Absolute Free Energy and Entropy of a Mobile Loop of the Enzyme Acetylcholinesterase

Mihail Mihailescu and Hagai Meirovitch*

Department of Computational Biology, University of Pittsburgh School of Medicine, 3059 BST3, Pittsburgh, Pennsylvania 15260

Received: January 12, 2009; Revised Manuscript Received: March 26, 2009

The loop 287−290 (Ile, Phe, Arg, and Phe) of the protein acetylcholinesterase (AChE) changes its structure upon interaction of AChE with diisopropylphosphorofluoridate (DFP). Reversible dissociation measurements suggest that the free-energy (F) penalty for the loop displacement is $\Delta F = F_{\text{free}} - F_{\text{bound}} \sim -4$ kcal/mol. Therefore, this loop has been the target of two studies by Olson's group for testing the efficiency of procedures for calculating F. In this paper, we test for the first time the performance of our "hypothetical scanning molecular dynamics" (HSMD) method and the validity of the related modeling for a loop with bulky side chains in explicit water. Thus, we consider only atoms of the protein that are the closest to the loop (they constitute the "template"), where the rest of the atoms are ignored. The template's atoms are fixed in the X-ray coordinates of the free protein, and the loop is capped with a sphere of TIP3P water molecules; also, the X-ray structure of the bound loop is attached to the free template. We carry out two separate MD simulations starting from the free and bound X-ray structures, where only the atoms of the loop and water are allowed to move while the template−water and template−loop (AMBER) interactions are considered. The absolute F_{free} and F_{bound} (of the loop and water) are calculated from the corresponding trajectories. A main objective of this paper is to assess the reliability of this model, and for this several template sizes are studied capped with 80−220 water molecules. We find that consistent results for the free energy (which also agree with the experimental data above) require a template larger than a minimal size and a number of water molecules approximately equal to the experimental density of bulk water. For example, we obtain $\Delta F_{\text{total}} = \Delta F_{\text{water}} + \Delta F_{\text{loop}} = -3.1 \pm 2.5$ and -3.6 ± 4 kcal/mol for a template consisting of 944 atoms and a sphere containing 160 and 180 waters, respectively. Our calculations demonstrate the important contribution of water to the total free energy. Namely, for water densities close to the experimental value, ΔF_{water} is always negative leading thereby to a negative ΔF_{total} (while ΔF_{loop} is always positive). Also, the contribution of the water entropy $T\Delta S_{\text{water}}$ to ΔF_{total} is significant. Various aspects related to the efficiency of HSMD are tested and improved, and plans for future studies are discussed.

I. Introduction

Calculation of the entropy S and (Helmholtz) free energy F constitutes a central problem in computer simulation, in spite of the significant progress achieved in the last 50 years.[1−10] This problem in particular is severe in structural biology due to the flexibility and the strong long-range interactions characterizing biomacromolecules, such as proteins. Thus, the potential energy surface of a protein $E(\boldsymbol{x})$, is rugged ($\boldsymbol{x}$ is the $3N$-dimensional vector of the Cartesian coordinates of the molecule's N atoms); i.e., the surface is "decorated" by a tremendous number of localized wells and "wider" ones defined over regions Ω_m (called microstates), each consisting of many localized wells. An example for a microstate is the α-helical region of a peptide; see further discussion in refs 11 and 12. A microstate Ω_m, which typically constitutes only a tiny part of the entire conformational space Ω, can be represented by a sample (trajectory) generated by a *local* molecular dynamics (MD)[13,14] simulation, starting from a structure belonging to Ω_m. MD studies have shown that a molecule will visit a localized well only for a very short time [several femtoseconds (fs)] while staying for a much longer time within a microstate,[15,16] meaning that the microstates are of a greater physical significance than the localized wells. Typically one is interested in calculating the free energy of the most stable microstates, rather than calculating the total free energy (of Ω). Identifying these microstates, in particular that with the global free-energy minimum, is the extremely daunting task of protein folding. Notice that the microstates discussed above are metastable states; however, nonstable microstates, such as a transition state, might also be of interest.

Differences in ΔF (ΔS) are commonly calculated by thermodynamic integration (TI) over physical quantities such as the energy, temperature, and the specific heat ("calorimetric TI"), as well as nonthermodynamic parameters [free-energy perturbation (FEP) is also included in this category].[1−10] This is a robust and highly versatile approach, which enables one to calculate a *small* difference in the binding F of two ligands **a** and **b** in the active site of a *large* enzyme solvated by water. However, while the mutation process leading from **a** to **b** is well controlled by TI, conformational changes in the entire protein (i.e., "jumps" of side chains among rotamers) occur constantly, and therefore, the results might converge only after extremely long simulation times. Also, it is sometimes difficult to control the size and shape of the active site after mutation and the correct position of **b** in it. In many cases, one is interested in calculating ΔF_{mn} between two microstates Ω_m and Ω_n (denoted for brevity m and n, respectively); however, if the structural variance between the microstates is significant, then the integration from m to n becomes difficult and for large molecules unfeasible. These

* Corresponding author. Phone: 412-648-3338. E-mail: hagaim@pitt.edu.

10.1021/jp900308y CCC: $40.75
Published on Web 05/12/2009

drawbacks of TI can be overcome to a large extent by methods that provide the absolute F_m (S_m) from a given sample; thus, one is required to carry out (only) two separate local MD simulations of microstates m and n, calculating directly the absolute F_m and F_n; hence their difference $\Delta F_{mn} = F_m - F_n$, where the complex TI process is avoided (however, this approach has its own limitations, since the fluctuation of S and E of an N-particle system grows as $N^{1/2}$. On the other hand, the fluctuation of the exact free energy is zero as discussed in section II.3, following eq 6. In practice, however, F is approximate, and its fluctuation is finite but typically smaller than that of S and E). For a more extensive discussion on TI and other techniques for calculating differences, ΔF (ΔS), see ref 7.

The absolute S and F can be calculated by harmonic[17−19] and quasi-harmonic[20−22,10] approximations and by more general methods that are not limited to harmonic conditions, such as the local states (LS)[23−25] and hypothetical scanning (HS)[26−28] methods of Meirovitch, and other techniques.[29,30] However, all these methods are not applicable as yet to diffusive systems such as explicit water. Notice that the absolute F can also be obtained with TI provided that a reference state R with known F is available and an efficient integration path $R \rightarrow m$ can be defined. A classic example is the calculation of F of liquid argon or water by integrating the free energy from an ideal gas reference state, where for such homogeneous systems TI constitutes a very efficient method. However, for nonhomogeneous systems such integration might not be trivial, and in models of peptides and proteins defining adequate reference states and integration paths is a standing problem.[7]

In recent years, we have developed a new method for calculating the absolute S and F from a single sample called the hypothetical scanning Monte Carlo (HSMC) (or HSMD, where MD is used).[11,12,31−36] HSMC(D) is based on the ideas of the LS and HS methods mentioned above. Namely, in all these methods each conformation i of a sample (generated by MC, MD, or any other technique) is *reconstructed* step-by-step (from nothing, see section II.4) using transition probabilities (TPs). The product of these TPs leads to an approximation for the correct Boltzmann probability P_i^B from which various free-energy functionals can be defined. The TPs of HSMC(D) are stochastic in nature calculated by MC or MD simulations, where *all* interactions are taken into account. In this respect HSMC(D) (unlike HS and LS) can be viewed as exact;[31] the only approximation involved is due to insufficient MC(MD) sampling. HSMC(D) has unique features: it provides *rigorous* lower and upper bounds for F, which enable one to determine the accuracy from HSMC(D) results alone without the need to know the correct answer. Furthermore, F can be obtained from a very small sample and, in principle, even from any single conformation (e.g., see results for argon in ref 31).

HSMC(D) has been developed systematically as applied to liquid argon, TIP3P water,[31,32] self-avoiding walks on a square lattice,[33] and peptides,[34−36] where for the first three models HSMC(D) results have been found to agree within error bars to TI results obtained by extensive MC or MD simulations. Also, for polyglycine molecules differences ΔF_{mn} and ΔS_{mn} for α-helix, extended, and hairpin microstates were calculated very reliably by HSMC.[34−36]

HSMD has also been applied to a mobile loop of the protein α-amylase,[11,12] where the system is modeled by the AMBER96 force field[37] alone and by AMBER96 with the GB/SA implicit solvent of Still and co-workers.[38] In this work only protein atoms close to the loop (the template) were considered; however, they were kept fixed in their X-ray crystallographic positions, while only the loop's atoms were moved by MD. A further development step of HSMD has been achieved recently, where the implicit solvent was replaced by an explicit solvent, i.e., the α-amylase loop was capped with 70 TIP3P[39] water molecules, which (together with the loop's atoms) were moved in the MD simulation. Because the application of HSMD to water has not been optimized yet, the contribution of water to the free energy was calculated by a TI procedure incorporated within the framework of HSMD; this procedure is called HSMD-TI. In this TI process, the water−loop interactions are gradually decreased to zero. Notice that the related statistical errors are relatively small because the loop structure is held fixed during the integration, and only the water molecules are moved by MD. Previous studies of $N_{\text{water}}^{\min}$, the minimal number of water molecules required for obtaining stable structures of proteins[40] and loops[41] (based on a root-mean-square deviation criterion), suggested that in some cases $N_{\text{water}}^{\min}$ might be small (even 5 or 10 for certain loops). However, for loops no systematic study has been carried out to determine the smallest template size and the minimal number of water molecules that would lead to consistent free-energy results, i.e., results that are unchanged for larger template/water systems.

This study is carried out here, where an additional goal is to optimize HSMD-TI further. Thus, HSMD-TI is applied to the four-residue loop 287−290 (Ile, Phe, Arg, and Phe) of the protein acetylcholinesterase (AChE) from *Torpedo california*. AChE degrades the neurotransmitter acetylcholine (ACh), producing choline and an acetate group. It is mainly found at neuromuscular junctions and cholinergic synapses in the central nervous system, where its activity serves to terminate synaptic transmission. AChE has a very high catalytic activity: each molecule degrades about 5000 ACh molecules per second. Reduction in the activity of the cholinergic neurons is a well-known feature of Alzheimer's disease. Thus, AChE inhibitors are employed to reduce the rate at which ACh is broken down, thereby increasing the concentration of ACh in the brain and combating the loss of ACh caused by the death of cholinergic neurons. AChE is also the target of many nerve gases, particularly organophosphate inhibitors (e.g., sarin), which block the function of AChE and, thus, cause excessive ACh to accumulate in the synaptic clefts. The excess ACh causes neuromuscular paralysis leading to death.[42]

Of interest is the reaction of the inhibitor diisopropylphosphorofluoridate (DFP) with AChE,[43−48] which leads to a displacement of the loop's backbone roughly by 4 Å. Moreover, comparison of experimental reversible dissociation constants measured for a variety of inhibitors of differing molecular size suggests that the free-energy penalty for the loop displacement is on the order of 4 kcal/mol (i.e., $F_{\text{free}} - F_{\text{bound}} \sim -4$ kcal/mol).[43,45,49]

The fact that the crystal structures of the free[50] and bound[43] enzyme are available makes this loop a convenient target for testing free-energy procedures, and indeed such tests were already carried out by Olson and collaborators.[51,52] In the present study, we shall also extend the scope of HSMD for an internal loop (the AChE loop is "hidden" in a cleft) with large side chains; the earlier investigation of amylase has been applied to an external loop consisting of small side chains.

Before describing the system and our procedures in detail, it should be stressed that our study relies heavily on the notion of a microstate introduced earlier. However, determining the *exact* limits of a microstate in conformational space is practically impossible, and therefore, it is commonly defined by an MC or MD sample initiated from a microstate's structure. Thus, the

microstate's size typically increases with simulation time t, and estimation of E, S, and F depends on t as well. Therefore, estimation of ΔE_{mn}, ΔS_{mn}, and ΔF_{mn} between two microstates should be conducted with care; for details, see refs 11 and 12.

II. Theory and Methodology

II.1. The Loop and the Protein's Template. As was pointed out above, we study the four-residue loop Ile, Phe, Arg, and Phe (287−290), of AChE in two microstates related to the free and bound loop structures. The starting point is the available crystal structures of AChE from the protein data bank (PDB), where the unbound (free) structure 2ace[50] is based on residues 4−535, and 2dfp, the structure of AChE with DFP, consists of residues 2−535 (the bound structure).[43] To be able to compare these structures, 2dfp was trimmed so that both structures consist of the 531 residues, 4−535, and crystal water molecules were retained.

A close analysis shows that these conformations differ overall by (all heavy atoms) a root-mean-square deviation (rmsd) of 1.22 Å, while the loop conformations differ by rmsd = 1.38, 2.91, and 2.71 Å for the backbone, side chains, and all loop heavy atoms, respectively, meaning that most of the deviation is due to side chain movement. The rmsd of the templates, i.e., all atoms excluding the loop's atoms, is 1.10 Å.

We used the program TINKER,[53] and the AMBER force field,[37] where Lys, Arg, and His are positively charged and Asp and Glu have a negative charge. Hydrogens were added to the free structure (including the crystal water), and the potential energy was minimized, with all heavy atoms restrained to their crystal positions by harmonic forces (with a force constant of 100 kcal/(molÅ^2)). Afterward, the loop and (TIP3P) water atoms were allowed to relax in the presence of a fixed template. The minimization eliminates bad atomic overlaps and strains in the original structures, while keeping the atoms reasonably close to the PDB coordinates.

Because the free and bound protein structures are similar, we adopt here the same strategy used in our previous studies,[11,12] i.e., the structure of the bound loop (with hydrogens added) is attached to the free template. It would be impossible to compare the free energies of the free and bound microstates keeping the DFP attached to the bound template because the two systems will be different. Thus, we calculate the free energy required to move the loop from the free to the bound microstate in the presence of the free template. In principle, one could carry out a similar study based on the bound template; however, many coordinates of the bound structure 2dfp appear with large B-factors above 40, while the free structure (2ace) is much better resolved (B-factor values lower than 30). Therefore, the calculations were carried out with the free template. Also, taking into account the whole protein (of 8284 atoms) would be computationally prohibitive; therefore (as in refs 11 and 12), the template size is reduced to the N_{tmpl} atoms closest to the loop, where the rest of the atoms of the protein are ignored. More specifically (see Figure 1), the center of mass of the backbone atoms of the free loop is calculated as a (3D) reference point denoted $\boldsymbol{x}_{\mathrm{cmb}}$, and a distance ($R_{\mathrm{tmpl}}$) is chosen. If the distance of any atom of a residue from $\boldsymbol{x}_{\mathrm{cmb}}$ is less than R_{tmpl}, then the entire residue is included in the template; otherwise, the residue is eliminated. To assess the minimal template size needed, we have studied $R_{\mathrm{tmpl}} = 11$, 12, and 13 Å corresponding to $N_{\mathrm{tmpl}} = 800$, 944, and 1100 protein atoms and 20, 30, and 40 crystal water molecules, respectively.

After defining the template, the water molecules and the orientations of the polar hydrogens of the free and bound loops (on the same template) are subjected to an optimization procedure. This procedure (where all the loops' heavy atoms are kept fixed) is carried out in several rounds each consisting of a 0.1 ns MD run (1 fs time step) at high temperature (1250 K) followed by energy minimization; the process is stopped after a fixed number of unsuccessful rounds (typically 100), namely rounds which do not reduce the energy further. Although the heavy atoms are fixed, considerable gain in potential energy is achieved. Next, keeping the template atoms fixed, the system energy is minimized where the loop and water molecules are allowed to move *freely*. The final free and bound structures constitute starting points for further analysis.

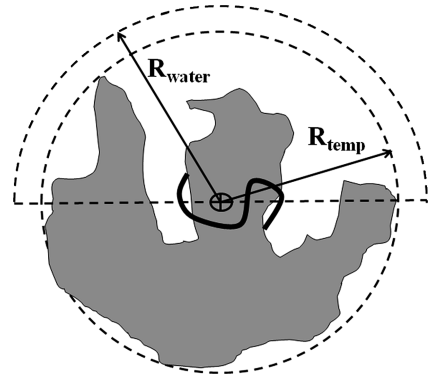

Figure 1. 2D diagram of the template and the spherical water restraining region. The loop is represented as the heavy black curve, where the symbol ⊗ denotes $\boldsymbol{x}_{\mathrm{cmb}}$, the center of mass of the loop backbone. $\boldsymbol{x}_{\mathrm{cmb}}$ is the center of the dashed (inner) circle (of radius = R_{tmpl}), which defines the edge of the (gray) template. $\boldsymbol{x}_{\mathrm{cmb}}$ is also the center of the larger (dashed) sphere of radius $R_{\mathrm{water}} = R_{\mathrm{tmpl}} + 1$ Å, where only the outer hemisphere of this sphere is shown. Approximately 30 crystal water molecules are inserted into the template, while the positions of the other waters are determined at random within the hemisphere. In the energy minimization and MD simulations, the waters are restricted to the hemisphere by strong harmonic potentials. "Drifting" of water molecules to the back of the template through the sides is mostly avoided because R_{water} is only 1 Å larger than R_{tmpl}.

II.2. Addition of Water. While our considerations thus far are based on the crystal structures, we seek to simulate the loop in solution. Therefore, it is not clear whether the positions of the crystal waters are relevant for the solution environment. In particular, water molecules that are caged within the crystal structure are expected to stay there during the simulation and, thus, can be considered as part of the template. Therefore, the number and arrangement of these waters should be globally optimized, practically by energy minimization, which, however, is a nontrivial problem (see below).

Our strategy is to add more water molecules to the already existing crystallographic waters. Thus, we define a sphere centered at $\boldsymbol{x}_{\mathrm{cmb}}$ with a radius R_{water} ($R_{\mathrm{water}} = R_{\mathrm{tmpl}} + 1$ Å), where waters are added at random to the hemisphere oriented toward the exterior of the template. To hold these waters around the loop, they are restrained with a flat-welled half-harmonic potential (with a force constant of 10 kcal mol^{-1} Å^{-2}) based on their distance from $\boldsymbol{x}_{\mathrm{cmb}}$. That is, if the distance of a water oxygen from $\boldsymbol{x}_{\mathrm{cmb}}$ is greater than R_{water}, then a harmonic restoring force is applied; otherwise the restraining force is zero. In a previous paper,[41] the minimal number of waters $N_{\mathrm{water}}^{\mathrm{min}}$ required to cap a loop has been studied with respect to the loop stability, i.e., the change in the loop's rmsd from the initial (PDB) structure during 4 ns MD simulations. The objective of the present paper is similar but based on a free-energy criterion; thus, we seek to find $N_{\mathrm{water}}^{\mathrm{min}}$ that for $N_{\mathrm{water}} \geq N_{\mathrm{water}}^{\mathrm{min}}$, $F_{\mathrm{free}} - F_{\mathrm{bound}}$ shows stability. To find $N_{\mathrm{water}}^{\mathrm{min}}$, we study $80 \leq N_{\mathrm{water}} \leq 220$ for different template sizes.

The fully hydrated system is then treated by a sequence of MD/minimization procedures similar to that outlined above, applied only to the hydrogen atoms of the loop, and *all* water molecules are restrained within the full sphere of radius R_{water}. Again, at the end of this optimization the energy is minimized where the loops' atoms and the water molecules are free to move. The conformation of the free loop changed minimally from its crystal structure (rmsd = 0.22, 0.63, and 0.57 Å for backbone, side chains, and all loop atoms, respectively), while the bound conformation changed more (rmsd = 1.16, 2.46, and 2.32 Å for backbone, side chains, and all loop atoms, respectively; see discussion of the results in Table 1).

For each of the optimized "free" and "bound" structures (with several template sizes N_{tmpl} and several levels of hydration defined by N_{water}) an MD run at 300 K is performed, where only the loop and water atoms are moved while the template atoms are kept fixed. Thus, 1000 loop/water configurations are collected by retaining a configuration every 0.5 ps along the 0.5 ns MD trajectory. An equilibration run of 0.5 ns is carried out prior to the production run. The total potential energy E_{total} is the sum of partial energies related to the loop and water (the template−template energy is constant and thus is ignored):

$$E_{total} = [E_{loop-loop} + E_{loop-tmpl}] + [E_{water-water} + E_{water-tmpl} + E_{water-loop}] = E_{loop} + E_{water} \quad (1)$$

where $E_{loop-loop}$ is the intraloop energy and $E_{loop-tmpl}$ is the energy due to loop−template interactions; these energies define the total loop energy E_{loop}, and the interactions related to water are defined in a similar way, where their total is defined as E_{water}. The reconstruction of the loop structure is carried out in internal coordinates; therefore, the loop conformations simulated by MD are transferred from Cartesians to the dihedral angles φ_i, ψ_i, and ω_i ($i = 1, N = 4$), the bond angles $\theta_{i,l}$ ($i = 1, N$, $l = 1, 3$), the side chain angles χ, and the corresponding bond angles. For convenience, all these angles (ordered along the backbone) are denoted by α_k, $k = 1, 37 = K$. We have argued in refs 11 and 36 that, to a good approximation, bond stretching can be ignored.

II.3. Statistical Mechanics of a Loop in Internal Coordinates. The partition function of the loop/water system is

$$Z_m = \int_m \exp[-E(\boldsymbol{x}_{loop}, \boldsymbol{x}^N)/k_B T]\, d\boldsymbol{x}_{loop}\, d\boldsymbol{x}^N \quad (2)$$

where $E(\boldsymbol{x}_{loop}, \boldsymbol{x}^N) = E_{tot}$ as defined in eq 1; $\boldsymbol{x}_{loop}$ is the Cartesian coordinates of the loop in microstate m (for brevity we omit the letter m in most equations); $\boldsymbol{x}^N$ is the $9N$ Cartesian coordinates of the water molecules (for brevity we denote in the theoretical section N_{water} by N ($N = N_{water}$)). However, it is convenient to change the variables of integration from $\boldsymbol{x}_{loop}$ to internal coordinates, which makes the integral dependent also on a Jacobian.[17,18,20] This transformation is applied under the assumption that the potentials of the bond lengths ("the hard variables") are strong, and therefore, their average values can be assigned to their corresponding Jacobian J, which to a good approximation can be taken out of the integral (however, see a later discussion in this section). For the same reason, one can carry out the integration over the bond lengths (assuming that they are not correlated with the α_k), and the remaining integral becomes a function of the K dihedral and bond angles (α_k)[17,18,20] and a Jacobian that depends only on the bond angles:

$$Z' = DZ = D \int_m \exp([-E_{loop}([\alpha_k]) - E_{water}([\alpha_k], \boldsymbol{x}^N)]/k_B T)\, d[\alpha_k]\, d\boldsymbol{x}^N \quad (3)$$

where $[\alpha_k] = [\alpha_1, ..., \alpha_K]$ and $d[\alpha_k] = d\alpha_1, ..., d\alpha_K$. $D(T)$ is a product of the integral over the bond lengths and their Jacobian J. Due to the strong bond stretching potentials, D is assumed to be the same (i.e., constant) for different microstates of the same loop, and therefore, ln D cancels and can be ignored in calculations of free energy and entropy differences. The Jacobian of the bond angles, which should appear under the integral, is omitted because we have shown[11] that it cancels out in entropy and free-energy differences (our main interest). The Boltzmann probability density corresponding to Z (eq 3) is

$$\rho^B([\alpha_k], \boldsymbol{x}^N) = \exp\{-E([\alpha_k], \boldsymbol{x}^N)/k_B T\}/Z \quad (4)$$

and the exact entropy S and exact free energy F (defined up to an additive constant) are

$$S = -k_B \int_m \rho^B([\alpha_k], \boldsymbol{x}^N) \ln \rho^B([\alpha_k], \boldsymbol{x}^N)\, d[\alpha_k]\, d\boldsymbol{x}^N \quad (5)$$

and

$$F = \int_m \rho^B([\alpha_k], \boldsymbol{x}^N)\{E([\alpha_k], \boldsymbol{x}^N) + k_B T \ln \rho^B([\alpha_k], \boldsymbol{x}^N)\}\, d[\alpha_k]\, d\boldsymbol{x}^N \quad (6)$$

It should be pointed out that the fluctuation of the *exact* F is zero,[54,55] because by substituting $\rho^B([\alpha_k])$ (eq 4) inside the curly brackets of eq 6 one obtains $E([\alpha_k]) + k_B T \ln \rho^B([\alpha_k]) = -k_B T \ln Z = F$, i.e., the expression in the curly brackets is constant and equal to F for any set $[\alpha_k]$ within m. This means that the free energy can be obtained from any single conformation if its Boltzmann probability density is known. However, the fluctuation of an approximate free energy (i.e., which is based on an approximate probability density) is finite, and it is expected to decrease as the approximation improves.[54,55,30−33] Because HSMC(D) provides an approximation for $\rho^B([\alpha_k], \boldsymbol{x}^N)$, it enables one, in principle, to estimate the free energy of the system from any single structure[31,33] Notice, however, that the calculation of $\rho^B([\alpha_k], \boldsymbol{x}^N)$ for a single conformation depends on the entire microstate as is also evident from the HSMC(D) procedure discussed later in section II.5.

With MD the bond-stretching energy is taken into account in eq 6 (and in free-energy functionals defined later), while the corresponding entropy is ignored. The contribution of this energy to the free energy becomes an additive constant, if one accepts the assumptions about the stretching energy and the corresponding Jacobian made prior to eq 6. This is a very good approximation; however, if the bond-stretching entropy should be considered, we have argued (in ref 11, section II.5) that it can be estimated approximately within the framework of HSMD.

II.4. Exact Future Scanning Procedure. HSMC(D) is based on the ideas of the exact scanning method where a system is constructed (from nothing) step-by-step using TPs. The product of these TPs is equal to the Boltzmann probability (eq 4) from which the entropy and free energy can be calculated. Practically, a loop/water configuration is generated by initially building a loop structure followed by the construction of a configuration of the surrounding water molecules. In this way a sample of statistically independent system configurations can be obtained.

For simplicity, this construction is described for a loop consisting of M Gly residues (with dihedral and bond angles denoted as α_k, $1 \le \alpha_k \le 6M = K$) in microstate m; the loop is surrounded by N_{water} water molecules moving within the volume defined by the sphere of radius R_{water}, the template, and the loop. Starting from nothing, a conformation of the loop is built first by defining the angles α_k step-by-step using TPs and adding the related atoms.[53] Thus, at step k, $k-1$ angles $\alpha_1, ..., \alpha_{k-1}$ have already been determined; these angles and the related structure

(the past) are kept constant, and α_k is defined with the *exact* TP density $\rho(\alpha_k|\alpha_{k-1}, ..., \alpha_1)$:

$$\rho(\alpha_k|\alpha_{k-1}, ...,\alpha_1) = Z_{\text{future}}(\alpha_k, ...,\alpha_1)/[Z_{\text{future}}(\alpha_{k-1}, ...,\alpha_1)] \quad (7)$$

where $Z_{\text{future}}(\alpha_k, ..., \alpha_1)$ is a future partition function. The term "future" indicates that the integration defining Z_{future} is carried out over the variables $\alpha_{k+1}, ..., \alpha_K$ and the $9N$ coordinates $\boldsymbol{x}^N$ of the water molecules, which will be determined in future steps of the build-up process. In this integration, the atoms treated in the past are held fixed in their coordinates (which are determined by $\alpha_1, ..., \alpha_k$), while $\alpha_{k+1}, ..., \alpha_K$ are varied in a restrictive way, where the corresponding conformations of the "future" part of the loop remain in microstate m. Thus

$$Z_{\text{future}}(\alpha_k, ...,\alpha_1) = \int_m \exp[-(E(\alpha_K, ...,\alpha_1,\boldsymbol{x}^N)/k_BT] \times d\alpha_{k+1}... d\alpha_K d\boldsymbol{x}^N \quad (8)$$

where E (eq 1) is the total potential energy of the loop/template/water system, which also imposes the loop closure condition. The product of the TPs (eq 7) leads to the (Boltzmann) probability density of the entire loop conformation:

$$\rho^{\text{B}}_{\text{loop}}(\alpha_K, ...,\alpha_1) = \prod_{k=1}^{K} \rho(\alpha_k|\alpha_{k-1}, ...,\alpha_1) \quad (9)$$

After the loop structure has been constructed, a configuration of water molecules is generated step-by-step, where the TP density $\rho_{\text{water}}(\boldsymbol{x}_k|\alpha_K, ..., \alpha_1, \boldsymbol{x}^{k-1})$ for placing water molecule k at $\boldsymbol{x}_k$ is defined in a similar way to eq 7 based on $Z_{\text{future}}([\alpha_k], \boldsymbol{x}^k)$ and $Z_{\text{future}}([\alpha_k], \boldsymbol{x}^{k-1})$, where the loop conformation is kept constant and the $k-1$ water molecules that have already been treated are fixed at their coordinates $\boldsymbol{x}^{k-1}$, and the summation in $Z_{\text{future}}(\boldsymbol{x}^k)$ is over the as yet undecided $N - k + 1$ water molecules. (Notice that $\boldsymbol{x}_k$ denotes the 9 Cartesian coordinates of water molecule k, while $\boldsymbol{x}^k$ denotes the set of Cartesian coordinates of the k molecules 1, 2, ..., k.) The Boltzmann probability density of the water is

$$\rho^{\text{B}}_{\text{water}}(\alpha_K, ...,\alpha_1,\boldsymbol{x}^N) = \prod_{k=1}^{N} \rho_{\text{water}}(\boldsymbol{x}_k|\alpha_K, ...,\alpha_1,\boldsymbol{x}^{k-1}) \quad (10)$$

and the probability density $\rho^{\text{B}}([\alpha_k], \mathbf{x}^N)$ of the loop/water configuration is the product of $\rho^{\text{B}}_{\text{loop}}([\alpha_k])$ and $\rho^{\text{B}}_{\text{water}}([\alpha_k], \boldsymbol{x}^N)$. One can define for m the loop entropy, S_{loop}:

$$S_{\text{loop}} = -k_B \int_m \rho^{\text{B}}_{\text{loop}}([\alpha_k]) \ln \rho^{\text{B}}_{\text{loop}}([\alpha_k]) d[\alpha_k] \quad (11)$$

S_{loop} is defined up to an additive constant. Extending the exact scanning procedure to side chains is straightforward.

This construction procedure (which is not feasible for a large loop/water system) provides the theoretical basis for HSMC(D). Thus, the exact scanning method is equivalent to any other exact simulation technique (in particular, to Metropolis MC or MD) in the sense that the large samples generated by such methods lead to the same averages and fluctuations. Therefore, one can assume that a given MC or MD 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; this is the basis for HSMC(D) (as well as the HS and LS methods).

II.5. The HSMC(D) Method. The theory of HSMD is again described as applied to a loop consisting of M Gly residues. One starts by generating a MD sample of microstate m with water molecules; the conformations are then represented in terms of the dihedral and bond angles α_k, $1 \le \alpha_k \le 6M = K$, and the variability range $\Delta\alpha_k$ is calculated

$$\Delta\alpha_k = \alpha_k(\text{max}) - \alpha_k(\text{min}) \quad (12)$$

where $\alpha_k(\text{max})$ and $\alpha_k(\text{min})$ are the maximum and minimum values of α_k found in the sample, respectively. $\Delta\alpha_k$, $\alpha_k(\text{max})$, and $\alpha_k(\text{min})$ enable one to verify that the sample has not "escaped" from microstate m.

System configuration ($[\alpha_k]$, $\boldsymbol{x}^N$) (denoted i for brevity) is reconstructed in two stages, where the loop structure is reconstructed first followed by the reconstruction of the water configuration. Thus, at step k of stage 1, $k-1$ angles $\alpha_{k-1}, ..., \alpha_1$ have already been reconstructed, and the TP density of α_k, $\rho(\alpha_k|\alpha_{k-1}, ..., \alpha_1)$ is calculated from a MD sample of n_f conformations (generated in Cartesian coordinates), where the *entire* future of the loop and water is moved by MD [i.e., the loop atoms defined by $\alpha_k, ..., \alpha_K$ and the water coordinates ($\boldsymbol{x}^N$)], while the past (the loop atoms defined by $\alpha_1, ..., \alpha_{k-1}$) are held fixed at their values in conformation i. A small segment (bin) $\delta\alpha_k$ is centered at $\alpha_k(i)$, and the number of visits of the future chain to this bin during the simulation, n_{visit}, is calculated; one obtains

$$\rho_{\text{loop}}(\alpha_k|\alpha_{k-1}, ...,\alpha_1) \approx \rho^{\text{HS}}(\alpha_k|\alpha_{k-1}, ...,\alpha_1) = n_{\text{visit}}/[n_f\delta\alpha_k] \quad (13)$$

where $\rho^{\text{HS}}(\alpha_k|\alpha_{k-1}, ..., \alpha_1)$ becomes exact for a very large n_f ($n_f \to \infty$) and a very small bin ($\delta\alpha_k \to 0$). This means that in practice $\rho^{\text{HS}}(\alpha_k|\alpha_{k-1}, ..., \alpha_1)$ will be somewhat approximate due to insufficient future sampling (finite n_f), a relatively large bin size $\delta\alpha_k$, an imperfect random number generator, etc. This equation is suitable for HSMC. However, for practical reasons, with HSMD a pair of angles should be treated simultaneously, where each pair consists of a dihedral angle and its successive bond angle (e.g., φ and the bond angle $N-C^{\alpha}-C'$). Thus, at each step both α_k and α_{k+1} are considered, and n_{visit} is increased by 1 only if α_k and α_{k+1} are both located within the limits of $\delta\alpha_k$ and $\delta\alpha_{k+1}$, respectively; also, for Arg we treat three consecutive χ angles, and in the future we plan to treat four angles. Therefore, for n consecutive angles eq 13 becomes

$$\rho^{\text{HS}}(\alpha_{k+n-1}, ...,\alpha_{k+1}, \alpha_k|\alpha_{k-1}, ...,\alpha_1) = n_{\text{visit}}/\left[n_f \prod_{j=k}^{j=k+n-1} \delta\alpha_j\right] \quad (14)$$

where in ref 11, we have shown that $\delta\alpha_k$ and $\delta\alpha_{k+1}$ can be optimized. Notice that with HSMD, the future loop conformations generated by MD at each step k remain, in general, within the limits of m, which is represented by the analyzed MD sample. The corresponding probability density is

$$\rho^{\text{HS}}(\alpha_K, ...,\alpha_1) = \prod_{k=1}^{K-n+1} \rho^{\text{HS}}(\alpha_{k+n-1}, ...,\alpha_{k+1}, \alpha_k|\alpha_{k-1}, ...,\alpha_1) \quad (15)$$

$\rho^{\text{HS}}([\alpha_k])$ defines an approximate entropy functional, denoted $S^{\text{A}}_{\text{loop}}$, which can be shown (using Jensen's inequality) to constitute a *rigorous* upper bound for S_{loop} (eq 11):[26,30]

$$S^{\text{A}}_{\text{loop}} = -k_B \int_m \rho^{\text{B}}([\alpha_k]) \ln \rho^{\text{HS}}([\alpha_k]) d[\alpha_K] \quad (16)$$

$\rho^{\text{B}}_{\text{loop}}$ (eq 9) is the Boltzmann probability density of $[\alpha_K]$ in m. Thus, for microstate m, $S^{\text{A}}_{\text{loop}}$ can be estimated from a Boltzmann

sample (of size n_s) generated by MD using the arithmetic average:

$$\bar{S}^{A}_{loop}(m) = -\frac{k_B}{n_s}\sum_{t=1}^{n_s} \ln\rho^{HS}(t,m) \quad (17)$$

where $\rho^{HS}(t,m)$ is the value of $\rho^{HS}([\alpha_k])$ obtained for configuration t of the sample of m. $\bar{S}^{A}_{loop}$ (with the bar) is an estimation of the ensemble average S^{A}_{loop} (eq 16); correspondingly, the ensemble averages of the energy are estimated from a sample of size n_s and should appear with a bar as well. However, from now on only estimations will be considered, and for simplicity all of them will appear without the bar, like the energies defined in eq 1. S^{A}_{loop} (eqs 16 and 17) constitutes a measure for the loop flexibility of a pure geometrical character, i.e., with no *direct* dependence on the interaction energy. When the *converged* value of S^{A}_{loop} is considered, it will be denoted by S_{loop}, which is expected to be the exact value within the statistical error. In the same way, the difference in the loop entropies between two microstates obtained for a specific set of parameters is denoted by ΔS^{A}_{loop}, while the converged difference is denoted by ΔS_{loop}, thus

$$\Delta S^{A}_{loop} = \bar{S}^{A}_{loop}(m) - \bar{S}^{A}_{loop}(n) \quad (18a)$$

where

$$\Delta S_{loop} = \bar{S}^{A}_{loop}(m) - \bar{S}^{A}_{loop}(n) \qquad \text{(converged)} \quad (18b)$$

The difference in the loop energy between two microstates is (see eq 1)

$$\Delta E_{loop} = E_{loop}(m) - E_{loop}(n) \quad (19)$$

One can also define a free-energy difference for the loop ΔF_{loop}:

$$\Delta F_{loop} = \Delta E_{loop} - T\Delta S_{loop} \quad (20)$$

To reconstruct the water configuration, one can use in principle the HSMC(D) procedure for fluids developed previously, which would lead to $\rho^{HS}_{water}([\alpha_k], \boldsymbol{x}^N)$ [as an approximation for $\rho^{B}_{water}([\alpha_k], \boldsymbol{x}^N)$ (eq 10)] and then to the contribution of the water configuration to the free energy:

$$F_{water}([\alpha_k], \boldsymbol{x}^N) = E_{water}([\alpha_k], \boldsymbol{x}^N) + k_B T \ln\rho^{HS}_{water}([\alpha_k], \boldsymbol{x}^N)$$

However, this procedure for fluids has not been optimized yet, and it is relatively time-consuming.

Alternatively as in ref 12, one can obtain $F_{water}([\alpha_k], \boldsymbol{x}^N)$ by a TI procedure based on the same reference state for the free and bound structures. In this state, the water−water and water−template interactions are preserved, but the (fixed) loop structure $[\alpha_k]$ does not "see" the surrounding waters, i.e., the loop−water interactions (electrostatic and Lennard-Jones) are switched off. These interactions are gradually increased (from zero) during an MD simulation of water (while the loop structure remains fixed at $[\alpha_k]$). For $[\alpha_k]$ of microstate m, one obtains from the integration $F^{TI}_{water}([\alpha_k], m)$, which is then averaged over the n_s sample configurations (as in eq 17). As described in the Appendix section V.1, the integration is carried out in two stages but in an *opposite* direction to that described above, i.e., first the charges are gradually decreased to zero, followed by a similar decrease in the Lennard-Jones (LJ) potential, leading to $F^{TI}_{water}([\alpha_k], m, \text{ch})$ and $F^{TI}_{water}([\alpha_k], m, \text{LJ})$, respectively. Denoting the set of $[\alpha_k]$ in the sample by t and omitting m one obtains

$$F^{TI}_{water}(m) = F^{TI}_{water}(\text{ch}) + F^{TI}_{water}(\text{LJ}) = \frac{1}{n_s}\sum_{t=1}^{n_s} F^{TI}_{water}(\text{ch}, t) + F^{TI}_{water}(\text{LJ}, t) \quad (21)$$

The difference in the free energy of water between m and n denoted ΔF_{water} is

$$\Delta F_{water} = F^{TI}_{water}(m) - F^{TI}_{water}(n) \quad (22)$$

ΔE_{water} (see eq 1) is

$$\Delta E_{water} = E_{water}(m) - E_{water}(n) \quad (23)$$

and

$$\Delta E_{total} = \Delta E_{water} + \Delta E_{loop} \quad (24)$$

The total free energy is

$$F_{total}(m) = F^{TI}_{water}(m) + E_{loop}(m) - TS_{loop}(m) \quad (25)$$

and the difference in the total free energy between microstates m and n is

$$\Delta F_{total} = \Delta E_{loop} - T\Delta S_{loop} + \Delta F_{water} \quad (26)$$

The difference in the water entropy between m and n is

$$T\Delta S_{water} = T[S^{TI}_{water}(m) - S^{TI}_{water}(n)] = \Delta E_{water} - \Delta F_{water} \quad (27)$$

where the corresponding difference in the total entropy is

$$T\Delta S_{total} = T\Delta S_{water} + T\Delta S_{loop} \quad (28)$$

Notice that all entropies and free energies are defined up to an additive constant.

II.6. The Reconstruction Procedure with HSMD. The HSMD reconstruction procedure needs further discussion. Thus, the MD simulation of the future chain at step k starts from the reconstructed conformation i, and every $g = 10$ fs, the current conformation is retained, where the n_{init} initial retained conformations are discarded for equilibration. The next n_f (retained) future conformations are represented in internal coordinates, and their contribution to n_{visit} (eqs 13 and 14) is calculated. An essential issue is how to guarantee an adequate coverage of microstate m, i.e., that the future chains will span its entire region (in particular the side-chain rotamers) while avoiding their "overflow" to neighboring microstates, conditions that will occur for a too small and a too large n_f, respectively. [Note that even at step k, where the "past" of the loop is kept fixed, the (future) unfixed part can leave the microstate during long MD simulations. Such an "overflow" is more likely to happen for small residues such as Gly and for small k.]

In previous work,[11,12,34−36] we developed procedures for keeping the loop in its original microstate and measures for estimating the extent of coverage of the microstate by the reconstructed samples (an important test for an adequate coverage is verifying that entropy differences are stable as n_f is increased.) In this paper these procedures have not been applied because the maximal $n_f$values used are not large and the microstates are concentrated (i.e., the $\Delta\alpha_k$ values of eq 12 are relatively small; see discussions in section III).

II.7. Calculation of Free-Energy and Entropy Differences. As has already been pointed out, our main interest is in the difference ΔF_{mn} (ΔS_{mn}) between microstates rather than in the absolute values themselves. For any practical set of n_f and bin sizes $\delta\alpha_k$, F^{A}_{m} (F^{A}_{n}) will be approximate, and thus, the difference $F^{A}_{m} - F^{A}_{n}$, might be approximate as well. However, if $F^{A}_{m} - F^{A}_{n}$

is found to be stable for significantly improving sets of parameters (i.e., better approximations), then the stable value can be considered as the correct difference (within the statistical errors). Indeed, in the application of HSMD to peptides[36] and loops,[11,12] relatively small values of n_f have already led to stable differences, meaning that the *systematic* errors in both F_m^A and F_n^A are comparable and, thus, are canceled in $F_m^A - F_n^A$ (we define the deviation, $F_m^A - F$ as the systematic error.) In ref 11, we have provided theoretical arguments supporting this error cancelation, which, however, should be verified for each system studied. Thus, using HSMC(D)-TI, the objective is not to obtain the most accurate F_m and F_n, but to minimize computer time by finding the *worst* F_m^A and F_n^A (i.e., the worst HSMC(D)-TI approximation) for which ΔF_{mn}^A is still correct within a required statistical error. This strategy is also applied to approximations used in the model. For example, the harmonic boundary conditions that keep water within the hemisphere impose errors that are expected to be comparable for both microstates; thus, one would anticipate them to get canceled in free-energy differences. Again, one should verify that the differences are stable for increasing values of N_{water} and R_{water}.

III. Results and Discussion

III.1. Simulation Details. As stated above and earlier, an essential aim of this work is to find the minimal template size and minimal number of water molecules required for a reliable modeling of the loop behavior (i.e., that free-energy differences for the minimal and larger models are approximately the same). We have tested three template sizes, defined by radii, R_{tmpl} = 11,12, and 13 Å, consisting of 783, 944, and 1141 atoms, respectively, where each template was studied with an increasing number of water molecules, N_{water}, ranging from 80 to 220.

For each pair (R_{tmpl}, N_{water}), the solvated free and bound structures were initially optimized, as described in sections II.1 and II.2, leading to two optimized structures. Then, starting from each optimized structure, a 1 ns MD trajectory was generated at 300 K, where the initial 0.5 ns part was used for equilibration and a sample of 1000 structures was generated from the last (half) part of the trajectory by retaining a structure to a sample every 0.5 ps. From these samples (which represent the free and bound microstates), the free energy and entropy are calculated.

These simulations and the reconstruction simulations (for generating the future samples) were carried out with the velocity-Verlet algorithm[57] based on a time step of 2 fs, where bonds involving hydrogens (including those of water) were frozen to their ideal values by the RATTLE algorithm;[57] the Berendsen[57] heat bath controlled the temperature. Cut-offs on long-range interactions were not imposed, and in the reconstruction process a structure was added to the sample every g = 10 fs, where the n_{init} = 250 initial structures (2.5 ps) were discarded for equilibration. The future samples were generated for several bin sizes, where results are presented for $\delta\alpha_k = \Delta\alpha_k/l$, l = 5, 10, 20, 30, 40, and 50, centered at α_k (i.e., $\alpha_k \pm \delta\alpha_k/2$) (eqs 12−14). If the counts of the smallest bin are smaller than 50, then the bin size is increased to the next size and, if necessary, to the next one, etc. In the case of zero counts, n_{visit} is taken to be 1; however, an event of zero counts is very rare.

III.2. Dihedral Angles for Different Microstates. In Table 1 we present results for α_k(min), α_k(max), and $\Delta\alpha_k$ (eq 12) for the backbone dihedral angles φ and ψ and the side chains χ. These values are based on the samples of 1000 conformations generated for the free and bound microstates for the template of R_{tmpl} = 12 Å and N_{water} = 160. The table shows that the $\Delta\alpha_k$ values are relatively small (in most cases smaller than 80°), and they are comparable to those obtained by other samples [based on different (R_{tmpl}, N_{water}) pairs]. This suggests that $|\Delta S_{loop}|$ (eq 18b) would be small as well (probably not larger than 2 kcal/mol; compare with $\Delta\alpha_k$ in Table 1 of ref 12). These small $\Delta\alpha_k$ values stem from the constraints imposed by the template on the inner loop.

For comparison we also provide in Table 1 the dihedral values of the crystal structures, α_k(crystal). These angles enable one to determine whether the samples have escaped from their original microstates. While exact definition of a microstate is practically unfeasible (see discussion in section I.3 of ref 11), we have accepted an "escape" criterion for a dihedral angle when α_k(crystal) + 60° is smaller than α_k(max) or α_k(crystal) − 60° is larger than α_k(min), i.e., if some angle values fall beyond the range α_k(crystal) ± 60°; these angles are bold faced in the table. The table reveals that three and two backbone angles of the free and bound microstates, respectively, have been escaped, but the deviations are small; the corresponding side-chain (escaped) angles are five and three, among them are χ^1−χ^4 of Arg (free) (where the deviation of χ^1 is small) and the slightly deviating χ^1 of Phe (bound). Thus, the original microstates are not completely retained (mainly for the side chains), but this might be expected since we model the loops in solution rather than in the crystal environment.

III.3. Determination of Minimal Values of N_{tmpl} and N_{water}. In Table 2 results are presented for $E_{loop-temp}$, $E_{loop-loop}$, E_{loop}, and E_{total} (eq 1) and for F_{water}^{TI}(ch), F_{water}^{TI}(LJ), and F_{water}^{TI} (eq 21), calculated as described in the Appendix, V.1; we also provide results for $F_{sum} = E_{loop} + F_{water}^{TI}$, which is the total free energy without the contribution of the loop entropy, S_{loop} (eqs 16 and 17). For each quantity, we have calculated the difference Δ between the free and bound values. These results were obtained for R_{tmpl} = 12 and R_{water} = 13 Å for 80 ≤ N_{water} ≤ 180. Similar results are presented in Table 3 for R_{tmpl} = 13 and R_{water} = 14 Å and 80 ≤ N_{water} ≤ 220.

To check the stability of these results and assess their statistical errors, they were calculated for an increasing sample size of n_s = 10, 20 (not shown) and 40 and 80 for the higher N_{water} values. In some calculations, the thermodynamic integration is doubled. The tables show that the results presented (in particular those for the larger N_{water}) are very stable. Statistical errors, $s/(n_s)^{1/2}$, where s is the standard deviation, are smaller than 2.5 and 1.5 kcal/mol for E_{total} and the other quantities, respectively. For n_s = 40 and 80, the conformations are selected from the trajectory every 12.5 and 6.25 ps, respectively, and the energy correlations are expected to be low. Notice that the correlations of the correct free energy are zero because every conformation leads to the correct result (see II.3); for an approximate free energy the correlations will be smaller than the energy correlations. The errors in Δ are differences between results obtained for the largest and smaller n_s values. The Δ results (in particular those for the larger N_{water}) are again very stable, i.e., the errors in most cases are relatively small, within 2 kcal/mol, due to cancelation of the individual errors in the difference; for example, F_{water}^{TI}(ch), F_{water}^{TI}(LJ), and thus, F_{water}^{TI} (eq 21) change significantly in going from a single to a double integration; however, this change is comparable in the free and bound calculations and gets canceled in Δ. It should be pointed out that to verify our error estimation for some of the larger values of N_{water} (R_{tmpl} = 12 and 13 Å), several TI runs (for calculating F_{water}^{TI}) were carried out starting from different sets of velocities.

To estimate the minimum values of R_{tmpl} and N_{water}, it is sufficient to study ΔF_{sum}, the difference in the sums of all

TABLE 1: Minimum and Maximum Values of Dihedral Angles α_k(min) and α_k(max) and Their Differences $\Delta\alpha_k$ (in degrees) for the Free and Bound Samples[a]

		free				bound			
residue	angle	α_k(crystal)	α_k(min)	α_k(max)	$\Delta\alpha_k$	α_k(crystal)	α_k(min)	α_k(max)	$\Delta\alpha_k$
Ser	ψ	165	−194	−172	22	172	−218	−176	42
	ω	179	**130**	**154**	24	177	−188	−157	23
Ile	φ	−120	−159	−119	40	−57	**−130**	**−71**	59
	ψ	138	**−210**	**−170**	40	−44	−58	−10	48
Phe	φ	70	30	80	50	−98	−153	−89	64
	ψ	43	**51**	**119**	68	130	73	144	71
Arg	φ	−139	−233	−150	83	−113	−149	−84	65
	ψ	136	−7	138	145	13	**−70**	**−28**	42
Phe	φ	−122	−140	53	190	−88	−84	16	100
				Side Chains					
Ile	χ^1	−118	−184	−110	74	−82	−130	−89	41
	χ^2	38	**−227**	**−171**	56	−47	−79	−15	64
	χ^3	180	−180	180	360	180	−180	−103	77
	$\chi^{2'}$	180	36	101	65	180	−199	−157	42
Phe	χ^1	−139	−160	−67	93	−62	**−153**	**−95**	58
	χ^2	−36	−99	2	97	−51	−103	−41	62
Arg	χ^1	−68	**−146**	**−77**	69	−72	−108	−61	47
	χ^2	−55	**−207**	**−158**	49	−71	**−195**	**−141**	54
	χ^3	−170	**33**	**102**	69	−50	−99	−45	54
	χ^4	−97	**50**	**133**	89	176	−209	−154	55
	χ^5	0	−33	39	72	0.4	−30	46	76
	χ^6	0	−51	26	77	0	−39	37	77
	$\chi^{6'}$	0	−30	34	64	0	−33	48	81
Phe	χ^1	−43	−77	−36	41	40	**−65**	**−30**	35
	χ^2	−80	−81	−1	82	−88	−79	−36	43

[a] α_k(min), α_k(max), and $\Delta\alpha_k$ are defined in eq 12; their values were calculated from samples of 1000 loop conformations (0.5 ps) generated for the free and bound microstates. The values of α_k(crystal) were calculated from the PDB crystal structures 2ace[50] and 2dfp[43] of the free and bound protein, respectively.

contributions to the free energies excluding that of S_{loop} (eqs 16 and 17), where ΔS_{loop} is expected to be small. The results of ΔF_{sum} in Table 2 are positive, 31, 30, and 14 kcal/mol for the low density water, N_{water} = 80, 100, and 120, respectively, meaning that the bound microstate is more stable (has lower free energy) than the free microstate (again, neglecting S_{loop}). On the other hand, as N_{water} is increased to 140, 160, and 180, ΔF_{sum} becomes negative, −8, −1, and −5 kcal/mol, respectively, i.e., the free loop becomes the most stable. We also provide a measure for the density of water, $\rho_{water} = N_{water}$/(hemisphere volume), i.e., $\rho_{water} = N_{water}/[2\pi(R_{water})^3/3]$, which increases to 0.0304, 0.0348, and 0.0391 Å^{-3} for N_{water} = 140, 160, and 180, respectively. These values are comparable to the experimental density of water, ρ_{water} = 0.0350 Å^{-3}, which corresponds to 154 waters in the hemisphere; however, these densities are somewhat approximate since the hemisphere is not totally empty but contains part of the template. Thus, the density is lower for waters arranged initially in crystal water positions or for those which are put randomly inside the template. Also, bulk water can move during the simulation to crevices inside the template, and some might "seep" to the back of the template; however, because $R_{water} - R_{tmpl}$ = 1 Å (see Figure 1) this escape of waters is avoided to a large extent, as can be learned from computer graphics.

The same behavior is shown in Table 3, where ΔF_{sum} is positive, 33, 21, 48, 14, and 14 kcal/mol for N_{water} = 80, 100, 120, 140, and 160, respectively; becoming negative, ΔF_{sum} = −20, −18, and −3 kcal/mol for N_{water} = 180, 200, and 220, i.e., for ρ_{water} = 0.0313, 0.0348, and 0.0383 Å^{-3}, respectively. Thus, as in Table 2, only ρ_{water} values close to the experimental density 0.0350 Å^{-3} (corresponding to 193 waters in the hemisphere) lead to a negative ΔF_{sum}, where the increase in N_{water} (as compared to R_{tmpl} = 12 Å) is due to the increase in R_{water} (from 13 to 14 Å).

We have also carried out calculations for R_{tmpl} = 11 and R_{water} = 12 Å but have found ΔF_{sum} to be positive (∼30 kcal/mol) for all values of N_{water}, even for $\rho_{water} \sim 0.0350$ Å^{-3}. This suggests that the template defined by R_{tmpl} = 11 Å (783 atoms) is too small. Indeed, computer graphics have shown that this template does not provide the required cover for the loop, and as a result some dihedral angles, especially for Arg, deviate significantly from their corresponding values in Table 1. On the other hand, the fact that water densities close to the experimental value ΔF_{sum} become negative for both R_{tmpl} = 12 and 13 Å suggests that a template defined by $R_{tmpl} \geq 12$ Å (944 atoms) would be adequate. The strongly fluctuating (negative) values ΔF_{sum} = −20, −18, and −3 kcal/mol obtained for R_{tmpl} = 13 Å and $N_{water} \geq 180$ probably reflect the difficulty to adequately optimize starting structures for MD simulations for these relatively large systems (see discussion in II.1 and II.2). Therefore, we calculate S_{loop} only for the smaller systems based on R_{tmpl} = 12 Å, where the (negative) ΔF_{sum} results are closer to each other.

Sometimes the energy rather than the free energy has been used in the literature as a criterion of stability; therefore, it is of interest to compare ΔF_{sum} to the corresponding values of ΔE_{total}. Tables 2 and 3 show that these quantities (while not equal) are correlated: both decrease as N_{water} increases with R^2 = 0.74 and 0.79, respectively; see Figures 2 and 3.

It should be pointed out that changes in N_{water} affect the energy values of the free microstate more than those of the bound one. Thus, in Table 2, $E_{loop-tmpl}$ is changed within the ranges δ_{free} = 102 − 52 = 50 and δ_{bound} = 128 − 120 = 8 kcal/mol, where the corresponding ranges for $E_{loop-loop}$ also differ significantly, δ_{free} = 46 − 10 = 36 versus δ_{bound} = −10 + 2 = 12 kcal/mol. Similar relations (but somewhat less pronounced) are observed in Table 3, where the ranges for $E_{loop-tmpl}$ are δ_{free} = 91 − 68 = 23 and δ_{bound} = 146 − 131 = 15 and for $E_{loop-loop}$ are δ_{free} =

TABLE 2: Results (in kcal/mol) for Energy and Free-Energy Components and Their Difference (Δ) between the Free and Bound Microstates for $R_{tmpl} = 12$ and $R_{water} = 13$ Å[a]

	$E_{loop-temp}$	$E_{loop-loop}$	F^{TI}_{water}(ch)	F^{TI}_{water}(LJ)	E_{loop}	F^{TI}_{water}	F_{sum}	E_{total}
			$N_{water} = 80$, $\rho_{water} = 0.0170$, $n_s = 40$					
free	−52	−46	−67	27	−98	−40	−138	−1207
bound	−128	−10	−48	17	−138	−31	−169	−1246
Δ^1	76 (1)	−36 (1)	−19 (1)	10 (1)	40 (1)	−9 (1)	31 (1)	39 (1)
			$N_{water} = 100$, $\rho_{water} = 0.0217$, $n_s = 40$					
free	−50	−45	−71	37	−95	−34	−129	−1389
bound	−128	−10	−48	27	−138	−21	−159	−1435
Δ^1	78 (1)	−35 (1)	−23 (1)	10 (1)	43 (1)	−13 (1)	30 (1)	46 (3)
			$N_{water} = 120$, $\rho_{water} = 0.0261$, $n_s = 40$					
free	−102	−25	−63	57	−127	−6	−133	−1646
bound	−128	−10	−50	41	−138	−9	−147	−1653
Δ^2	26 (1)	−15 (1)	−13 (2)	16 (1)	11 (2)	3 (3)	14 (2)	7 (1)
			$N_{water} = 140$, $\rho_{water} = 0.0304$, $n_s = 80$					
free	−75	−34	−76	99	−109	23	−86	−1788
bound	−118	−5	−50	95	−123	45	−78	−1791
Δ^3	43 (1)	−29 (1)	−26 (1)	4 (1)	14 (1)	−22 (1)	−8 (2)	3 (3)
			$N_{water} = 160$, $\rho_{water} = 0.0348$, $n_s = 80$ Double Integration					
free	−77	−10	−81	123	−87	42	−45	−1981
bound	−121	−1	−53	131	−122	78	−44	−1962
Δ^3	44 (1)	−9 (2)	−28 (1)	−8 (1)	35 (1)	−36 (2)	−1 (2)	−19 (5)
			$N_{water} = 180$, $\rho_{water} = 0.0391$, $n_s = 80$ Double Integration					
free	−71	−37	−84	148	−108	64	−44	−2155
bound	−119	1	−48	127	−118	79	−39	−2160
Δ^3	48 (1)	−38 (1)	−36 (1)	21 (4)	10 (1)	−15 (4)	−5 (4)	5 (2)

[a] The energies $E_{loop-tmpl}$ and $E_{loop-loop}$, their sum E_{loop}, and the total energy E_{total} are defined in eq 1. F^{TI}_{water}(ch) and F^{TI}_{water}(LJ) are free energies calculated by thermodynamic integration, where the charges and Lennard-Jones interactions, respectively, are gradually eliminated; they and their sum, F^{TI}_{water} are defined in eq 21. $F_{sum} = E_{loop} + F^{TI}_{water}$ is the total free energy (without the contribution of the loop entropy, S_{loop}). Δ is the difference free − bound. N_{water} is the number of capped water molecules. $\rho_{water} = N_{water}/[2\pi(R_{water})^3/3]$ is the density of water (in Å^{-3} units) in the hemisphere defined by the radius, R_{water}. n_s is the sample size. Statistical errors, $s/(n_s)^{1/2}$, where s is the standard deviation, are smaller than 2.5 and 1.5 kcal/mol for E_{total} and the other quantities, respectively. The errors in Δ appear in parentheses, e.g., −46 (2) = −46 ± 2; thus, the 2 kcal/mol in this example is the difference between the value which appears in the table and the value obtained for a smaller sample size, n_s: [1]for $n_s = 20$; [2]for $n_s = 17$; and [3]for $n_s = 40$. Double integration means that the MD simulation at each TI step is doubled; it is 40 ps for each λ_i step.

49 − 20 = 29 and $\delta_{bound} = 10 - 2 = 8$ kcal/mol. The results for F^{TI}_{water}(ch) show the same tendency, where $\delta_{free} = 84 - 63 = 21$ and $\delta_{bound} = 53 - 48 = 5$ (Table 2) and $\delta_{free} = 105 - 41 = 64$ and $\delta_{bound} = 62 - 47 = 15$ kcal/mol (Table 3). On the other hand, F^{TI}_{water}(LJ) increases in most cases with N_{water} for both microstates.

It is of interest to determine which energy components lead to the change of ΔF_{sum} from positive to negative at $N_{water} = 140$ (Table 2) and $N_{water} = 180$ (Table 3). Thus, for each energy (and free-energy) component in Table 2, we calculate two averages of Δ, one for the three lower values of N_{water} (80, 100, and 120) and the second for $N_{water} = 140$, 160, and 180 (these two averages are denoted by Δ_1). The calculations show that two components contribute to the decrease of ΔF_{sum}, while one component contributes slightly to its increase. More specifically, $\Delta_1(E_{loop-tmpl})$ decreases by 16 (from 60 to 44), $\Delta_1(F^{TI}_{water})$ decreases by 18 (from −6 to −24), while $\Delta_1(E_{loop-loop})$ increases (slightly) by 4 (from −29 to −25 kcal/mol). In Table 3, the first group consists of $N_{water} = 80$, 100, 120, 140, and 160 and the second group of $N_{water} = 180$, 200, and 220. Unlike in Table 2, $\Delta_1(E_{loop-tmpl})$ is practically the same for both groups; $\Delta_1(E_{loop-loop})$ increases slightly (as in Table 2) by 3 from −31 to −28, while $\Delta_1(F^{TI}_{water})$ decreases significantly by 41 (from ∼0 to −41 kcal/mol) and thus provides the sole contribution for the decrease of ΔF_{sum}. Thus, a consistent affect on the change of ΔF_{sum} is provided by the water component, F^{TI}_{water}. From a structural point of view, the results of Table 2 suggest that on average the loop moves (but not necessarily much) to decrease $\Delta_1(E_{loop-tmpl})$ [where $\Delta_1(F^{TI}_{water})$ is decreased as well]. In Table 3, the loop moves less, and its energy is unchanged, but $\Delta_1(F^{TI}_{water})$, which consists of the loop−water interactions, decreases significantly. It has been found difficult to verify this picture by a structural analysis. The consistent behavior of our model for the two templates (as a function of N_{water}) is reflected by the similar behavior of ΔF_{sum} for both templates (the Helmholtz free energy is the characteristic thermodynamic potential in the canonical ensemble and F_{sum} is very close to the total Helmholtz free energy). The fact that some components of ΔF_{sum} behave differently is not unexpected, since the template size, the number of buried crystal waters in the template, and the optimization of the water/template system are different for $R_{tmpl} = 12$ and 13 Å.

The discussion in the last two paragraphs demonstrates that N_{water} (which defines the water pressure) affects significantly the energy and free-energy components of the free and bound microstates (in particular it leads to the change in the sign of ΔF_{sum}). Hence, one would expect that these energetic changes will correlate with the local water density around the loop. Thus, we calculated (for $R_{tmpl} = 12$ Å) the average number of water molecules within spheres of radii 3, 4, 5, and 6 Å around the loop's residues. However, for each N_{water} studied ($N_{water} = 120$, 160, and 180) these numbers were found to be comparable for the free and bound microstates. On the other hand (as expected), the loop structures (in terms of dihedral angles) have been affected by N_{water} but not in a systematic way, and therefore, these changes are not discussed here.

TABLE 3: Results (in kcal/mol) for Energy and Free-Energy Components and Their Difference (Δ) between the Free and Bound Microstates for $R_{tmpl} = 13$ and $R_{water} = 14$ Å[a]

	$E_{loop-tmpl}$	$E_{loop-loop}$	F^{TI}_{water}(ch)	F^{TI}_{water}(LJ)	E_{loop}	F^{TI}_{water}	F_{sum}	E_{total}
			80 waters, $\rho_{water} = 0.0139$, $n_s = 40$					
free	−71	−41	−63	28	−112	−35	−147	−1247
bound	−141	−7	−47	15	−148	−32	−180	−1256
Δ^1	70 (1)	−34 (2)	−16 (1)	13 (1)	36 (2)	−3 (1)	33 (3)	9 (3)
			100 waters, $\rho_{water} = 0.0174$, $n_s = 40$					
free	−89	−49	−41	26	−138	−15	−153	−1469
bound	−142	−6	−47	21	−148	−26	−172	−1460
Δ^1	53 (1)	−43 (1)	6 (1)	5 (1)	10 (1)	11 (1)	21 (1)	−9 (5)
			120 waters, $\rho_{water} = 0.0209$, $n_s = 40$					
free	−89	−40	−58	61	−129	3	−126	−1668
bound	−144	−7	−52	29	−151	−23	−174	−1693
Δ^1	55 (1)	−33 (1)	−6 (1)	32 (1)	22 (1)	26 (1)	48 (1)	25 (1)
			140 waters, $\rho_{water} = 0.0244$, $n_s = 40$					
free	−86	−43	−71	+90	−129	19	−110	−1884
bound	−143	−6	−50	75 (4)	−149	25	−124	−1879
Δ^1	57 (2)	−37 (3)	−21 (1)	15 (1)	20 (1)	−6 (1)	14 (1)	−5 (1)
			160 waters, $\rho_{water} = 0.0278$, $n_s = 40$ Double Integration					
free	−91	−20	−81	84	−111	3	−108	−2075
bound	−146	−10	−50	84	−156	34	−122	−2087
Δ^2	54 (1)	−9 (1)	−31 (1)	0 (2)	45 (1)	−31 (1)	14 (2)	12 (3)
			180 waters, $\rho_{water} = 0.0313$, $n_s = 80$ Double Integrations					
free	−82	−21 (0)	−90	+94	−103	4	−99	−2285
bound	−131	−5 (4)	−49	106	−136	57	−79	−2274
Δ^3	48 (1)	−16 (1)	−41 (1)	−12 (1)	33 (1)	−53 (1)	−20 (2)	−11 (1)
			200 waters, ρ_{water}= 0.0348, n_s=40 Double Integration					
free	−68	−33	−106	122	−101	16	−85	−2478
bound	−136	−2	−60	126	−133	66	−67	−2452
Δ^2	68 (1)	−31 (1)	−46 (1)	−4 (1)	32 (1)	−50 (2)	−18 (2)	−33 (8)
			220 waters, $\rho_{water} = 0.0383$, $n_s = 40$					
free	−78	−42	−76	+137	−120	61	−59	−2584
bound	−134	−3	−62	+143	−137	81	−56	−2571
Δ^1	56 (1)	−39 (1)	−14 (1)	−6 (1)	17 (1)	−20 (1)	−3 (2)	−13 (3)

[a] For explanations, see Table 2. The errors in Δ appear in parentheses, e.g., −46 (2) = −46 ± 2; thus, the 2 kcal/mol is the difference between the value which appears in the table and the value obtained for a smaller sample size, n_s: [1]for $n_s = 20$; [2]for the largest difference obtained with $n_s = 20$ or $n_s = 40$ with single integration; and [3]for $n_s = 40$ with single integration.

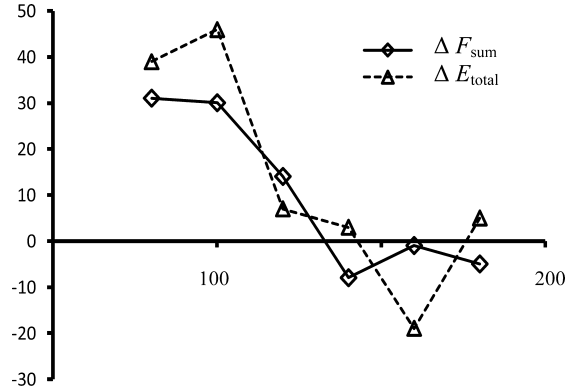

Figure 2. Graph demonstrating the correlation between results of Table 2 for $\Delta F_{sum} = \Delta E_{loop} + \Delta F^{TI}_{water}$ (eqs 19 and 22) and ΔE_{total} (eq 24), as a function of the number of water molecules, N_{water}.

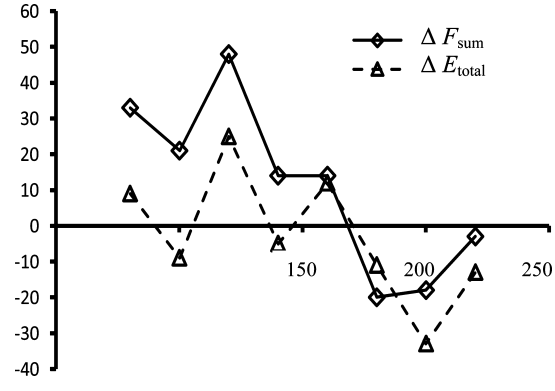

Figure 3. Graph demonstrating the correlation between results of Table 3 for $\Delta F_{sum} = \Delta E_{loop} + \Delta F^{TI}_{water}$ (eqs 19 and 22) and ΔE_{total} (eq 24), as a function of the number of water molecules, N_{water}.

III.4. Results for the Loop Entropy. Results for the loop entropy S^A_{loop} (eq 17), appear in Table 4 for $R_{tmpl} = 12$ and $R_{water} = 13$ Å. Two sets of results are presented for $N_{water} = 160$ and 180 for the free and bound microstates and for their difference $T[S^A_{loop}(\text{free}) - S^A_{loop}(\text{bound})] = T\Delta S^A_{loop}$ (see the discussion preceding eq 18a). These results were obtained by reconstructing $n_s = 80$ loop structures, distributed homogeneously along the entire sample of 1000 system configurations. The simulated future consists of the future part of the loop including all the surrounding water molecules. The results are presented for several values of n_f, the sample size of the future chains (eqs 13 and 14), where $n_f = 200$, 400, 1200, 1600, and 2000; these values of n_f are used for pairs of angles, such as a backbone dihedral and the successive bond angle. However, for the side chains we also reconstruct a single χ angle and triplets of successive χ angles (e.g., for Arg) for which the maximal n_f is 1000 and 4000 (rather than 2000), respectively (see eqs 13 and 14). The results are also presented as a function of bin size $\delta\alpha_k$

TABLE 4: HSMD Results (in kcal/mol) for the Loop Entropy and for Entropy Differences between the Free and Bound Microstates at T = 300[a]

		$N_{water} = 160$			$N_{water} = 180$		
		TS^{A}_{loop}			TS^{A}_{loop}		
bin size	n_f	free	bound	$T\Delta S^{A}_{loop}$	free	bound	$T\Delta S^{A}_{loop}$
$\Delta\alpha_k/5$	2000	64.3	61.9	2.4	63.3	63.2	0.1
$\Delta\alpha_k/10$	2000	61.2	59.5	1.7	60.0	60.2	−0.2
$\Delta\alpha_k/20$	2000	60.3	58.9	1.4	58.9	59.4	−0.5
$\Delta\alpha_k/30$	200	60.8	59.5	1.3	59.4	60.2	−0.9
	400	59.7	58.5	1.2	58.5	58.9	−0.4
	1200	60.0	58.8	1.2	58.6	59.2	−0.5
	1600	60.1	58.9	1.2	58.6	59.2	−0.6
	2000	60.1	58.9	1.2	58.7	59.3	−0.6
$\Delta\alpha_k/40$	200	60.7	59.6	1.2	59.3	60.3	−1.0
	400	59.7	58.5	1.1	58.5	58.9	−0.4
	1200	60.0	58.8	1.1	58.6	59.2	−0.6
	1600	60.0	58.9	1.2	58.6	59.2	−0.6
	2000	60.1	58.9	1.2	58.6	59.3	−0.6
$\Delta\alpha_k/50$	200	60.7	59.5	1.2	59.2	60.2	−1.0
	400	59.6	58.5	1.1	58.3	58.9	−0.6
	1200	59.9	58.8	1.1	58.5	59.1	−0.6
	1600	60.0	58.9	1.1	58.5	59.2	−0.7
	2000	60.1	58.9	1.2	58.5	59.2	−0.7
errors ≤		± 0.3	± 0.2		± 0.23	± 0.2	
converged		**60.1 ± 0.3**	**58.9 ± 0.2**	**1.2 ± 0.1**	**58.5 ± 0.23**	**59.2 ± 0.2**	**−0.7 ± 0.3**
TS^{QH}_{loop}		74.8 ± 0.3	72.1 ± 2	2.8 ± 0.8	73.0 ± 0.3	75.8 ± 1	−2.6 ± 2

[a] Results (based on $R_{tmpl} = 12$ and $R_{water} = 13$ Å) are presented for the loop entropy TS^{A}_{loop} (eqs 16 and 17) and for the differences $T\Delta S^{A}_{loop} = T[S^{A}_{loop}(free) - S^{A}_{loop}(bound)]$ (eq 18a); they were obtained by reconstructing 80 loop structures selected homogeneously from larger MD samples (of 1000 water/loop configurations) of the free and bound microstates obtained for $N_{water} = 160$ and 180. The results are calculated as a function of the bin size $\delta\alpha_k = \Delta\alpha_k/l$ (eq 12) and n_f (eqs 13 and 14), the sample size of the future chains used in the reconstruction process. S^{A}_{loop} is defined up to an additive constant that is expected to be the same for both microstates. The maximal statistical error in each column is denoted by errors ≤. The converged results for TS_{loop} and $T\Delta S_{loop}$ (eq 18b) are bold-faced. TS^{QH}_{loop} (eq A3) is the quasi-harmonic entropy, and these results were obtained from larger samples (see text for details).

$= \Delta\alpha_k/l$ (eqs 12−14) where $l = 30$, 40, and 50; while for $n_f = 2000$, we also provide results for bin sizes defined by $l = 5$, 10, and 20. The statistical errors were obtained from the fluctuations, and results obtained for a smaller sample of $n_s = 40$ configurations.

Being an upper bound, TS^{A}_{loop} (eq 18a) is expected to decrease as the approximation improves, i.e., with decreasing the bin size, an expectation that is fully satisfied. For example, for $N_{water} = 160$ and $n_f = 2000$, the TS^{A}_{loop}(free) values are 64.3, 61.2, 60.3, 60.1, 60.1, and 60.1 (kcal/mol + constant), i.e., they decrease for $l = 5$, 10, and 20 and converge to 60.1 (±0.3) kcal/mol for $l = 30$, 40, and 50. The same behavior is observed for all n_f values in the table, where in some cases the central values slightly decrease for $l = 40$ and 50 but should be considered as converged within the error bars. One would also expect TS^{A}_{loop} to decrease as n_f increases in each bin. The table reveals that such decrease always occurs in going from $n_f = 200$ to 400, but then a slight increase is observed for the larger n_f. This increase of TS^{A}_{loop} stems from the fact that for large n_f, the future part of the loop is expected to span larger regions of conformational space during the MD simulation and might also leave the original microstate. Therefore, the number of visits n_{visit} (eqs 13 and 14) to the bin decreases as compared to the number of visits obtained for $n_f = 200$ or 400, i.e., the probability decreases as well and S^{A}_{loop} increases.

This problem can be cured by dividing a (long) trajectory of size n_f into j shorter trajectories ("units") each based on $n'_f < n_f$ conformations, where $n_f = jn'_f$ and each unit starts from the reconstructed structure i with a different set of velocities followed by a short equilibration. In this procedure (which was carried out in our previous studies[11,12,34−36]), the future part of the loop is expected to remain within the original microstate. In any case, within the present statistical errors the results for TS^{A}_{loop} in Table 4 can be considered as converged to the correct values [for comparison, the four results for $TS^{A}_{loop}(m)$ (m stands for free or bound) for $N_{water} = 160$ and 180 based on smaller samples of $n_s = 40$ (obtained for $\Delta\alpha_k/50$, $n_f = 2000$) differ from those of Table 4 by 0.3, 0.2, 0.2, and 0.1 kcal/mol, respectively.] Moreover, we show below that differences in entropy ($T\Delta S_{loop}$), our main interest, are very stable. Finally, notice that our results are based on relatively small samples of $n_s = 80$ as compared to samples of $n_s \sim 600$ used in our previous studies, i.e., the present calculations led to a reduction in computer time by a factor of ∼7. Such a small sample is effective because it has been selected homogeneously from the larger sample of 1000 structures (based on a 0.5 ns MD trajectory).

The HSMD results for the entropy are compared to those obtained with the quasi-harmonic (QH) method from larger MD samples of 10 000 loop/water configurations, which are needed for achieving a reasonable precision. To avoid the "escape" of a sample from the original microstate, it consists of 10 separate samples of 1000 configurations (0.5 ns), each started from the same structure with a different set of initial velocities, where the initial trajectory of 0.5 ns is used for equilibration and is, thus, discarded. The central values of TS^{QH}_{loop} (eq A3 in the Appendix section) exceeded the HSMD results for TS_{loop} (for $n_s = 2000$) by ∼13−16 kcal/mol. These elevated results are in accord with S^{QH}_{loop} being an upper bound and are comparable to the overestimation of S^{QH}_{loop} values found in our previous studies.[11,12,34−36]

TABLE 5: Contributions (in kcal/mol) of the Loop and Water to the Energy, Entropy, and Free Energy and the Differences of These Values between the Free and Bound Microstates[a]

	E_{water}	TS_{water}	F^{TI}_{water}	E_{loop}	TS_{loop}	F_{loop}
			$N_{water} = 160$, $n_s = 80$ Double Integration			
free	-1893.3 ± 2		42.1 ± 3	-87.4 ± 1	60.1 ± 0.3	-147.5 ± 2
bound	-1839.7 ± 2		78.2 ± 3	-122.0 ± 1	58.5 ± 0.2	-180.5 ± 2
free−bound	ΔE_{water}	$T\Delta S_{water}$	ΔF_{water}	ΔE_{loop}	$T\Delta S_{loop}$	ΔF_{loop}
	-53.6 ± 3	-17.2 ± 2	-36.1 ± 2	$+34.6 \pm 1$	$+1.2 \pm 0.1$	$+33.0 \pm 1$
			$N_{water} = 180$, $n_s = 80$ Double Integration			
free	-2047.2 ± 2		64.5 ± 5	-107.6 ± 1	58.5 ± 0.2	-166.1 ± 1
bound	-2041.4 ± 2		79.3 ± 8	-118.1 ± 1	59.2 ± 0.2	-177.3 ± 1
free−bound	ΔE_{water}	$T\Delta S_{water}$	ΔF_{water}	ΔE_{loop}	$T\Delta S_{loop}$	ΔF_{loop}
	-5.8 ± 4	$+9.0 \pm 2$	-14.8 ± 4	$+10.5 \pm 1$	-0.7 ± 0.3	$+11.2 \pm 1$

[a] The water and loop energies, E_{water} and E_{loop} are defined in eq 1. F^{TI}_{water} (eq 21) is the water free energy obtained by a TI procedure; all these results are given with a higher precision than in Table 2. The loop entropy, TS_{loop} (eqs 16 and 17) and its difference $T\Delta S_{loop}$ (eq 18b) are taken from Table 4; $F_{loop} = E_{loop} - TS_{loop}$. $T\Delta S_{water}$ is obtained from $\Delta E_{water} - \Delta F_{water}$. In many cases, there is a decrease in the errors of the Δ values due to partial cancelation of errors in the differences (see text).

III.5. Differences in the Loop Entropy. As stated above, we are mostly interested in the results for the difference in entropy between the free and bound microstates, $T\Delta S_{loop}$ (eq 18b). Table 4 shows that for $N_{water} = 160$, the *converged* value is $T\Delta S_{loop}(n = 80) = 1.2 \pm 0.1$ kcal/mol, which "covers" the $T\Delta S^{A}_{loop}$ results (eq 18a) for $\Delta\alpha_k/l$, $l \geq 30$ for all n_f values, even for $n_f = 200$, i.e., the free microstate has the higher entropy. A table similar to Table 4 based on $n_s = 40$ (not shown) has led to $T\Delta S_{loop}(n_s = 40) = 1.3 \pm 0.2$ kcal/mol, which is equal to $T\Delta S_{loop}(n_s = 40)$ within the error bars. This result [which further supports our estimation of $T\Delta S_{loop}(n_s = 80)$] stems from the fact that the values of $S^{A}_{loop}(n_s = 40)$ for both microstates are systematically larger than those of $S^{A}_{loop}(n_s = 80)$, and these positive deviations are canceled in $T\Delta S^{A}_{loop}$.

For $N_{water} = 180$, $T\Delta S_{loop}(n_s = 80) = -0.7 \pm 0.3$ kcal/mol, representing the $T\Delta S^{A}_{loop}$ results for $\Delta\alpha_k/l$, $l \geq 30$ for all n_f values, i.e., the bound microstate, has the higher entropy. Here, the results for $S^{A}_{loop}(n_s = 80)$ for the free microstate are systematically higher than for $S^{A}_{loop}(n_s = 80)$ (as for N_{water}=160) while for the bound microstate, $S^{A}_{loop}(n_s = 40) < S^{A}_{loop}(n_s = 80)$, due to a significantly large contribution to the entropy of $S^{A}_{loop}(n_s = 80)$ by two structures that appear in the $n_s = 80$ sample but not in the $n_s = 40$ sample. In any case, based on $\Delta\alpha_k/l$, $l \geq 30$, we obtain $T\Delta S_{loop}(n_s = 40) = -0.3 \pm 0.3$ kcal/mol, which again is equal within the error bars to the value obtained from the $n_s = 80$ sample. Notice that the $T\Delta S^{QH}_{loop}$ values are significantly larger than their HSMD counterparts, while the signs are the same.

The computer time required to reconstruct a loop structure capped with $N_{water} = 160$ and 180 is 8.1 and 9.2 h CPU on a 2.1 GHz Athlon processor, meaning that the entire reconstructions required 1296 and 1472 h CPU. However, we have shown that considering only 10% ($n_f = 200$) of the maximal reconstruction samples and using smaller samples of $n_s = 40$ (rather than 80) have led to sufficiently accurate entropy differences, meaning that the total computer time can be reduced to 65 and 74 h CPU, respectively. We have generated the relatively large reconstruction samples to verify the convergence of the results.

In summary, the fact that $T\Delta S_{loop}$changes sign in going from $N_{water} = 160$ to 180 is not surprising since significant changes are also observed in other components of the energy in Tables 2 (e.g., $E_{loop-loop} = -10$ and -37 kcal/mol for $N_{water} = 160$ and 180, respectively). Also (like in previous studies), it is demonstrated that a limited future sampling in the reconstruction process (e.g., $n_f = 200$) is sufficient for obtaining the correct $T\Delta S_{loop}$, which enables one to reduce computer time significantly. This convergence of entropy differences stems from the cancelation (in $T\Delta S^{A}_{loop}$) of approximately equal systematic errors in S^{A}_{loop}(free) and S^{A}_{loop}(bound) as discussed in detail in section II.10 of ref 11.

III.6. Combined Results for the Entire Systems. In Table 5, we summarize the contribution of the loop and water to the free energy averaged over samples of of $n_s = 80$ configurations. We provide E_{water}, (eq 1) which includes the water−water, water−template, and water−loop interactions, and the contribution of water to the free energy, F^{TI}_{water} (eq 21). TS_{water} (which is not provided) can be obtained from E_{water} and F^{TI}_{water}. E_{loop}, which contains the loop−loop and loop−template interactions (eq 1), leads together with TS_{loop} (taken from Table 4) to F_{loop}. The entropy and free energy are defined up to additive constants, which are canceled in the differences (our main interest).

The table shows that for $N_{water} = 160$, the *absolute* values of ΔF_{water} and ΔF_{loop} are comparable; however, the contribution of $T\Delta S_{water}$ to ΔF_{water} is significant, being ~33% of ΔE_{water} (in absolute values), while the contribution of $T\Delta S_{loop}$ to ΔF_{loop} is small where $T\Delta S_{loop}$ constitutes only ~3.6% of ΔE_{loop}. For N_{water} =180, the situation is even more extreme, where the contribution of $T\Delta S_{water}$ to ΔF_{water} (in absolute values) is larger (by 155%) than that of ΔE_{water}, while the corresponding contribution of $T\Delta S_{loop}$ is again small (6.5%). In other words, for $N_{water} = 160$, E_{water}(free) < E_{water}(bound) significantly, and correspondingly also TS_{water}(free) < TS_{water}(bound) significantly. For $N_{water} = 180$, E_{water}(free) is only slightly smaller that E_{water}(bound), while TS_{water}(free) is *larger* than TS_{water}(bound). In any case, the effect of the entropy of water is significant. Also, the results of Table 5 and results in Tables 2 and 3 demonstrate the important contribution of water to the total free energy, where for water densities close to the experimental value ΔF^{TI}_{water} is always negative leading thereby to a negative ΔF_{total} (while ΔF_{loop} or ΔE_{loop} are always positive; see also Table 6).

The total contributions of $E_{total} = (E_{water} + E_{loop})$ and $TS_{total} = (TS_{water} + TS_{loop})$ to $F_{total} = (F_{water} + F_{loop})$ are summarized in Table 6 (again, for the entropy only, the results for $T\Delta S_{total}$ are given). For $N_{water} = 160$, ΔE_{total} and $T\Delta S_{total}$ are comparable with the same sign (as for water above) and, thus, lead to a small negative $\Delta F_{total} = -3.1 \pm 2.5$ kcal/mol. For $N_{water} = 180$, $\Delta F_{total} = -3.6 \pm 4$ kcal/mol is equal to the value obtained for $N_{water} = 160$ within a larger error; however, $\Delta F_{total}(180)$ is based on *positive* ΔE_{total} and $T\Delta S_{total}$ values (i.e., both the energy and entropy of the free microstate are larger than their bound counterparts). The fact that the entropic effects are significant means that (as the table also demonstrates) ΔE_{total} by itself does not constitute an adequate criterion of stability. We also provide

TABLE 6: Total Energy, Entropy, and Free Energy (in kcal/mol) at $T = 300$ K and Their Differences between the Free and Bound Microstates[a]

	E_{total}	TS_{total}	F_{total}	F_{sum}
	$N_{water} = 160$, $n_s = 80$ Double Integration			
free	-1980.7 ± 2		-105.4 ± 4	-45.3 ± 4
bound	-1961.7 ± 2		-102.3 ± 4	-43.8 ± 4
free−bound	ΔE_{total}	$T\Delta S_{total}$	ΔF_{total}	ΔF_{sum}
	-19.0 ± 2	-15.9 ± 4	-3.1 ± 2.5	-1.5 ± 2
	$N_{water} = 180$, $n_s = 80$ Double Integration			
free	-2154.8 ± 1		-101.6 ± 6	-43.1 ± 5
bound	-2159.5 ± 1		-98.0 ± 9	-38.8 ± 8
free−bound	ΔE_{total}	$T\Delta S_{total}$	ΔF_{total}	ΔF_{sum}
	$+4.7 \pm 2$	$+8.3 \pm 1.5$	-3.6 ± 4	-4.3 ± 4

[a] E_{total} (eq 1) and ΔE_{total} (eq 24) are the total energy and its difference for the free and bound microstates. F_{total} is the sum of the loop and water free energies, and its difference is ΔF_{total} (eq 26); $F_{sum} = F_{total} + TS_{loop}$ (F_{total} and F_{sum} are defined up to an additive constant). $T\Delta S_{total}$ (eq 28) is obtained from $\Delta E_{total} - \Delta F_{total}$. For information on the statistical error see Table 5 and the text.

in the table the results for ΔF_{sum} from Table 2, which are very close to those of F_{total} due to the small contribution of $T\Delta S_{loop}$, meaning that in these cases F_{sum} serves as a reliable measure of stability.

The above results for ΔF_{total} are equal to the experimental value, ~ -4 kcal/mol within the error bars. Furthermore, one would expect $|T\Delta S_{loop}(N_{water} = 140)|$ to be small, similar to the results obtained for $N_{water} = 160$ and 180 (this is based on values for $\Delta\alpha_k$ as those presented in Table 1). Therefore, $\Delta F_{total}(N_{water} = 140)$ is approximately represented by $\Delta F_{sum}(N_{water} = 140) = -8 \pm 2$ kcal/mol (see Table 2), which is close to the experimental value; the same is expected for the calculations based on $N_{water} = 220$ (for $R_{tmpl} = 13$ and $R_{water} = 14$ Å), where $\Delta F_{sum}(N_{water} = 220) = -3 \pm 2$ kcal/mol (Table 3). This agreement of our calculations with the experiment should be accepted with some caution because, for $R_{tmpl} = 13$ and $R_{water} = 14$ Å, the ΔF_{sum} values for $N_{water} = 180$ and 200 (-20 ± 2 and -18 ± 2 kcal/mol, respectively; see Table 4) are too low to lead to ΔF_{total} values close to the experimental result. However, the significant difference between the ΔF_{sum} results for $N_{water} = 180$, 200, and 220, suggests that the initial water configurations in the larger systems of $R_{tmpl} = 13$ and $R_{water} = 14$ Å have not sufficiently optimized (see section II.2). We find this global energy optimization to be the most uncertain, while the accuracy of the simulation results (including TI) is adequate.

IV. Summary and Conclusions

In the present paper, HSMD-TI has been applied to the mobile loop 287−290 (Ile, Phe, Arg, and Phe) of the protein acetylcholinesterase (AChE), where the difference in free energy between the free and bound structures of the loop, $F_{free} - F_{bound}$, has been estimated experimentally to be ~ -4 kcal/mol. In view of this result, the main objectives of the present paper have been related to modeling issues: do a fixed template and capped water constitute an adequate model? If they are, what is the minimal template size and minimal number of water molecules required to obtain reliable results? Another objective has been to further improve the efficiency of various components of HSMD-TI; in particular due to the fact that HSMD is applied for the first time to an internal loop consisting of residues with large side chains. To achieve these aims, we have carried out a systematic study consisting of several template sizes which are capped with 80−220 water molecules.

We have emphasized the difficulty to determine the optimal distribution of water. Thus, the hemisphere contains bulk water (which can be simulated adequately by MD) as well as *internal* water molecules that reside in crevices on the surface and inside the template; because some of these waters are practically immobile (by MD), they can be considered as part of the template, and their number, spatial positions, and orientations significantly affect the system energy. Using crystal water as internal ones might not always be adequate as our objective is to model the system in solution rather than in the crystal. Alternatively, one can seek to optimize (prior to the production simulations) the distribution of internal waters by energy minimization which, however, is an extremely difficult task. We have adopted a strategy where the initial configuration of water consists of $\sim$30 crystal water positions and water molecules distributed randomly in the hemisphere; this configuration is optimized by a procedure based on a series of high-temperature MD runs followed by energy minimizations. Clearly, finding the *global* energy minimum is practically unfeasible, but this is a general modeling problem not specific to HSMD-TI.

We have found that minimizing the energy of water before each integration step (during the TI process for calculating the contribution of water to the free energy) improves the convergence of the TI procedure significantly. Also, we have shown that satisfactory accuracy can be obtained by reconstructing a relatively small number of system configurations (80 or even 40), provided that they are selected homogeneously from the entire sample. This leads to a reduction in computer time by a factor of seven as compared to our previous studies. In the reconstruction of the side chain of Arg, three successive χ angles were treated successfully in each step, suggesting that four successive backbone angles (i.e., two pairs of a dihedral and the following bond angle) could be treated as well, which would decrease computer time further.

The limited number of atoms in the loop and waters and the (relatively small) *constant* template lead to relatively small statistical errors (which for an N-atom system increase as $N^{1/2}$). As in our previous studies (and in accord with theoretical arguments discussed in ref 11), we find that systematic errors in $S^{A}_{loop}(m)$ are canceled, to a large extent in differences ΔS^{A}_{loop} (eq 18a), and a similar cancelation occurs for the free energy of water ΔF_{water} and other energy components. It should be pointed out (that in agreement with our previous studies and ref 58) the quasi-harmonic approximation has been found to overestimate the entropy significantly, which might reflect strong long-range correlations and an harmonic effects within the loop due to the loop−template, loop−loop and loop−water interactions; also, the results for $|T\Delta S^{QH}_{loop}|$ overestimate the HSMD values for $|\Delta S_{loop}|$. Finally, notice that the calculations of the transition probabilities of different steps are completely independent, and they are also independent of the integration of water. Therefore, the reconstruction steps and TI of water can be fully parallelized.

The main conclusions from the present study (besides the above points which are mostly of a technical character) is that our approximate model is reliable, at least for the loop studied. This model is based on the same constant template for the free and bound microstates, where the loop is capped with a sphere of water molecules. By studying several template sizes and an increasing number of water molecules, we have found that to obtain consistent results for the free energy $\Delta F_{total} = F_{free} - F_{bound}$, the template should be larger than a minimal size, and the number of water molecules (in the hemisphere of $R_{water} -$

$R_{tmpl} = 1$ Å) should lead approximately to the experimental density of bulk water. Also, our results for ΔF_{total} agree with the experimental data ~ -4 kcal/mol. Our results demonstrate the important contribution of water to the total free energy, where for water densities close to the experimental value ΔF_{water} is always negative thereby leading to negative ΔF_{total} (while ΔF_{loop} is always positive). Also, the contribution of the water entropy, $T\Delta S_{water}$ to ΔF_{total} is significant. The next step would be to apply HSMD-TI to the free and bound loops attached to their own templates (defined by the PDB structures) rather than to a common template, as was done here.

V. Appendix

V.1. Thermodynamic Integration of Water. As described earlier in the TI process, the interaction energy (electrostatic and LJ) between a *fixed* loop structure and the (moving) water molecules is decreased gradually to zero (rather than increased from zero) at constant T and V, where the water−water and water−template potential energy is unchanged. For the (LJ) potential, we have used the shifted scaling potential introduced by Zacharias et al.:[59]

$$\varphi(r_{ij},\lambda) = \lambda 4\varepsilon\left[\frac{\sigma^{12}}{(r_{ij}^2 + \delta(1-\lambda))^6} - \frac{\sigma^6}{(r_{ij}^2 + \delta(1-\lambda))^3}\right] \quad \text{(A1)}$$

where the shift parameter, $\delta = 3$ Å^2, prevents the divergence of the potential (and its derivative) at small pair separations; a similar scaling function is used for the Coulomb interactions. The free-energy derivatives with respect to λ, $\partial F/\partial\lambda$ is

$$\frac{\partial F}{\partial \lambda} = \left\langle \frac{\partial E(\boldsymbol{x}^N,\lambda)}{\partial \lambda} \right\rangle_\lambda \quad \text{(A2)}$$

where the derivative of the energy is calculated analytically. The integration with respect to λ is carried out by dividing the range [1,0] into 20 equal integration bins $\Delta\lambda_i$. The ($\lambda = 1 \rightarrow \lambda = 0$) integration of the electrostatic interactions (i.e., charge elimination) is carried out first (in the presence of intact LJ interactions), followed by a $\lambda = 1 \rightarrow 0$ integration of the LJ interactions. Thus, the entire two-stage process is based on 40 $\partial F/\partial\lambda_i$ integration steps.

The MD simulation consists of a 2 fs integration step. Each ($\Delta\lambda_i$) step starts with energy minimization (based on λ_i) of the last structure obtained in the simulation of the $i - 1$ step, followed by 5 ps MD simulation for equilibration which is discarded; the next 20 ps of MD simulation, a configuration is retained every 0.02 ps, i.e., altogether 1000 configurations are used for evaluating $<\partial F/\partial\lambda_i>$. It should be pointed out that the energy minimization (which was not performed in paper 20) has contributed to a nice convergence of the integration results.

For a single loop structure, the free-energy integration requires approximately the same time for each procedure (electrostatic or LJ), i.e., $\sim 2 \times 9.2 = 18.4$ h CPU for $N_{water} = 160$ and $\sim 10.5 \times 2 = 21$ h CPU for $N_{water} = 180$ on a 2.1 GHz Athlon processor; these times refer to the *double* TI, where the MD simulation at each step is doubled, i.e., it is based on 40 ps for each λ_i (see Tables 2 and 3). These simulation lengths are adequate because the loop's conformation is kept fixed (as well as the template), and only the water molecules are moved by MD during integration.

V.2. The QH Methods. With the QH method introduced by Karplus and Kushick,[20] the Boltzmann probability density of structures defining a microstate is approximated by a multivariate Gaussian. Thus

$$S_{loop}^{QH}(m) = (k_B/2)\{N + \ln[(2\pi)^N \mathrm{Det}(\boldsymbol{\sigma})]\} \quad \text{(A3)}$$

where the covariance matrix, $\boldsymbol{\sigma}$, is obtained from a local MD (MC) sample, and N is (usually) the number of internal coordinates. Clearly, S^{QH} constitutes an upper bound for S since correlations higher than quadratic are neglected; also, an harmonic contributions are ignored, and QH is not suitable for diffusive systems such as water. While QH has been used extensively through the years, a systematic study of its performance has been carried out only recently by Gilson's group,[55] who have found that the performance of QH deteriorates significantly in Cartesian coordinates and when applied to more than one microstate.[7]

Acknowledgment. This work was supported by NIH grant 2-R01 GM066090-4 A2.

References and Notes

(1) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 43.
(2) Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395.
(3) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184.
(4) Meirovitch, H. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1998, Vol. 12, p 1.
(5) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047.
(6) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* **2003**, *107*, 9535.
(7) Meirovitch, H. *Curr. Opin. Struct. Biol.* **2007**, *17*, 181.
(8) Gilson, M. K.; Zhou, H.-X. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21.
(9) Foloppe, N.; Hubbard, R. *Curr. Med. Chem.* **2007**, *13*, 3583.
(10) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P. J.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. *Angew. Chem., Int. Ed.* **2006**, *45*, 406.
(11) Cheluvaraja, S.; Meirovitch, H. *J. Chem. Theory Comput.* **2008**, *4*, 192.
(12) Cheluvaraja, S.; Mihailescu, M.; Meirovitch, H. *J. Phys, Chem. B* **2008**, *112*, 9512.
(13) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1959**, *31*, 459.
(14) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature (London)* **1977**, *267*, 585.
(15) Elber, R.; Karplus, M. *Science* **1987**, *235*, 318.
(16) Stillinger, F. H.; Weber, T. A. *Scienc* **1984**, *225*, 983.
(17) Gō, N.; Scheraga, H. A. *J. Chem. Phys.* **1969**, *51*, 4751.
(18) Gō, N.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 535.
(19) Hagler, A. T.; Stern, P. S.; Sharon, R.; Becker, J. M.; Naider, F. *J. Am. Chem. Soc.* **1979**, *101*, 684.
(20) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325.
(21) Schlitter, J. *Chem. Phys. Lett.* **1993**, *215*, 617.
(22) Andricioaei, I.; Kaplus, M. *J. Chem. Phys.* **2001**, *115*, 6289.
(23) Meirovitch, H. *Chem. Phys. Lett.* **1977**, *45*, 389.
(24) Meirovitch, H.; Vásquez, M.; Scheraga, H. A. *Biopolymers* **1987**, *26*, 651.
(25) Meirovitch, H.; Koerber, S. C.; Rivier, J.; Hagler, A. T. *Biopolymers* **1994**, *34*, 815.
(26) Meirovitch, H. *Phys. Rev. A: St., Mol., Opt. Phys.* **1985**, *32*, 3709.
(27) Meirovitch, H.; Scheraga, H. A. *J. Chem. Phys.* **1986**, *84*, 6369.
(28) Meirovitch, H. *J. Chem. Phys.* **2001**, *114*, 3859.
(29) Hnizdo, V.; Darian, E.; Fedorowicz, A.; Demchuk, E.; Li, S.; Singh, H. *J. Comput. Chem.* **2007**, *28*, 65.
(30) Killian, B. J.; Kravitz, J. Y.; Gilson, M. K. *J. Chem. Phys.* **2007**, *127*, 024107.
(31) White, R. P.; Meirovitch, H. *J. Chem. Phys.* **2004**, *121*, 10889.
(32) White, R. P.; Meirovitch, H. *J. Chem. Phys.* **2006**, *124*, 204108.
(33) White, R. P.; Meirovitch, H. *J. Chem. Phys.* **2005**, *123*, 214908.
(34) Cheluvaraja, S.; Meirovitch, H. *J. Chem. Phys.* **2005**, *122*, 054903.
(35) Cheluvaraja, S.; Meirovitch, H. *J. Phys.Chem. B* **2005**, *109*, 21963.
(36) Cheluvaraja, S.; Meirovitch, H. *J. Chem. Phys.* **2006**, *125*, 024905.
(37) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; 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*, 517.
(38) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem.* **1997**, *101*, 3005.
(39) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 92.

(40) Steinbach, P. J.; Brooks, B. R. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9135.
(41) White, R. P.; Meirovitch, H. *J. Chem. Theory Comput.* **2006**, *2*, 1135.
(42) Rosenberry, T. L. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1975**, *43*, 103.
(43) Millard, C. B.; Kryger, G.; Ordentlich, A.; Greenblatt, H. M.; Harel, M.; Raves, M. L.; Segall, Y.; Barak, D.; Shafferman, A.; Silman, I.; Sussman, J. L. *Biochemistry* **1999**, *38*, 703.
(44) Millard, C. B.; Koellner, G.; Ordentlich, A.; Shafferman, A.; Silman, I.; Sussman, J. L. *J. Am. Chem. Soc.* **1999**, *121*, 988.
(45) Chiu, Y. C.; Main, A. R.; Dauterman, W. C. *Biochem. Pharmacol.* **1969**, *18*, 2171.
(46) Hart, G. J.; O'Brien, R. D. *Biochemistry* **1973**, *12*, 2940.
(47) Forsberg, A.; Puu, G. *Eur. J. Biochem.* **1984**, *140*, 153.
(48) Ordentlich, A.; Barak, D.; Kronman, C.; Flashner, Y.; Leitner, M.; Segall, Y.; Ariel, N.; Cohen, S.; Velan, B.; Shafferman, A. *J. Biol. Chem.* **1996**, *268*, 1708.
(49) Ordentlich, A.; Kronman, C.; Barak, D.; Stein, D.; Ariel, N.; Marcus, D.; Velan, B.; Shafferman, A. *FEBS Lett.* **1993**, *334*, 21.
(50) Raves, M. L.; Harel, M.; Pang, Y.-P.; Silman, I.; Kozikowski, A. P.; Sussman, J. L. *Nat. Struct. Biol.* **1997**, *4*, 57.
(51) Carlacci, L.; Millard, C. B.; Olson, M. A. *Biophys. Chem.* **2004**, *111*, 143.
(52) Olson, M. A. *Proteins* **2004**, *57*, 645.
(53) Ponder, J. W. *TINKER - software tools for molecular design*, Version*3.9*;Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine: St. Louis, MO, 2001.
(54) Meirovitch, H.; Alexandrowicz, Z. *J. Stat. Phys.* **1976**, *15*, 123.
(55) Meirovitch, H. *J. Chem. Phys.* **1999**, *111*, 7215.
(56) Meirovitch, H.; Vásquez, M.; Scheraga, H. A. *Biopolymers* **1988**, *27*, 1189.
(57) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarenden Press: Oxford, U.K., 1987.
(58) Chang, C. E.; Chen, W.; Gilson, M. K. *J. Chem. Theory Comput.* **2005**, *1*, 1017.
(59) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *100*, 9025.

JP900308Y