entropy

Article

Entropy and Free Energy of a Mobile Loop Based on the Crystal Structures of the Free and Bound Proteins

Mihail Mihailescu and Hagai Meirovitch *

Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, 3059 BST3, Pittsburgh, PA 15260, USA; E-Mail: Mihail.Mihailescu@strayer.edu

* Author to whom correspondence should be addressed; E-Mail: hagaim@pitt.edu; Tel.: +1-412-648-3338; Fax: +1-412-648-3163.

Abstract: A mobile loop changes its conformation from "open" (free enzyme) to "closed" upon ligand binding. The difference in the Helmholtz free energy, ΔF_{loop} between these states sheds light on the mechanism of binding. With our "hypothetical scanning molecular dynamics" (HSMD-TI) method $\Delta F_{\text{loop}} = F_{\text{free}} - F_{\text{bound}}$ where F_{free} and F_{bound} are calculated from two MD samples of the free and bound loop states; the contribution of water is obtained by a thermodynamic integration (TI) procedure. In previous work the free and bound loop structures were both attached to the same "template" which was "cut" from the crystal structure of the free protein. Our results for loop 287−290 of AcetylCholineEsterase agree with the experiment, $\Delta F_{\text{loop}} \sim -4$ kcal/mol if the density of the TIP3P water molecules capping the loop is close to that of bulk water, *i.e.*, $N_{\text{water}} = 140 - 180$ waters in a sphere of a 18 Å radius. Here we calculate ΔF_{loop} for the more realistic case, where two templates are "cut" from the crystal structures, 2dfp.pdb (bound) and 2ace.pdb (free), where $N_{\text{water}} = 40 - 160$; this requires adding a computationally more demanding (second) TI procedure. While the results for $N_{\text{water}} \leq 140$ are computationally sound, ΔF_{loop} is always *positive* (18 ± 2 kcal/mol for $N_{\text{water}} = 140$). These (disagreeing) results are attributed to the large average B-factor, 41.6 of 2dfp (23.4 Å^2 for 2ace). While this conformational uncertainty is an inherent difficulty, the (unstable) results for $N_{\text{water}} = 160$ suggest that it might be alleviated by applying different (initial) structural optimizations to each template.

Keywords: absolute entropy; free energy; mobile loop; ligand binding

1. Introduction

In recent years we have developed a new simulation method for calculating the absolute entropy, S and the absolute Helmholtz free energy, F called the hypothetical scanning Monte Carlo (HSMC) (or HSMD), where molecular dynamics (MD) is used [1–7]. HSMC(D) has been applied to systems of increasing complexity, where the most recent ones have been mobile loops in proteins [8-10]. Typically, a mobile loop has an "open" conformation in the unbound protein, which is changed to a "closed" conformation upon ligand binding, *i.e.*, the loop becomes a "lid" which covers the bound ligand. In these studies the main focus has been on entropy and free energy differences, ΔF_{loop} between these two loop conformations in the bound and free proteins, rather than on the absolute values themselves. (As discussed later, ΔF_{loop} not only sheds light on the mechanism of ligand binding, but in some cases it is an (ignored) component of the absolute free energy of binding.) In the present work we develop HSMD further, extending its applicability to more complex models, in particular to more complex models of loops, as explained below.

Before describing HSMC(D) and its planned enhancements in detail, it should be emphasized that calculation of S and F still constitutes a central problem in computer simulation, in spite of the significant progress achieved in the last 50 years [11–23]. This problem is in particular severe in structural biology due to the flexibility and strong long-range interactions characterizing bio-macromolecules such as proteins. More specifically, the potential energy surface of a protein, $E(\mathbf{x})$ is rugged ($\mathbf{x}$ is the $3N$-dimensional vector of the Cartesian coordinates of the molecule's N atoms), *i.e.*, this 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. A microstate Ω_m, (e.g., the α-helical region of a peptide) 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 [24,25] simulation. A molecule will visit a localized well only for a very short time [several femtoseconds (fs)] while staying much longer within a microstate [26,27] which is thus of a greater physical significance (for further discussions about microstates and their problematic definition in simulations, see [8,9]).

Typically one is interested in calculating the free energy of the most stable microstates rather than calculating the total free energy (of Ω). Thus, flexible protein segments (e.g., sidechains and surface loops), cyclic peptides and ligands bound to proteins can populate significantly several Ω_m in thermodynamic equilibrium, which should be identified and their populations, $p_m = \exp[-F_m/k_BT]$ calculated (k_B is Boltzmann constant and T is the absolute temperature). In particular, identifying the global free energy minimum of a protein is the daunting task of protein folding.

As stated above, we are primarily interested in entropy and free energy differences, which are commonly calculated by thermodynamic integration (TI) techniques where the integration can be carried out over physical quantities such as the energy, temperature, and the specific heat, as well as over non-thermodynamic parameters [free energy perturbation (FEP), and histogram analysis methods are also included in this category] [11–23]. While TI is a robust and highly versatile approach, it has some limitations; for example, if one seeks to calculate $\Delta F_{m,n} = \int dF$ between two microstates m and n with significant structural variance the integration from m to n becomes difficult and for large molecules unfeasible. This drawback of TI can be overcome to a large extent by methods that provide the *absolute F_m* (and 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. HSMC(D) (and other techniques [28–31] such as the quasi-harmonic approximation [32–34]) enable one to calculate the absolute F (and S) (for a more extensive discussion on TI and other techniques for calculating differences, ΔF (and ΔS) see [17]).

With HSMC(D) each conformation i of a given MD or MC sample 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} from which various free energy functionals can be defined. While the TPs of HSMC(D) are stochastic in nature (calculated by MD or MC simulations), all the system interactions are taken into account; in this respect HSMC(D) can be viewed as exact [3], where the only approximation involved is due to insufficient MC (MD) sampling for calculating the TPs. HSMC(D) has unique features: it provides rigorous lower and upper bounds for F, which enable one to determine the accuracy from HSMD 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 [3]).

The HSMC(D) methodology has been developed systematically, as applied to systems of increasing complexity. The initial (HSMC) calculations of liquid argon, TIP3P water [3], self-avoiding walks [5] and polyglycine molecules [6] have verified the validity of the theoretical predictions stated above by comparisons with accurate results obtained by other well established techniques; a subsequent application of HSMD to peptides has led to a 100 times reduction in computer time [7].

As mentioned earlier, HSMD has also been applied to mobile loops calculating $\Delta F_{m,n}$ (and $\Delta S_{m,n}$) between the open and closed microstates. First, we treated the 7-residue mobile loop, 304–310 (Gly-His-Gly-Ala-Gly-Gly-Ser) of the enzyme porcine pancreatic α-amylase [8,9], where the system is modeled by the AMBER96 force field [35] alone and by AMBER96 with the GB/SA implicit solvent of Still and coworkers [36]. In this work only protein atoms close to the loop (the template) within a sphere of radius, R_{tmpl} were considered, where they were kept fixed in their x-ray crystallography positions, while only the loop's atoms were moved by MD. Later the implicit solvent was replaced by explicit solvent, *i.e.*, the α-amylase loop was capped with 70 TIP3P [37] water molecules (within a sphere of radius $R_{\text{water}} = 14$ Å), 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 HSMD-TI process leads to a relatively small statistical error because only the water-loop interactions are involved, where their full values are gradually decreased to zero, for a fixed loop structure.

The number of water molecules used (70) is relatively small relying on a previous paper [38] where the minimal number of water molecules, $N_{\text{water}}^{\text{min}}$ required to cap a loop was studied. $N_{\text{water}}^{\text{min}}$ was determined from a stability criterion based on the loop's root mean square deviation (RMSD) from the Protein Data Bank (pdb) structure during 4 ns MD simulations. Surprisingly, $N_{\text{water}}^{\text{min}}$ (which depends on the loop's sequence and environment) was found to be relatively small on average 12 waters per residue, in accord with a study of Steinbach and Brooks [39] For the loop of amylase, which consists of

small residues, $N_{\text{water}}^{\min} = 70$ seemed to be adequate. Clearly, these conclusions should further be tested and corroborated by free energy calculations.

A detailed such study was carried out in a following paper [10] where HSMD-TI was applied to the loop 287–290 with the bulky residues, Ile, Phe, Arg, and Phe of acetylcholinesterase (AChE) from *Torpedo california*; reaction of AChE with the inhibitor diisopropylphosphorofluoridate (DFP) leads to a displacement of the loop's backbone roughly by 4 Å [40–45], and experimental evidence suggests that the free energy penalty for the loop displacement is on the order of 4 kcal/mol (*i.e.*, $\Delta F_{\text{loop}} = F_{\text{free}} - F_{\text{bound}} \sim -4$ kcal/mol) [40,42,46]. Therefore AChE has been an ideal system for checking the performance of HSMD-TI, and in particular for examining the minimal number of water molecules needed to cap the loop. Thus we carried out two sets of calculations for systems of increasing size defined by $R_{\text{tmpl}} = 12$ and $R_{\text{water}} = 13$ Å, and by $R_{\text{tmpl}} = 13$ and $R_{\text{water}} = 14$ Å, for an increasing number of water molecules, $N_{\text{water}} = 80, 100, 120, 180$ and $80, 100, 120, \ldots, 220$, respectively. We have found that to recover the experimental free energy difference the water density should be close to that of bulk water and our results for the first set, $\Delta F_{\text{loop}} = -3.1 \pm 2.5$ and -3.6 ± 4 kcal/mol for a sphere containing 160 and 180 waters are equal within error bars to the experimental value [10].

Because the crystal structures of the free (unbound) [47] and bound [40] proteins are similar, in [10], (but also in our previous loop studies [8,9] and studies by others [48,49]) the template was "cut" from the crystal structure of the unbound protein, where the loop conformation in the bound protein was attached to the free template by superimposing the two crystal structures. Obviously, it would be more realistic to carry out the calculations with respect to two templates, which are cut separately from the free and bound x-ray structures. However, this would require defining an additional TI procedure where the water-template interactions are also gradually eliminated. This TI is expected to be computationally demanding since, unlike the original TI where (relatively small) loop-water interactions are considered, the number of template-water interacting atoms is relatively large increasing with the increase of the system size.

AChE is expected to be a suitable system for studying the effect of two different templates because the experimental ΔF_{loop} is known. Therefore, the loop of AChE is treated here again, where the simulations are carried out with respect to two different templates which are cut from the free and bound crystal structures; this requires applying the (relatively) elaborate HSMD-TI procedure mentioned above. While the basic theory is similar to that described in [10], we derive it here in a new and shorter way for better clarity.

The foregoing (mainly technical) discussion did not touch on the relation of ΔF_{loop} to ligand binding. One question of interest in this area is whether the conformational change adopted by a mobile loop upon ligand binding is induced by the ligand (induced fit [50,51]) or alternatively, the free loop (in the apo protein) interconverts among different microstates, one of which is selected upon binding (selected fit [52]). While some recent studies favor the latter mechanism, there is no general consensus, and the adopted mechanism is probably system dependent. The present study is not aimed at solving this problem, which would require to apply the longest simulation possible to the entire AChE protein soaked in explicit water, as done, for example, in [53] for studying loop 6 of TIM. Notice, however, that in the case of induced fit ΔF_{loop} (in the apo protein) should be added to the total absolute free energy of binding of the protein-ligand complex. Thus, if the template remains unchanged (due to

binding), calculation of ΔF_{loop} carried out in [10] is expected to be adequate, whereas for a changed template the present calculation should be used. Indeed, in our recent study of the avidin-biotin complex ΔF_{loop} is taken into account (for the first time) in the calculation of the absolute free energy of binding [54].

However, even if the behavior of the loop of AChE is of a selected fit type (*i.e.*, the loop conformations in the apo protein also cover the structural region of the bound loop) the crystal structures and other experimental data suggest that the loop spends appreciable amount of time in its unbound microstate. Therefore, calculating ΔF_{loop}, which measures the difference in stability between the two (possibly) meta-stable microstates is of interest. For this purpose, generating the longest possible MD trajectories might be a disadvantage because the loop is expected to escape from the original microstates during the simulations. Therefore, ΔF_{loop} is calculated from relatively short trajectories (0.5 ns) where the loop remains in its original microstates (bound and free).

In this context it should be pointed out that 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 depend on t as well; in previous publications we have developed procedures for checking that the system remains in its microstate during a simulation. Thus, in the case of two microstates one should verify that the differences, $\Delta E_{\text{loop}}(t)$, $\Delta S_{\text{loop}}(t)$ and $\Delta F_{\text{loop}}(t)$ are stable during a long enough time, t; for more details see [8,9].

Finally, while the differences ΔE_{loop}, ΔS_{loop} and ΔF_{loop} between two loop microstates might be small, the fluctuations in E and S increase as $N^{1/2}$, where N is the number of atoms simulated. Therefore, simulating the entire protein might be prohibitive and to keep the fluctuations manageable the coordinates of the template in our calculations are held fixed. Notice, however, that this is not an inherent limitation as in our present application of HSMD-TI to the avidin-biotin complex [54] part of the template is also moved in the MD simulation since the size of the fluctuations remains under control.

2. Theory and Methodology

2.1. The Loop and the Protein's Template

As in [10] we study the 4-residue loop, Ile, Phe, Arg, 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 [47] is based on residues 4 to 535 and 2dfp, the structure of AChE with DFP consists of residues 2 to 535 (the bound structure) [40]. 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.

Moreover changing the bound structure also consisted in recovering the catalytic residue Serine (200) of this serine hydrolase (2ace) by remodeling the mono-isoprophyl-phosphoryl residue (MIS) (200) of the aged phosphorylated enzyme (2dfp). Remodeling consisted in removing the d-isopropyl-phosphoryl chain without changing the heavy atoms of the modified Serine residue and adding the corresponding hydrogen atoms.

A close analysis shows that these conformations differ overall by (all heavy atoms) 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, sidechains, and all loop heavy atoms, respectively. The RMSD of the templates, *i.e.*, all atoms excluding the loop's atoms, is 1.10 Å. It should be emphasized that the coordinates of the x-ray bound structure (2dfp) appear with significantly larger B-factors than those of the free structure (2ace).

Using the program TINKER [55], and the AMBER force field [35], where Lys, Arg, and His are positively charged and Glu and Asp have a negative charge. Hydrogens were added to the free and bound structures (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 1 kcal/Å^2). Afterwards the loop and (TIP3P) water atoms were allowed to relax in the presence of a fixed template. These minimizations eliminate bad atomic overlaps and strains in the original structures, while keeping the atoms reasonably close to the PDB coordinates.

The relatively weak force constant (1 kcal/Å^2) (as compared to 100 kcal/Å^2 used in [10] for 2ace) was chosen to allow more extensive changes in the bound structure due to its relatively large B-factors and the chemical change imposed on it by removing the phosphate group; correspondingly, larger structural changes were also allowed to occur in the free structure. To assess the change in the structure during minimization we list the various RMSD values (Å) in pairs, for the beginning and the end of the process: RMSD for all atoms, (0.98, 0.99), backbone atoms, (0.43, 0.41), sidechains, (1.20, 1.21), whole template, (0.85, 0.86), template backbone, (0.35, 032), template sidechains, (1.05, 1.06), whole loop (2.71, 2.87), loop backbone (1.39, 1.45), loop sidechains (2.92, 3.10). Thus, the strongest changes occurred for the loops.

Taking into account the whole protein (of 8,284 atoms) would be computationally prohibitive; therefore (as in [8-10]), the template size is reduced to the N_{tmpl} atoms closest to the loop, where the rest of the protein atoms are ignored. More specifically, the center of mass of the backbone atoms of the free loop (in the 2ace structure) is calculated as a (3D) reference point denoted $\mathbf{x}_{cmb}$ and a distance (R_{tmpl}) is chosen. If the distance of any atom of a residue from $\mathbf{x}_{cmb}$ is less than R_{tmpl}, the entire residue is included in the template; otherwise, the residue is eliminated. We use $R_{tmpl} = 12$ Å which was found in [10] to lead to a large enough template consisting of $N_{tmpl} = 944$ atoms; however, for the present (somewhat different) structure $N_{tmpl} = 961$, and exactly all these atoms were included in the template which was "cut" from the crystal structure of the bound protein (2dfp). We added to each template 33 crystal water molecules selected from those which are closest to the loop (30 in [10]).

After defining the templates, the water molecules and the orientations of the polar hydrogens of the free and bound loops (on different templates) were subjected to an optimization procedure. This procedure (where all the loop's heavy atoms are kept fixed), was carried out in several rounds, each consisting of a 0.1 ns MD run (1 fs time step) at high temperature (1,250 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 energy of each system was minimized where the loop and water molecules were allowed to move freely. Each of the final "free" and "bound" structures constitute starting points for a further analysis. The RMSD

values between the final (optimized) loop structure and its crystal structure are relatively small. For the free loop they are 0.18, 0.45, and 0.42 Å, and for the bound loop, 0.27, 0.47, and 0.44 Å for the backbone, sidechains, and all loop atoms, respectively.

2.2. 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 non-trivial problem (see below).

As in [10], our strategy is to add more water molecules to the already existing crystallographic waters. Thus, we define a sphere centered at $\mathbf{x}_{cmb}$ with a radius, R_{water} ($R_{water} = R_{tmpl} + 1$ Å) where waters are added at random to the hemisphere oriented towards 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 $\mathbf{x}_{cmb}$. That is, if the distance of a water oxygen from $\mathbf{x}_{cmb}$ is greater than R_{water} a harmonic restoring force is applied, otherwise the restraining force is zero.

Even though we have found in [10] that for the above R_{tmpl} and R_{water}, N_{water} should be 140–180, it is of interest to check this conclusion again for the present case of two different templates; therefore, we carry out calculations for $N_{water} = 40$, 60, 80, 100, 120, 140, and 160.

Each of the two fully hydrated systems is then treated by a sequence of MD/minimization procedure similar to that outlined above, applied only to the hydrogen atoms of the loop and *all* water molecules restrained within the full sphere of radius R_{water}. Again, at the end of this optimization the energy is minimized where the loop's atoms and the water molecules are free to move. The change in the loops' conformations from their crystal structures measured by the RMSD are relatively small; for the free loop we obtain 0.31, 1.03, and 0.94 Å for backbone, side chains and all loop atoms, respectively, where the corresponding results for the bound loop are 0.37, 1.06, and 0.98 Å (see the discussion related to Table 1).

For each of the optimized "free" and "bound" structures 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, 1,000 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):

$$E_{total} = [E_{loop\text{-}loop} + E_{loop\text{-}tmpl}] + [E_{water\text{-}water} + E_{water\text{-}tmpl} + E_{water\text{-}loop}] = E_{loop} + E_{water} \tag{1}$$

where $E_{loop\text{-}loop}$ is the intra loop energy, $E_{loop\text{-}tmpl}$ 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 $(i = 1, N_{res} = 4)$, the bond angles $\theta_{i,l}$ $(i = 1, N_{res}, l = 1,3)$, 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 [7,8] that to a good approximation bond stretching can be ignored.

2.3. Statistical Mechanics of a Loop in Internal Coordinates

The theory of HSMD-TI has been developed in previous publications; therefore, the present description is shorter, providing mainly equations that are related directly to the calculations.

The reconstruction of the loop structure (of $N_{res} = 4$ residues) 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 $(i = 1, N_{res})$, the bond angles $\theta_{i,l}$ $(i = 1, N_{res}, l = 1, 3)$, the side chain angles χ, and the corresponding bond angles. For convenience, all these angles (ordered along the backbone) are denoted by α_k, $k = 1,K$; We have argued in [7,8] that to a good approximation bond stretching can be ignored, thus the bond lengths are considered to be constant.

The partition function of the loop/water /template system is:

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

where $E(\mathbf{x}_{loop}, \mathbf{x}^N) = E_{total}$ is defined in Equation (1), $\mathbf{x}_{loop}$ is the Cartesian coordinates of the loop in microstate m and $\mathbf{x}^N$ is the $9N$ Cartesian coordinates of the water molecules; E_{total} also depends on the coordinates of the "frozen" template which are omitted for simplicity. For the same reason the letter m will be omitted in most of the equations and N_{water} will be replaced in the theoretical section by N $(N = N_{water})$. After changing the variables of integration from $\mathbf{x}_{loop}$ to internal coordinates, the integral becomes a function of the K dihedral and bond angles, α_k, $k = 1,K$ and a Jacobian, $\cos(\theta_{i,l})$ that depends (only) on each of the bond angles $\theta_{i,l}$ [27,28,32]:

$$Z = \int_m \exp\{[-E_{loop}([\alpha_k]) - E_{water}([\alpha_k], \mathbf{x}^N)] / k_B T\} d[\alpha_k] d\mathbf{x}^N \tag{3}$$

where $[\alpha_k] = [\alpha_1, ... \alpha_K]$ and $d[\alpha_k] = d\alpha_1 ... d\alpha_K$. In Equation (3) we have omitted a factor, which depends on the bond lengths and is *assumed* to be the same (*i.e.*, constant) for different microstates of the same loop and therefore does not affect entropy *differences* (e.g., see discussion in [10]). The Jacobian is also omitted for simplicity because we have shown [8] that it cancels out (within the error bars) in entropy and free energy differences. The Boltzmann probability density corresponding to Z [Equation (3)] is:

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

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

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

and:

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

It should be noted that the fluctuation of the *exact F* is zero [56,57], because by substituting $\rho^B([\alpha_k])$ [Equation (4)] inside the curly brackets of Equation (6) one obtains, $E([\alpha_k]) + k_B T \ln \rho^B([\alpha_k]) = -kT \ln Z = F$, *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. [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]. Still, the fluctuation of an approximate free energy (*i.e.*, which is based on an approximate probability density) is finite and is expected to decrease as the approximation improves [3–5,56,57]. Because HSMC(D) provides an approximation for $\rho^B([\alpha_k], \mathbf{x}^N)$, one can, *in principle*, estimate the free energy of the system from *any single* structure [3–5]. This is the reason why in practice reliable HSMC(D) results for *F* (but not necessarily for *S* and *E*) can be obtained from a relatively small sample.

2.4. Exact Future Scanning Procedure

It should first be pointed out that MC and MD are exact *dynamical* methods that enable one to sample system configuration *i* correctly with its Boltzmann probability, P_i^B, while the *value* of P_i^B is not provided (due to the dynamical character of these methods). [To simplify the discussion we use the discrete probability, P_i^B rather than the probability density $\rho^B([\alpha_k], \mathbf{x}^N)$ defined in Equation (4).] Thus, properties such as the energy that are "written" on *i* can easily be calculated, while a *direct* calculation of the absolute *S* is difficult because $\ln P_i^B$ is unknown (it depend not only on *i* but on the entire ensemble through the partition function *Z*, which cannot be obtained from a finite sample).

Unlike the dynamical MC and MD, the exact future scanning method [that constitutes the basis of HSMC(D)] is a growth procedure which enables one (at least in principle) generating any system configuration (from nothing) by determining the positions of the atoms step-by-step with the help of transition probabilities (TPs); the product of these TPs leads to the value of P_i^B hence to *S* and *F*. Practically, a loop/water/template configuration would be generated by initially building a loop structure (in the presence of moving water) followed by the construction of a configuration of the surrounding water molecules (in the presence of a *fixed* loop conformation). In this way a sample of statistically independent system configurations can be obtained.

For simplicity this construction is described for a loop without side chains, consisting of *M* Gly residues, *i.e.*, ordered along the chain are the 3*M* heavy atoms denoted $k' = 1...3M$ and the 6*M* dihedral and bond angles denoted α_k, $1 \le \alpha_k \le 6M = K$ with values within microstate *m*; the loop is surrounded by N_{water} water molecules moving within the volume defined by a sphere of radius, R_{water}, the template, and the loop. We seek to generate a configuration of the entire system by first generating a loop conformation and then a configuration of the water molecules.

With the scanning procedure the position of the heavy atoms ($\mathbf{x}_{loop}$) is determined step-by step. Thus, the position of the first atom $k' = 1$ is defined by the simultaneous determination of the first pair of dihedral and bond angles α_1, and α_2. The maximum range $\Delta\alpha_1\Delta\alpha_2$ which will keep the loop within *m*

is defined, and each of $\Delta\alpha_1$ and $\Delta\alpha_2$ is divided into n_b small bins (of sizes $\Delta\alpha_1/n_b$ and $\Delta\alpha_2/n_b$) denoted j_1 and j_2, $j_1 = 1\ldots n_b$, $j_2 = 1\ldots n_b$, respectively. A long MD simulation of the whole system (loop + water) is carried out within microstate m, where a conformation is retained every l fs leading to a huge sample of size n; then, the number of conformations n_{j1j2} that visit simultaneously the (double) bin j_1j_2 is calculated from which the corresponding TP is obtained, $p_{j1j2} = n_{j1j2}/n$ (or $\rho_{j1j2} = [n_{j1j2}/n]/[\Delta\alpha_1\Delta\alpha_2/n_b^2]$). A double bin is then selected by a random number according to the probabilities p_{j1j2} which defines the position of atom $k' = 1$ (and its hydrogen or oxygen). The position of this atom is not changed in the next steps of the build-up process, *i.e.*, it becomes part of the "past". The position of the second atom ($k' = 2$) is determined in the same manner from a long MD simulation of the future part of the system (*i.e.*, atoms $k' = 2\ldots 3M$ and water) where α_3 and α_4 are considered, bins $\Delta\alpha_3/n_b$ and $\Delta\alpha_4/n_b$ are defined, probabilities are calculated and a "lottery" (like above) determines the values of α_3 and α_4 which define the position of atom $k' = 2$; the process continues until the positions of all the loop's atoms (and their hydrogens or oxygens) have been determined. A configuration of the N_{water} molecules is then determined in a similar way step-by-step in the presence of the fixed loop structure previously constructed (for details see [3]). Obviously, the smaller are the bins the higher is the accuracy of the construction process, provided that the statistics is adequate, *i.e.*, that the (future) MD simulations are long enough; this stochastic scanning method becomes exact as the bin size $\rightarrow 0$ ($n_b \rightarrow \infty$) and $n \rightarrow \infty$. Notice that in applications of the (deterministic) scanning method to lattice models, only part of the future has been considered, (*i.e.*, only f steps ahead), where this part has been scanned completely; therefore, the corresponding TPs are approximate but deterministic (rather than stochastic), and accurate results were obtained by using an additional importance sampling procedure [58].

This procedure can be described more formally as follows: at step k' ($k = 2k'$), the positions of $k' - 1$ atoms have already been determined (from the values of the corresponding $k - 2$ angles $\alpha_{1, \ldots}, \alpha_{k-2}$) and they are kept fixed (defining the "past"); α_{k-1} and α_k (which will determine the position of atom k') are defined with the exact TP density $\rho(\alpha_{k-1}\alpha_k|\alpha_{k-2}, \cdots, \alpha_1)$:

$$\rho(\alpha_{k-1}\alpha_k|\alpha_{k-2}, \cdots, \alpha_1) = Z_{\text{future}}(\alpha_k\alpha_{k-1}, \cdots, \alpha_1)/[Z_{\text{future}}(\alpha_{k-2}, \cdots, \alpha_1)] \tag{7}$$

where $Z_{\text{future}}(\alpha_k, \cdots, \alpha_1)$ is a future partition function. The term "future" indicates that the integration defining Z_{future} is carried out over the positions of atoms $k' = k/2 + 1\ldots K/2$ (which affect angles, $\alpha_{k+1}, \cdots, \alpha_K$) and the $9N$ coordinates $\mathbf{x}^N$ of the water molecules (which will be determined in future steps of the build-up process). Notice that this integration is carried out in a restrictive way where the corresponding conformations (of the loop) remain within microstate m. Also, in this integration the atoms treated in the past ($1\ldots k' - 1$) (which were determined by $\alpha_1 \cdots \alpha_{k-2}$) are held fixed in their coordinates. For simplicity the integrations below are written over the angles rather than the Cartesian coordinates ($\mathbf{x}_{\text{loop}}$) of the loop atoms, $k' = k/2+1\ldots K/2$. Thus:

$$Z_{\text{future}}(\alpha_k, \cdots, \alpha_1) = \int_m \exp[-(E(\alpha_K, \cdots, \alpha_1, \mathbf{x}^N)/k_B T]d\alpha_{k+1} \cdots d\alpha_K d\mathbf{x}^N \tag{8}$$

where E (Equation (1)) is the total potential energy of the loop/template/water system, which also imposes the loop closure condition. The product of the TPs (Equation (7)) leads to the (Boltzmann) probability density of the entire loop conformation, $\rho_{\text{loop}}^B(\alpha_K, \cdots, \alpha_1)$. After the loop structure has been constructed a configuration of water molecules is generated step-by-step (in the presence of the

constant loop structure where the product of the corresponding TPs leads to the probability density of the water configuration, $\rho_{\text{water}}^{\text{B}}(\alpha_K, \cdots, \alpha_1, \mathbf{x}^N)$. The probability density $\rho^{\text{B}}([\alpha_k], \mathbf{x}^N)$ of the loop/water /template configuration is the product of $\rho_{\text{loop}}^{\text{B}}([\alpha_k])$ and $\rho_{\text{water}}^{\text{B}}([\alpha_k], \mathbf{x}^N)$. One can define for m "the loop entropy of mean force", S_{loop}:

$$S_{\text{loop}} = -k_{\text{B}} \int_m \rho_{\text{loop}}^{\text{B}}([\alpha_k]) \ln \rho_{\text{loop}}^{\text{B}}([\alpha_k]) d[\alpha_k] \tag{9}$$

where S_{loop} is defined up to an additive constant. Extending the exact scanning procedure to side chains is straightforward, where again the position of a side chain atom is defined by two angles as described above.

However, implementation of the exact scanning procedure (as described prior to Equation (7)) for generating Boltzmann weighted configurations of a large loop/water/template system would be inefficient, due to the need to calculate (at each step) a large set of accurate TPs (*i.e.*, for an extremely large number of small bins); this would require extremely long (future) MD simulations. Also, with long simulations it is difficult to guarantee that the loop will remain in microstate m (see discussion in [8]). However, the exact scanning procedure provides the theoretical basis for HSMC(D). Thus, the exact scanning method is equivalent to any other exact simulation technique (in particular, to MC or MD) in the sense that large samples generated by such methods lead to the same averages and fluctuations (the sample does not carry a memory of the simulation method with which it has been generated). 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. With HSMC(D) the efficiency problems discussed above for the exact (stochastic) scanning procedure are alleviated to a large extent since only a single TP is calculated at each step (rather than many TPs (p_j) required with the scanning method), as described below; also, because we are mainly interested in entropy differences, approximations (e.g., ignoring the Jacobians and bond stretching) can be applied without compromising the accuracy of the results.

2.5. Loop Reconstruction by HSMC(D)

The theory of HSMC(D) is described again for a loop consisting of M Gly residues. Notice that while HSMD and the exact scanning method (described in the previous section) have some resemblance, they are different. With the scanning method a loop conformation is generated with P_i^{B} whereas with HSMD an already generated structure (by MD) is reconstructed in order to obtain its probability. Thus, one starts by generating an MD sample of the loop in water in microstate m; the configurations of this sample will be reconstructed by HSMD (application of HSMC is similar). First, the conformations are 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:

$$\Delta\alpha_k = \alpha_k(\text{max}) - \alpha_k(\text{min}) \tag{10}$$

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*. Notice that in the discussion below we define the loop conformation by the set of angles $[\alpha_k]$ rather than by the positions of the loop atoms, $\mathbf{x}_{\text{loop}}$, which are determined by $[\alpha_k]$.

Each of the configurations (frames) $([\alpha_k], \mathbf{x}^N)$ (denoted *i* for brevity) of the sample is reconstructed in two stages, where the biotin structure is reconstructed first followed by the reconstruction of the water configuration $\mathbf{x}^N$. Because the position of atom *k'* is defined by a dihedral and a bond angle one has to calculate their TP simultaneously. Thus, at step *k'* (*k = 2k'*) of stage 1, the *k*−2 angles $\alpha_{k-2} \cdots \alpha_1$ have already been reconstructed and the TP density of $\alpha_{k-1}\alpha_k$, $\rho(\alpha_{k-1}\alpha_k | \alpha_{k-2}, \cdots, \alpha_1)$, is calculated from an MD run, where the entire future of the loop and water is moved [*i.e.*, the loop's atoms *k'*,*k'* + 1,…*K*/2 and their connected hydrogens, or oxygens and the water coordinates $\mathbf{x}^N$] while the past (loop's atoms 1,2,…,*k'*−1 and their connected hydrogens or oxygens) are held fixed at their values in conformation *i*. By considering a future conformation every 10 fs, a sample of size n_f is generated. Two small segments (bins) $\delta\alpha_{k-1}$ and $\delta\alpha_k$ are centered at $\alpha_{k-1}(i)$ and $\alpha_k(i)$, respectively, and the number of simultaneous visits, n_{visit}, of the future chain to these two bins during the simulation is calculated; one obtains (see Figure 1):

$$\rho_{\text{loop}}(\alpha_{k-1}\alpha_k | \alpha_{k-2}, \cdots, \alpha_1) \approx \rho^{\text{HS}}(\alpha_{k-1}\alpha_k | \alpha_{k-2}, \cdots, \alpha_1) = n_{\text{visit}} / [n_f \delta\alpha_{k-1} \delta\alpha_k] \tag{11}$$

where $\rho^{\text{HS}}(\alpha_{k-1}\alpha_k | \alpha_{k-2}, \cdots, \alpha_1)$ becomes exact for very large n_f ($n_f \to \infty$) and very small bins ($\delta\alpha_{k-1}$, $\delta\alpha_k \to 0$). This means that in practice $\rho^{\text{HS}}(\alpha_{k-1}\alpha_k | \alpha_{k-2}, \cdots, \alpha_1)$ will be somewhat approximate due to insufficient future sampling (finite n_f) and relatively large bin sizes. The same treatment can be applied to side chain atoms; however, notice that the sidechain of Arg is treated here approximately, where at each step three consecutive atoms are considered and their positions are defined only by the corresponding dihedral angles. In [8] 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 related to the loop is:
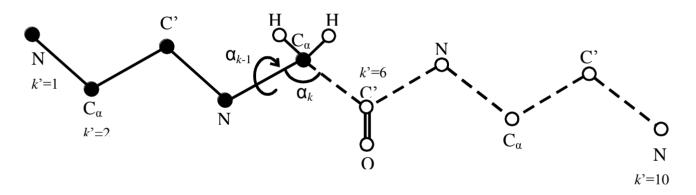
$$\rho^{\text{HS}}(\alpha_K, \cdots, \alpha_1) = \rho^{\text{HS}}([\alpha_k]) = \prod_{k=2,2}^{K} \rho^{\text{HS}}(\alpha_{k-1}\alpha_k | \alpha_{k-2}, \cdots, \alpha_1) \tag{12}$$

$\rho^{\text{HS}}([\alpha_k])$ defines an approximate entropy functional, denoted S_{loop}^A, which can be shown (using Jensen's inequality) to constitute a *rigorous* upper bound for S_{loop} (Equation (9)) [3,59]:

$$S_{\text{loop}}^A = -k_B \int_m \rho_{\text{loop}}^B([\alpha_k]) \ln \rho^{\text{HS}}([\alpha_k]) d[\alpha_K] \tag{13}$$

ρ_{loop}^B (Equation (9)) is the Boltzmann probability density of $[\alpha_K]$ in *m*. ($S_{\text{loop}}^A \geq S_{\text{loop}}$ is also known as the Gibbs' inequality). Being an upper bound suggests that S_{loop}^A will decrease as the approximation improves. In practice, however, this inequality is satisfied only if the probabilities are well defined (e.g., they are correctly normalized). This is pointed out because $\rho^{\text{HS}}(\alpha_K, \cdots, \alpha_1)$ is calculated stochastically and its error (and the error in $\bar{S}_{\text{loop}}^A$, the estimation of S_{loop}^A, see Equation (14)) increases with increasing system size (*i.e.*, increasing the number of TPs; see later discussions in *3.3*).

Figure 1. An illustration of the HSMD reconstruction process of conformation *i* of a peptide consisting of three glycine residues, where for simplicity the oxygens and most of the hydrogens are discarded. At step $k' = 6$ the TPs related to the "past" atoms $k' = 1...5$ (depicted by full spheres) have already been determined and these atoms are kept fixed in their positions at conformation *i*. At this step (6) one calculates the TPs of bond angle α_k (defined by C'-C$_\alpha$-N) and dihedral angle α_{k-1} (defined by C'-C$_\alpha$-N-C') which are related to C', where $k = 2k'$ and thus $k = 12$. The TPs are obtained from an MD simulation where the as yet unreconstructed atoms $k' = 6...10$ (the "future" atoms) are moved (depicted by empty spheres connected by dashed lines) while the past atoms $k' = 1...5$ are kept fixed; notice that the future part should remain within the limits of the microstate and future-past interactions are taken into account. Small bins $\delta\alpha_{k-1}$ and $\delta\alpha_k$ are centered at the values α_{k-1} and α_k in *i*. The TP is calculated from the number of visits (n_{visit} Equation 11) of the future part to $\delta\alpha_{k-1}$ and $\delta\alpha_k$ *simultaneously* during the simulation. After $\rho^{\text{HS}}(\alpha_{11}\alpha_{12}|\alpha_{10},\cdots,\alpha_1)$ (Equation 11) has been determined the coordinates of C' (and O) are fixed at their positions in *i*, *i.e.*, they become "past" atoms and the process continues.

$S_{\text{loop}}^{\text{A}}$ can be estimated (for microstate *m*) from a Boltzmann sample (of size n_{s}) generated by MD using the arithmetic average:

$$\bar{S}_{\text{loop}}^{\text{A}}(m) = -\frac{k_{\text{B}}}{n_{\text{s}}} \sum_{t=1}^{n_{\text{s}}} \ln \rho^{\text{HS}}(t,m) \tag{14}$$

where $\rho^{\text{HS}}(t,m)$ is the value of $\rho^{\text{HS}}([\alpha_k])$ obtained for configuration *t* of the sample of *m*. $\bar{S}_{\text{loop}}^{\text{A}}$ (with the bar) is an estimation of the ensemble average $S_{\text{loop}}^{\text{A}}$ (Equation (13)); 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 Equation (1). $S_{\text{loop}}^{\text{A}}$ (Equations (13) and (14)) constitutes a measure of a pure geometrical character for the loop flexibility, *i.e.*, with no *direct* dependence on the interaction energy. When the *converged* or the *best* value of $S_{\text{loop}}^{\text{A}}$ is considered, it will be denoted by S_{loop}; thus, $F_{\text{loop}} = E_{\text{loop}} - TS_{\text{loop}}$ is defined as the loop's contribution to the total free energy, where E_{loop} is defined in Equation (1). In the same way, the difference in the loop entropies between the free and

bound microstates obtained for a specific set of parameters is denoted by $\Delta S_{\text{loop}}^{\text{A}}$ while the converged difference is denoted by ΔS_{loop}:

$$\Delta S_{\text{loop}}^{\text{A}} = \overline{S}_{\text{loop}}^{\text{A}}(\text{free}) - \overline{S}_{\text{loop}}^{\text{A}}(\text{bound}) \tag{15a}$$

where:

$$\Delta S_{\text{loop}} = \overline{S}_{\text{loop}}^{\text{A}}(\text{free}) - \overline{S}_{\text{loop}}^{\text{A}}(\text{bound}) \qquad (\text{converged}) \tag{15b}$$

One can define a free energy difference for the loop, $\Delta F_{\text{loop}} = \Delta E_{\text{loop}} - T\Delta S_{\text{loop}}$, where ΔE_{loop} is obtained from Equation (1). Notice that (unlike in [10]) ΔS_{loop} and ΔF_{loop} are obtained for different templates; for simplicity their dependence on the templates is omitted.

2.6. Thermodynamic Integration of Water

To reconstruct the water configuration one can use, in principle, the HSMC(D) procedure for fluids mentioned previously [3], which for a fixed loop structure, $[\alpha_k]$ would lead to $\rho_{\text{water}}^{\text{HS}}([\alpha_k], \mathbf{x}^N, \text{tmpl})$ (as an approximation to $\rho_{\text{water}}^{\text{B}}([\alpha_k], \mathbf{x}^N, \text{tmpl})$; see discussion preceding Equation (9)) and then to the contribution of the water configuration to the free energy $F_{\text{water}}^{\text{HS}}([\alpha_k], \mathbf{x}^N, \text{tmpl}) = E_{\text{water}}([\alpha_k], \mathbf{x}^N, \text{tmpl}) + k_{\text{B}}T \ln \rho_{\text{water}}^{\text{HS}}([\alpha_k], \mathbf{x}^N, \text{tmpl})$, which depends on the free or bound template (tmpl). However, this procedure for fluids has not been optimized yet and it is relatively time consuming.

Alternatively, as in [9,10], one can obtain $F_{\text{water}}([\alpha_k], \mathbf{x}^N, \text{tmpl})$ (defined up to an additive constant) by a thermodynamic integration (TI) procedure based on the same reference state for the free and bound structures (*i.e.*, the above constant is the same for both structures). In this state only the water-water interactions and the harmonic restraining force that keeps the waters in the spherical volume are preserved, while the water-template and water-loop interactions [electrostatic and Lennard Jones (LJ)] are switched off, *i.e.*, the fixed loop structure $[\alpha_k]$ and the fixed template do not "see" the surrounding waters. Thus, ΔF_{water} (our main interest) can be obtained by a TI procedure (applied to each system) where these interactions are gradually increased (from zero) during an MD simulation of water (while the loop structure and template remain fixed). (Notice that while the templates are structurally different, the template-template interactions are ignored, therefore the templates are considered to be in the same state.)

In practice, however, it is more convenient to carry out the integration in an opposite direction, starting from the current water configuration with full water-loop and water-template interactions, which are gradually decreased to zero. More specifically, the water-loop interactions are integrated (eliminated) first (in the presence of intact water-template interactions), which leads to $F_{\text{water}}^{\text{TI}}([\alpha_k], m)$. The subsequent template–water integration (that depends only on the template) is denoted $F_{\text{water}}^{\text{TI}}(\text{tmpl}, m)$, where m stands for "free" or "bound". Each of the free and bound integrations are carried out in two stages, where in the first stage the electrostatic interactions (*i.e.*, the charges) are decreased to zero followed by decreasing to zero the Lennard Jones (LJ) interactions, where the latter is performed by decreasing a parameter λ from 1 to zero (see Section 3.3 and Appendix).

The water-loop integration is carried out n_s times for each of the loop structures $[\alpha_k]$ of the sample of size n_s. Denoting these $[\alpha_k]$ of the sample by t and keeping for brevity the letter m ("free" or "bound") only on the left side of the equation, leads to:

$$F_{\text{water}}^{\text{TI}}(\text{loop}, m) = F_{\text{water}}^{\text{TI}}(\text{ch, loop}) + F_{\text{water}}^{\text{TI}}(\text{LJ, loop}) = \frac{1}{n_s}\sum_{t=1}^{n_s} F_{\text{water}}^{\text{TI}}(\text{ch}, t) + F_{\text{water}}^{\text{TI}}(\text{LJ}, t) \qquad (16)$$

Unlike the different $[\alpha_k]$ in the sample, the template remains constant and therefore, in principle, only a single integration is needed. However, to decrease the statistical fluctuations, n_s' integrations (denoted by t) are carried out, each starts from a different water configuration and a set of velocities (denoted t) thus:

$$F_{\text{water}}^{\text{TI}}(\text{tmpl}, m) = F_{\text{water}}^{\text{TI}}(\text{ch, tmpl}) + F_{\text{water}}^{\text{TI}}(\text{LJ, tmpl}) = \frac{1}{n_s'}\sum_{t=1}^{n_s'} F_{\text{water}}^{\text{TI}}(\text{ch, tmpl}, t) + F_{\text{water}}^{\text{TI}}(\text{LJ, tmpl}, t) \qquad (17)$$

It should be pointed out again that the template-template energy is not considered in our model, therefore the two templates only differ by their geometry and the corresponding reference states are thus the same.

2.7. 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 $g = 10$ fs the current conformation is considered, where the n_{init} initial considered conformations are discarded for equilibration. For each of the next n_f (considered) future conformations the pair (α_{k-1}, α_k) $(k = 2k')$ are calculated and their contribution to n_{visit} (Equation (11)) 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 (atoms $1 \ldots k'-1$) is kept fixed, the (future) unfixed part (atoms k', $k'+1\ldots$) 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 [6–9] 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. 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 Equation (10) are relatively small; see discussions in *3.2*).

3. Results and Discussion

3.1. Simulation Details

The solvated free and bound structures (defined on the corresponding x-ray structures) were initially optimized as described in sections 2.1 and 2.2 leading to two optimized structures. Then, starting from each optimized structure, an 1 ns MD trajectory was generated at 300 K, where the initial 0.5 ns part

was used for equilibration and a sample of 1,000 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 on their different templates) the free energy and entropy are calculated.

All simulations were carried out with the velocity-Verlet algorithm [60] 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 [60]; the Berendsen [60] 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, δ, where results are presented for $\delta = \Delta\alpha_k/l$, $l = 5, 10, 20, 30, 40$, and 50, centered at α_k (*i.e.*, $\alpha_k \pm \delta/2$) (Equation (11)). 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, an event of zero counts is very rare.

3.2. Dihedral Angles for Different Microstates

In Table 1 we present results for α_k(min), α_k(max) and $\Delta\alpha_k$ (Equation (10)) for the backbone dihedral angles φ and ψ and the sidechains, χ. These values are based on the samples of 1,000 conformations generated for the free and bound microstates for $N_{water} = 140$ (rather than the results for $N_{water} = 160$, which show some inconsistencies, see *3.4*). The table reveals that the $\Delta\alpha_k$ values for the backbone of the free and bound microstates are relatively small (in most cases smaller than 80°) and they are comparable to those obtained by other samples (based on different N_{water} values). Somewhat larger $\Delta\alpha_k$ values were obtained for χ, where all these results are also comparable to those obtained in [10]; this suggests that ΔS_{loop} (Equation (15b)) would be small as well. 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 1.3 of [8]), we have accepted an "escape" criterion for a dihedral angle when α_k(crystal) $+ 60^o$ 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) $\pm 60^o$; these angles are bold-faced in the table. The table reveals that only one backbone angle, φ(Arg) of both the free and bound microstates has been escaped but with small deviations of 18° and 10°, respectively. The number of escaped sidechain angles is seven for the free microstate and five for the bound one. Thus, the original microstates are retained for the backbone but only partially for the side chains, which might be expected since we model the loops in solution rather in the crystal environment.

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 140 free and bound samples*.

Residue angle		α_k(crystal)	α_k(min)	α_k(max)	$\Delta\alpha_k$	α_k(crystal)	α_k(min)	α_k(max)	$\Delta\alpha_k$
		Free				Bound			
Ser	ψ	165	135	162	27	172	159	186	27
	ω	179	156	179	23	177	142	172	30
Ile	φ	−120	−150	−109	41	−57	−60	−8	52
	ψ	138	136	185	49	−44	−93	−31	62
Phe	φ	70	29	79	50	−98	−131	−69	62
	ψ	43	32	88	56	130	73	139	66
Arg	φ	−139	−217	−133	84	−113	−126	−43	83
	ψ	136	84	160	76	13	−37	16	53
Phe	φ	−122	−155	−82	73	−88	−93	−41	52
Side Chains									
Ile	χ^1	−118	−93	−38	55	−82	−65	−25	40
	χ^2	38	−98	−34	64	−47	−62	−17	45
	χ^3	180	−180	180	360	180	−179	180	359
	$\chi^{2'}$	180	31	91	60	180	14	77	63
Phe	χ^1	−139	−174	−119	55	−62	−86	−40	46
	χ^2	−36	−58	8	66	−51	−94	−31	63
Arg	χ^1	−68	−129	−51	78	−72	−88	−39	49
	χ^2	−55	−223	−148	75	−71	−228	−150	78
	χ^3	−170	30	104	74	−50	−36	90	126
	χ^4	−97	61	176	115	176	−183	−51	132
	χ^5	0	−37	35	72	0.4	−29	37	66
	χ^6	0	−42	42	84	0	−39	34	73
	$\chi^{6'}$	0	−37	35	72	0	−48	44	92
Phe	χ^1	−43	−62	−13	49	40	36	76	40
	χ^2	−80	−104	−41	63	−88	−132	−79	53

*α_k(min), α_k(max), and $\Delta\alpha_k$ are defined in Equation (10); their values were calculated from samples of 1,000 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 [47] and 2dfp [40] of the free and bound protein, respectively.

Table 2. HSMD results (in kcal/mol) for the loop entropy, $TS_{\text{loop}}^{\text{A}}$ and for entropy differences $TS_{\text{loop}}^{\text{A}}$ between the free and bound microstates at $T = 300$*.

Bin size	n_f	$N_{\text{water}} = 140$		$T\Delta S_{\text{loop}}^{\text{A}}$
		$TS_{\text{loop}}^{\text{A}}$		
		Free	Bound	
$\Delta\alpha_k/5$	2,000	63.4	63.3	0.1
$\Delta\alpha_k/10$	2,000	61.6	61.5.	0.1
$\Delta\alpha_k/20$	2,000	61.3	61.4	−0.1
$\Delta\alpha_k/30$	200	60.8	60.9	−0.1
`"`	400	61.0	61.1	−0.1
`"`	1,200	61.3	61.3	0.0
`"`	1,600	61.3	61.3	0.0
	2,000	61.3	61.4	−0.1
$\Delta\alpha_k/40$	200	60.8	60.9	−0.1
`"`	400	61.0	61.1	−0.1
`"`	1,200	61.2	61.3	−0.1
`"`	1,600	61.3	61.4	−0.1
	2,000	61.3	61.4	−0.1
$\Delta\alpha_k/50$	200	60.8	60.9	−0.1
`"`	400	61.0	61.1	−0.1
`"`	1,200	61.3	61.3	0.0
`"`	1,600	61.3	61.3	0.0
	2,000	61.3	61.3	0.0
Errors $\leq$		±0.3	±0.3	0.0
Converged		61.3 ±0.3	61.3 ±0.3	0.0 ±0.2
QH	10^4	72.1 ±1.0	71.5 ±1.1	0.6 ±2.0

*Results are presented for the loop entropy, $S_{\text{loop}}^{\text{A}}$ (Equations (13) and (14)) and for the differences, $T\Delta S_{\text{loop}}^{\text{A}} = T[S_{\text{loop}}^{\text{A}}(\text{free}) - S_{\text{loop}}^{\text{A}}(\text{bound})]$ (Equation (15a)]; they were obtained by reconstructing 80 loop structures selected homogeneously from larger MD samples (of 1,000 water-loop configurations) of the free and bound microstates obtained for $N_{\text{water}} = 140$. The results are calculated as a function of the bin size $\delta = \Delta\alpha_k/l$ (Equation (10)) and n_{f} (Equation (11)) the sample size of the future chains used in the reconstruction process.

3.3. Results for the Loop Entropy

Results for the loop entropy, $S_{\text{loop}}^{\text{A}}$ [Equations (13) and (14)] appear in Table 2 for $N_{\text{water}} = 140$ for the free and bound microstates and for their difference $T[S_{\text{loop}}^{\text{A}}(\text{free}) - S_{\text{loop}}^{\text{A}}(\text{bound})] = T\Delta S_{\text{loop}}^{\text{A}}$ [see the discussion preceding Equation (15a)]. These results were obtained by reconstructing $n_{\text{s}} = 80$ loop

structures, distributed homogeneously along the entire sample of 1,000 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 [Equation (11)], where $n_f = 200, 400, 1,200, 1,600,$ and $2,000$; these values of n_f are used for pairs of angles, such as a backbone dihedral and the successive bond angle. However, for the sidechains we also reconstruct a single χ angle and triplets of successive χ angles (e.g., for Arg) for which the maximal n_f is $1,000$ and $4,000$ (rather than $2,000$), respectively (see Equation (11)). The results are also presented as a function of bin size, $\delta = \Delta\alpha_k/l$ (Equations (10) and (11)) where $l = 30, 40$ and 50, while for $n_f = 2,000$ we also provide results for larger bin sizes defined by $l = 5, 10$ and 20. The statistical errors were obtained from the fluctuations (standard deviation).

Being an upper bound, $TS_{\text{loop}}^{\text{A}}$ [Equation (13)] is expected to decrease as the approximation improves, *i.e.*, with decreasing the bin size — an expectation that is fully satisfied. For example, for $n_f = 2,000$, the $TS_{\text{loop}}^{\text{A}}$(free) values are $63.4, 61.6, 61.3, 61.3, 61.3,$ and 61.3 (kcal/mol + constant), *i.e.*, they decrease for $l = 5, 10,$ and 20 and converge to 61.3 (±0.3) kcal/mol for $l = 30, 40$ and 50; a similar behavior is observed for $TS_{\text{loop}}^{\text{A}}$(bound). One would also expect $TS_{\text{loop}}^{\text{A}}$ to decrease as n_f increases in each bin. On the other hand, the table reveals that the central values of $TS_{\text{loop}}^{\text{A}}$ slightly increase in going from $n_f = 200$ to $1,200$ and then become constant. While most of these changes are within the error bars, they reflect an insufficient equilibration of 2.5 ps (see discussion following Equation (13)). Thus, eliminating the results for $n_f = 200$ (*i.e.*, increasing the equilibration to 4.5 ps) would lead to the expected decrease among the (new) $TS_{\text{loop}}^{\text{A}}$ results due to the expected increase in the (new) results for $n_f = 200$ and $1,000$ (however, to be consistent with [10] the results in the table are based on a 2.5 ps equilibration). In any case, within the error bars the $TS_{\text{loop}}^{\text{A}}$ values in Table 2 can be considered as converged; they are only slightly larger (by ~1 kcal/mol) than those of Table 4 of [10] and the errors in both tables are comparable.

We are mostly interested in the difference in entropy between the free and bound microstates, $T\Delta S_{\text{loop}}$ [Equation (15b)]. Table 2 shows that for all approximations $T\Delta S_{\text{loop}} \sim 0$ kcal/mol, *i.e.*, the entropy of both microstates is the same, while in [10] for 160 water molecules $T\Delta S_{\text{loop}}(n_s = 80) = 1.2 \pm 0.1$ kcal/mol, *i.e.*, the free microstate has slightly higher entropy than the bound one.

The HSMD results for the entropy are compared to those obtained with the quasi-harmonic (QH) method [32], *i.e.*, based on the equation:

$$S_{\text{loop}}^{\text{QH}}(m) = (k_{\text{B}}/2)\{N + \ln[(2\pi)^N \text{Det}(\boldsymbol{\sigma})]\} \tag{18}$$

where the covariance matrix, $\boldsymbol{\sigma}$, is obtained from a local MD sample and N is the number of internal coordinates used in the HSMD procedure. It should be pointed out that in a detailed study of QH Chang *et al.* [61] have found that the method might be unreliable when used in Cartesian coordinates or applied (in internal coordinates) to several microstates. On the other hand, QH was found suitable for treating a single microstate, while the convergence of the results is slow and large samples are typically needed; also, because QH takes into account only second order correlations $S_{\text{loop}}^{\text{QH}}(m)$ constitutes an upper bound for the correct S. Still, entropy differences $\Delta S_{\text{loop}}^{\text{QH}}(m)$ are expected to be reliable (see also [21]). Thus, to obtain reasonable precision, results for $S_{\text{loop}}^{\text{QH}}(m)$ (m is bound or free)

were obtained from MD samples of 10,000 loop-water configurations. To avoid the "escape" of a sample from the original microstate, it consists of 10 separate samples of 1,000 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 values of $TS_{\text{loop}}^{\text{QH}}$ exceed the HSMD results for TS_{loop} (for $n_{\text{s}} = 2{,}000$) by 11–10 kcal/mol for the free and bound microstates. These elevated results are in accord with $S_{\text{loop}}^{\text{QH}}$ being an upper bound and are comparable to the overestimation values found in our previous studies [6-10]. On the other hand, $S_{\text{loop}}^{\text{QH}}(\text{free}) - S_{\text{loop}}^{\text{QH}}(\text{bound}) = 0.6 \pm 2.0$ is in accord (within the error bars) with the HSMD result $\Delta S_{\text{loop}} = 0$.

As in [10], the computer time required to reconstruct a loop structure capped with $N_{\text{water}} = 160$ is 8.1 hours CPU on a 2.1 GHz Athlon processor, meaning that the entire reconstructions required 1,296 and 1,472 h CPU; this time is smaller for $N_{\text{water}} = 140$. However, we have shown that considering only 10% ($n_{\text{f}} = 200$) of the maximal reconstruction samples and using smaller samples of $n_{\text{s}} = 40$ (rather than 80) have led to sufficiently accurate entropy differences, meaning that the total computer time can be reduced to 65 h CPU, respectively. We have generated the relatively large reconstruction samples to verify the convergence of the results.

3.4. The Effect of Water

In Table 3 results are presented for the loop energy, $E_{\text{loop-tmpt}}$, $E_{\text{loop-loop}}$, and their sum, E_{loop} (Equation (1)), for the loop-water free energies (calculated by TI), $F_{\text{water}}^{\text{TI}}(\text{ch,loop})$, $F_{\text{water}}^{\text{TI}}(\text{LJ,loop})$, and their sum $F_{\text{water}}^{\text{TI}}(\text{loop})$ [Equation (16)], for the water-template free energies $F_{\text{water}}^{\text{TI}}(\text{ch, tmpl})$, $F_{\text{water}}^{\text{TI}}(\text{LJ, tmpl})$ and their sum $F_{\text{water}}^{\text{TI}}(\text{tmpl})$ (Equation (17)); we also provide results for $F_{\text{sum}} = E_{\text{loop}} + F_{\text{water}}^{\text{TI}}(\text{loop}) + F_{\text{water}}^{\text{TI}}(\text{tmpl})$ (which is the total free energy without the contribution of the loop entropy, S_{loop}, Equations (13) and (14)), and for the total energy, E_{total} (Equation (1)). For simplicity we have omitted the letter "*m*" from these quantities; however, each one of them is calculated for the free and bound microstates, and their difference, Δ (free bound) is also provided. These results are obtained for $40 \leq N_{\text{water}} \leq 160$. Details about the integrations are described in the Appendix.

To check the stability of these results, and assess their statistical errors they were calculated for an increasing sample size (not shown), where the maximal n_{s} is 80 ($N_{\text{water}} \leq 140$) and 40 for $N_{\text{water}} = 160$. Statistical errors, $s/(n_{\text{s}})^{1/2}$, where s is the standard deviation, are provided only for $N_{\text{water}} = 140$ and 160, which are larger than those obtained for $N_{\text{water}} < 140$. The largest error of 2.5 kcal/mol is for $E_{\text{total}}(N_{\text{water}} = 160)$, where the other errors are significantly smaller. The errors in the differences, Δ are again very stable, and smaller than 2.6 kcal/mol for $\Delta F_{\text{sum}}(N_{\text{water}} = 160)$; these errors are comparable to those obtained in [10]. It is in particular encouraging that small errors are found for the water-template free energies, $F_{\text{water}}^{\text{TI}}(\text{ch, tmpl})$, $F_{\text{water}}^{\text{TI}}(\text{LJ, tmpl})$ and their sum $F_{\text{water}}^{\text{TI}}(\text{tmpl})$, which are based on an integration of a significant number of atoms and are calculated for the first time in this paper.

Based on [10], one would expect to obtain the experimental difference in free energy, $\Delta F_{\text{loop}} = F_{\text{free}} - F_{\text{bound}} \sim -4$ kcal/mol for water densities close to that of bulk water, which for the present system occur for N_{water} ranging from 140 to 180. Still, it is of interest to check this expectation by calculating results also for smaller N_{water}. We have seen in *3.2* that for $N_{\text{water}} = 140$ $\Delta S_{\text{loop}} \sim 0$, where small ΔS_{loop} was found also in [10]. We therefore assume that $\Delta S_{\text{loop}} \sim 0$ for other sets of results

($N_{\text{water}} \neq 140$) meaning that ΔF_{sum} actually constitutes the difference in free energy between the two microstates. However, the results for ΔF_{sum} in the table are positive for all N_{water}, even for $N_{\text{water}} = 140$ and 160, *i.e.*, they differ significantly from the experimental value; in particular, for $N_{\text{water}} = 140$ $\Delta F_{\text{sum}} = 18 \pm 2$ kcal/mol.

Expecting ΔS_{loop} to be close to zero for all values of N_{water}, also means that the total entropy, $T\Delta S_{\text{total}} = \Delta E_{\text{total}} - \Delta F_{\text{sum}}$ is mainly due to water. The table shows that in all cases ΔE_{total} is positive, *i.e.*, the total energy of the free loop is larger than that of the bound one. Correspondingly, in all cases (besides for $N_{\text{water}} = 100$), ΔE_{total} is larger than ΔF_{sum}, *i.e.*, the entropy of the free loop (as expected) is also the largest, where $T\Delta S_{\text{total}}$ ranging from 5 ($N_{\text{water}} = 140$) to 42 kcal/mol for $N_{\text{water}} = 160$. Thus the entropy contributes significantly to the free energy where it constitutes between 20 and 64% of the total energy.

Before returning to the experimental-computational disagreement related to ΔF_{loop}, it is of interest to discuss the results for $N_{\text{water}} = 160$. These results (in particular those for the differences Δ) show an unexpected "jump" as compared to their counterparts for $N_{\text{water}} < 160$. Thus, while the results for ΔF_{sum} for $60 \leq N_{\text{water}} \leq 140$ are comparable (ranging between 15 and 26 kcal/mol), $\Delta F_{\text{sum}}(N_{\text{water}} = 160) = 95$ kcal/mol. Correspondingly, ΔE_{total} increases from 23 ($N_{\text{water}} = 140$) to 137 kcal/mol ($N_{\text{water}} = 160$). The main contributors for these results are the relatively large values of the sum of the loop energies, $\Delta E_{\text{loop}} = 98$ and $\Delta F_{\text{water}}^{\text{TI}}(\text{LJ, tmpl}) = 55$ kcal/mol; these values are significantly larger than their counterparts for $N_{\text{water}} < 160$. To check this point further we integrated the water-template interactions in our sample for the free microstate generated in [10] for $N_{\text{water}} = 160$ to find the still relatively large results, $\Delta F_{\text{sum}} = 43$ kcal/mol and $\Delta E_{\text{total}} = 134$ kcal/mol. In these calculations (which do not appear in the table) $\Delta E_{\text{loop}} = 117$ contributes significantly to the large ΔF_{sum}; on the other hand, $\Delta F_{\text{water}}^{\text{TI}}(\text{LJ, tmpl}) = -64$ kcal/mol is negative! It has been difficult to identify the configurationl elements that lead to these energetic changes even after examining the individual configurations numerically and by computer graphics. However, since the template is fixed, these results suggest that as the number of waters increase to $N_{\text{water}} = 160$ their configuration undergoes a significant change, probably due to the increase of pressure in the sphere. Thus, water molecules that stay in the bulk for $N_{\text{water}} < 160$ enter the template already in the equilibration stage (*i.e.*, when the water-template interactions are still intact) affecting thereby the results for $\Delta F_{\text{water}}^{\text{TI}}(\text{LJ, tmpl})$ and $\Delta F_{\text{water}}^{\text{TI}}(\text{ch, tmpl})$.

These configurational changes of water depend on the template which is different in [10] than in the present study. In this context it should be noted that there is a large cavity in the active site of the bound template as the catalytic residue was trimmed significantly to become a smaller Ser residue (see *2.1*). It should be noted, that this effect of water is expected and indeed is found to be less pronounced in the loop-water integrations, $F_{\text{water}}^{\text{TI}}(\text{ch,loop})$ and $F_{\text{water}}^{\text{TI}}(\text{LJ,loop})$ (Equation (16)). The problems involving the results for $N_{\text{water}} = 160$ are not expected to disappear to a large extent when a single template is used, because changes in the arrangement of water will affect approximately equally the results for the free and bound microstates and the corresponding deviations will mostly be cancelled in differences.

Unlike the non-consistent results for $N_{\text{water}} = 160$ discussed above, the results for $N_{\text{water}} < 160$ show a consistent behavior, and based on [10] one would expect the simulations with $N_{\text{water}} = 140$ to lead to the experimental, $\Delta F_{\text{loop}} = -4$ rather than $\Delta F_{\text{loop}} = +18$ kcal/mol. This disagreement might stem from

the uncertainty in the crystal structure of the bound protein, 2dfp, which is characterized by large B-factors (with an average of 41.6 Å^2) whereas for many atoms the values are much larger; for comparison, the average B-factors in 2ace, the crystal structure of free AChE, is 23.4 Å^2. The large cavity formed in the active site of the bound template mentioned above might also affect the results for the bound protein. These are inherent problems which might not be easy to overcome completely; however, one might alleviated them by applying different initial optimizations to the two templates. Thus, in this work we have used the same harmonic force constants of 1 kcal/(molÅ^2) to the free and bound x-ray structures, while it is plausible to anticipate that the force constant should be commensurate with the quality of the crystal structure, *i.e.*, stronger for well defined structures. The results for $N_{water} = 160$ provide some support for this assumption; thus, $F_{sum} = 43$ is better (*i.e.*, smaller) than $F_{sum} = 95$ kcal/mol where these values are based on templates optimized by the force constants, 100 ([10]) and 1 kcal/(molÅ^2) (here), respectively. However, the optimal force constants are unknown a-priori and their determination probably would require quite a few trials.

In summary: The fact that the experimental value, $\Delta F_{loop} = -4$ is known makes AChE a suitable system for checking the calculation of ΔF_{loop} based on two different templates. As previously discussed, such calculations are of interest, as they mimic better the experimental conditions, and for an enzyme with a mobile loop, ΔF_{loop} might be an important ingredient of the total absolute free energy of binding. Our calculations have revealed some instability in the results for $N_{water} = 160$ but have led to consistent results for $N_{water} \leq 140$; obviously, it is somewhat disappointing that $\Delta F_{loop} = -4$ kcal/mol has not been recovered for $N_{water} = 140$. However, this is not a failure of HSMD-TI per se but reflects the uncertainty in the crystal structure of the bound protein. One should also bear in mind that no other studies of a loop attached to two templates have been carried out thus far and besides our HSMD-TI papers [8–10,62] only few calculations of ΔF_{loop} (based on a single template) are available [48,49]. Thus, the present work should be considered as the first step and a basis for future more extensive studies, which can be pursued in several avenues: for example, HSMD-TI can be applied to a mobile loop with known ΔF_{loop} where the quality of the crystal structures of both, the bound and the free protein is high. As is already stated above, one can try optimizing the templates by changing the force constants in a systematic way for $N_{water} = 140$, and more work can be done to elucidate the problematic behavior of the results for $N_{water} = 160$, which again reflects problems in the simulation itself rather than in the application of HSMD-TI.

Table 3. Results in kcal/mol for energy and free energy components and their difference (Δ) between the free and bound microstates*.

	E loop energy			*F* loop-water (TI)			*F* template-water (TI)				
	template-loop	loop-loop	sum	CH	LJ	sum	CH	LJ	sum	*F* sum	*E* total
40 waters $n_s = 40$											
Free	−93	−48	−140	−60	8	−52	−242	−12	−254	−446	−432
Bound	−173	−49	−222	−13	−1	−14	−246	−11	−257	−493	−494
Δ	81	1	82	−47	9	−38	3	0	3	47	61

Table 3. *Cont.*

60 waters $n_s = 40$											
Free	−91	−51	−142	−60	12	−48	−313	−2	−315	−505	−675
Bound	−174	−47	−221	−8	15	7	−305	−4	−309	−523	−725
Δ	84	−4	79	−52	−2	−55	−8	2	−7	18	50
80 waters $n_s = 40$											
Free	−90	−49	−139	−67	45	−23	−350	11	−339	−500	−874
Bound	−168	−47	−215	−11	30	19	−354	29	−326	−521	−928
Δ	78	−2	76	−56	15	−41	5	−18	−13	21	54
100 waters $n_s = 80$											
Free	−108	−39	−147	−66	51	−15	−379	37	−342	−504	−1075
Bound	−168	−45	−213	−6	46	40	−395	38	−357	−530	−1134
Δ	60	6	66	−60	5	−55	16	−1	15	26	5
120 waters $n_s = 80$											
Free	−105	−39	−143	−64	59	−5	−426	73	−353	−500	−1296
Bound	−152	−40	−192	−24	63	40	−442	79	−363	−516	−1339
Δ	47	1	49	−40	−4	−44	16	−6	11	15	38
140 waters $n_s = 80$											
Free	−89 (0.5)	−32 (0.7)	−121 (0.9)	−75 (0.3)	82 (0.8)	8 (0.8)	−471 (0.4)	126 (0.6)	−345 (0.6)	−459 (1.3)	−1491 (1.4)
Bound	−128 (0.4)	−36 (0.7)	−164 (0.8)	−44 (0.4)	76 (0.6)	32 (0.8)	−479 (0.6)	135 (0.8)	−344 (0.9)	−477 (1.4)	−1515 (1.6)
Δ	39 (0.6)	4 (1.1)	43 (1.2)	−30 (1.5)	6 (1.0)	−24 (1.1)	8 (0.6)	−9 (1.0)	−1 (1.1)	18 (1.9)	23 (2.2)
160 waters $n_s = 40$											
Free	−82 (0.7)	−28 (0.9)	−111 (1.0)	−73 (0.6)	106 (1.0)	33 (1.3)	−523 (0.5)	332 (1.4)	−191 (1.5)	−269 (2.1)	−1524 (2.5)
Bound	−167 (0.8)	−42 (1.2)	−209 (1.2)	−8 (0.7)	96 (0.9)	88 (0.9)	−520 (0.5)	276 (1.6)	−244 (1.5)	−364 (1.6)	−1661 (2.4)
Δ	84 (0.7)	13 (1.2)	98 (1.2)	−65 (0.7)	10 (0.9)	−54 (1.0)	−2 (0.5)	55 (1.7)	53 (1.7)	95 (2.1)	137 (2.6)

*The energies, $E_{\text{loop-tmpl}}$, $E_{\text{loop-loop}}$, their sum, E_{loop}, and the total energy, E_{total} are defined in Equation (1). $F_{\text{loop-water}}(\text{TI})$ are $F_{\text{water}}^{\text{TI}}(\text{ch})$ and $F_{\text{water}}^{\text{TI}}(\text{LJ})$ (Equation (16)) which are free energies calculated by TI, where the loop-water charges and Lennard-Jones interactions, respectively, are gradually eliminated; their sum is $F_{\text{water}}^{\text{TI}}(\text{loop})$. Correspondingly, $F_{\text{tmpl-water}}(\text{TI})$ are $F_{\text{water}}^{\text{TI}}(\text{ch}, \text{tmpl})$ and $F_{\text{water}}^{\text{TI}}(\text{LJ}, \text{tmpl})$ [Equation (17)] which are free energies calculated by TI, where the template-water charges and Lennard-Jones interactions, respectively, are gradually eliminated; their sum is $F_{\text{water}}^{\text{TI}}(\text{tmpl})$. $F_{\text{sum}} = E_{\text{loop}} + F_{\text{water}}^{\text{TI}}(\text{loop}) + F_{\text{water}}^{\text{TI}}(\text{tmpl})$ is the total free energy (without the contribution of the loop entropy, S_{loop}). Δ is the difference, free bound. n_s is the sample size. Statistical errors, $s/(n_s)^{1/2}$, where s is the standard deviation, are given only for results based of a number of water molecules, $N_{\text{water}} = 140$ and 160; the errors for $N_{\text{water}} < 140$ are smaller than those obtained for $N_{\text{water}} = 140$ and 160.

4. Summary and Conclusions

This is an additional paper in a series of papers where our HSMD-TI method has been extended systematically for mobile loops in proteins. The method provides the absolute entropy and free energy and thus enables one calculating these quantities for microstates *m* and *n* with a large conformational variance without the need to carry out a complex thermodynamic integration from *m* to *n*. Mobile loops in enzymes can be crucial for the extent of binding, where their open conformation is moved to form a "lid" over the bound active site.

Thus far we have studied the 7-residue mobile loop, 304–310 (Gly-His-Gly-Ala-Gly-Gly-Ser) of the enzyme porcine pancreatic α-amylase, where the system was modeled by the AMBER96 force field alone and by AMBER96 with the GB/SA implicit solvent [8]. Later the implicit solvent was replaced by explicit solvent, *i.e.*, the α-amylase loop was capped with 70 TIP3P [9] water molecules. To determine the minimum size of the template and the minimum number of water molecules required to cap a loop, we have recently carried out a systematic study of the loop, 287–290 (Ile, Phe, Arg, and Phe) of AChE [10]. The conclusion is that to recover the experimental free energy difference the water density should be close to that of bulk water, and the minimum template size is of 1,000 atoms for this loop. In all these studies and in a recent study of a mobile loop of the protein streptavidin [62] the free and bound loop structures were attached to a common template that was "cut" from the crystal structure of the free loop. Obviously, one would seek to treat each loop attached to its own template. This has been the objective of the present paper.

As a first step of this study the heavy atoms of the two crystal structures were harmonically restrained (with a force constant of 1 kcal/Å^2) and the energy was minimized. Afterwards the loop and (TIP3P) water atoms were allowed to relax in the presence of a fixed template. The relatively weak force constant (1 kcal/Å^2) (as compared to 100 kcal/Å^2 used in [10] for the pdb structure 2ace) was chosen to allow more extensive changes in the bound structure due to its relatively large B-factors and the chemical change imposed on it by removing the phosphate group; correspondingly, larger structural changes were also allowed to occur in the free structure. Then two cuts from each crystal structures were performed, which contain exactly the same atoms; these templates and the attached loops were subjected to further optimization steps as described earlier.

The new computational component in this study is the annihilation (for each template) of the water-template interactions using a TI procedure. Unlike the loop-water TI, one would expect the accuracy of the water–template TI to depend strongly on the system size. Thus, for $N_{water} \leq 140$ the errors are relatively small and the results for the free-bound differences in the total free energy and energy, ΔF_{sum} and ΔE_{total} are comparable, while for $N_{water} = 160$ the errors become larger (in particular for the Lennard-Jones part of the water-template TI) and a significant increase in ΔF_{sum} and ΔE_{total} occurs. We attribute this jump in the results to changes in the configuration of water stemming from the increase in water pressure in the sphere as the number of waters grows.

Finally, calculation of ΔF_{loop} based on separate templates which are cut from the free and bound crystal structures is more realistic than using a single common template for both loop conformations. ΔF_{loop} should also be considered in the calculation of the absolute free energy of binding when the conformational change of the loop is due to ligand binding is of an induced fit type. Thus, while the present calculations show that recovering the experimental result, $\Delta F_{loop} = F_{free} - F_{bound} \sim -4$ kcal/mol

is not straightforward, to achieve improvement more research is needed, which will be based on the conclusions of the initial work carried out here, as specified in the summary of the previous section.

5. Appendix

5.1. Thermodynamic Integration of Water

As described earlier, two TI processes are carried out where the first is applied to the loop and water. Thus, 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*; the water-water and water-template potential energy is unchanged. After completing a loop-water TI (*i.e.*, the loop-water interactions are zero), a second TI process is performed where the water-template interactions (LJ and electrostatic) are decreased to zero; this TI process starts from the last water configuration obtained in the loop-water TI. For the (LJ) potential we have used the shifted scaling potential, introduced by Zacharias *et al*. [63]:

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

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:

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

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 ($\lambda = 1 \to \lambda = 0$) integration of the electrostatic interactions (*i.e.*, charge elimination) is carried out first (in the presence of intact LJ interactions) followed by a $\varepsilon = \lambda = 1 \to 0$ 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 (window) 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 step the production MD simulation, is of 20 ps in the loop-water TI, where a configuration is retained every 0.02 ps, *i.e.*, altogether 1,000 configurations are used for evaluating $<\partial F/\partial\lambda_i>$. For the template-water TI, the production simulation is longer, of 40 ps (2,000 configurations), due to the larger size of the template. It should be pointed out again that these two integrations are carried out for two different templates. The computer time is discussed in the Appendix of [10].

References and Notes

1. White, R.P.; Meirovitch, H. A simulation method for calculating the absolute entropy and free energy of fluids: Application to liquid argon and water. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9235-9240.
2. Cheluvaraja, S.; Meirovitch, H. Simulation method for calculating the entropy and free energy of peptides and proteins. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9241-9246.
3. White, R.P.; Meirovitch, H. Lower and upper bounds for the absolute free energy by the hypothetical scanning Monte Carlo Method: Application to liquid argon and water. *J. Chem. Phys.* **2004**, *121*, 10889-10904.
4. White, R.P.; Meirovitch, H. Free volume hypothetical scanning molecular dynamics method for the absolute free energy of liquids. *J. Chem. Phys.* **2006**, *124*, 204108-204113.
5. White, R.P.; Meirovitch, H. Calculation of the entropy of random coil polymers with the hypothetical scanning Monte Carlo Method. *J. Chem. Phys.* **2005**, *123*, 214908-214911.
6. Cheluvaraja, S.; Meirovitch, H. Calculation of the entropy and free energy by the hypothetical scanning Monte Carlo Method: Application to peptides. *J. Chem. Phys.* **2005**, *122*, 054903-054913.
7. Cheluvaraja, S.; Meirovitch, H. Calculation of the entropy and free energy of peptides by molecular dynamics simulations using the hypothetical scanning molecular dynamics method. *J. Chem. Phys.* **2006**, *125*, 024905-024913.
8. Cheluvaraja, S.; Meirovitch, H. Stability of the free and bound microstates of a mobile loop of α amylase obtained from the absolute entropy and free energy. *J. Chem. Theory Comput.* **2008**, *4*, 192-208.
9. Cheluvaraja, S.; Mihailescu, M.; Meirovitch, H. Entropy and free energy of a mobile loop in explicit water. *J. Phys, Chem. B* **2008**, *112*, 9512-9522.
10. Mihailescu, M.; Meirovitch, H. Absolute free energy and entropy of a mobile loop of the enzyme Acetyl Choline Esterase. *J. Phys. Chem. B* **2009**, *113*, 7950-7964.
11. Beveridge, D.L.; DiCapua, F.M. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Chem.* **1989**, *18*, 431-492.
12. Kollman, P.A. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395-2417.
13. Jorgensen, W.L. Free energy calculations: a breakthrough for modeling organic chemistry in solution. *Acc. Chem. Res*. **1989**, *22*, 184-189.
14. Meirovitch, H. Calculation of the free energy and entropy of macromolecular systems by computer simulation. *Rev. Compu.t Chem*. **1998**, *12*, 1-74.
15. Gilson, M.K.; Given, J.A., Bush, B.L.; McCammon, J.A. The statistical thermodynamic basis for computing of binding affinities: A critical review. *Biophys. J.* **1997**, *72*, 1047-1069.
16. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A qualitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535-9551.
17. Meirovitch, H. Recent developments in methodologies for calculating entropy and free energy of biological systems by computer simulation. *Curr. Opin. Struct. Biol.* **2007**, *17*, 181-186.

18. Gilson, M.K; Zhou, H.X. Calculation of protein-ligand binding affinities. *Ann. Rev. Biophys. Biomol. Struct.* **2007**, *36,* 21-42.

19. Foloppe, N.; Hubbard, R. Towards predictive ligand design with free-energy based computational methods? *Curr. Med. Chem.* **2007**, *13*, 3583-3608.

20. 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. Biomolecular modeling: Goals, Problems, Perspectives. *Angew. Chem. Int. Ed.* **2006**, *45*, 4064-4092.

21. Singh, N.; Warshel, A. A comprehensive examination of the contributions to the binding entropy of protein-ligand complexes. *Proteins* **2010**, *78*, 1724-1735.

22. Deng, Y.; Roux, B. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B* **2009**, *113,* 2234-2246.

23. Warshel, A; Sharma, P.K.; Kato, M.; Parson, W.W. Modeling electrostatic effects in proteins. *Biochim. Biophys. Acta.* **2006**, *1764*, 1647-1676.

24. Alder, B.J.; Wainwright, T.E. Studies of molecular dynamics. I. General method. *J. Chem. Phys.* **1959**, *31*, 459-466.

25. McCammon, J.A.; Gelin, B.R.; Karplus, M. Dynamics of folded proteins. *Nature* **1977**, *267*, 585-590.

26. Elber, R.; Karplus, M. Multiple conformational states of proteins a molecular dynamics analysis of myoglobin. *Science* **1987**, *235*, 318-321.

27. Stillinger, F.H.; Weber, T.A. Packing structures and transitions in liquids and solids. *Science*, **1984**, *225*, 983-989.

28. Gō, N.; Scheraga, H.A. Analysis of the contribution of internal vibrations to the statistical weights of equilibrium conformations of macromolecules. *J. Chem. Phys.* **1969**, *51*, 4751-4767.

29. Gō, N.; Scheraga, H.A. On the use of classical statistical mechanics in the treatment of polymer chain conformation. *Macromolecules* **1976**, *9*, 535-542.

30. Hagler, A.T.; Stern, P.S.; Sharon, R.; Becker, J.M.; Naider, F. Computer simulation of the conformational properties of oligopeptides. Comparison of theoretical methods and analysis of experimental results. *J. Am. Chem. Soc.* **1979**, *101*, 6842-6852.

31. Messer, B.M.; Roca, M.; Chu, Z.T.; Vicatos, S.; Kilshtain, A.V.; Warshel, A. Multiscale simulations of protein landscapes: Using coarse-grained models as reference potentials to full explicit models. *Proteins* **2010**, *78*, 1212-1227.

32. Karplus, M.; Kushick, J.N. Method for estimating the configurational entropy of macromolecules. *Macromolecules* **1981**, *14*, 325-332.

33. Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* **1993**, *215*, 617-621.

34. Andricioaei, I.; Kaplus, M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* **2001**, *115*, 6289-6292.

35. 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. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

36. Qiu, D.; Shenkin, P.S.; Hollinger, F.P.; Still, W.C. The GB/SA continuum model for salvation: A fast analytical method for the calculation approximate Born radii. *J. Phys. Chem.* **1997**, *101*, 3005-3014.

37. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.

38. White, R.P.; Meirovitch, H. Minimalist explicit solvation models for surface loops in proteins. *J. Chem. Theory Comput.* **2006**, *2*, 1135-1151.

39. Steinbach, P.J.; Brooks, B.R. Protein hydration elucidated by molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA.* **1993**, *90*, 9135-9139.

40. 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. Crystal Structures of aged phosphonylated acetylcholinesterase: Nerve agent reaction products at the atomic level. *Biochemistry* **1999**, *38*, 7032-7039.

41. Millard, C.B.; Koellner, G.; Ordentlich, A.; Shafferman, A.; Silman, I.; Sussman, J.L. Reaction products of acetylcholinesterase and VX reveal a mobile histidine in the catalytic triad. *J. Am. Chem. Soc.* **1999**, *121*, 9883-9884.

42. Chiu, Y.C.; Main, A.R.; Dauterman, W.C. Affinity and phosphorylat1on constants of a series of *O,O*-dialkyl malaoxons and paraoxons with acetylcholinesterase. *Biochem. Pharmacol.* **1969**, *18*, 2171-2177.

43. Hart, G.J., O'Brien, R.D. Recording spectrophotometric method for determination of dissociation and phosphorylation constants for the inhibition of acetylcholineesterase by organophosphates in the presence of substrate. *Biochemistry* **1973**, *12*, 2940-2945.

44. Forsberg, A.; Puu, G. Kinetics for the inhibition of acetylcholinesterase from the electric eel by some organophosphates and carbamates. *Eur. J. Biochem.* **1984**, *140*, 153-156.

45. Ordentlich, A.; Barak, D.; Kronman, C.; Flashner, Y.; Leitner, M.; Segall, Y.; Ariel, N.; Cohen, S.; Velan, B.; Shafferman, A. Dissection of the human acetylcholinesterase active center determinants of substrate specificity. Identification of residues constituting the anionic site, the hydrophobic site, and the acyl pocket. *J. Biol. Chem.* **1993**, *268*, 17083-17095.

46. Ordentlich, A.; Kronman, C.; Barak, D.; Stein, D.; Ariel, N.; Marcus, D.; Velan, B.; Shafferman, A. Engineering resistance to 'aging' of phosphylated human. Acetylcholinesterase. role of hydrogen bond network in the active center. *FEBS. Lett.* **1993**, *334*, 215-220.

47. Raves, M.L.; Harel, M.; Pang, Y.P.; Silman, I.; Kozikowski, A.P.; Sussman, J.L. Crystal Structures of aged phosphonylated acetylcholinesterase: Nerve agent reaction products at the atomic level. *Nat. Struct. Biol.* **1997**, *4*, 57-63.

48. Carlacci, L; Millard, C.B.; Olson, M.A. Conformational energy landscape of the acyl pocket loop in acetylcholinesterase: A Monte Carlo-generalized Born model study. *Biophys. Chem.* **2004**, *111*, 143-157.

49. Olson, M.A. Modeling loop reorganization free energies of acetylcholinesterase: A comparison model of explicit and implicit solvent models. *Proteins* **2004**, *57*, 645-650.

50. Getzoff, E.D.; Geysen, H.M.; Rodda, S.J.; Alexander, H.; Tainer, J.A. Lerner, R.A. Mechanisms of antibody binding to a protein. *Science* **1987**, *235*, 1191-1196.

51. Rini, J.M.; Schulze-Gahmen, U.; Wilson, I.A. Structural evidence for induced fit as a mechanism for antibody- antigen recognition. *Science* **1992**, *255*, 959-965.

52. Constantine, K.L.; Friedrichs, M.S.; Wittekind, M.; Jamil, H.; Chu, C.H.; Parker, R.A.; Goldfarb, V.; Mueller, L.; Farmer, B.T. Backbone and side chain dynamics of uncomplexed human adipocyte and muscle fatty acid-binding proteins. *Biochemistry* **1998**, *37*, 7965-7980.

53. Cansu, S; Doruker, P. Dimerization affects collective dynamics of triosphosphate isomerase. *Biochemistry* **2008**, *47*, 1358-1368.

54. General, I.J; Dragomirova, R; Meirovitch, H. Calculation of the Absolute free energy of binding: the effect of a mobile loop on the avidin/biotin complex. *J. Phys. Chem. B*. **2010**, in review.

55. 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, USA, 2001.

56. Meirovitch, H.; Alexandrowicz, Z. On the zero fluctuation of the microscopic free energy and its potential use. *J. Stat. Phys.* **1976**, *15*, 123-127.

57. Meirovitch, H. Simulation of a free energy upper bound, based on the anti-correlation between an approximate free energy functional and its fluctuation. *J. Chem. Phys*. **1999**, *111*, 7215-7224.

58. Meirovitch, H. The scanning simulation method for macromolecules. *Int. J. Mod. Phys.* **1990**, *1*, 119-145.

59. Meirovitch, H. Computer simulation of the free energy of polymer chains with excluded volume and with finite interactions. *Phys. Rev. A* **1985**, *32*, 3709-3715.

60. Allen, M.P.; Tildesley, D.J. *Computer Simulation of Liquids*; Clarenden Press: Oxford, UK, 1987.

61. Chang, C.E.; Chen, W.; Gilson, M.K. Evaluating the accuracy of the quasiharmonic approximation. *J. Chem. Theory. Comput.* **2005**, *1*, 1017-1028.

62. General, I.J.; Meirovitch, H. Relative stability of the open and closed conformations of the active site loop of streptavidin. *J. Chem. Phys*. **2010**, in review.

63. Zacharias, M.; Straatsma, T.P.; McCammon, J.A. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys*. **1994**, *100*, 9025-9031.