



Published in final edited form as:

J Phys Chem B. 2008 August 7; 112(31): 9512–9522. doi:10.1021/jp801827f.

Entropy and Free Energy of a Mobile Protein Loop in Explicit Water

Srinath Cheluvvaraja, Mihail Mihailescu, and Hagai Meirovitch*

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

Abstract

Estimation of the energy from a given Boltzmann sample is straightforward since one just has to average the contribution of the *individual* configurations. On the other hand, calculation of the *absolute* entropy, S (hence the absolute free energy F) is difficult because it depends on the entire (unknown) ensemble. We have developed a new method called, “the hypothetical scanning molecular dynamics” (HSMD) for calculating the absolute S from a given sample (generated by any simulation technique). In other words, S (like the energy) is “written” on the sample configurations, where HSMD provides a prescription of how to “read” it. In practice, each sample conformation, i is reconstructed with transition probabilities and their product leads to the probability of i , hence to the entropy. HSMD is an exact method where all interactions are considered and the only approximation is due to insufficient sampling. In previous studies HSMD (and HS Monte Carlo – HSMC) has been extended systematically to systems of increasing complexity, where the most recent is the 7-residue mobile loop, 304–310 (Gly-His-Gly-Ala-Gly-Gly-Ser) of the enzyme porcine pancreatic α -amylase modeled by the AMBER force field and AMBER with the implicit solvation GB/SA (paper I). In the present paper we make a step further and extend HSMD to the same loop capped with TIP3P explicit water at 300 K. As in paper I, we are mainly interested in entropy and free energy differences between the free and bound microstates of the loop, which are obtained from two *separate* MD samples of these microstates. The contribution of the loop to S and F is calculated by HSMD and that of water by a particular thermodynamic integration procedure. As expected, the free microstate is more stable than the bound microstate by a total free energy difference, $F_{\text{free}} - F_{\text{bound}} = -4.8 \pm 1$, as compared to -25.5 kcal/mol obtained with GB/SA. We find that relatively large systematic errors in the loop entropies, $S_{\text{free}}(\text{loop})$ and $S_{\text{bound}}(\text{loop})$ are cancelled in their difference which is thus obtained efficiently and with high accuracy, i.e. with a statistical error of 0.1 kcal/mol. This cancellation, which has been observed in previous HSMD studies, is in accord with theoretical arguments given in paper I.

I. Introduction

I.1. The difficulty in calculating the absolute entropy

The commonly used simulation methods, Metropolis Monte Carlo (MC)¹ and molecular dynamics^{2,3} (MD) enable one a direct calculation of quantities such as the energy E_i that are “written” on a simulated configurations i . However, these methods (due to their dynamic character) do not provide the *absolute* entropy S (hence the absolute Helmholtz free energy F , $F=E-TS$, where T is the absolute temperature) in a straightforward manner. More specifically, $S = -k_B \sum_i P_i^B \ln P_i^B$, where k_B is the Boltzmann constant and P_i^B is the Boltzmann probability of configuration i

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

$$P_i^B = \exp[-E_i/k_B T]/Z. \quad (1)$$

However, P_i^B depends not only on i but also on the whole ensemble through the partition function Z , ($Z = \sum \exp[-E_i/k_B T]$), which cannot be obtained directly from a finite sample. Therefore, in spite of the progress achieved during the years, calculation of S from a single MC or MD sample still remains a difficult problem.

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). HSMC(D) is based on ideas of previous methods suggested by Meirovitch, the local states (LS)⁴⁻⁶ and the hypothetical scanning (HS).⁷⁻⁹ HSMC(D) has been developed systematically as applied to liquid argon, TIP3P water,^{10,11} self-avoiding walks on a square lattice,¹² and peptides,¹³⁻¹⁵ where for the first three models HSMC(D) results have been found to agree within error bars to thermodynamic integration 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 HSMC¹¹⁸⁻¹²⁰ (IV.C-E).

Very recently HSMD has been applied successfully to a mobile loop of the protein α -amylase,¹⁶ where the system was modeled by the AMBER96 force field¹⁷ alone and the AMBER96 force field with the GB/SA implicit solvent of Still and coworkers;¹⁸ this paper (Ref. 16) is referred here as paper I. In the present paper we make a step further in the development of HSMD, where GB/SA is replaced by explicit water, i.e., the same loop is capped with TIP3P¹⁹ water molecules.

1.2. Microstates of biological macromolecules

Before discussing the loop and HSMC(D) further, one should emphasize the importance of the free energy in structural biology as a criterion of stability. Bio-macromolecules such as proteins have rugged potential energy surface, $E(\mathbf{x})$ (\mathbf{x} is the $3N$ -dimensional vector of the Cartesian coordinates of the molecule's N atoms), which is "decorated" by a tremendous number of *localized* wells and "wider" wells, defined over regions, Ω_m , which we call microstates; thus, each microstate consists of many localized wells (an example for a microstate is the α -helical region of a peptide). 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* MD simulation starting from a structure that belongs 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,^{20,21} meaning that the microstates are of a greater physical significance than the localized wells (see also I.4 below).

A central aim of computational structural biology is to identify the most stable microstates, i.e., those with the largest *conformational* partition function Z_m (or equivalently with lowest Helmholtz free energy, F_m)

$$F_m = -k_B T \ln Z_m = -k_B T \ln \int_m \exp[-E(\mathbf{x})/k_B T] d\mathbf{x} \quad (2)$$

where the integration is carried out over the limited microstate Ω_m , rather than over Ω (for simplicity we shall denote in most cases a microstate Ω_m by m). Thus, the protein folding problem is the notoriously difficult task of identifying the microstate with the global minimum F_m , which practically might be achieved by two challenging stages: (1) identifying an initial

set of microstates with expected high stability (e.g., based on an energetic criterion), and (2) calculating their relative populations, p_m/p_n [$p_m = \exp[-F_m/k_B T]/Z$], which leads to minimum F_m ,

$$p_m/p_n = Z_m/Z_n = \exp - [\Delta F_{mn}/k_B T] \quad (3)$$

where $\Delta F_{mn} = F_m - F_n$.

Calculation of relative populations is also required in problems which are less challenging than protein folding, i.e., in cases of *intermediate flexibility*, where a flexible protein segment (e.g., a side chain or a surface loop), a cyclic peptide, or a ligand bound to an enzyme populates significantly several microstates in thermodynamic equilibrium. It is of interest to know whether the conformational change adopted by a loop (a side chain, ligand, etc.) upon binding has been induced by the other protein (induced fit^{22,23}) or alternatively the free loop already interconverts among different microstates where one of them is selected upon binding (selected fit²⁴). This analysis requires calculating p_m values, which are also needed for a correct analysis of NMR and x-ray data of flexible macromolecules.^{25–28} Calculation of F is essential in many other biological processes. Thus, F determines the binding affinities of protein-protein interactions, it is an important factor in enzymatic reactions, electron transfer, and ion transport through membranes, and it leads to the solubilities of small molecules.

I.3. Advantages of the absolute F and S

The examples discussed above require calculating only the difference in free energy ΔF_{mn} (rather than the absolute F) which can be obtained, in principle, by applying thermodynamic integration (TI) techniques or a counting method which leads to $\Delta F_{mn} = -k_B T \ln[(\#m)/(\#n)]$, where $\#m$ ($\#n$) are the populations of m and n obtained from a long MD trajectory.^{29–35} In paper I¹⁶ we discuss these methods, emphasizing their limitations where m and n are separated by high energy barriers, which demonstrates the need for methods for calculating the *absolute F_m from a given sample*; with such methods, one will be able 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 or the long runs needed for the counting method are avoided.

The absolute S and F can be calculated by the harmonic^{36–38} and quasi-harmonic³⁹ approximations and can also be obtained by TI provided that a reference state R with known F_R is available and an efficient integration path $R \rightarrow m$ can be defined. However, for non-homogeneous systems such integration might not be trivial, and in models of peptides and proteins defining adequate reference states is a difficult problem (for further discussions on these and other methods, see paper I¹⁶ and Ref.35).

With the LS, HS and HSMC(D) techniques mentioned earlier, each conformation i of a sample (generated by MC MD or any other technique) is *reconstructed* step-by-step (from nothing) using transition probabilities (TPs). The product of these TPs leads to an approximation for the correct Boltzmann probability P_i^B (eq 1) 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. From this respect HSMC(D) (unlike HS and LS) can be viewed as exact;¹⁰ 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. 10).

I.4. Problems to define microstates by computer simulation

Thus far we have dealt with microstates (and their populations) without providing a *practical* definition for them. This, however, is not a straightforward task which has been ignored to a large extent in the literature but has been given considerable thoughts by us in the course of the years.^{5,6,26,40–43} (see also paper I¹⁶). To illustrate the problem assume a peptide model based on constant bond lengths and bond angles in a helical microstate Ω_h , i.e., the dihedral angles φ_i and ψ_i are expected to vary within relatively small ranges $\Delta\varphi_i$ and $\Delta\psi_i$ around $\varphi_i = -60^\circ$ and $\psi_i = -50^\circ$ (we ignore for a moment the side chains). However, the correct limits of Ω_h in terms of $[\varphi_i, \psi_i]$ are unknown because the strongly correlated angles define a complicated narrow “pipe” within the region, $\Delta\varphi_1 \times \Delta\psi_1 \times \Delta\varphi_2 \times \Delta\psi_2 \cdots \Delta\varphi_N \times \Delta\psi_N$. Obviously, these correlations are taken into account by an exact simulation method and thus, in practice, Ω_h can be defined (or more correctly, represented) by a *local* MC (MD) sample of conformations initiated from an α -helical structure.

However, this definition should be used with caution. Thus, a short simulation will span only a small part of Ω_h and this part will grow constantly as the simulation continues; correspondingly, the calculated average potential energy, E_h and the entropy S_h (obtained by any method) will both increase and the free energy, F_h is expected to change as well. As the simulation time is increased further, side chain dihedrals will “jump” to different rotamers, which according to our definition should also be included within Ω_h ; for a long enough simulation the peptide is expected to “leave” the α -helical region moving to a different microstate. Thus, *in practice*, the microstate size and the corresponding thermodynamic quantities depend on the simulation time t . Therefore, in practice there is always some arbitrariness in the definition of a microstate, which affects the calculated averages. This arbitrariness is severe with some methods and can be controlled (minimized) by others.

Because the size of m (n) depends on t , calculation of differences $S_m - S_n$ ($F_m - F_n$) from their absolute values [with QH, LS, or HSMC(D)] will depend on t as well where (as mentioned above) *typically* m and its energy and entropy all grow with t . To be able to carry out reliable estimation of ΔS_{mn} (ΔF_{mn} , etc.) we simulate both m and n for the same t looking for a range of t values where $\Delta F_{mn}(t)$, $\Delta S_{mn}(t)$ and $\Delta E_{mn}(t)$ are stable within the statistical errors [due to simultaneous increase of $E_m(t)$, $E_n(t)$, etc.]. With HSMC(D) one can calculate a series of improving approximations denoted $S^A(t_r)$ by increasing the reconstruction time t_r which leads to improving differences $\Delta S^A_{mn}(t_r)$ [and $\Delta F^A_{mn}(t_r)$]; if these differences converge within the statistical errors, the converged values are considered to be the correct differences due to cancellation of equal systematic errors in $S^A_m(t_r)$ and $S^A_n(t_r)$ (a similar procedure is applicable with LS) (see a detailed discussion in II.10 of paper I¹⁶).

Obviously, if m is less stable than n the t values should be adjusted (i.e., decreased) to fit the stability of m . If m is significantly larger than n , t_m should be large enough to allow an adequate coverage of m . However, if $\Delta S_{mn}(t)$ increases monotonically it constitutes a lower bound. If the microstate is restrictive, e.g., side chains should populate a single rotamer, the MD sample can be composed of several smaller samples each starts from the same structure with a different set of velocities. One should always verify that the samples remain in the original microstates and have not “escaped” to neighbor ones. We have developed methods for analyzing the stability of a microstate by calculating distribution profiles of dihedral angles.^{26,41,43}

I.5. A mobile loop in porcine pancreatic α -amylase

As in paper I¹⁶, we apply HSMD to two structures (microstates) of the flexible surface loop 304–310 (Gly-His-Gly-Ala-Gly-Gly-Ser) of the enzyme porcine pancreatic α -amylase (PPA). PPA is a single polypeptide chain of 496 amino acid residues^{44–47} consisting of three structural domains, domain A (residues 1–99, 170–404), domain B (residues 100–169) and

domain C (residues 405–496). Domain A adopts a $(\beta/\alpha)_8$ barrel structure and contains the three catalytic residues Asp197, Glu233 and Asp300. A deep cleft in this domain is accepted to be the substrate-binding site.^{44–50} An essential chloride ion and a calcium ion are located closer to this V-shaped depression and have been suggested to enhance the catalytic activity.^{50–53} (for more details about PPA see Paper I).

In the crystal structures of the free protein (PPA I⁴⁴ and II⁴⁷ which differ by two residues) the above loop has larger B-factors than the average B-factors of the atoms in the protein. However, in the crystal structures of PPA I complexed with acarbose⁴⁵ and PPA II complexed with V-1532⁴⁷ the B-factors of this loop are close to the average value in the protein where the loop has moved toward the active site. The maximum main-chain movement is ~ 5 Å at His305, which approaches the inhibitor from the solvent side to make a hydrogen bond with a glucose residue. The outcome of this movement is an apparent closure of the surface edge of the cleft.⁴⁵ Subsequently, several hypotheses have been put forward with respect to the function of the mobile loop in α -amylases, such as providing assistance in holding the glucose residues in a favorable orientation during catalysis,⁴⁵ or assisting in the transition state,⁵⁴ or inducing a trap-release mechanism of substrate and products.⁵⁵

In paper I we carried out two MD simulations, starting from the x-ray structures of the free and complexed PPA II, 1pif and 1pig, respectively,⁴⁷ which spanned the corresponding microstates and the entropy and free energy was calculated by HSMD. In this initial study the loop was modeled by the AMBER96 force field¹⁷ alone (where solvation effects are not considered) and by the AMBER96 and the highly approximate GB/SA implicit solvent.¹⁸ In the present work we make an additional step in the development of HSMD by extending it to the above loop capped by TIP3P water molecules.¹⁹

II. Theory and methodology

II.1. The loop and the protein's template

As was pointed out above we study the 7-residue loop, 304–310 of PPA in two microstates related to the free and bound loop structures; the starting point is the available crystal structures of PPA II, 1pif and 1pig,⁴⁷ respectively. Because the structures of these proteins are almost identical, we have chosen (as in paper I) to carry out the calculations with the 1pif structure, where the loop structure of 1pig is attached to the 1pif structure by superimposing the structure of 1pig on that of 1pif (the ligand was discarded). This might indicate whether the transition of the loop to the bound microstate constitutes a selected fit, i.e., whether this microstate is reachable in the free protein. PPA is a relatively large protein and it would be computationally unfeasible to include all of its atoms in the calculations. Therefore, exactly as in paper I, we consider only a *template* of 700 atoms (the same atoms for the bound and free structures) that are close to the loop where the rest of the protein's atoms are ignored. Using the same template as in paper I will allow comparing the present results to those obtained in paper I with implicit solvent. The construction of the template is described in some detail below.

Thus (see Figure 1 and Ref. 56), the center of mass of the loop backbone atoms (in the x-ray structure of 1pif) is calculated as a reference point denoted \mathbf{x}_{cmb} . A distance (R_{temp}) is chosen such that if the distance of any atom of a residue from \mathbf{x}_{cmb} is less than R_{temp} , the entire residue is included in the template. Otherwise, the residue is eliminated. As in paper I, $R_{\text{temp}} = 12$ Å (which leads to a template of 700 atoms). Then, the loop and template atoms are relaxed to a nearby geometry. This minimization is carried out using harmonic positional restraints with force constant of $5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, which are applied to all heavy atoms. This eliminates bad atomic overlaps and strains in the original structure, while keeping the atoms still reasonably close to the PDB coordinates. All these procedures are exactly the same as in paper I.

The treatment of water has been discussed in Ref. 56, where the effect of minimalist explicit solvation models for surface loops in proteins has been studied, following the original work of Steinbach and B. Brooks.⁵⁷ In Ref. 56 we performed MD simulations of surface loops capped by TIP3P water,¹⁹ in the presence of fixed templates, where the prime focus of the study was to check the performance of the small numbers (e.g., ~100) of water molecules employed. The number of water molecules, N was systematically varied, and convergence with large N was monitored to reveal the minimum number required for the loop to exhibit realistic (fully hydrated) behavior. It was found that the loop backbone can stabilize with a surprisingly small number of water molecules (as low as 5 molecules per amino acid residue). The side chains require somewhat larger N , roughly 12 water molecules per residue. The importance of this result lies in the fact that at this hydration level, computational times are comparable to those required for GB/SA,¹⁸ while the “minimalist explicit models” are expected to provide a viable and potentially more accurate alternative. In the present paper we have chosen to apply initially $N=70$ due to the fact that five out of the seven residues are Gly and Ala, which practically are without side chains. However, we have also carried out calculations using $N=120$ (see III.4).

To hold these waters around the loop they are restrained with a flat-welled half-harmonic potential (a force constant of $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$), based on the distance from the “center” of the loop region. That is, the distance of each water molecule (in practice, the oxygen atom) is measured from a restraining center denoted \mathbf{x}_{sph} . If this distance is greater than a prescribed distance, R_{cap} , a harmonic restoring force is applied, otherwise the restraining force is zero. A reasonable restraining center could be, for example, the center of mass of the loop backbone atoms (i.e. $\mathbf{x}_{\text{sph}} = \mathbf{x}_{\text{cmb}}$). However, we have found that our template is too “thin” and during MD simulations water molecules percolate through cavities in the template to its “back side”. To avoid this undesired situation we have chosen $R_{\text{cap}}=14 \text{ \AA}$ (i.e. R_{cap} is larger than R_{temp}) and as suggested in Ref 56, \mathbf{x}_{sph} was defined by,

$$\mathbf{x}_{\text{sph}} = \mathbf{x}_{\text{cmb}} + r_{\text{shift}}(\mathbf{x}_{\text{cmb}} - \mathbf{x}_{\text{cm}}) / |\mathbf{x}_{\text{cmb}} - \mathbf{x}_{\text{cm}}| \quad (4)$$

where \mathbf{x}_{cm} is the overall center of mass of the loop-template system. Here, the effect is to shift the center of the restraining sphere (\mathbf{x}_{sph}) toward the “loop side” of the loop-template system by $r_{\text{shift}} \text{ \AA}$ (see Figure 1). However, because changing r_{shift} affects the energy of the system one would seek to adopt its smallest value that still avoids the percolation of water. We carried out several MD simulations (500 and 1000 ps long) for $R_{\text{cap}}=14 \text{ \AA}$ and $r_{\text{shift}} = 2 - 7, 9, 15,$ and 25 \AA calculating the various energy components; it has been found that water seeping is eliminated for $r_{\text{shift}} \geq 4 \text{ \AA}$, but as a step of precaution have decided to use $r_{\text{shift}}=5 \text{ \AA}$, for which the energy results are similar to those obtained with $r_{\text{shift}} = 4 \text{ \AA}$.

As in paper I, the potential energy is defined by the AMBER96 force field,¹⁷ the His residue is protonated in the free and bound states. The reconstruction of the loop structure is carried out in internal coordinates, therefore, the conformations simulated by MD should be transferred from Cartesians to the dihedral angles $\varphi_i, \psi_i,$ and ω_i ($i=1, N=7$), the bond angles $\theta_{i,l}$ ($i=1, N, l=1,3$), the side chain angles χ and the corresponding bond angles. As in paper I we consider only three χ angles two of His and one of Ser, while the contribution of the side chain of Ala is ignored. Also, because the side chains are much shorter than the backbone and are not restricted by the loop closure condition, the effect of their bond angles on entropy *differences* is expected to be small and is thus ignored; we have argued in paper I that to a good approximation bond stretching can be ignored as well. For convenience, these angles (ordered along the backbone) are denoted by $\alpha_k, k=1, 45=K$. The MD runs are carried out with the package

TINKER,⁵⁸ where the loop and the capped TIP3P waters are free to move while the template is kept fixed in its x-ray coordinates and the total potential energy E is

$$E = E_{\text{loop}} + E_{\text{water}} \quad (5)$$

where E_{loop} includes the loop-loop and loop-template potential energy and E_{water} consists of the water-water, water-loop, and water-template interactions (the template-template energy is constant and thus is ignored).

II.2. Statistical mechanics of a loop in internal coordinates

The partition function of the loop/water system Z (eq 1) is an integration of $\exp[-E/k_B T]$ with respect to \mathbf{x}_{loop} , the Cartesian coordinates, of the loop and \mathbf{x}^N , the $3N$ coordinates of the N water molecules over a microstate m (for brevity we omit the letter m in most equations). However, it is convenient to change the variables of integration from \mathbf{x}_{loop} to internal coordinates, α_k , $k=1, K$ which makes the integral dependent also on a Jacobian, J , which for a linear chain has been shown to be a simple function of the bond angles and bond lengths independent of the dihedral angles.^{36,37,39} 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 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 correlations with the α_k) and the remaining integral becomes a function of the K dihedral and bond angles (α_k)^{36,37,39} and a Jacobian that depends only on the bond angles; the partition function is

$$Z' = DZ = D \int_m \exp(-E_{\text{loop}}([\alpha_k]) - E_{\text{water}}([\alpha_k], \mathbf{x}^N)) / k_B T d\alpha_1 \dots d\alpha_K d\mathbf{x}^N, \quad (6)$$

where $[\alpha_k] = [\alpha_1, \dots, \alpha_K]$. D is a product of the integral over the bond lengths and their Jacobian J . The Jacobian $[\Pi; \sin(\theta_j)]$ of the bond angles, θ_j ; that should appear under the integral is omitted for simplicity (However, in paper I we have shown that the Jacobian cancels out in entropy and free energy differences; therefore, we shall discard the Jacobian from future discussions.) We assume D 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 Boltzmann probability density corresponding to Z (eq 6) is (E is defined in eq 5)

$$\rho^B([\alpha_k], \mathbf{x}^N) = \exp\{-E([\alpha_k], \mathbf{x}^N) / k_B T\} / Z, \quad (7)$$

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

$$S = -k_B \int_m \rho^B([\alpha_k], \mathbf{x}^N) \ln \rho^B([\alpha_k], \mathbf{x}^N) d\alpha_1 \dots d\alpha_K d\mathbf{x}^N \quad (8)$$

and

$$F = \int_m \rho^B([\alpha_k], \mathbf{x}^N) \{E([\alpha_k], \mathbf{x}^N) + k_B T \ln \rho^B([\alpha_k], \mathbf{x}^N)\} d\alpha_1 \dots d\alpha_k d\mathbf{x}^N \quad (9)$$

It should be pointed out that the fluctuation of the *exact* F is zero;⁵⁹ thus (provided that the above assumptions about the bond lengths are correct) one can substitute the expression for $\rho_B([\alpha_k], \mathbf{x}^N)$ (eq 7) inside the curly brackets of eq 9 to obtain,

$$E([\alpha_k], \mathbf{x}^N) + k_B T \ln \rho^B([\alpha_k], \mathbf{x}^N) = -k_B T \ln Z = F, \quad (10)$$

i.e. the expression in the curly brackets is constant and equal to F for any set $([\alpha_k], \mathbf{x}^N)$ 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.^{9,40,59-61} Because HSMC(D) provides an approximation for $\rho^B([\alpha_k], \mathbf{x}^N)$, it enables one, in principle, to estimate the free energy of the system from *any single* structure [Notice, however, that calculation of $\rho^B([\alpha_k], \mathbf{x}^N)$ for a single conformation depends on the entire microstate as is also evident from the HSMC(D) procedure discussed later].

With MD the bond stretching energy is taken into account in eq 9 (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 paper I, II.6 that it can be estimated approximately within the framework of HSMD.

II.3. Exact future scanning procedure

HSMC(D) (as well as HS and LS) is based on the ideas of the *exact* scanning method where a system is constructed (from nothing) step-by-step using transition probabilities (TPs). The product of these TPs is equal to the Boltzmann probability (eqs 1 and 7) 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 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 α_k , $1 \leq \alpha_k \leq 6M=K$) in microstate m ; the loop is surrounded by N water molecules moving within the volume defined by the sphere of radius R_{cap} , 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 transition probabilities (TPs) and adding the related atoms;⁶² for example, the angle ϕ determines the coordinates of the two hydrogens connected to C^α , while the bond angle $N-C^\alpha-C'$ determines the position of C' . Thus, at step k , $k-1$ angles $\alpha_1, \dots, \alpha_{k-1}$ 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 $\rho(\alpha_k | \alpha_{k-1} \dots \alpha_1)$,

$$\rho(\alpha_k | \alpha_{k-1}, \dots, \alpha_1) = Z_{\text{future}}(\alpha_k, \dots, \alpha_1) / [Z_{\text{future}}(\alpha_{k-1}, \dots, \alpha_1)] \quad (11)$$

where $Z_{\text{future}}(\alpha_k, \dots, \alpha_1)$ is a future partition function. The term “future” indicates that the integration defining Z_{future} is carried out over the variables $\alpha_{k+1}, \dots, \alpha_K$ and the $3N$ coordinates \mathbf{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 $\alpha_1 \dots \alpha_k$), while $\alpha_{k+1}, \dots, \alpha_K$ are varied in a restrictive way where the corresponding conformations of the “future” part of the loop remain in microstate m . Thus

$$Z_{\text{future}}(\alpha_k, \dots, \alpha_1) = \int_m \exp[-(E(\alpha_k, \dots, \alpha_1, \mathbf{x}^N)/k_B T)] d\alpha_{k+1} \dots d\alpha_K d\mathbf{x}^N \quad (12)$$

where E (eq 5) is the total potential energy of the loop/template/water system, which also imposes the loop closure condition. The product of the TPs (eq 11) leads to the (Boltzmann) probability density of the entire loop conformation,

$$\rho_{\text{loop}}^B(\alpha_K, \dots, \alpha_1) = \prod_{k=1}^K \rho(\alpha_k | \alpha_{k-1}, \dots, \alpha_1). \quad (13)$$

After the loop structure has been constructed a configuration of water molecules is generated step-by-step, where the TP density for placing water molecule k at \mathbf{x}_k is

$$\rho_{\text{water}}(\mathbf{x}_k | \alpha_K, \dots, \alpha_1, \mathbf{x}^{k-1}) = Z_{\text{future}}(\alpha_K, \dots, \alpha_1, \mathbf{x}^k) / [Z_{\text{future}}(\alpha_K, \dots, \alpha_1, \mathbf{x}^{k-1})] \quad (14)$$

where the loop conformation is kept constant and the $k-1$ water molecules that have already been treated are fixed at their coordinates, \mathbf{x}^{k-1} and the summation in $Z(\mathbf{x}^k)$ is over the as yet undecided $N-k+1$ water molecules. (Notice that \mathbf{x}_k denotes the 3 Cartesian coordinates of water molecule k , while \mathbf{x}^k denotes the set of Cartesian coordinates of the k molecules $1, 2, \dots, k$). The Boltzmann probability density of the water is

$$\rho_{\text{water}}^B(\alpha_K, \dots, \alpha_1, \mathbf{x}^N) = \prod_{k=1}^N \rho_{\text{water}}(\mathbf{x}_k | \alpha_K, \dots, \alpha_1, \mathbf{x}^{k-1}) \quad (15)$$

and the probability density of the loop/water configuration is,

$$\rho^B([\alpha_k], \mathbf{x}^N) = \rho_{\text{loop}}^B([\alpha_k]) \rho_{\text{water}}^B([\alpha_k], \mathbf{x}^N) = \exp\{-E([\alpha_k], \mathbf{x}^N)/k_B T\} / Z_{\text{future}}(\alpha_1) \quad (16)$$

where $Z_{\text{future}}(\alpha_1) (= Z)$ is the partition function of the entire loop/water system for microstate m . Because $\rho_B([\alpha_k], \mathbf{x}^N)$ is known one can obtain the free energy from any single loop/water configuration (see eq 10). In addition to S (eq 8), one can define for m “the loop entropy of mean force”, S_{loop} ,

$$S_{\text{loop}} = -k_B \int_m \rho^B([\alpha_k]) \ln \rho^B([\alpha_k]) d[\alpha_k] \quad (17)$$

where $d[\alpha_k] \equiv d\alpha_1 \cdots d\alpha_K$; 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 Metropolis MC and MD) in the sense that 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.4. The HSMC(D) method

The theory of HSMD is again described as applied loop consisting of Gly residues. One starts by generating an MD sample of microstate m with water molecules; the conformations are then represented in terms of the dihedral and bond angles α_k , $1 \leq k \leq 6N=K$, and the variability range $\Delta\alpha_k$ is calculated,

$$\Delta\alpha_k = \alpha_k(\text{max}) - \alpha_k(\text{min}), \quad (18)$$

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 spans correctly the microstate m .

System configuration ($[\alpha_k]$, \mathbf{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 $\alpha_{k-1} \cdots \alpha_1$ have already been reconstructed and the TP density of α_k , $\rho(\alpha_k | \alpha_{k-1}, \dots, \alpha_1)$ is calculated from an MD sample of n_f conformations (generated in Cartesian coordinates), where the *entire* future of the loop and water is moved [i.e., the loop atoms defined by $\alpha_k, \dots, \alpha_K$ and the water coordinates (\mathbf{x}^N)] while the past (the loop atoms defined by $\alpha_1, \dots, \alpha_{k-1}$) 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,

$$\rho_{\text{loop}}(\alpha_k | \alpha_{k-1}, \dots, \alpha_1) \approx \rho^{\text{HS}}(\alpha_k | \alpha_{k-1}, \dots, \alpha_1) = n_{\text{visit}} / [n_f \delta\alpha_k] \quad (19)$$

where $\rho^{\text{HS}}(\alpha_k | \alpha_{k-1}, \dots, \alpha_1)$ becomes exact for very large n_f ($n_f \rightarrow \infty$) and a very small bin ($\delta\alpha \rightarrow 0$). This means that in practice $\rho^{\text{HS}}(\alpha_k | \alpha_{k-1}, \dots, \alpha_1)$ 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 consisting 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 located within the limits of $\delta\alpha_k$ and $\delta\alpha_{k+1}$, respectively; therefore eq 19 becomes

$$\rho^{\text{HS}}(\alpha_{k+1}, \alpha_k | \alpha_{k-1}, \dots, \alpha_1) = n_{\text{visit}} / [n_f \delta\alpha_k \delta\alpha_{k+1}], \quad (20)$$

where in paper I 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

$$\rho^{\text{HS}}(\alpha_K, \dots, \alpha_1) = \prod_{k=1}^K \rho^{\text{HS}}(\alpha_{k+1}; \alpha_k | \alpha_{k-1}, \dots, \alpha_1), \quad (21)$$

where in the product only odd values of k are used. $\rho^{\text{HS}}([\alpha_k])$ defines an approximate entropy functional, denoted $S_{\text{loop}}^{\text{A}}$ which can be shown using Jensen's inequality to constitute a *rigorous* upper bound for S_{loop} (eq 17),¹⁰

$$S_{\text{loop}}^{\text{A}} = -k_B \int_m \rho^{\text{B}}([\alpha_k]) \ln \rho^{\text{HS}}([\alpha_k]) d[\alpha_k]. \quad (22)$$

$\rho_{\text{loop}}^{\text{B}}$ (eq 13) is the Boltzmann probability density of $[\alpha_k]$ in m . Thus, for microstate m , $S_{\text{loop}}^{\text{A}}$ can be estimated from a Boltzmann sample (of size n_s) generated by MD using the arithmetic average,

$$\bar{S}_{\text{loop}}^{\text{A}}(m) = -\frac{k_B}{n_s} \sum_{t=1}^{n_s} \ln \rho^{\text{HS}}(t, m) \quad (23)$$

where $\rho^{\text{HS}}(t, m)$ is the value of $\rho^{\text{HS}}([\alpha_k])$ obtained for configuration t of the sample of m . $S_{\text{loop}}^{\text{A}}$ constitutes a measure of the loop flexibility of a pure geometrical character, i.e. with no *direct* dependence on the interaction energy. We denote the difference in the loop entropies obtained for a specific set of parameters by $\Delta S_{\text{loop}}^{\text{A}}$ but the converged difference, which is expected to be exact within the statistical errors, is denoted by ΔS_{loop}

$$\Delta S_{\text{loop}} = \Delta S_{\text{loop}}^{\text{A}} = \bar{S}_{\text{loop}}^{\text{A}}(m) - \bar{S}_{\text{loop}}^{\text{A}}(n) \quad (24)$$

In the same way one calculates the arithmetic averages of the energies over the n_s system configurations

$$\bar{E}_{\text{loop}}(m) = \frac{1}{n_s} \sum_{t=1}^{n_s} E_{\text{loop}}(t, m) \quad (25)$$

where $E_{\text{loop}}(t, m)$ is the loop-loop and loop-template interaction energy of loop conformation t . The corresponding difference is,

$$\Delta E_{\text{loop}} = \bar{E}_{\text{loop}}(m) - \bar{E}_{\text{loop}}(n) \quad (26)$$

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

$$\Delta F_{\text{loop}} = \Delta E_{\text{loop}} - T \Delta S_{\text{loop}} \quad (27)$$

To reconstruct the water configuration one can use the HSMC(D) procedure for fluids developed previously, which would lead to $\rho_{\text{water}}^{\text{HS}}([\alpha_k], \mathbf{x}^N)$ [as an approximation for $\rho_{\text{water}}^{\text{B}}([\alpha_k], \mathbf{x}^N)$ (eq 15)] and then to the contribution of the water configuration to the free energy $F_{\text{water}}([\alpha_k], \mathbf{x}^N) = E_{\text{water}}([\alpha_k], \mathbf{x}^N) + k_B T \ln \rho_{\text{water}}^{\text{HS}}([\alpha_k], \mathbf{x}^N)$. However, this procedure for fluids has not been optimized yet and it is relatively time consuming. Alternatively, one can obtain $F_{\text{water}}([\alpha_k], \mathbf{x}^N)$ by a thermodynamic integration (TI) procedure, where the water molecules are integrated from an ideal gas to their TIP3P form within the spherical volume (R_{cap}) and the presence of constant template and loop structure; however, this would be a complex procedure as well. Since the free and bound loop structures have the same template and because we are mainly interested in free energy differences, we have applied a much simpler TI procedure based on the same reference state for the two microstates. 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 $F_{\text{water}}^{\text{TI}}([\alpha_k], m)$ which is then averaged over the n_s sample configurations (see eq 23),

$$\bar{F}_{\text{water}}^{\text{TI}}(m) = \frac{1}{n_s} \sum_{t=1}^{n_s} F_{\text{water}}^{\text{TI}}(t, m) \quad (28)$$

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

$$\Delta F_{\text{water}} = \bar{F}_{\text{water}}^{\text{TI}}(m) - \bar{F}_{\text{water}}^{\text{TI}}(n) \quad (29)$$

In the same way one calculates the arithmetic averages of the water energy over the n_s system configurations

$$\bar{E}_{\text{water}}(m) = \frac{1}{n_s} \sum_{t=1}^{n_s} E_{\text{water}}(t, m) \quad (30)$$

where $E_{\text{water}}(t, m)$ is the water-water, water-template, and water-loop interaction energy of system configuration t . The corresponding differences are,

$$\Delta E_{\text{water}} = \bar{E}_{\text{water}}(m) - \bar{E}_{\text{water}}(n) \quad (31)$$

and

$$\Delta E_{\text{total}} = \Delta E_{\text{water}} + \Delta E_{\text{loop}} \quad (32)$$

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

$$\Delta F_{\text{total}} = \Delta E_{\text{loop}} - T\Delta S_{\text{loop}} + \Delta F_{\text{water}} \quad (33)$$

The difference in the water entropy between m and n is

$$T\Delta S_{\text{water}} = T[\bar{S}_{\text{water}}^{\text{TI}}(m) - \bar{S}_{\text{water}}^{\text{TI}}(n)] = \Delta E_{\text{water}} - \Delta F_{\text{water}}, \quad (34)$$

where the corresponding difference in the total entropy is

$$T\Delta S_{\text{total}} = T\Delta S_{\text{water}} + T\Delta S_{\text{loop}}. \quad (35)$$

It should be pointed out again that the dependence of ΔF_{total} (eq 33) (and $T\Delta S_{\text{total}}$, eq 35) on the bond stretching energy is through E_{loop} while this interaction is ignored in $S_{\text{loop}}^{\text{A}}$ (eqs 22 and 23). However, under the assumptions leading to eq 6 this is not expected to affect differences in free energy which are our main interest; see also paper I (II.6).

II.5. The reconstruction procedure with HSMD

The HSMD reconstruction procedure needs further discussions. Thus, the MD simulation of the future chain at step k starts from the reconstructed conformation i , and every g fs the current conformation is considered, where the n_{init} initial considered conformations are discarded for equilibration. The next n_f (considered) future conformations are represented in internal coordinates and their contribution to n_{visit} (eq 20) 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 “overflow” is more likely to happen for small residues such as Gly and for small k .) To be able to control the extent of coverage of m the following procedure has been applied: n_f has been divided into several (j) shorter repetitive procedures (“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 equilibration of size, n_{init} ; obviously, one would seek to determine the minimal values for n'_f , j , and n_{init} , which would keep the future chains within m while allowing its adequate sampling. A similar procedure was first suggested by Brady & Karplus⁶³ within the framework of the quasi-harmonic method, and was also used in implementations of the LS method to peptides.^{26,41}

In paper I (II.6) we have discussed (and applied) several measures which enable one to estimate the extent of coverage of the reconstructed samples of the future chains. Because the present $\Delta\alpha_k$ values are smaller than those in paper I (see Table 1) and we use the same n_f values, we did not consider it necessary to apply these measures here. However, it should be emphasized again that we are interested in an entropy difference, $\Delta S_{m,n}^A$ between two microstates, where $\Delta S_{m,n}^A$ is considered to be reliable (i.e., to lead to ΔS_{loop} , eq 24) if its results are found to be stable for a large range of the parameters n_{init} , n'_f , and j . From now on we shall replace in most cases n'_f by the word unit.

II.6. The Local States (LS) and the quasi-harmonic (QH) methods

With the LS method⁴⁻⁶ (applied to an N -residue polyglycine with $6N=K$ backbone angles, α_k) the ranges $\Delta\alpha_k$ (eq 18) are divided into l equal segments, where l is the discretization parameter. These segments are denoted by v_k , ($v_k=1, l$), where an angle α_k is represented by the segment v_k to which it belongs, and a conformation i is expressed by the corresponding vector of segments $[v_1(i), v_2(i), \dots, v_K(i)]$. The TP $\rho(\alpha_k|\alpha_{k-1}\dots\alpha_1)$ can be estimated only approximately by $n(v_k, \dots, v_{k-b})/\{n(v_{k-1}, \dots, v_{k-b})[\Delta\alpha_k/l]\}$, where $n(v_k, \dots, v_{k-b})$ is the number of times the local state [i.e., the vector (v_k, \dots, v_{k-b})] appears in the sample; b is the correlation parameter.

One obtains the approximate probability density, $\rho_i(b, l) = \prod_{k=1}^K p(v_k|v_{k-1}, \dots, v_{k-b})/(\Delta\alpha_k/l)$, the larger are b and l the better the approximation (for enough statistics). $\rho_i(b, l)$ defines a rigorous upper bound, S_{loop}^A (eq 22) where $\rho_i(b, l)$ replaces ρ^{HS} ; S_{loop}^A can be estimated by eq 23.

With the QH method introduced by Karplus and Kushick,³⁹ the Boltzmann probability density of structures defining a microstate is approximated by a multivariate Gaussian. Thus,

$$S_{loop}^{QH}(m) = (k_B/2) \{N + \ln[(2\pi)^N \text{Det}(\sigma)]\} \quad (36)$$

where the covariance matrix, σ , 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 during the years, a systematic study of its performance has been carried out only recently by Gilson's group⁶⁴ who have found that the performance of QH deteriorates significantly in Cartesian coordinates and when applied to more than one microstate.³⁵

III. Results and discussion

III.1. Simulation details

We carried out two MD simulations at $T=300$ K starting from the free and bound PDB structures (which were capped by 70 TIP3P waters); by considering a structure every 0.5 ps these simulations led to stable samples of $n_s=600$ conformations. These simulations and the reconstruction simulations (for generating the future samples) were carried out with the velocity-Verlet algorithm⁶⁵ 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;⁶⁵ the Berendsen⁶⁵ 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 four bin sizes, $\delta = \Delta\alpha_k/45$, $\Delta\alpha_k/30$, $\Delta\alpha_k/15$, and $\Delta\alpha_k/10$, centered at α_k (i.e., $\alpha_k \pm \delta/2$) (eqs 19 and 20). If the counts of the smallest bin are smaller than 50 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, zero counts is a very rare event.

To obtain the loop entropy, $\overline{S}_{\text{loop}}^A(m)$ (eq 23) the calculations are based on unit $n'_f=250$ (2.5 ps) and future sample sizes $n_f=250$ ($j=1$), 500 ($j=2$), 750 ($j=3$), and 1250 ($j=5$). To examine the convergence of the results for ΔS_{loop}^A (eq 24) we have extracted from the main samples (of the free and bound microstates) two sets (pairs) of partial samples. One set, which consists of $n_s=100$, was obtained by selecting every 6th conformation of the main samples; the corresponding conformations were reconstructed (as above) with $n_f=250, 500, 750$, and 1250. The second partial set (which was selected in a similar way) consists of $n_s=40$ but its conformations were reconstructed more extensively again with unit $n'_f=250$ (2.5 ps) but $n_f=1250$ ($j=5$), 2500 ($j=10$), 5000 ($j=20$), and 10^4 ($j=40$).

III.2. Results for the loop entropy

In Table 1 we present the values of $\Delta\alpha_k$ (eq 18) for the free and bound microstates obtained from the corresponding MD samples (of size 600). These values suggest that the two samples indeed are concentrated in conformational space, and the corresponding values for the χ angles of the two microstates are comparable. For comparison we also provide the corresponding values obtained in paper I (Table 6) for the same loop in the GB/SA implicit solvent. The present results, in most cases, are larger than those obtained in paper I for the loop in vacuum (not shown) but are smaller (and in some cases considerably smaller) than results obtained with implicit solvent as Table 1 reveals. Larger $\Delta\alpha_k$ values are expected to correlate with higher entropy and indeed the results for $T\overline{S}_{\text{loop}}^A(m)$ in Table 2 are smaller than those of paper I obtained with implicit solvent, as discussed below.

Table 2 contains results for the entropy, $T\overline{S}_{\text{loop}}^A(m)$ (eq 23) for the free and bound microstates. The statistical errors were obtained from the fluctuations and results obtained for partial samples. For comparison we have added in Table 2 results for $n_f=1250$ (appear with an underscore) obtained in paper I for a loop in implicit solvent. These results are larger (by ~ 4.2 and ~ 3.2 kcal/mol for the free end bound microstates) than the results for $n_f=1250$ obtained with explicit water, which is in accord with $\Delta\alpha_k(\text{implicit})$ being larger in most cases than $\Delta\alpha_k(\text{explicit})$ in Table 1.

Being an upper bound, one would expect $\overline{S}_{\text{loop}}^A(m)$ to decrease with decreasing bin size and increasing n_f – an expectation which is fully satisfied. Also, results for the same n for $\Delta\alpha_k/30$ and $\Delta\alpha_k/45$ are almost converged; thus,

$T\overline{S}_{\text{loop}}^A(\Delta\alpha_k/30, n_f=1250) - T\overline{S}_{\text{loop}}^A(\Delta\alpha_k/45, n_f=1250) = 0.13$ kcal/mol for both microstates, which is close to the statistical errors. On the other hand, results of the same bin (obtained for different n_f values) are not converged. The computer time required to reconstruct a loop structure capped with 70 water molecules using $n_f=500$ is ~ 1.6 h CPU on a 2.1 GHz Athlon processor, which is smaller than 3.6 h CPU required for reconstructing a structure in implicit solvent in paper I

The HSMD results for the entropy are compared in the table to those obtained with the LS and QH methods from larger MD samples of 25,000 loop-water configurations. These samples were obtained from 2.5 ns trajectories where a configuration is retained every 0.1 ps (where the 20 ps initial trajectory is discarded for equilibration). The central values of $T\overline{S}_{\text{loop}}^{\text{QH}}$ (eq 36)

exceed the HSMD results (for $n_s=600$) by ~ 14 and ~ 20 kcal/mol for the free and bound microstates, respectively, while the corresponding LS results (eq 23, using $b=1$, $l=10$, see II. 6) are even higher, exceeding the HSMD results by ~ 22 and ~ 24 kcal/mol. These elevated results are in accord with both $S_{\text{loop}}^{\text{QH}}$ and $S_{\text{loop}}^{\text{LS}}$ being upper bounds; however, they might also be affected by the longer MD trajectories generated for QH and LS than for HSMD, as discussed in the I.4. $S^{\text{LS}} > S^{\text{QH}}$ was also found in previous studies.^{13–15}

III.3. Differences in loop entropy

In paper I and ref ¹⁶ converging results for $T\Delta S^{\text{A}}$ were obtained already for unit 2.5 ps and $n_f \geq 500$ (and even for smaller n_f values using optimized bins). Because the $\Delta\alpha_k$ values for a loop in explicit water are smaller than those obtained with implicit solvent (Table 1) we have applied unit =2.5 ps and $n_f=250-1250$ also in the present study and to examine the convergence of the results for of $\Delta S_{\text{loop}}^{\text{A}}$ for smaller samples, we present in Table 3 also results for $n_s=200$, 100, and 40 as discussed in III.1. The results in the table are given for the two smallest bins of $\Delta\alpha_k/30$ and $\Delta\alpha_k/45$.

The statistical errors of the results for $T\Delta S_{\text{loop}}^{\text{A}}$ ($n_s=600$) are not larger than ± 0.06 kcal/mol and to simplify the comparison are not presented in the table. All the results for $n_f \geq 500$ converge to -1.3 kcal/mol within ± 0.1 kcal/mol, and even those for $n_f=250$ (-1.6 kcal/mol) deviate only by $0.3 - 1.3$ kcal/mol; furthermore, the results for $\Delta\alpha_k/15$, which are not provided in Table 3 are very good, -1.4 and -1.5 kcal/mol. This extent of convergence is comparable to that obtained in previous HSMD studies.^{15,16} Notice that the results for the first 200 conformations of the samples are equal to those based on $n_s=600$ (while providing a factor of 3 reduction in computer time). On the other hand, the results obtained for $n_s=100$ and 40 are higher by $0.3-0.4$ kcal/mol above the $n_s=600$ values. Still the merit of such calculations (that might provide further reduction in computer time) depends on the errors caused by the other components, i.e., the total energy and the water contribution to the entropy. It should be pointed out that for a loop in implicit solvent (paper I) $T\Delta S_{\text{loop}}^{\text{A}} = +0.3 \pm 0.1$, while in explicit water the loop entropy of the free microstate is lower than that of the bound microstate.

This convergence of entropy differences stems from the cancellation (in $T\Delta S_{\text{loop}}^{\text{A}}$) of approximately equal systematic errors in $S_{\text{loop}}^{\text{A}}$ (free) and $S_{\text{loop}}^{\text{A}}$ (bound) as discussed in detail in section II.10 of paper I. Thus, Table 2 shows that the worst approximations that still lead to a good $T\Delta S_{\text{loop}}^{\text{A}}$ value differ from the best ones by

$S_{\text{loop}}^{\text{A}}(m)(\Delta\alpha_k/15, n_f=250) - TS_{\text{loop}}^{\text{A}}(m)(\Delta\alpha_k/45, n_f=1250) = 2.1$ and 2.4 kcal/mole for the free and bound microstates, respectively; these differences constitute lower bounds because the correct TS_{loop} values might be significantly smaller than $TS_{\text{loop}}^{\text{A}}(m)(\Delta\alpha_k/45, n_f=1250)$. The large errors in the results for LS and QH do not allow calculating meaningful differences.

III.4. Thermodynamic integration of water

As described earlier, in the TI process the interaction energy [electrostatic and Lennard Jones (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.,⁶⁶

$$\phi(r_{ij}, \lambda) = \lambda 4\epsilon \left[\frac{\sigma^{12}}{(r_{ij}^2 + \delta(1 - \lambda))^6} - \frac{\sigma^6}{(r_{ij}^2 + \delta(1 - \lambda))^3} \right], \quad (37)$$

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

$$\frac{\partial F}{\partial \lambda} = \left\langle \frac{\partial E(x^N, \lambda)}{\partial \lambda} \right\rangle_{\lambda}, \quad (38)$$

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 16 equal integration bins $\Delta \lambda_i$. The ($\lambda=1 \rightarrow \lambda=0$) integration of the electrostatic interactions 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 32 $\partial F / \partial \lambda_i$ integration steps.

The MD simulation consists of a 2 fs integration step, where every 20 fs the current water configuration is added to the sample. For each ($\Delta \lambda_i$) step the initial simulation (5 ps) is used for equilibration and is thus discarded; the following 20 ps (1000 configurations) are used for evaluating $\langle \partial F / \partial \lambda_i \rangle$. Notice that in spite of the advanced scaling function (eq 37), in the last steps of the LJ integration (i.e., λ close to zero) the results always increased strongly; therefore, we have adopted the results integrated up to $\lambda=0.25$. For a single loop structure this free energy integration requires ~ 5 h CPU on a 2.1 GHz Athlon processor. As shown in Table 4, the free energy integrations over the two samples have given $\overline{F}_{\text{water}}^{\text{TI}}(m) \sim 31.2$ and 35.5 kcal/mol. (eq 28) for the free and bound microstates, respectively, thus leading to $\Delta F_{\text{water}} = -4.4 \pm 0.7$ kcal/mol (eq 29), i.e., the water provides higher stability to the free microstate. To check whether computer time can further be decreased, we also calculated $\overline{F}_{\text{water}}^{\text{TI}}(m)$ by considering the contribution of only 100 configurations of each sample (i.e., every 6th configuration was considered) obtaining $\Delta F_{\text{water}} = -5.0$ kcal/mol, which is within the error bars of the result based on the entire samples, while computer time is reduced by a factor of 6.

It should be pointed out that unlike an *NVT* system of pure water under periodic boundary conditions, the water system here is not homogeneous. Computer graphics has shown that the water molecules cover the loop and parts of the template while the outer region of the spherical volume is predominately vacant. Indeed, a crude calculation shows that the empty volume is $\sim 5000 \text{ \AA}^3$ which would require ~ 160 water molecules to obtain the experimental density of water. Thus, crevices in the template might remain empty or become occupied by waters for long simulation time, which leads to increased fluctuations of thermodynamic parameters. Therefore, a systematic improvement in the integration parameters (i.e., using up to 64 $\Delta \lambda_i$ steps and longer simulations for each TI step) has not decreased these fluctuations. We expect this picture to improve as the number of waters, N increases, and the effect of various minimalist models studied. We have already carried out preliminary simulations with $N=120$, but could not reach definite conclusions due to the dependence of the results on the parameters r_{shift} and R_{cap} . This problem will be studied systematically in the future by considering a larger template, where using r_{shift} might not be needed, because the percolation of water will be avoided.

In Table 4 we provide results for the free and bound microstates obtained for $\bar{E}_{\text{loop}}(m)$ (eq 25), $\bar{E}_{\text{water}}(m)$ (eq 30), their sum, $\bar{E}_{\text{total}}(m)$ and $\bar{F}_{\text{water}}^{\text{TI}}(m)$ (eq 28); we also provide the corresponding differences (free-bound), where the errors were estimated from results obtained for partial samples. It is of interest to point out that the results for $\bar{E}_{\text{loop}}(m) \sim -108$ and -106 for the free end bound microstates are in the same range of ~ -137 and -99 kcal/mol, respectively obtained in paper I for a loop in vacuum; however, the difference here $\Delta E_{\text{loop}} = -1.7$ is much smaller than in paper I (~ -38 kcal/mol) due to the effect of water. This low ΔE_{loop} (eq 26) together with a small entropy contribution, $T\Delta S_{\text{loop}} = -1.3$ leads to $\Delta F_{\text{loop}} = -0.4 \pm 1$ kcal/mol (eq 27), i.e., the loop contributes very little to the higher stability of the free microstate (see Table 5 where all the various differences are summarized).

Thus, the relatively high difference, $\Delta E_{\text{total}} = -7.2$ kcal/mol is contributed mainly by water, $\Delta E_{\text{water}} = -5.4$ kcal/mol (eq 31, and Tables 4 and 5). This energy difference for water is slightly counterbalanced by $T\Delta S_{\text{water}} = -1.0$ (Table 5) leading to $\Delta F_{\text{water}} = -4.4$ kcal/mol. A similar effect is demonstrated (Table 5) for the total energy values (due to the small contributions of the loop discussed above), i.e., $\Delta E_{\text{total}} = -7.2$, and $T\Delta S_{\text{total}} = -2.3$ where $\Delta F_{\text{total}} = -4.8$ kcal/mol.

Our results also show that $\Delta E_{\text{total}} = -7.2$ correctly predicts the higher stability of the free microstate, and this value is not significantly different from $\Delta F_{\text{total}} = -4.8$ kcal/mol -the correct measure of stability due to a relatively small entropy effect, $T\Delta S_{\text{total}} = -2.3$ kcal/mol. Notice that the free microstate was found to be the more stable microstate also in implicit water (paper I); however, the free energy difference there is significantly larger than here, $\Delta F_{\text{implicit}} = -25.5$ kcal/mol. This higher stability is expected because the bound loop structure was superimposed on the template of the free protein. However, one should bear in mind that this result is mainly due to water and it thus depends strongly on the model of water used, which is presently based on $r_{\text{shift}} = 5 \text{ \AA}$. Also, it is not clear how the number of water molecules and their density would affect the relative stability of the two microstates.

IV. Summary and conclusions

In paper I HSMD has been extended to a protein loop by treating the short loop (207–209) of pancreatic α -amylase modeled in vacuum and in the implicit solvent GB/SA. In the present paper we have made an important step further by extending HSMD to the same loop solvated by explicit water; treating the same loop in the free and bound microstates enables one to compare the effects of the different models. Computation of the entropy and free energy is divided into two stages, where the loop's entropy is calculated first by reconstructing its structures in the presence of moving waters; this is followed by the calculation of the free energy of the surrounding water in the presence of a fixed loop structure. As in previous studies, we have found that already small reconstruction samples of 500 structures lead to the correct entropy difference, $T\Delta S_{\text{loop}}$. Furthermore, the same difference was obtained from a partial sample of 200 configurations (rather than 600), which decreases computer time by a factor of 3, where other means to improve efficiency (discussed in the Summary of paper I but were not applied here) are expected to decrease computer time considerably further. The fast convergence of the results for $T\Delta S_{\text{loop}}$ supports (like previous calculations) theoretical arguments discussed in paper I that relatively large systematic errors in $S_{\text{loop}}^{\text{A}}(m)$ are cancelled to a large extent in differences, $\Delta S_{\text{loop}}^{\text{A}}$ (eq 24). The relatively small statistical errors stem from the small system (loop and water) moved by MD. 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 the TI of water can be fully parallelized.

Calculation of the free energy of water (second stage) was carried out successfully (and with relatively small error bars) by TI rather than HSMD, where again the difference in free energy ΔF_{water} obtained from a partial sample of 100 configurations is equal within the statistical error to that obtained from the entire sample. Thus, the application of our entire methodology to a loop capped with water has been successful. However, the performance of the water model used with respect to other models has not been investigated; in particular, the effect of the number of waters capped, their volume, density, and the size of the shifting parameter on ΔF_{total} and other thermodynamic parameters should be studied. Such a study is being carried out now with respect to the 4-residue mobile loop, 287–290 of the protein acetylcholinesterase (AChE) from *Torpedo californica*. Treating this loop has the advantage that ΔF_{total} has been estimated from experimental data and was calculated by several techniques.⁶⁷

In accordance with paper I, the quasi-harmonic approximation and the local states method 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.

Acknowledgements

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

References

1. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. J Chem Phys 1953;21:1087.
2. Alder BJ, Wainwright TE. J Chem Phys 1959;31:459.
3. McCammon JA, Gelin BR, Karplus M. Nature 1977;267:585. [PubMed: 301613]
4. Meirovitch H. Chem Phys Lett 1977;45:389.
5. Meirovitch H, Vásquez M, Scheraga HA. Biopolymers 1987;26:651. [PubMed: 3593889]
6. Meirovitch H, Koerber SC, Rivier J, Hagler AT. Biopolymers 1994;34:815. [PubMed: 8054467]
7. Meirovitch H. Phys Rev A 1985;32:3709. [PubMed: 9896540]
8. Meirovitch H, Scheraga HA. J Chem Phys 1986;84:6369.
9. Meirovitch H. J Chem Phys 2001;114:3859.
10. White RP, Meirovitch H. J Chem Phys 2004;121:10889. [PubMed: 15634040]
11. White RP, Meirovitch H. J Chem Phys 2006;124:204108. [PubMed: 16774320]
12. White RP, Meirovitch H. J Chem Phys 2005;123:214908. [PubMed: 16356071]
13. Cheluvvaraja S, Meirovitch H. J Chem Phys 2005;122:054903.
14. Cheluvvaraja S, Meirovitch H. J Phys Chem B 2005;109:21963. [PubMed: 16853854]
15. Cheluvvaraja S, Meirovitch H. J Chem Phys 2006;125:024905.
16. Cheluvvaraja S, Meirovitch H. J Chem Theory Comput 2008;4:192.
17. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. J Am Chem Soc 1995;117:5179.
18. Qiu D, Shenkin PS, Hollinger FP, Still WC. J Phys Chem 1997;101:3005.
19. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. J Chem Phys 1983;79:926.
20. Elber R, Karplus M. Science 1987;235:318. [PubMed: 3798113]
21. Stillinger FH, Weber TA. Science 1984;225:983. [PubMed: 17783020]
22. Getzoff ED, Geysen HM, Rodda SJ, Alexander H, Tainer JA, Lerner RA. Science 1987;235:1191. [PubMed: 3823879]
23. Rini JM, Schulze-Gahmen U, Wilson IA. Science 1992;255:959. [PubMed: 1546293]
24. Constantine KL, Friedrichs MS, Wittekind M, Jamil H, Chu CH, Parker RA, Goldfarb V, Mueller L, Farmer BT. Biochemistry 1998;37:7965. [PubMed: 9609689]
25. Kessler H, Matter H, Gemmecker G, Kottenhahn M, Bates JW. J Am Chem Soc 1992;114:4805.

26. Baysal C, Meirovitch H. *Biopolymers* 1999;50:329. [PubMed: 10397793]
27. Korzhnev DM, Salvatella X, Vendruscolo M, Di Nardo AA, Davidson AR, Dobson CM, Kay LE. *Nature* 2004;430:586. [PubMed: 15282609]
28. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE, Kern D. *Nature* 2005;438:117. [PubMed: 16267559]
29. Beveridge DL, DiCapua FM. *Annu Rev Biophys Biophys Chem* 1989;18:431. [PubMed: 2660832]
30. Kollman PA. *Chem Rev* 1993;93:2395.
31. Jorgensen WL. *Acc Chem Res* 1989;22:184.
32. Meirovitch, H. *Reviews in Computational Chemistry*. Lipkowitz, KB.; Boyd, DB., editors. Wiley-VCH; New York: 1998. p. 12p. 1
33. Gilson MK, Given JA, Bush BL, McCammon JA. *Biophys J* 1997;72:1047. [PubMed: 9138555]
34. Boreesch S, Tettinger F, Leitgeb M, Karplus M. *J Phys Chem B* 2003;107:9535.
35. Meirovitch H. *Curr Opin Struct Biol* 2007;17:181. [PubMed: 17395451]
36. Gô N, Scheraga HA. *J Chem Phys* 1969;51:4751.
37. Gô N, Scheraga HA. *Macromolecules* 1976;9:535.
38. Hagler AT, Stern PS, Sharon R, Becker JM, Naider F. *J Am Chem Soc* 1979;101:6842.
39. Karplus M, Kushick JN. *Macromolecules* 1981;14:325.
40. White RP, Meirovitch H. *J Chem Phys* 2003;119:12096.
41. Meirovitch H, Meirovitch E. *J Phys Chem* 1996;100:5123.
42. Meirovitch H, Hendrickson TF. *Proteins* 1997;29:127. [PubMed: 9329078]
43. Baysal C, Meirovitch H. *Biopolymers* 2000;53:423. [PubMed: 10738203]
44. Qian M, Haser R, Payan F. *J Mol Biol* 1993;231:785. [PubMed: 8515451]
45. Qian M, Haser R, Buisson G, Duee E, Payan F. *Biochemistry* 1994;33:6284. [PubMed: 8193143]
46. Qian M, Haser R, Payan F. *Protein Sci* 1995;4:747. [PubMed: 7613472]
47. Machius M, Vertesy L, Huber R, Wiegand G. *J Mol Biol* 1996;260:409. [PubMed: 8757803]
48. Brayer GD, Sidhu G, Maurus R, Rydberg EH, Braun C, Wang Y, et al. *Biochemistry* 2000;39:4778. [PubMed: 10769135]
49. Rydberg EH, Li C, Maurus R, Overall CM, Brayer GD, Withers SG. *Biochemistry* 2002;41:4492. [PubMed: 11914097]
50. Numao S, Maurus R, Sidhu G, Wang Y, Overall CM, Brayer GD, Withers SG. *Biochemistry* 2002;41:215. [PubMed: 11772019]
51. Steer ML, Levitzki A. *FEBS Letters* 1973;31:89. [PubMed: 4710004]
52. Levitzki A, Steer ML. *Eur J Biochem* 1974;41:171. [PubMed: 4856205]
53. Aghajari N, Feller G, Gerday C, Haser R. *Protein Sci* 2002;11:1435. [PubMed: 12021442]
54. Brayer GD, Luo Y, Withers SG. *Protein Sci* 1995;4:1730. [PubMed: 8528071]
55. Ramasubbu N, Paloth V, Luo Y, Brayer GD, Levine MJ. *Acta Crystallog sect D* 1996;52:435.
56. White RP, Meirovitch H. *J Chem Theory Comput* 2006;2:1135. [PubMed: 17429495]
57. Steinbach PJ, Brooks BR. *Proc Natl Acad Sci USA* 1993;90:9135. [PubMed: 8415667]
58. Ponder, JW. TINKER - software tools for molecular design, version 3.9. Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine; St. Louis, Mo: 2001.
59. Meirovitch H, Alexandrowicz Z. *J Stat Phys* 1976;15:123.
60. Meirovitch H. *J Chem Phys* 1999;111:7215.
61. Szarecka A, White RP, Meirovitch H. *J Chem Phys* 2003;119:12084.
62. Meirovitch H, Vásquez M, Scheraga HA. *Biopolymers* 1988;27:1189. [PubMed: 3219393]
63. Brady J, Karplus M. *J Am Chem Soc* 1985;107:6103.
64. Chang CE, Chen W, Gilson MK. *J Chem Theory Comput* 2005;1:1017.
65. Allen, MP.; Tildesley, DJ. *Computer Simulation of Liquids*. Clarendon Press; Oxford: 1987.
66. Zacharias M, Straatsma TP, McCammon JA. *J Chem Phys* 1994;100:9025.
67. Olson MA. *Proteins* 2004;57:645. [PubMed: 15481087]

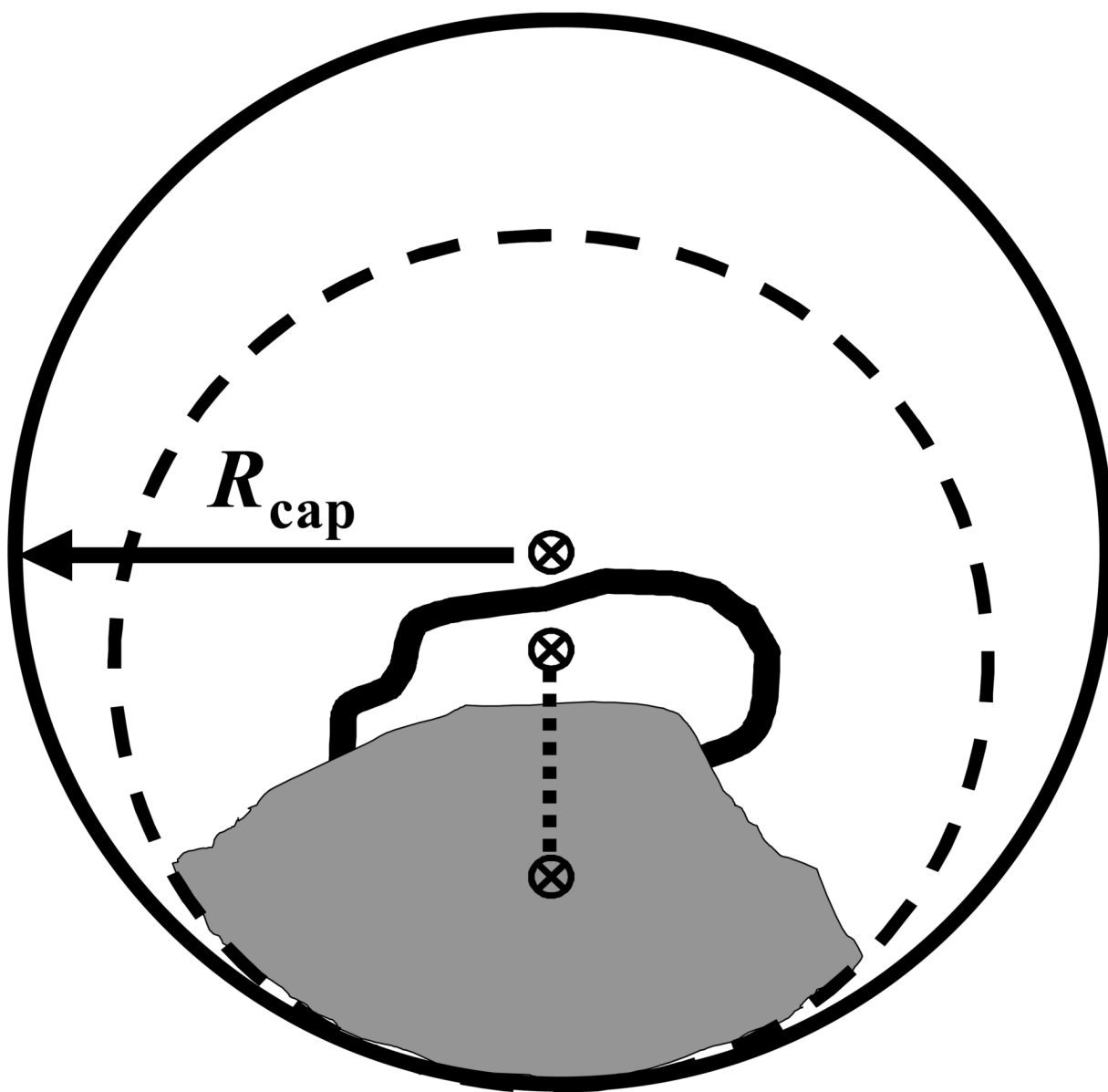


Figure 1.

A two dimensional diagram of the spherical water restraining region. The loop is represented as the heavy black curve, and the protein template is the region shown in gray. The dashed circle (radius = R_{temp}), defines the edge of the template. Three positions are marked with the symbol, \otimes , in the figure. These are, starting from the bottom, \mathbf{x}_{cm} , \mathbf{x}_{cmb} , and \mathbf{x}_{sph} . \mathbf{x}_{cm} is the center of mass of the loop and template atoms, while \mathbf{x}_{cmb} is the center of mass of the loop backbone. \mathbf{x}_{cm} and \mathbf{x}_{cmb} are connected by a dotted line which defines the vector direction (pointing from \mathbf{x}_{cm} to \mathbf{x}_{cmb}) that is used to determine the position of \mathbf{x}_{sph} . (That is, \mathbf{x}_{sph} is shifted away from the template by r_{shift} , see eq 4.) Water molecules are contained within a spherical region defined by the distance, R_{cap} , measured from \mathbf{x}_{sph} . This containment region is represented by the large outer circle. Note that generally, $R_{\text{cap}} > R_{\text{temp}}$, and therefore the edge of this circle (sphere in 3D) is shifted to keep the water molecules on the “loop side” of the model system.

Two sets of differences $\Delta\alpha_k$ (in degrees) between the minimum and maximum values of dihedral angles in the free and bound samples^a

TABLE 1

	Explicit Solvent						Implicit solvent									
	Free			Bound			Free			Bound						
	$\Delta\phi$	$\Delta\Psi$	$\Delta\chi$	$\Delta\phi$	$\Delta\Psi$	$\Delta\chi$	$\Delta\phi$	$\Delta\Psi$	$\Delta\chi$	$\Delta\phi$	$\Delta\Psi$	$\Delta\chi$				
Residue																
Gly 1	47	90	75	121	76	153	92	148								
His 2	86	104	107	70	139	130	125	105								
Gly 3	83	99	173	177	175	124	80	95								
Ala 4	133	87	96	132	131	94	143	288								
Gly 5	73	88	95	88	107	100	199	360								
Gly 6	133	98	116	81	126	109	285	267								
Ser 7	65	59	84	71	83	64	243	109								
χ^1 (His)	78		89		55		53									
χ^2 (His)	107		106		130		108									
χ^1 (Ser)	166		163		317		321									

* $\Delta\alpha_k$ are defined in eq 18. The explicit solvent results were calculated in the present study based on a sample of $n_S=600$ configurations of the loop and 70 TIP3P water molecules using the AMBER force field. The implicit solvent results have been obtained in paper 1¹⁶ from the entire sample of 500 conformations using the AMBER force field and the GB/SA implicit solvent.

TABLE 2HSMD results (in kcal/mol) for the entropy, TS_{loop}^A (eqs 22 and 23) at $T=300$ K for the free and bound microstates^a

		Free loop	Bound loop
Bin size	$n_f(j)$	TS_{loop}^A	TS_{loop}^A
$\Delta\alpha_k/15$	250 (1)	67.18 (4)	68.72 (4)
“	500 (2)	66.48 (7)	67.86 (8)
“	750 (3)	66.17 (4)	67.58 (8)
“	1250 (5)	65.74 (4)	67.19 (8)
“	<u>1250 (2)</u>	<u>69.9 (1)</u>	<u>70.3 (2)</u>
$\Delta\alpha_k/30$	250 (1)	67.04 (9)	68.61 (7)
“	500 (2)	66.22 (7)	67.61 (7)
“	750 (3)	65.77 (4)	67.15 (8)
“	1250 (5)	65.19 (4)	66.49 (3)
“	<u>1250 (2)</u>	<u>69.4 (1)</u>	<u>69.8 (2)</u>
$\Delta\alpha_k/45$	250 (1)	67.03 (4)	68.60 (5)
“	500 (2)	66.17 (7)	67.56 (7)
“	750 (3)	65.69 (4)	67.08 (8)
“	1250 (5)	65.06 (4)	66.36 (8)
“	<u>1250 (2)</u>	<u>69.4 (2)</u>	<u>69.7 (2)</u>
TS^{QH}		78.6 (1)	87 (6)
TS^{LS}		87.4 (1)	90 (7)

^aThe bin sizes are $\delta = \Delta\alpha_k/l$. n_f denotes the sample size of the future chains used in the reconstruction process, $n_f = \text{unit} \times j$, where j is the number of simulations of unit size applied at each reconstruction step. Generation of the samples (of $n_s=600$ conformations) and their reconstruction is based on the AMBER force field¹⁷ and 70 TIP3P water molecules.¹⁹ However, the underscored results for $n_f=1250$ (2) (unit=650) were obtained in paper I from samples of 200 conformations using the AMBER force field and the GB/SA implicit solvation. The statistical error in the last significant digit is given in parentheses, e.g., 65.06 (4) = 66.06 ± 0.04. S^{QH} (eq 36) is the quasi-harmonic entropy and S^{LS} is ΔS_{loop}^A obtained by the local states method using $b=2$ and the discretization parameter, $l=10$ (see II.6); these results were obtained from larger samples (see text for details). The entropy is defined up to an additive constant that is the same for both microstates.

Entropy differences, $T\Delta S_{\text{loop}}^A$ (in kcal/mol) at $T=300$ K between the free and bound microstates obtained by HSMD for different samples in explicit water^a

TABLE 3

Bin size	n_f	$T\Delta S_{\text{loop}}^A$			n_f	$T\Delta S_{\text{loop}}^A$
		$n_s = 600$	$n_s = 200$	$n_s = 100$		
$\Delta\alpha_k/30$	250	-1.6	-1.5 (1)	-1.1 (2)	1250	$n_s = 40$ -0.9 (3)
"	500	-1.4	-1.3 (1)	-1.0 (2)	2500	-0.9 (4)
"	750	-1.4	-1.4 (1)	-1.1 (2)	5000	-0.9 (2)
"	1250	-1.3	-1.3 (1)	0.9 (2)	10^4	-1.0 (2)
$\Delta\alpha_k/45$	250	-1.6	-1.5 (1)	-1.1 (2)	1250	-0.9 (3)
"	500	-1.4	-1.3 (1)	-1.0 (2)	2500	-0.9 (3)
"	750	-1.4	-1.4 (1)	-1.1 (2)	5000	-0.9 (2)
"	1250	-1.3	-1.3 (1)	0.9 (2)	10^4	-1.0 (2)

^a $T\Delta S_{\text{loop}}^A$ is defined in eq 24 and its results are given only for the two smallest bins, $\delta = \Delta\alpha_k/30$ and $\delta = \Delta\alpha_k/45$, using $\text{unit}=2.5$ ps. The table consists of two parts. The results in the left-hand part are based on $250 \leq n_f \leq 1250$ and are presented for the entire (two) samples ($n_s=600$), for the samples' first 200 conformations, and for $n_s=100$ by considering every 6th conformation of the entire sample. The results in the right-hand part are based on $1250 \leq n_f \leq 10^4$ using samples of $n_s=40$ by considering every 15th conformation of the entire sample. The statistical error is defined in Table 2; for $n_s=600$ it is smaller than ± 0.1 . Results for $T\Delta S_{\text{loop}}^{\text{QH}}$ and $T\Delta S_{\text{loop}}^{\text{L-S}}$ are not given due to their low accuracy (see Table 2). All calculations were carried out with the AMBER96 force field and 70 TIP3P water molecules.

Energy and free energy averages for the loop and water obtained from the samples ($n_s=600$) of the free and bound microstates at $T=300$ K^a

TABLE 4

	$\bar{E}_{\text{loop}}(m)$	$\bar{E}_{\text{water}}(m)$	$\bar{E}_{\text{total}}(m)$	$\bar{F}_{\text{water}}^{\text{II}}(m)$
Free	-108.3 ± 2	-1021.9 ± 2	-1130.8 ± 0.5	31.2 ± 0.5
Bound	-106.6 ± 0.7	-1016.5 ± 1.5	-1123.0 ± 1.5	35.5 ± 0.4
Free-Bound	ΔE_{loop}	ΔE_{water}	ΔE_{total}	ΔF_{water}
	-1.7 ± 0.9	-5.4 ± 2.5	-7.2 ± 1.2	-4.4 ± 0.7

^a $\bar{E}_{\text{loop}}(m)$ (eq 25) is the loop-loop and loop-template average energy for microstate m , $\bar{E}_{\text{water}}(m)$ (eq 30) is the water-water, water-loop, and water-template average energy for m where $\bar{E}_{\text{total}}(m)$ is their sum. $\bar{F}_{\text{water}}^{\text{II}}(m)$ (eq 28) is the average free energy of water for fixed loop structures. The corresponding differences are ΔE_{loop} (eq 31), ΔE_{water} (eq 26), ΔE_{total} (eq 32), and ΔF_{total} (eq 33).

All results are in kcal/mol, where $\bar{F}_{\text{water}}^{\text{II}}(m)$ is defined up to an additive constant which is the same for both microstates.

Summary of energy, entropy, and free energy differences (in kcal/mol) for the free and bound microstates at $T=300\text{ K}^a$

TABLE 5

ΔF_{water}	ΔE_{water}	$T\Delta S_{\text{water}}$	ΔE_{loop}	$T\Delta S_{\text{loop}}$	ΔF_{loop}
-4.4 ± 0.7	-5.4 ± 2.5	-1.0 ± 2	-1.7 ± 0.9	-1.3 ± 0.2	-0.4 ± 1
	ΔF_{total}	ΔE_{total}	$T\Delta S_{\text{total}}$		
	-4.8 ± 1	-7.2 ± 1.2	-2.3 ± 2		

^aDifferences (free-bound) in the following quantities: free energy of water, ΔF_{water} (eq 29), energy of water, ΔE_{water} (eq 31), entropy of loop, ΔE_{loop} (eq 26), entropy of water, $T\Delta S_{\text{water}}$ (eq 34), entropy of loop, $T\Delta S_{\text{loop}}$ (eq 24), total free energy, ΔF_{total} (eq 33), total energy, ΔE_{total} (eq 32), and total entropy, $T\Delta S_{\text{total}}$ (eq 35).