# Calculation of the entropy and free energy of peptides by molecular dynamics simulations using the hypothetical scanning molecular dynamics method

Srinath Cheluvaraja and Hagai Meirovitch[a)]
*Department of Computational Biology, University of Pittsburgh School of Medicine, 3059 BST3, Pittsburgh, Pennsylvania 15260*

(Received 6 March 2006; accepted 5 May 2006)

Hypothetical scanning (HS) is a method for calculating the absolute entropy $S$ and free energy $F$ from a sample generated by any simulation technique. With this approach each sample configuration is reconstructed with the help of transition probabilities (TPs) and their product leads to the configuration's probability, hence to the entropy. Recently a new way for calculating the TPs by Monte Carlo (MC) simulations has been suggested, where all system interactions are taken into account. Therefore, this method—called HSMC—is in principle exact where the only approximation is due to insufficient sampling. HSMC has been applied very successfully to liquid argon, TIP3P water, self-avoiding walks on a lattice, and peptides. Because molecular dynamics (MD) is considered to be significantly more efficient than MC for a compact polymer chain, in this paper HSMC is extended to MD simulations as applied to peptides. Like before, we study decaglycine in vacuum but for the first time also a peptide with side chains, $(\mathrm{Val})_2(\mathrm{Gly})_6(\mathrm{Val})_2$. The transition from MC to MD requires implementing essential changes in the reconstruction process of HSMD. Results are calculated for three microstates, helix, extended, and hairpin. HSMD leads to very stable *differences* in entropy $T\Delta S$ between these microstates with small errors of 0.1–0.2 kcal/mol ($T=100$ K) for a wide range of calculation parameters with extremely high efficiency. Various aspects of HSMD and plans for future work are discussed. © *2006 American Institute of Physics*. [DOI: 10.1063/1.2208608]

## I. INTRODUCTION

Calculation of the entropy $S$ and Helmholtz free energy $F$ ($F=E-TS$, where $E$ is the potential energy and $T$ is the absolute temperature) is of central interest in physics, chemistry, engineering, and biology.[1–5] $S$ is an essential thermodynamic property that constitutes a measure of order and is the main driving force in protein folding. The usual thermodynamic properties such as the pressure and the chemical potential can be derived from $F$,[6] which also serves as a criterion of stability, the lower is $F$ the higher the stability; this is, in particular, important in structural biology. The potential energy surface of peptides and proteins is rugged, i.e., "decorated" by a tremendous number of localized wells and "wider" ones, which are defined over regions $\Omega_m$ called microstates—each consists of many localized wells. A microstate can be obtained computationally by the local molecular dynamics[7,8] (MD) fluctuations around a structure (such as an $\alpha$ helix or a hairpin of a peptide). MD studies have shown that a molecule will visit a localized well only for a very short time (several femtaseconds) while staying for a much longer time within a microstate,[9,10] meaning that the microstates are of a greater physical significance than the localized wells. Thus, the aim of protein folding, for example, is to find the most stable microstate, i.e., that with lowest $F_m$.

However, flexible protein segments (e.g., surface loops), cyclic peptides, ligands bound to proteins, or side chains, can undergo *intermediate flexibility*, where they populate significantly several microstates $m$ in thermodynamic equilibrium. These populations $p_m$ are proportional to $\exp[-F_m/k_BT]$, where $F_m=-k_BT\ln Z_m=-k_BT\ln\int_m \exp[-E/k_BT]dx$, and $Z_m$ is the conformational partition function integrated over the microstate $\Omega_m$. It is of interest to know whether the conformational change adopted by a loop (a side chain, ligand, etc.) upon binding has been induced by the other protein (induced fit[11,12]) or alternatively the free loop interconverts among different microstates where one of them has been selected upon binding (selected fit[13]); this analysis requires calculating $p_m$, which is also needed for a correct analysis of NMR and x-ray data of macromolecules. Finally, the free energy determines the binding affinities of ligands interacting with active sites of enzymes, protein-protein interactions, and it is an important factor in enzymatic reactions.

While calculation of the absolute $F$ is difficult (due to the need to know the value of the sampling probability), in most cases (and in the examples discussed above) one is mostly interested in the ratio of populations $p_n/p_m=\exp-[\Delta F_{mn}/k_BT]$ between two microstates $m$ and $n$, which can be calculated in the most straightforward way by a *counting method*, i.e., from a long MD or Monte Carlo[14] (MC) simulation that "covers" both microstates. Thus, $\Delta F_{mn}=-k_BT\ln[(\#m)/(\#n)]$, where $\#m(\#n)$ is the population, i.e.,

a)Author to whom correspondence should be addressed. Electronic mail:hagaim@pitt.edu

 

the number of times the molecule visited microstate $m(n)$ during the simulation. Notice, however, that because of high energy barriers, the transition between microstates at room temperature might require long times, nanoseconds or more even for side chain rotamers, meaning that reliable sampling of $\#m(\#n)$ might become prohibitive. This problem can be alleviated by applying enhanced sampling techniques such as replica exchange[15] or multicanonical methods[16,17] (usually with principal component analysis); however, the conformational search capability of these methods is also limited and microstates of interest might be visited poorly or will not be visited at all.

Differences $\Delta S$ and $\Delta F$ are commonly calculated by thermodynamic integration (TI) over physical quantities such as the energy, temperature, and the specific heat,[18,19] as well as nonphysical parameters[1–5,20–27] (free energy perturbation methods, umbrella, and histogram analysis methods[28–30] are also included in this category). While this is a robust approach, if the structural variance of $m$ and $n$ is large (e.g., helical and hairpin states of a polypeptide) the integration from $m$ to $n$ becomes difficult and in many cases unfeasible.

Developing methods for calculating the *absolute* $F$ would remedy this problem to a large extent. Thus, one can carry out two separate long MD simulations of microstates $m$ and $n$ and calculating directly the absolute $F_m$ and $F_n$ and their difference $\Delta F_{mn}=F_m-F_n$ with high accuracy. Still, the absolute $F$ can also be obtained with TI provided that a reference state $r$ is available, where the free energy is known exactly and an efficient integration path between $r$ and $m$ (and $n$) can be defined. A classic example is the calculation of $F$ of liquid argon or water by integrating the free energy from an ideal gas reference state.[31,32] However, for nonhomogeneous systems such integration might not be trivial, and in models of peptides and proteins defining reference states that are close to the state of interest is a standing problem.[33–35] Furthermore, because MC (MD) simulations constitute models for dynamical processes, one would seek to calculate changes in $F$ and $S$ during a relaxation process, by assuming local equilibrium in certain parts along the trajectory; a classic example is simulation of protein folding.[36] Again, such information cannot be obtained by thermodynamic integration, and methods that estimate $S$ and $F$ directly from the trajectory of interest should be developed.

From the statistical mechanics point of view the absolute entropy (which leads to the absolute $F$) is related to the Boltzmann probability of system configuration $i$ $S\sim-\ln P_i^B$. However, the *value* of $P_i^B$ cannot be obtained in a straightforward manner from a MC or MD trajectory, therefore it has been commonly represented by a Gaussian[37–39] or a quasiharmonic approximation.[40,41]

Another approach for estimating the value of the sampling probability $P_i^B$ from a given sample has been suggested by Meirovitch. Two related techniques, the local states (LS) method[42–46] and the hypothetical scanning (HS) method,[31,47–50] were developed and applied to magnetic systems, polymers, fluids, and peptides. With this approach each sample configuration is reconstructed with the help of transition probabilities (TPs) and their product leads to the configuration's probability, hence to the entropy. Recently the HS has been further developed to a method called HSMC,[32,51] where the transition probabilities are calculated by MC simulations. HSMC takes into account *all* system interactions (i.e., short as well as long-range) and in this respect can be considered to be exact; the only approximation is due to insufficient MC sampling for calculating the TPs. This method provides rigorous upper and lower bounds for $F$, and $F$ can be obtained from a very small sample, even from a single conformation.

HSMC is a general technique that has been applied thus far very successfully to liquid argon,[32,51,52] TIP3P water,[51,52] peptides,[53–55] and self-avoiding walks on a lattice.[56,57] In particular, in Refs. 53 and 54 two models of polyglycine molecules of 10 and 16 residues, described by the AMBER force field[58] in vacuum were studied. One model is based on constant bond lengths and bond angles (the rigid model) and the other consists only of constant bond lengths (called there the flexible model). These models were simulated by MC in a helical, hairpin, and extended states and the corresponding $F_m$ and $S_m$ were calculated leading to very accurate results for $\Delta F_{m,n}=F_m-F_n$ ($\Delta S_{m,n}$), which are significantly better than those obtained with the LS and the quasiharmonic methods. In a subsequent paper[55] HSMC was applied to a model of decaglycine which is stretched by an external force.

With HSMC applied to a peptide, $S$ is calculated from a given MC sample by reconstructing each peptide conformation $i$ step by step, i.e., calculating successively a TP for each dihedral and bond angle along the chain and fixing the related atoms at their positions at $i$. Thus, at each step the chain's coordinates that have already been determined are kept fixed (the "frozen past") and the TP is obtained from a MC simulation of the "future" part of the chain whose TPs as yet have not been determined. It is important to verify that the simulated future part remains within the original microstate.

It is desirable to extend HSMC also to MD simulations. MD provides a model for dynamics and is considered to be a significantly more efficient method than MC for a compact polymer chain [notice, however, that Jorgensen and coworkers have been simulating protein-water systems efficiently with a MC procedure based on local conformational moves (e.g., see Refs. 59 and 60 and references cited therein)]; correspondingly, Hu *et al.* have shown recently that such procedures can be more efficient than MD at least for small peptides[61]). Thus, in this paper we extend HSMC to MD, where an essential part of the HSMD design is devoted for "harnessing" the simulated future chains to remain within the original microstate. HSMD is applied to decaglycine $(Gly)_{10}$ in the helix, extended, and hairpin microstates and the results are compared to those obtained with HSMC for the "flexible model" in our Ref. 54 (Table VI), which will be called Paper I throughout this article; also, to distinguish between the present (completely) flexible model of $(Gly)_{10}$ and the (partially) flexible model studied in Paper I (which is based on constant bond lengths) we call the latter "the flexible model I." We also study for the first time a peptide with side chains—$(Val)_2(Gly)_6(Val)_2$ in the helix and hairpin microstates.

## II. THEORY AND METHODOLOGY

### A. The peptides studied

We study two peptides, decaglycine $NH_2(Gly)_{10}CONH_2$ and $NH_2(Val)_2(Gly)_6(Val)_2CONH_2$, in vacuum defined by the AMBER96 force field,[58] where the charges of the end groups are neutralized. These models are simulated by MD in the helix, hairpin, and extended microstates. However, HSMC (as well as LS or the quasiharmonic method) is implemented naturally in internal coordinates; therefore the simulated conformations should be transferred from Cartesians to the dihedral angles $\varphi_i$, $\psi_i$, and $\omega_i$ and the bond angles $\theta_{i,l}$ ($i=1$, $N=10$, $l=1,3$); for the second molecule we also consider the four side chain angles $\chi_k$ of the four valine residues (in the next section we argue that to a good approximation bond stretching can be ignored). For convenience, these angles (ordered along the backbone) are denoted by $\alpha_k$, $k=1$, 60 (64).

### B. Statistical mechanics of a peptide in internal coordinates

The partition function of a peptide $Z$ is an integral over the function $\exp(-E/k_BT)$ ($E$ is the potential energy and $k_B$ is the Boltzmann constant) with respect to the Cartesian coordinates over the stable microstate $\Omega_0$ (e.g., a helical region). As has already been pointed out, to apply HSMC(D) one has to change the variables of integration from Cartesian to internal coordinates, which makes the integral dependent also on a Jacobian $J$. For a linear chain $J$ has been shown to be independent of the dihedral angles and it is a simple function of the bond angles and bond lengths.[37,38,40] For decaglycine the transformation from Cartesian to the internal coordinates, $\alpha_k$, $k=1$, $6N=60$ is applied under the assumption that the potentials of the bond lengths ("the hard variables") are strong and therefore their average values can be assigned to $J$, which to a good approximation can be taken out of the integral (however, see a later discussion). For the same reason one can carry out the integration over the bond lengths (assuming that they are not correlations with the $\alpha_k$) and the remaining integral becomes a function of the $6N$ dihedral and bond angles ($\alpha_k$) (Refs. 37, 38, and 40) and a Jacobian that depends only on the bond angles. The partition function becomes

$$Z' = DZ = D\int_{\Omega_0} \exp\{-[E([\alpha_k])]/k_BT\}d\alpha_1\cdots d\alpha_{6N}, \quad (1)$$

where $[\alpha_k]=[\alpha_1,\ldots,\alpha_{6N}]$. $D$ is a product of the integral over the bond lengths and their Jacobian $J$. The Jacobian $[\Pi_k \sin(\theta_k)]$ of the bond angles $\theta_k$ that should appear under the integral is omitted for simplicity. We *assume* $D$ to be the same (i.e., constant) for different microstates and therefore $\ln D$ cancels and can be ignored in calculations of free energy and entropy *differences*. The Boltzmann probability density corresponding to $Z$ [Eq. (1)] is

$$\rho^B([\alpha_k]) = \exp\{-[E([\alpha_k])]/k_BT\}/Z, \quad (2)$$

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

$$S = -k_B\int_{\Omega_0} \rho^B([\alpha_k])\ln \rho^B([\alpha_k])d\alpha_1\cdots\alpha_{6N} \quad (3)$$

and

$$F = \int_{\Omega_0} \rho^B([\alpha_k])[E([\alpha_k]) + k_BT\ln\rho^B([\alpha_k])]d\alpha_1\cdots\alpha_{6N}. \quad (4)$$

As discussed earlier in applications of HS, LS, and HSMC, the fluctuation of the exact $F$ is zero,[62] because the integrand, $E([\alpha_k])+k_BT\ln\rho^B([\alpha_k])=-kT\ln Z=F$, is constant and equal to $F$ for any set $[\alpha_k]$. This means that the free energy can be obtained from *any single* conformation if its Boltzmann probability density is known. Using the HSMC(D) method, it is possible to estimate the free energy of the system from any single structure. Notice that the fluctuation of an approximate free energy (i.e., based on an approximate probability density) is finite and it is expected to decrease as the approximation improves.[31,32,49,50,52,53,62]

It should be pointed out that in our previous implementation of HSMC the peptides were modeled by internal coordinates (rather than Cartesian coordinates) where the bond lengths were kept constant, and thus the energy and entropy of bond stretching were ignored (correspondingly, the MC variables were the dihedral and bond angles). With MD on the other hand, the bond stretching energy is taken into account in Eq. (4) (and in free energy functionals defined later) while the corresponding entropy is ignored. The contribution of this energy to the free energy becomes an additive constant if one accepts the assumptions about the stretching energy and the corresponding Jacobian made prior to Eq. (4). This is a very good approximation; however, if the bond stretching entropy should be considered, we argue later that it can be estimated approximately within the framework of HSMD by assuming that bond stretching is independent of the other interactions.

### C. Exact scanning procedure

The HSMC(D) method is based on the ideas of the *exact* scanning method, which is a step-by-step construction procedure for a peptide.[63,64] Thus, an $N$-residue conformation of polyglycine in the helical region ($\Omega_0$), for example, is built (using internal coordinates) by defining the angles $\alpha_k$ step by step with TPs and adding the related atoms;[64] for example, the angle $\varphi$ determines the coordinates of the two hydrogens connected to $C^\alpha$, and the position of $C'$. Thus, at step $k$, $k-1$ angles $\alpha_1,\ldots,\alpha_{k-1}$ have already been determined; these angles and the related structure (the past) are kept constant, and $\alpha_k$ is defined with the exact TP density $\rho(\alpha_k|\alpha_{k-1}\ldots\alpha_1)$,

$$\rho(\alpha_k|\alpha_{k-1}\cdots\alpha_1) = Z_{\text{future}}(\alpha_k\cdots\alpha_1)/[Z_{\text{future}}(\alpha_{k-1}\cdots\alpha_1)d\alpha_k], \quad (5)$$

where $d\alpha_k$ is a small segment centered at $\alpha_k$ and $Z_{\text{future}}(\alpha_k\ldots\alpha_1)$ is a future partition function defined over the helical region $\Omega_0$ by integrating over the future conformations defined by $\alpha_{k+1}\ldots d\alpha_{6N}$ (within $\Omega_0$) where the past angles, $\alpha_1\ldots\alpha_k$, are held fixed,

$$Z_{\text{future}}(\alpha_k, \ldots, \alpha_1) = \int_{\Omega_0} \exp - [(E(\alpha_{6N}, \ldots, \alpha_1)/k_B T\ ]d\alpha_{k+1} \cdots d\alpha_{6N}. \quad (6)$$

The product of the TPs [Eq. (5)] leads to the probability density of the entire conformation [Eq. (2)],

$$\rho^B(\alpha_{6N}, \ldots, \alpha_1) = \prod_{k=1}^{6N} \rho(\alpha_k|\alpha_{k-1} \cdots \alpha_1). \quad (7)$$

This construction procedure is not feasible for a large molecule and in practice can be carried out by scanning only a limited number of future angles;[63,64] however, the ideas of the exact scanning method constitute the basis for HSMC(D), as discussed below.

Thus, the exact scanning method is equivalent to MC and MD in the sense that large samples generated by all these methods lead to the same averages and fluctuations within the statistical errors. 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 by calculating the TP densities that *hypothetically* were used to create it step by step. This idea has been implemented initially in two different ways, by the LS and HS methods. However, an exact reconstruction of the TPs [Eq. (5)] is feasible only for a very small peptide. Therefore, calculation of future partition functions [Eq. (6)] by these methods has been carried out only approximately, by considering a partial future (or a limited past in the case of LS). As will be described later, with HSMC(D) the *entire* future is considered and in this respect the method can be considered to be exact.

### D. The HSMC method in internal coordinates

It would be beneficial to describe first the HSMC method in internal coordinates that has been developed in previous publications. In the first step the MC sample to be analyzed (of a given microstate) is visited and the variability range $\Delta\alpha_k$ is calculated, where $\alpha_k$ are the dihedral and bond angles, $1 \leqslant \alpha_k \leqslant 6N$,

$$\Delta\alpha_k = \alpha_k(\max) - \alpha_k(\min), \quad (8)$$

where $\alpha_k(\max)$ and $\alpha_k(\min)$ are the maximum and minimum values of $\alpha_k$ found in the sample, respectively. $\Delta\alpha_k$, $\alpha_k(\max)$, and $\alpha_k(\min)$ enable one to verify that the sample spans correctly its microstate and they help keeping the future chains within the limits of the microstate during the MC simulations as discussed below.

As mentioned in Sec. II C, the idea of the HS method is to reconstruct each sample conformation step by step obtaining the TP density of each $\alpha_k$ [Eq. (5)] by calculating the future partition functions $Z_{\text{future}}$ [Eq. (6)]. However, a systematic integration of $Z_{\text{future}}$ based on the entire future within the limits of $\Omega_0$ is difficult and becomes impractical for a large peptide where $\Omega_0$ is unknown. The idea of the HSMC method is to obtain the TPs [Eq. (5)] by carrying out MC simulations of the future part of the chain rather than by evaluating the integrals defining $Z_{\text{future}}$ [Eq. (6)] in a systematic deterministic way. Thus, at reconstruction step $k$ of conformation $i$ the TP density, $\rho(\alpha_k|\alpha_{k-1}\ldots\alpha_1)$, is calculated from $n_f$ MC steps (trials), where the entire future of the peptide can move by changing the future angles $\alpha_k, \ldots, \alpha_{6N}$ while the angles $\alpha_1, \ldots, \alpha_{k-1}$ and their related atoms (defining the past) are kept fixed at their values in conformation $i$. A small segment (bin) $\delta\alpha_k$ [see also Eq. (5)] is centered at $\alpha_k$ and the number of MC visits to this bin, $n_{\text{visit}}$, during the simulation is calculated; one obtains

$$\rho(\alpha_k|\alpha_{k-1} \cdots \alpha_1) \approx \rho^{\text{HS}}(\alpha_k|\alpha_{k-1} \cdots \alpha_1) = n_{\text{visit}}/[n_f \delta\alpha_k], \quad (9)$$

where the relation becomes exact for very large $n_f (n_f \to \infty)$ and a very small bin $(\delta\alpha_k \to 0)$ [see discussion in Paper I (Ref. 54)]. This means that in practice $\rho^{\text{HS}}(\alpha_k|\alpha_{k-1}\ldots\alpha_1)$ will be somewhat approximate due to insufficient future sampling (finite $n_f$), a relatively large bin size $\delta\alpha_k$, an imperfect random number generator, etc.; therefore, we denote this TP by HS (rather than by HSMC—for the sake of brevity). Notice that unlike the deterministic calculation of $Z_{\text{future}}$ [Eq. (6)] where the limits of $\Omega_0$ are in practice unknown, with HSMC the future structures generated by MC at each step $k$ remain in general within the limits of the microstate $\Omega_0$ defined by the analyzed MC sample. In some cases, however, the future samples might escape from this region; therefore, the $\alpha_k(\min)$ and $\alpha_k(\max)$ values [Eq. (8)] are used to keep the future structures within $\Omega_0$ by rejecting MC moves with angle values beyond those of $\alpha_k(\min)$ and $\alpha_k(\max)$. The corresponding probability density is

$$\rho^{\text{HS}}(\alpha_{6N}, \ldots, \alpha_1) = \prod_{k=1}^{6N} \rho^{\text{HS}}(\alpha_k|\alpha_{k-1} \cdots \alpha_1). \quad (10)$$

$\rho^{\text{HS}}([\alpha_k])$ defines approximate entropy and free energy functionals, $S^A$ and $F^A$, respectively,

$$S^A = -k_B \int \rho^B \ln \rho^{\text{HS}}([\alpha_k])d\alpha_1 \cdots \alpha_{6N}, \quad (11)$$

$$F^A = \langle E\rangle - TS^A = \langle E\rangle + k_B T \int \rho^B[\ln \rho^{\text{HS}}([\alpha_k])]d\alpha_1 \cdots \alpha_{6N}, \quad (12)$$

where $\langle E\rangle$ is the Boltzmann average of the potential (force field) energy estimated from the MC (or MD) sample and $\rho^B$ [Eq. (2)] is the Boltzmann probability density with which the sample has been generated. $S^A$ is estimated from a Boltzmann sample of size $n$ by the arithmetic average of the $\ln(\rho^{\text{HS}})$ values. As discussed in Sec. II B, the fluctuation (standard deviation) $\rho_F$ of the correct free energy is zero, while the approximate $F^A$ has finite fluctuation, $\sigma_A$ (estimated by its arithmetic average, $\overline{\sigma_A}$), which is expected to decrease as the approximation improves,[31,32,49,50,52,53,62]

$$\overline{\sigma_A} = \left[\frac{1}{n}\sum_{t=1}^{n} [\bar{F}^A - E_t - k_B T \ln \rho_t^{\text{HS}}]^2\right]^{1/2}. \quad (13)$$

$S^A$ and $F^A$ are expected to overestimate and underestimate, respectively, the correct values, where the fluctuation of $F^A$, $\sigma_A$ [Eq. (13)], does not vanish, but decreases as the approximation improves, i.e., as $n_f$ increases and/or $\delta\alpha_k$ decreases.

### E. The HSMD method

Unlike HSMC, HSMD is applied to a sample generated by MD. To verify that the sample conformations remain within the microstate $\Omega_0$ of interest (e.g., a helix) each of them is expressed in internal coordinates, $\alpha_k$'s [Eq. (8)]. Equation (9) can be used also with MD, where at step $k$ of the reconstruction procedure an MD simulation of the future chain starts from the reconstructed conformation $i$, and every $l$ fs the current conformation is considered; thus, the initial conformations generated are ignored for equilibration and the next $n_f$ future conformations are expressed in internal coordinates and their contribution to $n_{\text{visit}}$ [Eq. (9)] is calculated.

However, as with the MC implementation, an essential issue is to keep the future chains within the limits of the microstate $\Omega_0$—a condition that might be violated for large $n_f$; therefore, the above procedure has been changed by dividing it into several ($m$) shorter repetitive procedures ("units"), each based on $n_f' < n_f$ conformations where $n_f = mn_f'$, and each unit starts from the reconstructed structure $i$ with a different set of velocities; the unit size $n_f'$ (and the equilibration length) should be correctly chosen that it is small enough to keep the future chain within the microstate but allow an adequate sampling of this microstate; a similar procedure was first suggested by Brady and Karplus[65] within the framework of the quasiharmonic method and was also used in implementations of the LS method to peptides.[66,67] Another practical change from the HSMC implementation is the need to treat a pair of angles simultaneously, where each pair consists of a dihedral angle and its successive bond angle (e.g., $\varphi$ and the bond angle $N-C^{\alpha}-C'$). Thus, at each step both $\alpha_k$ and $\alpha_{k+1}$ are considered and each must be located within the limits of $\delta\alpha_k$ and $\delta\alpha_{k+1}$, respectively, in order to increase $n_{\text{visit}}$ by 1.

This MD implementation is based on three parameters, $\delta\alpha_k$, $n_f'$ and $m$, while only two parameters are needed for the MC implementation. The unit $n_f'$ should be adjusted where it can be increased as the microstate's stability increases. An adequate $n_f'$ should lead to smaller entropy as $m$ is increased or $\delta\alpha_k$ is decreased. In general one would attempt to apply the largest $n_f'$ that still satisfies these requirements. From now on we shall replace $n_f'$ by the word unit.

It should be pointed out again that in the case of HSMD $F^A$ includes the bond stretching energy while the corresponding entropy is ignored. However, under the assumptions leading to Eq. (1) this is not expected to affect differences in free energy which are our main interest. Still, if one seeks to include the bond stretching entropy, one can use a transition probability density $\rho(a_k)$ similar to Eq. (9) for the bond length $a_k$ which corresponds to the pair of atoms $k$ and $k+1$; considering the Jacobian, one obtains $\rho(a_k) \approx n_{\text{visit}}/[n_f 3^{-1}\delta(a_k^3)]$, where $\delta a_k$ is small compared to $a_k$. In this *approximation* the bond stretching is independent of the other interactions and thus $\rho_{\text{TP}}^{\text{HS}} = \rho^{\text{HS}}(\alpha_k|\alpha_{k-1}\ldots\alpha_1)\rho(a_k)$. Both probability densities can be calculated simultaneously, which in practice would not increase computer time.

### F. Upper bounds for the free energy

In addition to $F^A(\rho^{\text{HS}}([\alpha_k]))$ [Eq. (12)], which in practice is a lower bound, one can define another approximate free energy functional denoted $F^B$,[48]

$$F^B = \int_{\Omega_0} \rho^{\text{HS}}([\alpha_k])[E + k_B T \ln \rho^{\text{HS}}([\alpha_k])]d\alpha_1 \cdots d\alpha_{6N}. \tag{14}$$

According to the free energy minimum principle,[68] $F^B \geqslant F$ [Eq. (4)]. Thus, $F^B$ is an upper bound which approaches the correct free energy $F$ when $\rho^{\text{HS}} \to \rho^B$ [Eq. (2)]. It is necessary to rewrite Eq. (14) such that $F^B$ can be estimated by importance sampling from a (Boltzmann) sample of configurations generated with $\rho^B$ (rather than $\rho^{\text{HS}}$). It has been shown that

$$F^B = \frac{\int_{\Omega_0} \rho^B [\rho^{\text{HS}} \exp[E/k_B T](E + k_B T \ln \rho^{\text{HS}})]d\alpha_1 \cdots d\alpha_{6N}}{\int_{\Omega_0} \rho^B [\rho^{\text{HS}} \exp[E/k_B T]]d\alpha_1 \cdots d\alpha_{6N}}. \tag{15}$$

In practice $F^B$ is estimated as the ratio of simple arithmetic averages, which are accumulated for each of the quantities in the brackets in Eq. (15). It should be noted, however, that the statistical reliability of this estimation (unlike the estimation of $F^A$) decreases sharply with increasing system size, because the overlap between the probability distributions $\rho^B$ and $\rho^{\text{HS}}$ decreases exponentially [see discussion in Ref. 45]. With values for both $F^A$ and $F^B$, their average $F^M$ defined by

$$F^M = (F^A + F^B)/2, \tag{16}$$

often becomes a better approximation than either of them individually. This is provided that their deviations from $F$ (in magnitude) are approximately equal, and that the statistical error in $F^B$ is not too large. Typically, several improving approximations for $F^A$, $F^B$, and $F^M$ are calculated and their convergence enables one to determine the correct free energy with high accuracy.

It should be pointed out that the probability distribution defined by HSMC is stochastic as compared to the deterministic distribution (for a given sample) obtained by the LS method and the deterministic HS method. In Ref. 51 it is proved that the inequalities $F^A \leqslant F \leqslant F^B$ hold for the stochastic probabilities as well.

These conclusions hold also for HSMD provided that the assumptions leading to Eq. (1) are valid. In this case $F^B$ (like $F^A$) will be increased by an additive constant (contributed by the bond stretching energy) which will be canceled out in free energy differences of microstates. Because $E/k_BT + \ln \rho^{\text{HS}}$ is exponentiated in both the numerator and denominator of Eq. (15), if deviations from these assumptions occur, they will affect $F^B$ more significantly than $F^A$ and to observe the expected behavior of $F^B$ one might need to consider the bond stretching entropy as well.

### G. Exact expression for the free energy

As shown for fluids in Ref. 51, the denominator of $F^B$ in Eq. (15) defines an exact expression for the partition function,

$$\frac{1}{Z} = \frac{1}{Z}\int_{\Omega_0} \rho^B(\rho^{\mathrm{HS}}/\rho^B)[d\alpha_k]$$

$$= \int_{\Omega_0} \rho^B(\rho^{\mathrm{HS}} \exp[E/k_BT])[d\alpha_k]$$

$$= \int_{\Omega_0} \rho^B \exp[F^{\mathrm{HS}}/k_BT][d\alpha_k], \quad (17)$$

and an exact expression for the correct free energy $F$ denoted by $F^D$ is

$$F^D = k_BT \ln\left(\frac{1}{Z}\right) = k_BT \ln\left[\int_{\Omega_0} \rho^B \exp[F^{\mathrm{HS}}/k_BT][d\alpha_k]\right], \quad (18)$$

where $[d\alpha_k]=d\alpha_1 \ldots d\alpha_{6N}$ and $F^{\mathrm{HS}}/k_BT=(E[\alpha_k])/k_BT + \ln \rho^{\mathrm{HS}}[\alpha_k]$.

In practice, the efficiency of estimating $F$ by $F^D$ depends on the fluctuation of this statistical average, which is determined by the fluctuation of $F^{\mathrm{HS}}$ exponentiated. Obviously, as $F^{\mathrm{HS}} \to F$ (i.e., $\rho^H \to \rho^B$) all fluctuations become zero and $F$ can be obtained from a single configuration [see discussion following Eq. (4) and Ref. 51]. Therefore (as for $F^B$), the direct calculation of $F$ through $F^D$ will not be as statistically reliable as the corresponding calculation for the lower bound estimate, $F^A$, however, $F^D$ is expected to be more statistically reliable than $F^B$ which is defined as a ratio of two summations similar to that defining $F^D$. These conclusions hold also for HSMD provided that the assumptions leading to Eq. (1) are correct. The discussion in the preceding section II F regarding $F^B$ applies also to $F^D$.

### H. The local states method

We compare our results to those obtained by the LS method. With this method the ranges $\Delta\alpha_k$ [Eq. (8)] are divided into $l$ equal segments, where $l$ is the discretization parameter. We denote these segments by $\nu_k$, $(\nu_k=1,l)$. Thus, an angle $\alpha_k$ is now represented by the segment $\nu_k$ to which it belongs and a conformation $i$ is expressed by the corresponding vector of segments $[\nu_1(i), \nu_2(i), \ldots, \nu_{6N}(i)]$. Under this discretization approximation $\rho(\alpha_k|\alpha_{k-1}\ldots\alpha_1)$ can be estimated by

$$\rho(\alpha_k|\alpha_{k-1}\cdots\alpha_1) \approx n(\nu_k, \ldots, \nu_1)/\{n(\nu_{k-1}, \ldots, \nu_1)[\Delta\alpha_k/l]\}, \quad (19)$$

where $n(\nu_k, \ldots, \nu_1)$ is the number of times the *local states* [i.e., the partial vector $(\nu_k, \ldots, \nu_1)$ representing $(\alpha_k, \ldots, \alpha_1)$] appears in the sample. Because the number of local states increases exponentially with $k$ one has to resort to approximations based on smaller local states that consists of $\nu_k$ and the $b$ angles preceding it along the chain, i.e., the vector $(\nu_k, \nu_{k-1}, \ldots, \nu_{k-b})$, where $b$ is the correlation parameter. The sample is visited for the second time and for a given $b$ one calculates the number of occurrences $n(\nu_k, \nu_{k-1}, \ldots, \nu_{k-b})$ of all the local states from which a set of transition probabilities $p(\nu_k|\nu_{k-1}, \ldots, \nu_{k-b})$ are defined. The sample is then visited for the third time and for each member $i$ of the sample one determines the $6N$ local states and the corresponding transition probabilities, whose product defines an *approximate* probability density $\rho_i(b,l)$ for conformation $i$,

$$\rho_i(b,l) = \prod_{k=1}^{6N} p(\nu_k|\nu_{k-1}, \ldots, \nu_{k-b})/(\Delta\alpha_k/l), \quad (20)$$

the larger are $b$ and $l$ the better the approximation (for enough statistics). $\rho_i(b,l)$ allows one to define *rigorous* upper and lower bounds for the entropy and free energy, $S^A$ (Eq. (11)] and $F^A$ [Eq. (12)], respectively.

### I. The quasiharmonic approximation

With the quasiharmonic (QH)approximation [40,41] the entropy $S_{\mathrm{QH}}$ is given by

$$S_{\mathrm{QH}} = (1/2)6Nk_B + (1/2)k_B \ln[(2\pi)^{6N}\sigma], \quad (21)$$

where $\sigma$ is the determinant of the covariance matrix of the $6N$ dihedral and bond angles.

## III. RESULTS AND DISCUSSION

### A. Simulation details for $(Gly)_{10}$

To obtain stable MD samples of $(Gly)_{10}$ in the helix, hairpin, and extended microstates the temperature was lowered to 100 K and each sample was started from a specific energy minimized structure. Thus, the initial helix structure (i.e., before minimization) is defined by $\varphi_i=\psi_i=-55°$ and $\omega_i=180°$ and the extended structure is $\varphi_i=\psi_i=\omega_i=180°$, $i=1,10$; the initial hairpin structure is $\varphi_i=\psi_i=\omega_i=180°$ for $i=1,4$ and $i=7,10$, while $\varphi_5=60°$, $\psi_5=-30°$, $\omega_5=180°$, $\varphi_6=90°$, $\psi_6=0°$, $\omega_6=180°$, i.e., the hairpin creates a type I′ turn. The first 5000 MD steps were used for equilibration and then 300 000 production MD steps were performed with a step size of 1 fs. The velocity-Verlet algorithm[21] was used to generate the dynamics with the Berendsen[21] heat bath controlling the temperature. A configuration was retained for future analysis every 500 MD steps; in this way three samples, each of 500 structures, were generated for the three microstates of $(Gly)_{10}$. As has been discussed in Sec. II E, to keep the future chains within the limits of the microstate the trajectory of the future chain is assembled from units of smaller trajectories. Each different unit is generated by restarting the simulation from the same initial conformation (the future part of the restructured conformation $i$) but with different velocities and discarding the initial configurations for equilibration. In this way, we can obtain rather large samples in which the system remains within the limits of the microstate. A unit $(n_f')$ was formed by keeping configurations every ten MD steps (i.e., 10 fs), where we seek to find the largest unit for which $S^A$ decrease with decreasing the bin size $\delta\alpha_k$ and increasing $n_f$ [Eq. (9)]. Different unit sizes were explored and the results for $S^A$ and $F^A$ differ slightly with unit size although we verify that the differences in these quantities—our main interest—are independent of the unit size. Notice, however, that the unit size should be the same while comparing different microstates (but it can vary from

TABLE I. The differences (in deg) between the minimum and maximum values of the dihedral angles [Eq. (8)] of $(Gly)_{10}$ obtained from MD samples of 400 conformations of the helix, hairpin, and extended microstates.

| | Extended | | | Helix | | | Hairpin | | |
|---|---|---|---|---|---|---|---|---|---|
| Res. No. | $\Delta\varphi$ | $\Delta\psi$ | $\Delta\omega$ | $\Delta\varphi$ | $\Delta\psi$ | $\Delta\omega$ | $\Delta\varphi$ | $\Delta\psi$ | $\Delta\omega$ |
| 1 | 66 | 159 | 34 | 50 | 125 | 38 | 50 | 245 | 39 |
| 2 | 81 | 51 | 35 | 43 | 45 | 26 | 155 | 60 | 35 |
| 3 | 83 | 49 | 37 | 37 | 32 | 28 | 57 | 38 | 30 |
| 4 | 89 | 49 | 41 | 35 | 42 | 28 | 63 | 46 | 33 |
| 5 | 97 | 47 | 31 | 33 | 40 | 30 | 40 | 95 | 31 |
| 6 | 112 | 52 | 36 | 36 | 41 | 26 | 156 | 52 | 29 |
| 7 | 143 | 43 | 33 | 43 | 44 | 29 | 77 | 48 | 31 |
| 8 | 99 | 54 | 35 | 32 | 38 | 33 | 74 | 34 | 31 |
| 9 | 99 | 49 | 32 | 39 | 35 | 25 | 85 | 46 | 31 |
| 10 | 119 | 52 | 30 | 64 | 48 | 33 | 216 | 53 | 32 |

system to system). For $(Gly)_{10}$ we have studied $n_f'$=unit =1500 (15 ps), 2000 (20 ps), and 500 (5 ps), where the equilibration size is 500 (5 ps).

The TPs and their product, $\rho^{HS}$ [Eqs. (9) and (10)], were calculated by reconstructing each conformation step by step with MD simulations of the future part. As mentioned earlier, for MD simulations (unlike MC) the bin becomes two-dimensional and a two-dimensional TP density is measured replacing the one-dimensional TP used in HSMC. To check the convergence of the results they were calculated for four future sample sizes, $n_f$=2000, 3000, 4000, 6000, 12 000, 18 000, and 24 000 and for unit=500 also $n_f$=500 and 1000. The future samples were generated for four bin sizes, $\delta$ $=\Delta\alpha_k/15$, $\Delta\alpha_k/10$, $\Delta\alpha_k/5$, and 20°, centered at $\alpha_k$ (i.e., $\alpha_k\pm\delta/2$). Notice that as for the LS method, the bin size is proportional to $\Delta\alpha_k$. If the counts of the smallest bin are smaller than 50, the bin size is increased to the next size, and if necessary to the next one, etc. In the case of zero counts, $n_{visit}$ is taken to be 1; however, zero counts is a very rare event. For $(Gly)_{10}$ samples of $n$=400 structures were analyzed.

## B. Results for the entropy of $(Gly)_{10}$

In Table I we present the values of $\Delta\alpha_k$ [Eq. (8)] for the extended, helix, and hairpin microstates obtained from the corresponding MD samples. These values suggest that the samples indeed are concentrated in conformational space as expected.

It should first be pointed out that as for the dihedral angles, Eq. (9) was used with $\delta\alpha_k$ also for the bond angles, i.e., without considering the Jacobian component $[\Pi_k \sin(\theta_k)]$, because we have found that to a good approximation, the contribution of the Jacobian to the entropy cancels out in entropy and free energy differences, which are our main interest; this allows us to compare the HSMD results for the entropy and free energy to those obtained with the flexible model of Paper I (Ref. 54) which were calculated without the Jacobian as well. Table II contains the results of the entropy $S^A$ [Eq. (11)] for the three different microstates, where the results on the left hand side are for unit=1500 and they were obtained from samples of $n$=400 conformations; for comparison we also provide results for unit=2000 on the right hand side of the table based on smaller samples of $n$ =200 conformations. The results were calculated for four different future sample sizes $n_f$ and four bin sizes. However, the extent of convergence of these results is demonstrated by the best ones, i.e., those for the three smallest bin sizes, $\Delta\alpha_k/5$, $\Delta\alpha_k/10$, and $\Delta\alpha_k/15$, and therefore only they are

TABLE II. Entopy $TS^A$ ($T$=100 K) in kcal/mol [Eq. (11)] for three bin sizes $\Delta\alpha_k/i$ [Eq. (5)], and future samples sizes $n_f$ obtained with the HSMD method for the three microstates of $(Gly)_{10}$ with unit=1500 and 2000. $\Delta\alpha_k$ is defined in Eq. (8). $n$ is the number of MD conformations in a microstate sample. The statistical errors are given in parentheses, e.g., 32.83(3)=32.83±0.03. $S_{QH}$ is the quasiharmonic entropy [Eq. (21)] and $S_{LS}$ [Eqs. (11) and (20)] is $S^A$ obtained by the local states (LS) method using $b$=1 and $l$=10 (for details see text). $S_{flex}$ is the entropy ($S^A$) of the flexible model obtained in Table IV of Paper I by HSMC (Ref. 54). The entropy is defined up to an additive constant.

| Bin size | $n_f$ | Extended | Helix | Hairpin | Extended | Helix | Hairpin |
|---|---|---|---|---|---|---|---|
| | | Unit=1500 | $n$=400 | | Unit=2000 | $n$=200 | |
| $\Delta\alpha_k/5$ | 6 000 | 33.12(2) | 29.15(6) | 30.6 (2) | 33.18(7) | 29.3(1) | 30.7(2) |
| $\Delta\alpha_k/5$ | 12 000 | 33.15(4) | 29.18(6) | 30.6 (2) | 33.20(7) | 29.3(1) | 30.7(2) |
| $\Delta\alpha_k/5$ | 18 000 | 33.15(3) | 29.19(6) | 30.6 (2) | 33.21(6) | 29.3(1) | 30.7(2) |
| $\Delta\alpha_k/5$ | 24 000 | 33.16(3) | 29.20(6) | 30.6 (2) | 33.21(7) | 29.3(1) | 30.7(2) |
| $\Delta\alpha_k/10$ | 6 000 | 32.82(1) | 28.83(6) | 30.0 (2) | 32.90(8) | 29.0(1) | 30.2(2) |
| $\Delta\alpha_k/10$ | 12 000 | 32.89(3) | 28.90(5) | 30.1 (2) | 32.95(9) | 29.0(1) | 30.2(2) |
| $\Delta\alpha_k/10$ | 18 000 | 32.90(2) | 28.92(6) | 30.1 (2) | 32.97(8) | 29.0(1) | 32.2 () |
| $\Delta\alpha_k/10$ | 24 000 | 32.90(3) | 28.93(6) | 30.1 (2) | 32.97(8) | 29.0(1) | 30.2(2) |
| $\Delta\alpha_k/15$ | 6 000 | 32.74(2) | 28.74(7) | 29.9 (2) | 32.82(8) | 28.9(1) | 30.1(2) |
| $\Delta\alpha_k/15$ | 12 000 | 32.82(6) | 28.83(6) | 29.9 (1) | 32.89(9) | 29.0(1) | 30.1(2) |
| $\Delta\alpha_k/15$ | 18 000 | 32.84(2) | 28.85(7) | 30.0 (1) | 32.91(8) | 29.0(1) | 30.1(2) |
| $\Delta\alpha_k/15$ | 24 000 | 32.83(3) | 28.85(7) | 30.0 (1) | 32.90(8) | 29.0(1) | 30.1(2) |
| $TS_{flex}$(I[a]) | | 28.5 (3) | 24.4 (1) | 25.41(7) | | | |
| $TS_{QH}$ | | 33.5 (1) | 29.4 (1) | 31.8 (1) | | | |
| $TS_{LS}$ | | 34.8 (1) | 32.1 (2) | 34.9 (6) | | | |

[a]Reference 54.

TABLE III. SMD results for the free energy $F^A$ [Eq. (12)], the energy $E$, and their fluctuations for $(Gly)_{10}$. $F^A$ is a lower bound of the free energy and $\sigma_A$ [Eq. (13)] is its fluctuation. The HSMD results were obtained from samples of $n=400$ conformations for the smallest bin size, $\delta=\Delta\alpha_k/15$, but for all future sample size $n_f$. $F_{QH}$ [Eq. (21)] and $F_{LS}$ [Eq. (12) and (20)] are free energies obtained by the quasiharmonic approximation and the local states method, respectively, and are based on larger samples (see text). The average potential energy $E_{int}$ of the HSMD samples, and $E_{flex}$ [I (Ref. 54)], the energy of the flexible model [from Table VI of Paper I using HSMC (Ref. 54)] appear in the two bottom rows; $\sigma_E$ is the energy fluctuation (these results are in kcal/mol). All free energies (at $T=100$ K) are in kcal/mol and are defined up to an additive constant. The statistical error is defined in the caption of Table II.

| | Extended | | Helix | | Harpin | |
|---|---|---|---|---|---|---|
| HSMC/$n_f$ | $-F^A$ | $\sigma_A$, $\sigma_E$ | $-F^A$ | $\sigma_A$, $\sigma_E$ | $-F^A$ | $\sigma_A$, $\sigma_E$ |
| 6 000 | 76.65(4) | 1.30(7) | 110.59(6) | 1.46(2) | 94.1(2) | 1.42(3) |
| 12 000 | 76.73(4) | 1.29(7) | 110.67(6) | 1.45(3) | 94.2(2) | 1.41(3) |
| 18 000 | 76.74(4) | 1.29(7) | 110.70(6) | 1.44(3) | 94.2(2) | 1.40(3) |
| 24 000 | 76.74(4) | 1.28(6) | 110.70(6) | 1.43(3) | 94.2(2) | 1.40(3) |
| $-F_{QH}$ | 77.6 (2) | | 111.39(6) | | 96.1(4) | |
| $-F_{LS}$ | 78.9 (2) | | 112.7 (2) | | 99.1(2) | |
| $-e_{int}$ | 43.91(6) | 1.46(3) | 81.85(2) | 1.57(6) | 64.2(1) | 1.55(7) |
| $-E_{flex}$ (I[a]) | 56.0 (3) | 1.0 (3) | 96.2 (3) | 1.4 (2) | 79.1(5) | 1.3 (2) |

[a]Reference 54.

presented in the table. The statistical errors were obtained from the fluctuations and results obtained for partial samples.

One would expect $S^A$ to decrease with decreasing the bin size—an expectation that indeed is materialized in the results of Table II. It should be pointed out, however, that the decrease of $S^A$ in going from $\delta=\Delta\alpha_k/10$ to $\Delta\alpha_k/15$ is approximately 0.1 kcal/mol (or smaller) within a relatively large statistical error of up to ±0.2 kcal/mol. One would also expect $S^A$ to decrease as the sample size, $n_f$ of the future chains, increases. However, when $n_f$ is increased the chance also increases for the creation of future chains that fluctuate significantly and might even deviate from the limits of the microstate leading thus to a decrease in the value of $n_{visit}$ [Eq. (9)] and hence to a larger $S^A$. Indeed, this effect is observed in the table for $\delta=\Delta\alpha_k/10$ and $\Delta\alpha_k/15$ as $TS^A$ increases from $n_f=6000-12\,000$; however, for $n_f=18\,000$ and 24 000 (and in many cases also for $n_f=12\,000$) the values of $TS^A$ are practically equal (within the error bars). These results suggest that for the given sample size $n$ (which determines to a large extent the statistical errors) decreasing $\delta$ or increasing $n_f$ further would not lead in most cases to better (i.e., smaller) $S^A$. However, the fact that the same decrease in $S^A(n_f)$ is observed in going from $\Delta\alpha_k/10$ to $\Delta\alpha_k/15$ suggests that for $\Delta\alpha_k/15$ $S^A(n_f)$ is obtained with the same accuracy for the three microstates. This means that differences in $S^A(n_f)$ for these microstates (which is our main interest) are expected to lead to the correct values, because the equal errors in $S^A(n_f)$ would get canceled. We shall return to this issue later.

To demonstrate the effect of the unit size we have also calculated results for unit=2000, which, as expected, are shown to be slightly larger than their counterparts for unit =1500 due to the increase in the number of fluctuating future chains (as explained in the previous paragraph for the case of increasing $n_f$). As discussed later, using unit=2000 will not change the differences in $TS^A(n_f)$. We also provide results for $TS^A$ obtained in Paper I (Ref. 54) for the flexible model of decaglycine (i.e., with constant bond lengths) where they are shown to be lower by ~4 kcal/mol than the present MD results that are based on more flexible chains (see also discussion in the second paragraph of the Summary section).

The HSMD results for the entropy are also compared in the table with those obtained using the LS and QH methods. For this we generated for QH larger MD samples of 10 000, 10 000, and 5000 conformations for the extended, helix, and hairpin microstates, respectively, by retaining a conformation every 30 fs. For LS samples of size 18 000 were generated by retaining a conformation every 10 fs. As expected, the QH results (like those obtained in Paper I) are larger than the HSMC values—here by 0.7–1.8 kcal/mol. The LS results (calculated for $b=1$, $l=10$) are larger than the corresponding QH values, as has also been found in Paper I.

### C. Results for the free energy of $(Gly)_{10}$

Results for the free energy functional $F^A$ [Eq. (12)] and its fluctuation $\sigma_A$ [Eq. (13)] and the energies are presented in Table III. These results are given only for the smallest bin $\Delta\alpha_k/15$ because $F^A$ values for the other bins can be obtained from the entropies of Table II and the energies provided in the bottom of Table III. $F^A$ (like $S^A$) does not change within the error bars as $n_f$ is increased from 12 000 to 24 000 and the central values of the fluctuations, as expected, decrease as the approximation improves but this decrease is insignificant within the error bars.

The results for $F^B$ are not provided in the table because they do not behave as expected, i.e., they do not decrease as $n_f$ is increased or as the bin size is decreased. This "misbehavior" can be attributed to a too small sample size $n$, or might stem from the fact that the bond stretching energy is included in the potential energy while the corresponding entropy is not taken into account in $\rho^{HS}$ [Eq. (10)]. More specifically, while the differences between the bond stretching energies of conformations are of ~1 kcal/mol, these differences (divided by $RT$) increase to ~5 kcal/mol and affect the exponential terms in Eq. (15) without the corresponding

TABLE IV. Entopy $TS^A$ ($T$=100 K) in kcal/mol [Eq. (11)] for three bin sizes $\Delta\alpha_k/i$ [Eq. (5)] obtained with the HSMD method for the three microstates of $(Gly)_{10}$ with unit=1500 and 500 based on smaller future sample sizes, $n_f$. $\Delta\alpha_k$ is defined in Eq. (8). The HSMD results are based on samples of $n$=400 conformations. The boldfaced results were obtained for unit =500. The statistical errors are defined in the caption of Table II. $S_{QH}$ is the quasiharmonic entropy [Eq. (21)] and $S_{LS}$ [Eqs. (11) and (20)] is the local states (LS) entropy ($S^A$) obtained for $b$=1 and $l$=10. The entropy is defined up to an additive constant.

| Bin size | $n_f$ | Extended | Helix | Hairpin |
|---|---|---|---|---|
| $\Delta\alpha_k/5$ | 500 | 33.09(3) | 28.9 (1) | 30.4(2) |
| $\Delta\alpha_k/5$ | 1000 | 32.86(3) | 29.1 (1) | 30.3(2) |
| $\Delta\alpha_k/5$ | 2000 | 33.04(2) | 29.05(6) | 30.5(2) |
| $\Delta\alpha_k/5$ | 3000 | 33.11(1) | 29.13(6) | 30.5(2) |
| $\Delta\alpha_k/5$ | 4000 | 33.11(2) | 29.13(6) | 30.5(2) |
| $\Delta\alpha_k/5$ | 6000 | 33.13(3) | 29.15(6) | 30.5(2) |
| $\Delta\alpha_k/10$ | 500 | 32.99(3) | 28.9 (1) | 30.1(2) |
| $\Delta\alpha_k/10$ | 1000 | 32.56(3) | 28.5 (1) | 29.7(2) |
| $\Delta\alpha_k/10$ | 2000 | 32.69(2) | 28.65(6) | 29.9(2) |
| $\Delta\alpha_k/10$ | 3000 | 32.78(2) | 28.75(6) | 30.0(2) |
| $\Delta\alpha_k/10$ | 4000 | 32.80(2) | 28.77(6) | 30.0(2) |
| $\Delta\alpha_k/10$ | 6000 | 32.83(2) | 28.81(6) | 30.0(2) |
| $\Delta\alpha_k/15$ | 500 | 32.98(3) | 28.9 (1) | 30.0(2) |
| $\Delta\alpha_k/15$ | 1000 | 32.53(3) | 28.5 (1) | 29.6(2) |
| $\Delta\alpha_k/15$ | 2000 | 32.61(3) | 28.56(6) | 29.7(2) |
| $\Delta\alpha_k/15$ | 3000 | 32.68(2) | 28.66(5) | 29.8(2) |
| $\Delta\alpha_k/15$ | 4000 | 32.71(2) | 28.68(6) | 29.8(2) |
| $\Delta\alpha_k/15$ | 6000 | 32.72(2) | 28.70(7) | 29.8(2) |
| $TS_{QH}$ | | 33.5 (1) | 29.4 (1) | 31.8(1) |
| $TS_{LS}$ | | 34.8 (1) | 32.1 (2) | 34.9(6) |

compensation from the entropy term, $\rho^{HS}$. Still, the results obtained for $F^B$ are always larger than those for $F^A$ and thus probably provide upper bounds, but the deviations are relatively large ($F^B$=71.0, −106.4, and −87.3 kcal/mol for extended, helix, and hairpin, respectively). Therefore, it is not clear whether the $F^B$ results lead to improved approximations for the free energy, i.e., whether the average values $F^M$ [Eq. (16)] are better than those of $F^A$. We have also calculated results for $F^D$ which have been found to be smaller than the corresponding $F^B$ values but larger than those for $F^M$. While it would be good to have reliably behaving results for $F^B$ and $F^D$ we demonstrate below that one can obtain reliable differences in entropy (and free energy) which are our main interest from differences in $S^A$ (and $F^A$).

In Table III we also provide the average potential energies and fluctuations of the different microstates. As expected, the energy fluctuations are always larger than the corresponding free energy fluctuations. For comparison we also present the energy values from Paper I (Ref. 54) (Table VI) for the flexible model I studied there, which are ~15 kcal/mol lower than the present MD results.

### D. Differences in entropy and free energy of $(Gly)_{10}$

Computer time increases linearly with $n_f$, therefore it is of interest to check the effect of decreasing $n_f$ on the entropy results. In Table IV we provide results for $TS^A$ for unit =1500 with $n_f$=2000, 3000, 4000, and 6000 for samples of size $n$=400. We also present results for unit=500(5 ps) for $n_f$=500 and 1000 (results are boldfaced in the table). The results for unit=1500 behave the same way as in Table II, i.e., they decrease as the bin size decreases from $\delta=\Delta\alpha_k/5$ to $\Delta\alpha_k/15$ and are approximately constant within a bin. As expected, because of smaller $n_f$ values, the results of each bin in Table IV are always somewhat smaller than their counterparts in Table II.

Because we are mostly interested in entropy differences, in Table V we present the differences $T\Delta S^A$ for the three microstates for several $n_f$ values for unit=2000, 1500, and 500 (for the smallest bin) and also for the flexible model of Paper I. The table reveals that all these results are equal within the error bars, which are the largest for the flexible model and for unit=2000 for which the results are based on a relatively small sample size, $n$=200. On the other hand, even for unit=500 ($n_f$=500) the errors of 0.1 and 0.2 kcal/mol are relatively small while the computer time is 48 times smaller than that required for $n_f$=24 000. In fact, reconstructing a conformation of $(Gly)_{10}$ based on $n_f$ =24 000 requires 2.4 h CPU on a 2.4 GHz Athlon processor whereas a reconstruction based on $n_f$=500 requires 3 min CPU. One can still increase the integration step to 2 fs which would decrease this time further to 90 s. It should be pointed out that similar results for $T\Delta S^A$ were obtained for other $n_f$ values and for the second smallest bin ($\Delta\alpha_k/10$). The fact that the differences $T\Delta S^A$ are constant while the values of $S^A$ change within a range of 0.5–0.6 kcal/mol suggest that these differences would remain constant also for more and more accurate values of $S^A$ and thus they constitute the *correct* differences within the error bars. In other words, for a given approximation, for each microstate $j$ (e.g., a helix), $S^A(j)=S_{exact}(j)+\delta S$, where $\delta S$ is an error which is approxi-

TABLE V. Differences in the entropy $T\Delta S^A$ (kcal/mol) at $T$=100 K between different microstates obtained by HSMD for $(Gly)_{10}$. $n$ is the size of the reconstructed MD sample; $n_f$ is the sample size of the future chains. The statistical error is defined in Table II. Results for the flexible model (using HSMC) were taken from Table VI of Paper I (Ref. 54).

| | Unit=1500 $n$=400 | | | Unit=500 $n$=400 | | Unit=2000 $n$=200 | |
|---|---|---|---|---|---|---|---|
| | $n_f$=24 000 | $n_f$=6000 | $n_f$=2000 | $n_f$=1000 | $n_f$=500 | $n_f$=6000 | Flexible model (I[a]) |
| $T(S_{extend}-S_{hairpin})$ | 2.9(1) | 2.9(2) | 2.9 (2) | 2.9 (2) | 2.9 (2) | 2.8 (3) | 3.0 (3) |
| $T9S_{extended}-S_{helix}$ | 4.0(1) | 4.0(1) | 4.0 (1) | 4.0 (1) | 4.0 (1) | 3.9 (2) | 4.0 (3) |
| $T(S_{hairpin}-S_{helix})$ | 1.1(1) | 1.2(1) | 1.2 (1) | 1.1 (1) | 1.1 (1) | 1.2 (1) | 1.0 (2) |

[a]Reference 54.

TABLE VI. Differences in energy $\Delta E$ and free energy $\Delta F^A$ ($T=100$ K) in kcal/mol between the three microstates of $(Gly)_{10}$ obtained by HSMD and in Paper I (Ref. 54 for the flexible model using HSMC. The present $\Delta F^A$ results were calculated for a sample of $n=400$ conformations, future sample size $n_f=24\,000$, and bin size $\Delta\alpha_k/15$.

| Microstates | $\Delta E$ | $\Delta E$ (flexible I[a]) | $\Delta F$ (unit=1500) | $\Delta F$ (flexible I[a]) |
|---|---|---|---|---|
| Extended-hairpin | 20.30(7) | 23.1(3) | 17.4(1) | 20.1(3) |
| Extended-helix | 37.94(7) | 40.2(2) | 34.0(1) | 36.1(3) |
| Hairpin-helix | 17.64(6) | 17.0(2) | 16.5(2) | 16.0(2) |

[a]Reference 54.

mately the same for all the microstates $j$ and thus is canceled in the differences $\Delta S^A$.

The above results demonstrate the advantage of MD over the MC procedure used in Paper I. With MD the conformational changes at each step are carried out deterministically along the forces (by solving Newton's equation of motion) and hence they are imposed with similar efficiency on the different microstates. Thus, if the amount of MD sampling is changed the three microstates are affected equally and the corresponding changes in entropy are approximately the same. On the other hand, for low $n_f$ values the efficiency of the MC procedure depends on the simulated structure (more rejections occur for a compact one), where our procedure has been found to be most efficient for the extended microstate, and as discussed in Paper I, a relatively large $n_f=160\,000$ is needed for calculating reliably $T\Delta S^A$ values for the three microstates; in this case the reconstruction of a single structure requires 2.4 h CPU, which is 100 times larger than that requires with the shortest MD run ($n_f=500$, and 2 fs step size).

In Table VI we provide the differences in energy $\Delta E$ and free energy $\Delta F^A$ for the three microstates and their counterparts for the flexible model from Paper I. The table reveals that the two sets of $\Delta E$ values are similar, which explains the equality in the $T\Delta S^A$ values of the two models in Table V. As one would expect, the larger $\Delta E$ values correspond to the larger $T\Delta S^A$ values in Table V; however, changes in $\Delta E$ correspond to much smaller changes in $T\Delta S^A$, e.g., an increase of 17.6 kcal/mol in $\Delta E$ (in going from 20.30 to 37.94 kcal/mol, see Table VI) corresponds to an increase of 1.1 kcal/mol of $T\Delta S^A$ (in going from 2.9 to 4.0 kcal/mol for the first set of results in Table V). Thus, we have calculated the average bond stretching energy and have obtained 8.57, 9.95, and 9.23 kcal/mol for the extended, helix, and hairpin microstates, respectively. The differences between these values contribute very little to the $\Delta E$ values in the table and therefore the corresponding bond stretching entropies are expected to be small and thus will not contribute to the differences in Table V. This justifies our ignoring the bond stretching entropy from $S^A$ (but not necessarily from $F^B$ as previously discussed).

TABLE VII. The differences (in deg) between the minimum and maximum values of the dihedral angles [Eq. (8)] of $(Val)_2(Gly)_6(Val)_2$ obtained from MD sample of 400 conformations of the helix and hairpin microstates.

| | Helix | | | Hairpin | | |
|---|---|---|---|---|---|---|
| Res. No. | $\Delta\varphi$ | $\Delta\psi$ | $\Delta\omega$ | $\Delta\varphi$ | $\Delta\psi$ | $\Delta\omega$ |
| 1 | 42 | 62 | 31 | 48 | 38 | 28 |
| 2 | 37 | 44 | 24 | 26 | 36 | 30 |
| 3 | 44 | 48 | 23 | 50 | 32 | 29 |
| 4 | 41 | 47 | 29 | 36 | 42 | 26 |
| 5 | 36 | 42 | 25 | 31 | 47 | 31 |
| 6 | 37 | 45 | 25 | 30 | 44 | 38 |
| 7 | 46 | 45 | 23 | 37 | 133 | 35 |
| 8 | 44 | 58 | 27 | 101 | 42 | 32 |
| 9 | 55 | 41 | 24 | 30 | 37 | 34 |
| 10 | 37 | 50 | 33 | 34 | 115 | 40 |

### E. Results for $S^A$ and $F^A$ for $(Val)_2(Gly)_6(Val)_2$

The MD samples at $T=100$ K for $(Val)_2(Gly)_6(Val)_2$ were obtained in a similar way as described for $(Gly)_{10}$ but with the following changes. First, only the helix and hairpin microstates were studied, because the extended state was found to be unstable even at 100 K. Second, the step size was increased to 2 fs where bonds involving hydrogens were frozen to their ideal values by using the RATTLE algorithm.[21] Also, the microstates of $(Val)_2(Gly)_6(Val)_2$ are less stable than those of $(Gly)_{10}$ which required using smaller units of sizes 600 and 400. The $\Delta\alpha_k$ values for the two microstates

TABLE VIII. Entropy $TS^A$ ($T=100$ K) in kcal/mol [Eq. (11)] for three bin sizes $\Delta\alpha_k/i$ [Eq. (5)] and future sample sizes $n$ obtained with the HSMD method for the helix and hairpin microstates of $(Val)_2(Gly)_6(Val)_2$ based on unit=600 and 400. $\Delta\alpha_k$ is defined in Eq. (8). The HSMD results are based on samples of $n=400$ conformations. The statistical errors are defined in the caption of Table II. $S_{QH}$ is the quasi harmonic entropy [Eq. (21)] and $S_{LS}$ [Eqs. (11) and (20)] is the local states (LS) entropy ($S^A$) obtained for $b=1$ and $l=10$. The entropy is defined up to an additive constant.

| | $nf$ | Helix | Hairpin | Helix | Hairpin |
|---|---|---|---|---|---|
| | | Unit=600 $n$=400 | | Unit=400 $n$=200 | |
| $\Delta\alpha_k/5$ | 6 000 | 31.07 (3) | 30.6 (1) | 31.1 (1) | 30.8 (2) |
| $\Delta\alpha_k/5$ | 12 000 | 31.06 (2) | 30.6 (1) | 31.1 (1) | 30.7 (2) |
| $\Delta\alpha_k/5$ | 18 000 | 31.05 (2) | 30.6 (1) | 31.1 (1) | 30.7 (2) |
| $\Delta\alpha_k/5$ | 24 000 | 31.04 (3) | 30.6 (1) | 31.1 (1) | 30.7 (2) |
| $\Delta\alpha_k/10$ | 6 000 | 30.75 (2) | 30.2 (1) | 30.8 (1) | 30.4 (2) |
| $\Delta\alpha_k/10$ | 12 000 | 30.73 (2) | 30.2 (1) | 30.7 (1) | 30.3 (2) |
| $\Delta\alpha_k/10$ | 18 000 | 30.71 (3) | 30.1 (1) | 30.7 (1) | 30.3 (2) |
| $\Delta\alpha_k/10$ | 24 000 | 30.70 (3) | 30.1 (1) | 30.7 (1) | 30.3 (2) |
| $\Delta\alpha_k/15$ | 6 000 | 30.69 (2) | 30.1 (1) | 30.7 (1) | 30.3 (2) |
| $\Delta\alpha_k/15$ | 12 000 | 30.67 (2) | 30.1 (1) | 30.7 (1) | 30.2 (2) |
| $\Delta\alpha_k/15$ | 18 000 | 30.64 (2) | 30.1 (1) | 30.7 (1) | 30.2 (2) |
| $\Delta\alpha_k/15$ | 24 000 | 30.63 (2) | 30.0 (1) | 30.6 (1) | 30.2 (2) |
| $TS_{QH}$ | | 31.7 (1) | 31.1 (2) | | |
| $TS_{LS}$ | | 35.8 (5) | 36.2 (5) | | |

TABLE IX. HSMD results for the free energy $F^A$ [Eq. (12)], the energy $E$, and their fluctuations for $(Val)_2(Gly)_6(Val)_2$. $F^A$ is a lower bound of the free energy and $\sigma_A$ [Eq. (13)] is its fluctuation. The HSMD results were obtained from samples of $n=400$ conformations for the smallest bin size, $\delta=\Delta\alpha_k/15$, but for all future sample sizes $n_f$. $F_{QH}$ [Eq. (21)] and $F_{LS}$ [Eqs. (12) and (20)] are free energies obtained by the quasiharmonic approximation and the local states method, respectively, and are based on larger samples (see text). The average potential energy $E_{int}$ (in kcal/mol) and its fluctuation $\sigma_E$ appears in the bottom row. All free energies (at $T=100$ K) are in kcal/mol and are defined up to an additive constant. The statistical error is defined in the caption of Table II.

| | Helix | | Hairpin | |
|---|---|---|---|---|
| HSMC/$n_f$ | $-F^A$ | $\sigma_A$, $\sigma_E$ | $-F^A$ | $\sigma_A$, $\sigma_E$ |
| 6 000 | 163.9(1) | 1.76(3) | 160.05(7) | 1.59(4) |
| 12 000 | 163.9(1) | 1.75(3) | 160.03(5) | 1.58(4) |
| 18 000 | 163.9(1) | 1.75(3) | 160.02(5) | 1.56(3) |
| 24 000 | 163.9(1) | 1.74(3) | 160.00(5) | 1.56(3) |
| $-F_{QH}$ | 165.0(1) | | 161.0 (1) | |
| $-F_{LS}$ | 168.9(5) | | 166 (2) | |
| $-E_{int}$ | 133.2(1) | 1.8 (1) | 130.0 (1) | 1.72(2) |

are presented in Table VII, which are shown to be quite restricted. Not shown in the table are the values of $\Delta_{\chi_1}$ which are $\sim 30°$.

The results for $TS^A$ (at $T=100$ K) for unit=600 and 400 in Table VIII show the expected behavior, i.e., they increase with bin size and those for unit=400 are slightly smaller than their counterparts for unit=600 [see previous discussion regarding units of 1500 and 2000 for $(Gly)_{10}$]. For the given sample sizes studied, $n=400$ and 200, the results are converged, i.e., they do not change (within the statistical errors) with decreasing the bin size or increasing $n_f$. The table reveals that the entropies of the two microstates are close. Again, The QH and LS results constitute upper bounds, where as for $(Gly)_{10}$ the QH values are smaller than the corresponding LS ones. The QH results are based on 5000 and 2500 conformations for the helix and hairpin, respectively (a conformation was retained every 40 fs), while the LS results are based on samples of 24 000 conformations (every 10 fs).

In Table IX we provide the free energies $F^A$, the potential energies, and their fluctuations. It should first be pointed out that the energy difference between the two microstates, $\sim 3$ kcal/mol, is relatively small leading thus to a small difference $T\Delta S^A$ as discussed in detail below; correspondingly, the free energy differences are also small. The tendencies of the results of both $F^A$ and $\sigma_A$ are as expected (see discussion of Table III); in particular, the $\sigma_A$ values are smaller than the corresponding energy fluctuations, $\sigma_E$.

TABLE X. Entropy $TS^A$ ($T=100$ K) in kcal/mol [Eq. (11)] for three bin sizes $\Delta\alpha_k/i$ [Eq. (5)] obtained with the HSMD method for the helix and hairpin microstates of $(Val)_2(Gly)_6(Val)_2$ using unit=600 and smaller future sample sizes $n_f$. $\Delta\alpha_k$ is defined in Eq. (8). The HSMD results are based on samples of $n=400$ conformations. The statistical errors are defined in the caption of Table II. $S_{QH}$ is the quasiharmonic entropy [Eq. (21)] and $S_{LS}$ [Eqs. (11) and (20)] is the local states (LS) entropy ($S^A$) obtained for $b=1$ and $l=10$. The entropy is defined up to an additive constant.

| | $n_f$ | Helix | Hairpin |
|---|---|---|---|
| $\Delta\alpha_k/5$ | 2000 | 31.03(5) | 30.6(2) |
| $\Delta\alpha_k/5$ | 3000 | 31.03(5) | 30.5(2) |
| $\Delta\alpha_k/5$ | 4000 | 31.00(4) | 30.5(2) |
| $\Delta\alpha_k/5$ | 6000 | 30.96(4) | 30.5(1) |
| $\Delta\alpha_k/10$ | 2000 | 30.64(5) | 30.1(2) |
| $\Delta\alpha_k/10$ | 3000 | 30.63(5) | 30.1(2) |
| $\Delta\alpha_k/10$ | 4000 | 30.60(4) | 30.1(2) |
| $\Delta\alpha_k/10$ | 6000 | 30.54(4) | 30.0(1) |
| $\Delta\alpha_k/15$ | 2000 | 30.53(5) | 29.9(2) |
| $\Delta\alpha_k/15$ | 3000 | 30.53(5) | 29.9(2) |
| $\Delta\alpha_k/15$ | 4000 | 30.50(4) | 29.9(2) |
| $\Delta\alpha_k/15$ | 6000 | 30.45(4) | 29.9(2) |
| $TS_{QH}$ | | 31.7 (1) | 31.1(2) |
| $TS_{LS}$ | | 35.8 (5) | 36.2(5) |

### F. Entropy and free energy differences for $(Val)_2(Gly)_6(Val)_2$

Because we are highly interested in an efficient calculation of entropy and free energy differences, in Table X we present HSMD results for $TS^A$ based on unit=600 obtained for significantly smaller $n_f$ values (2000–6000) than in Table VIII. The statistical errors here are slightly larger than in Table VIII, while the $TS^A$ results in Table X are smaller than their counterparts in Table VIII due to smaller $n_f$ values that keep the future chains better within the limits of the microstates thus leading to larger $n_{visit}$ values [Eq. (9)]. The expected behavior of the results with decreasing bin size and increasing $n_f$ is observed.

In Table XI results for $T\Delta S^A$, $\Delta F^A$, and $\Delta E$ are presented for unit=600 and 400 for several $n_f$ values. The table shows that $T\Delta S^A=0.6\pm0.2$ remains unchanged as $n_f$ is decreased and it is equal within the error bars to $0.4\pm0.2$ kcal/mol obtained for unit=400 for a smaller sample of $n=200$. Again, as for $(Gly)_{10}$, a significant reduction in computer time can be achieved for calculating entropy and free energy differences. Thus, reconstruction of a single structure of $(Val)_2(Gly)_6(Val)_2$ using $n_f=24\,000$ requires 4.3 h CPU while for $n_f=2000$ it requires 21 min CPU, and as for $(Gly)_{10}$ this time can probably be reduced further by a factor of 4 to 5 min (using $n_f=600$ or 400).

TABLE XI. Differences in the entropy $T\Delta S^A$ (kcal/mol) at $T=100$ K between the helix and hairpin microstates obtained by HSMD for $(Val)_2(Gly)_6(Val)_2$. $n$ is the size of the reconstructed MD sample; $n_f$ is the sample size of the future chains. $\Delta E$ and $\Delta F^A$ are the differences in energy and free energy, respectively (in kcal/mol). The statistical error is defined in Table II.

| | | | Unit=600 $n=400$ | | | Unit=400 $n=200$ | | |
|---|---|---|---|---|---|---|---|---|
| Microstates | $\Delta E$ | $\Delta F^A$ | $n_f=24\,000$ | $n_f=6000$ | $n_f=2000$ | $n_f=24\,000$ | $n_f=12\,000$ | $n_f=2000$ |
| $T(S_{helix}-S_{hairpin})$ | −3.3(2) | −3.87(8) | 0.6(1) | 0.6(2) | 0.6(2) | 0.4(2) | 0.4(2) | 0.4(2) |

## IV. SUMMARY AND CONCLUSIONS

In our previous work [Paper I (Ref. 54)] the HSMC method has been applied initially to polyglycine molecules in vacuum simulated by MC. Because MD is considered to be significantly more efficient than MC and hence is the commonly used method in proteins; in the present paper our method has been extended to MD simulations and the new version, HSMD, has been applied to $(Gly)_{10}$ and for the first time to a peptide with side chains, $(Val)_2(Gly)_6(Val)_2$. As before, we calculate the entropy and free energy of three microstates, helix, extended, and hairpin, and find the results to be more accurate than those obtained by the QH and LS methods that both provide upper bounds for the entropy[69] (however, see also discussion in the next paragraph). We also compare our results to those obtained for the flexible model of $(Gly)_{10}$ in Paper I (Ref. 54) at $T=100$.

To keep the molecule within the limits of the microstates during the reconstruction process, the MD simulation is divided into several repeating "units" each unit starts from the reconstructed conformation $i$. This raises the following questions: (1) Is the unit long enough to cover the sampled microstate? (2) What is the dependence of the results on the unit size? To answer the first question we generated a 15 ps MD sample of $(Gly)_{10}$ and a 6 ps sample of $(Val)_2(Gly)_6(Val)_2$ by retaining a conformation every 10 fs (as in the reconstruction process), converted these conformations to internal coordinates, and calculated the $\Delta\alpha_k$ values [Eq. (8)] of the samples. We have found that the $\Delta\alpha_k$ sets thus obtained are comparable to those presented in Tables I and VII, respectively, suggesting that a suitable coverage is achieved by these units. On the other hand, it is evident that the results for the *absolute* $S$ and $F$ depend somewhat on unit size and there is no criterion to determine the correct value. However, this problem reflects the difficulty inherent in defining a microstate in conformational space by simulation, which affects all entropy methods. For example, to obtain reasonable precision with QH (or LS) significantly longer trajectories than those used with HSMD are required (see text), which are expected to span larger regions in space thus leading to an increase in entropy; therefore, the overestimation of the entropy results for QH and LS is probably also due to this effect of larger trajectories.

However, we have shown that differences in entropy $T\Delta S^A$ obtained from absolute values (calculated for the same conditions) are very stable for various unit sizes, bin sizes, and sample size $n_f$, where the latter values can be relatively small leading to extremely efficient calculations. The accuracy of $T\Delta S^A$ of 0.1–0.2 kcal/mol is very satisfying; in general, the validity of such results can be verified by increasing the accuracy of HSMD, i.e., decreasing the bin size, increasing $n_f$ and/or changing the unit size. We have also argued that the effect of bond stretching on differences $\Delta S^A$ can in general be neglected but also suggested an approximate way to take this contribution into account if necessary. The stable results for entropy differences and the high efficiency obtained in this work open the door for the application of HSMD to more complex systems. As a next step, HSMD will be applied to a flexible loop in a protein where solvent effects will be taken into account implicitly. Because HSMC(D) is applicable to water we intend in a later stage to apply it to a loop capped by explicit water.

## ACKNOWLEDGMENT

This work was supported by NIH Grant No. R01 GM66090.

[1] D. L. Beveridge and F. M. DiCapua, Annu. Rev. Biophys. Biophys. Chem. **18**, 431 (1989).
[2] P. A. Kollman, Chem. Rev. (Washington, D.C.) **93**, 2395 (1993).
[3] W. L. Jorgensen, Acc. Chem. Res. **22**, 184 (1989).
[4] T. P. Straatsma and J. A. McCammon, Annu. Rev. Phys. Chem. **43**, 407 (1992).
[5] H. Meirovitch, in *Reviews in Computational Chemistry*, edited by Kenny B. Lipkowitz and Donald B. Boyd (Wiley, New York, 1998), Vol. 12, p. 1.
[6] T. L. Hill, *Statistical Mechanics Principles and Selected Applications* (Dover, New York, 1956).
[7] B. J. Alder and T. E. Wainwright, J. Chem. Phys. **31**, 459 (1959).
[8] J. A. McCammon, B. R. Gelin, and M. Karplus, Nature (London) **267**, 585 (1977).
[9] R. Elber and M. Karplus, Science **235**, 318 (1987).
[10] F. H. Stillinger and T. A. Weber, Science **225**, 983 (1984).
[11] E. D. Getzoff, H. M. Geysen, S. J. Rodda, H. Alexander, J. A. Tainer, and R. A. Lerner, Science **235**, 1191 (1987).
[12] J. M. Rini, U. Schulze-Gahmen, and I. A. Wilson, Science **255**, 959 (1992).
[13] K. L. Constantine, M. S. Friedrichs, M. Wittekind, H. Jamil, C. H. Chu, R. A. Parker, V. Goldfarb, L. Mueller, and B. T. Farmer, Biochemistry **37**, 7965 (1998).
[14] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[15] A. E. Garcia and K. Y. Sanbonmatsu, Proteins **42**, 345 (2001).
[16] B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).
[17] K. Ikeda, O. V. Galzitskaya, H. Nakamura, and J. Higo, J. Comput. Chem. **24**, 310 (2003).
[18] I. R. MacDonald and K. Singer, J. Chem. Phys. **47**, 4766 (1967).
[19] J.-P. Hansen and L. Verlet, Phys. Rev. **184**, 151 (1969).
[20] W. G. Hoover and F. H. Ree, J. Chem. Phys. **47**, 4873 (1967).
[21] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).
[22] J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).
[23] R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).
[24] D. R. Squire and W. G. Hoover, J. Chem. Phys. **50**, 701 (1969).
[25] G. M. Torrie and J. P. Valleau, Chem. Phys. Lett. **28**, 578 (1974).
[26] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[27] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[28] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[29] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kolmann, J. Comput. Chem. **16**, 1339 (1995).
[30] S. Kumar, P. W. Payne, and M. Vásquez, J. Comput. Chem. **17**, 1269 (1996).
[31] A. Szarecka, R. P. White, and H. Meirovitch, J. Chem. Phys. **119**, 12084 (2003).
[32] R. P. White and H. Meirovitch, J. Chem. Phys. **119**, 12096 (2003).
[33] J. P. Stoessel and P. Nowak, Macromolecules **23**, 1961 (1990).
[34] C. Costenbrink and W. F. van Gunsteren, J. Comput. Chem. **24**, 1730 (2003).
[35] A. F. Voter, J. Chem. Phys. **82**, 1890 (1985).
[36] Y. Duan and P. A. Kollman, Science **282**, 740 (1998).
[37] N. Gō and H. A. Scheraga, J. Chem. Phys. **51**, 4751 (1969).
[38] N. Gō and H. A. Scheraga, Macromolecules **9**, 535 (1976).
[39] A. T. Hagler, P. S. Stern, R. Sharon, J. M. Becker, and F. Naider, J. Am. Chem. Soc. **101**, 6842 (1979).
[40] M. Karplus and J. N. Kushick, Macromolecules **14**, 325 (1981).
[41] O. L. Rojas, R. M. Levy, and A. Szabo, J. Chem. Phys. **85**, 1037 (1986).
[42] H. Meirovitch, Chem. Phys. Lett. **45**, 389 (1977).
[43] H. Meirovitch, Phys. Rev. B **30**, 2866 (1984).
[44] H. Meirovitch, M. Vásquez, and H. A. Scheraga, Biopolymers **26**, 651 (1987).
[45] H. Meirovitch, S. C. Koerber, J. Rivier, and A. T. Hagler, Biopolymers

**34**, 815 (1994).
[46] A. J. Chorin, Phys. Fluids **8**, 2656 (1996).
[47] H. Meirovitch, J. Phys. A **16**, 839 (1983).
[48] H. Meirovitch, Phys. Rev. A **32**, 3709 (1985).
[49] H. Meirovitch, J. Chem. Phys. **111**, 7215 (1999).
[50] H. Meirovitch, J. Chem. Phys. **114**, 3859 (2001).
[51] R. P. White and H. Meirovitch, J. Chem. Phys. **121**, 10889 (2004).
[52] R. P. White and H. Meirovitch, Proc. Natl. Acad. Sci. U.S.A. **101**, 9235 (2004).
[53] S. Cheluvaraja and H. Meirovitch, Proc. Natl. Acad. Sci. U.S.A. **101**, 9241 (2004).
[54] S. Cheluvaraja and H. Meirovitch, J. Chem. Phys. **122**, 054903-1 (2004).
[55] S. Cheluvaraja and H. Meirovitch, J. Phys. Chem. B **109**, 21963 (2004).
[56] R. P. White, J. Funt, and H. Meirovitch, Chem. Phys. Lett. **410**, 430 (2005).
[57] R. P. White and H. Meirovitch, J. Chem. Phys. **123**, 214908-1 (2005).
[58] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[59] J. P. Ulmschneider and W. L. Jorgensen, J. Chem. Phys. **118**, 4261 (2003).
[60] R. C. Rizzo, J. Tirado-Rives, and W. L. Jorgensen, J. Med. Chem. **44**, 145 (2001).
[61] J. Hu, A. Ma, and A. R. Dinner, J. Comput. Chem. **27**, 203 (2006).
[62] H. Meirovitch and Z. Alexandrowicz, J. Stat. Phys. **15**, 123 (1976).
[63] H. Meirovitch, J. Chem. Phys. **89**, 2514 (1988).
[64] H. Meirovitch, M. Vásquez, and H. A. Scheraga, Biopolymers **27**, 1189 (1988).
[65] J. Brady and M. Karplus, J. Am. Chem. Soc. **107**, 6103 (1985).
[66] H. Meirovitch and E. Meirovitch, J. Phys. Chem. **100**, 5123 (1996).
[67] C. Baysal and H. Meirovitch, Biopolymers **50**, 329 (1999).
[68] W. Gibbs, *Elementary Principles in Statistical Mechanics* (Yale University Press, New Haven, CT, 1902), Chap. XI.
[69] C.-E. Chang, W. Chen, and M. K. Gilson, J. Chem. Theory Comput. **1**, 1017 (2005).

## AUTHOR QUERIES — 513623JCP

#1 AU: PLS CHECK CHANGES MADE IN EQ. (7)
#2 AU: PLS CHECK CHANGES IN EQ. (10)
#3 Au: PLS CHECK CHANGES ON THE TERM “ / femtosecondslfemtosecondslfemtosecondslfemtosecondslfe