

# Convergence of folding free energy landscapes via application of enhanced sampling methods in a distributed computing environment

Xuhui Huang,<sup>1</sup> Gregory R. Bowman,<sup>2</sup> and Vijay S. Pande<sup>3,a)</sup>

<sup>1</sup>*Department of Bioengineering, Stanford University, Stanford, California 94305, USA*

<sup>2</sup>*Biophysics Program, Stanford University, Stanford, California 94305, USA*

<sup>3</sup>*Department of Chemistry, Stanford University, Stanford, California 94305, USA*

(Received 8 January 2008; accepted 20 March 2008; published online 28 May 2008)

We have implemented the serial replica exchange method (SREM) and simulated tempering (ST) enhanced sampling algorithms in a global distributed computing environment. Here we examine the helix-coil transition of a 21 residue  $\alpha$ -helical peptide in explicit solvent. For ST, we demonstrate the efficacy of a new method for determining initial weights allowing the system to perform a random walk in temperature space based on short trial simulations. These weights are updated throughout the production simulation by an adaptive weighting method. We give a detailed comparison of SREM, ST, as well as standard MD and find that SREM and ST give equivalent results in reasonable agreement with experimental data. In addition, we find that both enhanced sampling methods are much more efficient than standard MD simulations. The melting temperature of the F<sub>s</sub> peptide with the AMBER99 $\phi$  potential was calculated to be about 310 K, which is in reasonable agreement with the experimental value of 334 K. We also discuss other temperature dependent properties of the helix-coil transition. Although ST has certain advantages over SREM, both SREM and ST are shown to be powerful methods via distributed computing and will be applied extensively in future studies of complex bimolecular systems. © 2008 American Institute of Physics.

[DOI: 10.1063/1.2908251]

## I. INTRODUCTION

Adequately sampling the conformational space of biomolecules is important for understanding molecular functions. Computer simulation, such as molecular dynamics (MD) and Monte Carlo (MC), is a powerful technique for exploring conformation space. However, such simulations often become trapped in local energy minima when applied to complex biomolecular systems.<sup>1,2</sup> Recently, it has been shown that a large ensemble of individual MD trajectories generated by a distributed computing network makes it possible to study the folding of some miniproteins in explicit solvent.<sup>3–8</sup> Unfortunately, due to the local trapping problem, it is still difficult to satisfactorily sample the entire configuration space of a biomolecule in explicit solvent even with a distributed computing effort. Tempering methods such as simulated tempering<sup>9,10</sup> (ST) and parallel tempering [or replica exchange method (REM)]<sup>11–14</sup> were developed to overcome this kinetic trapping problem by inducing a random walk in temperature space. At high temperatures the free energy landscape is more or less flat, allowing broad sampling of configuration space. Conformations found at high temperatures may then be exchanged to lower temperatures in order to explore the relevant free energy minima. Both ST and parallel tempering have proved useful for sampling phase space and have been widely used to simulate biomolecular systems.<sup>2,15–21</sup> Therefore, the combination of the tempering methods and a distributed computing network such as FOLDING@HOME (FAH), currently with more than 200 000

computer processors (CPU's), should greatly enhance our ability to sample the conformation space of complex biomolecular systems.

In the REM, independent MD or MC simulations are performed in parallel at different temperatures. Exchanges of configurations between neighboring temperatures are attempted periodically and accepted according to the detailed balance condition.<sup>13,14</sup> The standard REM scheme requires synchronization and frequent communication between different processors, which makes it unsuitable for a heterogeneous distributed computing environment lacking direct communication between hosts. Rhee and Pande developed the multiplexed replica exchange method (MREM) for running replica exchange on distributed computers.<sup>22</sup> In MREM, multiplexed replicas with more than one independent simulation at each temperature are run, and exchanges of configurations between these multiplexed replicas are attempted periodically. This algorithm still requires synchronization between multiplexed replicas, resulting in short idle times, as in another serial algorithm called Asynchronous Replica Exchange.<sup>23</sup> Recently, Hagen *et al.* proposed another method called serial replica exchange method (SREM) that is suitable for distributed computing.<sup>24</sup> In SREM, a single, independent simulation performs a random walk in temperature space by making regular attempts to change temperatures. The transition probability for this move is determined by one potential energy from the simulation and a second one from a prestored potential energy distribution function (PEDF) at the new temperature. In contrast to MREM, SREM does not require any synchronization between different simulations,

<sup>a)</sup>Electronic mail: pande@stanford.edu.

so no CPUs are left idle. Thus SREM is substantially more efficient than MREM. The main difficulty with SREM is determining the PEDFs because SREM is only approximately correct unless the exact PEDFs are adopted.

ST, an inherent serial algorithm, is another enhanced sampling method suitable for distributed computing environments.<sup>9,10</sup> In ST, an expanded canonical ensemble, where canonical ensembles at different temperatures are weighted differently, is used to sample the conformation space. A single simulation is performed, again making periodic attempts to change temperatures according to a well defined transition probability determined by the detailed balance condition. In contrast to SREM, ST is an exact method. Each temperature is assigned a weight and the difference between the weights of neighboring temperatures determines the probability of making an exchange between them. If these weights are not determined properly then simulations will be constrained to a subset of temperature space and become inefficient.<sup>9,10,25,26</sup> Therefore, determining a good set of weights is essential for ST. A number of attempts have been made to obtain weights allowing the system to perform a random walk in temperature space.<sup>27-33</sup> However, these algorithms normally require an interactive procedure and relatively expensive trial simulations such as REM. In this study, we propose a simple method to compute good initial weights from short trial MD simulations. The combination of this scheme with an adaptive weighing method is shown to efficiently determine a set of weights enabling the system to perform a random walk in temperature space.<sup>31</sup>

In this paper, we apply both SREM and ST to the F<sub>s</sub> peptide in order to study the helix-coil transition using a distributed computing network. Our aim is to make an exact comparison between SREM, ST, and MD. We show that both enhanced sampling methods outperform standard MD and give reasonable agreement with experimentally determined thermodynamic properties of the F<sub>s</sub> peptide.

## II. SYSTEM AND METHODS

### A. Serial replica exchange

The SREM algorithm,<sup>24</sup> a serial version of the REM, is designed to run asynchronously on a worldwide distributed computing environment such as FAH.<sup>34</sup> We will first review the original REM. REM, also referred to as parallel tempering, involves running independent MD or MC simulations in parallel at different temperatures.<sup>13,14</sup> After a certain interval, an attempt is made to exchange the configurations of the neighboring replicas (closest neighbors in temperature space), typically as shown by the following

$$\begin{aligned} (X_i, T_i) &\rightarrow (X_j, T_i), \\ (X_j, T_j) &\rightarrow (X_i, T_j), \end{aligned} \quad (1)$$

where  $X_i$  and  $T_i$  are the configuration and the temperature of the  $i$ th replica, respectively. In the current study, we will focus on using MD to propagate each replica.

The transition probability for exchanging adjacent replicas  $i$  and  $j$  must satisfy the detailed balance condition,

$$P_i(X_i, p_i) P_j(X_j, p_j) P(i \rightarrow j) = P_j(X_j, p'_j) P_i(X_i, p'_i) P(j \rightarrow i), \quad (2)$$

where  $P(i \rightarrow j)$  is the transition probability for the exchange  $i \rightarrow j$  and  $P_i(X_j, p_j)$  is the probability of finding configuration  $X_j$  with momenta  $p_j$  at temperature  $i$ .

The equilibrium probability of state  $i$  for the canonical ensemble is

$$P_i(X_i, p_i) = \frac{1}{Z_i} \exp(-\beta_i H(X_i, p_i)), \quad (3)$$

where  $\beta_i = 1/(k_B T_i)$ ,  $p_i$  is the momenta, and  $Z_i$  is the partition function. The Hamiltonian  $H_i(X_i, p_i)$  is the sum of the kinetic energy ( $K$ ) and potential energy ( $U$ ) ( $H_i(X_i, p_i) = K(p_i) + U(X_i)$ ).

A rescaling of the momenta<sup>13</sup> ( $p'_i = \sqrt{T_j/T_i} p_i$ ) following the exchange causes the kinetic energy to cancel out in the detailed balance equation, and the transition probability can be rewritten as

$$\frac{P(i \rightarrow j)}{P(j \rightarrow i)} = \exp(-\Delta_{ij}), \quad (4)$$

where

$$\Delta_{ij} = -(\beta_j - \beta_i)(U(X_i) - U(X_j)). \quad (5)$$

If the Metropolis criteria is applied then one of the transition probabilities will always be one. Thus, they can be written as

$$\begin{aligned} P(i \rightarrow j) &= \min\{1, e^{-\Delta_{ij}}\}, \\ P(j \rightarrow i) &= \min\{1, e^{\Delta_{ij}}\}. \end{aligned} \quad (6)$$

From Eq. (6), it is easy to show that the average acceptance ratios for the forward and backward transitions in REM are equal,

$$\langle P(i \rightarrow j) \rangle = \langle P(j \rightarrow i) \rangle. \quad (7)$$

The above REM scheme requires synchronization and frequent communication between different processors. This requirement is not suitable for a distributed computing environment which consists of a heterogeneous network of distributed computers that are unable to communicate directly with one another. Recently, Hagen *et al.* noticed that in REM energy is the only property involved in the detailed balance condition. They derived the SREM based on this property.<sup>24</sup> In SREM, PEDFs,  $P(E, T_n)$ , are stored at different temperatures. The algorithm works as follows: A single replica is propagated with MC or MD at temperature  $T_n$  using a single processor. After a certain interval, it attempts to move to the neighboring temperature  $T_{n+1}$ . At this time, two potential energies  $E_n$  and  $E_{n+1}$  are required to determine the transition probability as shown in Eq. (5).  $E_n$  is the potential energy of the current simulation system and  $E_{n+1}$  is sampled from the energy distribution  $P(E, T_{n+1})$  at  $T_{n+1}$ . If the move is accepted, the replica will continue propagating at the new temperature  $T_{n+1}$ , otherwise the replica will stay at the initial temperature  $T_n$ . In this way, a single replica can walk through

the whole temperature space and thermodynamic averages can be computed at different temperatures.

The main difficulty with SREM is determining the PEDFs. SREM is only approximately correct unless the exact PEDFs are used. Hagen *et al.* suggest that the PEDFs be updated iteratively until they are stationary.<sup>24</sup> We apply SREM on a distributed computing network to study the helix-coil transition. In Sec. III, we will discuss the convergence of the PEDFs and show that SREM predicts correct thermodynamic properties for the F<sub>s</sub> peptide system.

## B. Simulated Tempering

ST is another enhanced sampling method suitable for a distributed computing environment because it is inherently a serial algorithm.<sup>9,10</sup> In ST, configurations are sampled from an expanded canonical ensemble in which the canonical ensemble with different temperatures are weighted differently. A generalized hamiltonian  $\Xi_n(X,p)$  is defined as

$$\Xi_n(X,p) = \beta_n H(X,p) - g_n, \quad (8)$$

where  $\beta_n = 1/(k_B T_n)$ ,  $H(X,p)$  is the Hamiltonian for the canonical ensemble at temperature  $T_n$ , and the *a priori* determined constant  $g_n$  is the weight for the temperature  $T_n$ .

In practice, ST works as follows: A single simulation is performed at a particular temperature using MC or MD, and after a certain interval, an attempt is made to change the configuration to another temperature among a list of choices  $(T_1, \dots, T_n)$  as shown by the following:

$$(X_i, T_i) \rightarrow (X_i, T_j), \quad (9)$$

where  $X_i$  is the configuration of the system and  $T_i$  is the temperature.

The transition probability for moving from  $T_i$  to  $T_j$  has to satisfy the following detailed balance condition:

$$P_i(X_i, p_i) P(i \rightarrow j) = P_j(X_i, p'_i) P(j \rightarrow i). \quad (10)$$

The probability of state  $i$  for the expanded canonical ensemble is

$$P_i(X_i, p_i) = \frac{1}{Z} \exp(-\Xi_i(X_i, p_i)), \quad (11)$$

where  $\beta_i = 1/(k_B T_i)$ ,  $p_i$  is the momenta, and  $Z$  is the partition function for the expanded canonical ensemble.

The momenta are rescaled just as in REM, again causing them to drop out of the detailed balance condition. The transition probabilities after applying the metropolis criteria are shown by the following:

$$P(i \rightarrow j) = \min\{1, \exp(-(\beta_j - \beta_i)U^i(X) + (g_j - g_i))\}, \quad (12)$$

$$P(j \rightarrow i) = \min\{1, \exp(-(\beta_i - \beta_j)U^j(X) + (g_i - g_j))\},$$

where  $U^i(x)$  and  $U^j(x)$  are potential energies sampled from the canonical ensembles at  $T_i$  and  $T_j$ , respectively.

In contrast to SREM, which is an approximate method unless the *exact* PEDFs are employed, ST is an exact method regardless of the selection of the weights. However, without proper weighting, ST will only explore a subset of temperature space and becomes inefficient. In ST, the “*optimal*”

weights are typically defined to be a set of weights which lead the system to perform a random walk in temperature space, resulting in uniform sampling of each temperature.<sup>9,10,25,26,30-33</sup> In Sec. III, we show that such *uniform* sampling does not necessarily produce *optimal* sampling, and could oversample at high temperatures and undersample at low temperatures or around the melting temperature. Thus, we refer to weights leading to uniform sampling as “*free energy*” weights rather than optimal weights in this study, for reasons explained in the next section.

*How to choose weights leading to uniform sampling?* In ST, the probability that temperature  $T_i$  is visited in the expanded canonical ensemble is

$$P_i = Z_i e^{g_i} / Z, \quad (13)$$

where  $Z_i$  and  $Z$  are partition functions for the canonical ensemble at  $T_i$  and for the expanded ensemble, respectively.

In the canonical ensemble, the partition function is related to the Helmholtz free energy,

$$f_i = -\frac{1}{\beta_i} \ln Z_i. \quad (14)$$

Uniform sampling where  $P_1 = P_2 = \dots = P_n$  can be achieved if

$$g_i = \beta_i f_i. \quad (15)$$

Since the weights leading to uniform sampling equal the unitless free energies at different temperatures, we refer to them as free energy weights. It is not an easy task to determine these free energy weights. In principle, these unitless free energies can be estimated from trial simulations using the weighted histogram analysis method.<sup>27,28,31</sup> However, this still requires an interactive procedure and relatively expensive trial simulations such as REM.<sup>31</sup>

In this study, we propose a simple method to compute approximate free energy weights from short trial MD simulations. Our method is based on the following property:<sup>25</sup> The “free energy” weights leading to uniform sampling must yield the same acceptance ratios for both forward and backward transitions from  $T_i$  to  $T_j$  as shown by the following:

$$\langle P(i \rightarrow j) \rangle = \langle P(j \rightarrow i) \rangle. \quad (16)$$

From Eq. (13), we can compute the average acceptance ratios,

$$\langle P(i \rightarrow j) \rangle = e^{g_j - g_i} \int_{-\infty}^{U_0} e^{-(\beta_j - \beta_i)U_i} P(U_i) dU_i + \int_{U_0}^{\infty} P(U_i) dU_i, \quad (17)$$

$$\langle P(j \rightarrow i) \rangle = \int_{-\infty}^{U_0} P(U_j) dU_j + e^{g_i - g_j} \int_{U_0}^{\infty} e^{-(\beta_i - \beta_j)U_j} P(U_j) dU_j,$$

where  $P(U_i)$  is the PEDF at  $T_i$  and  $U_0 = (g_i - g_j) / (\beta_i - \beta_j)$ .

PEDFs for each temperature are estimated from the short trial MD simulations by assuming the distributions are Gaussian.

$$P(U_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(U_i - \langle U_i \rangle)^2}{2\sigma_i^2}\right), \quad (18)$$

where  $\langle U_i \rangle$  and  $\sigma_i$  can be computed from the trial simulations.

By solving Eq. (16), we can obtain a set of near free energy weights with a given temperature list. This scheme is an extension of the method proposed by Park and Pande,<sup>25</sup> in which the weights were also computed by ensuring equal average acceptance ratios for the forward and backward transitions. However, in their original scheme, they only considered the average potential energy at each temperature. By taking into account the potential energy distribution our method provides greater efficiency, especially around the melting temperature where PEDFs are normally wider. To gain this advantage we assume that the PEDFs are Gaussian. This assumption has been shown to be valid in a variety of biological systems<sup>26,35</sup> including the Fs peptide studied here. More generally, we believe that the central limit theorem<sup>36</sup> implies the overall PEDFs are nearly Gaussian for any biological system in explicit solvent due to the large number of degrees of freedom.

Recently, Park<sup>26</sup> showed that the two methods for obtaining approximate free energy weights discussed above are consistent, i.e.,  $\langle P(i \rightarrow j) \rangle$  and  $\langle P(j \rightarrow i) \rangle$  become identical only if  $g_i = \beta_i f_i$ . These near free energy weights are still approximate considering the limited length of the trial MD simulations. For example, it is particularly difficult for these simulations to capture slow conformational changes, e.g., protein folding/unfolding. Therefore, in our study we combined the above scheme with an adaptive weighing method. The latter regularly updates the weights during the ST simulations using an adaptive weighted histogram analysis method (WHAM) by Bartels and Karplus.<sup>31</sup>

In ST, choosing free energy weights only ensures  $\langle P(i \rightarrow j) \rangle = \langle P(j \rightarrow i) \rangle$ . But the value of the acceptance ratio is determined by the temperature gap  $\Delta T_{ij} = T_i - T_j$ . Transition probabilities diminish quickly with increasing temperature gap. For example, when  $T_i \gg T_j$ ,  $\langle P(i \rightarrow j) \rangle = \langle P(j \rightarrow i) \rangle \approx 0$ , thus it will take forever to achieve uniform sampling even if the equilibrium probabilities are equal for the two states ( $P_i = P_j$ ). Therefore, the attempted moves in ST are usually between neighboring temperatures and a good temperature list will yield relatively large and equal acceptance ratios between them.

### C. Simulation details

In this study we examined the capped Fs (Acc-A5[AAAR+A]3A-NMe) peptide. The AMBER-99 $\phi$  potential,<sup>5</sup> a modified version of AMBER99,<sup>37</sup> was selected because it was found to better reproduce experimental helix-coil properties compared to earlier AMBER potentials.<sup>5</sup> Both the ST and SREM algorithms were implemented in a version of the GROMACS molecular dynamics simulation package<sup>38</sup>

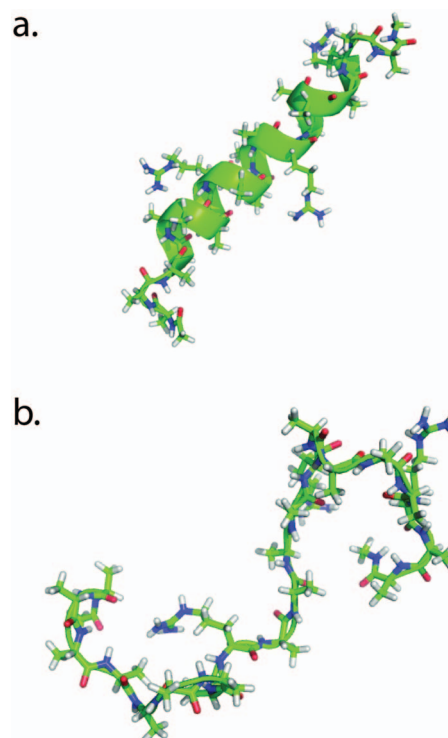


FIG. 1. (Color) The two initial configurations used for simulations of the capped Fs peptide (Ace-A5[AAAR+A]3A-NMe): (a) A prehelical structure and (b) a random-coil structure.

modified for the FAH (Ref. 34) infrastructure (<http://folding.stanford.edu>).

Two initial configurations are used, as shown in Fig. 1. The first is a prehelix structure with a helical content of 58% and the second is a random coil structure with no helical content. Each system is solvated in a 42 Å cubic box using 2039 TIP3P water molecules.<sup>39</sup> Three Cl<sup>-</sup> counterions were included to neutralize the charged peptide. The simulation systems were minimized using a steepest descent algorithm, followed by a 100 ps MD simulation applying a position restraint potential to the peptide heavy atoms. All simulations were conducted using constant NVT with a Nosé-Hoover thermostat having a coupling constant of 0.02 ps<sup>-1</sup>.<sup>40</sup> Long-range electrostatic interactions were treated using the reaction field method with a dielectric constant of 80. 9 Å cutoffs were imposed on nonbonded interactions. Neighbor lists were updated every ten steps. A 2 fs time step was used and covalent bonds involving hydrogen atoms were constrained with the LINCS algorithm.<sup>41</sup>

Starting from the two initial configurations, 1300 simulations were performed on a distributed computing environment using both the SREM and ST enhanced sampling methods. The total simulation time was aggregated to more than 250  $\mu$ s. The same 50 element, roughly exponentially distributed temperature list covering a range from 285 to 592 K was used for both SREM and ST. The initial temperatures were uniformly selected from the temperature list. Thus, there are 26 simulations starting from each temperature, each with a different set of initial velocities. In both SREM and ST simulations, exchanges were attempted every 2 ps. In SREM, potential energies were saved every 0.4 ps. To obtain

initial estimates of the PEDFs in SREM, we performed two hundred 500 ps SREM simulations where every move was accepted, resulting in a continuous walk in temperature space. The above simulations were first equilibrated for 2.5 ns at different initial temperatures. In a subsequent updating phase, slightly different from what was suggested by Hagen *et al.*,<sup>24</sup> we continued updating the PEDFs, though the updates became less frequent with time. The PEDFs were updated every 40 ns for about 40 iterations at first, then every 400 ns for about 20 iterations, and at last every 1000 ns. In ST, the initial weights were computed using the data obtained from 50 short 100 ps trial simulations. Subsequently, the weights were updated approximately every 400 ns by an adaptive weighing scheme.<sup>31</sup>

In the SREM/ST simulations run on a distributed computing environment described above, PEDFs or weights were updated based on data gathered from a heterogeneous network of computers with different rates of progress for each of a number of parallel simulations. Thus, it is difficult to do an exact comparison of the efficiency of SREM and ST using distributed computing data. To overcome this issue, we performed additional SREM, ST, and MD simulations on a local computer cluster. To ensure these simulations could be compared exactly, they were run with the same parameters and for the same duration. In addition, the same protocol was used for updating the SREM PEDFs and the ST weights. To be more specific, 50 24 ns simulations were run with SREM and ST. For MD 50 simulations were run with constant *NVT* at each of the three temperatures: 285, 308, and 592 K. For SREM and ST a 1 ns MD simulation with constant *NVT* was run for each of the 50 temperatures in our temperature list. We obtained initial weights for ST by solving Eq. (16) and PEDFs for SREM from these 1 ns simulations. These simulations were then extended using the SREM and ST algorithms. The PEDFs or weights were updated every 500 ps for six iterations. Then they were updated every 4 ns for five iterations. Thus, just as for MD at each of the three temperatures, a total of 50 24 ns simulations were run for both SREM and ST. The first 4 ns of each simulation was disregarded for equilibration purposes regardless of the algorithm used.

### D. Lifson-Roig helix-coil counting theory

Helical properties were computed using the Lifson–Roig helix-coil counting theory.<sup>42,43</sup> In this model, a residue is considered to be helical if  $\phi \in (-60 \pm x)^\circ$  and  $\psi \in (-47 \pm x)^\circ$ , where  $x$  is typically 30 (Refs. 5 and 18) but may be varied. We set  $x$  to 40°, and show that this choice gives the best agreement with the melting temperature using the AMBER-99 $\phi$  potential in Sec. III C. Following the Lifson–Roig model, a helical segment is defined as three or more consecutive helical residues because a minimum of three residues are required to form a helical hydrogen bond. Each segment has a length of  $n-2$ , where  $n$  is the total number of consecutive helical residues. Thus, the maximum helical length of our 21 residue F<sub>s</sub> peptide system is 19. This maximum is used to calculate the helical content as

$$N_c = \frac{\sum_{h=1}^{N_s} N_h}{N_{\max}}, \quad (19)$$

where  $N_c$  is the helical content,  $N_s$  is the number of helical segments,  $N_h$  is the length of segment  $h$ , and  $N_{\max}$  is the maximum possible helical length.

The Lifson–Roig model also gives average helix nucleation and propagation parameters called  $\langle \nu \rangle$  and  $\langle \omega \rangle$ , respectively. Mathematically,  $\langle \nu \rangle$  is the average statistical weight of a residue being in a helical state and is found as follows:

$$\langle \nu \rangle = \int_{\text{helical}} e^{F(\phi, \psi)/kT} d\phi d\psi, \quad (20)$$

where  $F(\phi, \psi)$  is the free energy of the given  $\phi, \psi$  angles.<sup>5</sup> The  $\langle \omega \rangle$  parameter is the statistical weight of a residue participating in a helical segment and thus requires a residue to be helical itself and belong to a group of at least three consecutive helical residues. It can be calculated as

$$\langle \omega \rangle = \int_{\text{helical}} e^{W(\phi, \psi)/kT} d\phi d\psi, \quad (21)$$

where  $W(\phi, \psi)$  is similar to  $F(\phi, \psi)$  but includes the interaction of the given residue with its neighbors when they form a helical segment.<sup>5</sup>

## III. RESULTS AND DISCUSSION

### A. Comparison of SREM, ST, and MD

To demonstrate the usefulness of the generalized ensemble methods examined in this work we first discuss the results of an exact comparison between SREM, ST, and standard MD. This comparison is based on simulation data generated using each of the algorithms with the same amount of sampling, as described in the Methods section. We also investigate the rate of the convergence for PEDFs in SREM and weights in ST. The data generated by the distributed computing network are shown to be well converged and used as the gold standard for this comparison.

#### 1. Convergence of the PEDF in SREM

As stated in Sec. II A, the main obstacle for SREM is determining the PEDFs. This algorithm cannot be used for production unless the PEDFs are converged. Thus, we first investigate the convergence of the PEDFs.

$\chi^2$  convergence measure. The  $\chi^2$  convergence measure was proposed by Hagen *et al.*<sup>24</sup> to verify the convergence of PEDFs.  $\chi^2$  is defined to be an integrated error,

$$\chi^2(t) = \sum_{i=1}^N (P_i(t) - P_i^{\text{ref}})^2, \quad (22)$$

where  $N$  is the number of bins in the potential histogram,  $P_i(t)$  is the value of the  $i$ th bin of the potential energy histogram generated by potential energies collected over time  $t$  at a particular temperature, and  $P_i^{\text{ref}}$  is the reference PEDF.

*Reference PEDFs.* In order to demonstrate convergence we need a set of converged PEDFs as a reference.

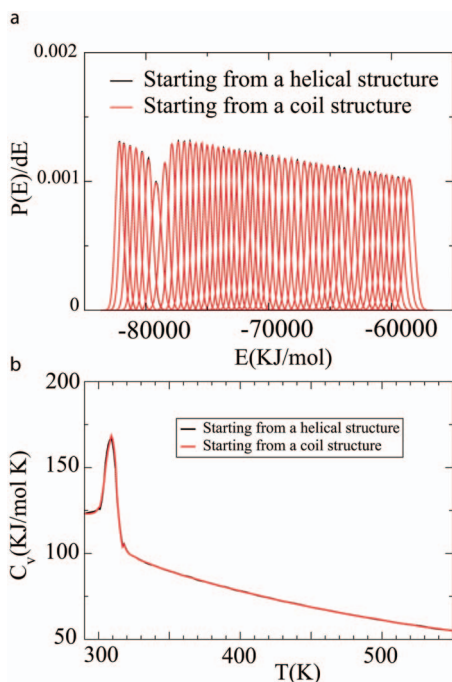


FIG. 2. (Color) (a). Potential energy distribution functions (PEDFs) generated from FOLDING@HOME data at each of the 50 temperatures used. The black and red curves show the converged PEDFs from simulations starting from a prehelix structure and a coil structure, respectively. (See Fig. 1 for the two initial configurations.) (b). Heat capacity ( $C_v = \partial U / \partial T$ ) as a function of temperature generated from Folding@home simulations starting from the two initial configurations.

Figure 2(a) shows two nearly identical sets of SREM PEDFs from massive distributed computing simulations. Each set is from one of the two starting configurations. To quantitatively test their convergence we applied an integrated difference measure ( $D^2(T)$ ) which is analogous to the  $\chi^2$  measure.<sup>24</sup>  $D^2(T) = \sum_{i=1}^N (P_i^{\text{helical}}(T) - P_i^{\text{coil}}(T))^2$ . The  $D^2(T)$  values at different temperatures are close to zero ( $\sim 10^{-9}$  or smaller), indicating strong convergence of the two sets of PEDFs. Section III B also shows that these massive distributed computing SREM simulations generate converged thermodynamics properties. We thus use these well converged PEDFs as our reference point.

**Convergence of PEDFs.** When the well converged PEDFs ( $P_i^{\text{FAH}}$ ) from distributed computing simulations are used as the reference distributions,  $\chi^2(t)$  should decay to zero. On the other hand, when PEDFs from initial trial simulations ( $P_i^{\text{trial}}$ ) are used as the reference,  $\chi^2(t)$  should grow to a plateau value. Using  $P_i^{\text{trial}}$  as the reference distribution has certain advantages, since it does not require a prior knowledge of the correct PEDFs.

In a previous study, Hagen *et al.* showed that the average  $\chi^2(t)$  over all temperatures is a good convergence measure for the alanine dipeptide system they studied.<sup>24</sup> However, we show here that such an average  $\chi^2(t)$  measure misrepresents the convergence of PEDFs for the Fs peptide system because PEDFs at different temperatures do not converge at the same rate. Figure 3 displays the average  $\chi^2(t)$  measures using both  $P_i^{\text{FAH}}$  and  $P_i^{\text{trial}}$  as reference distributions. Although PEDFs obtained from simulations starting from a helical structure converge slightly faster than those obtained from simulations

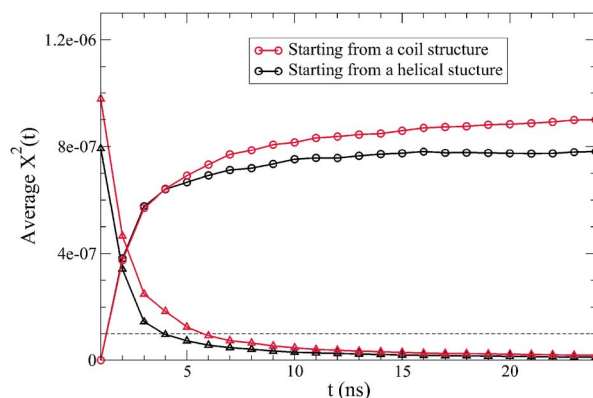


FIG. 3. (Color online) The  $\chi^2$  convergence measure averaged over all temperatures as a function of time. Triangles correspond to using  $P_i^{\text{FAH}}$  as the reference point while circles are used when  $P_i^{\text{trial}}$  is the reference. The black curves are from simulations starting from a prehelical structure, while gray (red) curves are obtained from simulations starting from a coil structure.

starting from a coil structure, both sets of average  $\chi^2(t)$  curves seem to converge after about 6 ns. Curves using  $P_i^{\text{FAH}}$  as the reference distribution decay to a value very close to zero ( $< 10^{-7}$ ) while curves using  $P_i^{\text{trial}}$  as the reference distribution start to reach a plateau. However, not all the PEDFs have converged within 6 ns, as shown in Fig. 4. This figure shows PEDFs from SREM simulations performed on the cluster as well as PEDFs from our distributed computing simulations at three representative temperatures: 285, 308, and 592 K. When  $T=308$  K, PEDFs obtained from 50 6 ns cluster simulations display a visible deviation from the corresponding curves obtained from distributed computing. In particular, the whole potential distribution corresponding to the coil starting structure [the center column of Fig. 4(b)] is clearly shifted to higher energy. On the other hand, all the PEDFs at 285 and 592 K have nearly converged. Most of the PEDFs at other temperatures are also converged, except a few around  $T=308$  K.

In Fig. 5, we show  $\chi^2(t)$  and average  $\chi^2(t)$  curves for our cluster simulations at three representative temperatures using  $P_i^{\text{FAH}}$  as the reference distribution. At 285 and 592 K all the  $\chi^2(t)$  curves are similar to the average  $\chi^2(t)$  plot and decay to a value below  $10^{-7}$  within 6 ns regardless of the initial configuration. However, the  $\chi^2(t)$  plot at 308 K decays much slower than those at other temperatures. At 308 K the  $\chi^2(t)$  plot starting from the prehelical structure converges at about 10 ns [see Fig. 5(a)] while the plot starting from the coil structure decays even slower and never falls below  $10^{-7}$  during our 24 ns SREM simulations [see Fig. 5(b)].

A further investigation shows that the melting temperature occurs around 308 K, which explains the slow convergence of the PEDFs at this temperature. At the melting temperature, the potential energy distribution is likely to be wider due to the more or less equal probability of finding the folded and unfolded states. As shown in Fig. 2(a), potential energy distributions around  $T=308$  K (the seventh curve from the left) are indeed wider than adjacent distributions. In addition, the value of the melting temperature is confirmed by the location of the dominant peak in the heat capacity plot, which occurs at around 310 K as shown in Fig. 2(b).

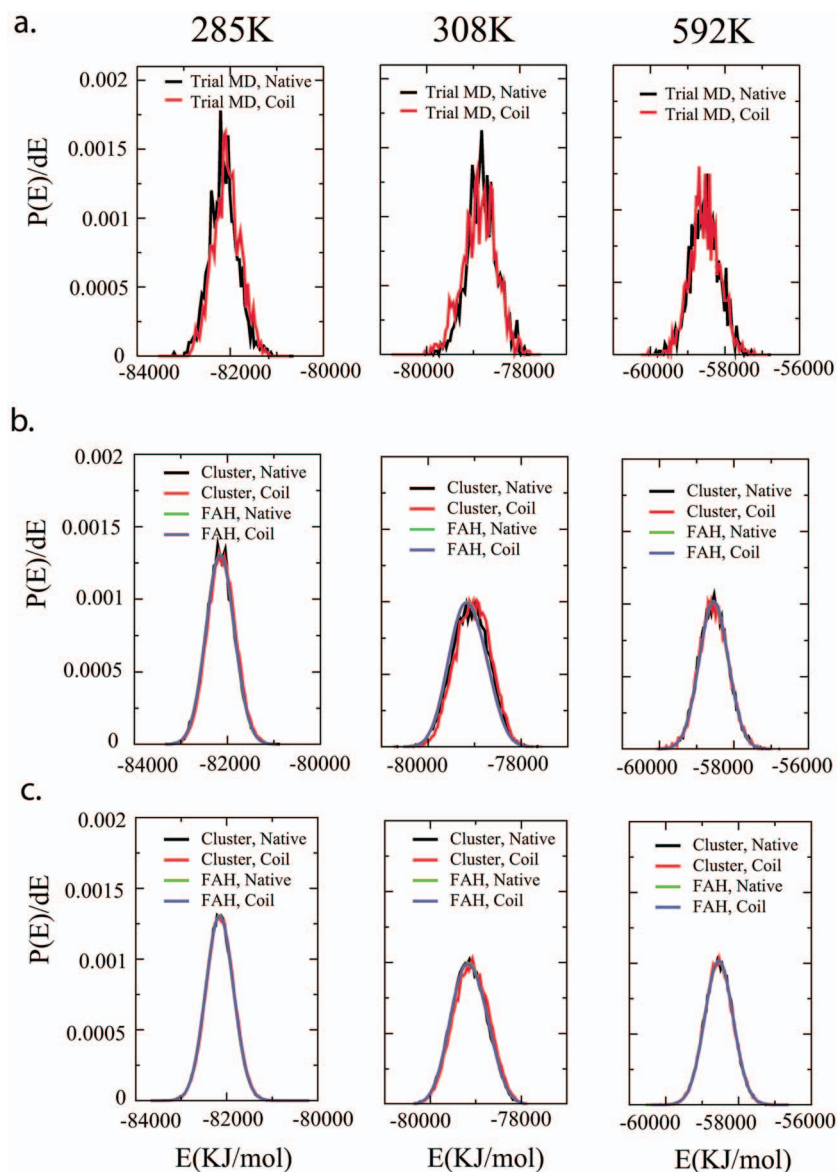


FIG. 4. (Color) Potential energy distribution functions (PEDFs) for three temperatures: 285 K (left column), 308 K (middle column), and 592 K (right column). (a) Rough PEDFs from the 1 ns trial MD simulations performed on the local cluster. The black and red curves correspond to simulations starting from a prehelix structure and a coil structure, repetitively. (b) Comparison of PEDFs obtained from 6 ns SREM simulations at each of the 50 temperatures on the local cluster with those from distributed computing simulations. The black and red curves show results from the local cluster simulations starting from the prehelix and coil structures, respectively. The green and blue curves show the results from FOLDING@HOME simulations again starting from a prehelix and coil structure, respectively. (c) The same as (b) except that 50 24 ns SREM simulations are used to compute the PEDFs.

*Uniform sampling convergence measure.* As stated above, PEDFs at different temperatures do not necessarily converge at the same rate. Furthermore, PEDFs converge slower around the protein melting temperature. Thus, the average  $\chi^2(t)$  is not a good convergence measure, and all 50  $\chi^2(t)$  plots must be monitored instead. In this section, we describe a new convergence measure based on how uniform the sampling is across temperature space.

As shown in Eq. (7), the average forward and backward transition probabilities from  $T_i$  to  $T_j$  in SREM are equal, implying that each simulation should spend an equal amount of time at each temperature given infinite sampling. Given finite sampling, approximately uniform sampling would also be achieved if there are no large gaps in the acceptance ratios between neighboring temperatures. Since all the acceptance ratios in our SREM simulations are between 20% and 40%, approximately equal sampling at each temperature should be obtained if the exact PEDFs are applied (see Fig. 11 for the acceptance ratios). However, if the PEDFs are not converged then  $\langle P_{ij} \rangle \neq \langle P_{ji} \rangle$ , resulting in nonuniform sampling. Thus,

uniform sampling could serve as an indirect convergence measure for PEDFs. We define  $u(t)$ , the average deviation from uniform sampling, as

$$u(t) = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \frac{n_i}{\langle n_i \rangle} - 1 \right)^2}, \quad (23)$$

where  $N$  denotes the number of temperatures,  $n_i$  represents the total amount of time spent at temperature  $T_i$  in a window of time centered at  $t$  over all the simulations, and  $\langle n_i \rangle$  is the average of  $n_i$  over all the temperatures. Given this definition,  $u(t)=0$  when the sampling is perfectly uniform and increases the more nonuniform the data are.

Figure 6 shows the amount of time our SREM simulations spent at each temperature during a series of 4 ns windows as well as the corresponding  $u(t)$  curves. When starting from a prehelical structure, the sampling is already roughly uniform in the second window (from 4 to 8 ns) but slightly more time is still spent at low temperatures before  $t=16$  ns. The corresponding  $u(t)$  plot is always below 0.4 and even falls below 0.2 after 16 ns, indicating that the PEDFs are

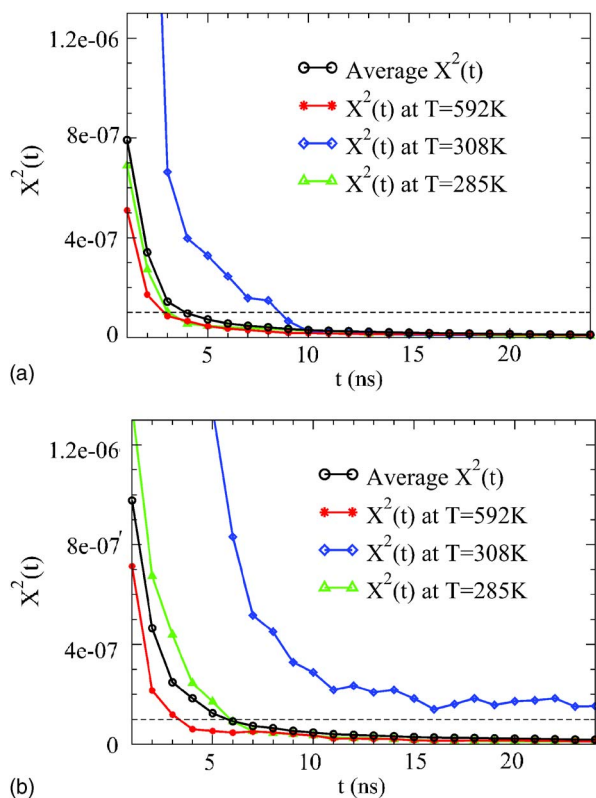


FIG. 5. (Color online)  $\chi^2$  convergence measure of the PEDFs as a function of time at three representative temperatures, 285 (green triangles), 308 (blue diamonds), and 592 K (red asterisks) are shown and compared to the averaged  $\chi^2$  (black circles). The plots are averaged over 50 SREM simulations, and  $P_i^{\text{FAH}}$  are used as the reference distributions. (a).  $\chi^2$  convergence measure obtained from simulations starting from a prehelical structure. (b) The same as (a) except that the results are obtained from simulations starting from a coil structure.

converged. The convergence time predicted by  $u(t)$  (16 ns) is slightly longer than that predicted by the average  $\chi^2(t)$  (10 ns) since uniform sampling can only be achieved after the converged PEDFs have been applied for a while. On the other hand, Fig. 6(b) clearly shows more sampling at low temperatures for simulations starting from a coil structure, though the sampling starts to become uniform as time progresses. The  $u(t)$  curve for these simulations decrease from 0.8 to 0.3, but never falls below 0.2, indicating that not

all the PEDFs have converged within the 24 ns simulations. This observation is consistent with the prediction from the temperature specific  $\chi^2(t)$  measures (as opposed to the average  $\chi^2(t)$ ).

The uniform sampling ( $u(t)$ ) measure has the advantage that only one function is necessary to evaluate the convergence of all of the PEDFs, whereas the  $\chi^2(t)$  measure requires the user to monitor one curve for each temperature. For the  $u(t)$  measure to be valid, the time window has to be long enough for the system to explore the whole temperature space. One additional caveat is that there should not be any large gaps in the acceptance ratios, otherwise it will take a long time for the system to reach uniform sampling. Finally, we note that the  $u(t)$  measure is still an average measure and, therefore, may smooth out one or two deviant temperatures. However, the above analysis demonstrates that it is a more sensitive measure than the  $\chi^2(t)$  metric and should provide a reasonable gauge of the level of convergence.

## 2. Convergence of the weights in ST

As stated in Sec. II B, ST is an exact method regardless of the selection of the weights. However, ST simulations will be constrained to a subset of temperature space without proper weighing. A set of free energy weights will lead the system to perform a random walk in temperature space. In this study, we combined a method to obtain near free energy weights from trial simulations as described in Sec. II B with an adaptive weighting scheme using WHAM (Ref. 31) to determine the free energy weights. In this section, we investigate the convergence properties of the weights in our ST simulations.

Figure 7 shows the time evolution of  $\Delta g = g_{i+1} - g_i$  during simulations on the cluster for three pairs of neighboring temperatures: One at low temperature, one at high temperature, and one around the melting temperature. The differences between weights are plotted rather than the weights themselves as it is the differences that determine the acceptance ratios, as shown in Eq. (13). The weights obtained from our massive distributed computing ST simulations starting from different initial configurations are well converged, as shown in Table I. We note that the weights for temperatures below the melting temperature are not converged as well as the other

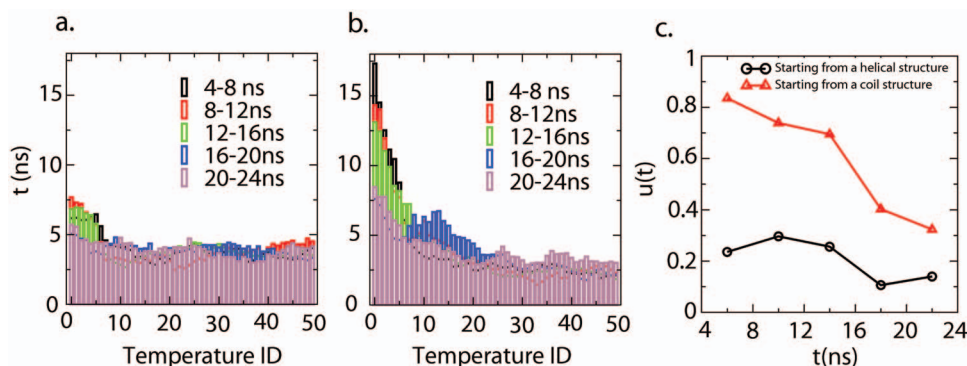


FIG. 6. (Color) (a). The amount of time each of the SREM simulations starting from a helical structure spent at each temperature. (b). The same as (a) except that data is collected from SREM simulations starting from a coil structure. (c). The uniform sampling convergence measure  $u(t)$  as a function of time obtained from simulations starting from a helical structure (black curve) and a coil structure (red curve). Each data point is averaged over the 50 SREM simulations using a time window of 4 ns.



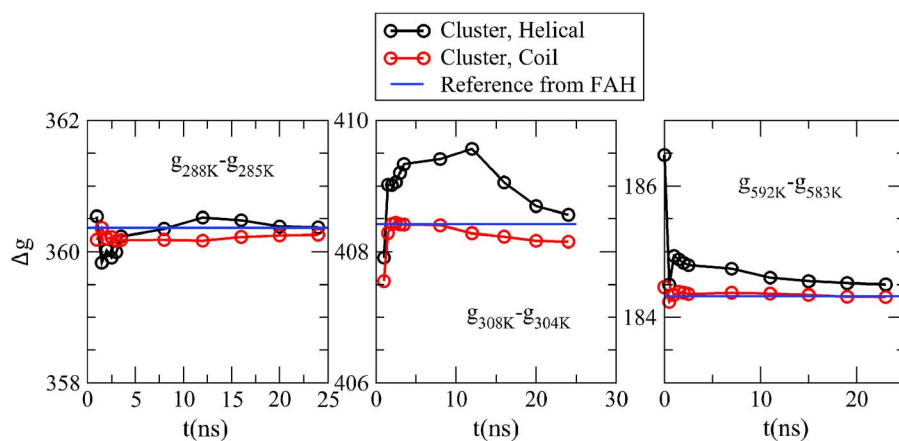


FIG. 7. (Color online)  $\Delta g = g_{i+1} - g_i$  from ST simulations as a function of time for three pairs of neighboring temperatures: 285 and 288 K (left column), 304 and 308 K (center column), and 583 and 592 K (right column). Black and gray (red) curves correspond to  $\Delta g$  values obtained from simulations starting from a helical and a coil structure, respectively. The well converged weights from distributed computing simulations are used as a reference (blue curve without symbols).

weights. Differences of up to 0.42 KJ/mol in free energy differences ( $\Delta f_{ji} = g_j / \beta_j - g_i / \beta_i$ ) are observed below the melting temperature but differences of not more than 0.1 KJ/mol occur elsewhere. This difference indicates that thermodynamic properties are more difficult to converge below the melting temperature. Regardless, the differences in  $\Delta f_{ji}$  are still much less than a  $KT$  so we consider them to be well converged (see Table I). Thus, we chose to use the weights from our distributed computing simulations starting from a prehelical configuration as our reference point for the convergence of weights from our cluster simulations. As shown in Fig. 7, the weights computed from short trial MD simulations are already close to the correct values (within 0.5%) for three temperature pairs. However, weights around the melting temperature ( $T=308$  K) still converge slower than those at other temperatures.

As discussed above, a set of converged weights will produce uniform sampling in ST. In Figs. 8(a) and 8(b), we again plot the amount of time spent at each temperature over a series of 4 ns windows. It is clear that the sampling is not uniform before 12 ns for simulations starting from either of the initial configurations. Before 12 ns, the system spends

considerably more time exploring high temperatures. This biased exploration of temperature space results from the fact that the weights are not converged. When the weights are not converged,  $\langle P_{ij} \rangle \neq \langle P_{ji} \rangle$ , resulting in nonuniform sampling. In Fig. 9, we plot  $P_{i \rightarrow i-1} - P_{i-1 \rightarrow i}$  for the transitions between all neighboring temperatures. When  $t < 12$  ns there is a significant difference between the forward and backward transition probabilities, especially around the melting temperature where  $i=6$ . After 12 ns the differences appear to be more randomly distributed around 0. These differences are probably due to the limitations of finite sampling rather than un-converged weights. To quantify the convergence of the weights we apply the same uniform sampling measure ( $u(t)$ ) as used for SREM. As shown in Fig. 8(c), the  $u(t)$  curves for each of the starting configurations decrease to a value below 0.2 after 16 ns, indicating that the weights are converged.

### 3. Convergence of helical properties

The convergence of various thermodynamic properties serves as another indirect measure of the convergence of our simulations. In theory, any thermodynamic property will de-

TABLE I.  $\Delta g = g_j - g_i$  obtained from distributed computing simulations starting from a helical structure (third column) and a coil structure (fourth column) at representative temperature pairs. In the fifth column, differences between free energy differences  $\Delta f_{ji} = g_j / \beta_j - g_i / \beta_i$  obtained from simulations starting from a helical structure and a coil structure are displayed.  $KT$  at temperature  $i$  is shown in the sixth column.  $\Delta f_{ji}(\text{helical}) - \Delta f_{ji}(\text{coil})$  is much smaller than  $KT$  at all temperature pairs.

$T_i$	$T_j$	$\Delta g_{ji}(\text{helical})$	$\Delta g_{ji}(\text{coil})$	$\Delta f_{ji}(\text{helical}) - \Delta f_{ji}(\text{coil})$ (KJ/mol)	$KT_i$ (KT/mol)
285	288	360.36	360.38	-0.06	2.37
288	292	466.48	466.39	0.22	2.39
292	296	451.13	450.95	0.42	2.43
296	300	436.41	436.30	0.28	2.46
300	304	422.25	422.29	-0.09	2.49
304	308	408.41	408.45	-0.09	2.53
308	312	394.65	394.69	-0.11	2.56
312	317	475.35	475.37	-0.06	2.59
317	322	457.33	457.33	-0.02	2.63
322	327	440.48	440.48	0.00	2.68
347	352	368.18	368.18	-0.01	2.88
399	405	314.38	314.39	-0.01	3.32
447	454	277.36	277.35	0.04	3.71
497	506	273.58	273.57	0.04	4.13
583	592	184.33	184.33	0.00	4.84

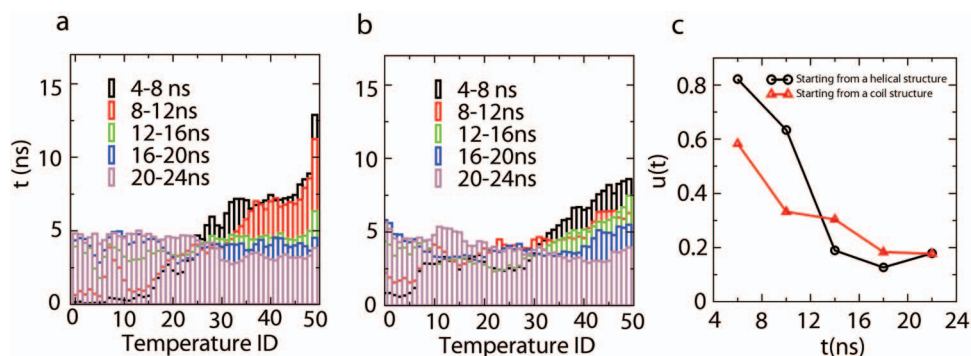


FIG. 8. (Color) (a). Amount of time the 50 ST simulations starting from a helical structure spent at each of the 50 temperatures. (b). The same as (a) except that data are collected from simulations starting from a coil structure. (c). The uniform sampling convergence measure  $u(t)$  as a function of time obtained from simulations starting from a helical structure (black curve) and a coil structure (red curve). Each data point is averaged over 50 ST simulations using a time window of 4 ns.

pend on the underlying PEDFs in SREM or weights in ST. In practice, however, thermodynamic properties may not converge at the same rate as these underlying variables. They are an attractive measure of convergence because, at least in this study, they are the primary quantities of interest. Measuring the convergence of thermodynamic properties also allows a direct comparison to MD.

In this study we will focus on the helical content. Other measures, such as number of helical residues, number of helical segments, and the maximum length of a single helix were also examined and showed similar trends (data not shown). To visualize the convergence of our simulations we plot the average helical content as a function of time for simulations starting from extended configurations and prehelical configurations, as shown in Fig. 10 for a few representative temperatures. Each point in Fig. 10 shows the average helical content over a 4 ns window with error bars corresponding to one standard deviation. The first 4 ns are omitted to allow for equilibration. The mean values and statistical uncertainties are computed by using WHAM (Ref. 28).

From Fig. 10(c) it is clear that both SREM and ST converge quickly at high temperatures and yield very small error bars. However, neither converges as quickly as the high temperature MD simulations. The small error bars make sense in light of the fact that the free energy surface is more smooth

at higher temperatures. In addition, the high fidelity of the convergence of MD compared to SREM and ST is probably due to the fact that the MD plot is for 50 24 ns simulations at constant temperature while 50 24 ns simulations are performed for all temperatures in SREM or ST.

At low temperatures SREM appears to converge more quickly than ST. This is not an inherent advantage of SREM, however. Rather, this phenomenon is the result of SREM being biased towards lower temperatures and ST being biased towards higher temperatures before their PEDFs and weights converge, as shown in Figs. 6 and 8. The fact that SREM converges faster than ST at lower temperatures in this case indicates that uniform sampling of temperature space may not be optimal. Various methods of exploiting this observation are now being implemented.

MD simulations starting from the prehelical configuration at low temperatures converge to the correct value quickly, most likely because the system tends to explore the free energy basin around the native structure. MD simulations starting from an extended configuration fail to converge to the correct value, however, because of the presence of large free energy barriers at low temperatures. The much faster convergence of our SREM and ST simulations starting from an extended state demonstrates that configurations accepted from high temperatures do indeed help sampling at lower temperatures.

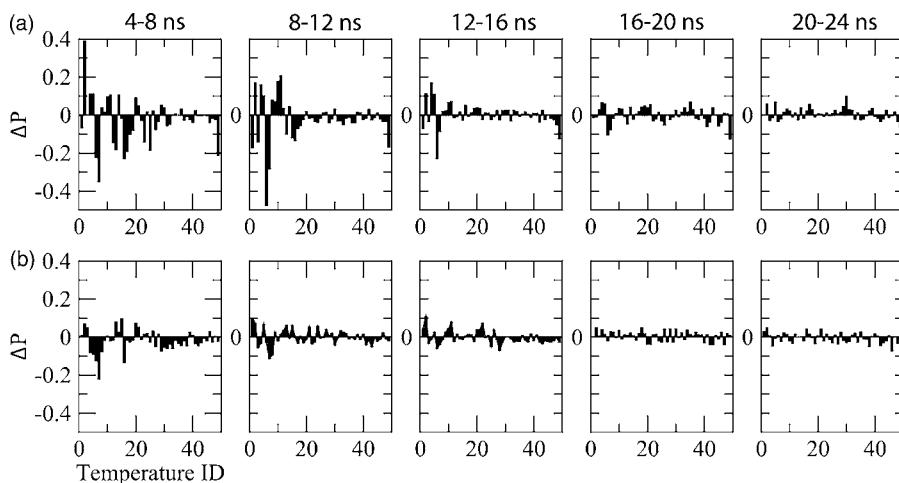


FIG. 9. Average probability difference ( $\Delta P$ ) between transitions to lower and higher temperatures over 50 ST simulations.  $\Delta P = P_{i \rightarrow i-1} - P_{i-1 \rightarrow i}$ . (a)  $\Delta P$  obtained from simulations starting from a helical structure. (b). The same as (a) except that  $\Delta P$  is obtained from simulations starting from a coil structure.

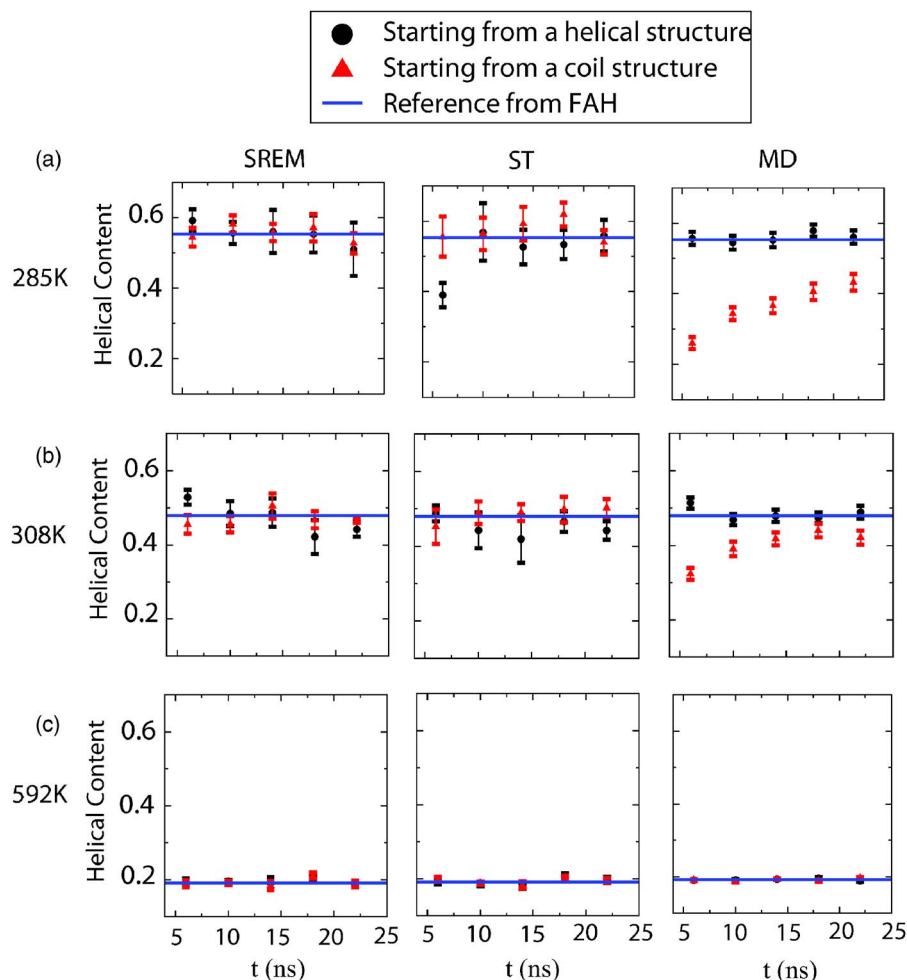


FIG. 10. (Color online) Comparison of helical content as a function of time between SREM, ST, and standard MD using the same amount of sampling at three temperatures: (a). 285, (b). 308, and (c) 592 K. (Black curve) Data obtained from simulations starting from a prehelical structure. (Gray/red curve) Data obtained from simulations starting from a prehelical structure. (Blue curve without symbols) The converged value from distributed computing simulations. The mean values and statistical uncertainties are estimated using WHAM. (Ref. 30).

Around the melting temperature both SREM and ST converge more slowly than at higher temperatures. SREM also converges more slowly than at lower temperatures, and this would probably be true for ST as well if it were not initially biased towards higher temperatures in this system. As for the PEDFs and weights, this slow convergence results from the broad energy distribution, more or less equal occurrence of the folded and unfolded states, and slow folding/unfolding transitions. Even at the melting temperature SREM and ST still converge within 10 ns, whereas MD fails to converge within the 24 ns simulations. Thus, both SREM and ST provide faster and more accurate thermodynamic predictions than MD. In this case they converge at least twice as quickly, though we cannot exactly quantify the advantage provided since our MD simulations fail to converge at all. Interestingly, the thermodynamic properties also converge faster than the PEDFs and weights (10 ns as opposed to 16 ns). Thus, convergence of the PEDFs and weights in SREM and ST is not necessarily a prerequisite for the convergence of thermodynamic properties of interest.

The weights in ST are converged after 16 ns regardless of the starting configuration. PEDFs from our SREM simulations converge after 16 ns starting from the prehelical configuration but take more than 24 ns to converge when started from a coil configuration. Regardless of these small differences, both algorithms are capable of reaching convergence within a small number of short simulations for this system

(50 24 ns simulations here). More interestingly, certain thermodynamic properties such as helical content converge even faster than the PEDFs and weights. Given the utility of these short simulations, a large number of simulations run in a distributed computing environment should be able to yield converged weights or PEDFs, even for a larger system. Thus, both of the enhanced sampling methods studied here should prove useful for studying the folding free energy landscapes of biomolecules in a distributed environment.

ST does have certain advantages over SREM. First, ST

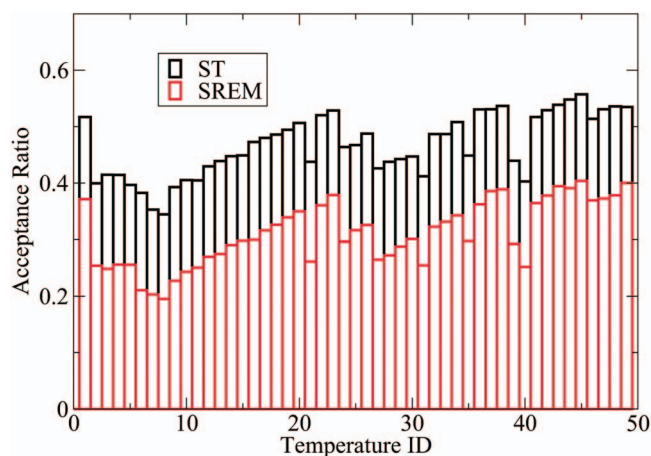


FIG. 11. (Color) Comparison of acceptance ratios from SREM and ST.

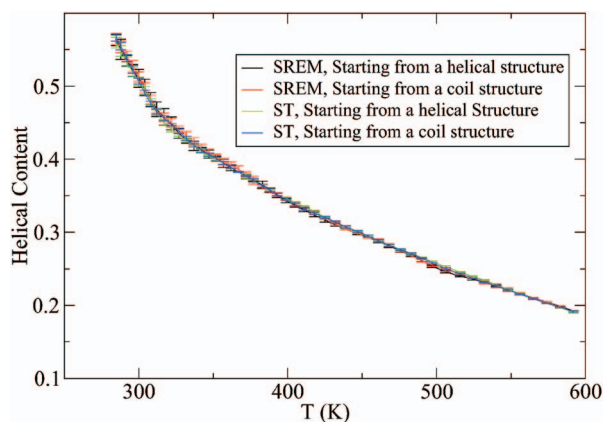


FIG. 12. (Color) Helical content as a function of temperature obtained from distributed computing simulations. (Black curve) SREM starting from a helical structure. (Red curve) SREM starting from a coil structure. (Blue curve) ST starting from a helical structure. (Green curve) ST starting from a coil structure. The mean values and statistical uncertainties are estimated using WHAM (Ref. 28).

is an exact method, while SREM is only approximately correct unless the exact PEDFs are used. Second, ST has higher acceptance ratios than SREM as shown in Fig. 11. This observation is in agreement with Park's recent analytic proof that ST should give higher acceptance ratios than SREM for any system given the same temperature list.<sup>26</sup> The primary difficulty with ST is determining the free energy weights. As discussed above, we show that obtaining initial weights from short trial simulations followed by refinement by an adaptive weighing method such as an adaptive WHAM (Ref. 31) very efficiently determines a set of free energy weights, which allows the system to perform a random walk in temperature space.

## B. Study of helical-coil transitions by SREM/ST in a distributed computing network

Distributed computing environments are able to generate a massive amount of sampling. Thus, converged PEDFs and weights can be more readily achieved, as shown in Fig. 2 and Table I. The convergence of these underlying variables leads to better convergence of thermodynamic properties. For ex-

ample, Fig. 12 shows helical content melting curves demonstrating that simulations starting from the prehelical and unfolded states converge excellently for both SREM and ST. Such exact convergence was not possible on our local computer cluster, where curves starting from different initial configurations converged only within very large error bars (data not shown).

In fact, Fig. 12 also shows that SREM and ST simulations run on a distributed computing environment give almost identical results. Each gives a melting temperature of about 310 K, in reasonable agreement with the experimental value of 334 K.<sup>44</sup> Again, this level of agreement was not found in our cluster simulations (data not shown). Thus, it appears that large-scale SREM simulations are able to yield converged PEDFs, leading to thermodynamic predictions with the same level of accuracy as ST even though SREM is an approximate method. However, for larger systems where converging the PEDFs may be more difficult, ST is probably still more useful since it is an exact method. Regardless, it is clear that SREM and ST are both powerful tools when employed on a distributed computing network.

At present there is no straightforward way to apply the  $u(t)$  convergence measure to FAH data due to its asynchronous nature. To ensure that uniform sampling was achieved 50 short 4 ns simulations were performed uniformly starting from different temperatures on a local cluster using the final FAH PEDFs in the case of SREM and the final FAH weights in the case of ST. Application of the  $u(t)$  metric to these data sets yielded values of 0.10 and 0.18 for SREM simulations starting from helical and coil structures, respectively, and values of 0.11 and 0.17 for ST simulations. These values are below the empirically determined cutoff of 0.2 used previously, indicating that near uniform sampling is achieved.

In addition to the average thermodynamic properties such as helical content, we also demonstrate the convergence of the free energy landscape starting from different initial configurations as projected onto the  $\phi$  and  $\psi$  axis. Figure 13 shows the convergence of the  $\phi$ - $\psi$  free energy landscape for both SREM and ST simulations at 308 K. Visual inspection of Fig. 13(a) shows that the free energy landscapes computed from ST simulations started from both prehelical and random

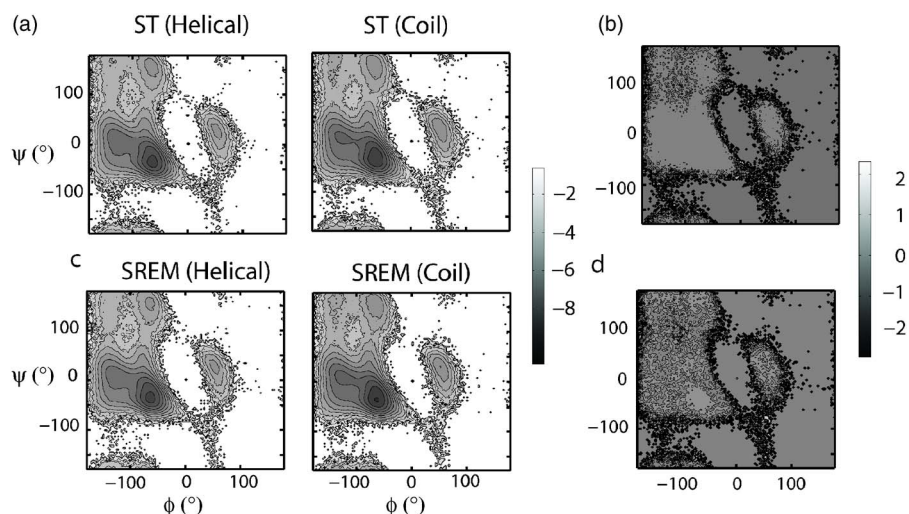


FIG. 13. (Color online) Free energy landscape projected onto the  $\Phi$  and  $\Psi$  axis in units of kcal/mol at 308 K. (a) Obtained from ST simulations started from the prehelical state (left figure) and from a random coil state (right figure). (b) The difference map between the free energy landscapes in (a). (c) The same as (a) except that free energy landscapes are obtained from SREM simulations. (d) The difference map between the free energy landscapes in (c).

coil states are almost identical. Figure 13(b) shows the difference map between the two free energy landscapes and quantitatively demonstrates the agreement between the two landscapes. Differences larger than  $kT$  mostly appear at the edges where the probability is low and therefore more variable. Results from SREM simulations are also converged as shown in Figs. 13(c) and 13(d). Furthermore, these free energy landscapes are in agreement with those from an earlier work employing ensemble MD simulations.<sup>5</sup>

The utility of the SREM and ST algorithms may be further increased by moving away from uniform sampling. Relative to short cluster simulations, our distributed simulations are approaching infinite sampling. In this case our melting curves are based on 12 times as much sampling while still throwing out 6 times as much data for equilibration. In fact, these numbers give an underestimate since other simulations not included in these calculations still contributed to the convergence of the PEDFs and weights. Aggregate simulation times for SREM and ST were 64.3 and 185.8  $\mu\text{s}$ , respectively. Even so, the error bars at lower temperatures, the primary area of interest, are still greater than those at higher temperatures as shown in Fig. 12. Adaptive sampling methods may serve to decrease the uncertainty at temperatures of interest even on a distributed computing network. One possibility would be to start more simulations from regions of temperature space that are not well converged. Another possibility would be to update the temperature list and or weights in order to bias simulations towards a specific set of temperatures.

### C. Temperature dependent properties

Sorin and Pande<sup>5</sup> define a helical residue as having  $\phi \in (-60 \pm 30)^\circ$  and  $\psi \in -47 \pm 30^\circ$ . They chose a cutoff of  $30^\circ$  because slight variations in this number caused the least difference in their results. Basing their cutoff on this variance relies on the assumption that the F<sub>s</sub> peptide is a two state folder. Two distinct state would imply some range of cutoffs that would give more or less the same results.

In this work, we use a cutoff of  $40^\circ$ . To arrive at this cutoff we started with the definition of the transition temperature, which is the temperature at which half the population is folded and the other half is unfolded. We assume that the folded state is completely helical while the unfolded state has no helical content, thus the helical content at the transition temperature would be 0.5. The helical content was then plotted as a function of temperature using cutoffs of  $30^\circ$ ,  $35^\circ$ ,  $40^\circ$ ,  $45^\circ$ , and  $50^\circ$ , as shown in Fig. 14. Both curves obtained from simulations starting from a helical structure and a coil structure converge well at every cutoff angle. The plots using a  $40^\circ$  cutoff are found to give 0.5 helical content at the transition temperature of 310 K, so  $40^\circ$  was selected as the optimal cutoff.

We also calculated the Lifson–Roig average helix nucleation and propagation propensity parameters,  $\langle\nu\rangle$  and  $\langle\omega\rangle$ , respectively, for our simulations. Our values for  $\langle\nu\rangle$  were an order of magnitude greater than the experimentally measured values (data not shown), in agreement with previous calculations.<sup>5,45</sup> This overestimate may be attributed to the

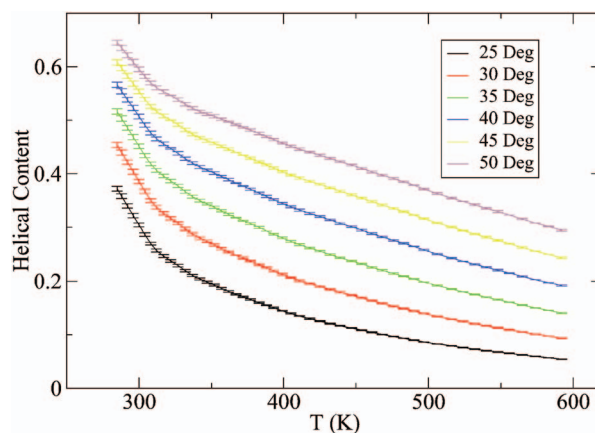


FIG. 14. (Color) Helical content as a function of temperature obtained from distributed computing ST simulations starting from a helical structure or a coil structure. Each curve uses a different cutoff angles to define a helical residue.

AMBER99 $\phi$  force field over stabilizing helices. As shown in Fig. 15, our results for  $\langle\omega\rangle$  are converged well for simulations starting from different configurations and more promising. At low temperatures we find reasonable agreement with the experimentally measured values. The overall trend of our  $\langle\omega\rangle$  curve is qualitatively similar to the experimentally measured curve, although our values fall off less quickly with increasing temperature. This can probably be attributed to some combination of the poor temperature dependence of the force field and the overstabilization of alpha helices at high temperatures.

### IV. CONCLUSION

Generalized ensemble methods such as the replica exchange method (REM) and simulated tempering (ST) have led to important advances in the sampling of conformations of biomolecular systems on computer clusters of all sizes. As demonstrated in this work, they can significantly speedup convergence relative to ensemble MD for the helical peptide studied here. However, it is still difficult to get convergence within the constraints of the computing power available on

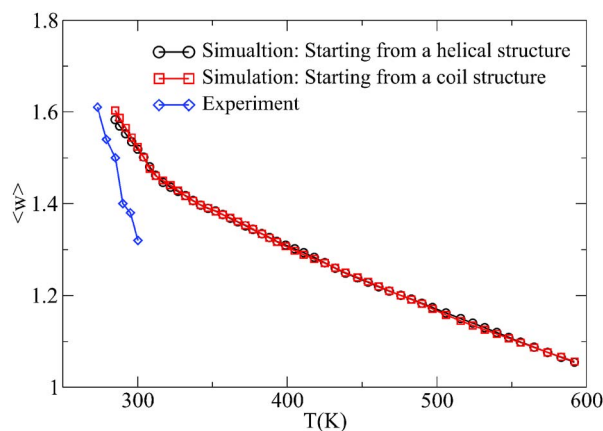


FIG. 15. (Color online) Average helix propagation parameters as a function of temperature obtained from distributed computing ST simulations starting from a helical structure (black circles), ST simulations starting from a coil structure (red squares), and from experiments (Ref. 47) (blue diamonds).

most local clusters. One common way of testing the convergence of such simulations is to determine whether the sampling is Boltzmann weighted. However, a simple thought experiment shows that it is possible to achieve Boltzmann weighted sampling of only a subset of the conformational space. A good measure of convergence is to achieve “reversible folding”<sup>46</sup> (i.e., multiple unfolding and refolding events in the same trajectory). However, even for a small protein, it is extremely difficult to achieve enough sampling to simulate reversible folding with all-atom force field in the explicit solvent. The difficulty is mostly due to the limit of the computing resources. With the large-scale distributed computing, we are able to gain higher level of convergence than those work using local clusters. Another less strict measure for the convergence is to run simulations starting from very different conformations, such as the prehelical and random coil states used in this work and compare the results as it demonstrates their robustness. Furthermore, it is important to look at a number of properties as some may converge faster than others. Besides thermodynamic properties, two obvious candidates are the PEDFs in SREM and the weights in ST.

Generalized ensemble methods have proved extremely powerful when taking these considerations into account in the context of a massive distributed computing environment. For ST, we demonstrate the efficacy of a new scheme for determining weights allowing the system to perform a random walk in temperature space that employs short trial simulations to make an initial estimate of the weights followed by massive simulations with an adaptive weighting method using WHAM. By making use of the extensive sampling capabilities of such an environment we are able to obtain converged energy distributions, weights, and temperature dependent thermodynamic properties. Both relatively short cluster simulations and large-scale distributed computing simulations demonstrate that SREM and ST converge to equivalent results within error on a small system. The performance of these methods may be improved through judicious use of nonuniform sampling. Thus, both methods could serve as powerful tools for further studies of more complicated systems on distributed computing networks. However, we note that ST may be more appropriate for larger systems because it is an exact solution, yielding valuable results even if a set of unconverged weights are used whereas SREM may not provide accurate predictions until after its PEDFs have converged.

## ACKNOWLEDGMENTS

This work would not have been possible without the FOLDING@HOME users who contributed invaluable processor time (<http://folding.stanford.edu>). We would like to thank Dr. Sanghyun Park, Dr. Guha Jayachandran, Dan Ensign, and Dr. Eric Sorin for their help on setting up SREM and ST at FOLDING@HOME. We would like to thank Dr. Morten Hagen and Professor Bruce Berne at Columbia University for helpful discussions. We would also like to thank A. Kulp for her excellent advice on comma placement and other editorial

comments. This work was funded by NIH R01-GM062868 and NSF MCB-0317072. Cluster resources were provided by NSF Award No. CNS-0619926. X.H. is supported by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant No. U54 GM072970. G.R.B. is supported by the NSF Graduate Research Fellowship Program. X.H. would also like to thank Professor Michael Levitt for his support.

- <sup>1</sup>K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
- <sup>2</sup>U. Hansmann and Y. Okamoto, *Curr. Opin. Struct. Biol.* **9**, 177 (1999).
- <sup>3</sup>C. Snow, H. Nguyen, V. S. Pande, and M. Gruebele, *Nature (London)* **420**, 102 (2002).
- <sup>4</sup>Y. Rhee, E. Sorin, G. Jayachandran, E. Lindahl, and V. S. Pande, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6456 (2004).
- <sup>5</sup>E. Sorin and V. S. Pande, *Biophys. J.* **88**, 2472 (2005).
- <sup>6</sup>G. Jayachandran, V. Vishal, A. E. Garcia, and V. S. Pande, *J. Struct. Biol.* **157**, 491 (2007).
- <sup>7</sup>G. Jayachandran, V. Vishal, and V. S. Pande, *J. Chem. Phys.* **124**, 164902 (2006).
- <sup>8</sup>E. J. Sorin and V. S. Pande, *J. Am. Chem. Soc.* **128**, 6316 (2006).
- <sup>9</sup>A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyainov, *J. Chem. Phys.* **96**, 1776 (1992).
- <sup>10</sup>E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451 (1992).
- <sup>11</sup>C. J. Geyer and E. A. Thompson, *J. Am. Stat. Assoc.* **90**, 909 (1995).
- <sup>12</sup>U. H. E. Hansmann, *Chem. Phys. Lett.* **281**, 140 (1997).
- <sup>13</sup>Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999).
- <sup>14</sup>R. H. Swendsen and J. S. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986).
- <sup>15</sup>R. Zhou, B. J. Berne, and R. Germain, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14931 (2001).
- <sup>16</sup>R. Zhou, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13280 (2003).
- <sup>17</sup>C. Simmerling, B. Strockbine, and A. E. Roitberg, *Proc. Natl. Acad. Sci. U.S.A.* **124**, 11258 (2002).
- <sup>18</sup>A. E. Garcia and K. Y. Sanbonmatsu, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2782 (2001).
- <sup>19</sup>C. Lin, C. Hu, and U. Hansmann, *Proteins* **52**, 436 (2003).
- <sup>20</sup>A. Schug, T. Herges, and W. Wenzel, *Proteins* **57**, 792 (2004).
- <sup>21</sup>W. Zheng, M. Andrec, E. Gallicchio, and R. Levy, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15340 (2007).
- <sup>22</sup>Y. Rhee and V. Pande, *Biophys. J.* **84**, 775 (2003).
- <sup>23</sup>E. Gallicchio, R. M. Levy, and M. Parashar, *J. Comput. Chem.* **29**, 788 (2008).
- <sup>24</sup>M. Hagen, B. Kim, P. Liu, R. A. Friesner, and B. J. Berne, *J. Phys. Chem. B* **111**, 1416 (2007).
- <sup>25</sup>S. Park and V. S. Pande, *Phys. Rev. E* **76**, 016703 (2007).
- <sup>26</sup>S. Park, “Weights and acceptance ratios in generalized ensemble simulations,” *Phys. Rev. E* (submitted).
- <sup>27</sup>S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, *J. Comput. Chem.* **13**, 1011 (1992).
- <sup>28</sup>J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill, *J. Chem. Theory Comput.* **3**, 26 (2007).
- <sup>29</sup>A. Mitsutake and Y. Okamoto, *Chem. Phys. Lett.* **332**, 131 (2000).
- <sup>30</sup>S. Park, D. L. Ensign, and V. S. Pande, *Phys. Rev. E* **74**, 066703 (2006).
- <sup>31</sup>C. Bartels and M. Karplus, *J. Comput. Chem.* **12**, 1450 (1997).
- <sup>32</sup>F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).
- <sup>33</sup>H. Li, M. Fajer, and W. Yang, *J. Chem. Phys.* **126**, 024106 (2007).
- <sup>34</sup>M. Shirts and V. S. Pande, *Science* **290**, 1903 (2000).
- <sup>35</sup>N. Rathore, M. Chopra, and J. J. de Pablo, *J. Chem. Phys.* **122**, 024111 (2005).
- <sup>36</sup>H. Tijms, *Understanding Probability: Chance Rules in Everyday Life* (Cambridge University Press, Cambridge, 2004).
- <sup>37</sup>J. Wang, P. Cieplak, and P. A. Kollman, *J. Comput. Chem.* **21**, 1049 (2000).
- <sup>38</sup>E. Lindahl, B. Hess, and D. van der Spoel, *J. Mol. Model.* **7**, 306 (2001).
- <sup>39</sup>W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).
- <sup>40</sup>W. Hoover, *Phys. Rev. A* **31**, 1695 (1985).
- <sup>41</sup>B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, *J. Comput. Chem.* **18**, 1463 (1997).

<sup>42</sup>S. Lifson and A. Roig, *J. Chem. Phys.* **34**, 1963 (1961).

<sup>43</sup>H. Qian and J. A. Schellman, *J. Phys. Chem.* **96**, 3987 (1992).

<sup>44</sup>S. Williams, T. P. Causgrove, R. Gilmanshin, K. S. Fang, R. H. Callender, W. H. Woodruff, and R. B. Dyer, *Biochemistry* **35**, 691 (1996).

<sup>45</sup>H. Nymeyer and A. Garcia, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13934 (2003).

<sup>46</sup>F. Rao and A. Caflish, *J. Chem. Phys.* **119**, 4035 (2003).

<sup>47</sup>C. A. Rohl and R. L. Baldwin, *Biochemistry* **36**, 8435 (1997).