JCTC Journal of Chemical Theory and Computation

Ensuring Mixing Efficiency of Replica-Exchange Molecular Dynamics Simulations

Mark J. Abraham and Jill E. Gready*

Computational Proteomics Group, John Curtin School of Medical Research, Australian National University, P.O. Box 334, Canberra, ACT, 2601, Australia

Received January 16, 2008

Abstract: We address the question of constructing a protocol for replica-exchange molecular dynamics (REMD) simulations that make efficient use of the replica space, assess whether published applications are achieving such "mixing" efficiency, and provide a how-to guide to assist users to plan efficient REMD simulations. To address our first question, we introduce and discuss three metrics for assessing the number of replica-exchange attempts required to justify the use of a replica scheme and define a "transit number" as the lower bound for the length of an efficient simulation. Our literature survey of applications of REMD simulations of peptides in explicit solvent indicated that authors are not routinely reporting sufficient details of their simulation protocols to allow readers to make independent assessments of the impact of the method on their results, particularly whether mixing efficiency has been achieved. Necessary details include the expected or observed replica-exchange probability, together with the total number of exchange attempts, the exchange period, and estimates of the autocorrelation time of the potential energy. Our analysis of cases where the necessary information was reported suggests that in many of these simulations there are insufficient exchanges attempted or an insufficiently long period between them to provide confidence that the simulation length justifies the size of the replica scheme. We suggest guidelines for designing REMD simulation protocols to ensure mixing efficiency. Two key recommendations are that the exchange period should in general be larger than 1 ps and the number of exchange attempts should be chosen to significantly exceed the transit number for the replica scheme.

1. Introduction

Interest from chemists in replica-exchange techniques for molecular dynamics simulations of biomolecules has grown rapidly since the seminal paper in 1999,[1] although the method was in use in physics before then.[2] One attractive feature of the method is its generality. Implementations of the algorithm have been developed that exchange temperatures,[1] modified nonbonded potentials,[3] the solvent contribution to the potential energy surface,[4] graininess of solute structure,[5–7] other Hamiltonian features,[8–10] λ values in free-energy calculations,[11] and bias potentials as used in meta-dynamics simulations.[12,13] Even pseudoexchange with a previously generated ensemble of structures[5,14] has been explored.

Techniques have been developed to optimize the distribution of the temperatures of the replicas in the conventional algorithm[15] to reduce bottlenecks in temperature exchange, and to produce a set of replica temperatures that span a given temperature range to yield a given average exchange probability.[16] Other work to improve exchange efficiency through the use of hybrid implicit−explicit solvent Hamiltonians during the exchange attempts[17] has been reported. There has been much discussion of the relative efficiency of canonical sampling of biomolecular systems using replica-exchange and fixed-temperature equilibrium sampling.[18–23] However, there has been no discussion of the dependence of the efficiency of the chosen replica scheme on the number of exchanges attempted over a simulation, either for conformational searching or for canonical sampling.

* Corresponding author e-mail: Jill.Gready@anu.edu.au.

10.1021/ct800016r CCC: $40.75
Published on Web 05/31/2008

There has been some discussion of the choice of period between exchange attempts (the *exchange period*). Zhang et al.[19] justify their choice of period (1 ps) by observing that, for their system, an alternative choice of 0.02 ps led to significant reduction in the traversal of temperature space by replicas, and that an alternative choice of 5 ps led to results qualitatively similar to those for 1 ps. They attributed the reduction in traversal to the reduced ability of replicas to stabilize in their new ensemble before being considered for exchange back to their former ensemble and recommended the use of an exchange period larger than 0.02 ps.

An alternative way of expressing the thought underlying this recommendation is that the probabilities for successive exchange events were time-correlated because the underlying potential energies and the configurations generating them were also time-correlated. Successive integration steps in molecular dynamics (MD) simulations of biomolecular systems occur in time steps of around 1−4 fs. These lead only to small changes in position and velocity coordinates and, because of the nature of the mathematical functions in the force fields, correspondingly small changes in potential energy. It follows that the correlation of an observable such as the potential energy with former values of itself (the *autocorrelation*) will be appreciable and will decay with increasing lengths of time between observations.

When measuring the value of an observable f such as potential energy in an MD simulation, we would like to make a set of Q independent (i.e., uncorrelated) measurements f_i. From these, we can form the estimator of the mean, $\bar{f} = 1/Q \sum_{i=1}^{Q} f_i$. The variance of that estimator, $\bar{f}$ (the square of the standard error), is then given by $\sigma^2(\bar{f}) = \sigma^2(f)/Q$ for $\sigma^2(f)$, the variance of f. It can be shown[24,25] that the variance of $\bar{f}$ for Q correlated samples collected at intervals Δt is given by

$$\sigma^2(\bar{f}) = \frac{2\tau_{\mathrm{int},f}\sigma^2(f)}{Q\Delta t} \tag{1}$$

for $\tau_{\mathrm{int},f}$, the *integrated autocorrelation time*. $\tau_{\mathrm{int},f}$ may be computed with

$$\tau_{\mathrm{int},f} = \Delta t\left(\frac{1}{2} + \sum_{t=1}^{\infty}\frac{C_f(t)}{C_f(0)}\right) \tag{2}$$

where $C_f(t)$ is the *autocorrelation function* of f. This latter function describes the extent to which successive measurements of f are correlated in time and is defined by

$$C_f(t) = \langle f(\tau)\, f(t+\tau)\rangle \tag{3}$$

averaging over all time origins τ. It is clear that the variance of the estimate of $\bar{f}$ of the correlated sample is a factor of $2\tau_{\mathrm{int},f}/\Delta t$ larger than it would be if the samples were independent (when $C_f(t)$ is a δ function). An alternative interpretation of eq 1 is that there are only $Q\Delta t/2\,\tau_{\mathrm{int},f}$ samples of f that are effectively statistically independent. It follows that the length of a molecular dynamics simulation for measuring f should be such that $Q\Delta t \gg \tau_{\mathrm{int},f}$ independent samples can be gathered. Conformational searching with replica-exchange molecular dynamics (REMD) seems most likely to be effective when successive exchange attempts make measurements of the potential energy U that are independent. Accordingly, the chosen exchange period should be at least as large as $\tau_{\mathrm{int},U}$.

There is further support for the above recommendation of Zhang et al.[19] from the findings of Periole and Mark,[22] who performed a comparative study of exchange periods of 0.1, 0.5, 2, and 5 ps in 10 ns REMD simulations over 20 replicas of a β-heptapeptide in explicit solvent. In this study, they observed a greater number of successful exchanges for 0.1 ps, which they attributed to an excessive correlation between successive exchange attempts. Further, they observed a higher frequency of "back exchange" following successful exchanges for both 0.1 and 0.5 ps, again suggesting correlation. These authors concluded that attempting replica exchanges with periods that are this short is not helpful in achieving the overall goal of a normal REMD simulation. They also tried a simulation where velocities were not scaled after successful exchanges and found that the correlation phenomena were not strongly affected by the use of velocity scaling. Further, they also presented results which indicated that the relative efficiency of various replica-exchange schemes for reaching the folding equilibrium across the generalized ensemble was greatest when the exchange period was large enough to prevent such correlation, but no larger.

Recent work by Sindhikara et al.[26] makes the claim that sampling efficiency in REMD simulations increases with a decreasing exchange-attempt period, even below the minimum period identified in refs 19 and 22. This claim is based on their observations of the variation of the dihedral angle root-mean-square deviation (rmsd) distribution of some model peptides with the exchange-attempt period, in each case compared with a reference REMD simulation. They use implicit-solvent simulations and chose as their comparison metric a 36 × 36 2D histogram of the φ and ψ dihedral angles. These were normalized, and the rmsd between the test and reference histograms was formed and then averaged for each residue. They did not discuss alternative metrics. It is possible for averaging over time and residue to mask significant variation in the distribution of dihedral angles. However, it is not possible to assess whether the observed variation in the values of the metric is significant, although this could be done by demonstrating how much variation occurs between replicates of the reference REMD simulation. It would be of interest to see if the above claims can be substantiated for explicit-solvent simulations or for other metrics.

Sindhikara et al. also demonstrated that the distribution of potential energies remains Boltzmann over the range of exchange-attempt periods they studied. They then argued that, because the average probability of an exchange acceptance[27,28] $\bar{P}_{\mathrm{acc}}$ may be expressed as

$$\bar{P}_{\mathrm{acc}}(\beta_1,\beta_2) = \int_1\int_2 P(U_1)\,P(U_2)\min(1, \mathrm{e}^{\Delta\beta\Delta U})\,\mathrm{d}U_2\,\mathrm{d}U_2 \tag{4}$$

and the probability distributions of potential energies $P(U)$ are independent of the exchange-attempt period, this implies that the value of this integral is also necessarily independent of the exchange-attempt period. This is not true for exchange-

attempt periods shorter than the autocorrelation time of the potential energy. Following a successful exchange, $P(U_1)$ and $P(U_2)$ are correlated even though they are both samples from correct Boltzmann distributions. That correlation will increase with decreasing exchange-attempt period. Accordingly, there must be a dependence of $\overline{P}_{acc}$ on the exchange-attempt period, which agrees with the findings of Periole and Mark above.[22]

In this work, we introduce the distinct but related concepts of *thermodynamic efficiency* and *mixing efficiency* of a REMD simulation. Such a simulation attains thermodynamic efficiency when it converges the observables of the ensembles of interest, and mixing efficiency when each replica has experienced the full range of replica conditions; that is, the replica-exchange system is "well mixed". We note that it is quite possible to achieve either thermodynamic or mixing efficiency without achieving the other. In a case where the free energy surface is flat with respect to the range of conformations accessible to replicas, the extra sampling ability of the higher-range replicas is wasted, and thermodynamic efficiency will be attained if the simulation is long enough. Even if mixing efficiency was attained in such a case by simulating long enough with replicas spaced closely enough, the simulation would overall still be wasteful of resources. In a case where the free energy surface is rugged with respect to the range of conformations accessible to replicas, it can be difficult to attain thermodynamic efficiency, even if, as above, mixing efficiency is attained.

Although the existence of thermodynamic efficiency of REMD is just as difficult to assess as ensemble convergence in normal MD simulations, recent work by Lyman and Zuckerman has shown some progress.[29] They proposed a method for assessing structural convergence in a knowledge- and parameter-free way and suggested that it could be applied to trajectories from replica-exchange simulations as they traverse temperature space. We agree with their expectation that, in an efficiently mixed replica-exchange simulation, the "decorrelation time" they define would be uniform across the replicas. They also expected that this decorrelation time will be shorter for replica-exchange simulations than for standard ones. In this work, we will consider the impact of the size of the replica-exchange scheme on the decorrelation time.

Thermodynamic and mixing efficiency both contribute to determining the total cost of a REMD simulation. An effective REMD simulation needs a replica scheme whose range of conditions is sufficient for a convergence of observables and needs to be long enough to achieve this convergence. A REMD simulation that is both effective and minimizes the total cost requires that the range of conditions in the replica scheme be sufficient but not excessive,[23] that it be no longer than required for observable convergence, and that it be long enough for good mixing to occur. Our objective in this work is to provide guidance on assessing the third of these requirements, that is, ensuring mixing efficiency. Accordingly, we will introduce and compare three metrics for assessing mixing efficiency, survey some REMD simulations reported in the literature, make some observations about replica-exchange periods and potential energy autocorrelation times, and provide some practical guidelines for future REMD simulations.

2. Methods

2.1. Measuring Mixing Efficiency of REMD Simulations. As a basis for proposing three metrics for measuring the efficiency of replica-exchange schemes, we start by briefly reviewing replica-exchange methods. These methods are examples of generalized ensemble schemes, in which multiple independent simulations are carried out using different ensembles. It is desired that each independent simulation samples states according to some weight function that corresponds to a thermodynamic ensemble of choice. Periodically, exchanges of configurations between pairs of ensembles are attempted. For the generalized ensemble to converge to its equilibrium state, the condition of detailed balance must be observed for such exchanges. This can be satisfied by the use of a generalized Metropolis criterion for exchanges where, for Boltzmann-weighted states, the probability of exchange between replicas i and $i + 1$ from ensembles with temperatures and potential energies T_i, T_{i+1}, U_i, and U_{i+1} respectively, is

$$P(\text{exchange}) = \min\left\{1, \exp\left[\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_{i+1}}\right)(U_i - U_{i+1})\right]\right\} \quad (5)$$

where k_B is Boltzmann's constant. The temperatures are usually chosen such as to keep the average exchange probability constant over replica space. Accordingly, and in the absence of known bottlenecks in the free energy surface,[15] the temperatures are distributed exponentially over a chemically relevant range.

2.1.1. Metric One. We start with the straightforward Metric One. It is normally true that one of the extremal ensembles of a replica-exchange scheme is the target ensemble from which the simulation data are to be collected (the *minimum*), and the other extreme is expected to have the greatest capacity to enable the generalized ensemble to sample the target ensemble in an efficient manner (the *maximum*; e.g., Observation 3 in ref 20). We define a *transit* to be the process of a replica visiting the maximum and subsequently the minimum. A given replica scheme will only result in an efficient computation when there are sufficient exchange events to ensure significant probability that replicas have undergone transit(s). Given an estimate of the average probability of exchange occurring across replica space, it is possible to calculate the number of exchange attempts that are required for a given probability of at least one transit occurring. We define the *transit number* of the replica scheme to be the number of replica-exchange attempts required for a 95% probability that at least one replica has visited the maximum before visiting the minimum.

This transit number may be interpreted as the minimum number of exchange attempts that is necessary in order to begin to justify the use of that particular replica scheme. Any simulation that attempts fewer exchanges will not be making effective use of the replica space, and thus computing power. We note, however, that the transit number merely establishes a lower bound and implies nothing for the convergence of

the ensemble, or the autocorrelation time of any observables. In practice, in order to explore the conformational space efficiently and thoroughly, enough exchange attempts will be required such that multiple replicas will have the opportunity to undergo many transits. That number of exchange attempts will be dependent on the nature of the generalized free energy surface. Depending on the nature of that surface and the number of replicas, it may not be necessary to achieve transits in order to achieve sufficient sampling of the target ensemble. However it *is* necessary to achieve transits in order to sample efficiently, because otherwise the high-temperature replicas are not useful, and the computational cost could be reduced by eliminating those replicas. One might think it "fairer" to define transits as those between a top and bottom subset of replicas, particularly in cases where data are collected from more than the lowest-temperature replica, or where the higher temperatures are known to be excessive. Because all replicas that reach an extremum have done so after entering the subset for that extremum, such a replica system will have an effective transit number identical with that of a smaller system with single-replica subsets at the extrema.

It is straightforward to model the time evolution of a replica scheme with M replicas with average exchange probability p. The "odd" pairs of replicas are the pairs (1, 2), (3, 4),..., ($2n+1$, $2n + 2$) for $0 \leq n \leq {}^1/_2(M - 2)$, and the "even" pairs are (2, 3), (4, 5),..., ($2n$, $2n + 1$) for $0 \leq n \leq {}^1/_2(M - 1)$, where the integers indicate the index of the ensemble within the ordered set. In line with the implementation of the REMD method in the molecular simulation package GROMACS,[30–32] we limit our study to a scheme where only the odd pairs may exchange at one attempt, and only the even pairs at the next, alternating indefinitely. There are alternatives here, such as permitting exchange attempts between all neighboring replicas, or all possible replicas. If an exchange attempt was made between replicas i and $i + 1$ and a further attempt was made at that time between $i + 1$ and $i + 2$, then the probabilities for all of these exchanges would depend on the energy of all three replicas. Exchange between i and $i + 2$ would be much more unlikely than between neighboring pairs, however. Considering more than neighboring pairs thus increases the complexity of the computation at each exchange attempt for very few extra successful exchanges. In a practical REMD simulation, this increase in complexity would increase the length of time spent by the parallel computing system testing for exchanges (from M sends and receives of single messages, to $2M$ or M^2), with a corresponding decrease in efficiency of the whole calculation.

So, when developing a model of a REMD simulation, we can assign each replica to an ensemble, and note that after each exchange attempt, on average, each replica will be in the same ensemble with probability $1 - p$ and at its exchange partner's ensemble with probability p. After further exchanges, the distribution of all of the replicas across the generalized ensemble is known probabilistically. In order to study the transit number as a function of M and p, it is not necessary to track the evolution of the probability distribution of each of the M replicas. It suffices to "mark" all of the probability density at the maximum after each exchange attempt and to track only the amount of marked and unmarked density present in each ensemble. Likewise, on reaching the minimum, any marking of probability density is removed, so that it is able to return to the maximum unmarked and transit again (however, it is unlikely that such density will affect the transit number because the simulation will likely be too short to permit this). It should be clear that the origin of any probability density is not of concern for measuring mixing efficiency. Neither is it of concern whether the pathway from the maximum to the minimum was direct or meandering, fast or slow; only whether it has been to the maximum before reaching the minimum matters. A running total of the amount of density from which marking was removed can be kept, and the transit number can then be calculated.

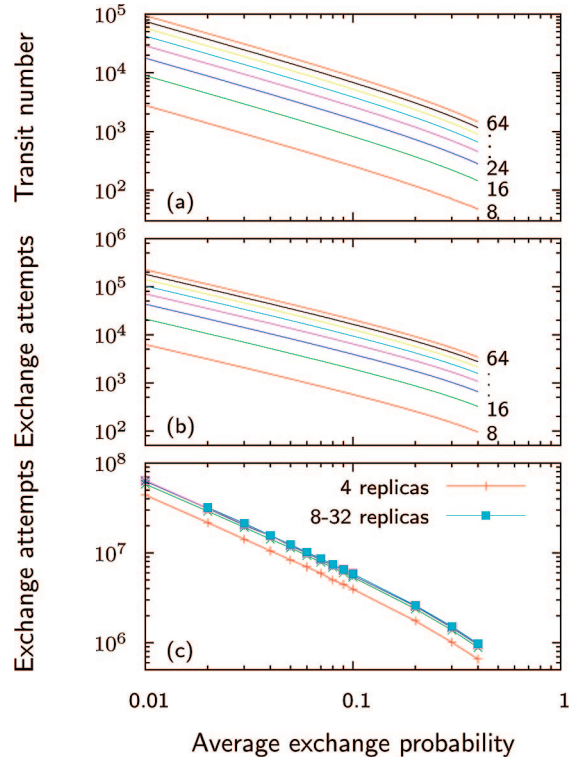

Figure 1. Illustrations of metrics (a) one, (b) two, and (c) three. These show either transit numbers or the number of exchange attempts required for convergence of the indicated simulated replica-scheme sizes and average exchange probabilities. For c, the mean number of exchange attempts required to achieve convergence of the observed replica distribution is shown, and for each datum, at least 1000 iterations were performed, so that in all cases the standard error of the sample mean was less than 1% of the sample mean. Thus, the error bars on these data are insignificant on the scale of the ordinate axis and are not shown. All axes are logarithmic.

A C++ program was written to implement the above algorithm. An initial density of 1 was assigned to each ensemble, and the transit number was observed after at least 0.95 density had been unmarked. The transit number was found for a range of average exchange probabilities and replica-scheme sizes. The results are plotted in Figure 1a.

The gradients of the lines on the log−log plot of Figure 1a clearly indicate the inverse proportionality between average exchange probability and transit number over a common range of probabilities and numbers of replicas. The spacing between the lines indicates that the transit number increases sublinearly with the number of replicas also (data not shown). Together, these trends illustrate that the minimum length of a simulation with an efficient replica-exchange scheme increases with the number of replicas and in roughly inverse proportion with the exchange probability.

It would be possible to calculate the number of exchange attempts necessary for a given probability of a number of transits occurring. Because multiple transits by different replicas can overlap in time, this will not be a simple function of the transit number. It is not clear how many transits indicate an efficient replica-exchange scheme.

We note also that the above is only a satisfactory model of an efficient REMD simulation when the latter achieves independence of replica-exchange events by simulating for a sufficiently long period between exchange attempts.

2.1.2. Metric Two. A second metric to test for efficient conduct of a REMD simulation can be derived from the expectation that, in the limit of a large number of exchange attempts, the probability that a replica in ensemble $i \in \{1,..., M\}$ began the simulation in any ensemble $j \in \{1,..., M\}$ approaches $1/M$. This convergence can again be studied in a simulation. We define the *replica probability matrix* $p(k)$ as the matrix whose $p(k)_{ij}$ elements are the probabilities that, after k exchange attempts, the replica in ensemble i began the simulation in ensemble j. $p(0)_{ij}$ is a diagonal matrix with unity on the leading diagonal, and in the limit of large k, all elements approach $1/M$.

A second C++ program was written to characterize this behavior. The criterion for assessing convergence after k exchange attempts over M replicas was the rmsd of $p(k)$ from its limiting value. This value is initially $\sqrt{(M-1)/M}$ and approaches zero as the replica system mixes. The convergence criterion was defined as

$$\sqrt{\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\left(\frac{1}{M} - p(k)_{ij}\right)^2} < 10^{-3} \qquad (6)$$

The number of exchange attempts necessary to achieve this convergence for a range of probabilities and replica-scheme sizes is shown in Figure 1b. Note again that the number of exchange attempts required to converge the replica probability matrix does not ensure that all replicas will have been exposed to all ensembles equally; it merely sets a lower bound below which reasonably equal exposure cannot have occurred.

The same kind of proportionality as for Metric One is evident, with the number of steps for convergence inversely proportional to average exchange probability.

The arbitrary nature of the convergence criteria chosen in constructing the plots in Figure 1a and b means that the actual number of exchange attempts on the ordinate axes is not significant. The value of these plots is in making clear how the effectiveness of the replica-exchange scheme depends on the number of replicas and the average exchange probability.

2.1.3. Metric Three. We carried out one further simulation of a replica-exchange process. This study followed the progress of the replicas and recorded the number of exchange attempts needed to converge the observed frequency of each replica in each ensemble to $1/M$. The criterion for assessing convergence after k exchange attempts over M replicas was the approach of the rmsd to zero, namely

$$\sqrt{\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\left(\frac{1}{M} - f(k)_{ij}\right)^2} < 10^{-3} \qquad (7)$$

where $f(k)_{ij}$ is the observed relative frequency with which the replica that began the simulation in ensemble i is in ensemble j after k exchange attempts. The number of exchange attempts required for convergence over a number of iterations was noted, and the mean and standard error of this sample were calculated for a range of exchange probabilities and sizes of replica schemes, as shown in Figure 1c. This plot again shows that the number of exchange attempts required increases inversely with the average exchange probability. However, in this model of replica exchange, there is very little dependence on the number of replicas; the reasons for this are not clear.

2.2. MD Protocol for Measuring Potential Energy Autocorrelation Time. The 357-atom, 23-residue N-terminal fragment of the N-terminal H4 histone tail studied by Lins and Rothlisberger[33] was solvated in a cubic box of normal TIP3P[34] water of side length 5.9 nm with nine chloride ions to neutralize the total charge. The all-atom CHARMM22 forcefield[35] in GROMACS 3.3.1[30–32] was used to model the peptide, with smooth particle-mesh Ewald (PME)[36] treatment of long-range electrostatics in an NpT ensemble. Berendsen coupling[37] to independent thermal baths was used for solute and nonsolute groups, both with a time constant of 0.5 ps, and the Berendsen algorithm was also used for pressure coupling with a time constant of 1 ps. SETTLE[38] and LINCS[39] were used to constrain all bond lengths in the solvent and solute, respectively. The system was equilibrated for 10 ps in NVT with position restraints and then for 50 ps in NpT, beginning from a canonical α helix. The length of the NpT production simulation was 20 ns. In the PME algorithm, the real-space summation was cut off at 1.0 nm, and the Ewald parameter was chosen to bound the error in the real-space summation such that $\mathrm{erfc}(\beta r)/r < 10^{-5}$ for r greater than the above cutoff. In the reciprocal-space approximation, the β-spline interpolation order was 6, and a Fourier grid spacing of 0.10 nm was used. van der Waals interactions were modeled with a 6,12 Lennard-Jones potential truncated at a maximum interaction length of 1.2 nm, with a long-range dispersion correction applied.

3. Analysis of Efficiency of REMD Simulations Reported in the Literature

We performed a thorough survey of published temperature REMD simulations on protein or peptide systems using explicit solvent models. Although all authors reported the number of replicas, temperature range, and simulation length, there was wide variation in the reporting of the expected and observed ranges of the frequency of successful exchange

attempts. Several authors made no mention at all of these latter values, which makes it extremely difficult to assess whether the use of the particular REMD protocol was justified. Some authors neglected to record either the number of exchange attempts or the period between them. For those who did report these quantities, their values fell into ranges of 400−500 000 and 0.01−5 ps, respectively. A significant number of authors did not report the frequency with which they collected data from the simulation for their analyses; however, this is not necessarily pertinent to the validity of their conclusions.

In the following sections, we discuss several aspects of this survey. We propose and test a method for estimating the average exchange probability where the authors did not report it. We then use this method to increase the size of our data set for a comparison of published simulations, using Metric One to indicate the extent to which these simulations achieved mixing efficiency.

3.1. Estimating Average Exchange Probabilities. Although we wanted to study the number of exchange attempts used in reported REMD simulations as a function of the number of replicas and the average exchange probability, we were hampered by the shortage of published data with a complete description. Consequently, we then aimed to extend the size of this data set available in the literature, by testing a method of estimating the average exchange probability to use in cases where that value was not given. It is possible to estimate the frequency of exchanges in the following way,[40] by using properties of the molecular system and the reported temperature range. From the equipartition theorem, we can approximate

$$U_i = U_{i+1} N_{\mathrm{df}} \frac{c}{2} k_{\mathrm{B}} (T_i - T_{i+1}) \tag{8}$$

where N_{df} is the number of degrees of freedom for the system, and c is 2 for protein−water systems.[40] If replica temperatures are distributed approximately exponentially (as is commonly the case in REMD simulations), so that successive temperatures have a constant ratio $1 + \varepsilon$, that is, $T_{i+1} = (1 + \varepsilon)T_i$, then from the definition of the replica-exchange probability in eq 5 we have

$$P(\text{exchange}) = \min\left(1, \exp\left[-\frac{\varepsilon^2 N_{df}}{1+\varepsilon}\right]\right) \tag{9}$$

In order to test the above estimation process, a data set of 33 simulations from 21 papers[18,41–60] from 2001 was selected, representing cases where the authors reported either expected or observed exchange probabilities and enough descriptive data to generate an *a priori* estimate of the exchange probabilities. The value of ε, where not reported, was inferred from the reported number of replicas and temperature range, under the supposition that the temperatures were distributed exponentially. The number of degrees of freedom of a condensed-phase system is significantly lower than the familiar $3N - 6$ used in single-molecule analyses. Many degrees are correlated, and some are eliminated methodologically, such as by the use of bond constraint algorithms. Not all authors reported their use of bond constraints in either the solute or solvent parts of their systems. The assumption of constraints on all bonds implies the approximation of $N_{\mathrm{df}} \approx 2N_{\mathrm{atoms}}$, which leads to poor correlation between reported and estimated exchange probabilities for the same simulation. Approximating $N_{\mathrm{df}} \approx N_{\mathrm{atoms}}$ led to a better correlation, as shown in Figure 2. This demonstrates some qualitative predictive value for this crude approximation. Hence, we used it to generate estimates of the average exchange probability in cases where the authors had not specified the information.

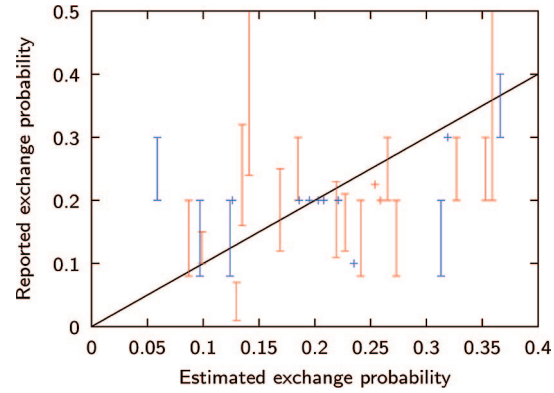

Figure 2. Comparison of estimated and reported average exchange probabilities. Reported replica-exchange probabilities for simulations described in refs 18, 41–60 are compared *a priori* with our estimates based on eq 9. Cases where the source reported only the replica-exchange probabilities expected by the authors in constructing the replica-exchange scheme are shown in blue, and those where the observed probabilities were reported are shown in red. Cases where the source reported a range of probabilities are indicated with a vertical bar, but those where the source only reported an average probability are plotted with a single point. The black line indicates perfect correlation. A version of this figure with the sources of the data points labeled is given in the Supporting Information.

3.2. Comparison of Mixing Efficiency of Simulations from the Literature. A second, intersecting data set of 41 simulations from 23 papers[22,41–54,56–59,61–64] was constructed, for which the authors had reported the exchange attempt frequency, the exchange period, and either the expected or observed exchange probability, or a sufficient atomic description to estimate that probability using eq 9. For these simulations, the reported number of exchange attempts and the exchange period were plotted in Figure 3 against the average expected, observed, or estimated exchange probability, and the transit number lines from Figure 1a are shown for reference.

It is desirable for a simulation to have an exchange attempt period that is long enough to ensure uncorrelated exchanges, and for that simulation to attempt sufficient exchanges for exploration of the replica space to take place, that is, to mix efficiently. By our construction in section 2.1.1, the transit number is a lower bound below which a simulation cannot be considered efficient from a computational point of view. It is clear from Figure 3b that all of the simulations used sufficient exchange attempts to exceed the transit number for their exchange probability and number of replicas. However, with one exception, all simulations were within 2 orders of magnitude above the transit number. We contend

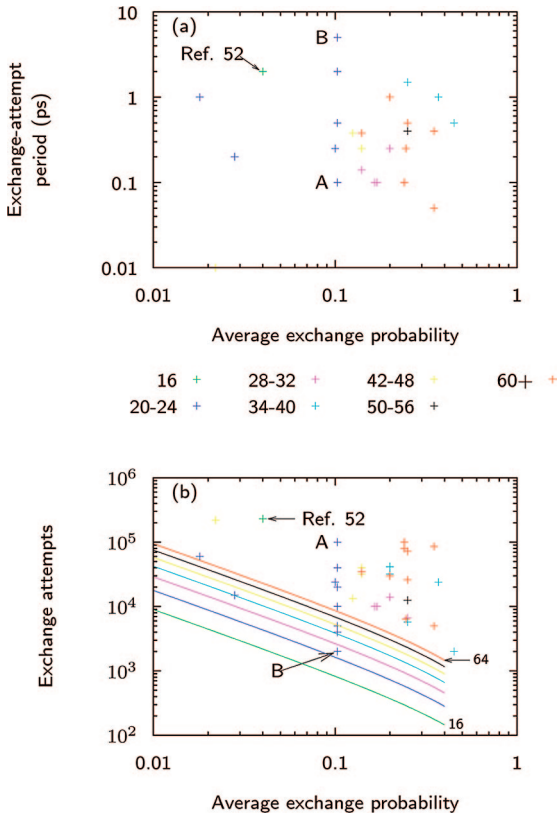

Figure 3. Comparison of replica scheme parameters from the literature. The (a) exchange period and (b) number of exchange attempts, reported in refs 22, 41–54, 56–59, 61–64, are plotted as a function of the average exchange probability (reported or estimated) and number of replicas (indicated by different symbols). Also superposed in b are some of the transit number lines from Figure 1a, indicatively labeled by replica number. The exchange attempt data are grouped according to the number of replicas in the simulation, in subranges of width 8. The corresponding transit number line has the same color for comparison. All axes are logarithmic. A version of b with the sources of the data points labeled is given in the Supporting Information.

that many transits are necessary to achieve mixing efficiency, and that this requires a simulation to exceed the transit number by a considerable degree. Accordingly, we recommend that a simulation exceed the transit number by at least 2 orders of magnitude.

Figure 3a illustrates that some simulations exceeded the transit number, but they did so with a questionably short exchange attempt period, as mentioned earlier and discussed further below. Of interest was the group of simulations with an average exchange probability of 0.1 using 24 replicas, which were all from the comparative study of Periole and Mark discussed earlier.[22] The data with high numbers of exchange attempts had correspondingly low exchange periods (labeled A in Figure 3), and vice-versa (labeled B in Figure 3). Thus, the appearance in these simulations of either a suitably long exchange attempt period in Figure 3a or high number of exchange attempts in Figure 3b is offset by a deficit in the other quantity, and none approaches the quality of the exceptional simulation.[52]

That exceptional simulation[52] mentioned above had an average exchange probability of 0.04, a number of exchange attempts of 230 000, and an exchange period of 2 ps. We believe that the other 40 simulations studied here used a number of exchange attempts that was too few to make effective use of their entire replica space, or they did so with a questionably low exchange-attempt period. Consequently, these simulations did not mix efficiently, even if they achieved thermodynamic efficiency by sampling the ensembles in a way suitable for the observables needed for their conclusions.

Some authors provided plots of the random walks in temperature space taken by one or several of their replicas as an indication that the replica-exchange method was working properly. Although helpful, we suggest these should be accompanied by a report of (say) the number of transits for the generalized ensemble. This could be sobering, particularly in cases where "random walk plots" reveal only one or two transits per replica.

The scaling behavior we have demonstrated for our metrics shows that the simulation length required for mixing efficiency increases with the number of replicas. This has implications for the decorrelation time of Lyman and Zuckerman[29] referred to earlier. The simulation length required for convergence of the individual replica trajectory decorrelation times will increase as the number of replicas increases. The value to which the decorrelation times converge can only decrease, however. Thus, we anticipate the existence of a negative correlation between our *a priori* transit number and their *a posteriori* decorrelation time. It would be interesting to test for this correlation by undertaking some REMD simulations on suitable model peptides. Various authors have observed the tendency of REMD simulations on peptides to predict an unphysical flattening of melting curves;[44,52] a method to address this issue partially exists.[65] Inadequate mixing of replicas could contribute to this effect. We propose that a suitable adaptation of this "decorrelation time" method evaluated at constant temperature, despite the discontinuities, might prove to be a useful indicator of per-replica-temperature thermodynamic convergence.

Finally, we need to reiterate that the data underlying Figure 1a−c were all generated on the assumption that each exchange attempt was statistically independent from all other attempts. This will only be true in a real REMD simulation if the period of simulation time between exchange attempts is greater than the autocorrelation time of the potential energy of the system. This is almost certainly not true for some of the shorter exchange periods for the peptide systems in our sample literature data set, illustrated in Figure 3a. If it were true, then sufficient configurational sampling would likely be occurring at the minimum, and in such cases, application of the replica-exchange method would not be useful. As no data are available to calculate such autocorrelation times, we cannot comment further on the data in Figure 3, except to observe that, for a system where the exchange probabilities are correlated between successive exchanges, the number of exchange attempts required to exceed the transit number will be higher than that for the uncorrelated case. Thus, the general trend, where the number of exchange attempts reported by these authors is possibly too few to justify the number of replicas (and, thus, the temperature space spanned),

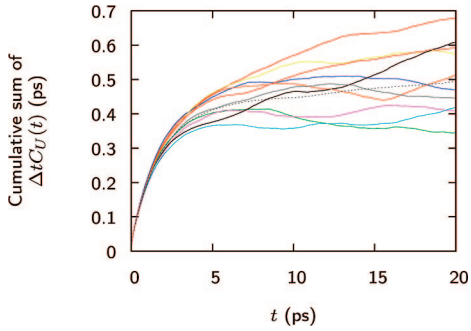

Figure 4. Behavior of potential energy autocorrelation functions. The product of the sampling interval Δt and the cumulative sum of the potential energy autocorrelation time $C_U(t)$ is plotted against time t for 0−20 ps. Plots for 1 ns autocorrelation functions based on 10 disjointed 2 ns fragments are shown in color, and the sum for the 10 ns autocorrelation function based on the whole 20 ns simulation is shown as a dotted line.

is likely to be even worse than it appears in Figure 3. We explore the behavior of the autocorrelation of potential energy further in the next section.

3.3. Estimation of Potential Energy Autocorrelation Time. To investigate further the interpretation of Periole and Mark[22] of the possible impact of correlation on the mixing efficiency of simulations, as discussed in the Introduction and above, we performed a MD simulation without replica exchange to estimate the autocorrelation time of the potential energy. The system studied was a 23-residue, 357-atom N-terminal fragment of the N-terminal H4 histone tail used in the comparative analysis of electrostatic treatments by Lins and Rothlisberger.[33] Over the 20 ns simulation, the initial helix unfolded progressively from the N-terminal end, retaining a helix turn at the C-terminal end. This is in qualitative agreement with the findings of Lins and Rothlisberger (see Figure 2C of ref 33) for this peptide under PME electrostatics.

The potential energy U of this system was recorded at $\Delta t = 0.01$ ps intervals, and the autocorrelation function $C_U(t)$ was found over the time range $\{0... P\Delta t\}$ from a simulation fragment of length $Q\Delta t$ (with $P \leq Q/2$) using

$$C_U(j\Delta t) = \frac{1}{P}\sum_{i=0}^{Q-1-P} U(i\Delta t)\, U[(i+j)\Delta t] \qquad (10)$$

Although other algorithms for computing an autocorrelation function exist,[40] this one has the virtue that the same number of simulation observations (P) contribute to each point of the function, and so the expectation of statistical error is consistent along its length. The 20 ns simulation was broken into 10 disjointed 2 ns fragments, and a 1 ns autocorrelation was derived for each (thus $Q = 2 \times 10^5$ and $P = 10^5$ for each). At large times, the exponential decay of $C_U(t)$ is swamped by the constant noise from the simulation. This is clearly seen in Figure 4, which plots the product of Δt and the cumulative sum of $C_U(t)$ against t.

Determining an integrated autocorrelation time can be difficult because the variance of the truncated-series estimator of $\tau_{\mathrm{int},f}$ from eq 2 diverges as $t \rightarrow \infty$ because of the aforementioned noisy behavior of $C_f(t)$ at large t. Thus, the series needs to be truncated before this noise dominates. Averaging over the 10 2 ns autocorrelation functions produced above, and applying eq 2 with time truncated at 10 ps, we find $\tau_{\mathrm{int},u} = 0.47$ ps. Our choice of 10 ps is justified by the curves in Figure 4, which suggests that noise is dominant at or about this time. Further, we note that the value of $C_U(t)$ first becomes negative (indicating dominance of noise) at 9.96 ps and 11.96 ps, respectively, for the 10 ns autocorrelation and the average over 10 2 ns autocorrelation functions. This noise could be reduced by extending the length of the autocorrelation; this is confirmed by the plot based on a 10 ns autocorrelation, which is dotted in Figure 4. However, the present results are sufficient for the purpose of indicating the size of the integrated autocorrelation time of the potential energy.

Thus, we conclude that successive measurements of U for such peptide-in-water systems are not independent unless at least around 0.5 ps has passed. This agrees with the findings of Zhang et al.[19] and Periole and Mark[22] that an exchange period under 1 ps is too short for the replica-exchange attempts to be regarded as independent. As may be seen in Figure 3a, there are several published studies performed on simulations that have strong similarities to the one studied here, for which the replica-exchange periods are below 1 ps. At the very least, these replica schemes did not achieve the mixing efficiency indicated by the number of exchanges they attempted.

4. How-To Guide for REMD with Suitable Mixing Efficiency

On the basis of our analyses and observations, we have developed a general strategy guide for designing REMD simulations so as to ensure a demonstrable level of mixing efficiency. This is described below.

1. Choose the simulation system and parameters to arrive at a full description, including the thermodynamic ensemble, solvent model, ionic concentration, and periodic box size.

2. Having judged the nature of the expected free energy surface, choose a range of temperatures over which to distribute replicas that is sufficient to permit replicas to cross the relevant activation barriers, and not greatly excessive.[23]

3. Given a number of available processors, *either* distribute a given number of replicas over the temperature range either exponentially or according to some algorithm to deal with generalized free energy surface bottlenecks,[15] and either estimate or test for the average exchange probability, *or*, given a desired average exchange probability, find a distribution of replicas that achieves this, either empirically or using an algorithm[16] as implemented at http://folding.bmc.uu.se/remd/index.php.

4. Calculate or estimate the potential energy autocorrelation time of the system at the temperature(s) of interest, and choose an exchange period at least this long. As a general rule for all-atom simulations of peptides in explicit solvents, we recommend no smaller than 1 ps.

5. Given the average exchange probability and the number of replicas, look up the transit number in Figure 1a, and choose a number of exchange attempts considerably larger

than this. Together with the exchange period, this determines the minimum simulation length for mixing efficiency.

If the above scheme leads to a simulation cost that is too great for the available resources, then reducing the size of the replica range is the simplest way to reduce both the number of replicas and the length of the simulation. The impact of such a reduction on the thermodynamic efficiency of the method will normally be difficult to judge *a priori*.

5. Conclusion

In summary, from the literature survey we conducted for this work, it is clear that authors are not routinely reporting information necessary for readers to assess the quality of the data underlying the analysis. The expected or observed replica-exchange probability, together with the total number of exchange attempts, the exchange period, and the estimates of the autocorrelation time of the potential energy should be given. We have provided a step-by-step guide to help in constructing protocols for REMD simulations that can mix efficiently. Two key recommendations are that the exchange period should not be smaller than about 1 ps and the number of exchange attempts should be chosen to significantly exceed the transit number for the replica-exchange scheme.

Replica-exchange methods offer great promise for improving the ability of computational chemists to sample thermodynamic ensembles of protein and peptide systems with efficiency. After a slow start in the literature, their recent implementation in common MD simulation programs has made them easily accessible, and they have become a popular method to improve sampling efficiency. However, the lesson of our work is that it is necessary to use these methods in a thoughtful and measured fashion in order to ensure that the desired increase in sampling is actually realized with mixing efficiency, and to properly report the simulation details.

Acknowledgment. We acknowledge support from the Australian National University Institute for Advanced Studies (IAS) block grant, and the APAC (Australian Partnership for Advanced Computing) for computer-time grants. We thank the reviewers for their constructive comments.

Supporting Information Available: Versions of Figures 2 and 3 labeled by reference number. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(2) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.

(3) Affentranger, R.; Tavernelli, I.; Di Iorio, E. E. *J. Chem. Theory Comput.* **2006**, *2*, 217–228.

(4) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749–13754.

(5) Lyman, E.; Ytreberg, F. M.; Zuckerman, D. M. *Phys. Rev. Lett.* **2006**, *96*, 028105.

(6) Christen, M.; van Gunsteren, W. F. *J. Chem. Phys.* **2006**, *124*, 154106.

(7) Liu, P.; Voth, G. A. *J. Chem. Phys.* **2007**, *126*, 045106.

(8) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058–9067.

(9) Jang, S. M.; Shin, S.; Pak, Y. *Phys. Rev. Lett.* **2003**, *91*, 8305.

(10) Cheng, X. L.; Cui, G. L.; Hornak, V.; Sinnnerling, C. *J. Phys. Chem. B.* **2005**, *109*, 8220–8230.

(11) Woods, C. J.; Essex, J. W.; King, M. A. *J. Phys. Chem. B.* **2003**, *107*, 13703–13710.

(12) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435–13441.

(13) Piana, S.; Laio, A. *J. Phys. Chem. B.* **2007**, *111*, 4553–4559.

(14) Li, H. Z.; Li, G. H.; Berg, B. A.; Yang, W. *J. Chem. Phys.* **2006**, *125*, 144902.

(15) Trebst, S.; Troyer, M.; Hansmann, U. H. *J. Chem. Phys.* **2006**, *124*, 174903.

(16) Patriksson, A.; van der Spoel, D. *Phys. Chem. Chem. Phys* **2008**, *10*, 2073–2077.

(17) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comput.* **2006**, *2*, 420–433.

(18) Sanbonmatsu, K. Y.; Garcia, A. E. *Proteins* **2002**, *46*, 225–234.

(19) Zhang, W.; Wu, C.; Duan, Y. *J. Chem. Phys.* **2005**, *123*, 154105.

(20) Zuckerman, D. M.; Lyman, E. *J. Chem. Theory Comput.* **2006**, *2*, 1200–1202.

(21) Zuckerman, D. M.; Lyman, E. *J. Chem. Theory Comput.* **2006**, *2*, 1693.

(22) Periole, X.; Mark, A. E. *J. Chem. Phys.* **2007**, *126*, 014903.

(23) Nymeyer, H. *J. Chem. Theory Comput.* **2008**, *4*, 626–636.

(24) Wolff, U. *Comput. Phys. Commun.* **2004**, *156*, 143–153.

(25) Madras, N.; Sokal, A. D. *J. Stat. Phys.* **1988**, *50*, 109–186.

(26) Sindhikara, D.; Meng, Y.; Roitberg, A. E. *J. Chem. Phys.* **2008**, *128*, 024103.

(27) Kofke, D. A. *J. Chem. Phys.* **2002**, *117*, 6911–6914.

(28) Rathore, N.; Chopra, M.; de Pablo, J. J. *J. Chem. Phys.* **2005**, *122*, 024111.

(29) Lyman, E.; Zuckerman, D. M. *J. Phys. Chem. B* **2007**, *111*, 12876–12882.

(30) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(31) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.

(32) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

(33) Lins, R. D.; Rothlisberger, U. *J. Chem. Theory Comput.* **2006**, *2*, 246–250.

(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(35) Mackerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(36) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(37) Berendsen, H. J.; Postma, J. P.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(38) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.

(39) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(40) van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L.; Feenstra, K. A.; van Drunen, R.; Berendsen, H. J. Gromacs User Manual, version 3.3; University of Groningen: Groningen, The Netherlands, 2005.

(41) Garcia, A. E.; Sanbonmatsu, K. Y. *Proteins* **2001**, *42*, 345–354.

(42) Zhou, R. H.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931–14936.

(43) Garcia, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782–2787.

(44) Garcia, A. E.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13898–13903.

(45) Ghosh, T.; Garde, S.; Garcia, A. E. *Biophys. J.* **2003**, *85*, 3187–3193.

(46) Gnanakaran, S.; Garcia, A. E. *Biophys. J.* **2003**, *84*, 1548–1562.

(47) Zhou, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280–13285.

(48) Garcia, A. E. *Polymer* **2004**, *45*, 669–676.

(49) Nguyen, P. H.; Stock, G.; Mittag, E.; Hu, C. K.; Li, M. S. *Proteins* **2005**, *61*, 795–808.

(50) Nguyen, P. H.; Mu, Y. G.; Stock, G. *Proteins* **2005**, *60*, 485–494.

(51) Nishino, M.; Sugita, Y.; Yoda, T.; Okamoto, Y. *FEBS Lett.* **2005**, *579*, 5425–5429.

(52) Seibert, M. M.; Patriksson, A.; Hess, B.; van der Spoel, D. *J. Mol. Biol.* **2005**, *354*, 173–183.

(53) Yoshida, K.; Yamaguchi, T.; Okamoto, Y. *Chem. Phys. Lett.* **2005**, *412*, 280–284.

(54) Baumketner, A.; Bernstein, S. L.; Wyttenbach, T.; Lazo, N. D.; Teplow, D. B.; Bowers, M. T.; Shea, J. E. *Protein Sci.* **2006**, *15*, 1239–1247.

(55) Gnanakaran, S.; Nussinov, R.; Garcia, A. E. *J. Am. Chem. Soc.* **2006**, *128*, 2158–2159.

(56) Juraszek, J.; Bolhuis, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15859–15864.

(57) Mu, Y. G.; Nordenskiold, L.; Tam, J. P. *Biophys. J.* **2006**, *90*, 3983–3992.

(58) Sugita, Y.; Miyashita, N.; Yoda, T.; Ikeguchi, M.; Toyoshima, C. *Biochemistry* **2006**, *45*, 11752–11761.

(59) Wickstrom, L.; Okur, A.; Song, K.; Hornak, V.; Raleigh, D. P.; Simmerling, C. L. *J. Mol. Biol.* **2006**, *360*, 1094–1107.

(60) Zhang, J.; Qin, M.; Wang, W. *Proteins* **2006**, *62*, 672–685.

(61) Gustiananda, M.; Liggins, J. R. R.; Cummins, P. L. L.; Gready, J. E. E. *Biophys. J.* **2004**, *86*, 2467–2483.

(62) Kawashima, Y.; Sugita, Y.; Yoda, T.; Okamoto, Y. *Chem. Phys. Lett.* **2005**, *414*, 449–455.

(63) Nymeyer, H.; Woolf, T. B.; Garcia, A. E. *Proteins* **2005**, *59*, 783–790.

(64) Baumketner, A.; Bernstein, S. L.; Wyttenbach, T.; Bitan, G.; Teplow, D. B.; Bowers, M. T.; Shea, J. E. *Protein Sci.* **2006**, *15*, 420–428.

(65) van der Spoel, D.; Seibert, M. M. *Phys. Rev. Lett.* **2006**, *96*, 0238102.

CT800016R