Serial Replica Exchange

Morten Hagen,† Byungchan Kim,† Pu Liu,† Richard A. Friesner, and B. J. Berne*

Department of Chemistry and Center for Biomolecular Simulation, Columbia University, New York, New York 10027

Received: July 14, 2006; In Final Form: December 7, 2006

Parallel tempering (or the replica exchange method (REM)) is a powerful method for speeding up the sampling of conformational states of systems with rough energy landscapes, like proteins, where stable conformational states can be separated by large energy barriers. The usual implementation of the REM is performed on local computer clusters (or parallel processors) where the different replicas must be run synchronously. Here, we present serial replica exchange (SREM), a method that is equivalent to the standard REM in terms of efficiency yet runs *asynchronously* on a distributed network of computers. A second advantage is the method's greatly enhanced fault tolerance, which enables the study of biological systems on worldwide distributed computing environments, such as Folding@Home.[1] For proof of concept, we apply the SREM to a single alanine dipeptide molecule in explicit water. We show that the SREM reproduces the thermodynamic and structural properties determined by the REM.

1. Introduction

Sampling the conformation space of complex systems, such as proteins, is a notoriously difficult problem in structural biology and theoretical chemistry. The difficulty arises from infrequent crossings of high energy barriers between local energy minima, leading to local trapping for long times and concomitant quasi-ergodicity in the sampling. The parallel tempering (or replica exchange method) was devised to overcome the problem of quasi-ergodicity.[2−4] In this method, molecular dynamics or Monte Carlo simulations are performed simultaneously on independent replicas each at different temperatures. Exchanges of configuration between nearest neighbor replicas are attempted periodically and accepted with a well-defined acceptance probability which is consistent with detailed balance. By occasionally obtaining the configuration from high temperature replicas, the replicas at low temperature can overcome their energy barriers. In addition to reducing the problem of quasi-ergodicity, the temperature dependence of various thermodynamic properties can be obtained. The replica exchange method (REM) has proven to be a useful method for sampling phase space,[5,6] and has been widely adopted.

In the standard replica exchange method, the number of replicas needed increases as $O(f^{1/2})$, where f is the system's total number of degrees of freedom.[7] Even for a relatively small biomolecular system consisting of one β-hairpin protein molecule dissolved in water (4342 atoms in all), 64 replicas were needed to cover the temperature range between 270 and 695 K with a nonvanishing acceptance ratio for replica exchange.[8] This severely restricts the applicability of the REM to reasonably small systems, unless one has access to either a massively parallel computer or a very large dedicated cluster on which a large number of replicas can run synchronously. Furthermore, in its usual implementation, the REM is a synchronous algorithm in that each replica must evolve for the same amount of time before the replica exchange is attempted. In this article, we present a strategy for running the equivalent of the REM on one computer or on a collection of distributed computers, which allows one to use these computers in an asynchronous manner. We call this the serial replica exchange method (SREM).

In addition, we show that this same strategy can be applied to another replica exchange method, replica exchange with solute tempering[9] (REST). Because the number of CPUs needed in this new scheme depends only on the degrees of freedom (DOF) in the solute molecule, and not the number of DOF of the water molecules, the number of CPUs needed can be greatly reduced; for example, for a single alanine dipeptide in an explicit water box, only 3 replicas were required, instead of the 16 needed in the ordinary REM. Clearly, REST also requires synchronous use of the processors. We show how the strategy embodied in the SREM can also be applied to REST, and we call the resulting algorithm "serial replica exchange with solute tempering", or SREST.

Since the REM (and REST) requires synchronization between different replicas in order to exchange replicas, running it on a heterogeneous network of nondedicated distributed computers is intrinsically inefficient, as its overall performance is determined by the slowest replica. With the advent of distributed networks such as Folding@Home and SETI@Home, it is important to free the REM from reliance on either a large local cluster or a massively parallel computer, the current requirement for simulation of large biological systems. For example, Folding@Home currently has more than 180 000 hosts, substantially more than any other computing cluster. The enormous computational power of Folding@Home makes sub-millisecond simulations of some miniproteins with explicit water possible.[10] The SREM and SREST provide a simple approach to harness the inherent power in such heterogeneous computer resources for sampling conformational states and computing thermodynamic properties of complex biological systems.

For large scale applications, biological systems in particular, the number of replicas/processors required in the REM, which generally scales as $O(f^{1/2})$, can become exceedingly large. This

* To whom correspondence should be addressed. E-mail: berne@chem.columbia.edu.

† These authors contributed equally to this work and are listed in alphabetical order.

10.1021/jp064479e CCC: $37.00
Published on Web 01/24/2007

has significant implications for the fault tolerance of the application, as the synchronous REM is typically run on inexpensive, loosely coupled computing clusters where the individual failure rate can be quite high. Indeed, if any one processor fails, then the entire simulation with N processors will fail. Hence, the time to failure for the REM is the minimum of the time to failure for N *independent* processors.[11] This implies that the rate of failure for the REM is N times higher than that for the SREM.

The SREM and SREST are described in the next section, followed by their application to an alanine dipeptide solvated in explicit water. The results show that the new serial methods reproduce the thermodynamic properties with nearly the same efficiency as the REM.

2. Methods

2.1. Replica Exchange. Replica exchange (or parallel tempering) involves running Monte Carlo (or constant temperature molecular dynamics) for a certain number of passes (or time steps) in parallel on a set of replica systems each at a different temperature, $\{T_1, T_2, \ldots, T_N\}$, where the temperatures are ordered from the lowest T_1 to the highest T_N. At the end of this period, an attempt is made to exchange the configurations of a pair of neighboring replicas and this exchange is accepted with a probability that satisfies detailed balance

$$a(i \rightarrow f) = \begin{cases} 1 & \text{if } \Delta_{nm} \leq 0 \\ \exp(-\Delta_{nm}) & \text{if } \Delta_{nm} > 0 \end{cases} \quad (1)$$

where $\Delta_{nm} = \Delta\beta\Delta E$, $\Delta\beta = 1/(kT_m) - 1/(kT_n)$, $\Delta E = E(X_n) - E(X_m)$, and $E(X_n)$ is the system energy for the nth replica with configuration X_n. The process is then repeated. The highest temperature, T_N, is chosen so that its replica can rapidly cross the potential energy barriers. Because configurations sampled at the high temperatures can, in principle, eventually exchange with the low temperature replicas, the low temperature systems will experience jumps between potential basins separated by high barriers, something they would not be able to do easily in ordinary Monte Carlo or molecular dynamics (MD).

For replica exchange with solute tempering (REST), different replicas can have not only different temperatures but also different potential functions, $\{E_1(X_1), E_2(X_2), \ldots, E_N(X_N)\}$, where X_n represents the configurational coordinates of the nth replica system, and for our choice of these energy surfaces, the acceptance probability differs from the above in that $\Delta E = E_p(X_n) - E_p(X_m) + {}^1/_2(E_{pw}(X_n) - E_{pw}(X_m))$, where $E_p(X_n)$ is the energy within the protein group (central group) and $E_{pw}(X_n)$ is the interaction energy between the protein group and the water group (surrounding group).

From the acceptance probability (cf. eq 1 for the REM and REST), only the *energy*, a *one-dimensional* property, is required in order to determine whether the attempted replica exchange is successful or not. The SREM and SREST are derived from this property.

Suppose one knows the potential energy distribution (PEDF), $P_n(E; T_n)$, at the temperatures, $\{T_1, T_2, \ldots, T_N\}$. Suppose further that one initiates a Monte Carlo (or constant temperature MD) walk at temperature T_n, runs this for a certain desired number of steps, and determines its energy to be E. From the energy distribution $P_{n+1}(E; T_{n+1})$, one could sample an energy E' for a system at the neighboring temperature, attempt a move to this neighboring temperature, say T_{n+1}, and accept or reject this move using the same acceptance probability as in eq 1. If this move is accepted, the walk is now continued at the new temperature, and otherwise, it is continued at the old temperature, and so on. In this way, a single walker samples configurations at all of the temperatures and averages can be determined at each temperature.

We believe that the SREM is as efficient as the REM; that is, if we run N independent SREM simulations for the same amount of time as each of the N replicas in the REM, then the overall information obtained is equivalent to that obtained from the REM. In the Results section, we demonstrate that this is approximately true for our model applications.

2.2. Determination of the PEDF. The main issue is that we do not know the required energy distributions a priori. However, it is possible to obtain first approximations to the PEDFs, even without communication between processors, by performing separate runs at each temperature, that is, without attempting temperature jumps. Another approach, recently proposed by Mitsutake and Okamoto,[12,13] is to perform a short preliminary replica exchange run. Alternatively, one could initially employ Wang−Landau sampling[14−16] at each temperature to obtain PEDFs of very high accuracy. However, it was recently pointed out[17] that, even with substantial modifications of the original algorithm[18] to allow for nonparametric estimation of the *continuous* distributions that typically occur in systems of biological relevance, Wang−Landau sampling is highly dependent on proper tuning of the various parameters, particularly the schedule for reducing the modification factor, f. The same study[17] argued that, in comparison, replica exchange is much less sensitive to the choice of parameters.

As a result, we have chosen the following approach based on replica exchange, which we deem sufficiently flexible to handle most cases of interest:

(i) Generate an approximate energy distribution at each temperature using the REM (or REST) or MD.

(ii) Propagate the system for a certain amount of time using the SREM (or SREST) and collect the potential energies at each temperature.

(iii) Update the PEDFs by using the sampled potential energies.

(iv) Repeat steps ii and iii until the PEDFs are stationary.

Once the energy distributions have converged, we can stop updating them and begin the production run. In this production phase, every replica run on a different processor will, in principle, sample all of the different temperatures, without communicating with any other processor. The procedure outlined for accepting temperature jumps is essentially the same as that in standard replica exchange, implying that the system should spend a roughly equal amount of time at each temperature level.

2.3. Representation of the PEDF. The PEDFs for biological systems are continuous and, according to the central limit theorem, approximately Gaussian. They are thus excellent candidates for parametric density estimation. However, for low dimensional systems, this is not the case. To demonstrate the general applicability of our method, we have consequently chosen to focus on nonparametric density estimation.[19]

For continuous distributions, the kernel density estimator[20] is the method of choice, often using the Gaussian kernel function. Another frequently used approach is the orthogonal series estimator.[21] However, regardless of the choice of statistical method, there is a tradeoff between the precision and the accuracy of the estimated PEDFs. Either way, the estimated PEDFs will have some inherent error relative to the true PEDFs. For this reason, and to avoid unnecessary detail, we limit the discussion to the simplest possible nonparametric estimator, the histogram with fixed bin width. While this is *not* an optimal

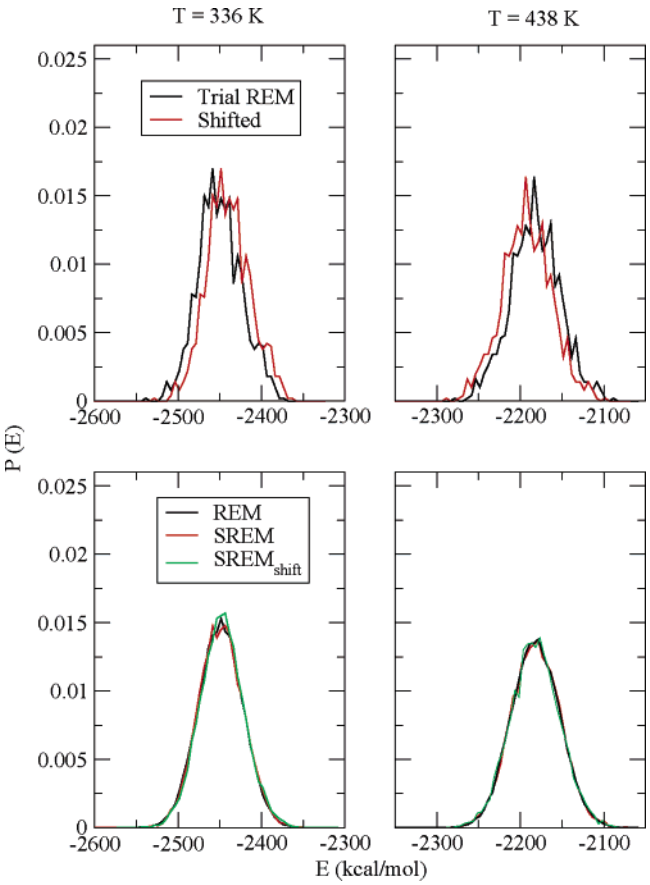

Figure 1. Potential energy distributions for two temperatures, 336 K (left column) and 438 K (right column). The black and red curves in the top panels correspond to the original rough PEDFs from the trial REM simulations and the shifted counterpart. The black, red, and green curves in the bottom panels show the converged PEDFs from a 9 ns REM simulation, the 2 ns aging process of the SREM simulation, and the 2 ns aging process of the SREM$_{shift}$ simulation, respectively. The SREM and SREM$_{shift}$ results are averaged over 16 independent runs.

density estimator,[19] it has been found to work well for Gaussian samples.[22] In the Appendix, we investigate the effect of bin width on non-Gaussian samples and conclude that histograms are viable as general purpose estimators.

2.4. Verification of the PEDF. The algorithm provided above states that the updating phase should continue until the PEDFs are stationary. To verify this, we employ the integrated squared error,[19] a chi-square type measure,

$$\chi^2(t) = \sum_{i=1}^{N} (\bar{P}_i(t) - \bar{P}_i^{\mathrm{ref}})^2 \qquad (2)$$

where $\bar{P}_i(t)$ is the ith element of the potential energy histogram averaged over time t, N is the number of bins in the PEDF histogram, and $\bar{P}_i^{\mathrm{ref}}$ is the reference PEDF. Assuming that $\bar{P}_i^{\mathrm{ref}}$ is the true distribution, we can continually monitor the evolution of the PEDFs. However, this has the drawback that determining $\bar{P}_i^{\mathrm{ref}}$ typically requires a lengthy REM simulation, thus defying the main purpose of the SREM. Instead, we propose to use the $\chi^2(t)$ measure above, with $\bar{P}_i^{\mathrm{ref}}$ being the approximate PEDF determined by the short initial REM simulation. This quantity can be monitored during the simulation without any additional overhead. The idea is that once $\chi^2(t)$ appears to plateau, the PEDFs have become stationary, and the updating phase can be terminated. This sounds too good to be true, and, to some extent, it is. The criterion only guarantees that the PEDFs have become

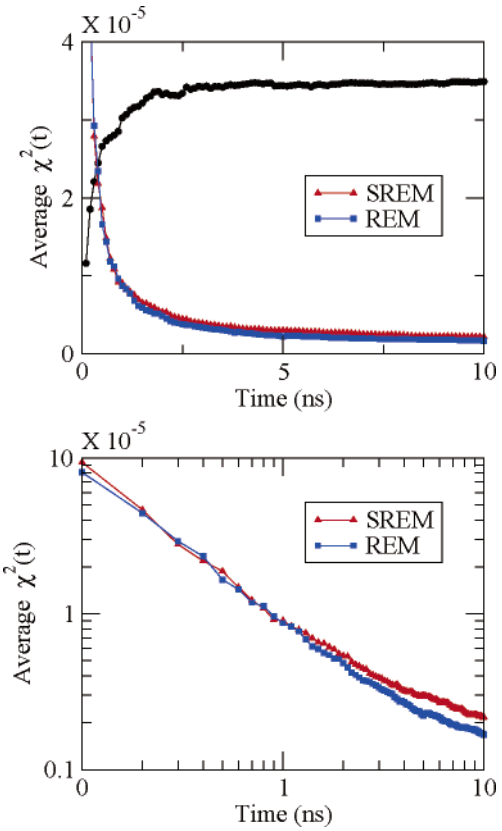

Figure 2. Convergence measure of the potential energy distributions for the REM and SREM, as a function of the simulation time. The curves of $\chi^2(t)$, using $\bar{P}_{\mathrm{initial}}^{\mathrm{ref}}$ and $\bar{P}_{\mathrm{final}}^{\mathrm{ref}}$ as the reference states, are plotted in black and blue (REM)/red (SREM), respectively. The $\chi^2(t)$ values are averaged over all temperature levels. The SREM values are obtained from 16 concatenated independent runs. We stop updating the PEDFs after 2 ns in each simulation.

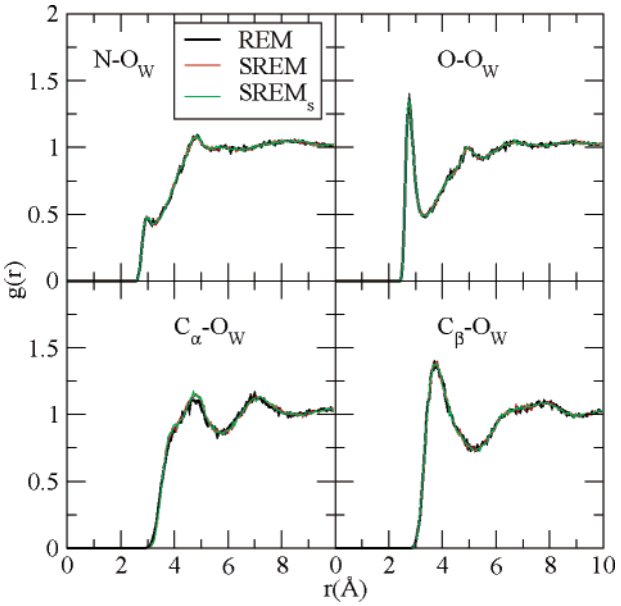

Figure 3. Comparison of the radial distribution functions (RDFs) between selected heavy atoms of alanine dipeptide and oxygens of water molecules at 300 K. RDFs from REM, SREM, and SREM$_{shift}$ simulations are plotted in black, red, and green, respectively. The selected heavy atoms are indicated in the legends. The SREM and SREM$_{shift}$ results are averaged over 16 independent runs.

stationary, perhaps with some locked in deviation from the true distribution. However, since the replicas are, using the detailed balance criterion above, able to move to higher temperatures

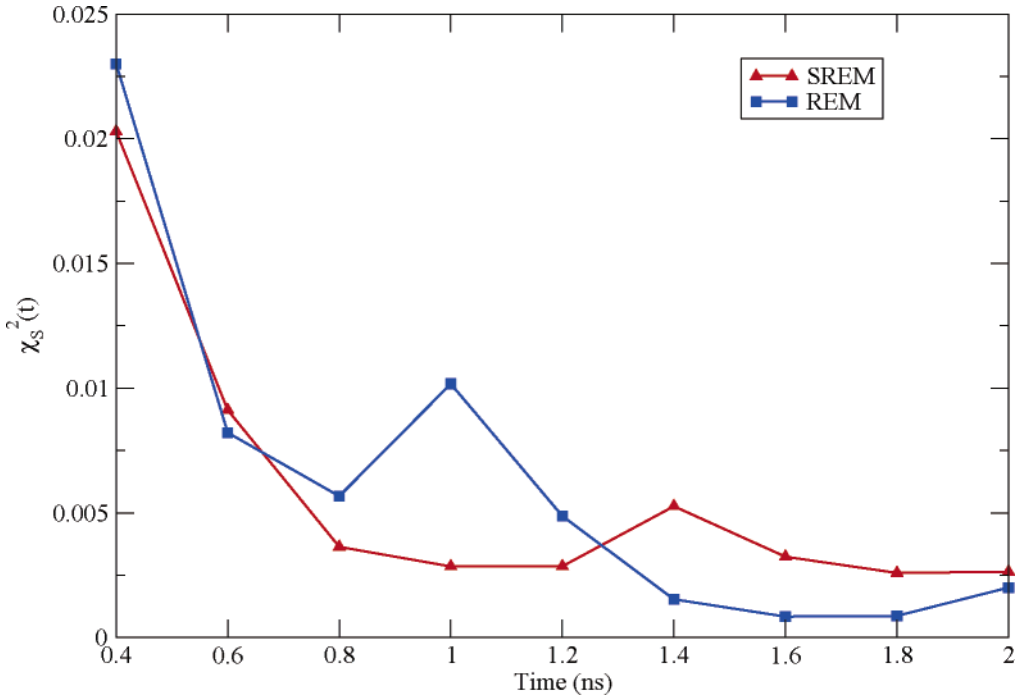

Figure 4. Convergence measure of the spatial distributions for the REM and SREM at 300 K, as a function of the simulation time. The curves of $\chi_S^2(t)$, using a 9 ns REM simulation as the reference, are plotted in blue (REM) and red (SREM). The SREM values are obtained from 16 concatenated independent runs.

during the updating phase, we do not expect any lingering issues of quasi-ergodicity. Hence, once stationary, and barring exceptional quasi-ergodicity that persists even with the use of a replica exchange scheme, the PEDFs will be very close to their true distribution. Insofar as quasi-ergodicity might persist, we believe that this condition would also afflict the production run, whether using the REM, the SREM, or some other simulation scheme. The inability to detect such quasi-ergodicity without prior knowledge of the true distributions is manifest in all simulations, and is thus not particular to the SREM.

3. Results

3.1. Peptide in Water. As a test case, we have applied the SREM to an aqueous solution consisting of one small peptide molecule, alanine dipeptide, solvated in 256 TIP4P[23] water molecules. The potential model for alanine dipeptide was taken to be the OPLS-AA/L force field.[24,25] We have simulated the system with cubic periodic boundary conditions using the P3ME method[26−28] for calculating the electrostatic interactions. The internal geometries of water molecules and the bond lengths of the alanine dipeptide were constrained with the RATTLE algorithm,[29] allowing a time step of 2 fs. The system temperature was controlled by Nose−Hoover chain thermostats.[30]

To obtain initial estimates of the potential energy distribution functions (PEDFs), we ran a REM simulation for 100 ps using 16 replicas over a range of temperatures from 300 to 600 K. The temperature spacing was distributed roughly exponentially, with the corresponding exchange acceptance ratios between neighboring replicas ranging from 22 to 29%. Replica exchanges were attempted every 2 ps during the simulation. Subsequently, 16 independent SREM simulations were performed using the PEDFs obtained from the initial REM simulation. During the SREM aging phase, the PEDFs were updated every 100 ps for 2 ns, at which point they became roughly stationary. Subsequently, an SREM production run of 9 ns was performed to collect the thermodynamic and structural data. For comparison, we ran a standard REM simulation with 16 replicas, also for 9 ns.

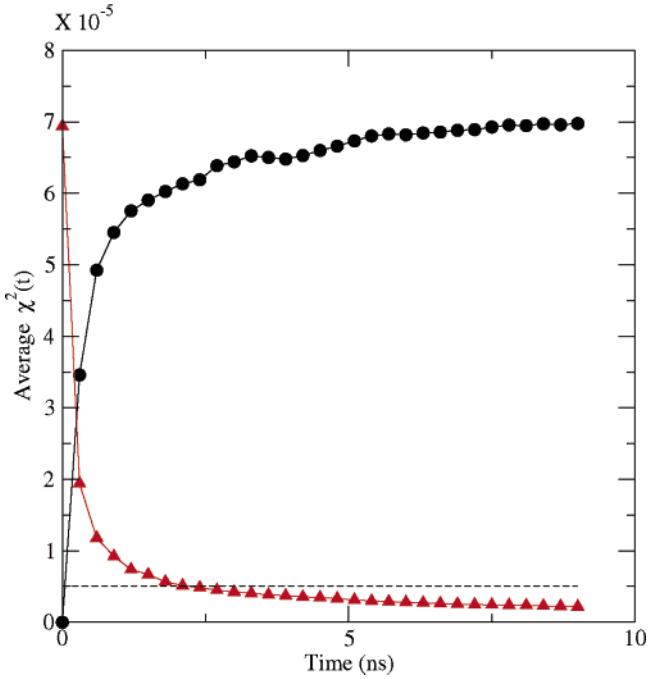

Figure 5. Convergence measure of the potential energy distributions for $SREM_{shift}$ simulations, as a function of the SREM simulation time. The curves of $\chi^2(t)$, using $\bar{P}^{ref}_{initial}$ and $\bar{P}^{ref}_{final}$ as the reference states, are plotted in black and red, respectively. We stop updating the PEDFs after 2 ns in each simulation. The $\chi^2(t)$ values are obtained from 16 concatenated independent runs and averaged over all temperature levels.

We investigate the convergence properties of the SREM PEDFs at two representative temperatures, 336 and 438 K, in Figure 1, where we compare the initial PEDFs from the trial REM simulation (the black curve in the top panel) with the PEDFs resulting from the 2 ns SREM aging phase (the red curve in the bottom panel) and that of a 9 ns REM reference simulation

TABLE 1: Comparison of the Average Percentage Populations, and Their Standard Deviations, of Four Regions in the Ramachandran Plot at 300 K[a]

region	REM	SREM	$SREM_{shift}$
P_{II}	0.45 ± 0.08	0.47 ± 0.07	0.46 ± 0.07
α_R	0.23 ± 0.08	0.20 ± 0.06	0.21 ± 0.05
β	0.17 ± 0.05	0.19 ± 0.03	0.18 ± 0.03
α'	0.15 ± 0.04	0.14 ± 0.04	0.15 ± 0.04

[a] The results were collected over 9 ns for each of the REM, SREM, and $SREM_{shift}$ simulations.

(the black curve of the bottom panel). Although the PEDFs from the initial short REM simulation are rough and noisy, the PEDFs from the 2 ns aging phase of SREM simulations have nearly converged to those of the 9 ns REM simulation. The PEDFs of the SREM at other temperatures show a similar behavior, suggesting that the final results are independent of the choice of the initial PEDFs, as described above.

To investigate how quickly the PEDFs approach equilibrium, we employ the $\chi^2(t)$ measure defined in the previous section and compare two different choices of the reference PEDF, that is, the PEDF of the initial 100 ps trial simulation, $\bar{P}^{ref}_{initial}$, versus the PEDF of a 9 ns REM reference simulation, $\bar{P}^{ref}_{final}$. If the SREM simulation eventually reproduces the correct PEDF for the corresponding ensemble, $\chi^2(t)$ should gradually decay to zero when $\bar{P}^{ref}_{final}$ is used as the reference state. On the other hand, when $\bar{P}^{ref}_{initial}$ is used as the reference state, $\chi^2(t)$ should grow until it plateaus at a value corresponding to the χ^2 measure of $\bar{P}^{ref}_{final}$ with respect to $\bar{P}^{ref}_{initial}$. The top panel of Figure 2 displays both $\chi^2(t)$ measures, concatenating 16 indepdent SREM simulations and averaging over all temperature levels. To compare the SREM with the REM, we compute the chi-square measure of a second, independent 9 ns REM simulation versus the 9 ns REM reference simulation. The lower panel reveals a minor difference in the convergence properties of the SREM and REM, which is related to the discrete representation of the PEDF in the SREM. We address this issue in the Appendix.

If the simulation time is shorter than the characteristic time of barrier crossing, the PEDFs from the short trial REM simulation will not be accurate. To probe this issue, we have performed SREM simulations initialized with PEDFs that deviate from the true PEDFs. This is illustrated in Figure 1, where we have shifted each of the initial PEDFs by an arbitrary constant, denoted by $SREM_{shift}$. Starting from the shifted PEDFs, we have performed 16 SREM simulations. The PEDFs were updated periodically during a 2 ns aging phase, followed by a 9 ns production run for data collection. Figure 1 also shows the PEDFs from the 2 ns $SREM_{shift}$ aging stage. Although the exchange is initially controlled by the incorrect PEDFs, $SREM_{shift}$ eventually reproduces the PEDFs obtained from the REM. Figure 5 demonstrates that the convergence measure has roughly the same behavior as that without the shift, that is, when compared to Figure 2.

To test the accuracy of the new scheme, we compare the radial distribution functions (RDFs) of the SREM to those of the REM at 300 K. Figure 3 shows the RDFs between some selected heavy atoms of alanine dipeptide and oxygens of water molecules. The two RDFs (the black curve for the 9 ns REM simulation and the red curve for the 9 ns production run of the SREM within each subplot) are very similar to each other. The results clearly show that SREM sampling correctly determines the structural properties described by the RDF. As a further measure, we compare the populations of the pronounced peaks in the Ramachandran plots, that is, the population distributions

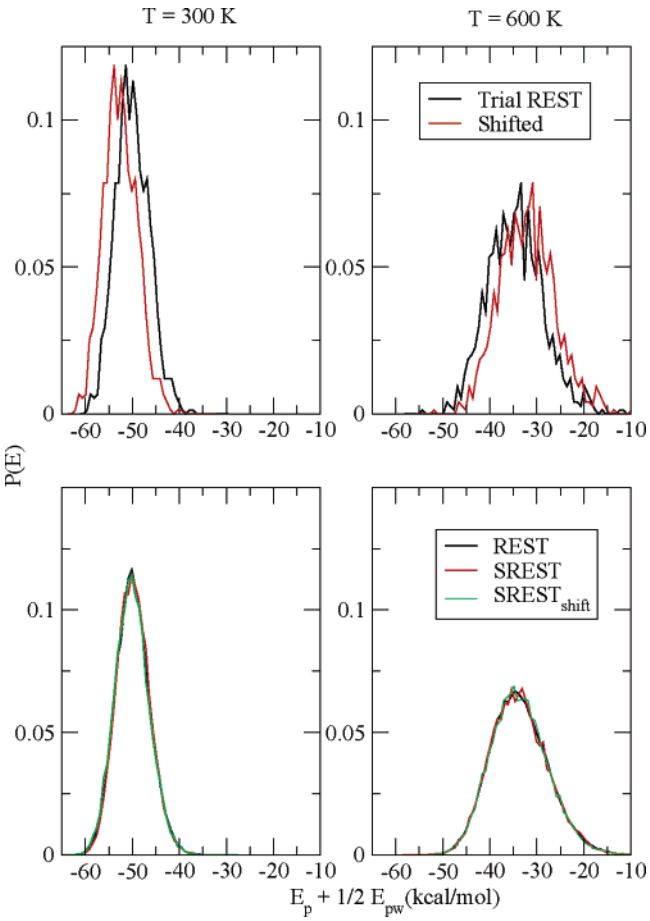

Figure 6. Potential energy distributions for two temperatures, 300 K (left column) and 600 K (right column). The black and red curves in the top panels correspond to the original rough PEDFs from the trial REST simulations and the shifted counterpart. The black, red, and green curves in the bottom panels show the converged PEDFs from a 9 ns REST simulation, the 2 ns aging process in SREST, and the 2 ns aging process in the $SREST_{shift}$ simulation, respectively. The SREST and $SREST_{shift}$ results are averaged over three independent runs.

of the ϕ and ψ backbone dihedral angles of alanine dipeptide, of the SREM to those of the REM at 300 K. As in our previous work, four regions, P_{II}, α_R, β, and α', ordered by populations, were identified as highly populated regions. Table 1 shows that the populations for the REM, SREM, and $SREM_{shift}$ simulations are comparable within the uncertainty. For completeness, we have computed a chi-square measure related to the spatial Ramachandran plot, the integrated squared error of the spatial density,

$$\chi_S^2(t) = \sum_{j=1}^{K} (\rho_j(t) - \rho_j^{exact})^2 \qquad (3)$$

where $K = 16$ is the number of bins in the Ramachandran plot of the peptide system. An approximation to the exact density, ρ_j^{exact}, is obtained from a 9 ns REM reference simulation. To provide a meaningful measure, the REM reference simulation must be substantially longer than the simulations for which we compute $\chi_S^2(t)$. Hence, in Figure 4, we show the convergence of the SREM simulation up to 2 ns.

In summary, it appears that SREM sampling of both energetic and structural properties is not greatly affected by inaccuracies in the initial PEDFs, provided that adequate aging runs are made. Indeed, our procedure of periodically updating the PEDFs during the aging phase was designed to allow the PEDFs to relax to the accurate PEDFs.

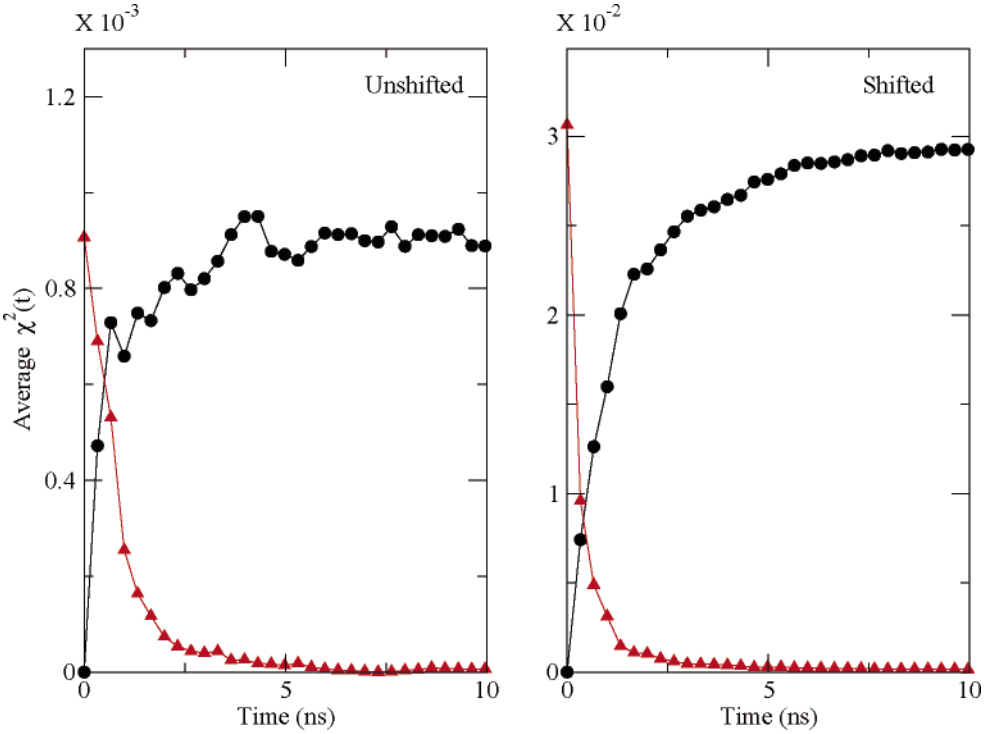

Figure 7. Convergence measure of the potential energy distributions as a function of the CPU time at 300 K from SREST (left panel) and $SREST_{shift}$ (right panel) simulations. The curves of $\chi^2(t)$, using $\bar{P}^{ref}_{initial}$ and $\bar{P}^{ref}_{final}$ as the reference states, are plotted in black and red, respectively. We stop updating the PEDFs after 2 ns in each simulation. The $\chi^2(t)$ values are obtained from three concatenated independent runs and averaged over all temperature levels.

3.2. Serial Replica Exchange with Solute Tempering. To illustrate how the SREM will allow the study of large scale biomolecular systems, we have combined the strategy behind the SREM with another recent method, replica exchange with solute tempering (REST),[9] which was developed to reduce the number of replicas required in the REM for the simulation of proteins solvated in explicit solvent. This combined method, called serial replica exchange with solute tempering (SREST), was also used on the alanine dipeptide system, but now solvated in 512 TIP4P[23] water molecules. This compares to the 256 TIP4P molecules used in the SREM simulations, while the force field and other simulation parameters were the same. With a target temperature for SREST of 300 K, only three replicas with temperatures of 300, 420, and 600 K were needed to generate reasonable acceptance ratios, ranging from 23 to 29%. The initial PEDFs were obtained from a 200 ps REST simulation. It should be noted that the effective energy used in the exchange of replicas in REST is $E_p + {}^1/_2E_{pw}$, not the standard E_{total} mentioned in section 2. After a 2 ns aging run in which the PEDFs were updated at an interval of 100 ps, we performed three independent SREST simulations for a further 9 ns. For comparison, we performed a REST simulation for 9 ns. As in the previous section, we checked the dependence on the initial PEDFs, by performing a SREST simulation with shifted initial PEDFs, which we refer to as $SREST_{shift}$.

Figure 6 shows the PEDFs initially used in the SREST simulation and the PEDFs resulting from the corresponding simulations. As already verified for the SREM, SREST generates accurate PEDFs even though the initial PEDF generated by the REM is noisy or shifted. Figure 7 clearly shows that the PEDFs converge to the reference PEDFs obtained with the 9 ns REST simulation, even when the initial PEDFs are shifted. We also observe that the convergence behavior, evidenced by both SREST and $SREST_{shift}$, is very similar to the SREM simulations. Once again, we have verified that the structural properties, as embodied by the RDF, obtained with SREST and $SREST_{shift}$ are comparable to the results of the REST simulation at the target temperature 300 K. Agreement for several heavy atom pairs can be seen in Figure 8, which is equivalent to Figure 3.

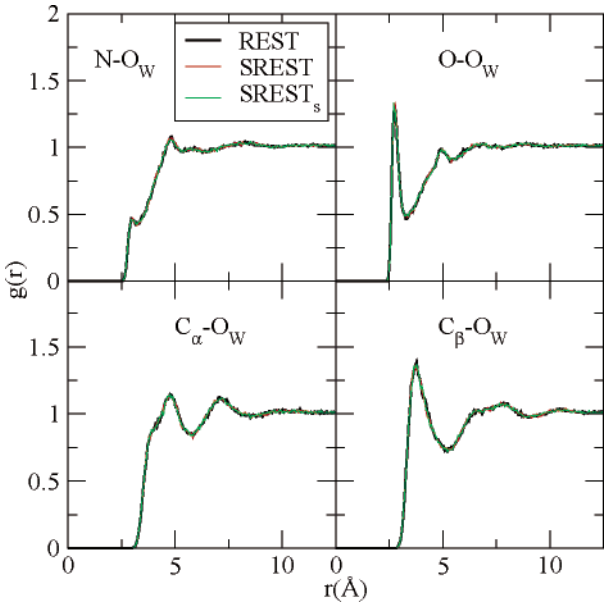

Figure 8. Comparison of the radial distribution functions (RDFs) between selected heavy atoms of alanine dipeptide and oxygens of water molecules at 300 K. RDFs from SREST, $SREST_{shift}$, and REST simulations are plotted in black, red, and green, respectively. The SREST and $SREST_{shift}$ results are averaged over three independent runs.

4. Discussion

We have introduced serial replica exchange (SREM), a serial version of the standard replica exchange method (REM),

designed to run *asynchronously* on a distributed computing cluster of inexpensive nodes with high failure rates. The method has been demonstrated to be equivalent to the standard replica exchange method in terms of accuracy and efficiency and greatly superior to the REM in terms of fault tolerance, especially for large systems, which require many replicas.

Initially, we determine an approximate potential energy distribution function (PEDF) from a short run of the REM, or possibly standard molecular dynamics (MD) without exchanges. This is followed by an aging phase and finally the production run. Neither of these require intensive message passing or synchronization between processors. Hence, the SREM is ideally suited to distributed computing environments, such as Folding@Home or the worldwide grid (www.grid.org). During the aging phase, the energy distribution at each temperature is periodically updated, and accurate PEDFs are gradually obtained. Since the amount of information collected is proportional to the number of processors used, the more processors that are used, the faster the PEDFs converge. To allow the replicas to change their temperature, a detailed balance criterion, which involves sampling new energies according to the PEDFs, is used for this phase as well as for the production run.

Application of the SREM to solvated alanine dipeptide shows that accurate potential energy distributions are gradually obtained, whether starting from rough estimates or even wrong estimates of the PEDFs. We have also demonstrated that correct thermodynamic and structural properties can be determined, as embodied in the Ramachandran populations and the radial distribution functions for heavy atom pairs at 300 K. Combination of the SREM with another recent method, replica exchange with solute tempering (REST),[9] significantly reduces the number of CPUs required. This method, called SREST, is expected to scale well to very large systems on distributed computing clusters. As for the SREM, we have verified that SREST obtains the correct structural properties at 300 K.

Other strategies exist for running replica exchange on distributed computers. One in particular is the multiplexed replica exchange method.[31] In contrast to that method, the SREM does not require synchronization between the different walkers (processors), leaves no CPUs idle, and thus has substantially lower overhead. Another method suitable for distributed environments is simulated tempering,[32] which is inherently a serial algorithm. The primary issue with that method is the determination of the free energy constants, which has historically required a somewhat involved iterative procedure. More recently, Mitsutake and Okamoto[12,13] have proposed to use a short initial replica exchange run, similar in spirit to the SREM described here, to determine these free energies. Simulated tempering has certain advantages, chiefly that it is an exact method even if the free energies are only obtained in some approximate manner. It may also be that the acceptance rates for simulated tempering are sometimes larger, as Mitsutake and Okamoto[12] have suggested. We have not explicitly compared the SREM and SREST with simulated tempering but take the view that for some applications, especially large scale biological systems, determining even approximate free energies can be substantially more difficult than the problem one is really trying to solve.

It is clear that the detailed balance criterion employed by the SREM and SREST is mathematically correct, provided that the *exact* potential energy distribution functions are employed. However, it is equally clear that unless the exact PEDFs are obtained, the method is only approximately correct. That does not imply that the SREM and SREST are any less rigorous than

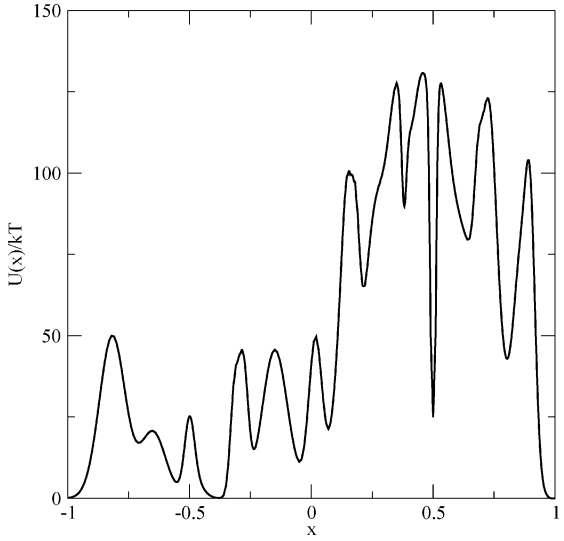

Figure 9. One-dimensional model system with periodic boundary conditions. The potential energy is in units of the thermal energy, kT.

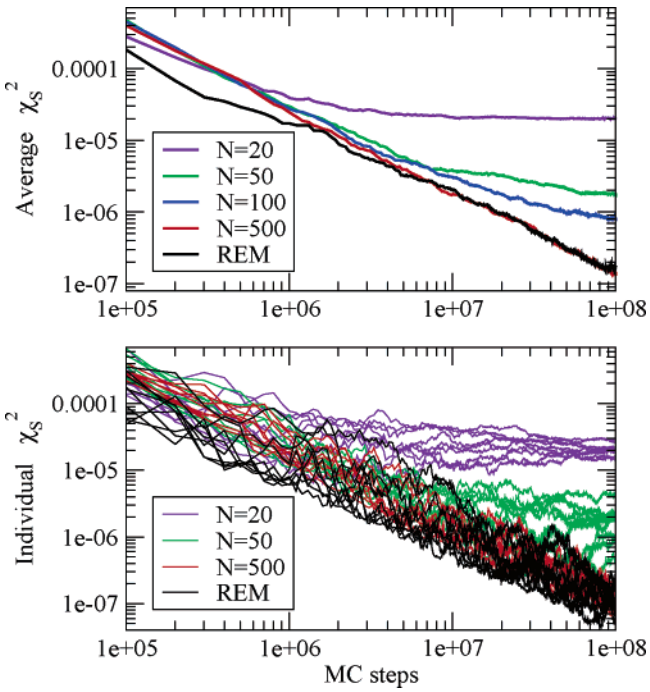

Figure 10. Integrated squared error of the spatial density, χs^2, shown for the REM and several versions of the SREM. The top panel demonstrates the behavior of the SREM versus the REM in terms of χs^2, averaged over 50 individual trajectories. The lower panel shows nine individual trajectories for the REM and SREM, with $N = 20$, 50, and 500.

other sampling methods, all of which are required to run "infinitely long" to produce correct statistics. Indeed, in the limit of an infinitely long run, the length of the aging phase will also tend to infinity, and hence, the aged PEDFs will tend to the exact PEDFs. As a note of caution, however, there is a difference between the exact PEDF and the exact *discrete* PEDF. Unless one is willing to consider infinitely many bins or a continuous density estimator, for example, Gaussian kernels, one can at most hope to obtain the exact discrete PEDF. For large scale systems, we believe that the exact PEDF is sufficiently smooth that it is not perceptibly different from the exact discrete PEDF. Whether any of these issues will be detected in practice depends on the application and on which quantity is being measured. We believe that other sources of error, such as the finite length

of a run, are much more significant. Indeed, when applied to our model system, we have found that there are no discernible differences in the thermodynamic and structural properties in comparison to the same properties computed with standard replica exchange. In the Appendix, we examine the issue of discretization in further detail.

We expect that once the SREM and SREST are implemented on a worldwide distributed computing platform, large scale biological systems solvated in explicit water can be simulated up to time scales of milliseconds, and possibly even seconds.

Acknowledgment. This work has been supported by grants to B.J.B. from the NSF (CHE16896) and the NIH (GM43340) and to R.A.F. from the NIH (GM52018). This work was partially supported by the National Computational Science Alliance, using the Xeon Linux Cluster under grant MCA95C007N, and by a SUR grant from the IBM Corporation for the purchase of an IBM LINUX CLUSTER, which was also used. We would like to thank one of our reviewers for suggesting that we generate the initial PEDFs with the Wang−Landau method.

Appendix

As argued in the Methods section, we chose to represent the PEDFs with histograms of fixed bin width because of their simplicity. However, in the discussion above, we acknowledged that for a finite number of bins there is always some difference between the exact PEDF and the exact *discrete* PEDF. In this appendix, we examine the effect of this distinction.

To elucidate the sometimes minute differences between REM and SREM simulations, we have chosen to design a simple one-dimensional system for which we can compute the exact *spatial* density profile. This system is illustrated in Figure 9. For both the REM and SREM, we have used a set of three replicas at kT, 17.3 kT, and 300 kT. Given the simplicity of the system, we have used Monte Carlo sampling within each ensemble, with a maximum step size of 0.1. Finally, as in our study of the peptide system, we employ a measure similar to the chi-square measure, the integrated squared error of the spatial density,

$$\chi_S^2(t) = \sum_{j=1}^{K} (\rho_j(t) - \rho_j^{\text{exact}})^2 \qquad (4)$$

to demonstrate the sampling efficiency of the REM and SREM. Here, $K = 1000$ is the number of bins used to calculate $\chi_S^2(t)$ with respect to the spatial coordinate, which is independent of the number of bins in the PEDF histogram, N. The normalized densities are calculated only for the replica at kT, and there is no averaging over the three ensembles. On the other hand, we have calculated $\chi_S^2(t)$ for several individual trajectories, each starting from independent initial conditions.

In Figure 10, we demonstrate the effect of the number of bins in the SREM histogram or, equivalently, the bin width. Each PEDF histogram spans the range −kT to 131 kT and is divided into N bins of equal width, where N is either 20, 50, 100, or 500. We observe that as N increases, the behavior of the SREM simulation approaches that of the REM simulation. The top panel demonstrates that, with $N = 500$ and averaged over 50 individual trajectories, there is no discernible difference in the $\chi_S^2(t)$ values for the REM and SREM. This is confirmed in the lower panel, which shows nine individual trajectories for each of four simulations, the REM and SREM with three different values of N.

The PEDFs in this one-dimensional system are decidedly non-Gaussian and thus expected to be less amenable to simple histogram estimation than Gaussian PEDFs. Given the apparent success, we conclude that simple fixed bin width histograms are adequate as general purpose nonparametric estimators of PEDFs for use in the SREM and SREST.

References and Notes

(1) Shirts, M.; Pande, V. *Science* **2000**, *290*, 1903−1904.
(2) Swendsen, R. H.; Wang, J. S. *Phys. Rev. Lett.* **1986**, *57*, 2607−2609.
(3) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604−1608.
(4) Tesi, M. C.; van Rensburg, E. J. J.; Orlandini, E.; Whittington, S. G. *J. Stat. Phys.* **1996**, *82*, 155−181.
(5) Okamoto, Y. *J. Mol. Graphics Modell.* **2004**, *22*, 425−439.
(6) Paschek, D.; Garcia, A. E. *Phys. Rev. Lett.* **2004**, *93*, 238105.
(7) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058−9067.
(8) Zhou, R.; Berne, B. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777−12782.
(9) Liu, P.; Kim, B.; Friesner, R.; Berne, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749−13754.
(10) Rhee, Y.; Sorin, E.; Jayachandran, G.; Lindahl, E.; Pande, V. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6456−6461.
(11) Voter, A. *Phys. Rev. B* **1998**, *57*, 13985−13988.
(12) Mitsutake, A.; Okamoto, Y. *Chem. Phys. Lett.* **2000**, *332*, 131.
(13) Mitsutake, A.; Okamoto, Y. *J. Chem. Phys.* **2004**, *121*, 2491−2504.
(14) Wang, F.; Landau, D. *Phys. Rev. Lett.* **2001**, *86*, 2050−2053.
(15) Wang, F.; Landau, D. *Phys. Rev. E* **2001**, *64*, 056101.
(16) Wang, F.; Landau, D. *Comput. Phys. Commun.* **2002**, *147*, 674−677.
(17) Poulain, P.; Calvoa, F.; Antonie, R.; Broyer, M.; Dugourd, P. *Phys. Rev.* **2006**, *E73*, 056704.
(18) Zhou, C.; Schulthess, T.; Torbrugge, S.; Landau, D. *Phys. Rev. Lett.* **2006**, *96*, 120201.
(19) Izenman, A. *J. Am. Stat. Assoc.* **1991**, *86*, 205−224.
(20) Cacoullos, T. *Ann. Inst. Stat. Math.* **1966**, *18*, 178−189.
(21) Cencov, N. *Sov. Math.* **1962**, *3*, 1559−1562.
(22) Scott, D. *Biometrika* **1979**, *66*, 605−610.
(23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(24) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(25) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem.* **2001**, *B105*, 6474−6487.
(26) Hockney, R. W.; Eastwood, J. W. *Computer Simulations Using Particles*; IOP: Bristol, U.K., 1988.
(27) Deserno, M.; Holm, C. *J. Chem. Phys.* **1998**, *109*, 7678.
(28) Zhou, R.; Harder, E.; Xu, H.; Berne, B. J. *J. Chem. Phys.* **2001**, *115*, 2348.
(29) Andersen, H. C. *J. Comput. Phys.* **1983**, *52*, 24−34.
(30) Martyna, G.; Klein, M.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635.
(31) Rhee, Y.; Pande, V. *Biophys. J.* **2003**, *84*, 775−786.
(32) Marinari, E.; Parisi, G. *Europhys. Lett.* **1992**, *19*, 451−458.