average with a fixed number of solvent molecules in the binding pocket. The double bracket implies a GCMC average over the number of solvent molecules. The calculated binding free energy for camphor to P450 is −8.25 kcal/mol, in good agreement with the experimental value of −7.75 kcal/mol.[52] In contrast, calculations with a fixed water configuration (holo or apo) give results with huge errors. When the calculation is carried out with the water molecules of the apo state, then the ligand has to push its way into the cavity and the resulting free energy is +12.60 kcal/mol, which is incorrect by almost 20 kcal/mol. On the other hand, when the calculation is carried out with the water molecules of the holo state, then there is a preformed empty cavity to receive the ligand and the resulting free energy is −14.25 kcal/mol, which is now too favorable by about 6 kcal/mol.

C. Putting Binding Free Energy Calculations to the Test. While many of the recent studies demonstrate that free energy simulations provide a satisfying physics-based perspective of ligand binding,[34,36−44,54,89,90] it is important to ascertain the accuracy and predictive power that can be achieved with current methods.

One particularly nice study recently combined both computations and experiments to examine the binding of a large number of ligands to the nonpolar L99A cavity of T4L.[39] In retrospective tests, computed standard binding free energies for 13 ligands had an rms error of about 1.9 kcal/mol relative to previously determined experimental values. In blind prospective tests, binding orientations and affinities were predicted for a set of five uncharacterized compounds identified by docking as putative binders. The calculations discriminated between several true binders and decoys, recognized the one nonbinder, accurately predicted ligand-bound orientations, correctly ranked the ligand-binding affinities, and quantitatively predicted binding free energies. The main conclusions from this extensive effort are that alchemical free energy methods are more accurate than docking, are able to distinguish binders from nonbinders, and can make successful predictions of bound orientations and binding affinities.[39] The authors also noted that accounting for protein conformational change via deliberate sampling of the rotameric state of the side chain of Val111 was important for accuracy.

There have also been some extensive studies of the ligand binding FKPB12.[40−42] This system is more difficult to manipulate than T4L, but it has the virtue of being representative of a real pharmaceutical target binding drug-like compounds. Computed binding free energies for a series of 8 FK506-related

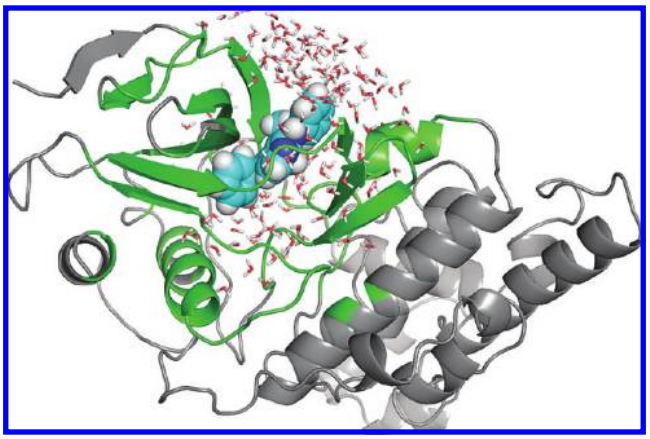

Figure 7. JNK kinase bound with ligand 19 from the OpenEye statistical assessment of the modeling of proteins and ligands (SAMPL) challenge.[116] The gray parts are treated as a mean-field approximation with generalized solvent boundary potential. The computations were carried out according to the same protocol presented in previous studies.[89]

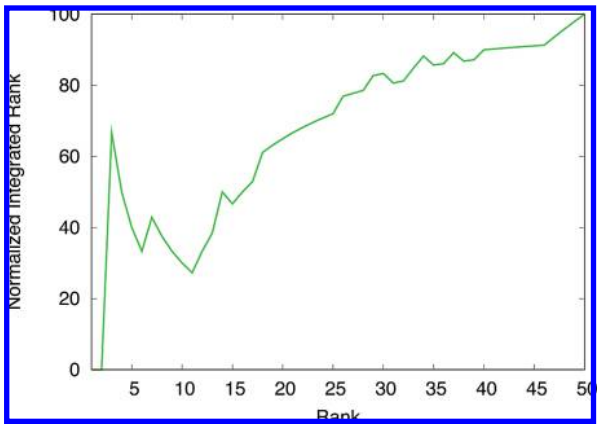

Figure 8. Cumulative percentage of enrichment $f(n)$ plotted as a function of ligand number n, given by $f(n) = 100 \sum_i^n \sum_j^n \delta[L^{exptl}(i), L^{calcd}(j)]/n$, $(1 \le n \le N)$, where $\delta[k, l]$ is a Kroenecker delta, $N = 50$, and $L^{exptl}(i)$ and $L^{calcd}(j)$ are the experimental and calculated lists of the N ligands ordered in decreasing affinity, respectively. The top five ligands from the computations are 40, **27**, **10**, 47, and 8; the experimental top five are **10**, 1, **27**, 11, and 38. The worse 10 ligands from the computations are 9, **43**, **51**, **6**, **5**, 21, **23**, 56, 31, and **17**; the experimental worst 10 are **51**, 4, **5**, **6**, **17**, **23**, 29, 30, 36, and **43**.

ligands display an average rms error of about 2−2.5 kcal/mol relative to previously determined experimental values.[40−42] Recent results on the same series with more extensive sampling show some improvement, with an rms error of about 1.4 kcal/mol.[44] Therefore, the results of retrospective tests on FKBP12 are now approaching an accuracy similar to that of T4L. It is worth noting that some of the results have been obtained from simulations of a fully solvated protein with PBC,[40,41,44] and some were based on a reduced system embedded in a GSBP mean-field surrounding.[42]

Recently, we have used free energy methods to participate in the OpenEye statistical assessment of the modeling of proteins and ligands (SAMPL) challenge.[116] In a blind test, the binding free energies of 50 neutral compounds to the JNK kinase were computed. The coordinates of all of the protein−ligand complexes were provided, so the purpose of this effort is to test the performance of standard binding free energy computations as a scoring strategy. Figure 7 shows the JNK kinase bound in complex with ligand 19. The computations were carried out according to a previously published protocol;[89] details are given in the caption. Most of the jobs were set up and submitted via automated scripts with default protonation states for the protein residues, trying to minimize human intervention as much as possible. Figure 8 shows the enrichment plot. Though there are clearly some false positives and false negatives, it shows a good separation of the binding and nonbinding compounds. Among the 50 compounds, the free energy computations correctly predicted 2 of the top 5 binders, as well as 6 of the 10 worst binders. Five out of nine nonactive compounds were also correctly identified. The computed binding free energies range from −16 to −3 kcal/mol, while the experimental values range from −8.6 to −5.5 kcal/mol. The error in some cases is considerable; e.g., the computed binding free energy for ligand 10 is −15.7 kcal/mol, whereas the experimental value is −8.7 kcal/mol. Thus, the binding affinity is strongly overestimated for a number of ligands, though this does not undermine the usefulness of the ranking itself. On the basis of the lessons from T4L, it is likely that this is caused by an inadequate sampling of the conformational changes in the protein.[36,39]

IV. Outlook

The theoretical foundations for the computation of standard binding free energies are well established. One can broadly choose between alchemical free energy perturbation methods, in which the ligand is decoupled progressively from its surrounding, and PMF-based methods, in which the ligand is pulled away from the binding site. The former are preferable in the case of ligands that bind to buried sites and cavities, while the latter are more advantageous for charged ligands that bind to the surface of a protein. For both of these computational approaches, various restraining potentials may be activated and released during the simulation to enhance and aid configurational sampling. While there is ample room for improvement, the recent studies demonstrate that standard binding free energy computations could be applied with success to increasingly challenging problems.

Computations of standard binding free energies provide an ultimate test, informing us about the ability of current computational methods to address issues of molecular recognition. In particular, the recent efforts have served to expose many of the approximations and shortcomings that often can be ignored or disregarded in calculations of relative free energies. On the other hand, it is important to point out that alchemical calculations of relative free energies, with their slightly more modest and practical goals, are already in a position to play an important role in drug design and lead optimization.[51]

The treatment of very large conformational changes in the receptor induced by ligand binding remains one of the biggest challenge in calculations of standard binding free energies. Striking examples are provided by the HIV protease,[117−119] or ligand-activated membrane receptors such as the ionotropic glutamate receptor (iGluR).[120,121] In principle, an adequate sampling of conformational changes could be achieved via exceedingly long unbiased MD simulations, though this approach rapidly becomes computationally prohibitive. At the present time, one must rely on a deliberate sampling strategy of the relevant degrees of freedom,[36,38,39,42,90,92] and thus on some prior identification of those degrees of freedom. In practice, this requires calculating a PMF along some chosen order parameter with biased simulations. In some cases, the order parameter may be relatively simple, such as the dihedral angle used to control the rotameric state of Val111 in the T4L/L99A system,[38] or could be relatively complicated, such as the combination of interdomain distances used to control the large conformational

change induced by the binding of glutamate to iGluR.[121] The rmsd relative to a reference conformation is a useful order parameter to control a flexible ligand.[36,42,55] Extending the rmsd to a protein macromolecule is possible[122,123] but limited. A general approach for mapping arbitrarily complex conformational changes onto a simple order parameter is lacking at present. This is currently an outstanding problem in computational biophysics.[124] It may be possible to make progress with the advance of methods aimed at finding the reaction path between the conformations of a macromolecular system.[124−127] The treatment of large protein conformation remains an outstanding issue in binding free energy calculations.

Apart from issues of conformational sampling, it is important to recall that the accuracy of computations is determined ultimately by the underlying atomic force field. Because they depend on a free energy difference between the bound and unbound states, computations can benefit from a cancelation of errors. For example, the standard binding free energy will remain roughly correct even if the van der Waals interaction of a ligand with the solvent and with the protein deviate systematically from the correct values. Nevertheless, dependable results will require that the force field represents the ligand, the protein, and the solvent as accurately as possible. Current biomolecular force fields with fixed partial charges, such as CHARMM,[22] AMBER,[23] OPLS,[24] and GAFF[25] with the AM1-BCC charge model,[128,129] provide atomic models able to yield solvation free energies that are in reasonably good agreement with experiments for amino acids and a wide range of small molecules.[26,27,29−31] Nonetheless, it is clear that it will be necessary to take the effect of induced electronic polarizability into account to achieve reliable results of high accuracy.[130] Additional complexities somewhat related to the force field include, among other factors, changes in protonation states upon binding,[131] and the existence of tautomers[132] will require special attention. In considering the accuracy of various force fields in the future, it will be helpful to draw more clearly the distinction between extremely rapid automatic assignment of force field parameters for an arbitrary ligand aimed at large database screening[25,128] and more computationally intensive physics-based approaches aimed at generating models of high accuracy.[133]

The computational methods are now approaching the point where making accurate prediction and addressing issues of molecular recognition will be achievable. While there is still much to be done, the methods are already bearing fruits and the path toward progress is very clear.

Acknowledgment. We would like to thank Hideaki Fujitani, Vijay Pande, Michael Shirts, David Mobley, Ken Dill, Devleena Shivakumar, Hyung-Jun Woo, and Jiayao Wang for their help. This work was supported by the National Science Foundation through Grant MCB-0630140. This work was supported by the U.S. Department of Energy (DOE), Office of Basic Energy Sciences, under Contract no. DE-AC02-06CH11357, and by Grant no. 2006-264-R2 from the Laboratory-Directed Research and Development (LDRD) program at Argonne National Laboratory.

References and Notes

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

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

(3) Garvie, C. W.; Wolberger, C. Recognition of specific DNA sequences. *Mol. Cell* **2001**, *8* (5), 937–946.

(4) Shoichet, B. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862–865.

(5) Shoichet, B.; Leach, A.; Kuntz, I. Ligand solvation in molecular docking. *Proteins* **1999**, *34*, 4–16.

(6) Carlson, H.; Masukawa, K.; Rubins, K.; Bushman, F.; Jorgensen, W.; Lins, R.; Briggs, J.; McCammon, J. Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* **2000**, *43*, 2100–2114.

(7) Schneider, G.; Bohm, H. J. Virtual screening and fast automated docking methods. *Drug Discovery Today* **2002**, *7* (1), 64–70.

(8) Price, D.; Jorgensen, W. Computational binding studies of human pp60c-src SH2 domain with a series of nonpeptide, phosphophenyl-containing ligands. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2067–2070.

(9) Wesolowski, S.; Jorgensen, W. Estimation of binding affinities for celecoxib analogues with COX-2 via Monte Carlo-extended linear response. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 267–270.

(10) Beveridge, D. L.; Dicapua, F. M. Free energy via molecular simulation - applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biol.* **1989**, *18*, 431–492.

(11) Straatsma, T.; McCammon, J. Computational alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407–435.

(12) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.

(13) Simonson, T.; Archontis, G.; Karplus, M. Free energy simulations come of age: Protein-ligand recognition. *Acc. Chem. Res.* **2002**, *35*, 430–437.

(14) Postma, J.; Berendsen, H. Thermodynamics of cavity formation in water - a molecular-dynamics study. *Symp. Chem. Soc.* **1982**, *17*, 55–67.

(15) Jorgensen, W.; Ravimohan, C. Monte-carlo simulation of differences in free-energies of hydration. *J. Chem. Phys.* **1985**, *83*, 3050–3054.

(16) Warshel, A. Dynamics of reactions in polar-solvents - Semiclassical trajectory studies of electron-transfer and proton-transfer reactions. *J. Phys. Chem.* **1982**, *86*, 2218–2224.

(17) Tember, B.; McCammon, A. Ligand receptor interactions. *Comput. Chem.* **1984**, *8*, 281–283.

(18) Warshel, A. Simulating the Energetics and Dynamics of Enzymatic Reactions. *Pontificiae Academiae Scientiarum Scripta Varia* **1982**, *55*, 59–81.

(19) Lybrand, T.; McCammon, A.; Wipff, G. Theoretical calculation of relative binding-affinity in host guest systems. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 833–835.

(20) Wong, C. F.; McCammon, J. A. Dynamics and design of enzymes and inhibitors. *J. Am. Chem. Soc.* **1986**, *108*, 3830–3832.

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

(22) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dumbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

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

(24) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom froce field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(25) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(26) 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.

(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) Wan, S.; Stote, R. H.; Karplus, M. Calculation of the aqueous solvation energy and entropy, as well as free energy, of simple polar solutes. *J. Chem. Phys.* **2004**, *121*, 9539–9548.

(29) Mobley, D.; Dumont, E.; Chodera, J.; Dill, K. Comparison of charge models for fixed-charge force fields: Small-molecule hydration free energies in explicit solvent. *J. Phys. Chem. B* **2007**, *111*, 2242–2254.

(30) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small molecule hydration free energies in explicit solvent: An extensive test of fixed-charge atomistic simulations. *J. Phys. Chem. B*, in press.

(31) Shivakumar, D.; Deng, Y.; Roux, B. Calculation of absolute solvation free energies of small molecules using explicit and implicit solvent model. *J. Chem. Theory Comput.*, in press.

(32) Nicholls, A.; Mobley, D.; Guthrie, J.; Chodera, J.; Bayly, C.; Cooper, M.; Pande, V. Predicting small-molecule solvation free energies: an informal blind test for computational chemistry. *J. Med. Chem.* **2008**, *51*, 769–779.

(33) Maccallum, J. L.; Tieleman, D. P. Calculation of the water-cyclohexane transfer free energies of neutral amino acid side-chain analogs using the OPLS all-atom force field. *J. Comput. Chem.* **2003**, *24*, 1930–1935.

(34) 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.

(35) 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.

(36) Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255–1273.

(37) 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.* **2006**, *125*, 084902.

(38) Mobley, D. L.; Chodera, J. D.; Dill, K. A. Confine-and-release method: Obtaining correct binding free energies in the presence of protein conformational change. *J. Chem. Theory Comput.* **2007**, *3*, 1231–1235.

(39) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.* **2007**, *371*, 1118–1134.

(40) Shirts, M. R. Calculating precise and accurate free energies in biomolecular systems. Ph.D. thesis, Stanford University, 2005.

(41) 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.

(42) Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular dynamics simulations with restraint potentials. *Biophys. J.* **2006**, *91*, 2798–2814.

(43) Lee, M. S.; Olson, M. A. Calculation of absolute portein-ligand binding affinity using path and endpoint approaches. *Biophys. J.* **2006**, *90*, 864–877.

(44) Jayachandran, G.; Shirts, M. R.; Park, S.; Pande, V. S. Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics. *J. Chem. Phys.* **2006**, *125*, 084901.

(45) Rajamani, R.; Good, A. C. Ranking poses in structure-based lead discovery and optimization: current trends in scoring function development. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 308–15.

(46) Seifert, M. H. J.; Kraus, J.; Kramer, B. Virtual high-throughput screening of molecular databases. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 298–307.

(47) Kontoyianni, M.; Madhav, P.; Suchanek, E.; Seibel, W. Theoretical and practical considerations in virtual screening: a beaten field. *Curr. Med. Chem.* **2008**, *15*, 107–16.

(48) Kim, J.; Hamilton, A.; Bailey, C.; Domaoal, R.; Domoal, R.; Wang, L.; Anderson, K.; Jorgensen, W. FEP-guided selection of bicyclic heterocycles in lead optimization for non-nucleoside inhibitors of HIV-1 reverse transcriptase. *J. Am. Chem. Soc.* **2006**, *128*, 15372–15373.

(49) Thakur, V.; Kim, J.; Hamilton, A.; Bailey, C.; Domaoal, R.; Wang, L.; Anderson, K.; Jorgensen, W. Optimization of pyrimidinyl- and triazinyl-amines as non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5664–5667.

(50) Zeevaart, J.; Wang, L.; Thakur, V.; Leung, C.; Tirado-Rives, J.; Bailey, C.; Domaoal, R.; Anderson, K.; Jorgensen, W. Optimization of azoles as anti-human immunodeficiency virus agents guided by free-energy calculations. *J. Am. Chem. Soc.* **2008**, *130*, 9492–9499.

(51) Jorgensen, W.; Thomas, L. T. Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *J. Chem. Theory Comput.* **2008**, *4*, 869–876.

(52) Helms, V.; Wade, R. C. Computational alchemy to calculate absolute protein-ligand binding free energy. *J. Am. Chem. Soc.* **1998**, *120*, 2710–2713.

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

(54) 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.

(55) 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.

(56) 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.

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

(58) Bjerrum, N. Untersuchungen über Ionenassoziation. I. der Einfluss der Ionenassoziation auf die Akativität der Ionen bei Mittleren Assoziationsgraden. *K. Dan. Vidensk. Selsk.* **1926**, *7*, 1–48.

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

(60) Hunenberger, P.; van Gunsteren, W. F. Experimental and theoretical approach to hydrogen-bonded diastereomeric interactions in a model complex. *J. Am. Chem. Soc.* **1997**, *119*, 7533–7544.

(61) Jorgensen, W. L. Free energy calculations: a breakthrough for modeling organic chemistry in solution. *Acc. Chem. Res.* **1989**, *22*, 184–189.

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

(63) 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.

(64) Beutler, T. C.; van Gunsteren, W. F. Molecular dynamics free energy calculation in four dimensions. *J. Chem. Phys.* **1994**, *101*, 1417–1422.

(65) Pomès, R.; Eisenmesser, E.; Post, C.; Roux, B. Calculating excess chemical potentials using dynamic simulations in the fourth dimension. *J. Chem. Phys.* **1999**, *111*, 3387–3395.

(66) Rodinger, T.; Howell, P. L.; Pomès, R. Absolute free energy calculations by thermodynamic integration in four spatial dimensions. *J. Chem. Phys.* **2005**, *123*, 34104.

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

(68) Beveridge, D. L.; DiCapua, F. M. Free energy via molecular simulation: Applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biomol. Struct.* **1989**, *18*, 431–492.

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

(70) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889–897.

(71) 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.

(72) Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. Ligand binding affinities from MD simulations. *Acc. Chem. Res.* **2002**, *35*, 358–365.

(73) Sham, Y. Y.; Chu, Z. T.; Tao, H.; Warshel, A. Examining methods for calculations of binding free energies: LRA, LIE, PDLD-LRA, and PDLD/S-LRA calculations of ligands binding to an HIV protease. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 393–407.

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

(75) Axelsen, P.; Li, D. Improved convergence in dual-topology free energy calculations through use of harmonic restraints. *J. Comput. Chem.* **1998**, *19*, 1278–1283.

(76) Leitgeb, M.; Schröder, C.; Boresch, S. Alchemical free energy calculations and multiple conformational substates. *J. Chem. Phys.* **2005**, *122*, 84109.

(77) van Duijneveldt, S.; Frenkel, D. Computer-simulation study of free-energy barriers in crystal nucleation. *J. Chem. Phys.* **1992**, *96*, 4655–4668.

(78) Patey, G. N.; Valleau, J. P. A Monte Carlo method for obtaining the interionic potential of mean force in ionic solution. *J. Chem. Phys.* **1975**, *63*, 2334–2339.

(79) 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.

(80) Chandler, D.; Pratt, L. Statistical-mechanics of chemical-equilibria and intramolecular structures of nonrigid molecules in condensed phases. *J. Chem. Phys.* **1976**, *65*, 2925–2940.

(81) Roux, B.; Simonson, T. Implicit solvent models. *Biophys. Chem.* **1999**, *78*, 1–20.

(82) Roux, B.; Andersen, O.; Allen, T. Comment on "Free energy simulations of single and double ion occupancy in gramicidin A". *J. Chem. Phys.* **2008**, *128*, 227101.

(83) Roux, B.; Simonson, T. Implicit solvent models. *Biophys. Chem.* **1999**, *78*, 1–20.

(84) King, G.; Lee, F.; Warshel, A. Microscopic simulation of macroscopic dielectric constants of solvated proteins. *J. Chem. Phys.* **1991**, *95*, 4365–4377.

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

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

(87) 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.

(88) 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.

(89) Deng, Y.; Roux, B. Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations. *J. Chem. Phys.* **2008**, *128*, 115103.

(90) 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.

(91) Chipot, C.; Rozanska, X.; Dixit, S. Can free energy calculations be fast and accurate at the same time? Binding of low-affinity, non-peptide inhibitors to the SH2 domain of the src protein. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 765–770.

(92) Gan, W.; Roux, B. Binding specificity of SH2 domains: Insight from free energy simulations. *Proteins: Struct., Funct., Bioinf.*, in press.

(93) 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.

(94) Kissinger, C. R.; Parge, H. E.; Knighton, D. R.; Lewis, C. T.; Pelletier, L. A.; Tempczyk, A.; Kalish, V. J.; Tucker, K. D.; Showalter, R. E.; Moomaw, E. W.; Gastinel, L. N.; Habuka, N.; Chen, X.; Maldonado, F.; Barker, J. E.; Bacquet, R.; Villafranca, J. E. Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex. *Nature* **1995**, *378*, 641–644.

(95) Griffith, J. P.; Kim, J. L.; Kim, E. E.; Sintchak, M. D.; Thomson, J. A.; Fitzgibbon, M. J.; Fleming, M. A.; Caron, P. R.; Hsiao, K.; Navia, M. A. X-ray structure of calcineurin inhibited by the immunophilin-immunosuppressant FKBP12-FK506 complex. *Cell* **1995**, *82*, 507–522.

(96) 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.

(97) Olano, L. R.; Rick, S. W. Hydration free energies and entropies for water in protein interiors. *J. Am. Chem. Soc.* **2004**, *126*, 7991–8000.

(98) 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.

(99) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. The maximal affinity of ligands. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9997–10002.

(100) Meirovitch, H. Recent developments in methodologies for calculating the entropy and free energy of biological systems by computer simulation. *Curr. Opin. Struct. Biol.* **2007**, *17*, 181–186.

(101) Chang, C. A.; Chen, W.; Gilson, M. K. Ligand configurational entropy and protein binding. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1534–1539.

(102) Hummer, G.; Szabo, A. Free energy surfaces from single-molecule force spectroscopy. *Acc. Chem. Res.* **2005**, *38*, 504–513.

(103) Collins, M. D.; Hummer, G.; Quillin, M. L.; Matthews, B. W.; Gruner, S. M. Cooperative water filling of a nonpolar protein cavity observed by high-pressure crystallography and simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16668–16671.

(104) Yin, H.; Hummer, G.; Rasaiah, J. C. Metastable water clusters in the nonpolar cavities of the thermostable protein tetrabrachion. *J. Am. Chem. Soc.* **2007**, *129*, 7369–7377.

(105) Li, Z.; Lazaridis, T. The effect of water displacement on binding thermodynamics: Concanavalin A. *J. Phys. Chem. B* **2005**, *109*, 662–670.

(106) Lu, Y.; Yang, C.-Y.; Wang, S. Binding free energy contribution of interfacial waters in HIV-1 protease/inhibitor complexes. *J. Am. Chem. Soc.* **2006**, *128*, 11830–11839.

(107) Li, Z.; Lazaridis, T. Water at biomolecular binding interfaces. *Phys. Chem. Chem. Phys.* **2007**, *9*, 573–581.

(108) Helms, V.; Wade, R. C. Hydration energy landscape of the active site cavity in cytochrome P450cam. *Proteins: Struct., Funct., Genet.* **1998**, *32*, 381–396.

(109) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. Classification of water molecules in protein binding sites. *J. Am. Chem. Soc.* **2007**, *129*, 2577–2587.

(110) Mancera, R. L. Molecular modeling of hydration in drug design. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 275–280.

(111) Oprea, T. I.; Hummer, G.; García, A. E. Identification of a functional water channel in cytochrome P450 enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2133–2138.

(112) Cojocaru, V.; Winn, P. J.; Wade, R. C. The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta* **2007**, *1770*, 390–401.

(113) García, A. E.; Hummer, G. Water penetration and escape in proteins. *Proteins: Struct., Funct., Genet.* **2000**, *38*, 261–272.

(114) Resat, H.; Mezei, M. Grand canonical ensemble Monte Carlo simulation of the dCpG/proflavine crystal hydrate. *Biophys. J.* **1996**, *71*, 1179–1190.

(115) 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.

(116) http://sampl.eyesopen.com/.

(117) Hornak, V.; Okur, A.; Rizzo, R.; Simmerling, C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 915–920.

(118) Hornak, V.; Okur, A.; Rizzo, R.; Simmerling, C. HIV-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state. *J. Am. Chem. Soc.* **2006**, *128*, 2812–2813.

(119) Hornak, V.; Simmerling, C. Targeting structural flexibility in HIV-1 protease inhibitor binding. *Drug Discovery Today* **2007**, *12*, 132–138.

(120) Armstrong, N.; Mayer, M.; Gouaux, E. Tuning activation of the AMPA-sensitive GluR2 ion channel by genetic adjustment of agonist-induced conformational changes. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5736–5741.

(121) Lau, A.; Roux, B. The free energy landscapes governing conformational changes in a glutamate receptor ligand-binding domain. *Structure* **2007**, *15*, 1203–1214.

(122) Banavali, N.; Roux, B. Free energy landscape of A-DNA to B-DNA conversion in aqueous solution. *J. Am. Chem. Soc.* **2005**, *127*, 6866–6876.

(123) Banavali, N.; Roux, B. The N-terminal end of the catalytic domain of SRC kinase Hck is a conformational switch implicated in long-range allosteric regulation. *Structure* **2005**, *13*, 1715–1723.

(124) Elber, R. Long-timescale simulation methods. *Curr. Opin. Struct. Biol.* **2005**, *15*, 151–156.

(125) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String method in collective variables: minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **2006**, *125*, 24106.

(126) Christen, M.; van Gunsteren, W. F. On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: A review. *J. Comput. Chem.* **2008**, *29*, 157–166.

(127) Pan, A.; Sezer, D.; Roux, B. Finding transition pathways using the string method with swarms of trajectories. *J. Phys. Chem. B* **2008**, *112*, 3432–3440.

(128) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.

(129) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(130) Jiao, D.; Golubkov, P.; Darden, T.; Ren, P. Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6290–6295.

(131) 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.

(132) Pospisil, P.; Ballmer, P.; Scapozza, L.; Folkers, G. Tautomerism in computer-aided drug design. *J. Recept. Signal Transduction Res.* **2003**, *23*, 361–371.

(133) Anisimov, V.; Lamoureux, G.; Vorobyov, I.; Huang, N.; Roux, B.; MacKerell, A. Determination of electrostatic parameters for a polarizable force field based on the classical Drude oscillator. *J. Chem. Theory Comput.* **2005**, *1*, 153–168.

JP807701H