Calculation of the entropy and free energy by the hypothetical scanning Monte Carlo method: Application to peptides

Srinath Cheluvaraja and Hagai Meirovitch[a)]
Center for Computational Biology and Bioinformatics and Department of Molecular Genetics and Biochemistry, University of Pittsburgh School of Medicine, W1058 BST, Pittsburgh, Pennsylvania 15261

A new approach, the hypothetical scanning Monte Carlo (HSMC), for calculating the *absolute* entropy, S, and free energy, F, has been introduced recently and applied first to fluids (argon and water) and later to peptides. In this paper the method is further developed for peptide chains in vacuum. S is calculated from a given MC sample by reconstructing each sample conformation i step-by-step, i.e., calculating transition probabilities (TPs) for the dihedral and bond angles and fixing the related atoms at their positions. At step k of the process the chain's coordinates that have already been determined are kept fixed (the ''frozen past'') and $TP(k)$ is obtained from a MC simulation of the ''future'' part of the chain whose TPs as yet have not been determined; when the process is completed the contribution of conformation i to the entropy is, $S_i \sim -\ln \Pi_k TP(k)$. In a recent paper we studied polyglycine chains, modeled by the AMBER force field with constant bond lengths and bond angles (the rigid model). Decaglycine $[(Gly)_{10}]$ was studied in the helical, extended, and hairpin microstates, while $(Gly)_{16}$ was treated only in the first two microstates. In this paper the samples are increased and restudied, $(Gly)_{16}$ is also investigated in the hairpin microstate, and for $(Gly)_{10}$ approximations are tested where only part of the future is considered for calculating the TPs. We calculate upper and lower bounds for F and demonstrate that like for fluids, F can be obtained from multiple reconstructions of a *single* conformation. We also test a more realistic model of $(Gly)_{10}$ where the bond angles are allowed to move (the flexible model). Very accurate results for S and F are obtained which are compared to results obtained by the quasiharmonic approximation and the local states method. Thus, differences in entropy and free energy between the three microstates are obtained within errors of 0.1–0.3 kcal/mol. The HSMC method can be applied to a macromolecule with any degree of flexibility, ranging from local fluctuations to a random coil. The present results demonstrate that the difference in stability, $\Delta F_{mn} = F_m - F_n$ between *significantly different* microstates m and n, can be obtained from two simulations only without the need to resort to thermodynamic integration. Our long-term goal is to extend this method to any peptide and apply it to a peptide immersed in a box with explicit water. © *2005 American Institute of Physics.* [DOI: 10.1063/1.1835911]

I. INTRODUCTION

The hypothetical scanning (HS) method is a general technique for calculating the *absolute* entropy, S and free energy F by computer simulation. The method has been initially applied to Ising models[1] and lattice polymer models,[2–6] and recently has been extended to fluids by two procedures, the *grand canonical* HS (HSGC) (Ref. 7) and the Monte Carlo (MC) HS (HSMC).[8] HSMC has been further developed to a method named *complete* HSMC,[9] which enables one, in principle, to obtain S and F exactly. Complete HSMC was applied very successfully to liquid argon, water,[9,10] and peptides.[11] In these papers we have provided an extensive discussion on the importance and the difficulty in calculating S and F by computer simulation (see also Refs. 12–15, and references cited therein). In short, the commonly used simulation techniques, the Metropolis MC method[16] and molecular dynamics[17,18] enable one to sample configuration i with

its Boltzmann probability, P_i^B, however, they do not provide the *value* of this probability in a direct way (e.g., such as the energy) and therefore the absolute entropy $S \sim -\ln P_i^B$ is unknown. Still, differences in $F(S)$ between two states can be obtained by the thermodynamic integration approach, but only when the absolute $F(S)$ of one state is known can $F(S)$ of the other be obtained. The HS approach and another method of Meirovitch, the local states (LS) method,[19–23] are unique in that they provide (at least approximately) the value of the sampling probability; hence S and F can be obtained.

The ability to obtain the absolute free energy is in particular important in computational structural biology. The energy surface of a protein, commonly defined by a force field, is highly rugged, consisting of a tremendous number of local minima,[24] where the native structure corresponds to the localized energy well with the lowest F. However, molecular dynamics simulations have shown[25,26] that even a protein with a well-defined structure fluctuates significantly within a region called *wide microstate* (e.g., the conformational region of an α-helix of a peptide) that typically consists of

a)Author to whom correspondence should be addressed. Electronic mail: hagaim@pitt.edu; Phone: 412-648-3338.

many localized energy wells. A peptide, protein, or protein segments such as surface loops, can exhibit an *intermediate flexibility*, where several wide microstates are populated significantly at thermodynamic equilibrium. It is essential to be able to identify these wide microstates, m, calculate F_m, which lead to their relative populations, and to weighted averages of various quantities that can be compared with experimental values.[27,28] F_m is in particular useful if m and n differ significantly; then, calculating the difference, $\Delta F_{mn} = F_m - F_n$ is straightforward. On the other hand, calculating ΔF_{mn} by the commonly used thermodynamic integration approach might be difficult or prohibitive (see Refs. 12–15, and references therein).

With complete HSMC applied to a peptide, S is calculated from a given MC sample by reconstructing each peptide conformation step-by-step, i.e., calculating transition probabilities (TPs) for the dihedral angles and fixing the related atoms at their positions. 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. In a recent paper this method was applied to the polyglycine molecules, $(Gly)_{10}$ and $(Gly)_{16}$ in vacuum described by a simplified model based on constant bond lengths and bond angles (the rigid model). Decaglycine $[(Gly)_{10}]$ was studied in the helical, extended, and hairpin microstates, while $(Gly)_{16}$ was treated only in the first two microstates.

In this paper we study the same molecules but the scope of the calculations is extended significantly. First, the samples are increased, $(Gly)_{16}$ is also investigated in the hairpin wide microstate, and for $(Gly)_{10}$ approximations are tested where only part of the future is considered for calculating the TPs. We calculate upper and lower bounds for F and demonstrate that like for fluids, F can be obtained from reconstructions of a *single* conformation. We also test a more realistic model of $(Gly)_{10}$ where the bond angles are allowed to change (the flexible model). Accurate results for S and F are obtained which are compared to results obtained by the quasiharmonic (QH) approximation[29,30] and the local states method. Thus, differences in entropy and free energy between these microstates are obtained within errors of 0.1–0.3 kcal/mol. Our long-term goal is to extend the complete HSMC method to any peptide and apply it to a peptide immersed in a box with explicit water.

II. THEORY AND METHODOLOGY

A. The models studied

We study two models of polyglycine, $NH_2(Gly)_N CONH_2$, in vacuum defined by the AMBER96 force field,[31] where the charges of the end groups are neutralized. One model (the rigid model) is based on constant bond angles and bond lengths, i.e., a conformation is determined by the dihedral angles ϕ_i, ψ_i, and ω_i ordered along the chain, which are denoted for simplicity by α_k, $k = 1, 3N$, where N is the number of residues. This model is studied with $N = 10 [(Gly)_{10}]$ and $N = 16 [(Gly)_{16}]$, i.e., the corresponding numbers of internal coordinates are 30 and 48. The

second model of $(Gly)_{10}$ (the flexible model) is more realistic allowing the three bond angles of each residue to vary as well; thus, the total number of internal coordinates is 60. Each model is simulated with the Metropolis MC method[16] within three wide microstates, helix, extended, and hairpin, and the entropy and free energy are calculated from the generated samples using the HSMC method. Notice that for brevity we shall refer in most cases to "wide microstates" as microstates omitting the word "wide."

It should be pointed out that both MD and MC are most straightforwardly carried out in Cartesian coordinates, while application of the local states (LS) and HS approaches is based on internal coordinates; therefore, in previous LS studies conformations generated by MD were transferred into internal coordinates before analysis.[22] However, we have found MC simulations in Cartesian coordinates (i.e., for a *fully* flexible model) to be extremely inefficient, and therefore have chosen to treat the rigid and flexible models described above, where ϕ_i, ψ_i, and ω_i, and the bond angles are the natural variables which can be used in the MC simulations. This internal variables representation of a peptide provided by the program TINKER (Ref. 32) has led indeed to a MC method that is significantly more efficient than that based on Cartesian coordinates; thus, for the present models the application of HSMC, LS, and QH is direct. In what follows, for simplicity, the various methods will be described as applied to the present models of polyglycine, where the extension to a peptide with side chains is straightforward.

B. Statistical mechanics of a peptide in internal coordinates

The partition function of a peptide, Z, is an integral over the function $\exp(-E/k_B T)$ (E is the potential energy and k_B the Boltzmann constant) with respect to the Cartesian coordinates over the whole conformational space, Ω. However, for a stable wide microstate the integration is carried out only over the limited region Ω_0 that defines the wide microstate (helix, hairpin, etc.). As said above, to apply HSMC or LS, 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 is a simple function of the bond angles and bond lengths.[29,33,34] In previous LS studies of linear and cyclic peptides, and surface loops in proteins, an approximate transformation to dihedral and bond angles was adopted where the bond lengths were kept constant (see below).[22]

Our rigid model is defined by transforming the Cartesian coordinates (of a *fully* flexible model) into internal coordinates under the assumption that the potentials of the bond lengths and bond angles ("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. It should be noted that while the contribution of bond stretching to the *absolute* entropy is not small, it is expected to be similar for different wide microstates of the same molecule. Therefore, to a good approximation, the contribution of bond stretching to the differences $\Delta S_{m,n}$ and $\Delta F_{m,n}$ be-

tween wide microstates m and n cancels out. Again, because of the strong potentials of the hard variables (and assuming that the bond lengths, bond angles, and dihedral angles are uncorrelated) one can carry out the integration over the bond lengths and bond angles, and the remaining integral becomes a function of the $3N$ dihedral angles (α_k).[29,33,34] For our flexible model, the $3N$ bond angles and their Jacobian are also considered. An expression for the partition function that with small modification suits the two models is

$$Z' = DZ = D\int_{\Omega_0} \exp\{-E([\alpha_k])/k_B T\}d\alpha_1\cdots d\alpha_{6N}, \quad (1)$$

where $[\alpha_k] = [\alpha_1,\dots,\alpha_{6N}]$. For the rigid model the prefactor D is a product of J and the integral over the bond lengths and bond angles, and the integration in Eq. (1) is carried out only over the $3N$ dihedral angles α_k (i.e., $3N$ replaces $6N$). D depends on the absolute temperature T and the units in which the bond lengths and bond angles are expressed. For the flexible model α_k denotes the dihedral and bond angles, and the integration is carried out over all of them $(6N)$. D is a product of J of the bond lengths and the integral over these variables. The Jacobian $[\Pi_k \sin(\theta_k)]$ of the bond angles, θ_k that should appear under the integral is omitted for simplicity. For calculating $\Delta S_{m,n}$ and $\Delta F_{m,n}$ of two wide microstates of the same molecule, $\ln D$ cancels and can be ignored (notice, however, that D contributes to the absolute F and S). For these models the Boltzmann probability density corresponding to Z [Eq. (1)] is

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

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

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

It is easy to show that the fluctuation (standard deviation), σ_F, of F is zero, because the integrand, $E([\alpha_k]) + k_B T \ln \rho^B([\alpha_k]) = -kT\ln Z = F$, is constant for any set $[\alpha_k]$. This means that the free energy can be obtained from *any single* conformation if its energy and Boltzmann probability density are known. Using the HSMC method, we shall estimate the free energy of the rigid model of $(\mathrm{Gly})_{10}$ from single structures. Notice that for an approximate probability density the fluctuation is finite and it is expected to decrease as the approximation improves.[5,7,10,35]

C. Exact and approximate scanning procedures

The *exact* scanning method is a step-by-step construction procedure for a peptide.[36,37] Thus, an N-residue conformation of polyglycine in the helical region (Ω_0), for example, is built by defining the angles α_k step-by-step with transition probabilities (TPs) and adding the related atoms;[36] for example, the angle ϕ determines the coordinates of the

two hydrogens connected to C^α, and the position of C'. Thus, at step k, $k-1$ angles $\alpha_1,\dots,\alpha_{k-1}$ (i.e., for the flexible model α_k include dihedral and bond angles) have already been determined; these angles and the related structure (the past) are kept constant, and α_k should be defined with the exact TP density $\rho(\alpha_k|\alpha_{k-1}\cdots\alpha_1)$,

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

where $d\alpha_k$ is a small segment centered at α_k, and $Z_f(\alpha_k\cdots\alpha_1)$ is a future partition function defined over the helical region Ω_0 by integrating over the future conformations defined by $\alpha_{k+1}\cdots d\alpha_{6N}$ (within Ω_0) where the past angles, $\alpha_1\cdots\alpha_k$, are held fixed,

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

The probability density of the entire conformation is

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

One can also define an approximate scanning procedure in which only partial future based on f future angles is scanned, i.e., the integration in Eq. (6) is carried out with respect to $d\alpha_{k+1}\cdots d\alpha_f$, where $f < 6N$. The approximation introduced can be removed by importance sampling, which will not be discussed here (e.g., see discussion in Ref. 37 for self-avoiding walks on a lattice).

The exact scanning method is equivalent to the MC and MD procedures in the sense that large samples generated by all these methods lead to the same averages and fluctuations within the statistical errors. Thus, one can assume that a given MC 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 can be implemented 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)] has been carried out thus far only approximately by both the HS and LS methods. As described later, the HSMC method enables one to scan the future partially as well as *completely*. Because some elements of the LS method are implemented within the framework of the HSMC method we describe the LS method first.

D. The LS method

In the first step the MC sample (of a given wide microstate) is visited and the variability range $\Delta\alpha_k$ is calculated, where α_k are the dihedral and bond angles, $1 \leq k \leq 6N$ (in the case of constant bond angles only the $3N$ dihedral angles are considered),[19–22,27,28]

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

where $\alpha_k(\max)$ and $\alpha_k(\min)$ are the maximum and minimum values of α_k found in the sample, respectively. Next, the ranges $\Delta\alpha_k$ are divided into l equal segments, where l is the discretization parameter. We denote these segments by v_k, $(v_k=1,l)$. Thus, an angle α_k is now represented by the segment v_k to which it belongs and a conformation i is expressed by the corresponding vector of segments $[v_1(i), v_2(i), ..., v_{6N}(i)]$. Under this discretization approximation $\rho(\alpha_k|\alpha_{k-1}\cdots\alpha_1)$ can be estimated by

$$\rho(\alpha_k|\alpha_{k-1}\cdots\alpha_1) \approx n(v_k, \cdots, v_1) / \{n(v_{k-1}, ..., v_1)$$
$$\times [\Delta\alpha_k/l]\}, \qquad (9)$$

where $n(v_k, ..., v_1)$ is the number of times the *local state* [i.e., the partial vector $(v_k, ..., v_1)$ representing $(\alpha_k, ..., \alpha_1)$] 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 v_k and the b angles preceding it along the chain, i.e., the vector $(v_k, v_{k-1}, ..., v_{k-b})$; b is called the correlation parameter. The sample is visited for the second time and for a given b one calculates the number of occurrences $n(v_k, v_{k-1}, ..., v_{k-b})$ of all the local states from which a set of transition probabilities $p(v_k|v_{k-1}, ..., v_{k-b})$ 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

$$\rho_i(b,l) = \prod_{k=1}^{6N} p(v_k|v_{k-1}, ..., v_{k-b}) / (\Delta\alpha_k/l), \qquad (10)$$

the larger are b and l the better the approximation (for enough statistics). $\rho_i(b,l)$ allows defining an approximate entropy and free energy functional, S^A and F^A, which constitute *rigorous* upper and lower bounds for the correct values, respectively,[21]

$$S^A = -k_B \int \rho^B \ln \rho(b,l) d\alpha_1\cdots\alpha_{6N} \qquad (11)$$

and

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

where $\langle E \rangle$ is the Boltzmann average of the potential energy, estimated from the MC sample and ρ^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 $\bar{S}^A$,

$$\bar{S}^A = -\frac{k_B}{n} \sum_{t=1}^{n} \ln \rho_t(b,l). \qquad (13)$$

As discussed in Sec. II B,[5,7,10,35] the fluctuation (standard deviation) σ_F of the correct free energy is zero, while the approximate F^A has finite fluctuation, σ_A (estimated by $\overline{\sigma_A}$), which is expected to decrease as the approximation improves,

$$\overline{\sigma_A} = \left[\frac{1}{n} \sum_{t=1}^{n} [\bar{F}^A - E_t - k_B T \ln \rho_t(b,l)]^2 \right]^{1/2}. \qquad (14)$$

It should be noted that Eqs. (12)–(14) also hold for the HSMC procedures described later, where $\rho(b,l)$ is replaced by ρ^{HS}.

The LS method can be applied to any chain flexibility, i.e., it is not limited to harmonic or quasiharmonic fluctuations.[29,30,33,34,38] Thus, free energy differences between wide microstates with significant structural differences can be calculated, which is a difficult task with methods based on thermodynamic integration.

E. Approximate HS method

As discussed 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 α_k [Eq. (5)] by calculating the future partition functions Z_{future} [Eq. (6)]. However, a systematic integration of Z_{future} based on the *complete* future within the limits of Ω_0 is difficult and becomes impractical for a large peptide where Ω_0 is unknown; therefore, thus far, HS was applied to self-avoiding walks (SAWs) on a lattice, where Ω_0 is the entire space and Z_{future} is calculated approximately by enumerating only future SAWs of f steps (i.e., $\alpha_k, ...\alpha_{k+f-1}$), rather than the whole future of $N-k+1$ steps.[2,4–6] The approximate HS method was also applied to the freely-jointed chain.[3]

F. The HSMC method

The idea of the complete HSMC method is to obtain the TPs [Eq. (5)] from MC simulations of the future part of the chain rather than by evaluating the integrals defining Z_{future} [Eq. (6)]. Thus, at reconstruction step k of conformation i the TP density, $\rho(\alpha_k|\alpha_{k-1}\cdots\alpha_1)$ is calculated from n_f MC steps (trials),[16] where the *entire* future of the peptide can move by changing the future angles $\alpha_k, ..., \alpha_{6N}$ while the angles $\alpha_1, ..., \alpha_{k-1}$ 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 α_k and the number of MC visits to this bin during the simulation, n_{visit}, is calculated; one obtains,

$$\rho(\alpha_k|\alpha_{k-1}\cdots\alpha_1) \approx n_{visit} / [n_f \delta\alpha_k], \qquad (15)$$

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 Ref. 10). The product of these TP densities leads to the probability density of the entire chain [Eq. (7)]. Notice that unlike the deterministic calculation of Z_{future} [Eq. (6)], where the limits of Ω_0 are in practice unknown, with the complete HSMC procedure, the future structures generated by MC at each step k remain in general within the limits of the wide microstate Ω_0 defined by the analyzed MC sample. In some cases, however, the future samples were found to escape from this region; therefore, before applying the HSMC method, the LS method is applied to the analyzed sample and the $\alpha_k(\min)$ and $\alpha_k(\max)$ values [Eq. (8)] are calculated; they are then

used to keep the future structures within Ω_0 by rejecting MC moves with angle values beyond those of $\alpha_k(\min)$ and $\alpha_k(\max)$.

While complete HSMC considers the entire future, in practice $\rho(\alpha_k|\alpha_{k-1}\cdots\alpha_1)$ [Eq. (15)] 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, the corresponding probability density {approximating ρ^B [Eq. (7)]} will be denoted by $\rho^{\text{HS}}([\alpha_k])$, which defines approximate entropy and free energy functionals, S^A and F^A, where $\rho^{\text{HS}}([\alpha_k])$ replacing $\rho(b,l)$ in Eqs. (11) and (12), respectively. S^A and F^A are expected to overestimate and underestimate, respectively the correct values, where the fluctuation of F^A, σ_A [Eq. (14)] does not vanish, but decreases as the approximation improves, i.e., as n_f increases and/or $\delta\alpha_k$ decreases.

We shall also study a version of HSMC that was originally applied to argon,[8] where only f future angles are simulated by MC during the reconstruction procedure; the corresponding probability density will be denoted by $\rho^{\text{HS}}([\alpha_k],f)$. Obviously, this method unlike complete HSMC is always approximate even for very large n_f and very small $\delta\alpha_k$ and therefore will be called here *approximate HSMC*.

G. 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,[2]

$$F^B = \int_{\Omega_0} \rho^{\text{HS}}([\alpha_k])[E + k_B T \ln \rho^{\text{HS}}([\alpha_k])]d\alpha_1 \cdots d\alpha_{6N}. \tag{16}$$

According to the free energy minimum principle,[39] $F^B \geq F$ [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 should be noted, however, that the above inequality is rigorously satisfied for complete HSMC while for *approximate* HSMC it can only be proven[2] that $F^B \geq F^A$. However, in several applications of the HS method the relation $F^B \geq F$ was found to hold where only partial future was considered; this can be verified if F^B decreases systematically as the approximation is improved (e.g., as f is increased). It is necessary to rewrite Eq. (16) such that F^B can be estimated by importance sampling from a (Boltzmann) sample of configurations generated with ρ^B (rather than ρ^{HS}). It has been shown that

$$F^B = \frac{\int_{\Omega_0} \rho^B [\rho^{\text{HS}} \exp(E/k_B T)(E + k_B T \ln \rho^{\text{HS}})]d\alpha_1 \cdots d\alpha_{6N}}{\int_{\Omega_0} \rho^B [\rho^{\text{HS}} \exp(E/k_B T)]d\alpha_1 \cdots d\alpha_{6N}}. \tag{17}$$

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. (17). 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 ρ^B and ρ^{HS} decreases exponentially (see discussion in Ref. 22).

With values for both F^A and F^B, their average, F^M, defined by

$$F^M = (F^A + F^B)/2, \tag{18}$$

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. 10 we have proven that the inequalities $F^A \leq F \leq F^B$ hold for the stochastic probabilities as well. We have also shown there that one can calculate F^A and F^B from a sample generated by n HSMC reconstructions of a *single* conformation and applied this to configurations of argon and TIP3P water molecules; this will be checked for polyglycine as well.

H. Exact expression for the free energy

As shown for fluids in Ref. 10, the denominator of F^B in Eq. (17) defines an exact expression for the partition function,

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

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

$$= \int_{\Omega_0} \rho^B \exp[F^{\text{HS}}/k_B T][d\alpha_k] \tag{19}$$

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

$$F^D = k_B T \ln\left(\frac{1}{Z}\right) = k_B T \ln\left[\int_{\Omega_0} \rho^B \exp[F^{\text{HS}}/k_B T][d\alpha_k]\right], \tag{20}$$

where $[d\alpha_k] = d\alpha_1 \cdots d\alpha_{6N}$ and $F^{\text{HS}}/k_B T = E[\alpha_k]/k_B T + \ln \rho^{\text{HS}}[\alpha_k]$.

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^{HS} exponentiated. Obviously, as $F^{\text{HS}} \to F$ (i.e., $\rho^{\text{HS}} \to \rho^B$) all fluctuations become zero and F can be obtained from a single configuration [see discussion following Eq. (4) and Ref. 10]. 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. It should be noticed that for approximate HSMC, where f is small ρ^{HS} might be defined over a region in conformational space that is larger than Ω_0; in this case F^D is not exact.

TABLE I. The differences (in deg) between the minimum and maximum values of the dihedral angles [Eq. (8)] of $(Gly)_{10}$ for the rigid and flexible models in six MC samples.[a]

No.	Extended						Helix						Hairpin					
	$\Delta\varphi$		$\Delta\psi$		$\Delta\omega$		$\Delta\varphi$		$\Delta\psi$		$\Delta\omega$		$\Delta\varphi$		$\Delta\psi$		$\Delta\omega$	
1	47	53	142	117	23	25	43	43	48	117	23	40	35	49	57	114	21	35
2	61	68	55	49	23	26	20	38	43	45	21	25	37	78	30	52	21	28
3	57	72	43	59	23	28	28	32	33	54	22	28	37	60	36	42	26	28
4	68	70	51	60	26	31	22	36	25	55	21	26	54	55	89	30	21	27
5	58	66	46	59	25	28	30	67	35	59	19	28	59	26	65	77	24	31
6	68	89	46	52	24	32	25	33	29	61	20	26	31	100	27	54	18	26
7	61	66	45	52	23	23	22	34	46	50	16	31	30	80	43	40	32	30
8	66	62	42	51	25	25	27	40	34	44	19	24	39	41	31	40	22	26
9	60	77	54	48	25	25	30	42	48	39	19	28	41	53	37	53	23	30
10	65	63	47	52	26	28	49	63	360	47	26	27	56	66	32	41	28	25

[a]For each angle, the first and second entries are for the rigid and flexible models, based on samples of 700 and 600 conformations, respectively.

I. The quasiharmonic approximation

With the quasiharmonic approximation[29,30] the entropy, S_{QH} is given by,

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

where σ is the determinant of the covariance matrix of the $6N$ dihedral and bond angles. For the model with constant bond angles $3N$ replaces $6N$ and the matrix is correspondingly smaller.

III. RESULTS AND DISCUSSION

A. The rigid model of $(Gly)_{10}$

1. Simulation and computational details

In Ref. 11 we have already studied $(Gly)_{10}$ with constant bond angles in three wide microstates, helix, hairpin, and extended. Samples of these wide microstates were generated by the Metropolis MC procedure[16] at 100 K where a trial structure is obtained by randomly changing all the 30 dihedral angles, each within $\pm 1°$ of its current value. These simulations were started from helical, extended, and hairpin structures obtained by minimizing the energy of the corresponding structures, $\phi_k = \psi_k = -55°$, and $\omega_k = 180°$, $\phi_k = \psi_k = \omega_k = 180°$, and two extended strands of four residues connected by a type I′ turn. The first 5000 steps were used for equilibration and then 50 000 MC steps were carried out for each microstate where after every 100 steps the current structure was retained for future analysis; in this way three equal samples of 500 structures were generated. It should be pointed out that preliminary simulations at 300 K resulted in unstable samples (i.e., the structures escaped from their wide microstates); therefore, the temperature was decreased to 100 K, where the helix and extended simulations were found to be very stable, while the hairpin sample remained stable only up to the first 50 000 MC steps. The corresponding $\Delta\alpha_k$ values [Eq. (8)] are relatively small (see Table I), representing relatively concentrated samples. Notice, however, that due to correlations each wide microstate is significantly smaller than the corresponding region, $\Delta\alpha_1 \times \Delta\alpha_2 \times \cdots \times \Delta\alpha_{30}$.

In the present study each of these samples was increased to 700 conformations, using the same MC procedure described above but with an additional restriction that is aimed at keeping these samples within the corresponding microstates. Thus, if an angle α_k of a trial set of dihedral is larger than $\alpha_k(\max)$ or smaller than $\alpha_k(\min)$ it is rejected; we shall refer to this restriction as *the geometrical restriction*. Results were calculated with complete HSMC method (as in Ref. 11) and also with the approximate HSMC procedure based on partial futures of $f = 3$ and 4 dihedral angles.

The TPs and their product, ρ^{HS} [Eqs. (7) and (15)] were calculated by reconstructing each conformation step-by-step with MC simulations of the future part, where the geometrical restriction described above is applied as well. To check the convergence of the results they were calculated for four future sample sizes, $n_f = 20\,000$, $40\,000$, $80\,000$, and $160\,000$, generated by retaining a conformation every 10 MC steps, and for four bin sizes, $\delta = \Delta\alpha_k/15$, $\Delta\alpha_k/10$, $\Delta\alpha_k/5$, and 20° centered at α_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 ($\delta = \Delta\alpha_k/5$); the same is applied to the second size bin. In the case of zero counts, n_{visit} is taken to be 1; notice, however that zero counts is a very rare event.

2. Results for the entropy

Results for the entropy (TS) appear in Table II for various n_f values but only for two bin sizes, which is, however, sufficient for checking convergence. All the HSMC results are based on samples of 700 structures and the statistical errors were obtained from the fluctuations and results obtained for partial samples. The accuracy of HSMC can always be improved by decreasing the bin size and increasing the future sample size, meaning that correspondingly $S^A(\rho^{HS})$ [Eq. (11)] is expected to decrease [provided that ρ^{HS} is defined on the same conformational space (i.e., the wide microstate) that was generated by the MC simulation. Indeed, for each bin the entropy decreases (or remain constant) as n_f increases, where the only exception is the entropy for the helix based on the smallest sample, $n_f = 20\,000$, which is slightly smaller than the entropies of the larger samples; this probably stems from a HSMC probability density that is de-

TABLE II. Entropy TS^A ($T = 100$ K) in kcal/mol [Eq. (11)] for the rigid model of $(Gly)_{10}$ for various bin sizes [Eq. (5)] and future sample sizes, n_f, obtained with the HSMC method for different approximations, f.[a]

Bin size	n_f	Extended	Helix	Hairpin
		f=entire future		
$\Delta\alpha_k/10$	20 000	20.23 (5)	16.06 (3)	18.18 (7)
$\Delta\alpha_k/10$	40 000	20.10 (4)	16.13 (2)	18.06 (8)
$\Delta\alpha_k/10$	80 000	20.02 (2)	16.14 (2)	17.9 (1)
$\Delta\alpha_k/10$	160 000	19.99 (3)	16.15 (2)	17.8 (2)
$\Delta\alpha_k/15$	20 000	20.21 (4)	15.98 (2)	18.15 (8)
$\Delta\alpha_k/15$	40 000	20.10 (4)	16.07 (3)	18.05 (8)
$\Delta\alpha_k/15$	80 000	20.02 (2)	16.08 (2)	17.9 (1)
$\Delta\alpha_k/15$	160 000	19.99 (3)	16.09 (2)	17.8 (2)
		f=4 dihedrals		
$\Delta\alpha_k/10$	20 000	20.20 (3)	16.10 (2)	18.4 (2)
$\Delta\alpha_k/10$	40 000	20.08 (5)	16.18 (2)	18.4 (4)
$\Delta\alpha_k/10$	80 000	20.01 (3)	16.19 (2)	18.4 (4)
$\Delta\alpha_k/10$	160 000	19.97 (2)	16.19 (2)	18.4 (4)
$\Delta\alpha_k/15$	20 000	20.19 (4)	16.03 (2)	18.3 (3)
$\Delta\alpha_k/15$	40 000	20.08 (5)	16.12 (2)	18.4 (3)
$\Delta\alpha_k/15$	80 000	20.01 (4)	16.14 (2)	18.4 (4)
$\Delta\alpha_k/15$	160 000	19.97 (2)	16.13 (2)	18.4 (4)
		f=3 dihedrals		
$\Delta\alpha_k/10$	20 000	20.18 (6)	16.27 (5)	18.7 (2)
$\Delta\alpha_k/10$	40 000	20.10 (6)	16.39 (5)	18.6 (2)
$\Delta\alpha_k/10$	80 000	20.01 (4)	16.38 (5)	18.6 (2)
$\Delta\alpha_k/10$	160 000	19.98 (3)	16.37 (7)	18.5 (2)
$\Delta\alpha_k/15$	20 000	20.16 (6)	16.20 (5)	18.6 (2)
$\Delta\alpha_k/15$	40 000	20.09 (6)	16.34 (6)	18.6 (2)
$\Delta\alpha_k/15$	80 000	20.01 (4)	16.33 (7)	18.6 (2)
$\Delta\alpha_k/15$	160 000	19.98 (3)	16.32 (8)	18.5 (2)
TS_{QH}		19.86 (4)	16.15 (3)	17.79 (4)
TS_{LS}		19.98 (3)	17.50 (3)	19.29 (3)

[a]$\Delta\alpha_k$ is defined in Eq. (8). The HSMC results are based on a sample of 700 conformations. f is the number of dihedral angles considered in the future MC sampling. The statistical errors are given in parentheses, e.g., 19.96 (3) = 19.96±0.03. S_{QH} is the quasiharmonic entropy [Eq. (21)] and S_{LS} [Eqs. (11) and (13)] is the local states (LS) entropy obtained for $b = 1$ and $l = 10$. The entropy is defined up to an additive constant.

fined on only a partial region of the helical wide microstate due to insufficient sampling, i.e., the bin is overpopulated.

The entropy for the extended microstate is shown to converge, where its values for $n_f = 80\,000$ and 160 000 are the same within the error bars for *all* the approximations (i.e., f=[entire future], f=4 and 3) and for the two bin sizes. This perfect convergence (that also includes the not-shown results for the approximations f=2 and $\Delta\alpha_k/5$) reflects the short-range angular correlations characterizing an extended structure. In other words, for the extended microstate the approximation based on $n_f = 80\,000$, f=2, and $\Delta\alpha_k/5$ leads to the exact result within the statistical errors.

The range of interactions in a helix is larger than in the extended structure due to the 1–4 hydrogen bonds, which is reflected in a less than perfect convergence, and an increase (even though not large) in the entropy values as f is decreased, i.e., as the approximation worsens. Thus, for each bin the results are the same for the three largest n_f values, while they decrease (improve) slightly by 0.06 kcal/mol with decreasing the bin size and probably are not yet completely converged. However, this difference is smaller than the statistical errors, and within the accuracy of the usual force fields, entropy and free energy differences smaller than 0.1 kcal/mol are in general ignored; therefore, even the helix results can be considered as converged. Notice that the f =4 results constitute a very good approximation as they are larger than the results for f=[entire future] only by 0.06 kcal/ mol, while the f=3 results deviate more significantly by ∼0.22 kcal/mol.

Unlike the helix, results for the hairpin pertaining to the same n_f in the two bins are equal, meaning that smaller bins are not required. However, for f=[entire future] the results for a given bin decrease as n_f increases with relatively large statistical errors such that 17.9 (1) and 17.8 (2) obtained for $n_f = 80\,000$ and 160 000, respectively, can be considered as equal. For the approximations f=4 and 3 the entropy converges, again with large statistical errors of up to 0.4 and 0.2 kcal/mol, respectively. These inferior results of the hairpin stem from the maximal range of interactions characterize this structure, where the first and last residues are in closed proximity. Correspondingly, the best results for f=4 and 3 are higher (by 0.6 and 0.7 kcal/mol, respectively) than those obtained for f=[entire future]. Therefore, such approximations are suitable for structures based on short or medium range interactions and should be used with caution when long-range interactions exist.

It is of interest to compare the HSMC results to those obtained by other methods. For that we increased the samples of the three wide microstates from 500 to 30 000 structures by imposing the geometrical restriction (related to $\Delta\alpha_k$) on the MC procedure. We applied the quasiharmonic (QH) approximation[29,30] to a subsample of 8000. We also applied the LS method (with correlation parameter, $b = 1$ and $l = 10$) to the entire increased sample. The QH results presented in Table II are very close to the complete HSMC values (i.e., f=[entire future]), probably because the three samples are approximately quasi-harmonic. The LS and HSMC entropies are equal for the extended microstate because the angular correlations along the chain are short and $b = 1$ already captures most of them. On the other hand, the range of these correlations increases for the helix and the hairpin, and the LS entropies, as expected, become slightly larger (upper bounds) than the HSMC values.

3. Results for the free energy

In Table III HSMC results are presented for the free energy functionals $F^A(\rho^{HS})$ [Eq. (12)], F^B, [Eqs. (16) and (17)], F^M [Eq. (18)], and F^D [Eq. (20)] obtained for the rigid model of $(Gly)_{10}$ for f=[entire future], f=4 and 3. These results are given only for the smallest bin, because F^A values for the bin, $\Delta\alpha_k/10$ can be obtained from the entropies of Table II and the energies provided in the bottom of Table III. Also, we have checked that the results for F^B in bin $\Delta\alpha_k/10$ have the expected trend, i.e., they decrease as n_f increases and their values are not smaller than those of bin, $\Delta\alpha_k/15$ within the statistical errors. F^A follows the trends of S^A in Table II, i.e., in general it increases as the approximation improves besides for the cases where S^A increases or remains constant, as discussed in Sec. III A 2; in particular, for the approximations f=4 and 3 of the hairpin the tendencies of the F^A results cannot be observed due to large statistical

054903-8   S. Cheluvaraja and H. Meirovitch

TABLE III. HSMC results for various free energy functionals obtained for the rigid model of $(Gly)_{10}$.[a]

HSMC/n_f	Extended			Helix			Hairpin		
	$-F^A$	$-F^B$	σ_A	$-F^A$	$-F^B$	σ_A	$-F^A$	$-F^B$	σ_A
				f=entire future					
20 000	74.76 (3)	73.50 (7)	0.62 (2)	98.49 (2)	97.1 (2)	0.46 (2)	84.59 (4)	82.8 (2)	0.84 (1)
40 000	74.65 (3)	73.98 (5)	0.40 (2)	98.58 (2)	97.6 (2)	0.30 (4)	84.49 (3)	83.2 (4)	0.64 (3)
80 000	74.57 (3)	74.20 (4)	0.25 (2)	98.59 (1)	97.9 (2)	0.21 (4)	84.33 (3)	83.4 (4)	0.44 (3)
160 000	74.54 (2)	74.35 (3)	0.17 (2)	98.60 (1)	98.0 (2)	0.17 (1)	84.25 (3)	83.4 (4)	0.33 (3)
$-F^D$ $-F^M$	74.46 (2)	74.45 (3)		98.39 (8)	98.3 (1)		83.9 (1)	83.8 (2)	
				f=4 dihedrals					
20 000	74.74 (4)	73.50 (8)	0.60 (2)	98.54 (2)	96.9 (4)	0.48 (3)	84.8 (4)	83.0 (9)	0.79 (3)
40 000	74.63 (4)	74.07 (5)	0.40 (2)	98.63 (2)	97.8 (2)	0.32 (4)	84.9 (4)	83.1 (4)	0.67 (3)
80 000	74.56 (3)	74.23 (5)	0.25 (2)	98.64 (2)	97.8 (2)	0.25 (4)	84.9 (4)	83.7 (4)	0.61 (3)
160 000	74.52 (2)	74.35 (4)	0.16 (2)	98.64 (2)	97.8 (2)	0.21 (4)	84.9 (4)	83.8 (4)	0.59 (4)
$-F^D$ $-F^M$	74.32 (2)	74.43 (3)		98.24 (8)	98.2 (1)		83.9 (1)	84.2 (4)	
				f=3 dihedrals					
20 000	74.71 (5)	73.3 (1)	0.59 (2)	98.71 (2)	96.7 (5)	0.55 (2)	85.1 (4)	82.9 (9)	1.09 (6)
40 000	74.64 (4)	74.00 (6)	0.42 (2)	98.84 (3)	97.5 (3)	0.43 (1)	85.1 (4)	83.2 (4)	1.01 (4)
80 000	74.56 (3)	74.09 (5)	0.26 (2)	98.83 (4)	98.0 (2)	0.36 (4)	85.0 (4)	83.1 (4)	0.91 (3)
160 000	74.53 (2)	74.36 (3)	0.17 (2)	98.83 (5)	98.1 (2)	0.32 (1)	85.0 (4)	83.2 (4)	0.82 (1)
$-F^D$ $-F^M$	74.32 (2)	74.45 (3)		98.30 (7)	98.5 (1)		83.8 (1)	84.1 (4)	
				Single conformations; f=entire future					
Conf. 1	74.53 (2)	74.43 (5)	0.135 (2)	98.65 (2)	98.61 (2)	0.088 (2)	84.31 (4)	84.0 (2)	0.318 (3)
Conf. 2	74.53 (2)	74.45 (5)	0.124 (2)	98.52 (2)	98.49 (2)	0.071 (2)	84.12 (4)	83.9 (2)	0.205 (3)
				QH and LS methods					
$-F_{QH}$	74.50 (4)			98.72 (5)			84.66 (4)		
$-F_{LS}$	74.68 (1)			100.08 (1)			86.14 (1)		
-Energy	54.55 (2)		0.72 (2)	82.51 (3)		0.77 (4)	66.44 (8)		0.67 (5)

[a]F^A [Eq. (12)] and F^B [Eqs. (16) and (17)] are lower and upper bounds of the free energy, respectively, and σ_A [Eq. (14)] is the fluctuation of F^A. These HSMC results were obtained from samples of 700 conformations for three approximations defined by the number, f of future dihedral angles considered. These results are presented only for the smallest bin size, $\delta = \Delta\alpha_k/15$, but for all future sample sizes n_f. The results for F^M [Eq. (18)]—the average of F^A and F^B, and for F^D [Eq. (20)]—the exact free energy functional, are calculated for $\delta = \Delta\alpha_k/15$ and n_f= 160 000 only. F_{QH} [Eq. (21)] and F_{LS} [Eq. (12)] are free energies obtained by the quasiharmonic approximation and the local states method, respectively, and are based on larger samples (see text). The average energies of the HSMC samples appear 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.

errors. The fluctuations of F^A, σ_A, as expected, always decrease as the approximation improves, i.e., as the bin size is decreased (not shown) and n_f is increased. For f=[entire future] and n_f= 160 000 σ_A is smaller than the energy fluctuations (that are also provided) by 4.2, 4.5, and 2 for the extended, helix, and hairpin, respectively, where for the extended microstate this ratio holds also for f=4 and 3. For the helix and hairpin, on the other hand, σ_A increases as f is decreased (due to the worsening approximation), and the corresponding ratios for the best results of f=3 decrease to 2.4 and 0.8.

Of special interest are the results for F^B, which for f=[entire future] are expected to provide an upper bound, while for a partial future scanning can only be shown to overestimate F^A. For the extended microstate the F^B values decrease systematically as n_f is increased (even for f=4 and 3) meaning that they are upper bounds. The best results deviate by $\Delta = F^B - F^A = -74.35 + 74.54 = 0.19$ kcal/mol. The F^B results for the larger bin, $\Delta\alpha_k/10$ (not shown) are comparable to the corresponding values for the smallest bin ($\Delta\alpha_k/15$) presented in Table III. A similar behavior of F^B is observed also for a partial future sampling based on f=2 (results not shown). For each approximation f of the helix,

F^B is shown to decrease monotonically as n_f is increased, with errors that are significantly larger than those obtained for the extended microstate, and larger Δ (0.6 kcal/mol) with a higher error bars. Again, for each f, the F^B results for the bin $\Delta\alpha_k/10$ (not shown) are comparable to the corresponding results of Table III. For the hairpin—the most difficult structure to handle—F^B behaves as an upper bound only for f=[entire future], the best approximation, and the statistical errors are larger than those obtained for the helix and extended microstates.

For n_f= 160 000 of the three approximations, f=[entire future], f=4, and 3, we have calculated results for F^D [Eq. (20)], and for comparison also for F^M [Eq. (18)], the average of F^A and F^B. For f=[entire future] the F^D results are equal within the error bars to the corresponding F^M values. These results for F^D are considered to be exact within the error bars. For the helix and hairpin, the F^D results for f=4 and 3 are equal to those of f=[entire future] within relatively large statistical errors and they are close to their F^M counterparts; no change of F^M for different f can be observed due again to large statistical errors. However, the errors for the extended results for f=4 and 3 are small, which allows detecting tendencies of F^D. Thus, for f=4 and 3 $F^D = -74.32$ is slightly

larger than $F^B = -74.35$ and $F^D = -74.46$ obtained for $f=$[entire future].

In accordance with the entropy, the QH results slightly underestimate the F^D values for $f=$[entire future], where the deviation increases in going from the extended to the helix and to the hairpin (with deviation of 0.8 kcal/mol for the latter). The LS result for the extended microstate is lower by 0.22 kcal/mol than the F^D value and this deviation increases to 1.7 and 2.2 kcal/mol for the helix and hairpin, respectively, due to increase in correlations. Notice, however, that part of the above deviations stem from the energy components of F_{QH} and F_{LS} that are calculated from much larger samples than the sample size of 700 used for calculating F^D.

4. Reconstructions of single conformations

As in Ref. 10, we demonstrate here that the free energy (but not the entropy) functionals can be obtained by carrying out n reconstructions of a single conformation (rather than one reconstruction for each of n different conformations). Thus, from the sample of each microstate we have chosen two conformations with energies within a 0.5 kcal/mol of the average energy obtained from the sample of 700 conformations (which appear in the bottom of Table III). Each of these conformations was then reconstructed $n=100$ times using $f=$[entire future]. Indeed, the results for F^A and F^B obtained for this relatively small n are already close to those obtained from the whole sample of size 700 using $f=$[entire future], where the results for F^B and the fluctuations are systematically lower than those obtained for the whole sample based on $f=$[entire future]. For the three configurations defined as conf. 2 in Table III we also calculated F^D and obtained -74.49 (3), -98.51 (3), and -84.03 (4) kcal/mol for the extended, helix, and hairpin, respectively, which are equal within the error bars to the F^D results obtained from the whole samples of 700 conformations using $f=$[entire future]; this constitutes another check for the reliability of the single conformations results. As discussed in Ref. 10, the ability to obtain the free energy from a single structure will be useful when extremely small samples are available for analysis.

5. Differences in S and F of the three microstates

The main interest in this study is to determine the relative stability of the three wide microstates. In the upper parts of Table IV we present results for the differences, $T\Delta S^A$, ΔF^A, and ΔE between the helix, extended, and hairpin microstates for the rigid model of $(\mathrm{Gly})_{10}$. For $f=$[entire future] the results are the same as those obtained in Ref. 11 (from a samples of 400 structures) which demonstrates that they are statistically reliable. Moreover, within their uncertainty of 0.1–0.2 kcal/mol, the differences are not changed also for the three bin sizes, for $n_f=40\,000-160\,000$, for samples as small as 200 conformations, and for the helix-extended differences also for 100 conformations. This demonstrates that in practice HSMC can be quite efficient. For the model studied the helix is the most stable, where its free energy is lower by 14.3 and 24.1 kcal/mol than that of the hairpin and extended microstates, respectively. These differences are mostly governed by the energy differences, 11.9

TABLE IV. Differences in the entropy, $T\Delta S^A$, the free energy, ΔF^A and the energy, ΔE, between the three wide microstates (these properties are denoted R).[a]

	$T\Delta S^A$	ΔF^A	ΔE
$(\mathrm{Gly})_{10} f=$entire future			
R(extended)-R(hairpin)	2.2 (2)	9.7 (1)	11.9 (1)
R(extended)-R(helix)	3.9 (1)	24.1 (1)	27.95 (6)
R(hairpin)-R(helix)	1.7 (2)	14.3 (1)	16.1 (1)
$(\mathrm{Gly})_{10} f=4$ dihedrals			
R(extended)-R(hairpin)	1.6 (3)	10.3 (2)	11.9 (1)
R(extended)-R(helix)	3.8 (1)	24.1 (1)	27.95 (6)
R(hairpin)-R(helix)	2.3 (3)	13.8 (2)	16.1 (1)
$(\mathrm{Gly})_{10} f=3$ dihedrals			
R(extended)-R(hairpin)	1.5 (1)	10.4 (2)	11.9 (1)
R(extended)-R(helix)	3.7 (1)	24.3 (1)	27.95 (6)
R(hairpin)-R(helix)	2.2 (2)	13.9 (2)	16.1 (1)
$(\mathrm{Gly})_{16} f=$entire future			
R(extended)-R(hairpin)	6.2 (3)	36.5 (3)	42.7 (4)
R(extended)-R(helix)	7.1 (2)	55.7 (2)	62.9 (3)
R(hairpin)-R(helix)	1.6 (2)	19.2 (2)	20.1 (3)
$(\mathrm{Gly})_{10}$ flexible model $f=$entire future			
R(extended)-R(hairpin)	3.0 (3)	20.1 (3)	23.1 (3)
R(extended)-R(helix)	4.0 (3)	36.1 (3)	40.2 (2)
R(hairpin)-R(helix)	1.0 (2)	16.0 (2)	17.0 (2)

[a]The four sets of results at the top are for the rigid model (constant bond lengths and bond angles); the set at the bottom is for the flexible model. These HSMC results (obtained at $T=100$ K) are in kcal/mol. The statistical error is defined in Table II.

and 27.95 kcal/mol, where the $T\Delta S^A$ values are only 1.7 and 3.9 kcal/mol, respectively. We have also calculated the values of ΔF^D and estimated their errors from differences calculated for samples of increasing size and obtained 9.5 (1), 23.9 (1), and 14.5 (2) kcal/mol, which are equal within the error bars to the corresponding values those in Table IV.

As expected (see discussion for Table II) for $f=4$ and 3 the entropy differences (hence ΔF^A) between the helix and extended microstates are equal within the error bars to the value obtained with $f=$[entire future]. However, the S^A results for the hairpin increase in going from $f=$[entire future] to $f=4$ and 3 (see Table II), hence the corresponding values of ΔS^A and ΔF^A in Table IV involving the hairpin differ from those obtained for $f=$[entire future]. This demonstrates that for a partial future sampling to be reliable the value of f should cover the range of correlations along the chain, which in the case of the hairpin include the entire chain; therefore, for the hairpin the larger is f the better the approximation.

B. The rigid model of $(\mathrm{Gly})_{16}$

It is of interest to test the performance of HSMC for larger peptides; therefore in Ref. 11 we also applied it to the rigid model of $(\mathrm{Gly})_{16}$. Two samples of size 600 each spanning the extended and helical wide microstates were generated by MC [as described for $(\mathrm{Gly})_{10}$], where 400 and 600 conformations of them were reconstructed by complete HSMC, respectively. In the present paper we simulate $(\mathrm{Gly})_{16}$ in the hairpin state, and obtain the entropy from the sample by reconstructing each of its 600 conformations; we

TABLE V. HSMC results for the upper bound of the entropy, TS^A [Eq. (11)] and various free energy functionals for the rigid model of (Gly)$_{16}$.[a]

Bin size	n_f			Extended TS^A		Helix TS^A		Hairpin TS^A	
$\Delta\alpha_k/10$	20 000			32.9 (3)		24.54 (4)		25.4 (3)	
$\Delta\alpha_k/10$	40 000			32.3 (2)		24.73 (5)		25.9 (2)	
$\Delta\alpha_k/10$	80 000			32.07 (6)		24.74 (5)		25.9 (2)	
$\Delta\alpha_k/10$	160 000			31.94 (3)		24.73 (5)		25.7 (3)	
$\Delta\alpha_k/15$	20 000			32.9 (3)		24.50 (4)		25.4 (3)	
$\Delta\alpha_k/15$	40 000			32.4 (2)		24.72 (5)		25.9 (2)	
$\Delta\alpha_k/15$	80 000			32.08 (7)		24.74 (5)		25.9 (2)	
$\Delta\alpha_k/15$	160 000			31.94 (3)		24.73 (5)		25.7 (3)	
S_{QH}				33.20 (6)		26.67 (3)		27.0 (1)	
S_{LS}				32.80 (4)		26.90 (6)		27.62 (3)	
		$-F^A$	σ_A	$-F^A$	$-F^B$	σ_A	$-F^A$	$-F^B$	σ_A
$\Delta\alpha_k/15$	20 000	100.4 (3)	1.28 (3)	154.81 (3)	152.71 (3)	0.69 (7)	135.6 (2)	132.2 (9)	0.9 (5)
$\Delta\alpha_k/15$	40 000	99.8 (2)	0.76 (5)	155.03 (2)	153.13 (4)	0.47 (6)	136.0 (1)	133.7 (6)	0.74 (7)
$\Delta\alpha_k/15$	80 000	99.51 (8)	0.51 (6)	155.05 (2)	153.60 (6)	0.33 (4)	136.02 (5)	134.7 (3)	0.55 (7)
$\Delta\alpha_k/15$	160 000	99.37 (4)	0.37 (5)	155.03 (1)	153.80 (8)	0.24 (2)	135.90 (3)	134.9 (3)	0.43 (5)
$-F^D -F^M$		99.0 (1)	98.9 (8)	154.6 (2)	154.4 (5)		135.4 (1)	135.4 (2)	
$-F_{QH}$		100.73 (6)		157.40 (4)			137.9 (1)		
$-F_{LS}$		100.34 (4)		157.60 (7)			138.4 (1)		
-Energy		67.43 (2)	0.94 (10)	130.31 (8)		1.13 (10)	110.2 (4)		0.9 (2)

[a]F^A [Eq. (12)] and F^B [Eqs. (16) and (17)] are lower and upper bounds of the free energy, respectively, and σ_A [Eq. (14)] is the fluctuation of F^A. These *complete* HSMC results were obtained from samples of 600 conformations for f=[entire future]. These results are presented only for the smallest bin size, $\delta=\Delta\alpha_k/15$, but for all future sample sizes n_f. The results for F^M [Eq. (18)]—the average of F^A and F^B, and for F^D [Eq. (20)]—the exact free energy functional, are calculated for $\delta=\Delta\alpha_k/15$ and $n_f=160\,000$ only. F_{QH} [Eq. (21)] and F_{LS} [Eq. (12)] are free energies obtained by the quasiharmonic approximation and the local states method, respectively, and are based on larger samples (see text). All free energies and TS^A (at $T=100$ K) are in kcal/mol and are defined up to an additive constant. The average energies of the HSMC samples appear in the bottom row. The statistical error is defined in the caption of Table II.

also provide new results for the extended microstate based on an increased sample size of 600. The dihedral angle values for the helix and extended samples are concentrated around their canonical values with deviations $\Delta\alpha_k$ [Eq. (8)] very close to those obtained for (Gly)$_{10}$ in Table I, where significant differences exist only for $\Delta\psi$ of the first and last residues; on the other hand, for (Gly)$_{16}$ the α_k values for the hairpin are more concentrated than those obtained for (Gly)$_{10}$, where the corresponding ranges of $\Delta\psi$, for example, are 23.5°–70.4° versus 31°–89°.

1. Results for the entropy and free energy

Results obtained with f=[entire future] for TS^A, F^A, its fluctuation σ_A, F^B, F^D, and F^M appear in Table V, which is structured as a combined Tables II and III. The table reveals that in *all* cases the results for σ_A decrease as the approximation improves and they are smaller than the corresponding energy fluctuations by 2.5, 4.7, and 2.1 for the extended, helix, and hairpin microstates, respectively. These ratios for the helix and hairpin are slightly larger than the corresponding values for (Gly)$_{10}$, while the ratio for the extended microstate is significantly smaller than 4.2 obtained for (Gly)$_{10}$.

The TS^A results for the helix [unlike for (Gly)$_{10}$] show a perfect convergence with respect to both bin size and n_f [also for the results of bin $\Delta\alpha_k/5$ (not shown)] probably due to further stabilization caused by the formation of extra hydrogen bonds, meaning that $n_f=40\,000$ is sufficient. Notice that the smallest TS^A value for $n_f=20\,000$ (and for the hairpin, see below) stems from insufficient sampling, which

leads to poorly occupied bins. Correspondingly, the results for F^A show apparent convergence and those for F^B, while exhibiting the expected decrease as n_f increases, are not yet fully converged, meaning that the correct free energy is located within the range $F^B-F^A=1.13$ kcal/mol above F^A, probably close to the F^A value. Indeed, the value of $F^D[154.6(2)]$ is slightly smaller than that of F^M, the average of F^B and F^A. The fact that $F^A\neq F^D$ means that both functionals (or one of them) are not yet converged, i.e., the statistical errors are larger than those provided. It should be noted that the free energy, energy, and entropy per residue for the helix of (Gly)$_{16}$ and (Gly)$_{10}$ are close.

For (Gly)$_{16}$ the entropy of the extended microstate, which has shown a perfect convergence for (Gly)$_{10}$, is converged only with respect to bin size but for each bin the results decrease as n_f increases and n_f larger than 160 000 is needed to reach convergence; for $n_f=320\,000$ the expected extrapolated result is TS$^A=31.90$, which is used in calculating the differences for (Gly)$_{16}$ in Table IV. Also, the results for F^B do not decrease systematically with increasing n_f, and therefore they (and those for F^M) are not presented in Table V. This worsened convergence of the extended results probably stems from the decrease in the stability of the longer molecule, which is reflected by the increase of the energy per residue from −5.45 for (Gly)$_{10}$ to −4.2 kcal/mol for (Gly)$_{16}$. Another measure for the decrease in the accuracy of the free energy is the decrease in the ratio (energy fluctuation)/σ_A from 4.2 for (Gly)$_{10}$ to 2.5 for (Gly)$_{16}$. The F^D value, which is expected to be exact, is 0.37 kcal/mol higher than the best F^A result.

The energy and entropy (TS^A) per residue of the hairpin slightly decrease (by ~0.2 kcal/mol) in going from $(Gly)_{10}$ to $(Gly)_{16}$ which is in accord with the corresponding decrease in the $\Delta\alpha_k$ values previously discussed. The TS^A (hence F^A) results (excluding those for $n_f=20\,000$; see earlier discussion) are equal within a relatively large statistical error, and those of F^B show the expected decrease as n_f is increased but with statistical errors that are larger than those detected for F^A, where $F^B-F^A\approx1$ kcal/mol. For the hairpin $F^D=F^M$, and they are ~0.5 kcal/mol above F^A.

As expected, for all microstates the LS results ($b=1$, $l=10$) for TS overestimate the HSMC values, and the deviations are ~1, 2.2, and 2 kcal/mol for the extended, helix, and hairpin microstates, respectively. Notice that as in Ref. 11, the LS results of the first two microstates were obtained from MC samples of $5\cdot10^4$ structures generated with the geometrical restriction, where a conformation was retained every 10 MC steps; the sample size of the hairpin is $3\cdot10^4$, where a structure was retained every 100 MC steps. The QH results are also larger than the corresponding S^A (HSMC) values by 1.3, 2, and 1.3 kcal/mol, suggesting that anharmonic effects for $(Gly)_{16}$ are significantly larger than for $(Gly)_{10}$. The QH results for the extended and helix were obtained from samples of size 16 000 and the hairpin from size 32 000, where a structure was retained every 100 MC steps. In accordance with the entropy, the results for F_{QH} and F_{LS} underestimate the HSMC values. Notice, however, that the corresponding energies are slightly different from those based on the HSMC samples of 600 conformations.

2. Differences in S and F

In Table IV the differences, $T\Delta S^A$, ΔF^A, and ΔE for the extended and helix and hairpin microstates are presented with acceptable errors of 0.2–0.3 kcal/mol (see previous discussion). We have also calculated the corresponding values of ΔF^D and estimated their errors from differences calculated for samples of increasing size, obtaining 36.38 (2), 55.60 (4), and 19.22 (6) kcal/mol, which are equal to those in Table IV but have smaller statistical errors. It should be pointed out that the results for S^A and the energy of the helix and hairpin scale with increasing peptide size, while the energy of the extended state does not; therefore, ΔE, and ΔF^A do not scale in going from $(Gly)_{10}$ to $(Gly)_{16}$.

In summary, with HSMC calculating the TP density of each angle involves a certain error and therefore the total error in the entropy is expected to grow with increasing molecular size while using the same set of parameters (bin size, n_f, etc.). This effect is demonstrated by the worsening results of the extended microstate as the number of variables increases from 30 [$(Gly)_{10}$] to 48 [$(Gly)_{16}$]. For the helix and hairpin, on the other hand, this effect is compensated by the enhanced stability of these structures due to formation of additional hydrogen bonds; thus, the accuracy of the entropy for $(Gly)_{16}$ for the hairpin remains approximately the same as for [$(Gly)_{10}$] and it is even improved slightly for the helix.

C. $(Gly)_{10}$ with variable bond angles

We carried out three MC simulations of $(Gly)_{10}$ with flexible geometry, i.e., the bond angles are allowed to vary in addition to the dihedral angles, and the total number of variables is thus 60. The MC simulations (at 100 K) are carried out as described for the rigid model, but to take into account the Jacobian, trial bond angles are selected at random within the range $\cos[\theta_0(k)]\pm\delta$, where $\theta_0(k)$ is the current value of bond angle k ($k=1,3N$) and $\delta=0.001$; the chosen cosine values are then translated into bond angles through the arccosine function. MC simulations of the helix, extended, and hairpin microstates have yielded three samples of 600 structures each that have been reconstructed by complete HSMC and the corresponding entropy and free results are summarized in Tables IV and VI. Also, some information about these microstates is given in Table I where results for $\Delta\alpha_k$ [Eq. (8)] are presented for the rigid and flexible models. Table I reveals that for the three microstates the $\Delta\alpha_k$ values of the flexible model are larger in most cases than those of the rigid model, where the $\Delta\alpha_k$ values of the bond angles (not shown) range from 3° to 13°. While the bending energy of the flexible model is positive, the additional degrees of freedom allow further optimization of the nonbonded interactions (in particular the hydrogen bond energy) and the average energies of the helix and hairpin microstates of the flexible model are thus significantly lower (by ~12 kcal/mol) than those of the rigid model. This effect is much less pronounced for the extended microstate (its energy decreases by ~1.5 kcal/mol only), which lacks hydrogen bonds and its interactions are of a short-range character (compare the energy results in Tables III and VI).

1. Results for the entropy and free energy

The entropy was calculated without the Jacobian component [$\Pi_k\sin(\theta_k)$] because to a good approximation, this contribution is canceled out in differences in entropy and free energy which is our main interest. As discussed earlier for $(Gly)_{16}$ the increase in the number of variables from 30 for the rigid model to 60 for the flexible model is expected to increase the total error of the latter accumulated during the reconstruction process. Therefore, it would be required to decrease further the smallest bin size used for the rigid model, and the fact that the $\Delta\alpha_k$ values are larger for the flexible than for the rigid model suggests that the future sample size, n_f, should be enhanced as well.

However, for comparison with the rigid model, the entropy in Table VI was obtained with the same bin sizes and n_f values that were used for the rigid model and as expected, it decreases as the approximation improves. The errors are larger than those obtained for the rigid model and the convergence with respect to bin size and future sampling has not been fully reached. For both bins TS(80 000)−TS(160 000) ≈0.2 kcal/mol, and TS(160 000) decreases by ≈0.2 kcal/mol in going from $\Delta\alpha_k/10$ to $\Delta\alpha_k/15$. To check the effect of the future sampling we have carried out limited reconstruction runs for 100 conformations based on $n_f=40\,000-320\,000$. For the helix, hairpin and extended the differences TS(160 000)−TS(320 000) are 0.04, 0.10, and 0.15 kcal/mol, respectively, meaning that within our approximation

TABLE VI. HSMC results for the flexible model of $(Gly)_{10}$.[a]

Bin size	n_f	Extended TS^A	Helix TS^A	Hairpin TS^A
$\Delta\alpha_k/10$	20 000	30.3 (2)	25.1 (1)	26.31 (7)
$\Delta\alpha_k/10$	40 000	29.4 (3)	24.9 (1)	26.13 (8)
$\Delta\alpha_k/10$	80 000	28.9 (3)	24.7 (1)	25.78 (5)
$\Delta\alpha_k/10$	160 000	28.7 (3)	24.6 (1)	25.58 (7)
$\Delta\alpha_k/15$	20 000	30.1 (2)	25.0 (2)	26.17 (7)
$\Delta\alpha_k/15$	40 000	29.2 (3)	24.8 (1)	26.01 (7)
$\Delta\alpha_k/15$	80 000	28.7 (3)	24.6 (1)	25.64 (5)
$\Delta\alpha_k/15$	160 000	28.5 (3)	24.4 (1)	25.41 (7)
S_{QH}		30.28 (2)	26.65 (2)	27.84 (3)
S_{LS}		31.9 (2)	28.9 (2)	30.7 (2)

Bin size	n_f	$-F^A$	σ_A	$-F^A$	σ_A	$-F^A$	σ_A
$\Delta\alpha_k/15$	20 000	86.1 (2)	1.7 (3)	121.2 (2)	1.5 (2)	105.3 (4)	1.6 (2)
$\Delta\alpha_k/15$	40 000	85.2 (2)	1.2 (3)	121.0 (2)	1.3 (2)	105.1 (4)	1.4 (2)
$\Delta\alpha_k/15$	80 000	84.8 (1)	1.0 (2)	120.7 (2)	1.2 (2)	104.8 (3)	1.2 (2)
$\Delta\alpha_k/15$	160 000	84.6 (1)	0.9 (2)	120.6 (2)	1.1 (1)	104.5 (3)	1.1 (1)
$-F^D \quad -F^M$		82.2 (7)	82.9 (8)	118.7 (8)	119.2 (7)	102.6 (5)	103.2 (5)
$-F_{QH}$		86.84 (5)		123.33 (2)		107.78 (7)	
$-F_{LS}$		88.4 (2)		125.6 (2)		110.6 (2)	
	-Energy	56.0 (3)	1.0 (2)	96.2 (3)	1.4 (2)	79.1 (5)	1.3 (2)

[a]For explanations, see the caption of Table V. All HSMC results were calculated with f=[entire future].

TS(160 000) can already be considered as converged with respect to future sampling for the first two microstates, whereas the extended state is expected to converge for n_f = 640 000. Because we are mainly interested in entropy and free energy differences, they can be calculated reliably from the data by subtracting 0.1 kcal/mol from the entropy of the extended microstate and assuming that the effect of the bin size is the same for all microstates (see below).

The F^A results, as expected, increase monotonically with increasing n_f and they are larger than the corresponding values in bin $\Delta\alpha_k/10$ (not shown), while the fluctuations, σ_A always decrease with ratios (energy fluctuation)/σ_A that are significantly smaller than those obtained for the rigid models, indicating a worse approximation for the flexible model. For the helix and hairpin, F^B decreases with increasing n_f for each bin but the corresponding results of bin $\Delta\alpha_k/10$ are smaller than those of bin $\Delta\alpha_k/15$; for the extended microstates even this order within each bin is not satisfied. Therefore, we do not provide the results for F^B in Table VI. However, we give the results for F^D, which are expected to be statistically more reliable than those for F^B, and for comparison provide also results for F^M, based on the best results for F^A and F^B, i.e., for n_f=160 000. The results for F^D and F^M are equal to each other with relatively large statistical errors and as expected, they are larger than those of F^A.

2. Comparison with other methods

The table demonstrates again that as the number of variables increases (or chain length increases) anharmonic effects become stronger, which is reflected by quasiharmonic results, S_{QH} for the entropy that overestimate the S^A values by 1.8–2.4 kcal/mol (thus, the correct values even more); these S_{QH} values were obtained from samples of 120 000 conformations. The LS results were calculated from samples of size of 80 000 and they are based on a correlation param-

eter, b=2 and a discretization parameter, l=15, which define a better approximation than that used for S_{LS} in Tables II and V. S_{LS}, as expected overestimates S^A and with the above approximation also S_{QH}. The QH and LS results for the free energy constitute lower bounds for the free energy corresponding to the entropy results.

3. Differences in F and S

Results for $T\Delta S^A$, ΔF^A, and ΔE for the flexible model appear at the bottom of Table IV. It is shown that the entropy differences are close to those obtained for the rigid model, while the free energy differences involving the extended microstate increased significantly due to the significant decrease in the energy of the helix and hairpin compared to the rigid model. To further verify the reliability of these differences that are based on S^A, we calculated them based on F^D and obtained 20 (1), 36.6 (10), and 16.1 (3) which are equal within relatively large statistical errors to the corresponding values in Table V.

IV. SUMMARY AND CONCLUSIONS

We have applied here the *complete* HSMC method (i.e., f=entire future) to the rigid model of $(Gly)_{10}$ and $(Gly)_{16}$ and the flexible model of $(Gly)_{10}$; for the rigid model of $(Gly)_{10}$ we have also tested an *approximate* HSMC method where a limited future based on f=4 and 3 dihedral angles (rather than f=entire future) was simulated by MC for the calculation of the transition probabilities. To be able to represent the wide microstates of the helix, extended, and hairpin by stable MC samples the simulation temperature was decreased to 100 K, and samples of 600 and 700 conformations were generated. In addition to the entropy and free energy functionals, S^A and F^A that were studied in Ref. 11, and in practice are upper and lower bounds, respectively, we

have studied the functionals, F^B and F^D, where the former is an upper bound and the latter an exact expression for the free energy. As in Ref. 10, we have also demonstrated that the free energy can be obtained by reconstructions of a single conformation.

As expected, the approximate HSMC provides good approximations as long as f covers the range of the angular correlation along the chain; thus, $f = 3$ and 4 are suitable for the extended microstate where short-range correlations (interactions) exist, but becoming too small for the helix and hairpin, which are characterized by medium and long-range interactions, respectively. Still, approximate HSMC (as the LS method) is expected to generate adequate results for *differences* in entropy and free energy among microstates with a similar range of correlations. In most cases the results for F^A, F^B, F^D, and $F^M [(F^A + F^B)/2]$ are found to be consistent, i.e., the first two are lower and upper bounds, respectively, and $F^D \sim F^M$. Most importantly, differences of entropy and free energy between these three microstates are obtained with relatively small errors of up to 0.3 kcal/mol.

An important property of the HSMC method (and the HS approach in general) is the fact that it provides tools that allows one to determine the accuracy of the results (e.g., upper and lower bounds) without the need to compare them with results of other methods. Still, support for the reliability of HSMC is the fact that $S(\text{HSMC})$ is always smaller than $S(\text{LS})$ where the latter constitute an upper bound. Also, it is of interest to point out that for the rigid model of $(\text{Gly})_{10}$ the quasiharmonic (QH) results are very close to the complete HSMC values, suggesting that the samples are approximately quasiharmonic. However, for $(\text{Gly})_{16}$ and the flexible model of $(\text{Gly})_{10}$, where the number of variables increases from 30 to 48 and 60, respectively, the QH results for the entropy overestimate the HSMC values by up to 2.3 kcal/mol, which suggests that anharmonic effects are significant and the QH approximation becomes unsuitable. This emphasizes the importance of the complete HSMC method, which can be applied to a macromolecule with any degree of flexibility, ranging from local fluctuations to a random coil, where side chains visit all the available rotamers, for example.

At this stage of development of complete HSMC, reconstructing a single conformation of $(\text{Gly})_{10}$ defined by the rigid model and based on $n_f = 160\,000$ requires ~ 90 min CPU time on a 2.6 GHz Athlon processor, meaning that an $n_f = 40\,000$ run, which is sufficient for providing the 0.1–0.2 kcal/mol accuracy, requires 23 min CPU; for the rigid model of $(\text{Gly})_{16}$ the time increases by a factor of ~ 2.2 and for the flexible model of $(\text{Gly})_{10}$ it is 2.4 h CPU. As discussed earlier, increasing the number of variables increases the total error in the reconstruction probability, as is demonstrated by the results for $(\text{Gly})_{16}$ and the flexible model of $(\text{Gly})_{10}$, therefore for larger peptides the bins' sizes should be decreased and the future sample size, n_f increased. This can be achieved efficiently by adopting, for example, "the two-stage sampling" technique applied first to fluids in Ref. 10, where the transition probability becomes a product of two probabilities. The first one is defined by Eq. (15) for a relatively large bin (bin 1) and the second is a conditional probability obtained by forcing the angle at step k to stay at bin 1 while measuring its visits (counts) to a smaller bin (bin 2) contained within bin 1. This technique will be studied in future work.

The present results demonstrate that the difference in stability, $\Delta F_{mn} = F_m - F_n$ between *significantly different* microstates m and n can be obtained from two simulations only without the need to resort to thermodynamic integration. Our long-term goal is to develop software that enables one to apply the method to a general peptide consisting of any sequence of amino acid residues, where the simulations are performed by molecular dynamics, which is expected to be significantly more efficient than MC. The method will be applied to surface loops in proteins modeled by a force field and implicit as well as explicit solvent; in the latter case the peptide will be reconstructed first, where the water molecules will be added to a volume containing a frozen peptide structure. The complete HSMC method is expected to become an important tool for studying various problems involving intermediate flexibility.

ACKNOWLEDGMENTS

[1] H. Meirovitch, J. Phys. A **16**, 839 (1983).
[2] H. Meirovitch, Phys. Rev. A **32**, 3709 (1985).
[3] H. Meirovitch and H. A. Scheraga, J. Chem. Phys. **84**, 6369 (1986).
[4] H. Meirovitch, J. Chem. Phys. **97**, 5816 (1992).
[5] H. Meirovitch, J. Chem. Phys. **111**, 7215 (1999).
[6] H. Meirovitch, J. Chem. Phys. **114**, 3859 (2001).
[7] A. Szarecka, R. P. White, and H. Meirovitch, J. Chem. Phys. **119**, 12084 (2003).
[8] R. P. White and H. Meirovitch, J. Chem. Phys. **119**, 12096 (2003).
[9] R. P. White and H. Meirovitch, Proc. Natl. Acad. Sci. U.S.A. **101**, 9235 (2004).
[10] R. P. White and H. Meirovitch, J. Chem. Phys. **121**, 10889 (2004).
[11] S. Cheluvaraja and H. Meirovitch, Proc. Natl. Acad. Sci. U.S.A. **101**, 9241 (2004).
[12] D. L. Beveridge and F. M. DiCapua, Annu. Rev. Biophys. Biophys. Chem. **18**, 431 (1989).
[13] P. A. Kollman, Chem. Rev. (Washington, D.C.) **93**, 2395 (1993).
[14] W. L. Jorgensen, Acc. Chem. Res. **22**, 184 (1989).
[15] H. Meirovitch, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd (Wiley, New York, 1998), Vol. 12, p. 1.
[16] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[17] B. J. Alder and T. E. Wainwright, J. Chem. Phys. **31**, 459 (1959).
[18] J. A. McCammon, B. R. Gelin, and M. Karplus, Nature (London) **267**, 585 (1977).
[19] H. Meirovitch, Chem. Phys. Lett. **45**, 389 (1977).
[20] H. Meirovitch, Phys. Rev. B **30**, 2866 (1984).
[21] H. Meirovitch, M. Vásquez, and H. A. Scheraga, Biopolymers **26**, 651 (1987).
[22] H. Meirovitch, S. C. Koerber, J. Rivier, and A. T. Hagler, Biopolymers **34**, 815 (1994).
[23] A. J. Chorin, Phys. Fluids **8**, 2656 (1996).
[24] M. Vásquez, G. Némethy, and H. A. Scheraga, Chem. Rev. (Washington, D.C.) **94**, 2183 (1994).
[25] F. H. Stillinger and T. A. Weber, Science **225**, 983 (1984).
[26] R. Elber and M. Karplus, Science **235**, 318 (1987).
[27] H. Meirovitch and E. Meirovitch, J. Phys. Chem. **100**, 5123 (1996).
[28] C. Baysal and H. Meirovitch, Biopolymers **50**, 329 (1999).
[29] M. Karplus and J. N. Kushick, Macromolecules **14**, 325 (1981).

[30] O. L. Rojas, R. M. Levy, and A. Szabo, J. Chem. Phys. **85**, 1037 (1986).

[31] 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).

[32] J. W. Ponder, TINKER, software tools for molecular design, version 3.9 (1999).

[33] N. Gō and H. A. Scheraga, J. Chem. Phys. **51**, 4751 (1969).

[34] N. Gō and H. A. Scheraga, Macromolecules **9**, 535 (1976).

[35] H. Meirovitch, M. Vásquez, and H. A. Scheraga, Biopolymers **27**, 1189 (1988).

[36] H. Meirovitch, J. Chem. Phys. **89**, 2514 (1988).

[37] H. Meirovitch and Z. Alexandrowicz, J. Stat. Phys. **15**, 123 (1976).

[38] A. T. Hagler, P. S. Stern, R. Sharon, J. M. Becker, and F. Naider, J. Am. Chem. Soc. **101**, 6842 (1979).

[39] T. L. Hill, *Statistical Mechanics Principles and Selected Applications* (Dover, New York), 1986.