TIGER2: An improved algorithm for temperature intervals with global exchange of replicas

Xianfeng Li,¹ Robert A. Latour,^{1,a)} and Steven J. Stuart²

¹*Department of Bioengineering, Clemson University, Clemson, South Carolina 29634, USA*

²*Department of Chemistry, Clemson University, Clemson, South Carolina 29634, USA*

(Received 19 January 2009; accepted 10 April 2009; published online 7 May 2009)

An empirical sampling method for molecular simulation based on “temperature intervals with global exchange of replicas” (TIGER2) has been developed to reduce the high demand for computational resources and the low computational efficiency of the conventional replica-exchange molecular dynamics (REMD) method. This new method overcomes the limitation of its previous version, called TIGER, which requires the assumption of constant heat capacity during quenching of replicas from elevated temperatures to the baseline temperature. The robustness of the TIGER2 method is examined by comparing it against a Metropolis Monte Carlo simulation for sampling the conformational distribution of a single butane molecule in vacuum, a REMD simulation for sampling the behavior of alanine dipeptide in explicit solvent, and REMD simulations for sampling the folding behavior of two peptides, (AAQAA)₃ and chignolin, in implicit solvent. The agreement between the results from these conventional sampling methods and the TIGER2 simulations indicates that the TIGER2 algorithm is able to closely approximate a Boltzmann-weighted ensemble of states for these systems but without the limiting assumptions that were required for the original TIGER algorithm. TIGER2 is an efficient replica-exchange sampling method that enables the number of replicas that are used for a replica-exchange simulation to be substantially reduced compared to the conventional REMD method. © 2009 American Institute of Physics.

[DOI: [10.1063/1.3129342](https://doi.org/10.1063/1.3129342)]

I. INTRODUCTION

The replica-exchange molecular dynamics (REMD) method has been extensively used over the past decade for searching the conformational phase space of complex molecular systems.^{1–7} The REMD method conducts a number of parallel simulations at different temperature levels and periodically attempts to exchange configurations between adjacent levels with an exchange criterion that guarantees Boltzmann sampling. By sampling and exchanging over the temperature space, the molecular system is able to escape from local energy minima, thus overcoming the kinetic trapping problem experienced in conventional MD simulations.^{8–12} Still, practical difficulties remain in the application of REMD, notably its poor scaling with system size. REMD must use a large number of relatively closely spaced temperature levels to create sufficient overlap between the potential energy probability distributions for replicas at neighboring temperatures in order to obtain an acceptable rate of replica swapping. The number of replicas, N_r , needed to span a designated temperature range is proportional to $\sqrt{N}$, with N being the total number of degrees of freedom of the system.¹³ The use of a large number of replicas in a REMD simulation results in a further complication due to the time required for the diffusion of replicas over the numerous temperature levels.¹ Modeled as a random walk¹⁴ over temperature space, the time cost of diffusion from the highest to the lowest temperature level scales as N_r^2 . There-

fore, the addition of each replica quadratically increases the number of swaps that must be attempted before a Boltzmann-weighted ensemble of states can be obtained at the temperature level of interest, thus substantially increasing the time required for convergence. These two limitations make the use of REMD particularly problematic for application to large systems. Given that the single-replica simulation cost increases (at least) linearly with system size, this results in simulations with wall clock times that scale as $O(N^2)$ and total computational cost that scales as $O(N^{2.5})$.

To overcome the limitations of the REMD method with respect to the high demand of computational resources and the low efficiency caused by slow diffusion across the temperature space, we previously developed a new sampling method that we named “temperature intervals with global energy reassignment” (TIGER).¹⁵ If the assumptions underlying this method are met, TIGER generates a Boltzmann-weighted ensemble of states at the baseline temperature of interest using a novel replica-exchange scheme where the number of replicas used is greatly reduced. Unlike REMD, which depends on exchange between thermally adjacent replicas, TIGER uses a quenching cycle in which all replicas are quenched to the same baseline temperature at the end of each sampling cycle, considered for exchange using an algorithm that satisfies detailed balance (if underlying assumptions are met), and then globally reassigned and reheated to the designated temperature levels. By this approach, the quenching procedure removes energy differences that are purely thermal, thus providing the ability to greatly reduce the number

^{a)}Electronic mail: latourr@clemson.edu.

of replicas needed for a given size of the system and temperature range to be spanned. In addition, the global reassignment procedure eliminates the diffusion problem that is inherent in the conventional REMD method, thus substantially enhancing sampling efficiency. This method satisfies the detailed balance condition under the assumptions that quenching is conducted in a manner which ensures that the general conformational basin of the system (i.e., the dihedral angles of a protein) is maintained during the quenching process and the amount of thermal energy removed during quenching is not state dependent (i.e., the heat capacity of the system is state independent as long as state transitions do not occur during quenching). TIGER was successfully applied for sampling the conformational space of alanine dipeptide in explicit solvent, represented by the 3-site rigid TIP3P water model,¹⁵ a system for which these assumptions hold. However, subsequent simulations with more complex systems, such as peptide folding with implicit solvation, demonstrated that the assumptions underlying the method are not easily maintained, resulting in a failure to generate a proper Boltzmann-weighted ensemble.

In this work, we have revised the previous TIGER method to provide a new empirically based algorithm that maintains the same advantages of TIGER without requiring its limiting assumptions. The new algorithm, which we call “temperature intervals with global exchange of replicas,” or TIGER2, does not rigorously satisfy detailed balance; however, as shown by Manousiouthakis and Deem,¹⁶ detailed balance represents an overly strict condition that is sufficient but not necessary to provide a Boltzmann-weighted ensemble of sampled states. Instead, it is sufficient (and necessary) to satisfy the weaker “balance” condition.¹⁶ This balance condition relaxes the requirement that individual pairs of states be in local equilibrium and requires only that the sampling provided by an algorithm converges to a Boltzmann-weighted distribution of states and maintains Boltzmann weighting as sampling is continued. The ability of this revised method to generate Boltzmann-like statistics at the baseline temperature is demonstrated empirically in this work for a designated set of molecular systems first by comparison to a standard Metropolis Monte Carlo (MC) simulation of a simple atomistic model of a single butane molecule in vacuum. The performance of TIGER2 is then further illustrated by applying it to successfully sample the conformational distribution of alanine dipeptide in explicit water and to fold two small peptides, (AAQAA)₃ and chignolin, under implicit solvation conditions in comparison to REMD simulations. The simulation results of TIGER2 closely compare with the Metropolis MC and REMD simulation results as well as the available experimental measurements, thus demonstrating the utility of this new advanced sampling method.

II. METHODS

A. The REMD method

In brief, a REMD simulation constructs a Boltzmann-weighted ensemble by generating independent trajectories at different temperatures, T_m ($m = 1, \dots, N_r$), with N_r being the

number of temperature levels or replicas. An individual replica in this ensemble is represented by $(\mathbf{X}(T_m), \mathbf{P}(T_m))$, where $\mathbf{X}(T_m)$ and $\mathbf{P}(T_m)$ represent the positions and the momenta of all atoms of the m th replica. The acceptance probability for an attempt to exchange two replicas $(\mathbf{X}(T_i), \mathbf{P}(T_i))$ and $(\mathbf{X}(T_j), \mathbf{P}(T_j))$ is

$$\min[1, \exp(-(\beta_j - \beta_i)(E_i - E_j))], \quad (1)$$

where E_i and E_j are the potential energies of replicas i and j , respectively, and β_i and β_j represent $(k_B T_i)^{-1}$ and $(k_B T_j)^{-1}$, respectively, with k_B being Boltzmann’s constant. After a successful exchange, the momenta of the swapped replicas are updated by

$$\mathbf{P}(T_i) = \sqrt{\frac{T_i}{T_j}} \mathbf{P}(T_j), \quad \mathbf{P}(T_j) = \sqrt{\frac{T_j}{T_i}} \mathbf{P}(T_i). \quad (2)$$

The acceptance ratio for the swapping process decreases rapidly with increases in system size or the temperature intervals between replicas. As the number of replicas is increased for a given temperature range, the acceptance probability increases, but this comes at a cost of increasing the amount of time required for the diffusion of replica configurations from the high to the low temperature levels. Further details of the REMD method can be found elsewhere.^{1,2}

B. The TIGER2 method

Similar to REMD, a TIGER2 simulation involves a number of replicas (N_r) running in parallel at different temperature levels. However, in contrast to REMD, which involves paired exchanges between neighboring temperature levels, TIGER2 is composed of a series of cycles where the potential energies of all replicas are compared at the same time and at the same baseline temperature using a standard Metropolis MC decision algorithm.¹⁷ Each cycle contains four stages: (1) rapid heating from a baseline temperature (T_B) to the replica temperature (T_m) by rescaling the momenta of the atoms within the replica by a factor of $\sqrt{T_m/T_B}$ and thermally equilibrating, (2) molecular dynamics sampling at constant temperature (T_m), (3) rapid quenching back down to T_B by rescaling the momenta by a factor of $\sqrt{T_B/T_m}$ followed by thermal equilibration (note that the coldest replica is simulated at the baseline temperature, i.e., $T_1 = T_B$, and thus does not need to be quenched), and (4) replica comparison, selection, and global reassignment to temperature levels for the next cycle.

In order to understand how the set of quenched states must be appropriately generated and compared with the sampled state at T_B , we must first consider the requirements that are necessary for the condition of detailed balance to be satisfied. For a Metropolis MC simulation, the detailed balance condition can be described by the following equation:¹⁰

$$P(X_i)P(X_j|X_i)\Pi_{i \rightarrow j} = P(X_j)P(X_i|X_j)\Pi_{j \rightarrow i}, \quad (3)$$

where $P(X_i)$ and $P(X_j)$ are the probabilities of states X_i and X_j in the ensemble, respectively, $P(X_j|X_i)$ and $P(X_i|X_j)$ are the conditional probabilities of attempting a move to state X_j from state X_i and to state X_i from state X_j , respectively, and $\Pi_{i \rightarrow j}$ and $\Pi_{j \rightarrow i}$ are the acceptance probabilities for an

attempted move from state X_i to X_j and from X_j to X_i , respectively. For a Metropolis MC simulation, a symmetric process is almost always used to generate candidate moves, with the result that $P(X_j|X_i)=P(X_i|X_j)$. These terms are therefore often omitted in the detailed balance expression. However, for TIGER2, the second state (i.e., the one selected from the collection of quenched states) is generated completely independently of the first state (i.e., the state sampled at T_B) and quenching from the various elevated-temperature levels to T_B leads to a different probability distribution than that which is sampled at the baseline temperature. Therefore, for the TIGER2 algorithm, the $P(X_j|X_i)$ is not equal to $P(X_i|X_j)$ and is not analytically accessible, so the algorithm cannot be shown to satisfy the detailed balance condition.

However, while an algorithm satisfying detailed balance is guaranteed to eventually generate a Boltzmann distribution of states, detailed balance is an overly strict condition to ensure a valid Boltzmann-sampling method. As was pointed out by Manousiouthakis and Deem,¹⁶ the weaker balance condition (together with a move algorithm that generates adequate sampling) is sufficient and necessary for an algorithm to provide a Boltzmann distribution of states. The balance condition simply requires that the Markov process generates a Boltzmann distribution of states after a given number of sampling cycles and then maintains the Boltzmann distribution invariant once it is achieved. Briefly, if the starting distribution of states of a system is designated as X and the Boltzmann distribution of states is denoted by X^* , the balance condition is simply that

$$A^n X = X^* \quad \text{and} \quad A X^* = X^*, \quad (4)$$

where A is the transition matrix of states and n is the number of sampling cycles applied for the system to converge to a Boltzmann-weighted ensemble of states. Some widely used MC methods^{18–23} that obey the balance condition but not the detailed balance condition have been demonstrated to generate proper Boltzmann distributions. As pointed out by Manousiouthakis and Deem,¹⁶ several of these algorithms are too complex to enable the detailed balance condition to be analytically proven, and in such cases empirical evidence has been sufficient for most researchers. We provide similar empirical evidence in this paper that demonstrates that the TIGER2 algorithm generates and maintains a Boltzmann distribution to within statistical error, thus demonstrating consistency with the balance condition for a set of distinctly different types of molecular systems.

In TIGER2, stage 4 of the cycle (i.e., replica comparison, selection, and global reassignment to the temperature levels) consists of two substeps: (A) one state from among the set of (N_r-1) quenched states is randomly selected with a probability of $1/(N_r-1)$ as an exchange candidate, and this randomly selected state is then compared with the state sampled from the production run at the baseline temperature. The success of this exchange is determined by the standard Metropolis sampling criterion,¹⁷ with probability

$$\min\{1, \exp[-(E_i - E_B)/k_B T_B]\}, \quad (5)$$

where E_i is the energy of the randomly selected quenched

state and E_B is the energy of the baseline state, and (B) the nonselected replicas (including the replica not accepted by the Metropolis MC comparison) are then reassigned to higher temperature levels according to their potential energies; i.e., a higher potential energy state is assigned to a higher temperature level. Operationally, this reassignment method enables a replica that is trapped in a high-energy local state to undergo sampling at a higher temperature, thereby increasing the rate at which it is able to cross the energy barriers that confine it in its local energy state. Effectively, this algorithm results in a Metropolis MC-like method that is able to use elevated-temperature MD simulation for move generation in a manner that satisfies the balance condition. Alternatively, this same process could be followed using Metropolis MC for the production portion of the sampling cycle performed at each temperature level (i.e., stage 2) combined with MD simulation to perform the quenching and heating parts of the cycle.

The quenching and the accompanied thermal equilibration phase in TIGER2 are accomplished over approximately a 1–2 ps time interval. This time period is found to be long enough for intramolecular and intermolecular vibrational relaxations to take place (~ 100 fs) (Ref. 24) and sufficiently close to the 0.5–1.0 ps hydrogen-bond relaxation time and the 2–5 ps rotational relaxation time^{25–27} for the restructuring of small solvent molecules such as TIP3P water.

C. Computation setup

Three distinctly different model systems were used to evaluate the performance of TIGER2 and to demonstrate its ability to provide a close approximation to Boltzmann sampling for each type of system: (1) *n*-butane in vacuum, (2) alanine dipeptide in explicit water represented by the TIP3P model,²⁸ and (3) two small peptides with secondary structure in implicit water represented by the Generalized Born model with simple switching function (GBSW) model.^{29,30} All simulations were carried out in the canonical ensemble (NVT) with the CHARMM program.^{31,32} The PAR22 (Ref. 32) force field was used for model (1), whereas the CMAP force field³³ was applied for models (2) and (3). The Nosé–Hoover thermostat^{34–36} was used to control temperature. The SHAKE algorithm³⁷ was used to constrain all bond lengths containing hydrogen, and the leapfrog algorithm¹⁰ was used with a time step of 2 fs to integrate the equations of motion in the dynamics simulations.

1. *n*-butane in vacuum

This system is composed of an *n*-butane molecule in vacuum. For this simple system, a Metropolis MC simulation was used to sample the conformation distribution for comparison with TIGER2. The Metropolis MC simulation, which contained 5×10^6 samples, was performed at a temperature of 180 K to randomly sample the torsional angles and single atom displacements (with a relative sampling weight of 2:1). A 50 ns TIGER2 test using four replicas was conducted at temperatures exponentially spaced from 180 to 600 K. The

TIGER2 cycle contained 1.0 ps heating and thermal equilibration to the replica temperature (T_m), 1.0 ps sampling at the replica temperature (T_m), and 1.2 ps quenching and thermal equilibration to the baseline temperature (T_B).

2. Alanine dipeptide in explicit water

The Langevin piston method³⁸ was incorporated in these simulations (zero mass of the pressure piston was used throughout the calculation), with a relatively high value of 25 ps⁻¹ for the friction coefficient being used for the heating and quenching phases to slow down the pressure piston response and a smaller value of 10 ps⁻¹ was used for the dynamic sampling. The system was simulated with cubic periodic boundary conditions and the particle-mesh Ewald method³⁹ was used to calculate electrostatic interactions.

The simulated system contained an alanine dipeptide (Ace-Ala-Nme) and 544 TIP3P (Ref. 28) water molecules in a cubic box with side length of 25.6 Å. All simulations started from the same conformation of the alanine dipeptide positioned in the α_R region of the Ramachandran plot.⁴⁰ Two 16 ns trajectories were generated using the REMD and TIGER2 algorithms, respectively. The REMD simulation was conducted with 24 replicas, with the temperature levels distributed exponentially between 300 and 600 K. Exchanges between adjacent replicas were attempted every 1 ps. Only four parallel replicas at temperature levels of 300, 377, 476, and 600 K were used in the TIGER simulation. The trajectory of each replica contained 5000 sampling cycles. Each cycle consisted of 1.0 ps heating and thermal equilibration to the assigned temperature level, 1.0 ps dynamics sampling at that temperature, and 1.2 ps heating and thermal equilibration to T_B , in sequential order, yielding a total cycle time of 3.2 ps and a total simulation time of 16 ns.

3. (AAQAA)₃ and chignolin peptides in implicit water

Peptide folding in solution has been the focus of many recent computational studies. The folding behaviors were studied in this work for (1) (AAQAA)₃ [(Ala-Ala-Gln-Ala-Ala)₃], which is an alanine-based synthetic model peptide containing 15 amino acid residues, and (2) chignolin, which is composed of 10 residues with a sequence of Gly-Tyr-Asp-Pro-Glu-Thr-Gly-Trp-Gly.

In these simulations, both peptides were capped with the N-terminal acetyl (Ace) and C-terminal methylamino (Nme) capping groups. The solvent effect was implicitly represented by the generalized Born model⁴¹ based on an integral equation approach and a simple smooth switching function (GBSW).²⁹ Default GBSW parameters were used with a 0.6 Å smoothing length along with 50 Lebedev angular integration points and 24 radial integration points up to 20 Å for each atom.^{29,30} The nonpolar solvation energy was estimated from the solvent-exposed surface area using a phenomenological surface tension coefficient of 0.005 kcal mol⁻¹ Å². An extended conformation of the peptide was used as the starting structure in both REMD and TIGER2 simulations. The REMD simulations were conducted with ten replicas at temperature levels distributed exponentially between 270 and 550 K for (AAQAA)₃ and with

FIG. 1. Probability density distributions for the central dihedral angle of butane at 180 K obtained from MC (dashed) and TIGER2 (solid) simulations. For the two distributions, the converged T/G ratios are 6.71 and 6.58, respectively. Insert: An all-atom representation of n -butane.

eight replicas at temperature levels distributed exponentially between 300 and 550 K for chignolin. Exchanges between adjacent replicas were attempted every 1.0 ps. Only three replicas at temperature levels of 270, 385, and 550 K were used in the TIGER2 simulation for (AAQAA)₃, with an exponential temperature distribution, similar to an REMD simulation. Each TIGER2 cycle consisted of 1.0 ps heating and thermal equilibration to the assigned temperature level, 2.0 ps dynamics sampling at that temperature, and 2.0 ps quenching and thermal equilibration to T_B in sequential order, yielding a total cycle time of 5.0 ps. The TIGER2 simulation for chignolin was conducted with four replicas at temperature levels of 300, 367, 449, and 550 K, with the same heating, dynamics, and quenching times as for (AAQAA)₃. A 40 ns trajectory was produced for each peptide in the REMD and TIGER simulations.

III. RESULTS AND DISCUSSION

For all TIGER2 simulations, the stated simulation time represents the total time of the simulation including the heating and quenching components of the cycle, with sampling statistics collected from the baseline replica.

A. Butane in vacuum

Figure 1 displays the probability distributions for rotation about the C₂–C₃ bond (insert) calculated with the Metropolis MC and TIGER2 methods at the baseline temperature of 180 K. The ratios of the area in the trans region (T) centered at 180° to the area of gauche regions (G) centered at 120° and 240° for the two methods are 6.71 and 6.58, respectively. These ratios can be also converted into free energy differences between the trans and gauche states ($\Delta G_{T/G}$) by

$$\Delta G_{T/G} = -RT \ln[P(T)/P(G)], \quad (6)$$

where $P(T)/P(G)$ represents the ratio of the probabilities of the butane molecule being in a trans versus a gauche state during the simulations, which we will subsequently refer to as the “ T/G ratio.” This expression results in $\Delta G_{T/G}$ values

FIG. 2. Conformation distributions for the alanine dipeptide in TIP3P water at sampling times of $t=0.06$, 0.62 , and 16 ns for REMD (upper) and TIGER2 (lower) simulations.

of -0.678 and -0.671 kcal/mol, respectively, with a $\Delta\Delta G$ value between the MC and TIGER2 simulations thus being only -0.007 kcal/mol. Plots of the T/G ratio versus simulation time and comparative analyses of the energy distributions obtained by each sampling method are presented in the supplemental information (see Fig. S.1, Fig. S.4, and Table S.1).⁴²

B. Alanine dipeptide with explicit solvation

For the alanine dipeptide, the main features of the backbone dihedral angle ϕ/ψ distribution are three regions corresponding to conformations designated $C7_{eq}$, $C5$, and $C7_{ax}$ with local energy minima centered around $-77^\circ/66^\circ$, $-159^\circ/165^\circ$, and $78^\circ/-56^\circ$, respectively, as predicted with the CMAP force field in the gas phase.³³ An additional three very shallow local minima designated β_2 , α_R , and α_L and located near $-138^\circ/23^\circ$, $-60^\circ/-45^\circ$, and $64^\circ/35^\circ$, respectively, are found in quantum mechanical calculations but do not present distinct minima with CMAP.³³

Both of the alanine dipeptide simulations using REMD and TIGER2 were started in the α_R region. The ϕ/ψ distributions that were accumulated at 300 K over sampling times of $t=0.06$, 0.62 , and 16 ns are shown in Fig. 2. Note that the ϕ/ψ coordinates of the six conformers shift somewhat in aqueous solution from their gas-phase values,³³ with the largest changes in the $C7_{eq}$ and $C7_{ax}$ conformers. The TIGER2 algorithm demonstrates a substantial improvement in conformational sampling compared to the REMD method. It is observed that by 0.06 ns, the $C7_{eq}/C5$, β_2/α_R , and α_L regions were all sampled in the TIGER2 simulation, whereas only

FIG. 3. Evolution of the population ratios of the $C7_{eq}/C5$ (solid), α_R/β_2 (dashed), α_L (dashed dotted), and $C7_{ax}$ (dotted) regions calculated from (a) REMD and (b) TIGER2 simulations.

the β_2/α_R regions were sampled in the REMD simulation. The time needed to begin to sample the $C7_{ax}$ region exceeded 0.62 ns in both simulations.

Based on the 16 ns trajectories obtained from the REMD and TIGER2 simulations, we calculated the population ratios of conformers located in four regions corresponding to $C7_{eq}/C5$, α_R/β_2 , α_L , and $C7_{ax}$ in order to quantitatively contrast the development of sampling obtained with the two methods. The boundaries between these regions of the ϕ/ψ map are indicated in Fig. 5 of Ref. 15. The population fractions of the different regions of the Ramachandran plot as a function of the simulation time obtained from both methods are displayed in Fig. 3. Since all simulations started from a similar conformation located in the α_R region, the initial α_R/β_2 fraction is 1.0 , with this value then decreasing as sampling transitioned to other regions. Since REMD requires time for replica diffusion over the temperature phase space, a longer amount of sampling was needed for the stability of the α_R/β_2 and $C7_{eq}/C5$ ratios [Fig. 3(a)]. However, the TIGER2 method [Fig. 3(b)] led to an earlier plateau in the α_R/β_2 and $C7_{eq}/C5$ ratios. Based on the last 4 ns trajectory of each simulation, we calculated the population ratios of conformers located in the $C7_{eq}/C5$, α_R/β_2 , α_L , and $C7_{ax}$ regions. The mean values of these population ratios are given in Table I, with each simulation method providing a similar set of ratios for each designated conformational region. Comparative analyses of the energy distributions obtained by each sampling method are presented in the supplemental information (see Fig. S.5 and Table S.2).⁴²

C. (AAQAA)₃ with implicit solvation

The model peptide (AAQAA)₃ is known to fold into a stable α -helix in aqueous solution with an α -helical content of about 50% at 274 K, as measured by circular dichroism spectropolarimetry⁴³ and NMR (Ref. 44) methods. Computer simulations have been carried out to study the folding

TABLE I. Population ratios averaged over the last 4 ns of each trajectory and their associated standard deviations (SD) for the $C7_{eq}/C5$, α_R/β_2 , α_L , and $C7_{ax}$ regions from the REMD and TIGER2 simulations.

Method	$C7_{eq}/C5$	SD	α_R/β_2	SD	α_L	SD	$C7_{ax}$	SD
REMD	0.460	0.002	0.461	0.003	0.060	0.002	0.019	6.0×10^{-4}
TIGER2	0.473	0.003	0.453	0.004	0.050	0.002	0.024	8.8×10^{-4}

FIG. 4. (Color) Superposition of a representative TIGER2 folded structure (blue) on the reference folded structure from REMD simulation (red). The all-atom RMSD between the two structures is 1.08 Å. The plot was created using the VMD program (Ref. 49).

mechanism and the stability of the peptide in water.^{45,46} Recently, this model system has been extensively used as a benchmark in the development of new simulation methods^{47–49} and refined force field parameters.³⁰ We therefore selected this peptide to test the performance of TIGER2 in comparison with a REMD simulation.

Both the REMD and TIGER2 simulations were started from the same fully extended structure of (AAQAA)₃. In the REMD simulation, using ten replicas, the acceptance ratio for neighboring-replica exchanges ranged from 0.14 to 0.20, falling in the targeted range for a conventional REMD simulation,^{1–5} whereas in the TIGER2 simulation, using only three replicas, the acceptance ratio between replicas was about 0.15. The peptide folded into a stable α -helix in both simulations. Figure 4 shows the superposition of the minimum energy structures obtained from both the REMD and TIGER2 simulations. The main chain and all-atom root mean-square deviation (RMSD) between the two structures are 0.268 and 1.08 Å, respectively. All backbone atoms are superimposed very well, but small distortions are observed at the N-terminal region and the side chain of Gln3 and Gln8 residues. Using the minimum energy REMD structure as a reference, the RMSD for backbone C $_{\alpha}$ atoms (C $_{\alpha}$ RMSD) and the difference of the potential energy (ΔE) were calculated for all structures obtained from the two simulations. Figure 5 displays the correlation of C $_{\alpha}$ RMSD with ΔE from both simulations at 270 K. The folded structures (with RMSD ≤ 1.0 Å) have lower average energies, with complete folding of the peptide occurring at 5.9 ns for the REMD simulation (top panel) and at 13 ns for the TIGER2 simulation (bottom panel). Although in this case REMD was found to fold the peptide faster than the TIGER2 simulation, it must be emphasized that the TIGER2 simulation was conducted with only three replicas compared to the ten replicas that were used for the REMD simulation.

FIG. 5. Relative potential energy (ΔE) vs C $_{\alpha}$ RMSD at various stages during 40 ns REMD (top) and TIGER2 (bottom) simulations of (AAQAA)₃. ΔE and RMSD were calculated with respect to the reference helical structure from the REMD simulation. Snapshots were taken every 1.0 ps from the baseline temperature (270 K) ensembles.

The accuracy of TIGER2 was further compared with the REMD results based on the fractional helicity of each residue of the peptide at 270 K, averaged over the last 20 ns (i.e., from 20 to 40 ns) of the trajectories. These are plotted in Fig. 6 along with the experimentally measured values.⁴⁴ It is observed that both the TIGER2 and REMD simulations resulted in very similar residue helicities at each residue, with both methods predicting higher values than the experimental measurements. The deviation of the simulation from the experiment may be caused by small imperfections in the force field. The average residue helicities calculated by REMD and TIGER2 are 0.62 and 0.60, respectively. Plots of the temporal evolution of the secondary structure of the peptide at 270 K for the REMD and TIGER2 simulations and comparative analyses of the energy distributions obtained by each sampling method are provided in the supplementary information (see Fig. S.2, Fig. S.6, and Table S.3).⁴²

FIG. 6. Residue helicity at 270 K obtained from experiment (squares) (Ref. 41) and REMD (triangles) and TIGER2 (circles) simulations. The residue helicities were calculated using snapshots from 20 to 40 ns for each simulation.

FIG. 7. Relative potential energy (ΔE) vs C_α RMSD from the native structure for REMD (top) and TIGER2 (bottom) simulations of chignolin. For each simulation, the last plot was created based on the last 25 ns of the trajectory. Snapshots were taken every 1.0 ps from the baseline temperature (270 K) ensembles.

D. Chignolin

Chignolin is a relatively recently designed peptide that forms a stable β -hairpin in water.⁵⁰ The experimentally determined β -hairpin fraction is about 60% at 300 K.⁵⁰ Although it is composed of only ten residues, it meets the definition of a protein, folding into a unique structure and having a cooperative thermal transition between its unfolded and folded states. The folding properties of the peptide have recently been studied by computer simulation methods.^{51,52} In our study, the TIGER2 algorithm was further tested by comparing simulation results with the corresponding REMD results and the experimental observations for this peptide model.

Both REMD and TIGER2 simulations were started from the same fully extended structure of chignolin and run for 40 ns. In the REMD simulation using eight replicas, the acceptance ratios for neighboring-replica exchanges ranged from 0.16 to 0.20, whereas in the TIGER2 simulation using four replicas, the acceptance ratio for global exchanges between replicas was about 0.22. The peptide folded into a stable β -hairpin in both simulations. The RMSD for backbone C_α atoms (C_α RMSD) and the difference in potential energy (ΔE) were calculated relative to the experimental NMR structure of chignolin (the first entry of PDB structure 1UA0) and are displayed in Fig. 7. Figure 7 shows that complete folding of the peptide was sampled within 8.2 ns in the REMD simulation (top panel) and within 2.3 ns in the TIGER2 simulation (bottom panel). It is also noted that the REMD simulation results show that the peptide was temporarily “stuck” in an intermediate folded state during the first 8.2 ns of the simulation, with this sampling problem being overcome after 15 ns. As indicated in the lower panel, over-sampling of this temporarily stabilized state was avoided in the TIGER2 simulation. A further analysis of the trajectories indicates that the fraction of folded conformations (C_α RMSD < 1.0 Å) is 0.492 and 0.488 for REMD and TIGER2 simulations, respectively, both of which are close to the experimentally measured value of 0.6.⁵⁰ Plots of the temporal

FIG. 8. (Color) Superposition of the structures with the lowest potential energy obtained from REMD (yellow) and TIGER2 (blue) simulations on the native structure (red) of chignolin peptide. The backbone-atom RMSD from the native structure is 0.548 Å for the REMD structure and 0.595 Å for the TIGER2 structure. The plot was created by the VMD program (Ref. 49).

evolution of the secondary structure of the peptide at 300 K for the REMD and TIGER2 simulations and comparative analyses of the energy distributions obtained by each sampling method are provided in the supplementary information (see Fig. S.3, Fig. S.7, and Table S.4).⁴²

The structure with the lowest potential energy was selected from each simulation. The superposition of these calculated lowest-energy structures on the native structure is shown in Fig. 8, which shows excellent superposition not only of the backbone atoms but also for the side-chain atoms of all three structures (REMD, TIGER2, and native). The aromatic rings of Tyr2 and Trp9 contact each other closely in the calculated structures, similar to the experimentally determined structure. The backbone-atom RMSD from the native structure obtained from the REMD (yellow) and TIGER2 (blue) structures are 0.548 and 0.595 Å, respectively. The aromatic rings of Tyr2 and Trp9 superimpose well between the two calculated structures, but both slightly deviate from the aromatic rings in the native structure. The RMSD values between the REMD and TIGER2 structures were 0.375 and 1.264 Å for the main chain and all-atom comparisons, respectively.

A further test of the structures generated by the TIGER2 simulation was performed by comparing the solvent accessible surface area (SASA) of residues between the calculated structures and the native structure (the first entry of PDB structure 1UA0). SASA provides insight into structural properties such as hydrophobic packing or potential modulating effects arising from the peptide's local environment. It is proportional to the solvent-exposed fraction of the side chains (SCEF). Generally, the residues in a peptide can be divided into two classes based on whether the degree of exposure of the side chains is greater or less than 0.3. This threshold value was chosen based on the modulation of pH dependence properties of side chains.^{53–55} Figure 9 plots the results of SCEF calculated from the lowest potential energy REMD and TIGER2 structures and the native structure and

FIG. 9. Solvent-exposed fractions of side chains calculated from the structures with the lowest potential energy obtained from REMD (triangles) and TIGER2 (circles) simulations and the native structure (squares) of the chignolin peptide.

shows that most of the SCEF values of the calculated structures closely follow the values of the native structure except at Glu5, Thr8, and Gly10 in the TIGER2 structure and Asp3 in the REMD structure. Furthermore, all SCEF values are greater than 0.3, indicating that all residues in the chignolin peptide are exposed to solvent.

E. Replica number requirements with TIGER2 compared to conventional REMD

In a conventional REMD simulation, the potential energy difference between neighboring replicas is due to a combination of both thermal energy and the higher probability that the higher temperature replica will be in a higher-energy conformational state. This net difference in energy must be kept sufficiently low in order to obtain a desired level of replica exchange, which is controlled by the number of replicas that are used to span a given range of temperature in the simulation. The TIGER2 method substantially alleviates this problem by quenching all replicas down to the baseline temperature before comparisons are made, removing the differences in thermal energy between replicas while retaining their energy differences that are related to conformation. This results in a condition where the potential energy distributions in the quenched replicas overlap one another to a much greater extent than before quenching, thus enabling a given range of temperature to be spanned with a substantially reduced number of replicas. It should also be realized, however, that the amount of phase space that will be explored in a TIGER2 simulation over a given amount of simulated time will be directly proportional to the number of replicas used. Thus the number of replicas to be used for a given system should be selected as a balance between the number of processors available for use and the time available to perform the simulation.

F. Boltzmann-weighted sampling with TIGER2

As shown for the four example systems presented above, the TIGER2 algorithm is able to provide sampling that closely approximates a Boltzmann-weighted distribution of states and is thus consistent with the balance condition for this set of molecular systems. While we have not yet found a system for which this algorithm does not appear to appropriately satisfy the balance condition, this does not mean that this algorithm will provide Boltzmann sampling for all types of molecular systems. At this time, we therefore recommend

that this algorithm be first validated for a desired type of molecular system by comparison with another accepted sampling method prior to full-scale implementation. Once validated, however, this method should prove useful to provide substantial savings in computational resources for applications where a large series of simulations is required for the same general type of molecular system or where the size of a given system is extended to the point where it is considered too large for the conventional REMD method.

IV. CONCLUSIONS

The TIGER algorithm¹⁵ has been reinvestigated and improved in this work, with the new, empirically motivated algorithm being named TIGER2. As demonstrated by its ability to closely match the conventional Metropolis MC or REMD results for each of the four systems evaluated (single butane in vacuum, alanine dipeptide with explicit solvation, and (AAQAA)₃ and chignolin with implicit solvation), the refined TIGER2 scheme is able to generate Boltzmann-like statistics without the limiting conditions that are required in the previous TIGER algorithm. These results demonstrate that the TIGER2 algorithm satisfies the balance conditions for sampling these types of systems but with substantially fewer replicas than are required by REMD. The TIGER2 method also effectively overcomes the replica-diffusion problem inherent in the REMD method by using global replica exchange, thus providing improved sampling efficiency. In addition, and potentially most importantly, this algorithm enables computational resources to be utilized much more efficiently by allowing the number of replicas used in a simulation to be greatly reduced compared to REMD for a given size system and temperature range to be spanned. This provides users with the freedom to select the number of replicas that they will use in a simulation based on the computing resources at hand, as opposed to being constrained by the size of the system to be modeled. This feature is in distinct contrast to an REMD simulation where the minimum number of replicas required to simulate a large molecular system is determined by the system size and may exceed the number of central processing units available to a user at a given facility. Because this is an empirical sampling method, it is recommended that it be validated for a given type of molecular system prior to widespread application. Further studies are currently underway to evaluate the use of this algorithm for other types of molecular systems and to explore the influence of factors such as temperature range, temperature spacing, and replica number on the sampling distributions provided by the TIGER2 algorithm.

ACKNOWLEDGMENTS

This work was supported by “RESBIO—The National Resource for Polymeric Biomaterials” funded under NIH Grant No. P41 EB001046-01A1 and the Engineering Research Centers Program of the National Science Foundation through the Center for Advanced Engineering Fibers and Films (CAEFF) funded by NSF Award No. EEC-9731680. Partial support was also provided by NIH under Grant Nos. R01 GM074511-01A1 and R01 EB006163-0131.

- ¹Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999).
- ²A. E. García and K. Y. Sanbonmatsu, *Proteins* **42**, 345 (2001).
- ³R. H. Zhou, B. J. Berne, and R. Germain, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14931 (2001).
- ⁴P. Liu, B. Kim, R. A. Friesner, and B. J. Berne, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13749 (2005).
- ⁵M. Andrec, A. K. Felts, E. Gallicchio, and R. M. Levy, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6801 (2005).
- ⁶P. Liu, X. H. Huang, R. H. Zhou, and B. J. Berne, *J. Phys. Chem. B* **110**, 19018 (2006).
- ⁷A. Baumketner and J. E. Shea, *J. Mol. Biol.* **362**, 567 (2006).
- ⁸A. Rahman and F. H. Stillinger, *J. Chem. Phys.* **55**, 3336 (1971).
- ⁹F. H. Stillinger and T. A. Weber, *Phys. Rev. B* **31**, 5262 (1985).
- ¹⁰M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).
- ¹¹R. Car and M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985).
- ¹²M. Matsumoto, S. Saito, and I. Ohmine, *Nature (London)* **416**, 409 (2002).
- ¹³H. Fukunishi, O. Watanabe, and S. Takada, *J. Chem. Phys.* **116**, 9058 (2002).
- ¹⁴E. Cinlar, *Introduction to Stochastic Processes* (Prentice-Hall, Englewood Cliffs, NJ, 1975).
- ¹⁵X. F. Li, C. P. O'Brien, G. Collier, N. A. Vellore, F. Wang, R. A. Latour, D. A. Bruce, and S. J. Stuart, *J. Chem. Phys.* **127**, 164116 (2007).
- ¹⁶V. I. Manousiouthakis and M. W. Deem, *J. Chem. Phys.* **110**, 2753 (1999).
- ¹⁷N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- ¹⁸W. H. Wong and F. Liang, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 14220 (1997).
- ¹⁹A. Z. Panagiotopoulos, *Mol. Phys.* **61**, 813 (1987).
- ²⁰D. Frenkel and B. Smit, *Understanding Molecular Simulations* (Academic, San Diego, 1996), p. 50.
- ²¹E. Marinari, G. Parisi, and J. Riuz-Lorenzo, in *Spin Glasses and Random Fields*, in Directions in Condensed Matter Physics Vol. 12, edited by A. Young (World Scientific, Singapore, 1998).
- ²²G. Thorleifsson and M. Falcioni, *Comput. Phys. Commun.* **109**, 161 (1998).
- ²³P. G. Bolhuis, C. Dellago, and D. Chandler, *Faraday Discuss.* **110**, 421 (1998).
- ²⁴A. Nitzan, *Nature (London)* **402**, 472 (1999).
- ²⁵C. P. Lawrence and J. L. Skinner, *J. Chem. Phys.* **118**, 264 (2003).
- ²⁶D. Laage and J. T. Hynes, *Science* **311**, 832 (2006).
- ²⁷K. Ohta and K. Tominaga, *Bull. Chem. Soc. Jpn.* **78**, 1581 (2005).
- ²⁸W. L. Jorgensen, *J. Phys. Chem.* **87**, 5304 (1983).
- ²⁹W. Im, S. Lee Michael, and C. L. Brooks, *J. Comput. Chem.* **24**, 1691 (2003).
- ³⁰J. H. Chen, W. Im, and C. L. Brooks III, *J. Am. Chem. Soc.* **128**, 3728 (2006).
- ³¹B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).
- ³²A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wierkiewicz-Kuczera, D. Yin, and M. Karplus, *J. Phys. Chem. B* **102**, 3586 (1998).
- ³³A. D. MacKerell, Jr., M. Feig, and C. L. Brooks III, *J. Comput. Chem.* **25**, 1400 (2004).
- ³⁴H. C. Andersen, *J. Chem. Phys.* **72**, 2384 (1980).
- ³⁵S. Nosé, *Mol. Phys.* **52**, 255 (1984).
- ³⁶W. G. Hoover, *Phys. Rev. A* **31**, 1695 (1985).
- ³⁷W. F. van Gunsteren and H. J. C. Berendsen, *J. Mol. Phys.* **34**, 1311 (1977).
- ³⁸S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, *J. Chem. Phys.* **103**, 4613 (1995).
- ³⁹T. Darden, D. York, and L. Pedersen, *J. Chem. Phys.* **98**, 10089 (1993).
- ⁴⁰G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, *J. Mol. Biol.* **7**, 95 (1963).
- ⁴¹B. N. Dominy and C. L. Brooks, *J. Phys. Chem.* **103**, 3765 (1999).
- ⁴²See EPAPS Document No. E-JCPSA6-130-016919 for supplementary data for the simulations of butane and (AAQAA)₃ and chignolin peptides. For more information on EPAPS, see <http://www.aip.org/pubservs/epaps.html>.
- ⁴³J. M. Scholtz, E. J. York, J. M. Stewart, and R. L. Baldwin, *J. Am. Chem. Soc.* **113**, 5102 (1991).
- ⁴⁴W. Shalongo, L. Dugad, and E. Stellwagen, *J. Am. Chem. Soc.* **116**, 8288 (1994).
- ⁴⁵W. A. Shirley and C. L. Brooks III, *Proteins* **28**, 59 (1997).
- ⁴⁶P. Ferrara, J. Apostolakis, and A. Caflisch, *J. Phys. Chem. B* **104**, 5000 (2000).
- ⁴⁷S. S. Sung and X. W. Wu, *Proteins* **25**, 202 (1996).
- ⁴⁸S. A. Hassan and E. L. Mehler, *Int. J. Quantum Chem.* **83**, 193 (2001).
- ⁴⁹Y. Liu and D. L. Beveridge, *Proteins* **46**, 128 (2002).
- ⁵⁰S. Honda, K. Yamasaki, Y. Sawada, and H. Morii, *Structure* **12**, 1507 (2004).
- ⁵¹D. Satoh, K. Shimizu, S. Nakamura, and T. Terada, *FEBS Lett.* **580**, 3422 (2006).
- ⁵²S. Kannan and M. Zacharias, *Proteins* **66**, 697 (2007).
- ⁵³E. L. Mehler, *J. Phys. Chem.* **100**, 16006 (1996).
- ⁵⁴E. L. Mehler and F. A. Guarnieri, *Biophys. J.* **77**, 3 (1999).
- ⁵⁵X. F. Li, S. A. Hassan, and E. L. Mehler, *Proteins* **60**, 464 (2005).