JCTC Journal of Chemical Theory and Computation

Exchange Often and Properly in Replica Exchange Molecular Dynamics

Daniel J. Sindhikara,[†] Daniel J. Emerson,[‡] and Adrian E. Roitberg*[,‡]

Department of Physics, School of Science, Nagoya University, Nagoya, Aichi 464-8602, Japan, Quantum Theory Project and Department of Chemistry, University of Florida, Gainesville, Florida 32611

Received May 29, 2010

Abstract: Previous work by us showed that in replica exchange molecular dynamics, exchanges should be attempted extremely often, providing gains in efficiency and no undesired effects. Since that time some questions have been raised about the extendability of these claims to the general case. In this work, we answer this question in two ways. First, we perform a study measuring the effect of exchange attempt frequency in explicit solvent simulations including thousands of atoms. This shows, consistent with the previous assertion, that high exchange attempt frequency allows an optimal rate of exploration of configurational space. Second, we present an explanation of many theoretical and technical pitfalls when implementing replica exchange that cause "improper" exchanges resulting in erroneous data, exacerbated by high exchange attempt frequency.

1. Introduction

For molecular dynamics (MD) and Monte Carlo (MC) simulations alike, the need for sufficient sampling is fundamental. Since researchers are limited to the computational power available, there is a huge need for sampling efficiency in MD and MC algorithms. Two main factors limit sampling efficiency: phase space diffusion and kinetic trapping. Both factors are system dependent, but kinetic trapping can be avoided by utilizing an enhanced sampling technique to increase barrier-hopping. Such techniques may use a coupling to an alternate ensemble where the barrier heights are reduced. This can be done in either serial or parallel manner. Serial methods such as simulated tempering utilize estimated statistical weight factors to determine probabilities of switching between ensembles.[1] Conversely, parallel methods, e.g., parallel tempering (PT) also known as the replica exchange method, switch between ensembles using exactly known probabilities by exchanging two replicas.[2,3] These two types of methods (including all variants thereof) utilize an MC move to change ensembles after a period of simulating the system within an ensemble with either MD or MC. It is the choice of the simulator how long one must simulate within an ensemble before attempting the MC switch to another.

To date, two popular opinions exist about the exchange attempt frequency (EAF) between ensembles. One is that exchanges should be performed often, but no more often than the potential energy autocorrelation time.[4,5] Periole et al.[4] studied explicit solvent REMD simulations with EAFs of 10, 2, 0.5, and 0.2 ps^{-1} (the MD time step was not reported in this work). They found that the 10 ps^{-1} simulation had correlated exchanges and thus only reported convergence for the three smaller EAF simulations.[4] Of these three, the sampling efficiency grew with EAF. Abraham et al. similarly argued against using a high EAF due to correlated exchanges, and also showed that "mixing" increased with EAF.[5] Yang et al. argued that the sampling efficiency gained by the increased traversing of energy space using high EAF has little effect on conformational sampling and is not worth the computational cost (they used a Perl wrapper outside an MD engine to perform exchanges and hence there is a substantial computational overhead to starting and stopping the runs for the invividual replicas) for their simulations of alanine-dipeptide and Ala-Pro.[6] This argument seems to dispute the efficiency of REMD in general.

The other argument is that EAFs should be maximized to attempt exchanges every few steps. Our previous work made this argument based on an array of simulations of peptides

* Corresponding author e-mail: roitberg@qtp.ufl.edu.
[†] Department of Physics, Nagoya University.
[‡] Quantum Theory Project and Department of Chemistry, University of Florida.

10.1021/ct100281c
Published on Web 07/28/2010

in implicit solvent and a reduced analytical model.[7] Other works have achieved identical conclusions.[8–10] So what is the cause of the discrepancy?

Here, we present further and complementary evidence to our previous study, and show, unequivocally, that EAF should be maximized in all simulations as long as the exchanges are done correctly. In the context of REMD, we will concisely explain the conditions in which an exchange is "proper" and argue that the contradictory results presented above can be directly assigned to wrongly performed exchanges.

2. Theory

In PT, also known as temperature replica exchange molecular dynamics (T-REMD), several copies (replicas) of a system are simulated simultaneously at a ladder of temperatures. The temperatures typically range from about the one most relevant to the system to a temperature high enough to easily make pass energetic barriers. Periodically, a proposal is made to exchange conformations between replicas adjacent in temperature. The probability of exchange between a replica in conformation, *i,* and temperature, *n,* with a replica in conformation, *j*, at replica, *m*, is based on the difference in energies of the two systems as well as the difference in temperatures as shown in eq 1. The formal derivation uses the a Boltzmann weighted population as the limiting distribution after convergence, as well as imposing detailed balance between the ensembles before and after an exchange.[3] Note that in this formulation, there is no mention or use for different EAFs, and should work regardless, if implemented correctly. As usual with MC algorithms, once a move is accepted, the question of the new structure belonging or not to the ensemble has a trivial answer: of course it is part of the ensemble, the positive answer to the MC question guarantees it.

$$P_{\mathrm{acc}} = \exp(-(E_{\mathrm{j}} - E_{\mathrm{i}})(\beta_{\mathrm{n}} - \beta_{\mathrm{m}})) \qquad (1)$$

Here, E represents the total energy and $\beta = 1/k_{\mathrm{B}}T$ is the inverse temperature. The E can be replaced by the potential energy V in order to increase the acceptance rate by rescaling the velocities by a factor of $\sqrt{T_{\mathrm{new}}/T_{\mathrm{old}}}$ to match the new temperature in the case of an accepted exchange.[3] This step is crucial: it analytically cancels out the kinetic energy terms and makes the exchange in terms of potential energy alone correct. Note that some implementations use velocity reassignment rather than rescaling.[11,12] Velocity reassignment, however, scrambles all pre-exchange momentum and may reduce the rate of barrier crossing. Nadler et al. showed an increased acceptance rate using amplified rescaling factors.[13] After the simulation is complete, time series data at a temperature (or temperatures) of interest are collected, or a reweighting technique such as WHAM[14,15] or MBAR[16] is used to collect data from all temperatures.

There are many factors in REMD that have previously been scrutinized. For example, the ideal distribution temperatures at which replicas are held. Most commonly, the temperatures are distributed exponentially.[17] This distribution provides a uniform acceptance ratio in the limit of constant heat capacity. Trebst et al. decscribed a technique to further optimize this distribution of temperatures.[18] The lowest temperature is usually near the temperature of highest interest. Zheng et al. showed that the maximum temperature should be just above the temperature where the activation free energy vanishes and any higher replicas than that would decrease the efficiency due to non-Arrhenius behavior.[19] Nymeyer recently proved this result analyitically.[20]

Rosta et al. showed that use of thermostats that do not sample the proper canonical space (such as the Berendsen thermostat) leads to incorrect results.[21] We also showed recently that poor implementation of REMD with respect to pseudorandom number seeds can cause negative results.[22] Recent work in systems with sharp phase transitions indicate that one should use different EAF for different replicas, according to the canonical correlation time at that temperature.[23]

Our previous work showed that larger EAFs are better than smaller ones.[7] Though the result was only for implicit solvent systems, we believe the result is general and fundamental. Simply put, attempting replica exchanges more often allows for more (computationally inexpensive) MC moves; in MC simulations, it is obviously better to have performed more MC moves than less. Thus, as long as the move is thermodynamically correct and computationally inexpensive, it is worth doing.

In our previous work on EAF, we also highlighted the need for computationally efficient exchange attempt calculations.[7] Efficient exchanging can be achieved by performing exchanges within the MD engine rather than continuous calls of the MD engine from an external wrapper. The computational costliness of an exchange attempt should be the chief concern for simulators as it is the only potential limiting factor for EAF.

3. Methods

For this work, two polyalanine peptides (Ace-A_n-*N*me, for $n = 3,7$) were simulated using REMD. Each peptide was capped by an acetyl group (Ace) at the *N*-terminus and an *N*-methylamine (*N*me) group at the C-terminus. All simulations were performed using AMBER 10 molecular simulation package with the AMBER ff99SB parameter set.[24,25] The SHAKE algorithm in which all bonds involving hydrogen were constrained giving a 2 fs time step.[26] Calculations were performed in a canonical ensemble with the Langevin thermostat (collision frequency of 1 ps^{-1}). Each molecule was solvated in an octahedral box of TIP3P water molecules. A total of 1301 water molecules were added to $(Ala)_7$ (total system size 3985 atoms) and 761 water molecules were added to $(Ala)_3$ (total size of 2325 atoms). Each system was run with REMD with temperatures exponentially distributed with a theoretical acceptance probability of 20%. For $(Ala)_3$, Twenty-four temperatures were used ranging from 293.2 to 495.5 K; for $(Ala)_7$, and twenty-eight temperatures were used ranging from 294.9 to 467.9 K. For each system, 10 ns-long test runs of varying EAFs were compared to a reference 100 ns-long REMD run. The test runs utilized an array of EAFs (Ala_3: 0.01, 0.02, 0.2, 2, 25, 62.5, and 250 ps^{-1}; Ala_7: 0.001, 0.01, 0.1, 2, 20, and 250 ps^{-1}). Since the time step is 2 fs,

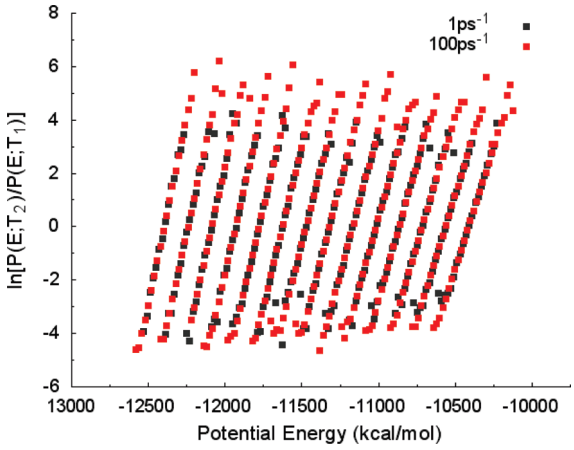

Figure 1. Logarithm of ratio of population between adjacent temperatures for simulations with different EAFs. Each of the 14 lines visible represents the overlap between two adjacent temperatures (of the 28 total, only every other neighboring pair is shown). Since the points from each EAF lie on the same line, the overlap region between the two temperature distributions must be identical.

the fastest exchange for both Ala_3 and Ala_7 was every 2 MD steps. The reference run utilized a "moderate" EAF of 0.1 ps^{-1} (every 5000 MD steps). Each test run was compared to the reference run to judge its convergence to the "correct" value.

4. Results and Discussion

Much of the analysis following is similar to our previous work[7] on this new data set. Here we will highlight the most important findings.

4.1. Thermal Equilibration at High EAF. To check if EAF affects thermal equilibration, energetic properties were compared between a moderate EAF of 1 ps^{-1} and a high EAF of 100 ps^{-1} for the $(Ala)_7$ simulations. The overlap in the potential energy distributions was calculated to check the similarity. For all 28 temperatures, the lowest overlap between distributions at the two EAFs is 0.996 indicating that the distributions are nearly identical. Figure 1 shows the logarithm of the ratio of energy distributions between neighboring temperatures for the $(Ala)_7$ simulations at both 1 ps^{-1} and 100 ps^{-1} (only every other pair of temperatures is shown). Since the points for either simulation lie along the same line, the two simulations must be sampling the same Boltzmann distribution.

4.2. Effect of EAF on Sampling Efficiency. Figure 2 shows the average rmsd of the dihedral population for both $(Ala)_3$ and $(Ala)_7$. This rmsd represents the difference between the test and the reference simulations, computed by constructing 36 × 36 histograms for each residue into 10°x10° bins. These histograms were normalized into populations and the rmsd between the histograms from test and reference runs was calculated and later averaged over all residues in each peptide we studied.

Consistent with the results from the implicit solvent study,[7] the trend is that for the larger system, the error asymptotically decreases with EAF. While there is no clear effect of EAF on efficiency for the Ala_3 simulations, mostly due to the fact

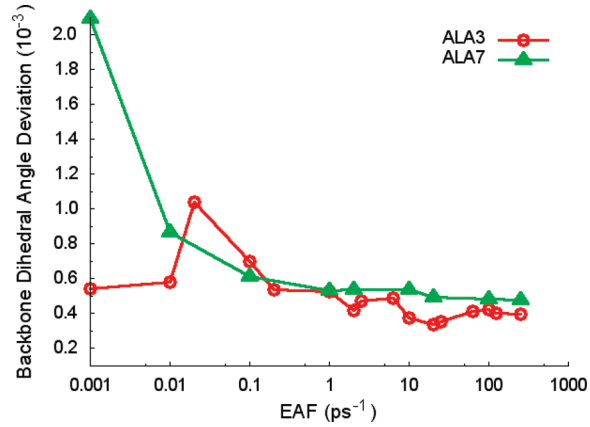

Figure 2. Backbone dihedral angle rmsd between 10 ns-long test runs and 100 ns long reference run.

that it is a very simple system to converge, the error in the Ala_7 simulations clearly decays with EAF.

Figure 3a,b shows linear-log plots of the deviation vs time for the $(Ala)_3$ and $(Ala)_7$ simulations respectively. While there is some line crossing, the errors in the simulations with the largest EAFs tend to reduce the fastest. The line crossing also indicates that preliminary runs to optimize EAF may not suggest the correct asymptotic behavior. For example, in the $(Ala)_7$ simulation, the simulation with the smallest EAF (0.001 ps^{-1}) seems to be on par with the rest after only 0.5 ns of simulation; eventually, however, it becomes clear that it is the least efficient of all the simulations.

Another indicator of sampling efficiency in REMD is the total number of times a replica visits both the lowest and the highest temperature during the 10 ns simulations.[18,19,27–30] This "round trip number" for a fixed simulation time represents the speed of diffusion of a replica in temperature space and is thus indicative of the freedom of the simulation to sample. Figure 4 shows the round trip number for both the $(Ala)_3$ and $(Ala)_7$ simulations. The trend is clearly that the higher EAF simulations move faster through temperature space. It may be of interest to note that the dependence of number of round trips on EAF is logarithmic rather than linear as would be expected with freely diffusing systems. This may be caused partially by energy autocorrelation times as well as other factors that cause nonlinearity even in slow EAF simulations as clearly seen in our previous study.[7]

4.3. 1-Dimensional REMC System. Finally, to present a simple system that has the features of a larger, more complex set, we constructed a 1D test system using replica exchange Monte Carlo (REMC). In this system, it is easy to test the fundamental effect of EAF on replica exchange simulations. The energy is a simple 1D double well as in eq 2.

$$U(x) = x^2 - |x| \tag{2}$$

Two temperatures were used (300 K and 350 K) using metropolis Monte Carlo with a MC step evenly distributed within [−0.15,+0.15]. A total of 1 000 000 MC steps were taken at each temperature and replica exchanges were attempted at various EAFs ranging from 0.0001 to 1 per MC step (one exchange attempted every 10 000 MC steps, or one exchange attempted at every single MC step). The error

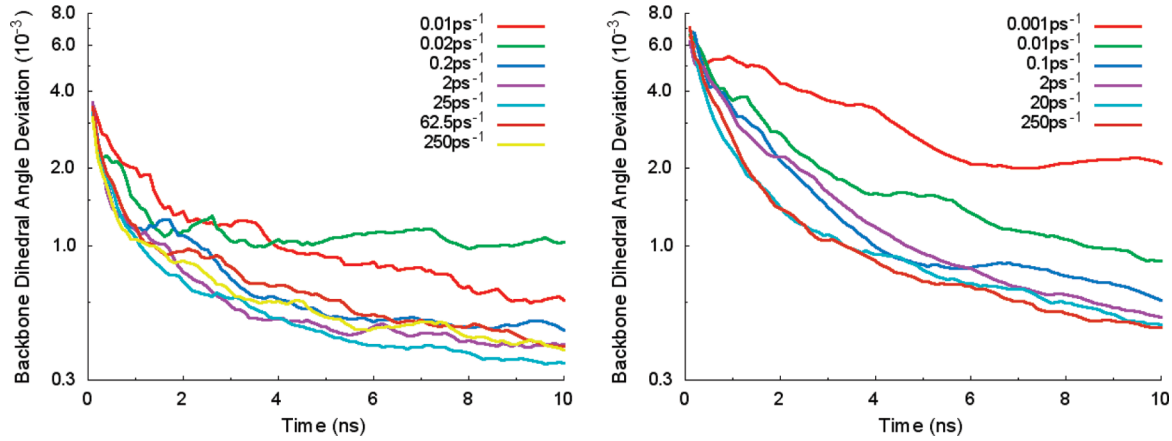

Figure 3. (a,b) Linear-log plot of deviation vs time for various EAFs. (a) $(Ala)_3$. (b) $(Ala)_7$.

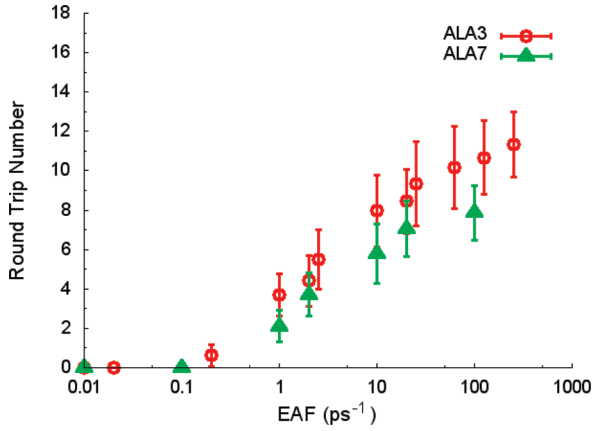

Figure 4. Average number of round trips per replica between the lowest and highest temperature for both the $(Ala)_3$ and $(Ala)_7$ simulations within 10 ns. The error is calculated as the standard deviation between individual replica counts.

Table 1. Population Error Vs Correct Distribution[a]

EAF (per MC step)	0.0001	0.001	0.01	0.1	1
population error (10^{-3})	677	590	255	39	8

[a] Error is RMSD between bin populations. The bin width is 0.1.

versus the correct, analytical solution is shown in Table 1. The trend is the same as with the molecular simulations in this and previous studies, clearly supporting the idea that larger EAFs are better than small ones.

None of these results should be surprising considering the fact that replica exchanges are not fundamentally different than any other MC move. Just as in MC, the more moves, the better, such is the case of replica exchange attempts.

4.4. Recommendations. Considering the many critiques of REMD mentioned in Section 2, along with the complementary evidence above, we emphasize that the most proper way to perform replica exchanges is as follows:

- Use a canonical thermostat (such as Langevin, Andersen, or Nose-Hoover).[21]
- Utilize a reasonable temperature distribution (as described in the Theory section).[18–20]
- Rescale velocities after the exchange if using potential energy in the MC exchange criterion.[3]
- Utilize an exchange algorithm within the MD engine.
- Attempts exchanges as often as possible.*

We note that the simulator should always question the computational efficiency of an exchange in their particular algorithm and hardware. Algorithms with computationally expensive exchanges (such as those performed in an external wrapper) may not afford frequent exchanges and should be upgraded. The computationally efficient exchanges in AMBER10 (and above) allow for high EAFs and thus we found that exchanging every 10−100 steps is a good balance between sampling and computational efficiency. Furthermore, if any of the criteria above are violated, the use of high EAF could exacerbate the respective detrimental effects (as described above and within the cited references).

5. Conclusions

Despite strong evidence in prior studies that REMD using high EAFs are optimal for fastest sampling, questions remained about the consistency of this result for explicit solvent systems. Here, a similar study to the original, implicit solvent study[7] shows that the trend also holds for explicit solvent systems. This new work should resolve many remaining questions about the effect of EAF on REMD simulations. Combined with the extensive analysis of the study, we strongly conclude that for any system where REMD exchanges are performed properly, as described above, maximum sampling efficiency can be obtained by attempting to exchange as often as possible. We also detailed many of the ways REMD simulations were performed improperly.

Acknowledgment. The authors acknowledge the University of Florida High-Performance Computing Center and NSF Large Allocations Resource Committee through Grant Nos. TG-MCA05S010 and UT-NTNL0002 for providing computational resources and support that have contributed to the research results reported within this work. Work at University of Florida was funded by the National Science Foundation Grant No. CHE-0822-935. The authors wish to thank Kevin Hauser for his initial work on these systems.

References

(1) Marinari, E.; Parisi, G. Simulated tempering—A new Monte-Carlo scheme. *Europhys. Lett.* **1992**, *19* (6), 451–458.

(2) Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **1997**, *281* (1−3), 140–150.

(3) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314* (1−2), 141–151.

(4) Periole, X.; Mark, A. E. Convergence and sampling efficiency in replica exchange simulations of peptide folding in explicit solvent. *J. Chem. Phys.* **2007**, *126* (1), 014903.

(5) Abraham, M. J.; Gready, J. E. Ensuring mixing efficiency of replica-exchange molecular dynamics simulations. *J. Chem. Theory. Comput.* **2008**, *4* (7), 1119–1128.

(6) Yang, L. J.; Shao, Q.; Gao, Y. Q. Comparison between integrated and parallel tempering methods in enhanced sampling simulations. *J. Chem. Phys.* **2009**, *130* (12), 124111.

(7) Sindhikara, D.; Meng, Y. L.; Roitberg, A. E. Exchange frequency in replica exchange molecular dynamics. *J. Chem. Phys.* **2008**, *128* (2), 024103.

(8) Opps, S. B.; Schofield, J. Extended state-space Monte Carlo methods. *Phys. Rev. E* **2001**, *6305* (5), 056701.

(9) Zhang, C.; Ma, J. P. Comparison of sampling efficiency between simulated tempering and replica exchange. *J. Chem. Phys.* **2008**, *129* (13), 134112.

(10) Rosta, E.; Hummer, G. Error and efficiency of replica exchange molecular dynamics simulations. *J. Chem. Phys.* **2009**, *131* (16), 165102.

(11) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. *J. Chem. Theory Comput.* **2007**, *3* (1), 26–41.

(12) Yang, L. J.; Grubb, M. P.; Gao, Y. Q. Application of the accelerated molecular dynamics simulations to the folding of a small protein. *J. Chem. Phys.* **2007**, *126* (12), 125102.

(13) Nadler, W.; Hansmann, U. H. E. Optimizing replica exchange moves for molecular dynamics. *Phys. Rev. E* **2007**, *76* (5), 057102.

(14) Ferrenberg, A. M.; Swendsen, R. H. Optimized Monte-Carlo Data-Analysis. *Phys. Rev. Lett.* **1989**, *63* (12), 1195–1198.

(15) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules 0.1. The method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021.

(16) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129* (12), 124105.

(17) Kofke, D. A. On the acceptance probability of replica-exchange Monte Carlo trials. *J. Chem. Phys.* **2002**, *117* (15), 69116914.

(18) Trebst, S.; Troyer, M.; Hansmann, U. H. E. Optimized parallel tempering simulations of proteins. *J. Chem. Phys.* **2006**, *124* (17), 174903.

(19) Zheng, W. H.; Andrec, M.; Gallicchio, E.; Levy, R. M. Simulating replica exchange simulations of protein folding with a kinetic network model. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (39), 15340–15345.

(20) Nymeyer, H. How efficient is replica exchange molecular dynamics? An analytic approach. *J. Chem. Theory. Comput.* **2008**, *4* (4), 626–636.

(21) Rosta, E.; Buchete, N. V.; Hummer, G. Thermostat artifacts in replica exchange molecular dynamics simulations. *J. Chem. Theory Comput.* **2009**, *5* (5), 1393–1399.

(22) Sindhikara, D. J.; Kim, S.; Voter, A. F.; Roitberg, A. E. Bad seeds sprout perilous dynamics: Stochastic thermostat induced trajectory synchronization in biomolecules. *J. Chem. Theory Comput.* **2009**, *5* (6), 1624–1631.

(23) Bittner, E.; Nussbaumer, A.; Janke, W. Make life simple: Unleash the full power of the parallel tempering algorithm. *Phys. Rev. Lett.* **2008**, *101* (13), 130603.

(24) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26* (16), 1668–1688.

(25) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65* (3), 712–725.

(26) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *N*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(27) Fenwick, M. K.; Escobedo, F. A. Expanded ensemble and replica exchange methods for simulation of protein-like systems. *J. Chem. Phys.* **2003**, *119* (22), 11998–12010.

(28) Rathore, N.; Chopra, M.; de Pablo, J. J. Optimal allocation of replicas in parallel tempering simulations. *J. Chem. Phys.* **2005**, *122* (2), 024111.

(29) Earl, D. J.; Deem, M. W. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.* **2005**, *7* (23), 3910–3916.

(30) Katzgraber, H. G.; Trebst, S.; Huse, D. A.; Troyer, M. Feedback-optimized parallel tempering Monte Carlo. *J. Stat. Mech.* **2006**, *2006*, P03018.

CT100281C