JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

Calculation of the Absolute Free Energy of Binding and Related Entropies with the HSMD-TI Method: The FKBP12-L8 Complex

Ignacio J. General, Ralitsa Dragomirova, and Hagai Meirovitch*

Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, 3059 BST3, Pittsburgh, Pennsylvania 15260, United States

ABSTRACT: The hypothetical scanning molecular dynamics (HSMD) method is used here for calculating the absolute free energy of binding, ΔA^0, of the complex of the protein FKBP12 with the ligand SB2 (also denoted L8)—a system that has been studied previously for comparing the performance of different methods. Our preliminary study suggests that considering long-range electrostatics is imperative even for a hydrophobic ligand such as L8. Therefore, the system is modeled by the AMBER force field using Particle Mesh Ewald (PME). HSMD consists of three stages applied to both the ligand−solvent and ligand−protein systems. (1) A small set of system configurations (frames) is extracted from an MD trajectory. (2) The entropy of the ligand in each frame is calculated by a reconstruction procedure. (3) The contribution of water and protein to ΔA^0 is calculated for each frame by gradually increasing the ligand-environment interactions from zero to their full value using thermodynamic integration (TI). Unlike the conventional methods, the structure of the ligand is kept fixed during TI, and HSMD is thus free from the end-point problem encountered with the double annihilation method (DAM). Therefore, the need for applying restraints is avoided. Furthermore, unlike the conventional methods, the entropy of the ligand and water is obtained directly as a byproduct of the simulation. In this paper, in addition to the difference in the internal entropies of the ligand in the two environments, we calculate for the first time the *external* entropy of the ligand, which provides a measure for the size of the active site. We obtain $\Delta A^0 = -10.7 \pm 1.0$ as compared to the experimental values −10.9 and −10.6 kcal/mol. However, a protein/water system treated by periodic boundary conditions grows significantly with increasing protein size, and the computation of ΔA^0 would become expensive by all methods. Therefore, we also apply HSMD to FKBP12-L8 described by the GSBP/SSBP model of Roux's group (implemented in the software CHARMM) where only part of the protein and water around the active site are considered and long-range electrostatic effects are taken into account. For comparison, this model was also treated by the double decoupling method (DDM). The two methods have led to comparable results for ΔA^0, which are somewhat lower than the experimental value. The ligand was found to be more confined in the active site described by GSBP/SSBP than by PME, where its entropy in solvent is larger than in the active site by 1.7 and by 5.5 kcal/mol, respectively.

I. INTRODUCTION

The central aim of this paper is to further develop our hypothetical scanning molecular dynamics (HSMD) method for calculating the absolute free energy of binding. To emphasize the significance of our method, we first discuss the importance of the problem and the properties of existing techniques.

The development of simulation methods for calculating the affinity (free energy) of molecular binding is important for both academic and practical reasons as tools for elucidating the mechanisms of complex biological processes, such as the action of hormones, the recognition of antigens by the immune system, the catalysis of chemical reactions by enzymes, and the action of drugs. Therefore, such methods can be used in rational drug design, leading to various therapeutic means. It is important in particular to devise highly accurate methods for calculating the free energy of binding based on detailed molecular interactions and rigorous statistical mechanics; such methods are required in the refinement stage of screening procedures based on simplified (fast) scoring functions. Of special interest in this category are methods for calculating the *absolute* (standard) free energy of binding of a ligand to a protein.

A great deal of work has been done in this direction, where various techniques have been developed and applied to a wide range of problems (see refs 1−22 and references cited therein). A central *rigorous* approach is based on thermodynamic cycles, where the interactions between the ligand and its environment are decreased to zero in both, the active site and the bulk solution (water), using thermodynamic integration (TI) or free energy perturbation (FEP) procedures. However, this approach, called the double annihilation method (DAM),[1−5] is not straightforward since during the final stages of TI the ligand leaves the active site and starts wandering within the volume, which makes it extremely difficult to obtain converged results; also in most of the DAM applications, the effect of the standard state was not applied (see ref 12 as a recent example). Still, this method has been used successfully in recent years,[8,12] where in a later study[13] the effect of the standard state has been considered.

The end-point problem has been rigorously solved by adding restraints which hold the ligand in the active site; the corresponding bias introduced is later removed by releasing the restraints. Because of the additional integration step involved, this procedure is sometimes called the double decoupling method (DDM).[3−5] DDM has been developed systematically in the past

Received: July 13, 2011
Published: October 27, 2011

15 years,[7,9,15] where various implementation issues have been improved and a large number of complexes have been successfully studied (see review in ref 9). Historically, a (translational) restraining potential was introduced rigorously first by Hermans and Shankar[1] and extended later to a more complex system by Roux's group.[18] Hermans and Wang were also the first to introduce an angular restraint,[19] and more restraints were added later by Roux and Deng[20,9] (see also refs 4 and 5).

Therefore, it is desirable to enrich the arsenal of methods available in this field by developing *alternative* techniques that are free from the problems discussed above and provide, in addition to the binding free energy, other information, such as the entropy. A step in this direction has been made in our recent publication, where HSMD has been extended for the first time to a binding problem, treating the avidin−biotin complex.[22] HSMD is an *exact* technique for calculating the *absolute* entropy and free energy, which has been developed in our group over the past several years and has been tested systematically on models of increasing complexity, liquid argon,[23] TIP3P water,[23] self-avoiding walks,[24] and peptides,[25] where HSMD results were compared to those obtained by well established techniques (more accurately, HSMC was used in these calculations where Monte Carlo replaces MD). While HSMD is a general technique, which is applicable to any system (including fluids), for a protein in explicit water, the contribution of water to the free energy is calculated much more efficiently by a TI (or FEP) procedure that has been incorporated within the framework of HSMD; the combined method—called HSMD-TI (or HSMD-FEP)—has been applied successfully to several mobile loops in proteins.[26−30]

HSMD-TI consists of three stages applied to both the ligand−solvent and ligand−protein systems. (1) A small set of system configurations (frames) is extracted from an MD trajectory. (2) The ligand's entropy in each frame is calculated by a reconstruction procedure. (3) The contribution of water and protein to the binding free energy is calculated for each frame by gradually increasing the ligand−environment interactions from zero to their full value using TI. Because during TI the structure of the ligand is kept fixed, the end-point problem (of DAM) does not exist and the need for applying restraints (DDM) is avoided. Furthermore, unlike DAM and DDM, several partial entropies of the system are obtained directly as byproducts of the simulation, and the method is, in particular, suitable to handle large flexible ligands. In this study, we calculate for the first time the external entropy of the ligand in the active site, which constitutes a measure for the global movement of the bound ligand, thereby providing estimation for the size of the active site. All of these entropies provide microscopic insights into the binding mechanism and are thus important in rational drug design.

In our initial avidin−biotin study[22] (using TINKER[31]), we were mainly interested in checking the performance of HSMD-TI; therefore, the modeling has been somewhat limited. In particular, only a spherical part of the tetrameric avidin, consisting of residues within a distance of 18 Å from the center of the binding site, were considered, and in most of the calculations, this part (the template) was kept fixed during the simulations. Also, a comparable water sphere was used. Finally, long-range electrostatic effects were ignored. Still, including the contribution of a mobile loop has led to the experimental absolute free energy of binding.

The aim of this paper is to develop HSMD-TI further by applying it to the protein FKBP12 bound to the ligand SB3 (see below), where this system is described by models of various complexities. FKBP12 is a small protein (107 residues), which

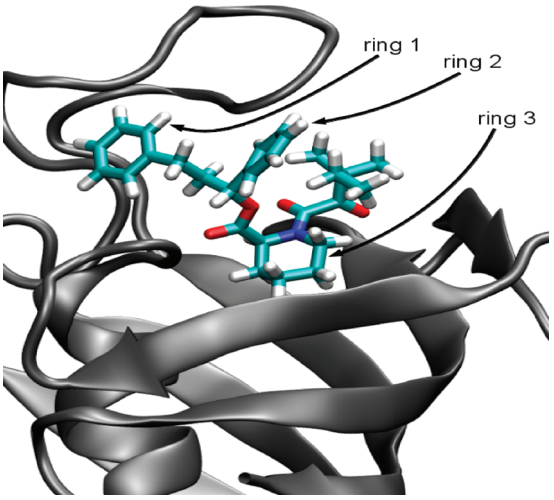

Figure 1. Illustration of the FKBP12−L8 complex. The three rings of the ligand are numbered; ring 3 is positioned at the bottom of the active site.

catalyzes the cis−trans isomerization of peptidyl−prolyl bonds.[32] FK506 is a key drug used for immunosuppression in organ transplant. It binds strongly to FKBP12,[33] and the FKBP12/FK506 complex, in turn, binds and inhibits calcineurin, thus blocking the signal transduction pathway for the activation of T-cells.[34,35]

Crystal structures of FKBP12 in complex with several ligands are available,[36−38] and the binding constants of the FK506-related ligands with FKBP12 have been experimentally determined.[37] Therefore, this system has served as a rich platform to test and validate different computational strategies for estimating binding free energies.[8,11,13,39−42] In particular, a set of eight ligands complexed with FKBP12 (out of the 27 studied in ref 37) have become the target for absolute free energy calculations by several groups who applied different methods. Wang et al.[40] have used DDM with their models—Spherical Solvent Boundary Potential (SSBP[43]) and Generalized Solvent Boundary Potential (GSBP[44]). GSBP is an example of what we call in this paper a "model of partial structure", i.e., a model where only part of the protein close to the active site is considered, and this part is covered by a relatively small sphere of water. Notice, however, that both GSBP and SSBP take into account long-range electrostatics (reaction field). Pande's group and Fujitani et al. used DAM,[8,11,13,39] where the (entire) complex and the ligand in the bulk were immersed in a box of explicit water and long-range electrostatic effects were taken into account by periodic boundary conditions with Particle Mesh Ewald (PME).[45]

We have decided to study the SB3 ligand, which is also called L8 (68 atoms, Figure 1), because in the set of eight ligands, it is the smallest for which the crystal structure of the complex is known. As discussed later in section III.1, in preliminary studies, L8−FKBP12 was described by a model of partial structure handled by the AMBER 10 package[46] with an empirical potential energy—the AMBER force field,[47] TIP3P water,[48] and the General AMBER Force Field (GAFF)[49] for the ligand. Because the ligand is hydrophobic, we had reasons to believe that applying long-range electrostatic interactions would not be necessary (see section III.1); however, the results were found to be unsatisfactory. Therefore, we have decided to treat the FKBP−L8 system (from now on we omit "12") with two models that consider long-range electrostatics—PME implemented in the software package

AMBER[46] and the SSBP/GSBP[43,44] modeling, which is incorporated within the program CHARMM.[50]

II. THEORY AND METHODOLOGY

II.1. Theory of Binding. Imagine a *dilute* solution of a protein (P) and a ligand (L) in a volume V in equilibrium with their complex (PL), P + L ↔ PL. The equilibrium constant, K_b, is defined by the equilibrium concentrations (denoted [ ]) of these components, $K_b = [LP]/[P][L]$, where K_b leads to the absolute (standard) free energy of binding, ΔA^0. Since our system is defined in the *NVT* ensemble, we obtain[4,5]

$$\Delta A^0 = -k_B T \ln \frac{V Z_{PL,N} Z_{0,N}}{V^0 Z_{P,N} Z_{L,N}} = -k_B T \ln \frac{\overline{Z}_{PL,N} Z_{0,N}}{8\pi^2 V^0 \overline{Z}_{P,N} \overline{Z}_{L,N}} \quad (1)$$

where T is the absolute temperature, k_B is the Boltzmann constant, and N stands for the number of solvent molecules (water). $Z = V8\pi^2\overline{Z}$, where $\overline{Z}_{PL,N}$, $\overline{Z}_{P,N}$, and $\overline{Z}_{L,N}$ are the conformational partition functions of the complex, protein, and ligand all in water; $Z_{0,N}$ is the partition function of N water molecules in the volume. The bar means that P and L in $\overline{Z}$ are defined by *internal coordinates*, where the integration ($V8\pi^2$) over the external coordinates (e.g., a reference atom and three Euler angles defined by two more atoms) has already been carried out, where V is the system's volume and $V^0 = 1660$ Å^3 is the standard volume. ΔA^0 is expressed in terms of $\overline{Z}$ because $\overline{Z}$ does not depend explicitly on V, and with HSMD-TI we mainly calculate internal entropies and free energies. Notice that the ligand moves in the active site; i.e., $\overline{Z}_{PL,N}$ includes a localized ligand partition function where its coordinates can also be divided into internal and external, where the contribution of the latter will be calculated by HSMD (rather than analytically as in the solvent). ΔA^0 is expressed in terms of configurational (Helmholtz) free energies, $F = -k_B T \ln \overline{Z}$ (and $F_{0,N} = -k_B T \ln Z_{0,N}$) and an additional term

$$\Delta A^0 = (F_{PL,N} - F_{P,N}) - (F_{L,N} - F_{0,N}) + k_B T \ln(8\pi^2 V^0) = \Delta F_p - \Delta F_{sol} + k_B T \ln(8\pi^2 V^0) \quad (2)$$

ΔF_p and ΔF_{sol} are free energy differences defined for the protein and solvent environments, respectively, which are calculated by HSMD-TI. Also, the absolute Gibbs free energy $\Delta G^0 \sim \Delta A^0$, since $\Delta G^0 = \Delta A^0 + P^0 \Delta \overline{V}_{PL}$ where $P^0 \Delta \overline{V}_{PL}$ is small and can be neglected.[4,5]

II.2. The Philosophy of HSMD. The HSMD method (as well as HSMC) enables one to calculate the *absolute* entropy and free energy from a sample generated by MD,[51,52] MC,[53] or any other simulation technique. HSMD is based on the fact that a system configuration can, in principle, be generated *exactly* also by a step-by-step (growth) procedure, where particles are added gradually to an initially empty volume using transition probabilities (TPs). A trivial example is an *ideal chain* of N bonds on a square lattice, i.e., a chain without the excluded volume interaction. An ideal chain can be simulated *exactly* by the dynamical MC method where the entropy, however, is unknown; alternatively, a chain configuration, i, can be constructed (from nothing) as a random walk where a bond's direction (out of four) is chosen with TP = 1/4 and added at each step. Here, the Boltzmann probability is known, $P_i^B = (1/4)^N$, and thus the entropy is known as well ($S = -k_B \ln P_i^B$). Clearly, large samples constructed by MC or as random walks are equivalent in the sense that they lead to the same thermodynamic averages and fluctuations.

While an exact growth procedure (called the exact scanning method[54,30]) can be defined for any system (e.g., water, protein in water), application of such a procedure will in general be very inefficient. However, relying on the equivalence among exact simulation methods mentioned above, one can assume that a given MD (MC) sample has rather been generated by the exact scanning method, which enables one to reconstruct each conformation i by calculating the TP densities that *hypothetically* were used to create it step-by-step. Application of this *hypothetical scanning* (HS) procedure is much more efficient than the (direct) scanning method.

In practice, the product of the TPs leads to an approximation, P_i, for the correct Boltzmann probability P_i^B, where from P_i various free energy functionals (F) can be defined. However, no inherent approximation is applied in the calculation of the TPs. That is, all of the system interactions are taken into account, and the only approximation involved is due to insufficient MD sampling for their calculation. In this respect, HSMD is considered to be an exact method.[23] More specifically, we calculate $S(n_f) = -k_B \Sigma P_i^B \ln P_i(n_f)$, which constitutes a rigorous upper bound for the correct entropy S; the larger the sample size, n_f, the better is P_i and the smaller is $S(n_f)$. Thus, one can follow the convergence of $S(n_f)$ to S (from above) as n_f is increased until the increase in $S(n_f)$ becomes smaller than the statistical error; i.e., $S(n_f)$ is *exact* within this error (see detailed discussions related to eqs 4, 6, and 7 below). In practice, this convergence will not be attained satisfactorily (even though in a recent publication, results for $S(n_f)$ of biotin were found to be close to convergence). However, we are not interested in $S(n_f)$ itself but in the entropy difference $[\Delta S(n_f)]$, where in the application for binding, it is between the ligand's entropy in solvent and in the active site. Typically, $\Delta S(n_f)$ converges very rapidly as a function of n_f (due to cancellation of errors) with a statistical error that is smaller than other errors involved in the calculation of ΔA^0 (see refs 22−30 and a note in ref 55). HSMD (as well as HSMC) has unique features: it provides rigorous lower and upper bounds for F, which enables one to determine the accuracy from HSMD results alone without the need to know the correct answer (however, not all of these features are used in the present application). Furthermore, F can be obtained from a very small sample and in principle even from *any* single conformation (see next section, and ref 23).

II.3. Fluctuations. The fact that HSMD provides an approximation for P_i^B means that the *absolute* entropy can be directly obtained and that the free energy, F, can in principle be calculated from a single conformation and in practice from a small sample. This stems from the fact that if the energy, E_i, and P_i^B ($= \exp(-E_i/k_B T)/Z$) of *any* structure i are known, the *total* free energy of the system, F, is known as well, since $F_i = E_i + k_B T \ln P_i^B = E_i + k_B T(-E_i/k_B T - \ln Z) = -k_B T \ln Z = F$; in other words, the *exact* F has zero fluctuation.[56,57] (While this may look strange, one should bear in mind that the calculation of P_i^B for a single conformation depends on the entire ensemble through its normalization factor, which is the partition function.) Still, for an approximate P_i the fluctuation of $F(P_i)$ is finite, but it is expected to decrease as the approximation improves, i.e., as $P_i \rightarrow P_i^B$ where the required sample size decreases as well.[56,57,23] Notice, however, that unlike the free energy discussed above, the entropy (and energy) cannot

be obtained from an *arbitrary* single structure but should be calculated from a Boltzmann sample. We shall return to this matter in section II.6.

II.4. The Reconstruction Procedure for Binding. The process starts by carrying out two *production* MD *runs* of the ligand in water and the protein−ligand complex in water. From these trajectories, we determine two sets of n_s equally spaced frames (snapshots) for later reconstruction and TI calculations by HSMD-TI. As discussed earlier, in the reconstruction process, one seeks to calculate the Boltzmann probability density P_i^B related to each of the frames, as a product of transition probabilities (TPs). For simplicity, we describe this procedure as applied to the ligand in water, where the ligand can be any organic molecule (e.g., L8). For this system (ligand in water), it is convenient to reconstruct the ligand conformation first, followed by the reconstruction of the configuration, $\mathbf{x}^N$, of the N water molecules.[22,26−30]

In the next step, we identify the K (internal) dihedral and bond angles ordered along the chain where their set is denoted by $[\alpha_k]$, $k = 1...K$. We then denote the ordered heavy atoms (and polar hydrogens) along the chain by $k' = 1...K/2$; we shall see below that in the reconstruction process the position of atom k' is determined solely by the dihedral and bond angles, $k - 1$ and k ($k = 2k'$), respectively, and the bond length $r_{k'}$ with the Jacobian $r_{k'}^2 \sin \alpha_k$. The calculation of the TPs depends on the fluctuations in these three coordinates. However, bond stretching is ignored because in general it contributes very little to entropy *differences*[58−60] (notice, however, that bond stretching can straightforwardly be implemented within HSMD; see eq 9 below); therefore, in practice, the only variables considered are α_{k-1} and α_k, and for the sake of completeness in the calculations, we use the Jacobian $\langle r \rangle^2 \sin \alpha_k$, where the constant $\langle r \rangle = 1.6$ Å is an average value that appears for the ligand in the protein and the solvent environments and thus gets canceled in entropy differences. We then calculate the variability range

$$\Delta\alpha_k = \alpha_k(\max) - \alpha_k(\min) \tag{3}$$

where $\alpha_k(\max)$ and $\alpha_k(\min)$ are the maximum and minimum values of α_k found in each sample.

Each of the system configurations (frames; $[\alpha_k]$,$\mathbf{x}^N$; denoted i for brevity) is reconstructed in two stages, where the ligand structure is reconstructed first followed by the reconstruction of the water configuration. Because the position of atom k' is defined by a dihedral and a bond angle, one has to calculate their TP simultaneously. Thus, at step k' ($k = 2k'$) of stage 1, $k - 2$ angles $\alpha_{k-2} \cdots \alpha_1$ have already been reconstructed, and the TP density of $\alpha_{k-1}\alpha_k$, $\rho(\alpha_{k-1}\alpha_k|\alpha_{k-2}, ..., \alpha_1)$, is calculated from an MD run, where the *entire future* of the ligand and water is moved [i.e., ligand's atoms k', $k' + 1$, ...$K/2$ and their connected hydrogens, and all the water molecules] while the past (ligand's atoms 1, 2, ..., $k' - 1$ and their connected hydrogens) are held fixed at their values in conformation i (see Figure 2 in ref 29 and a note in ref 61). By considering a future conformation every 20 fs, a sample of size n_f is generated. Two small segments (bins) $\delta\alpha_{k-1}$ and $\delta\alpha_k$ are centered at $\alpha_{k-1}(i)$ and $\alpha_k(i)$, respectively, and the number of *simultaneous* visits, n_{visit}, of the future chain to these two bins during the simulation is calculated. One obtains[22,26−30]

$$\rho_{ligand}(\alpha_{k-1}\alpha_k|\alpha_{k-2}, ..., \alpha_1) \approx \rho^{HS}(\alpha_{k-1}\alpha_k|\alpha_{k-2}, ..., \alpha_1) = n_{visit}/[n_f \delta\alpha_{k-1}\delta\alpha_k J] \tag{4}$$

where $\rho^{HS}(\alpha_{k-1}\alpha_k|\alpha_{k-2}, ..., \alpha_1)$ becomes exact for very large n_f ($n_f \to \infty$) and very small bins ($\delta\alpha_{k-1}, \delta\alpha_k \to 0$). This means that in practice $\rho^{HS}(\alpha_{k-1}\alpha_k|\alpha_{k-2}, ..., \alpha_1)$ will be approximate due to insufficient future sampling (finite n_f) and relatively large bins (where their optimal size depends on n_f). The Jacobian is $J = \langle r \rangle^2 \sin \alpha_k$. The corresponding probability density related to the ligand's conformation is

$$\rho^{HS}(\alpha_K, ..., \alpha_1) = \rho^{HS}([\alpha_k]) = \prod_{k=1,2}^{K-1} \rho^{HS}(\alpha_k\alpha_{k+1}|\alpha_{k-1}, ..., \alpha_1) \tag{5}$$

$\rho^{HS}([\alpha_k])$ defines an approximate entropy functional denoted S^A_{ligand}, which can be shown (using Jensen's inequality, see ref 23) to constitute a *rigorous* upper bound for the correct S_{ligand}

$$S^A_{ligand} = -k_B \int_m \rho^B_{ligand}([\alpha_k]) \ln \rho^{HS}([\alpha_k]) J \, d[\alpha_K] \tag{6}$$

$\rho^B_{ligand}([\alpha_k])$ is the Boltzmann probability density of $[\alpha_k]$, and J is the Cartesian to the internal coordinates Jacobian. ($S^A_{ligand} \geq S_{ligand}$ is also known as the Gibbs' inequality). Being an upper bound suggests that S^A_{ligand} will decrease as the approximation improves. It should be noted that S^A_{ligand} is defined by our procedure and can be viewed as the conformational internal "entropy of mean force", which constitutes a measure of a pure geometrical character for the flexibility of the ligand; it is estimated by $\overline{S}^A_{ligand}$ from an MD (Boltzmann) sample of size n_s

$$\overline{S}^A_{ligand} = -(k_B/n_s) \sum_{t=1}^{n_s} \ln \rho^{HS}(t) \tag{7}$$

Notice again that S^A_{ligand} is the internal entropy of the ligand. The internal entropy of the ligand in the active site is calculated in the same way, where, however, the future includes the waters as well as the protein atoms, which are all moved by MD in the reconstruction process. We denote the entropies of the ligand in the protein and the solvent (water) environments by $S^A_{ligand}(p)$ and $S^A_{ligand}(sol)$, respectively, where our main interest is in their *converged* difference ΔS_{ligand}, which is expected to be the *exact* difference within the statistical error (see discussion in section II.2 and ref 55)

$$\Delta S_{ligand} = S^A_{ligand}(sol) - S^A_{ligand}(p) \text{ converged} \tag{8}$$

Thus, we calculate $S^A_{ligand}(sol)$ and $S^A_{ligand}(p)$ for increasing n_f and decreasing bins, verifying that both entropies decrease monotonically as the approximation improves; i.e., both approach the correct values from above (notice again that the commanding parameter is n_f where the bin size should correspond to the given statistics (n_f), and it cannot be decreased independently). Typically, the convergence of ΔS^A_{ligand} is much faster than that of the individual entropies, due to cancellation of comparable errors in $S^A_{ligand}(sol)$ and $S^A_{ligand}(p)$. Thus, one can obtain ΔS_{ligand} in a desired accuracy, when the changes in the improved values of ΔS^A_{ligand} are smaller than a given statistical error (notice that this error also depends on the sample size, n_s, and other simulation parameters). The range of errors obtained in our previous work has been 0.2−1 kcal/mol. Therefore, HSMD is considered to be an exact procedure.

Two comments should be made. First, besides S_{ligand}, one should also include in ΔA^0 (eq 2), the contribution of the ligand−ligand energy, $E_{ligand-ligand}$, which is averaged over the samples of the two environments (of size n_s) leading to

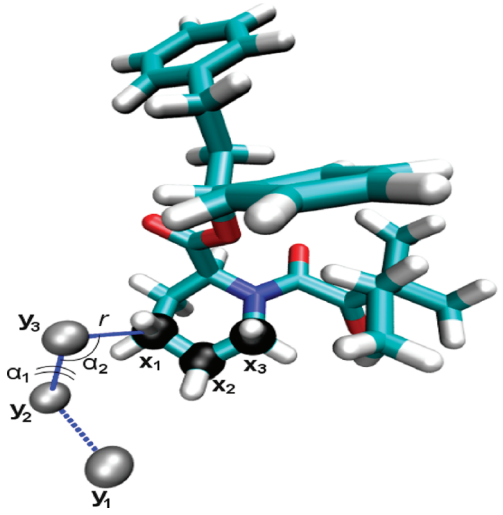

Figure 2. Reconstruction of the external coordinates. $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are the coordinates of three (successive) reference atoms on ring 3 of the ligand; $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$ are three fixed positions in space. The position $\mathbf{x}_1$ is defined by "dihedral angle" α_1, "bond angle" α_2, and $r=|\mathbf{x}_1\text{-}\mathbf{y}_3|$. In the reconstruction of the internal coordinates the right branch is treated first.

$E_{\text{ligand}-\text{ligand}}(\text{sol})$ and $E_{\text{ligand}-\text{ligand}}(\text{p})$. Second, notice that the effect of bond stretching can be considered by defining a bin $\delta r_{k'}$ around the existing k'th bond length, $r_{k'}$ (with the related Jacobian, $J = rk'^2 \sin \alpha_k$), and eq 4 becomes

$$\rho_{\text{ligand}} = n_{\text{visit}}/[n_f \delta\alpha_{k-1}\delta\alpha_k \delta r_{k'} J] \tag{9}$$

where n_{visit} is the number of *simultaneous* visits of the future chain to the *three* bins.

II.5. Entropy of the External Coordinates. To calculate the external entropy within the framework of HSMD, one defines initially three successive atoms of the ligand (with coordinates $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ in the laboratory frame) where the TPs for conformation i are based on a small cube (bin) of volume V_{cube} around $\mathbf{x}_1$ and a small bin for each of the Euler angles. More specifically, to be consistent with the reconstruction of the internal coordinates, the position of atom 1 ($\mathbf{x}_1$) is expressed by spherical coordinates. Thus, one determines three fixed points in space with coordinates $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$, which together with $\mathbf{x}_1$ define a "dihedral angle" α_1 (based on $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{y}_3$, and $\mathbf{x}_1$), a "bond angle" α_2 (based on $\mathbf{y}_2$, $\mathbf{y}_3$, and $\mathbf{x}_1$), a distance $r = |\mathbf{x}_1 - \mathbf{y}_3|$, and their corresponding bins (see Figure 2); these bins together with the Jacobian, $r^2 \sin(\alpha_2)$, define $V_{\text{cube}}(i)$ of ligand conformation i. The contribution of atom 1 to S_{external} [denoted $S_{\text{external}}(1)$] is based on eq 9, where the occupancy, n_{visit}, of $\mathbf{x}_1$ in $V_{\text{cube}}(i)$ is obtained by generating n_f future conformations of the entire ligand (and environment); $S_{\text{external}}(1)$ is averaged over the sample of n_s frames. Notice that for the ligand in solvent such calculation would be very time-consuming and not necessary because the result TP = $1/V$ is known analytically.

After atom 1 has been reconstructed, it becomes fixed at its position in conformation i and the contribution of atom 2 is calculated. Thus (assuming that $r = |\mathbf{x}_2 - \mathbf{x}_1|$ is constant), one defines a dihedral angle α_3 (based on $\mathbf{y}_2$, $\mathbf{y}_3$, $\mathbf{x}_1$, and $\mathbf{x}_2$) and a bond angle α_4 (based on $\mathbf{y}_3$, $\mathbf{x}_1$, and $\mathbf{x}_2$) with their bins and the Jacobian, $r^2 \sin(\alpha_4)$, where $r^2 = 1$. [$r^2 = 1$ because we calculate only the angular contribution, i.e., for a partial spherical surface of radius $r = 1$; this is compatible with the solvent side, where a (very long) such simulation would lead to the (analytically known) TP = $1/4\pi$.] On the protein side, one obtains $S_{\text{external}}(2)$ through eq 4.

Finally, the contribution of atom 3 is considered by defining the pair α_5 ($\mathbf{y}_3$, $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$) and α_6 ($\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$) (assuming that $r = |\mathbf{x}_3 - \mathbf{x}_2|$ is constant). However, only the effect of the dihedral α_5 is considered (using eq 4) by defining an artificially large bin for the bond angle α_6 where the corresponding contribution is $S_{\text{external}}(3)$. Here, the Jacobian is $r^2 \sin(\alpha_6) = 1$, which is again compatible with the solvent side where a long reconstruction simulation would lead to the analytically known TP = $1/2\pi$.

One can define the external entropy for the ligand in solvent as $-k_B T \ln[1/(V8\pi^2)]$ where $1/(V8\pi^2)$ is the product of the above three TPs; however, this entropy is already included in eq 1. Correspondingly, the sum of the three external entropies calculated above for the ligand in the protein environment leads to S_{external}

$$S_{\text{external}} = S_{\text{external}}(1) + S_{\text{external}}(2) + S_{\text{external}}(3) \tag{10}$$

where S_{external} constitutes a measure for the conformational freedom of the ligand in the active site; hence, it provides some estimation for the size of the latter.

It should be pointed out that because with HSMD the whole future is scanned, the reconstruction procedure can in principle be started from any atom, i.e., from the "first" atom of a chain-like ligand, from the "last" one (in an opposite direction), or from any middle atom, where the order in which the two branches are reconstructed is arbitrary. In practice, a large enough n_f is expected to lead to close results for $S^{\text{A}}_{\text{ligand}} + S_{\text{external}}$ based on different starting atoms (this choice of atoms might lead to different computational efficiencies; however, this issue has not been studied by us as yet). Also, because the reconstruction process treats all atoms in a successive manner, the three reference atoms should be nearest neighbors along the ligand, and the position of the first atom should preferably be chosen in such a way that S_{external} will best express the global movement of the ligand in the active site. While an ideal first (global) coordinate would be the center of mass, this coordinate is inadequate due to its constant change during reconstruction. We shall see later that to achieve this goal the reference atoms of L8 are defined close to the middle of the chain.

II.6. HSMD-TI. After a ligand conformation i (say, in solvent) has been reconstructed, one should reconstruct the water configuration $\mathbf{x}^N(i)$ in the presence of a *fixed* ligand structure,[23] which would lead to $\rho_{\text{water}}(i)$ and to the contribution, F_i, to the free energy, F:

$$F_i = [E_{\text{water}-\text{water}}(i) + E_{\text{water}-\text{ligand}}(i) - k_B T \ln \rho_{\text{water}}(i)] + [E_{\text{ligand}-\text{ligand}}(i) - TS^{\text{A}}_{\text{ligand}}(i)] = F_{\text{water}}(i) + F_{\text{ligand}}(i) \tag{11}$$

where, for large n_f [defined also for $\rho_{\text{water}}(i)$], $F_i \rightarrow F$ (see discussion in section II.3). However, reconstructing $\rho_{\text{water}}(i)$ would be a time-consuming process that would be necessary for calculating the absolute value of $F_{L,N}$ (eq 2). Nevertheless, $F_{\text{water}}(i)$ can also be obtained by a TI procedure where an ideal gas (in N, V, and T) is gradually transformed first to liquid water, leading to $F_{0,N}$ (eq 2); then the ligand−water interactions are turned on and are integrated to their full value, leading to $F_i^{\text{TI}}(\text{sol})$, which for a long integration is equal to $F_{\text{water}}(i)$ (eq 11). Therefore,

the water contribution to the *difference* $\Delta F_{sol} = F_{L,N} - F_{0,N}$ (eq 2) can be calculated for each of the n_s configurations i by a TI procedure in which the already reconstructed ligand conformation $[\alpha_k]$ is kept fixed and the water−ligand interactions are increased gradually from zero to their full values, leading to $F_i^{TI}(sol)$; in practice, however, it is easier to decrease these interactions to zero (using $\lambda \rightarrow 0$), obtaining the negative of $F_i^{TI}(sol)$. This elimination of the ligand−environment interactions by TI is expected to be physically more reliable than building them from zero. In the latter case, one can encounter unphysical situations, where, for example, too many water molecules are trapped in the active site after integration. Thus, we carry out only the elimination TI but for quite a few *different* ligand structures, averaging the results where their standard deviation defines the error.

This suggests that, in principle, ΔF_{sol} can be obtained from *any single* configuration i by using large enough n_f for calculating $S_{ligand}(i)$ (eqs 6 and 7) and long enough integration, which leads to accurate $F_i^{TI}(sol)$ (see discussion in section II.3). In practice, however, the results are averages over a relatively *small* sample size $10 \leq n_s \leq 100$, where the errors are defined by $sd/(n_s)^{1/2}$, where sd stands for the standard deviation (see section II.3); $sd \rightarrow 0$ as n_f and the integration time increase. In the protein environment, one obtains in the same way $F_i^{TI}(p)$ by eliminating the ligand−protein and ligand−water interactions. Technically, the charges are eliminated first, followed by elimination of the Lennard-Jones (LJ) potential, using a soft-core potential. With $TS_{external}$, ΔA^0 (eq 2) becomes

$$\begin{aligned}\Delta A^\circ &= \Delta F_p - \Delta F_{sol} + k_B T \ln(8\pi^2 V^0) - TS_{external} \\ &= [E_{ligand-ligand}(p) - TS_{ligand}(p) + F^{TI}(p)] \\ &\quad -[E_{ligand-ligand}(sol) - TS_{ligand}(sol) + F^{TI}(sol)] \\ &\quad + k_B T \ln(8\pi^2 V^0) - TS_{external} = \Delta E_{ligand-ligand} - T\Delta S_{ligand} \\ &\quad + \Delta F^{TI} + k_B T \ln(8\pi^2 V^0) - TS_{external} \qquad (12)\end{aligned}$$

where all of the quantities (defined earlier) are averages over n_s snapshots and Δ denotes differences in the corresponding variables (protein− solvent); the errors of ΔA^0 and its different components are obtained from the standard deviations (sd) divided by $(n_s)^{1/2}$. Notice again that $V^0 8\pi^2$ and the term in the logarithm of $S_{external}$ have the same dimensions, which disappear in the difference.

III. RESULTS AND DISCUSSION

III.1. Preliminary Studies without Long-Range Electrostatics. Because L8 is an uncharged hydrophobic ligand, we had reasons to believe that describing the complex with a good model of partial structure (where the whole protein is considered, it is covered with water, and cut-offs on nonbonded interactions are not imposed), long-range electrostatic interactions would not be significant, and they would get canceled in free energy differences. Previous binding studies of models of partial structure have been found to be successful.[3,62]

Thus, we studied initially this system using the AMBER 10 package,[46] where due to its small size, the whole complex was considered (PDB code 1fkg, 1974 atoms). Around the active site, we defined a TIP3P[48] water sphere of 22 Å radius (658 molecules), which covered the whole protein (because the protein was not soaked in a large container of water, we consider this system a model of partial structure). For simulations of the solvated ligand, a sphere of 22 Å radius was used, containing 1310 waters. The protein was modeled by the AMBER99 force field[47] and the ligand by GAFF[49] (with the AM1-BCC partial charges). *No cut-offs* on nonbonded interactions were imposed. While this model includes a large part of the system, the HSMD-TI results obtained for the absolute free energy were too high (in the best case favoring unbinding by 5 kcal/mol) even after several attempts to optimize the structure of the ligand in the complex by minimizing the ligand−environment interaction energy prior to the production run. No improvement could be achieved also by charging the neutral histidines or adding chlorine ions to neutralize the total charge of the system.

As mentioned in the Introduction, these studies have led to the conclusion that long-range electrostatics should be considered, and below we describe the application of HSMD-TI to L8−FKBP modeled by periodic boundary conditions with PME. Results are also provided later for this system modeled by SSBP/GSBP.[43,44]

III.2. Studies with Long-Range Electrostatics (PME). To apply PME, we also used the AMBER 10 package along with the AMBER99 and GAFF force fields (including TIP3P water) and the crystal structure described above. The unit cell of the periodic system was defined by constructing a truncated octahedron around the protein which was filled with 4581 water molecules. The minimum distance between the protein and the walls of the cell was 11 Å (this cell size is comparable to that used in previous studies of L8; also, to be compatible with these studies, the three histidine residues have positive charge). To neutralize the system, four chloride counterions were added at random locations. The periodic boundary conditions were defined by the PME algorithm,[45] with a cutoff of 10 Å. This system was optimized in several stages, starting with a short simulated annealing protocol consisting of energy minimization, where 10^4 steepest descent and 10^4 conjugate gradient steps were performed. Then, using MD, the system was heated to 600 K over 12 ps (using the Berendsen thermostat[63] with a 2 ps time constant), slowly cooled down to 100 K (28 ps), cooled down further (5 ps) with a smaller time constant (1.0 ps), and finally driven to 0 K with an even smaller time constant of 0.1 ps for another 5 ps. The time step during this process was 1 fs, and the SHAKE[63] algorithm was applied to all hydrogens in all our studies.

Next, we performed a temperature and pressure equilibration to 300 K and 1 atm, for 50 ps, using a Berendsen thermostat time constant of 1.5 ps and a weak coupling isotropic barostat with a relaxation time of 2 ps; the time step here was 2 fs. Finally, a 2 ns constant volume production run at 300 K (time step of 2 ps) was carried out, where the first 0.4 ns trajectory was used for equilibration. From the latest 1.6 ns, a frame was extracted every 40 ps, thus obtaining a sample of $n_s = 40$ frames. These frames were analyzed by our HSMD-TI procedure.

A similar procedure was applied to the solvent system consisting of L8 without the protein, which was solvated with 631 water molecules in a truncated octahedron. No counterions were added to this neutral system. Again, 40 frames were determined for this system as well. It should be pointed out that we have checked the movement of the ligand in the 2 ns MD trajectories using computer graphics. Rings 1 and 2 and the linear part of the peptide that starts from ring 3 (see Figure 1) were found to fully rotate in both the solvent and protein, where the ligand in solvent showed a random coil behavior. We therefore consider the 2 ns trajectories to be reasonably long for the present study, which is mostly concentrated on the development of HSMD-TI for various realistic models.

Table 1. Results for the Internal Entropy S^{A}_{ligand} of the Ligand in the Solvent and Protein Environments and the Difference $\Delta S^{A}_{ligand} = S^{A}_{ligand}(sol) - S^{A}_{ligand}(p)$ Obtained with AMBER99/PME[a]

bin size, δ	n_f	$TS^{A}_{ligand}(sol)$	$TS^{A}_{ligand}(p)$	$T\Delta S^{A}_{ligand}$
$\Delta\alpha_k/30$	1000	3.2	0.8	2.4
	2000	−10.5	−12.7	2.2
	3000	−18.4	−20.2	1.8
	4000	−23.4	−25.3	1.9
	5000	−27.3	−29.2	1.9
	6000	**−30.4**	**−32.3**	**1.9**
$\Delta\alpha_k/60$	1000	3.2	0.9	2.3
	2000	−10.7	−12.9	2.2
	3000	−18.5	−20.3	1.8
	4000	−23.5	−25.4	1.9
	5000	−27.5	−29.3	1.8
	6000	**−30.7**	**−32.4**	**1.7**
$\Delta\alpha_k/90$	1000	3.1	0.8	2.3
	2000	−10.7	−12.8	2.1
	3000	−18.5	−20.3	1.8
	4000	−23.6	−25.5	1.9
	5000	−27.8	−29.5	1.7
	6000	**−30.9**	**−32.6**	**1.7**
converged				**1.7 ± 0.2**

[a] The results were obtained by reconstructing $n_s = 40$ structures of L8 selected homogeneously from MD samples of 1.6 ns for the solvent and in the protein environments. The results are calculated as functions of $\delta = \Delta\alpha_k/l$ and n_f (eq 4)—the bin and sample size of the future chains, respectively. S^{A}_{ligand} is defined up to an additive constant that is expected to be the same for both environments. The (best) results for $n_f = 6000$ are bold-faced.

III.3. Calculation of Entropy. L8 consists of three rings denoted 1−3 (Figures 1 and 2), where ring 3 is located deep in the active site as a basis for the two subchains which point to the outside of the pocket. Therefore, we defined the second carbon of ring 3 as the first reference atom (with coordinates $\mathbf{x}_1$, see section II.5). This allows defining a physically meaningful external entropy which is related to the global movement of L8 in the active site. We first discuss results for the internal entropy. We reconstructed $n_s = 40$ L8 conformations (denoted i) in both, the solvent and the protein, where 31 atoms ($k' = 1, ..., 31$) and 62 angles α_k ($1 \le k \le K = 62$) participate in the reconstruction (see section II.4). Each reconstruction step (out of 31) starts from conformation i with a 120 ps production run where a future L8 conformation is stored every 20 fs for a later analysis; thus, the total sample for each step consists of $n_f = 6000$ future conformations, where the first 200 are usually dropped as part of the equilibration. The number of counts, n_{visit} (eq 4), for each pair of bins is calculated, leading to $TP_{k'}$, where the product of the 31 TPs is the distribution, ρ^{HS} (eq 5), which leads to the entropy, S^{A} (eqs 6 and 7). This reconstruction is significantly larger than the 14-atom reconstruction performed previously for biotin.[22]

In practice, the calculation is done in two stages, where in stage 1 we carry out the reconstruction simulations. Thus, for the 31 reconstructed atoms, $31n_f$ future chains are generated for snapshot i, and their coordinates are stored in a file for a later analysis in stage 2. Stage 1 can be performed in a straightforward way with any of the available programs, AMBER, TINKER, CHARMM, etc. In stage 2, the files generated in stage 1 are read by an analysis program, which enables one to calculate the transition probabilities and to study the behavior of S^{A}_{ligand} (eq 7), ΔS^{A}_{ligand} (eq 8), and $S_{external}$ (eq 10) as a function of various parameters (e.g., bin size, n_s, and n_f) without the need to carry out additional (stage 1) runs. This free program with a tutorial and explanations appears at http://www.ccbb.pitt.edu/Faculty/meirovitch/reconstruction-web/reconstruction-web.html.

III.4. Results for the Internal Entropy. Results for S^{A}_{ligand} (eqs 6 and 7) are presented in Table 1. It should be noted that the angles (α_k) are calculated in radians, which can lead to negative entropies (in contrast to our previous studies where using degrees led to positive entropies[22,25−29]). This is not unexpected, as $S^{A}_{ligand}(sol)$ is defined up to an additive constant, and we are interested only in entropy differences ΔS_{ligand} (eq 8) where this constant cancels out. As expected by theory, the results for S^{A}_{ligand} decrease systematically as the approximation improves, i.e., with increasing n_f, but they remain unchanged as a function of δ. However, these results have not converged even for $n_f = 6000$. On the other hand, the corresponding results for ΔS^{A} show convergence to 1.7 ± 0.2 kcal/mol, which is thus considered to be the exact result. This relatively small difference might seem surprising, but as pointed out in section III.2, computer visualization shows that the two subchains (branches) of the ligand that protrude from the active site have significant conformational freedom, where rings 1 and 2 perform full rotations like in solvent.

III.5. Results for the External Entropy. The contributions of the three reference atoms to the external entropy appear in Table 2 as a function of n_f and only a single bin size $\Delta\alpha_k/60$, because exactly the same results (within the statistical errors) were obtained for bin sizes within the range $\Delta\alpha_k/30$ to $\Delta\alpha_k/90$. As expected, for each atom, the entropy decreases as the approximation improves, i.e., as n_f increases; however, the results have not completely converged even by increasing the maximal n_f to 12 000. The space covered by these variables can be estimated from eq 4, using the results obtained for the largest bin, $\Delta\alpha_k$, for which $n_{visit}/n_f = 1$, or more specifically by calculating $\exp(TS^{A}/0.6)$, where TS^{A} stands for the results in the table and $k_BT = 0.6$ at 300 K. Thus, atom 1 visits a volume of ∼14 Å^3 where the "bond angle" and "dihedral angle" related to atom 2 cover together 7.2% of 4π (=12.6) and the range of change of the dihedral angle of atom 3 is ∼48°. However, the results for atom 2 and 3 should be considered as lower bounds since the predecessor atoms, 1 and 2, respectively are held fixed.

III.6. TI Results. To each of the $n_s = 40$ frames of the solvent sample we applied a TI procedure where the ligand−water interactions were turned off gradually for a *fixed* L8 structure; for the 40 frames of the protein environment, we decoupled both the ligand−water and ligand−protein interactions. Using a parameter λ, $0 \le \lambda \le 1$, the electrostatic interactions were decoupled first followed by decoupling the Lennard-Jones (LJ) potentials (in the presence of zero electrostatic interactions). In all, 30 λ values (windows) were used, 13 for the electrostatic interactions (λ = 0.01, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.95, and 0.99) and 17 for LJ (λ = 0.01, 0.03, 0.07, 0.10, 0.15, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.85, 0.90, 0.93. 0.97, and 0.99). In the LJ integration, we used a soft-core potential based on $\delta = 3$ Å^2.[29,64] For a given frame (i), each integration step (window) always started from the (initial) structure of i according to the corresponding step potential energy [$E(\lambda)$], followed by a 120 ps production run, where the initial 20 ps are discarded for equilibration.

Table 2. External Entropy for the Three Reference Atoms Based on AMBER99/PME[a]

bin size, δ	n_f	$TS^A_{ligand}(1)$	$TS^A_{ligand}(2)$	$TS^A_{ligand}(3)$	TS^A_{ligand} (total)
$\Delta\alpha_k/60$	2000	1.3	−0.1	0.7	1.9
	4000	1.0	−0.6	0.2	0.6
	6000	0.8	−0.8	0.0	0.0
	8000	0.7	−0.9	−0.2	−0.4
	10000	0.6	−1.1	−0.3	−0.8
	12000	**0.5**	**−1.2**	**−0.4**	**−1.1**
$\Delta\alpha_k$	12000	1.6	−0.06	−0.1	1.4
space		14.4 Å^3	0.90 out of	48°	
covered			4π = 12.6 (7.2%)		

[a] The results are based on the 40 structures of L8 in the active site used to calculate the internal entropy (Table 1); they are presented for bin sizes $\delta = \Delta\alpha_k/60$ and $\Delta\alpha_k$ (eq 3) and n_f (eq 4)—the sample size of the future chains with a maximal value of 12 000. $S^A_{ligand}(1)$ is the entropy related to the volume occupied by the first atom. $S^A_{ligand}(2)$ is related to the "bond angle" and "dihedral angle" of atom 2 and $S^A_{ligand}(3)$ to the dihedral angle of atom 3. In the last row, we provide the space covered by the variables of each atom. The statistical error for $-TS^A_{ligand}$(total) is 0.5 kcal/mol. The (best) results for n_f = 12000 are bold-faced, where $TS^A_{ligand}(\text{total}) = TS_{external}$.

Table 3. AMBER99/PME Results for the Free Energy, F^{TI} in kcal/mol, Obtained by TI for the Protein and the Solvent Environments[a]

	F^{TI}(ch)	F^{TI}(LJ)	F^{TI}
protein	−25.9 ± 0.4	−15.6 ± 0.9	−41.5 ± 0.7
solvent	−20.2 ± 0.2	−0.2 ± 0.2	−20.4 ± 0.2
Δ = prot − sol	−5.7 ± 0.4	−15.4 ± 0.5	−21.1 ± 0.5

[a] F^{TI}(ch) and F^{TI}(LJ) are free energies calculated by TI by eliminating the electrostatic and LJ interactions, respectively, based on n_s = 20 structures for each environment. F^{TI} is their sum.

The results for F^{TI} obtained by eliminating the charge (ch) and the LJ also include ligand−ligand contributions,[65] and because the conformations are fixed, these contributions constitute part of the ligand−ligand interaction energy. To calculate these energies, we generated two samples of the ligand in vacuum by removing the protein and water and applied TI to both (vacuum) samples. We have found the average result in solvent to be lower by 0.2 kcal/mol than the protein result. These very close results were subtracted from the original F^{TI} values, and the subtracted values are provided in Table 3, leading to $\Delta F^{TI} = F^{TI}(p) - F^{TI}(s) = -21.1$ kcal/mol (this result is expected to be converged as −20.6 was obtained for 60 ps runs). The ligand−ligand energy $E_{ligand-ligand}$ in Table 4 considers only the total bonded energy of the ligand, $E_{bond}(p) = 64.6$ and $E_{bond}(s) = 64.0$, i.e., $\Delta E_{bond} = 0.6$ kcal/mol.

Table 4 presents several energetic and entropic components that lead to our estimation of the absolute binding free energy, $\Delta A^0 = -10.7 \pm 1.0$ kcal/mol, which within the error bars is equal to the experimental values −10.9 (ref 37) and −10.6 (ref 66) kcal/mol. Notice, however, that the results for $-TS_{external}$ (Table 2) have not been converged, and we have estimated this uncertainty by specifying an increased statistical error of ±0.5 kcal/mol to $TS_{external}$. This nonconvergent result is due to insufficient future sampling (too small n_f), whereas the number of frames studied, n_s = 40, is adequate. In fact, as discussed in sections II.3 and II.6, HSMD-TI requires a relatively small sample (number of frames). Indeed, the result for ΔF^{TI} (eq 12) is based on n_s = 20 frames, where $n_s \sim 15$ has led to the same result. While the entropy in principle requires larger samples than the free energy, converging results for ΔS_{ligand} have been obtained for $n_s \sim 40$.

Our result, $\Delta A^0 = -10.7 \pm 1.0$ kcal/mol, should be compared to results obtained by other methods. Pande's group applied DAM to a set of eight ligands bound to FKBP using the AMBER and GAFF force fields, where the ligand−environment interactions were eliminated with FEP rather than TI. In their first paper,[8] they obtained for L8 $\Delta A^0 = -7.3$ kcal/mol without considering the standard state contribution. Much better results have been obtained for these eight ligands in their following paper,[13] where ΔA^0(L8) = −10.5 ± 0.26 kcal/mol. However, we feel that the active site volume, V^b, defined in ref 13 is too large depending strongly on the ligand size. Indeed, for all eight ligands, V^b is larger thav V^0 where the maximal ratio is $V^b/V^0 \sim 2.8$.

In a third study by Fujitani et al.,[12] the force field was changed and the standard state correction was not applied; for L8, they obtained $\Delta A^0 = -10.1$ kcal/mol. Using DDM and the SSBP/GSBP modeling, Wang et al.[40] obtained $\Delta A^0 = -10.3$ kcal/mol with errors of ±0.4 and ±1.2 kcal/mol depending on the initial equilibration method. It should be pointed out that they used a flat bottom restraining potential on all symmetric units (rings) which might affect the entropy difference, ΔS (eq 8), hence the binding free energy. We also mention the results of Lamb et al.,[42] who applied LIE[67] with various sets of parameters obtaining results for ΔA^0(L8) between −10.1 and −10.3 kcal/mol. It should be noted that in our calculations none of the restrictions mentioned above have been imposed.

III.7. DDM Results Obtained with SSBP/GSBP. The use of PME requires treating the entire protein soaked in (typically) thousands of water molecules; hence for a large protein application of HSMD-TI (as well as other techniques), is expected to become time-consuming. This is the reason why we have studied initially the avidin−biotin complex with a model of partial structure, where only protein atoms and waters close to the active site were considered. However, the unsatisfactory results (discussed in section III.1) obtained for FKBP−L8 suggest that taking into account long-range electrostatic effects is imperative. Models of partial structure that take into account long-range electrostatics (the reaction field) have been developed, in particular by Warshel's group (the program MOLARIS[68−70]) and the SSBP/GSBP models mentioned earlier.[9,43,44]

Therefore, we decided to apply HSMD-TI to the complex FKBP−L8 modeled by GSBP, where L8 in a solvent is modeled by SSBP; the relevant computer programs are included in the software package CHARMM.[50] While this system has already been studied by Wang et al.[40] using DDM, we have decided to apply DDM again where some of the parameters differ from those used by Wang et al. This would help us get a feel for the sensitivity of the model and establish a set of results that will be compared to later results obtained by HSMD-TI.

First, instead of using the GAFF parameters,[49] the force-field parameters for the ligand were taken from the CHARMM general force field,[71] and the charges of the histidine residues were kept neutral. Most importantly, our DDM procedure is based on a single harmonic distance restraint with a force constant, K_r = 7 kcal/(molÅ^2) (as compared to several different restrains applied by Wang et al.[40]); finally, no flat-bottom restraining potential was imposed on all symmetric units (rings).[40]

Table 4. Energetic, Entropic, and Free Energy Components (in kcal/mol) Which Contribute to the Absolute Free Energy of Binding, ΔA^0 (eq 12), Obtained for the Protein and Solvent with AMBER99/PME[a]

	$k_BT\ln(8\pi^2V^0)$	$-TS_{external}$	$-TS_{ligand}$	$E_{ligand-ligand}$	F^{TI}	total
protein	7.0	1.1 ± 0.5	32.3 ± 0.5	64.6 ± 0.3	−41.5 ± 0.7	63.2 ± 1.6
solvent			30.6 ± 0.6	64.0 ± 0.6	−20.4 ± 0.2	73.9 ± 1.3
Δ = prot − solv	7.0	1.1 ± 0.5	1.7 ± 0.2	0.6 ± 0.7	−21.1 ± 0.5	−10.7 ± 1.0

[a] The table is organized according to eq 12. Results for TS_{ligand}, $TS_{external}$, and F^{TI} are taken from Tables 1, 2, and 3, respectively; most of the components are defined up to an additive constant, and only their difference has a physical meaning. The absolute free energy of binding is $\Delta A^0 = 10.7 \pm 1.0$ kcal/mol and is defined on the right-hand side of the bottom row.

Table 5. Free Energy Results in kcal/mol Obtained by Applying DDM to the SSBP/GSBP/CHARMM Model[43,44a]

	$\Delta A(FEP)_{charge}$	$\Delta A(FEP)_{LJ\text{-}attractive}$	$\Delta A(FEP)_{LJ\text{-}repulsive}$	ΔA_R	ΔA_τ^0	total
solvent	15.0 ± 0.3	40.5 ± 0.2	−49.0 ± 0.2			6.2 ± 0.5
protein	14.7 ± 0.3	59.2 ± 0.2	−49.6 ± 0.5	1.5 ± 0.1	−5.0	20.8 ± 0.6
Δ = solv − prot	0.1 ± 0.4	−18.7 ± 0.3	0.6 ± 0.5	−1.5 ± 0.1	5.0	−14.6 ± 0.8

[a] The elimination of the ligand-environment interactions was obtained by free energy perturbation (FEP) in three steps, treating first the electrostatic interactions then the attractive and repulsive LJ interactions. ΔA_R and ΔA_τ^0 and $\Delta A^0 = -14.6$ kcal/mol are defined in eqs 13 and 14. For each environment, the results are averages of five FEP runs started from different ligand or protein−ligand structures.

The results which appear in Table 5 are averages of five FEP runs. Notice that with the GSBP/SSBP software the ligand−environment interactions are eliminated in the following order: electrostatic, attractive LJ, and repulsive LJ. The equations below are based on the notation of ref 5 (eqs 29−31), but to simplify the presentation, we define several more notations. Thus, $\Delta A_S(FEP) \equiv \Delta A_I^0$ (ref 5) and ΔA_p (FEP) stand for the free energy change due to the elimination of the ligand−environment interactions in the solvent and the protein, respectively. ΔA_R is the free energy due to the application of the restraint, and ΔA_τ^0 is the free energy due to the release of the restraint plus the effect of the standard state volume. Z is the partition function, $Z = 8\pi^2V\overline{Z}$ (see eq 1)

$$\Delta A_I^0 = -k_BT\ln\frac{Z_{L,0}Z_{0,N}}{Z_{L,N}} = \Delta A_S(FEP)$$

$$\Delta A_{II}^{*,0} = -k_BT\ln\frac{Z_{P\ldots L,N}}{Z_{PL,N}} = \Delta A_R + \Delta A_P(FEP)$$

$$\Delta A_\tau^0 = -k_BT\ln\frac{V^0Z_{P,N}Z_{L,0}}{VZ_{P\ldots L,N}} = -k_BT\ln\frac{V^0K_r^{3/2}}{(2\pi k_BT)^{3/2}} \quad (13)$$

where the free energy of binding ΔA^0 is

$$\Delta A^0 = \Delta A_I^0 - \Delta A_{II}^{*,0} - \Delta A_\tau^0 \quad (14)$$

The systems were simulated by Langevin dynamics at $T = 300$ K with a friction coefficient of 5 ps^{-1}. For each environment, five initial configurations were selected from the last 2 ns of a 4 ns run (where the first 2 ns were used for equilibration). The solvent system (described by SSBP) consists of 1000 TIP3P water molecules within a sphere of ∼20 Å, whereas the protein system (described by GSBP) contains 325 waters within a sphere of ∼15 Å. The elimination of the L8−water interactions was carried out in 20 windows for each type of interaction (electrostatic, attractive LJ, and repulsive LJ) where for each window the first 60 ps were used for equilibration and the next 120 ps for production. The different FEP results appear in Table 5 where $\Delta A^0 = -14.4$ kcal/mol is expected to be a converged value, as $\Delta A^0 = -15.3$ kcal/mol has been obtained from an initial limited study based on 30 ps equilibration and 30 ps production (FEP) runs.

The fact that the binding free energy, −14.5, is still lower by ∼3.5 kcal/mol from the experimental value (hence from the very good result obtained by Wang et al.[40]) can be attributed to the following: (1) Only a single (distance) restraint was applied which cannot restrict adequately the conformational freedom of the ligand in the active site. A better restriction was achieved by Wang et al. who used several restraints (among them an orientational restraint). (2) Wang et al. applied a flat-bottom restraining potential on all symmetric units (rings). (3) They used the GAFF parameters and probably charged histidines. However, another reason for the computational/experimental disagreement might be an inaccurate water density around the protein in our calculations. We have found that the results are sensitive to this density, which is difficult to apply accurately to the relatively small protein system. This is probably the reason why Roux and Deng have introduced a grand-canonical procedure to control the water density.[72] Moreover, below, we demonstrate that application of HSMD-TI to the same model (FKBP−L8 described by SSBP/GSBP) leads to a ΔA^0 value comparable to that obtained above by DDM, which lends further support for the water density effect.

III.8. HSMD-FEP Results Obtained with SSBP/GSBP. We have also applied HSMD-TI to the SSBP/GSBP model, where the method actually becomes HSMD-FEP. We used the same parameters defined above for DDM, but studying for the perturbation 10 (rather than five) configurations for each environment. The main difference is that the elimination of the ligand−environment interactions has been applied to fixed ligand structures selected from the 2 ns initial Langevin runs of the protein and solvent systems discussed in section III.7.

The results in Table 6 for the internal entropy TS_{ligand} (eq 7) are not converged, but their difference $T\Delta S_{ligand}$ (eq 8) are converged nicely to 5.5 ± 0.2 kcal/mol, which is larger than the 1.7 kcal/mol obtained in the AMBER/PME calculations (Table 1). This suggests that the ligand in the active site is more restricted by

SSBP/GSBP than by AMBER/PME. This is also in accord with the smaller external entropy, $E_{external}$ (eq 10), of −3.3 kcal/mol obtained in Table 7 than the −1.1 kcal/mol obtained in Table 2. Notice that, as in Table 2, the external entropy in Table 7 has not been converged, and an error of 0.5 kcal/mol has been assigned to it.

All of these results are summarized in Table 8. Notice that in contrast to Table 5, the results in Table 8 under the three columns of ΔA(FEP) are the negative values obtained in the actual FEP (see the end of the paragraph following eq 11).

Table 6. HSMD Results for the Internal Entropy S^A_{ligand} of the Ligand in the Solvent and Protein Environments and the Difference $\Delta S^A_{ligand} = S^A_{ligand}(sol) - S^A_{ligand}(p)$ Obtained for the SSBP/GSBP/CHARMM Model[43,44a]

bin size, δ	n_f	TS^A_{ligand}(sol)	TS^A_{ligand}(p)	$T\Delta S^A_{ligand}$
$\Delta\alpha_k/60$	2000	−11.5	−15.9	4.4
	4000	−24.0	−29.1	5.1
	6000	−31.4	−36.7	5.3
	8000	−36.7	−42.0	5.3
	10000	**−40.7**	**−46.2**	**5.5**
converged				**5.5 ± 0.2**

[a] The results were obtained by reconstructing n_s = 20 structures of L8 selected homogeneously from the Langevin dynamics trajectories generated for the results of Table 5. The results are shown for different sample sizes of the future chains, n_f (eq 4), but only for one bin size $\delta = \Delta\alpha_k/l$ where l = 60, because results for l = 30 and 90 are similar. S^A_{ligand} is defined up to an additive constant that is expected to be the same for both environments. The (best) results for $n_f = 10^4$ are bold-faced.

Table 7. External Entropy for the Three Reference Atoms Obtained for the SSBP/GSBP/CHARMM Model[a]

bin size, δ	n_f	$TS^A_{ligand}(1)$	$TS^A_{ligand}(2)$	$TS^A_{ligand}(3)$	TS^A_{ligand} (total)
$\Delta\alpha_k/60$	2000	1.0	−0.8	1.1	1.3
	4000	0.7	−1.2	−0.1	0.4
	6000	0.5	−1.5	−0.8	−1.8
	8000	0.3	−1.7	−1.4	−2.8
	10000	**0.2**	**−1.8**	**−1.8**	**−3.4**
$\Delta\alpha_k$	10000	1.5	−0.7	−0.2	0.6
space		12.2 Å^3	0.3 out of	41°	
covered			4π = 12.6 (2.5%)		

[a] The results are based on the 20 structures of L8 in the active site used to calculate the internal entropy (Table 6); they are presented for bin sizes $\delta = \Delta\alpha_k/60$ and $\Delta\alpha_k$ (eq 3) and n_f (eq 4)—the sample size of the future chains with a maximal value of 10^4. For the meaning of $S^A_{ligand}(1)$, $S^A_{ligand}(2)$, and $S^A_{ligand}(3)$, see the caption of Table 2. In the last row, we provide the space covered by the variables of each atom. The statistical error for $-TS^A_{ligand}$(total) is 0.5 kcal/mol. The (best) results for $n_f = 10^4$ are bold-faced, where TS^A_{ligand}(total) = $TS_{external}$.

The table shows that the absolute free energy of solvation is $\Delta A^0 = -14.9 \pm 1.5$ kcal/mol, in accord with $\Delta A^0 = -14.6 \pm 0.8$ obtained in Table 5 using DDM, but with a larger error due to the uncertainty in the result for $TS_{external}$.

The efficiency of HSMD-TI(FEP) can be judged by comparing it to that of the DDM procedure applied to FKBP−L8 modeled by SSBP/GSBP.[40] First, each method is based on a pair of calculations where the ligand−environment interactions are annihilated (or created) in the protein and solvent environments using FEP, n_s = 10 and 5, such pairs were carried out by us and in ref 40, respectively. For each pair, two reconstructions are performed by HSMD, where in ref 40, three restraints were built and released by FEP (which requires six integrations in total). It is difficult to compare exactly the times involved in the reconstructions and for treating the restraints, as for example converging results for ΔS_{ligand} were obtained also for n_s = 10 and n_f = 6000 (rather than n_s = 20 and 10 000, see Table 6). Therefore, one can roughly say that DDM and HSMD-FEP have comparable efficiency; however, HSMD-FEP provides more information, the entropies ΔS_{ligand} and $S_{external}$.

III.9. Summary and Conclusions. HSMD-TI is a new method for calculating the absolute free energy of binding which does not suffer from the end point problem and is independent of DAM and DDM. It is of interest to view HSMD-TI from the perspective of DDM. Thus, to apply DDM efficiently, one seeks to limit the conformational freedom of the ligand in the active site by imposing various restraints, the stronger the restraints, the longer the time required for building them up from zero by TI (or FEP). With HSMD-TI, each ligand structure studied is fixed, and restraints are not needed. However, eliminating (by TI) the ligand−environment interactions for a fixed structure leads to some entropy loss (as compared to DAM and DDM), which is recovered by the reconstruction process. Thus, unlike DDM and DAM, HSMD-TI provides the difference in the internal entropy of the ligand in the two environments; one also calculates the external entropy, which constitutes an unbiased measure for the global movement of the bound ligand, providing thereby estimation for the size of the active site. Finally, the fact that HSMD-TI leads to ΔA^0 not as a result of two integrations (protein and solvent) but as a sum of entropic, energetic, and free energy components, enables one to gain a more complete picture of the various forces that determine the complex stability. Thus, HSMD-TI provides deep microscopic insights into the binding mechanism which are important from the academic point of view as well as for rational drug design.

In this paper, the scope of the theory has been extended, where we elaborate on the correlation between free energy fluctuations and the sample size, n_s. We have provided a more complete description of the reconstruction of the internal coordinates, pointing out the freedom in determining the first reference atom, and the order of the treated atoms. Still, one would seek to select

Table 8. Free Energy Results in kcal/mol Obtained by Applying HSMD-FEP to the SSBP/GSBP/CHARMM Model[a]

	$k_BT \ln(8\pi^2V^0)$	$-TS_{external}$	$-TS_{ligand}$	$\Delta A(FEP)_{charge}$	$\Delta A(FEP)_{LJ\text{-}attractive}$	$\Delta A(FEP)_{LJ\text{-}repulsive}$	$E_{ligand-ligand}$	total
protein	7.0	3.4 ± 0.5	46.2	−17.7 ± 0.5	−62.5 ± 0.5	42.9 ± 0.8	189.2 ± 2.0	208.5 ± 1.3
solvent			40.7	−15.5 ± 0.1	−38.1 ± 0.3	47.4 ± 0.3	188.9 ± 1.9	223.4 ± 0.6
Δ = prot − solv	7.0	3.4 ± 0.5	5.5 ± 0.2	−2.2 ± 0.8	−24.4 ± 1.0	−4.5 ± 1.0	0.3 ± 0.9	−14.9 ± 1.5

[a] The elimination of the ligand-environment interactions was obtained by free energy perturbation (FEP) in three steps, treating first the electrostatic interactions, then the attractive and repulsive LJ interactions. $\Delta A^0 = -14.9 \pm 1.5$ kcal/mol is obtained by eq 12. For each environment, the FEP results are averages of 10 runs started from different structures, while the entropy was obtained in Tables 6 and 7 by reconstructing 20 structures. $E_{ligand-ligand}$ is the intraligand energy.

a first reference atom which leads to a physically meaningful external entropy that adequately expresses the global movement of the ligand in the active site. The reconstruction of the external coordinates has been described in detail, as well as the incorporation of bond stretching within the framework of the reconstruction process; potential situations where this latter effect should be considered have been discussed.

The unsatisfactory (preliminary) HSMD-TI results obtained by applying the AMBER99−GAFF−TIP3P potentials to the FKBP−L8 complex described by a *finite* model of partial structure (i.e., a model which is not based on periodic boundary conditions) suggest that long-range electrostatic effects cannot be ignored. Indeed, the excellent result for ΔA^0 obtained in this paper demonstrates the importance of long-range electrostatics, the effectiveness of PME, and the high performance of HSMD-TI. Also, as discussed in sections II.3, II.6, and III.6, HSMD-TI requires relatively small sample sizes.

Finally, we have also tested the SSBP/GSBP model,[43,44] which takes into account long-range electrostatics and as a model of partial structure has the potential to be computationally less demanding than PME. We applied both DDM and HSMD-FEP to FKBP-L8 modeled by SSBP/GSBP and found comparable results for ΔA^0, which, however, are slightly lower than the experimental value. This disagreement has been attributed mainly to inaccurate water density around the protein in our calculations. To develop HSMD-TI further, we plan to apply it in the next step to the complex FKBP12−FK506, where the ligand FK506 is significantly larger than L8.

AUTHOR INFORMATION

Corresponding Author

*Phone: (412) 648-3338. E-mail: hagaim@pitt.edu.

ACKNOWLEDGMENT

This work was supported by NIH grant 2-R01 GM066090-4.

REFERENCES

(1) Hermans, J.; Shankar, S. *Isr. J. Chem.* **1986**, *27*, 225–227.
(2) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742–3746.
(3) Miyamoto, S.; Kollman, P. A. *Proteins* **1993**, *16*, 226–245.
(4) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047–1069.
(5) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* **2003**, *107*, 9535–9551.
(6) Zhou, H.-X.; Gilson, M. K. *Chem. Rev.* **2009**, *109*, 4092–4107.
(7) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. *J. Mol. Biol.* **2007**, *371*, 1118–1134.
(8) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. *J. Chem. Phys.* **2005**, *123*, 084108–5.
(9) Deng, Y.; Roux, B. *J. Phys. Chem. B* **2009**, *113*, 2234–2246.
(10) Singh, N.; Warshel, A. *Proteins* **2010**, *78*, 1724–1735.
(11) Singh, N.; Warshel, A. *Proteins* **2010**, *78*, 1705–1723.
(12) Fujitani, H.; Tanida, Y.; Matsuura, A. *Phys. Rev. E* **2009**, *79*, 021914–12.
(13) Jayachandran, G.; Shirts, M. R.; Park, S.; Pande, V. S. *J. Chem. Phys.* **2006**, *125*, 084901–12.
(14) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2004**, *126*, 7683–7689.
(15) Mobley, D. L.; Chodera, J. D.; Dill, K. A. *J. Chem. Theory Comput.* **2007**, *3*, 1231–1235.
(16) Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* **2010**, *114*, 10235–10253.
(17) Mobley, D. L.; Dill, K. A. *Structure* **2009**, *17*, 489–498.
(18) Roux, B.; Nina, M.; Pomes, R.; Smith, J.-C. *Biophys. J.* **1996**, *71*, 670–681.
(19) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707–2714.
(20) Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1255–1273.
(21) Chen, P.; Kuyucak, S. *Biophys. J.* **2009**, *96*, 2577–2588.
(22) General, I. J.; Dragomirova, R.; Meirovitch, H. *J. Phys. Chem. B* **2011**, *115*, 168–175.
(23) White, R. P.; Meirovitch, H. *J. Chem. Phys.* **2004**, *121*, 10889–10904.
(24) White, R. P.; Meirovitch, H. *J. Chem. Phys.* **2005**, *123*, 214908–11.
(25) Cheluvaraja, S.; Meirovitch, H. *J. Chem. Phys.* **2005**, *122*, 054903–14.
(26) Cheluvaraja, S.; Meirovitch, H. *J. Chem. Theory Comput.* **2008**, *4*, 192–208.
(27) Cheluvaraja, S.; Mihailescu, M.; Meirovitch, H. *J. Phys. Chem. B* **2008**, *112*, 9512–9522.
(28) Mihailescu, M.; Meirovitch, H. *J. Phys. Chem. B* **2009**, *113*, 7950–7964.
(29) General, I. J.; Meirovitch, H. *J. Chem. Phys.* **2011**, *134*, 025104–17.
(30) Meirovitch, H. *J. Mol. Recognit.* **2010**, *23*, 153–172.
(31) Ponder, J. W. *TINKER − Software Tools for Molecular Design*, version 5.0; Washington University School of Medicine: St. Louis, MO, 2009.
(32) Harrison, R. K.; Stein, R. L. *J. Am. Chem. Soc.* **1992**, *114*, 3464–3471.
(33) Schreiber, S. L. *Science* **1991**, *251*, 283–287.
(34) 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. H.; Maldonado, F.; Barker, J. E.; Bacquet, R.; Villafranca, J. E. *Nature* **1995**, *378*, 641–644.
(35) 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. *Cell* **1995**, *82*, 507–522.
(36) Van Duyne, G. D.; Standaert, R. F.; Karplus, P. A.; Schreiber, S. L.; Clardy, J. *Science* **1991**, *252*, 839–842.
(37) Holt, D. A.; Luengo, J. I.; Yamashita, D. S.; Oh, H.; Konialian, A. L.; Yen, H.; Rozamus, L. W.; Brandt, M.; Bossard, M. J.; Levy, M. A.; Eggleston, D. S; Liang, J.; Schultz, L. W.; Stout, T. J.; Clardy, J. *J. Am. Chem. Soc.* **1993**, *115*, 9925–9938.
(38) Wilson, K. P.; Yamashita, M. M.; Sintchak, M. D.; Rotstein, S. H.; Murcko, M. A.; Boger, J.; Thomson, J. A.; Fitzgibbon, M. J.; Black, J. R.; Navia, M. A. *Acta Crystallogr., Sect. D* **1995**, *51*, 511–521.
(39) Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Pande, V. S. *J. Phys. Chem. B* **2007**, *111*, 13052–13063.
(40) Wang, J.; Deng, Y.; Roux, B. *Biophys. J.* **2006**, *91*, 2798–2814.
(41) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. *Biophys. J.* **2004**, *86*, 67–74.
(42) Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L. *Bioorg. Med. Chem.* **1999**, *7*, 851–860.
(43) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050–9063.
(44) Im, W.; Bernèche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924–2937.
(45) Darden, T. A.; York, D. M.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089–92.
(46) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; , J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; A. Kovalenko, A.; Kollman, P. A. *AMBER 11*; University of California: San Francisco, CA, 2010.
(47) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(48) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(49) Wang, J.; Wolf, R .M.; Calldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Compu. Chem.* **2004**, *25*, 1157–1174.
(50) 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.; Wio2rkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(51) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1959**, *31*, 459–466.
(52) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585–590.
(53) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
(54) Meirovitch, H. *J. Chem. Phys.* **1988**, *89*, 2514–2522.
(55) The practical meaning of an exact method (such as Metropolis Monte Carlo[53]) is that for a given statistical error Δ around the correct value of a property E, there is a sample size n_Δ in which for $n > n_\Delta$ the corresponding error $\Delta_n < \Delta$; thus, practically, the values E_n for $n > n_\Delta$ will converge, i.e., will be equal within the error Δ (one should use this criterion cautiously for systems with rugged conformational space). This criterion also applies to HSMD, where the accuracy of $\ln P_i$ for a ligand depends only on the reconstruction sample size n_f, and in principle a desired accuracy can be obtained by increasing n_f adequately. While in most cases obtaining convergence for $\ln P_i$ will be too time-consuming, entropy *differences* (i.e., in $\langle \ln P_i \rangle$ for the ligand in the protein and solvent environments) converge rapidly.
(56) Meirovitch, H.; Alexandrowicz, Z. *J. Stat. Phys.* **1976**, *15*, 123–127.
(57) Meirovitch, H. *J. Chem. Phys.* **1999**, *111*, 7215–7224.
(58) Hnizdo, V.; Darian, E.; Fedorowicz, A.; Demchuk, E.; Li, S.; Singh, H. *J. Comput. Chem.* **2007**, *28*, 655–668.
(59) Killian, B. J.; Kravitz, J. Y.; Gilson, M. K. *J. Chem. Phys.* **2007**, *127*, 024107–16.
(60) Hnizdo, V.; Tan, J.; Killian, B. J.; Gilson, M. K. *J. Comput. Chem.* **2008**, *29*, 1605–1614.
(61) We distinguish here between a physical time related to the original 2 ns trajectory from which the n_s frames were extracted and a procedural time which is related to the step-by-step reconstruction procedure of a single frame. The words future and past refer to the procedural time, where in the reconstruction of atom k' the past atoms $(1, 2, ..., k'-1)$ are held fixed, while the future atoms $k', k'+1, ...K/2$ and all of the water molecules are free to move in the MD simulation.
(62) Wang, J.; Dixon, R.; Kollman, P. A. *Proteins* **1999**, *34*, 69–81.
(63) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarenden Press: Oxford, U. K., 1987.
(64) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *100*, 9025–9031.
(65) Steinbrecher, T.; Mobley, D. L.; Case, A. C. *J. Chem. Phys.* **2007**, *127*, 21410813.
(66) Hamilton, G.; Steiner, J. *Curr. Pharm. Des.* **1997**, *3*, 405–428.
(67) Åquist, J.; Medina, C.; Samuelsson, J. -E. *Protein Eng.* **1994**, *7*, 385–391.
(68) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. *Biochim. Biophys. Acta* **2006**, *1764*, 1647–1676.
(69) King, E.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647–3661.
(70) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161–185.
(71) Vanommeslaeghe, K; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerel, A. D., Jr. *J. Comput. Chem.* **2010**, *31*, 671–690.
(72) Deng, Y.; Roux, B. *J. Chem. Phys.* **2008**, *128*, 115103–8.