JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

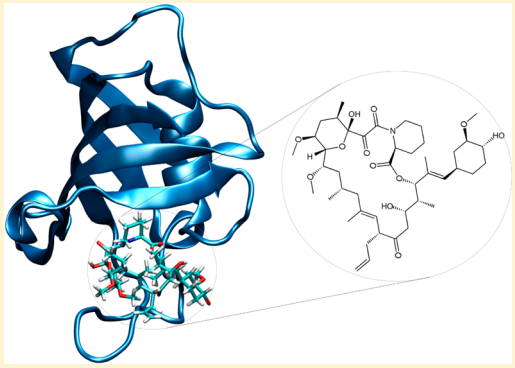
Absolute Free Energy of Binding and Entropy of the FKBP12-FK506 Complex: Effects of the Force Field

Ignacio J. General[†] and Hagai Meirovitch*[,†,‡]

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

[‡]Department of Physics, Bar-Ilan University, Ramat Gan, 52900, Israel

ABSTRACT: The hypothetical scanning molecular dynamics (HSMD) method combined with thermodynamic integration (HSMD-TI) has been extended recently for calculating ΔA^0—the absolute free energy of binding of a ligand to a protein. With HSMD-TI, ΔA^0 is obtained in a new way as a sum of several components, among them is ΔS_{ligand}—the change in the conformational entropy as the ligand is transferred from the bulk solvent to the active site—this entropy is obtained by a specific reconstruction procedure. This unique aspect of HSMD (which is useful in rational drug design) is in particular important for treating large ligands, where ΔS_{ligand} might be significant. Technically, one should verify that the results for ΔS_{ligand} converge—a property that might become more difficult for large ligands; therefore, studying ligands of increasing size would define the range of applicability of HSMD-TI for binding. In this paper, we check the performance of HSMD-TI by applying it to the relatively large ligand FK506 (126 atoms) complexed with the protein FKBP12, where $\Delta A^0 = -12.8$ kcal/mol is known experimentally as well as the crystal structure of the complex. This structure was initially equilibrated by carrying out a 100 ns molecular dynamics trajectory, where the system is modeled by the AMBER force field, TIP3P water, and Particle Mesh Ewald. HSMD-TI calculations were carried out in three conformational regions defined by the intervals [0.2,2], [2,5], and [5,100] ns along the trajectory, where *local* equilibration of the total energy has been observed; we obtained $\Delta A^0 = -13.6 \pm 1.1$, -16.6 ± 1.4, and -16.7 ± 1.4 kcal/mol, respectively indicating the following: (1) The second and third regions belong to the same conformational subspace of the complex, which is different from the [0.2,2] ns subspace. (2) The unsatisfactory result for ΔA^0 obtained in the well equilibrated (hence theoretically preferred) latter regions reflects the nonperfect modeling used, which however (3) has led to the experimental ΔA^0 in the [0.2,2] ns region close to the crystal structure. Keeping the complex near its crystal structure has been a successful approach in the literature. To check this avenue further, we applied harmonic restraints on backbone atoms and obtained unsatisfactory results for ΔA^0, suggesting that implementation of this approach is not straightforward. Converging results for ΔS_{ligand} were obtained in all regions, where the result $\Delta S_{\text{ligand}}([0.2,2]) = 7.1 \pm 1.2$ kcal/mol is less region dependent than ΔA^0 and is relatively large probably due to the large ligand.



I. INTRODUCTION

Calculation of ΔA^0—the absolute (standard) free energy of binding of a ligand to a protein—is of central interest in structural biology and drug design. Therefore, a great deal of work has been done in this area, where various techniques have been developed and applied to a wide range of problems (see refs 1−25 and references cited therein). In particular, it has been desirable to devise highly accurate methods for calculating ΔA^0 based on detailed molecular interactions and rigorous statistical mechanics; such methods are required in the refinement stage of screening procedures for strongly bound ligands based on simplified (fast) scoring functions.

I.1. Conventional Approaches for Calculating ΔA^0. A basic *rigorous* approach for calculating ΔA^0 is based on thermodynamic cycles, where the interactions between the ligand and its environment are decreased to zero in both the active site (the complexed system) and the bulk solution (water), the uncomplexed system, using thermodynamic integration (TI) or free energy perturbation (FEP) procedures. However, it has been argued that this approach, called the double annihilation method (DAM),[1−5,8,12,13] is not straightforward because in the final stages of TI (weak interactions) the ligand might leave the active site starting to "wander" within the inhomogeneous volume (due to the protein), which would make it difficult to obtain converged results; this effect is sometimes called "the end-point problem." However, DAM has been studied extensively by Pande's group[8,13] and Fujitani et al.[12] as applied to a set of ligands bound to the protein FKBP12, where a good agreement with the experiment has been achieved.

Received: June 8, 2013
Published: September 3, 2013

dx.doi.org/10.1021/ct400484u | *J. Chem. Theory Comput.* 2013, 9, 4609−4619

The end-point problem has been rigorously solved by adding harmonic restraints which hold the ligand in the active site during the gradual elimination of the ligand−environment interactions; these restraints are applied by a suitable TI procedure, and the corresponding bias introduced is later removed by releasing the restraints, again with TI. Because of the additional integration steps involved, this procedure is called the "double decoupling method" (DDM).[4,5] DDM has been developed systematically over the past 20 years,[1,9,18−20,24,25] where various implementation issues have been improved in particular by Roux's group (e.g., optimization of typically several force constants of the harmonic restraints), and a large number of complexes have been successfully studied (see a review by Deng and Roux[9]).

ΔA^0 can also be obtained by calculating the potential of mean force (PMF) of a ligand, which is initially far from the protein and is gradually moved to the active site. This approach, first suggested by Jorgensen,[21] has been further developed by Woo and Roux, who have established a rigorous procedure based on restraints and calculation of the PMF by umbrella sampling.[22] They have demonstrated the advantage of this approach for handling a highly charged ligand, which according to them would be extremely difficult to treat with DAM and DDM.

I.2. The HSMD-TI Method. In the past several years, a new general method for calculating the entropy, S, and the conformational Helmholtz free energy, F, has been developed by our group—the "hypothetical scanning molecular dynamics" (HSMD). HSMD has been developed systematically for systems of increasing complexity, liquid argon,[26] TIP3P water,[26] self-avoiding walks,[27] peptides,[28] and mobile loops in proteins.[29−32] To enhance efficiency (for calculating F of water), TI has been incorporated within the framework of HSMD, leading to the combined method, HSMD-TI, which has been extended recently to problems in protein−ligand binding.

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 conformational entropy of the ligand, $S^{A}_{ligand}(i)$, in each frame i is calculated by a reconstruction procedure (in the ligand−protein system, three atoms of the protein are held fixed). (3) The contribution of water and protein to ΔA^0 is obtained 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* (and the protein can move only locally), the end-point problem (encountered with DAM) does not exist and the need for applying restraints (as with DDM) is avoided. With HSMD-TI, ΔA^0 is obtained as a sum of several components, including ΔS_{ligand}—the change in the averages of $S^{A}_{ligand}(i)$ as the ligand is transferred from the bulk solvent to the active site, which is of a particular interest for large flexible ligands, where it might be significant. Another ingredient of ΔA^0 is the external entropy, $S_{external}$, which measures the global movement of the bound ligand in the active site. Calculation of these entropies, which is not shared by other techniques, provides microscopic insights into the binding mechanism and is thus important in rational drug design.

HSMD-TI has already been applied to a complex of the protein FKBP12 (107 residues) with the 68-atom hydrophobic ligand SB3 (also called L8), where the system was modeled by the AMBER force field[33] for the protein, and the General AMBER force field (GAFF)[34] for the ligand with AM1-BCC[35,36] partial charges. The system was immersed in a large container of TIP3P water[37] where long-range electrostatics was considered by periodic boundary conditions with Particle Mesh Ewald (PME).[38] Our result,[39] $\Delta A^0 = -10.7 \pm 1.0$, agrees very well with the experimental values −10.9 and −10.6 kcal/mol.[40,41] HSMD was also applied in two papers to the avidin−biotin complex, but the computation/experiment relation still remains an open question since we have treated biotin as neutral while it is expected to be positively charged.[42,43] Thus, it is still important to investigate the performance HSMD-TI for binding as applied to larger complexes.

I.3. Objectives. The aim of this paper is to check the performance of HSMD-TI for a complex consisting of a relatively large ligand, where the efficiency of the reconstruction of the ligand (which leads to ΔS_{ligand}) is of special interest. More specifically, it is crucial to verify that the results for ΔS_{ligand} converge as a function of the reconstruction time for ligands of an increasing size, where this convergence might be more difficult to achieve; thus, the largest "behaving" ligands in this respect would define the range of applicability of HSMD-TI for binding. A step in this direction is done here (see also the end of section II.4). For that, we use the same force field as in ref 39 and apply HSMD-TI again to the protein FKBP12 but now bound to FK506, a large ligand of 126 atoms, which is a key drug used for immunosuppression in organ transplant. FK506 binds strongly to FKBP12[44] ($\Delta A^0 = -12.8$ kcal/mol[40]), and the FKBP12/FK506 complex, in turn, binds and inhibits calcineurin, thus blocking the signal transduction pathway for the activation of T-cells.[45,46]

It should be noted that crystal structures of FKBP12 in complex with several ligands are available,[40,47,48] and some of their binding constants have been experimentally determined.[40] Therefore, these complexes have served as a rich platform to test and validate different computational strategies for estimating binding free energies.[8,11,13,49−52] In particular, a set of eight ligands complexed with FKBP12 (out of the 27 studied in ref 40) have become the target for absolute free energy calculations by several groups who applied different methods. Wang et al.[24] have used DDM with their models—Spherical Solvent Boundary Potential (SSBP[24]) and Generalized Solvent Boundary Potential (GSBP[53]). Pande's group[8,13,55] and Fujitani et al.[12] used DAM, while in ref 13 some restraint for holding the ligand in the active site has been applied. 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 PME.[38]

While the crystal structure of the FKBP12/FK506 is known, to make it compatible with the force field used, the complex should be equilibrated by MD prior to the application of HSMD-TI. Clearly, for a perfect modeling, a long equilibration is expected to move the system to its correct solution structure, where application of HSMD-TI should lead to the experimental value of ΔA^0. Indeed, relatively long equilibration times have been applied in the literature, where Fujitani et al.,[12] for example, obtained good results for ΔA^0 using a 20 ns equilibration time for complexes of FKBP12, among them FKB12/FK506. However, since in practice the force fields are never perfect, a long equilibration might drive the system to regions in the conformational space that differ from the experimental region. This effect can be somewhat mitigated by equilibrating the complex only *locally* around the crystal structure—an approach that has been adopted, for example,

by Warshel's[11] and Jorgensen's groups[53,56] quite successfully. To check these different approaches, we generated an MD trajectory of 100 ns, ignoring its beginning part for initial equilibration. Then, HSMD-TI was applied in three intervals along this trajectory ([0.2,2], [2,5], and [5,100] ns) where a *local* equilibration of the total energy has been observed; the ([0.2,2] ns region—the closest to the crystal structure—has led to the best ΔA^0 result. However, studying this avenue further by harmonically restraining the heavy backbone atoms of FKBP12 to their crystal structure positions has shown that implementing this approach is not straightforward and should be performed with caution.

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, $K_b = [LP]/[P][L]$, and 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. 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., $\overline{Z}_{PL,N}$ include 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. Theoretical Aspects of HSMD. While the theory of HSMD-TI for binding has been described in detail before,[39,42,43,56] some theoretical aspects should be re-emphasized for the present paper. Thus, assume a system of N particles consisting of configurations i with energies E_i and partition function Z. If the Boltzmann probability, P_i^B [$P_i^B = \exp(-E_i/k_B T)/Z$] of *any configuration* i is known, the free energy, F_i, of i is known, and it is equal to F—the total free energy of the system, $F_i = E_i - TS_i = E_i + k_B T \ln P_i^B = E_i + k_B T(-E_i/k_B T - \ln Z) = -k_B T \ln Z = F$. Thus, the entropy component, S_i, is exactly adjusted to E_i where both lead to F, making it a property with zero fluctuation[57,58] (thus the fluctuation $\sim N^{1/2}$ of the entropy and energy is canceled in F.) To illustrate this situation further, assume an Ising model of N spins, where the ensemble is based on the 2^N different spin configurations i, with energies E_i. At temperature T slightly higher than the critical temperature, T_c, the typical equilibrium configurations are composed of large clusters of equal spins (up or down) leading to zero magnetization. However, $F_i(T) = F(T)$ does not hold only for a configuration of this (typical) region but for any other configuration i, whether it is a random configuration (typical for $T = \infty$) or an ordered configuration with high magnetization (typical at low T, $T < T_c$). Notice, however, that this unique property is not shared by the *average* entropy and energy, which in practice are defined over the typical region at T and in simulations can only be obtained from a Boltzmann sample at T.

This analysis is also applicable to any subspace k where Z_k, hence F_k, is defined only over the configurations defining k. As an example, consider a simple model of a polypeptide in vacuum based on constant bond lengths and bond angles, where a configuration is defined only by the dihedral angles, φ, ψ, and ω. In principle, one can define two regions in conformational space (denoted by us as microstates) for the α-helical and hairpin structures with the corresponding partition functions $Z_{helix}(T)$ and $Z_{hairpin}(T)$ and free energies $F_{helix}(T)$ and $F_{hairpin}(T)$. Now, any *single* conformation i from the α-helix microstate would lead to $F_{helix}(i,T) = F_{helix}$ where the same applies to any i from the hairpin region, $F_{hairpin}(i,T) = F_{hairpin}$. Thus, if one has a way for calculating P_i^B, the difference in the free energy between these polypeptide microstates can be obtained from any two structures of these microstates. The fact that F can be obtained from *any single conformation* can be useful in computer simulations as discussed below and in later sections.

The above discussion remains mainly academic unless one provides a practical way for calculating P_i^B, which is not trivial since P_i^B for a single conformation depends on the entire ensemble through its normalization factor, Z. Therefore, unlike the energy E_i, which can be directly obtained from an MC or MD sample, a direct calculation of P_i^B is not straightforward. In this respect, HSMD is unique as it defines a reconstruction procedure (see section II.4) that enables one calculating a probability P_i which constitutes an approximation for P_i^B; P_i leads to the following relations: (1) The fluctuation of $F(P_i)$ [$F(P_i) = E_i + k_B T \ln P_i$] is not zero, and thus $F(P_i)$ is not calculated from a single configuration but by averaging over a *small* sample of size n. (2) $F(P_i)$ is a lower bound, $F(P_i) \le F(P_i^B)$ [and $S(P_i) \ge S(P_i^B)$], but as the computer time for calculating P_i is increased, P_i can be improved indefinitely, $P_i \to P_i^B$, $F(P_i) \to F(P_i^B)$ [$S(P_i) \to S(P_i^B)$] and the fluctuation of $F(P_i)$ decreases to zero. (3) In practice, one is not interested in the absolute values of the entropy and the free energy but in their differences between two microstates a and b (e.g., an α-helical and a hairpin as discussed above), where the errors in $F_a(P)$ and $F_b(P)$ are typically comparable and are canceled to a large extent in the differences $\Delta F_{ab} = F_a(P) - F_b(P)$ and ΔS_{ab}, see II.4.

II.3. Application of HSMD-TI—General Picture. We delineate here the practical stages involved in calculating ΔA^0, which are applied to both the ligand–solvent and ligand–

protein environments. (1) A small set of n system configurations (frames) are extracted from an initial MD trajectory generated for each environment. (2) The ligand's structure in frame i is reconstructed step-by-step with the help of transition probabilities (TPs, also called conditional probabilities) between successive dihedral and bond angles along the chain; the product of these TPs leads to an approximation P_i for P_i^B (the theoretical basis for the reconstruction process is discussed in detail in section II.2 of ref 39). The *internal* conformational entropy of the ligand, $S^A_{ligand}(i)$ is thus obtained, $S^A_{ligand}(i) = \sim k_B \ln P_i$. The contribution of the internal entropy to ΔA^0 is $T\Delta S_{ligand}$—the difference in the averages of $S^A_{ligand}(i)$ in the two environments, i.e., the change in the ligand's internal entropy as the ligand is transferred from the bulk solvent to the active site. (3) Since the ligand moves in the active site, one has to also calculate its external entropy, $S_{external}$, which is an additional ingredient of ΔA^0. (4) The contribution of water and protein to ΔA^0 is obtained for each frame by gradually increasing the ligand−environment interactions from zero to their full value using TI, where during TI the ligand is fixed in its structure at i; a flowchart of the process appears in ref 42.

II.4. The Reconstruction Procedure for Binding. For simplicity, the reconstruction of the TPs (which lead to P_i) is described as applied to the ligand in water, where the ligand can be any organic molecule. The contribution of the configuration, $\mathbf{x}^N$, of the N water molecules will be carried out later by applying TI.[29−32,39,42,43,56]

Thus, for frame i, $1 \le i \le n$, one defines the TPs in terms of internal coordinates (for a more detailed discussion, see ref 39). First, one determines three reference atoms of the ligand which are kept fixed in their original positions in frame i (the entropic contribution of these atoms, $V8\pi^2$, has already been considered in eqs 1 and 2). The remaining $K/2$ heavy atoms are ordered along the chain and are denoted by k', $k' = 1...K/2$. The related K dihedral and bond angles are denoted $[\alpha_k]$, $k = 1...K$. Since bond stretching is ignored,[59−61,39] the position of atom k' (with respect to atom $k' - 1$) is determined solely by the preceding dihedral and bond angles, α_{k-1} and α_k ($k = 2k'$), respectively. 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.

The TPs are calculated atom-by-atom in $K/2$ steps, where during the process we distinguish between "past" and "future" atoms. Past atoms at step k' are atoms $1, ...k' - 1$ whose TPs have already been determined; the TPs of the future atoms, k', $k' + 1, ...K/2$ should still be determined. To calculate the TP density $\rho(\alpha_{k-1}\alpha_k|\alpha_{k-2}, ..., \alpha_1)$ of atom k' ($k = 2k'$), we carry out an MD run, where the future atoms (k', $k'+ 1, ...K/2$) are allowed to move together with all the N water molecules, while the past atoms are kept fixed at their positions at frame i. By extracting a future conformation (of atoms k', $k' + 1, ...K/2$) every 20 fs, a sample of size n_f future (partial) chains 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 chains to these two bins during the simulation is calculated; one obtains[29−32,39,42,43]

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

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 ($\delta\alpha_{k-1}$, $\delta\alpha_k \to 0$). J is the Jacobian (see ref 39). The product of the TPs (eq 4) defines a probability density, $\rho^{HS}(\alpha_K, ..., \alpha_1)$ related to the ligand's conformation, which leads to an *approximate* entropy functional, $S^A_{ligand} = -k_B \int_m \rho^B_{ligand}([\alpha_k]) \ln \rho^{HS}([\alpha_k])J\, d[\alpha_K]$. S^A_{ligand} is a *rigorous* upper bound for the correct entropy,[26,32] and $\rho^B_{ligand}([\alpha_k])$ is the Boltzmann probability density (if the ligand resides in a microstate m, e.g., an α-helical region, the integration is carried out over m). S^A_{ligand} is estimated by the arithmetic average, $\overline{S}^A_{ligand}$, calculated over the MD (Boltzmann) sample of size n

$$\overline{S}^A_{ligand} = -(k_B/n)\sum_{t=1}^{n} \ln \rho^{HS}(t) \tag{5}$$

The *internal* entropy of the ligand in the active site is calculated in a similar way. First, one determines three reference atoms of the protein far from the active site which are kept fixed in their original positions in frame i (the entropic contribution of these atoms, $V8\pi^2$, has already been considered in eqs 1 and 2). Then, the $K/2$ atoms considered for the ligand in solvent are treated where at step k' the "future" includes atoms k', $k'+ 1, ...K/2$, the solvent (water and counterions), and the protein atoms which are all moved by MD in the reconstruction process (the movement in the active site of the first three atoms of the ligand contributes the external entropy, see next section). For simplicity, we ignore from now on the bar in $\overline{S}^A_{ligand}$ and denote the *average* internal entropies of the ligand in the protein and the solvent 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:

$$\Delta S_{ligand} = S^A_{ligand}(sol) - S^A_{ligand}(p) \quad \text{converged} \tag{6}$$

Thus, we calculate $S^A_{ligand}(sol)$ and $S^A_{ligand}(p)$ for increasing n_f and decreasing bins, verifying that both entropies decrease as the approximation improves; i.e., both approach the correct values from above (notice again that n_f and the bin size are not independent parameters; thus, for a given n_f the bin size should be the smallest for which the number of visits, n_{visit} (eq 4) is statistically sufficient). Typically, the convergence of $\Delta S^A_{ligand}(i)$ is much faster than that of the individual entropies, due to the 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.

However, as the ligand's size increases, the convergence of $S^A_{ligand}(n_f)$ worsens; i.e., the required values of n_f increase as well, which may also affect the convergence of ΔS^A_{ligand}; therefore, it is crucial to verify for each ligand studied that the convergence of ΔS^A_{ligand} has been attained. Moreover, it is of interest to find the range of applicability of HSMD-TI for binding, i.e., the largest ligands that in practice would still lead to converged results of ΔS^A_{ligand}; in this paper, we make a step in this direction.

II.5. The External Entropy. Calculation of the external entropy is described in detail in ref 39, section II.5, including Figure 2 there; here we only provide the technical aspects of

the calculation. This entropy of the ligand in the active site[42,43] is obtained by reconstructing the TPs of the three reference atoms defined for the ligand in solvent, with positions $\mathbf{x}_1(i)$, $\mathbf{x}_2(i)$, and $\mathbf{x}_3(i)$ in the laboratory frame. For that, one also determines three fixed coordinates in the laboratory frame, $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$, where the relative position of $\mathbf{x}_1$ is defined by a "dihedral angle" α_1, "bond" angles α_2, and the distance $r = |\mathbf{x}_1 - \mathbf{y}_3|$ (Figure 2, ref 39). TP(1) is obtained by defining three bins for these variables which replace the two bins in the denominator of eq 4, where n_{visit} is the number of times the three bins are visited *simultaneously*. TP(2) consists of a dihedral and a bond angle (based on $\mathbf{y}_2$, $\mathbf{y}_3$, $\mathbf{x}_1$, and $\mathbf{x}_2$, where $|\mathbf{x}_1 - \mathbf{x}_2|$ is assumed to be constant), and TP(3) is based only on a dihedral angle; $S_{external}(i) = -k_B \ln \mathrm{TP}(i)$, i = 1, 3, and one obtains

$$S_{external} = S_{external}(1) + S_{external}(2) + S_{external}(3) \quad (7)$$

Notice that besides S^A_{ligand}, one should also include in ΔA^0 (eq 2) the contribution of the intraligand energy, $E_{intraligand}(i)$, which is averaged over the samples of the two environments (of size n) leading to $E_{intraligand}(sol)$ and $E_{intraligand}(p)$.

II.6. HSMD-TI. We discuss first the ligand–solvent environment, where the ligand has already been reconstructed, and its structure at i is now fixed. In principle, one could also reconstruct the water configuration in i (in the presence of the fixed ligand structure), which would lead to the probability density of the water configuration, $\rho_{water}(i)$, and to the total probability density, ρ_i of the ligand–water structure, $\rho_i = \rho^{HS}([\alpha_k])\rho_{water}(i)$, where $\rho^{HS}([\alpha_k])$ is the ligand's probability density defined in the paragraph following eq 4. Since the energy of i is known, this would lead to $F_{L,N}$ (eq 2)—the free energy of the system, $F_{L,N} = E_i + k_BT \ln \rho_i$; if one is interested in a subspace, this analysis can be applied, in principle, to *any* ligand–solvent configuration in the subspace, even one with a negligible Boltzmann probability at the T studied, as discussed in section II.2. However, since the reconstruction of the water configuration would be time-consuming, we apply instead the more efficient TI. To calculate $F_{L,N}$ (eq 2), one starts from the system of N waters and the fixed ligand structure (in i) where the ligand–water interactions are zero; the corresponding free energy is thus $F_{0,N} + F_{ligand}(sol,i)$, where $F_{ligand}(sol,i) = S^A_{ligand}(i) + E_{intraligand}(i)$. Then the ligand–water interactions are switched on gradually (by changing a parameter λ) using TI, which leads to $F^{TI}(sol,i)$; ΔF_{sol} (eq 2) becomes

$$\begin{aligned}\Delta F_{sol} &= F_{L,N} - F_{0,N} \\ &= [F_{0,N} + F_{ligand}(sol, i) + F^{TI}(sol, i)] - F_{0,N} \\ &= F^{TI}(sol, i) + F_{ligand}(sol, i)\end{aligned} \quad (8)$$

In the corresponding equation for the protein environment, $F_{ligand}(p,i)$ also includes $S_{external}(i)$

$$\begin{aligned}\Delta F_p &= F_{PL,N} - F_{P,N} \\ &= [F_{P,N} + F_{ligand}(p, i) + F^{TI}(p, i)] - F_{P,N} \\ &= F^{TI}(p, i) + F_{ligand}(p, i)\end{aligned} \quad (9)$$

ΔF_{sol} and ΔF_p in eqs 8 and 9 are obtained for very large n_f, very small bins [where $S^A_{ligand}(i)$ (eqs 5 and 6) and $S_{external}$ for i (eq 7) become exact], and very long integration [leading to accurate $F^{TI}(sol,i)$]. While the index i was introduced originally to define different system configurations (e.g., ligand + water), in eqs 8 and 9 i represents only the ligand's structure. Thus, any ligand structure generated in a Boltzmann sample of the ligand–solvent system (and the ligand–protein–solvent system) will lead to the same ΔF_{sol} (and ΔF_p). More specifically, if each of these systems resides in a subspace, its free energy in the subspace can be obtained from a *single* ligand conformation generated in this subspace. In practice, however, the contribution of different conformations i is expected to slightly fluctuate; thus, while for example the results for ΔS_{ligand} (eq 6) and $F^{TI}(p,i) - F^{TI}(sol,i)$ (eqs 8 and 9) show convergence (see later discussions of Table 1 and Figure 2, respectively), each convergence is within some statistical error. Therefore, the results are averaged over a relatively *small* sample size, $n \leq 20$, where the standard deviation decreases to zero as n_f and the integration time increases indefinitely.

Table 1. Results for the Internal Entropy of the Ligand in Solvent $S^A_{ligand}(sol)$, in the Active Site of the Protein $S^A_{ligand}(p)$, and the Difference $\Delta S^A_{ligand} = S^A_{ligand}(sol) - S^A_{ligand}(p)$, where the Complex Is Locally Equilibrated in the [0.2,2] ns Region[a]

bin size, $\delta = \Delta\alpha_k/l$	n_f	$TS^A_{ligand}(sol)$	$TS^A_{ligand}(p)$	$T\Delta S^A_{ligand}$
$\Delta\alpha_k/90$	2000	−4.4	−11.4	7.0
	4000	−27.9	−35..0	7.1
	6000	−41.1	−47.7	6.6
	8000	−50.1	−57.0	6.9
	10000	−57.4	−64.1	6.7
	12000	**-63.0**	**-70.1**	**7.1**
$\Delta\alpha_k/120$	2000	−4.8	−11.1	6.3
	4000	−27.5	−34.8	7.3
	6000	−41.2	−48.3	7.1
	8000	−50.3	−57.5	7.2
	10000	−57.5	−64.6	7.1
	12000	**−63.1**	**−70.2**	**7.1**
$\Delta\alpha_k/150$	2000	−4.8	−11.5	6.7
	4000	−27.9	−34.8	6.9
	6000	−41.4	−48.3	6.9
	8000	−50.5	−57.5	7.0
	10000	−57.8	−64.7	6.9
	12000	**−63.4**	**−70.5**	**7.1**
converged				**7.1 ± 1.2**

[a]The results were obtained by reconstructing n_s = 10 structures selected homogeneously from MD samples of 1.8 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 = 12 000 are bold-faced. The statistical errors for the results of $TS^A_{ligand}(sol)$ and $TS^A_{ligand}(p)$ are ±1.6 and 1.2 kcal/mol for the solvent and protein, respectively. For [0.2,2] and the other regions where n = 5 was used, the errors are larger ±2.2 for $TS^A_{ligand}(p)$ and ±1.5 kcal/mol for $T\Delta S^A_{ligand}$ (see result in Table 4).

Again, for simplicity the practical calculation is described for the ligand in a solvent. Thus, for *each* of the n frames the ligand–solvent interactions should be increased from zero to their full values using a changing parameter λ. However, it is easier to go in the opposite direction, where these interactions are eliminated gradually by TI. Technically, the ligand charges are eliminated first (using a parameter λ: 1→0) followed by the elimination of the Lennard-Jones interactions using a soft-core potential. These TI results for the n frames are averaged, and the average value constitutes the *water* contribution to ΔF_{sol}

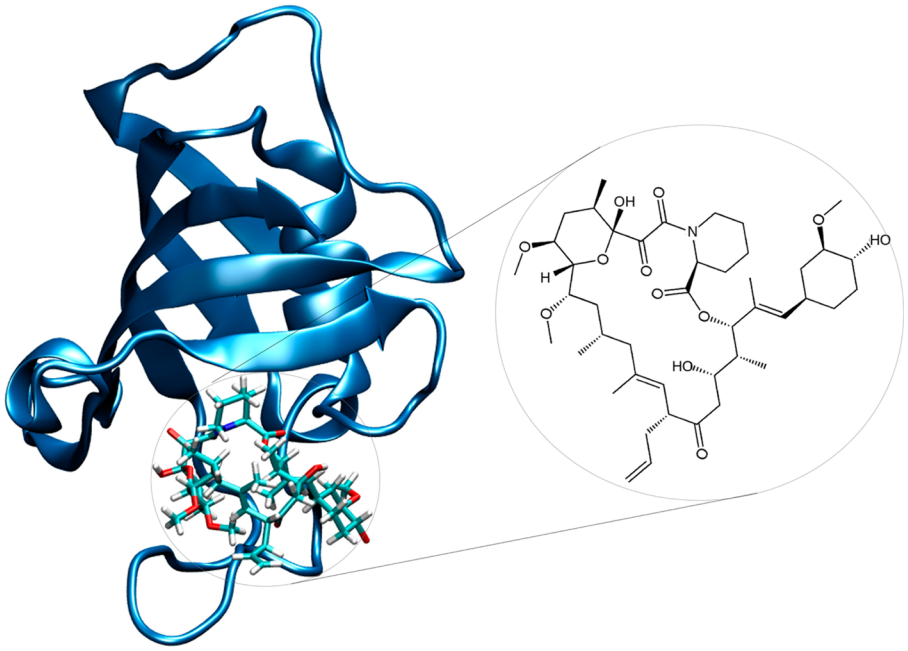


Figure 1. 3D representation of the FKBP12-FK506 complex, with a 2D image of the ligand.

(eqs 2 and 8). Exactly the same procedure is applied to each of the n frames of the complex, which leads to the TI contribution to ΔF_p (eqs 2 and 9). We emphasize again that with HSMD-TI the elimination of the ligand−environment interactions is performed for the *fixed* ligand structure defined by each of the frames, while the other atoms of the system (water, protein, and counterions) move in the MD simulations; this differs from the DAM and DDM methods where during TI the ligand moves as well.[4,5] The sum of all the components described above leads to ΔA^0 (eq 2):

$$\begin{aligned}\Delta A^0 &= \Delta F_p - \Delta F_{sol} + k_B T \ln(8\pi^2 V^0) \\ &= [E_{intraligand}(p) - TS_{ligand}(p) - TS_{external} + F^{TI}(p)] \\ &\quad - [E_{intraligand}(sol) - TS_{ligand}(sol) + F^{TI}(sol)] \\ &\quad + k_B T \ln(8\pi^2 V^0) \\ &= \Delta E_{intraligand} - T\Delta S_{ligand} - TS_{external} + \Delta F^{TI} \\ &\quad + k_B T \ln(8\pi^2 V^0) \end{aligned} \quad (10)$$

where all the quantities are averages over n snapshots and Δ denotes differences in the corresponding variables (protein and solvent); the errors of ΔA^0 and its different components are obtained from the standard deviations (sd) divided by $(n)^{1/2}$.

II.7. Modeling and Optimization of the Systems. The FKBP12−FK506 complex was treated with the AMBER 10 package,[62] where the protein was modeled by the AMBER99 force field[33] and the ligand by GAFF[34] (with the AM1-BCC partial charges[35,36]) where long-range electrostatics were taken into account by PME.[38] The unit cell of the periodic system (based on pdb code 1fkf, 1991 atoms[46]) was defined by constructing a truncated octahedron around the protein which was filled with 3592 water molecules. The minimum distance between the protein and the walls of the cell was 8 Å. (This cell size is comparable to that used in previous studies of FK506; also, to be compatible with these studies, the three histidine residues have positive charge.) To neutralize the system, 19 randomly selected water molecules were replaced by 10 chloride and nine sodium ions. As mentioned above, the periodic boundary conditions were defined by the PME algorithm,[38] with a cutoff of 10 Å.

The system was first energy minimized, with 10^4 steepest descent and 10^4 conjugate gradient steps. Then, using MD, it was heated to 298 K over 100 ps, followed by a 50 ps *NPT* run, where the temperature was kept at 298 K and pressure at 1 atm. The Berendsen thermostat[63] used in these procedures had a time constant of 1.5 ps, and the weak coupling isotropic barostat had a relaxation time of 2 ps. The time step employed was 2 fs, and the SHAKE[63] algorithm was applied to all hydrogens in all of our studies. Finally, a 100 ns constant volume production run at 300 K (time step of 2 ps) was carried out, where the first 0.2 ns trajectory was used for an initial equilibration. Three intervals along the trajectory were defined ([0.2,2], [2,5], and [5,100] ns; from now on the term "ns" will be omitted), where for each one a set of up to n = 10 equally separated frames was extracted and used for our HSMD-TI analysis. For each of the extracted frames, and for the rest of the analysis, three successive atoms of the protein were fixed at their positions in structure i, in order to remove the center-of-mass translation and rotation.

We have also applied HSMD-TI to a restricted complex, where the heavy backbone atoms of the protein are harmonically restrained to their crystal structure positions by force constants, k = 1 and 10 kcal/(mol·Å^2). In these cases, the restraints were applied also during the initial equilibration, which was carried out with the same parameters and times used in the nonrestricted case described above.

A similar procedure was applied to the solvent system consisting of FK506 without the protein, which was solvated with 754 water molecules in a truncated octahedron. No counterions were added to this neutral system, where 10 frames were extracted for its analysis.

III. RESULTS AND DISCUSSION

III.1. Details about the Reconstruction. FK506 is a cyclic compound containing three rings, one being three atoms away from the circumference of the main cycle (Figure 1). One of the rings places itself deepest in the active site. Therefore, we defined one of the carbon atoms in this ring as the first reference atom (with coordinates $\mathbf{x}_1$, see section II.5). This allows for the definition of a physically meaningful external entropy which is best related to the global movement of FK506 in the active site (see ref 39, end of section II.5).

Details about the entropy calculation are given for the [0.2,2] region, where n = 10. FK506 conformations (denoted i) are reconstructed in both the solvent and the protein; 57 atoms (k' = 1, ..., 57) and 114 angles α_k ($1 \le k \le K = 114$) participate in this reconstruction (see section II.4). Each reconstruction step (out of 57) starts from conformation i with a 240 ps production run where a future FK506 conformation is stored every 20 fs for a later analysis; thus, the total sample for each step consists of n_f = 12 000 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 57 TPs is the distribution, ρ^{HS} (see paragraph following eq 4), which leads to the internal entropy, S^A_{ligand} (eqs 5 and 6), and $S_{external}$ (eq 7). More details about the reconstruction appear in refs 32, 39, 42, and 43. A free program with a tutorial and explanations for calculating the TPs, S^A_{ligand} (eqs 5 and 6) and $S_{external}$ (eq 7), appears at http://www.ccbb.pitt.edu/Faculty/meirovitch/reconstruction-web/reconstruction-web.html.

III.2. Results for the Internal Entropy. Results for TS^A_{ligand} (eqs 5 and 6) are presented in Table 1. It should be noted that the angles (α_k) are calculated in radians, which can lead to negative entropies. This is not unexpected as TS^A_{ligand} is defined up to an additive constant, and we are interested only in entropy differences $T\Delta S_{ligand}$ (eq 6) where this constant cancels out. As expected by theory, the results for TS^A_{ligand} decrease systematically as the approximation improves, i.e., with increasing n_f; however, these results have not converged even for n_f = 12 000. On the other hand, for each bin size, $\delta = \Delta\alpha_k/l$ studied (l = 90–150), the corresponding results for $T\Delta S^A_{ligand}$ show convergence to the same value. Thus, we consider $T\Delta S_{ligand}(l = 90) = 7.1 \pm 1.2$ kcal/mol to be the exact result within the error bars. This is a relatively large difference between the internal conformational entropy in the two environments, which also reflects the relatively large size of FK506; this should be compared to the much smaller value $\Delta S_{ligand} = 1.7. \pm 0.2$ kcal/mol we obtained for the smaller ligand, L8.[39] The results in the other regions for TS^A_{ligand} and $T\Delta S^A_{ligand}$ behave in a similar way to those in the [0.2,2] region where the converged values, $T\Delta S_{ligand}$, are comparable (see Table 4).

III.3. Results for the External Entropy. The contributions of the three reference atoms (i) to the external entropy $TS_{external}(i)$ (for the [0.2,2] region) appear in Table 2 as a function of n_f and only for a single bin size, $\Delta\alpha_k/90$, as similar results were obtained for other bin sizes. As expected, for each atom the entropy decreases as the approximation improves; however, for i = 1, 2, the results show convergence; i.e., they are equal for n_f = 10 000 and 12 000. On the other hand, for $TS_{external}(3)$ the results are still decreasing even for n_f = 12 000, which leads to the corresponding decrease in the results for the total external entropy, $TS_{external}$ (eq 7). Clearly, nonconverging results might be problematic as they affect the result for ΔA^0. However, since the above decrease is very slight, one can extrapolate that the convergence of $TS_{external}$ will occur no more than 0.3 kcal/mol below the result for n_f = 12 000. While this is a relatively small value, to emphasize the importance of the $TS_{external}$ convergence, in Table 4 we subtract 0.3 kcal/mol from the result for $TS_{external}(n_f = 12\,000)$. A total of 0.3 kcal/mol is also subtracted from all the $TS_{external}(n_f = 12\,000)$ values obtained for the other regions where a similar behavior of the $TS_{external}(i)$ results has been observed.

Table 2. External Entropy for the Three Reference Atoms for the [0.2,2] ns Region[a]

bin size, δ	n_f	$TS^A_{ligand}(1)$	$TS^A_{ligand}(2)$	$TS^A_{ligand}(3)$	total ≡ $TS_{external}$
$\Delta\alpha_k/90$	2000	2.4	0.0	0.6	3.0
	4000	2.4	−0.5	0.2	2.1
	6000	2.4	−0.8	−0.1	1.5
	8000	2.4	−0.9	−0.3	1.2
	10000	2.3	−1.0	−0.4	0.9
	12000	**2.3**	**−1.0**	**−0.5**	**0.8**
$\Delta\alpha_k$	12000	3.1	−0.3	−0.3	2.6
space covered		175.3 Å^3	0.6		
			out of 4π = 12.6 (4.8%)	11.4°	

[a]The results are based on the 10 structures of FK506 in the active site used to calculate the internal entropy (Table 1); they are presented for bin sizes $\delta = \Delta\alpha_k/l$, l = 90, and l = 1 (eq 3; results for l = 30, 60, etc. were found to be the same as for l = 90). The results are presented for different values of n_f (eq 4)—the sample size of the future chains—with a maximal value of n_f = 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 $S^A_{ligand}(3)$ is the entropy related to dihedral angle of atom 3. The (best) results, i.e., those for the largest n_f, n_f = 12 000, are bold-faced. The results for $S^A_{ligand}(1)$ and $S^A_{ligand}(2)$ show convergence, unlike those for $S^A_{ligand}(3)$ which are still slightly decreasing leading to the corresponding decrease of the results for $TS_{external}$. Such a decrease of $TS_{external}$ has been found for all the regions; therefore, in Table 4 we extrapolate all the results for $TS_{external}$ by decreasing them further by 0.3 kcal/mol (see text). In the last row, we provide the space covered by the variables of each atom. The error in $TS_{external}$ is thus ±0.3 kcal/mol.

The space covered by these variables can be estimated from eq 4, using the results obtained for the largest bin, $\Delta\alpha_k$ (eq 3), for which n_{visit}/n_f = 1, or more specifically by calculating $\exp(TS^A/0.6)$, where $TS^A(n_{visit}/n_f = 1)$ stands for the results in the table and k_BT = 0.6 at 300 K. Thus, atom 1 visits a volume of ∼175.3 Å^3 where the "bond angle" and "dihedral angle" related to atom 2 cover together 4.8% of 4π (=12.6) and the range of change of the dihedral angle of atom 3 is ∼11.4°. However, the results for atoms 2 and 3 should be considered as lower bounds since the predecessor atoms, 1 and 2, respectively, are held fixed.

To check the effect of the number of frames on the results, we have also carried out calculations in the protein environment for a smaller sample of n = 5. We have found that both samples (n = 5 and 10) in the [0.2,2] region lead to the same value ΔA^0 within the error bars (see Table 4). Therefore, n = 5 was also used in the complex calculations of the [2,5] and [5,100] regions and in simulations based on harmonic restraints. Notice that the results in the solvent environment are always based on n = 10; these results have been found to be

dx.doi.org/10.1021/ct400484u | J. Chem. Theory Comput. 2013, 9, 4609–4619

stable along a long trajectory. All of these entropy calculations are summarized in Table 4 and will be discussed later.

III.4. TI Results. To each of the $n = 10$ frames of the solvent sample, we applied a TI procedure where the ligand−water interactions were turned off gradually for a *fixed* FK506 structure; for the 10 frames of the protein environment, we decoupled both the ligand−solvent and ligand−protein interactions. Using a parameter λ, $0 \le \lambda \le 1$, the electrostatic interactions were decoupled first (by decreasing the ligand's charge to zero) 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[32,64] with δ = 3 Å^2. For a given frame (i), each integration step (window) always started from the (initial) structure of i using the corresponding step potential energy $[E(\lambda)]$, followed by a 2 ns production run, where the initial 20 ps are discarded to equilibrate the system with respect to λ. As demonstrated in Figure 2, this relatively large production run for each window is required for equilibrating the result for F^{TI} (LJ).

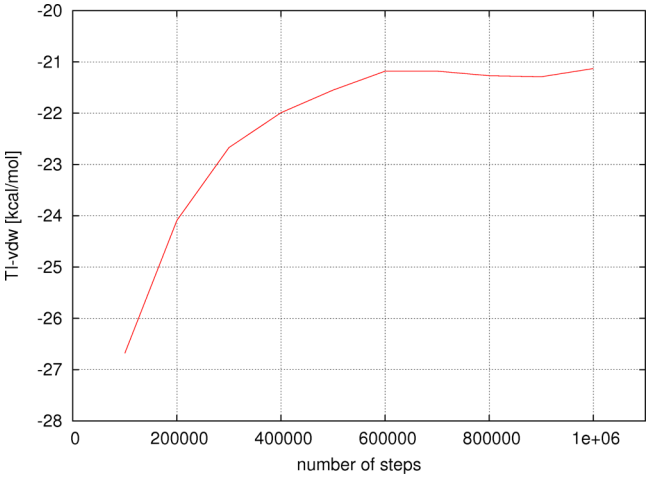


Figure 2. A plot describing the convergence of F^{TI} (LJ) as a function of the number of TI steps.

The results in Table 3 for ΔF^{TI}(ch), ΔF^{TI}(LJ), and ΔF^{TI} for n = 5 and 10 (for the [0.2,2] region) are very close with a maximal deviation of 0.7 kcal/mol for ΔF^{TI}; these results for the regions [2,5] and [5,100] are also relatively close with deviations of 0.9, 0.9, and 1.9 kcal/mol for ΔF^{TI} (ch), ΔF^{TI}(LJ), and ΔF^{TI}, respectively. The two sets of results obtained for harmonic restraints differ more significantly from each other with a maximum deviation of 6.6 kcal/mol for ΔF^{TI}. Ignoring the results for k = 1, which are significantly smaller than the rest the values of ΔF^{TI}(ch), ΔF^{TI}(LJ), and ΔF^{TI} in the table, these values differ within the ranges of 2.8, 3.0, and 4.8 kcal/mol, respectively.

III.5. Results for ΔA^0. Table 4 presents the various energetic and free energy components whose sum leads to ΔA^0 (see eq 10). First, notice that the results for $TS_{external}$ are smaller by 0.3 kcal/mol from the raw values (see section III.3). The values in the table are comparable for the regions [0.2,2], [2,5], and [5,100] (ranging from 0.6 to 1.2) but are smaller and

Table 3. Free Energy Components Obtained by TI (in kcal/mol) for Different Regions of the [0.2,100] ns Trajectory[a]

region	ΔF^{TI} (ch)	ΔF^{TI} (LJ)	ΔF^{TI}
[0.2,2] ns (n = 5)	−5.2	−18.2	−23.4
[0.2,2] ns (n = 10)	−4.9	−17.8	−22.7
[2,5] ns	−6.8	−17.9	−24.7
[5,100] ns	−7.7	−19.8	−27.5
restraints k = 10	−6.5	−20.8	−27.3
restraints k = 1	−9.9	−24.0	−33.9

[a][0.2,2], [2,5], [5,100] ns stand for regions along the 100 MD trajectory (with no restraints). The results in the solvent environment are based on a sample of n = 10 structures, while in most cases n = 5 is used for the complex; only for the [0.2,2] ns region are the complex results for n = 10 also provided. k [in kcal/(mol·Å^2)] is the force constant for the harmonic restraints applied to the backbone heavy atoms. Δ is the difference between the complex and solvent results. ΔF^{TI}(ch) and ΔF^{TI}(LJ) were obtained by thermodynamic integration for the electrostatic (ch) and Lennard-Jones interactions, respectively; ΔF^{TI} is their sum. T = 300 is the absolute temperature. The statistical errors for [0.2,2], n = 10, are ±1.0, ±2.2, and ±1.0 kcal/mol for ΔF^{TI} (ch), ΔF^{TI} (LJ), and ΔF^{TI}, respectively. The corresponding errors for the other cases (n = 5) are up to 0.3 kcal/mol higher.

larger by 1.8 kcal/mol from these values for k = 10 and 1, respectively. We have already pointed out that the two sets of results in Table 3 for the [0.2,2] region are slightly different with a lower $\Delta F^{TI}(n = 5)$ value. However this value is compensated by the higher values for $T\Delta S_{ligand}(n = 5)$ and $-TS_{external}(n = 5)$, and thus the results for ΔA^0 for n = 5 and n = 10 become equal within the error bars, which justifies our use of n = 5 for most of the calculations in the protein environment. The value $\Delta A^0(n = 10) = -13.6 \pm 1.1$ is equal within the error bars to the experimental value −12.8 kcal/mol.[40] For comparison, Fujitani et al.[12] using extensive simulations obtained ΔA^0 = −10.1 and −12.1 kcal/mol for the force fields AMBER99+GAFF+AM1BCC and FF-FOM(GAFF+RESP), respectively, demonstrating the effect of the force field on ΔA^0. An early work of Pande's group[8] has led to ΔA^0 = ∼−9.5 kcal/mol, while more recently this group obtained $\Delta A^0 = -11.8 \pm 1.14$ using AMBER99+AM1BCC and GAFF.[13] Finally, Wang et al.[51] using SSBP[53] and GSBP[23] obtained $\Delta A^0 = -10.1 \pm 1.2$ and -10.8 ± 3.0 kcal/mol when starting from the X-ray structure of the complex or from an optimal structure of it obtained by MD, respectively.

The corresponding components for the [2,5] and [5,100] regions are again different but a similar compensation as described above also occurs for them. More specifically, the lower values of ΔF^{TI}([5,100]) and $\Delta E_{intraligand}$([5,100]) appear with a lower internal entropy, $\{T\Delta S_{ligand}([5,100]) > T\Delta S_{ligand}([2,5])\}$, and lower external entropy, $\{TS_{external}([5,100]) < TS_{external}([2,5])\}$; i.e., the ligand in the [5,100] region is the less flexible. As has been found for the [0.2,2] region, this compensation leads to the same total free energy, $\Delta A^0 = -16.7 \pm 1.4$ kcal/mol for both the [2,5] and [5,100] regions, where however, this result is significantly lower (by 3.9 kcal/mol) than the experimental value.

In sections II.2 and II.6, we have argued that for a given subspace, ΔF_{sol} and ΔF_p can in principle be obtained from any single configuration i where the extensive reconstruction and integration are carried out only over this space (for less extensive simulations averaging over a small samples is performed.) Thus, the fact that the [2,5] and [5,100] regions lead to the same ΔA^0 suggests that these regions belong to the

dx.doi.org/10.1021/ct400484u | J. Chem. Theory Comput. 2013, 9, 4609−4619

Table 4. Components of the Free Energy of Binding and Their Sum, ΔA^0 (in kcal/mol) for the Different Regions along the [0.2,100] ns Trajectory[a]

region	$\Delta E_{\text{intraligand}}$	ΔF^{TI}	$-T\Delta S_{\text{ligand}}$	$-TS_{\text{external}}$	$k_BT \ln(8\pi^2V^0)$	total $\equiv \Delta A^0$	ΔA^0 (exp.)
[0.2,2] ns (n = 5)	−4.5	−23.4	7.4	−0.3	7	−13.8	−12.8
[0.2,2] ns (n = 10)	−4.5	−22.7	7.1	−0.5	7	−13.6	−12.8
[2,5] ns	−3.4	−24.7	5.3	−0.8	7	−16.6	−12.8
[5,100] ns	−4.0	−27.5	8.1	−0.3	7	−16.7	−12.8
restraints k = 10	4.5	−27.3	8.6	1.5	7	−5.7	−12.8
restraints k = 1	−3.2	−33.9	7.6	−2.6	7	−25.1	−12.8

[a]The results in the solvent environment are based on a sample of n = 10 structures, while in most cases n = 5 is used for the complex; only for the [0.2,2] ns region are the complex results for n = 10 also provided. Δ means complex minus solvent results. $\Delta E_{\text{intraligand}}$ is the intraligand energy. ΔF^{TI} was obtained by thermodynamic integration (see Table 3), and T = 300 K is the absolute temperature. ΔS_{ligand} is the difference in internal entropies of the ligand. The results for the external entropy, S_{external}, are smaller by 0.3 kcal/mol from those obtained for n_f = 12 000 (see section III.3). The sum of the various components leads to ΔA^0 (see eq 10). ΔA^0 (exptl) is the experimental value.[40] The statistical errors for [0.2,2], n = 10, are 0.7 and 1.1 kcal/mol for $\Delta E_{\text{intraligand}}$ and ΔA^0, respectively. The corresponding errors for the other cases (n = 5) are up to 0.3 kcal/mol larger.

same structural subspace of the protein−ligand complex. The results for the entropy seem to contradict this conclusion as $T\Delta S_{\text{ligand}}([5,100]) - T\Delta S_{\text{ligand}}([2,5])$ = 2.8 kcal/mol; however, as discussed in section II.2, unlike the free energy, the entropy should be averaged over the entire [2,100] trajectory, while the above results are based on small and somewhat localized samples of n = 5 structures each.

It should be emphasized that the errors of ΔA^0 in the various regions are based on the assumption that the corresponding results for $T\Delta S_{\text{ligand}}$ and ΔF^{TI} and other quantities have been converged within some statistical uncertainties which define the errors for ΔA^0 (see section II.6). Thus, the significantly different ΔA^0 values, 13.6 ± 1.1 and 16.7 ± 1.1, obtained in the [0.2,2] and [2,100] regions, respectively, suggest that these regions belong to different conformational subspaces. On the basis of statistical mechanics, the complex is expected to be better optimized in the [5,100] region than in the [0.2,2] region and thus leads to the better results for ΔA^0; the fact that this is not materialized probably stems from deficiencies in the force field used. However, such deficiencies seem to become ineffective to a large extent in the [0.2,2] region where a very good result for ΔA^0 is obtained based on structures of the complex which are close to the crystal structure. While it is not clear whether this is an inherent property of this force field, a partial support is provided by our earlier study of FKBP12-L8 (ref 39) where a very good result for ΔA^0 was obtained with the present force field and a relatively short equilibration of [0.4,2] ns; however, the effect of a longer equilibration was not studied there. The present results, while based only on a single complex, pose a practical dilemma for a researcher in the field, whether to perform a global optimization or to prefer equilibrating locally close to the crystal structure.

Both approaches have been adopted in the literature. For example, Fujitani et al., who calculated the absolute binding free energy (ΔA^0) of FKBP12 complexed with various ligands, used 20 ns initial MD simulations for equilibrating the complex.[12] On the other hand, in calculations of *relative* free energy of complexes of FKBP12, Lamb and Jorgensen[56] used an MC sampling procedure based on "variations of all bond angles and dihedrals of the ligand and protein side chains as well as overall rotation and translation of the ligand and water molecules. *The protein backbone atoms were held fixed in their crystallographic positions.*" In a following paper of this group,[53] the *absolute* binding free energy (ΔA^0) of several FKBP12 complexes was studied using the same methodology and modeling, with a very good agreement with the experiment. Very good results for the absolute free energy of binding were obtained also by Singh and Warshel,[10] who applied weak (distance) harmonic restraints [k = 0.01 kcal/(mol·Å^2)] to the crystallographic positions of all atoms within a spherical radius of 18 Å around the center of the active site; a metal ion was restrained there stronger with k = 10 kcal/(mol·Å^2).

To investigate the latter avenue, we have applied HSMD-TI to two models of FKBP12-FK506 where the backbone atoms of FKBP12 are restrained to their crystallographic positions using k = 1 and k = 10 kcal/(mol·Å^2). Most of the results for k = 10 in Table 4 are comparable to their counterparts besides for the positive values, $\Delta E_{\text{intraligand}}$ = 4.5 (which probly reflects difficulty in optimizing $E_{\text{intraligand}}$ due to the strong restraints and $-TS_{\text{external}}$ = 1.5, which led to $\Delta A^0(k = 10)$ = −5.7 kcal/mol—the highest overestimation of the experimental value in the table. The flexibility of the ligand in the active site is the lowest, as its internal (and external) entropy is the smallest, $T\Delta S_{\text{ligand}}$ = 8.6; this lower flexibility probably makes it difficult for the ligand to optimize its intraenergy, $E_{\text{intraligand}}$, which is reflected in the positive value of $\Delta E_{\text{intraligand}}$. On the other hand, a much better optimization occurs with the weaker restraints based on k = 1 kcal/(mol·Å^2) where $\Delta E_{\text{intraligand}}$ indeed becomes relatively low closer to the other results.

For k = 1, the outliers are the lowest value in Table 4, ΔF^{TI} = −33.9, and the highest external entropy, TS_{external} = 2.6 kcal/mol; these results constitute the main factors for the extremely low free energy of binding, $\Delta A^0(k = 1)$ = −25.1, which became even somewhat smaller by increasing the sample to n = 10. Thus, while one might suggest that the optimal k should lie somewhere between 1 and 10 kcal/(mol·Å^2) finding its value is beyond the aims of this study. We just show that determining the correct strength of the harmonic restraint is not straightforward and probably needs an early calibration.

IV. SUMMARY AND CONCLUSIONS

In this work, we have used HSMD-TI for calculating the absolute free energy of binding (ΔA^0) of the complex FKBP12−FK506 consisting of a large ligand. With HSMD-TI, ΔA^0 is obtained as a sum of several components, among them ΔS_{ligand}—the difference in the internal conformational entropy of the ligand as it is transferred from the bulk solvent to the active site; ΔS_{ligand} is important in rational drug design and its calculation (based on a reconstruction procedure) is unique to HSMD-TI. Thus, an essential goal of this study has been to examine the performance of the reconstruction procedure for a large ligand. Calculations in three regions

along a [0.2,100] initial equilibration MD trajectory have led to converging results for $T\Delta S_{ligand}$, where all of them are relatively large probably due to the large size of FK506; for the [0.2,2] region, we obtain $T\Delta S_{ligand}$ = 7.1 ± 1.2 kcal/mol, a value that cannot be corroborated by the experiment, as experimental results for $T\Delta S_{ligand}$ are unavailable.

Unlike the entropy, for many complexes, including FKBP12−FK506, the experimental ΔA^0 is known, which can help evaluate the effects of the simulation method, the extent of sampling, as well as the quality of the force field used. Thus, with a *perfect* force field, an extensive equilibration of the (solvated) crystal structure would drive it to the (typically close) solution structure, where application of HSMD-TI is expected to lead to the experimental ΔA^0. Therefore, the unsatisfactory result, ΔA^0 = −16.7 kcal/mol obtained in the longest part of the trajectory ([2,100]) probably reflects limitations in the present force field, which, however, become almost ineffective in the [0.2,2] region, i.e., closer to the crystal structure, where ΔA^0 = −13.6 kcal/mol. As was pointed out earlier, restricting the complex close to the crystal structure has become a practical and successful approach by groups who used different force fields, simulation methods, boundary conditions, etc.[10,11,52,55] However, this approach should be adopted with caution since the required strength of the harmonic restraints is not known a priori, as is demonstrated by our unsatisfactory results based on relatively strong and weak restraints. Still, in some studies in the literature, extensive equilibration has been applied, which has led to good agreement with the experiment, suggesting that the sensitivity of the results to the initial equilibration depends on both the specific modeling and the system in hand. Also, as was pointed out in previous studies, long "production" simulations might be needed for a flexible system to adequately sample the different microstates that the system visits.[7,15,17]

In summary, we have shown that HSMD-TI can successfully handle a complex with a large ligand. We have emphasized the sensitivity of the calculated ΔA^0 to the quality of modeling and discussed limitations in the force field used, which are probably typical to other force fields, which currently do not allow recovering the experimental ΔA^0 with an accuracy of 1 kcal/mol—the target accuracy accepted in the field. This is also an essential conclusion of the several SAMPL (Statistical Assessment of Modeling of Proteins and Ligands) blind tests for calculating ΔA^0 carried out in recent years (see refs 65−67 and references cited therein). Therefore, while rigorous methods for calculating ΔA^0 are already available, a major stumbling block for the theoretical prediction of the standard free energy of binding is the lack of a highly accurate modeling, and efforts should be devoted in the community for improving the existing force fields. As for HSMD-TI, it would be of interest to examine its validity for larger ligands than FK506, where the effect of the extent of equilibration on the results for ΔA^0 will be studied further.

AUTHOR INFORMATION

Corresponding Author

*E-mail: hagaim@pitt.edu.

Notes

The authors declare no competing financial interest.

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) Jorgensen, W. L. *J. Am. Chem. Soc.* **1989**, *111*, 3770−3772.
(22) Woo, H. -J.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6825−6830.
(23) Im, W.; Bernèche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924−2937.
(24) Wang, J.; Deng, Y.; Roux, B. *Biophys. J.* **2006**, *91*, 2798−2814.
(25) Jiang, W.; Roux, B. *J. Chem. Theory Comput.* **2010**, *6*, 2559−2565.
(26) White, R. P.; Meirovitch, H. *J. Chem. Phys.* **2004**, *121*, 10889−10904.
(27) White, R. P.; Meirovitch, H. *J. Chem. Phys.* **2005**, *123*, 214908−11.
(28) Cheluvaraja, S.; Meirovitch, H. *J. Chem. Phys.* **2005**, *122*, 054903−14.
(29) Cheluvaraja, S.; Meirovitch, H. *J. Chem. Theory Comput.* **2008**, *4*, 192−208.
(30) Cheluvaraja, S.; Mihailescu, M.; Meirovitch, H. *J. Phys. Chem. B* **2008**, *112*, 9512−9522.
(31) Mihailescu, M.; Meirovitch, H. *J. Phys. Chem. B* **2009**, *113*, 7950−7964.
(32) General, I. J.; Meirovitch, H. *J. Chem. Phys.* **2011**, *134*, 025104−17.
(33) 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.
(34) Wang, J.; Wolf, R. M.; Calldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(35) Jakalian, A.; Bush, B. L.; Jack, D B.; Bayly, C. I. *Comput. Chem.* **2000**, *21*, 79−157.
(36) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623−41.
(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(38) Darden, T. A.; York, D. M.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089−92.
(39) General, I. J.; Dragomirova, R.; Meirovitch, H. *J. Chem. Theory Comput.* **2011**, *7*, 4196−4207.

(40) 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.
(41) Hamilton, G.; Steiner, J. *Curr. Pharm. Des.* **1997**, *3*, 405−428.
(42) General, I. J.; Dragomirova, R.; Meirovitch, H. *J. Phys. Chem. B.* **2011**, *115*, 168−175.
(43) General, I. J.; Dragomirova, R.; Meirovitch, H. *J. Phys. Chem. B.* **2012**, *116*, 6628−6636.
(44) 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.
(45) 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.
(46) Van Duyne, G. D.; Standaert, R. F.; Karplus, P. A.; Schreiber, S. L.; Clardy, J. *Science* **1991**, *252*, 839−842.
(47) Van Duyne, G. D.; Standaert, R. F.; Karplus, P. A.; Schreiber, S. L.; Clardy, J. *Science* **1991**, *252*, 839−842.
(48) 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.
(49) 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: Biol. Crystallogr.* **1995**, *51*, 511−521.
(50) Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Pande, V. S. *J. Phys. Chem. B* **2007**, *111*, 13052−13063.
(51) Wang, J.; Deng, Y.; Roux, B. *Biophys. J.* **2006**, *91*, 2798−2814.
(52) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. *Biophys. J.* **2004**, *86*, 67−74.
(53) Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L. *Bioorg. Med. Chem.* **1999**, *7*, 851−860.
(54) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050−9063.
(55) Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Pande, V. S. *J. Phys. Chem. B* **2007**, *111*, 13052−13063.
(56) Lamb, M. L.; Jorgensen, W. L. *J. Med. Chem.* **1998**, *41*, 3928−3939.
(57) Meirovitch, H. *J. Mol. Recognit.* **2010**, *23*, 153−172.
(58) Meirovitch, H.; Alexandrowicz, Z. *J. Stat. Phys.* **1976**, *15*, 123−127.
(59) Meirovitch, H. *J. Chem. Phys.* **1999**, *111*, 7215−7224.
(60) Hnizdo, V.; Darian, E.; Fedorowicz, A.; Demchuk, E.; Li, S.; Singh, H. *J. Comput. Chem.* **2007**, *28*, 655−668.
(61) Killian, B. J.; Kravitz, J. Y.; Gilson, M. K. *J. Chem. Phys.* **2007**, *127*, 024107−16.
(62) Hnizdo, V.; Tan, J.; Killian, B. J.; Gilson, M. K. *J. Comput. Chem.* **2008**, *29*, 1605−1614.
(63) 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.; J. Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 11*; University of California: San Francisco, CA, 2010.
(64) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*: Clarenden Press: Oxford, United Kingdom, 1987.
(65) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *100*, 9025−9031.
(66) Skillman, A. G.. *J. Comput. -Aided Mol. Des.* **2012**, *26*, 473−474.
(67) Muddana, H. S.; Varnado, C. D.; Bielawski, C. W.; Urbach, A. R; Isaacs, L.; Geballe, M. T.; Gilson, M. K. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 475−487.
(68) Sulea, T.; Hogues, H.; Purisima, E. O. *J. Comput. -Aided Mol. Des.* **2012**, *26*, 617−633.