Communication: Simulated tempering with fast on-the-fly weight determination

Phuong H. Nguyen, Yuko Okamoto, and Philippe Derreumaux

Communication: Simulated tempering with fast on-the-fly weight determination

Phuong H. Nguyen,[1,a)] Yuko Okamoto,[2,b)] and Philippe Derreumaux[1,3,c)]

[1]*Laboratoire de Biochimie Théorique, UPR 9080, CNRS Université Denis Diderot, Sorbonne Paris Cité IBPC, 13 rue Pierre et Marie Curie, 75005 Paris, France*
[2]*Department of Physics, Nagoya University, Nagoya, Aichi 464-8602, Japan*
[3]*Institut Universitaire de France, 103 Bvd Saint-Germain, 75005 Paris, France*

We propose an efficient method to enhance sampling in computer simulations by combining the simulated tempering algorithm with a fast on-the-fly weight determination scheme. The weights are self-updated via a trapezoid rule during the simulated tempering simulation. With our proposed scheme, simulated tempering requires neither prior trial simulations nor complicated update schemes. The advantage of our method over replica exchange molecular dynamics has been demonstrated with the study of the folding of the 20-residue alanine peptide and the aggregation of a trimer formed by the Alzheimer's peptide fragment Aβ_{16-22}. © 2013 American Institute of Physics. [http://dx.doi.org/10.1063/1.4792046]

Convergence of the configuration space sampling is a primary concern of any simulation study of complex systems. At physiological temperature, a canonical Boltzmann simulation tends to get trapped in deep minima indefinitely, rendering the simulation ineffective. To accelerate the sampling, there has been considerable progress in developing a new class of simulation methods, referred to as the generalized-ensemble algorithms.[1–4] One of widely used generalized-ensemble algorithms is simulated tempering (ST).[5,6] The basic idea is that by coupling low temperature simulations with high temperature ones, one hopes to transfer the improved sampling at the higher temperature to the lower temperature. Although comparative studies between ST and replica exchange (RE)[7,8] have concluded that ST gives a higher rate of exchanging between high and low temperature states as well as a higher rate of traversing the potential space,[9–11] RE is used more often than ST. RE is simple in theory and implementation, while ST requires the determination of *a priori* unknown weight parameters to ensure a uniform random walk in temperature space and this is non-trivial and very tedious for complex systems. It is, therefore, desirable to develop efficient methods to determine accurate ST weight parameters. A common strategy is to perform short trial simulations to estimate weights from average potential energies[6,12–14] or via multiple-histogram reweighting.[9,15,16] However, both approaches bear the risk that these weights are of insufficient accuracy, resulting in poor or non-optimal ST performances. As a remedy, several schemes have been suggested to refine those initial rough weights during the subsequent ST production run.[11,17,18] The proposed procedures are, however, still complicated and computationally expensive.

In this Communication we propose a simple, yet practical scheme to perform a ST simulation without a prior iteration of trial simulations that guarantees random walk in temperature space. The simulation starts with zero weight parameters, which are then updated on-the-fly as the simulation progresses. The scheme can be very easily coded and automated, and therefore a wide range of applications is foreseen.

In the ST simulation, temperature itself becomes a dynamical variable which could take discrete values T_m ($T_1 < T_2 < \cdots < T_M$). Probability distribution of a state at temperature T_m and potential energy E is given by the following generalized canonical distribution:

$$W_{\text{ST}}(E, \beta_m) = \exp(-\beta_m E + f_m), \quad (1)$$

where $\beta_m = 1/k_B T_m$ (k_B is the Boltzmann constant). If the weight parameters f_m are chosen as $f_m = -\ln\left(\int dE n(E)\exp(-E/k_B T_m)\right)$ ($n(E)$ is the density of states, hence, f_m are dimensionless Helmholtz free energy at T_m), then it follows immediately that the distribution of temperature is flat, i.e., a free random walk in temperature space is realized, which in turn induces a random walk in potential energy space and allows the system to escape from local energy minima. Once the ST weight parameters f_m are determined, the ST algorithm consists of repeating the following two steps:[5,6] (i) perform a canonical Monte Carlo (MC) or molecular dynamics (MD) simulation at constant T_m for a certain number of steps, and (ii) update the current temperature T_m to a new value T_n, while the configurations are fixed, according to the following Metropolis-like transition probability:

$$w(T_m, T_n) = \min\left(1, \exp(-[(\beta_n - \beta_m)E_m - (f_n - f_m)])\right),$$
$$(2)$$

where E_m is the potential energy at temperature T_m. If MD is employed in step (i), the momentum rescaling is necessary,[16] as in Replica-Exchange MD (REMD).[19]

As seen, ST requires the determination of the weight parameters f_m. By modifying the original recursion formulae[6,12,13] slightly, Park and Pande recently derived

---
a)Email: phuong.nguyen@ibpc.fr.
b)Email: okamoto@phys.nagoya-u.ac.jp.
c)Email: philippe.derreumaux@ibpc.fr.

a simple formula which yields good estimates to the weight parameters from preparatory simulations:[14]

$$f_{n+1} = f_n + (\beta_{n+1} - \beta_n)(\bar{E}_{n+1} + \bar{E}_n)/2, \quad (3)$$

where $\bar{E}_n$ is the average potential energy at temperature T_n. By setting $f_1 = 0$, all remaining weight parameters are determined from Eq. (3), assuming that averaged potential energies $\bar{E}_n$ at temperatures T_n are known. Here, we show that the ST weight parameters can be updated on-the-fly, directly using Eq. (3), as follows:

- Start the ST simulation at the lowest temperature T_1, accumulating potential energy $\bar{E}_1(t) = \frac{1}{t} \int_0^t d\tau E_1(\tau)$ and estimating $f_2(t) = (\beta_2 - \beta_1)\bar{E}_1(t)/2$. During this time, switching to temperature T_2 is repeatedly attempted.
- Once the trajectory at T_2 is sampled, accumulate potential energy $\bar{E}_2$, then update f_2 and f_3.
- Now, the ST simulation runs with T_1, T_2, and T_3 using weights f_2 and f_3. Once the trajectory at T_3 is sampled, calculate potential energy $\bar{E}_3$, then update f_3 and f_4.
- Continue this procedure as the simulation progresses and eventually all weight parameters are updated according to Eq. (3).

Technically, this procedure can be easily automated, coded, and is very less time-consuming. At the beginning, the weight parameters are continuously updated and therefore the detailed balance condition is continuously broken. As the simulation progresses, the weight parameters are self-updated and eventually converge to the true values of the system's dimensionless free energies. The initial stage of the ST trajectory, where the weight parameters are significantly changed, is excluded in the analysis. Theoretically, two questions arise. First, are the weights updated if the system is trapped at one temperature? If $\beta_m \approx \beta_n$, then $\bar{E}_m \approx \bar{E}_n$, and Eq. (2) (with Eq. (3)) can be approximately written as

$$w(T_m, T_n) = \min(1, \exp(-(\beta_n - \beta_m)(E_m - \bar{E}_m))). \quad (4)$$

As seen from Eq. (4), the fluctuation of the instantaneous energy E_m around the averaged value $\bar{E}_m$ drives the temperature transition. The system tends to change to a lower temperature T_n if $E_m < \bar{E}_m$, and to a higher temperature, otherwise. Thus, a random walk in temperature space can be always realized if the temperature intervals are appropriately chosen, depending on energy fluctuations. Second, is the distribution of temperatures really flat since Eq. (3) does not take into account higher orders of the cumulant expansion of the energy difference between two temperatures? In practice, Park showed that Eq. (3) is an excellent approximation for systems with many degrees of freedom,[20] and we show that inclusion of higher orders is not necessary for the three systems studied here with our on-the-fly weight updating. We also remark that we can use the original recursion formulae[6,12,13] instead of Eq. (3) in our on-the-fly scheme.

In order to understand the essential physics and to demonstrate the validity as well as the good performance of the new method, we first applied it to a 1D system whose exact weights and conformational distribution are known. The sys-
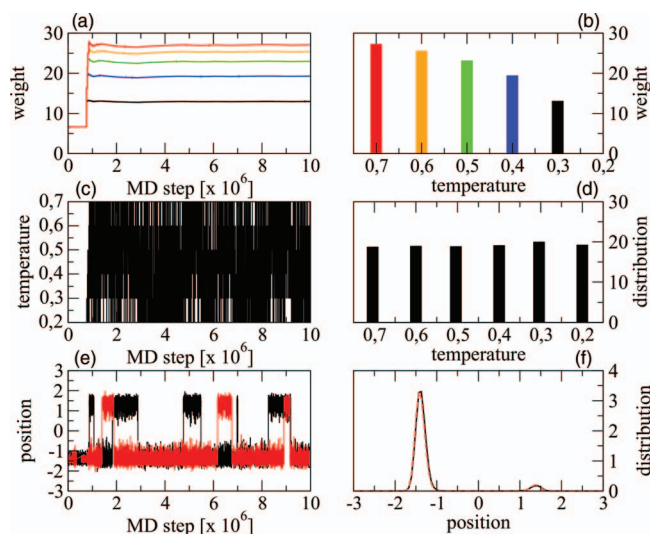

FIG. 1. ST simulation of the 1D system with on-the-fly weights determination. (a) Time evolution of the weights during ST. (b) Exact weights. (c) Change of temperatures during ST. (d) Temperature distribution. Time evolution (e) and distribution (f) of the positions of both particles at $T = 0.2$. (f) Results obtained from ST (black) and analytical calculation (red).

tem consists of 2 non-interacting particles, each has the potential energy of the form $U(x) = [(x+1)^2 - 1][(x-1)^2 - 0.9]$. It is an asymmetrical double well potential with two minima at $x \approx -1.5$ and ≈ 1.5. For simplicity, we set $k_B = 1$, and mass of particle $m = 1$. The ST simulation was performed using 6 temperatures with uniform interval, $T = 0.2, \ldots, 0.7$, for 10^7 steps. The MD time step was 0.002 in reduced unit. The initial 10^5 MD steps were discarded from the analysis. The Nosé-Hoover thermostat was used to maintain the system's temperature. The exact weights were calculated by performing numerical integration $f = -\ln[\int_{-\infty}^{\infty} dx \exp(-U(x)/k_B T)]$. As seen from Fig. 1(a), the simulation starting at $T_1 = 0.2$ spent about 7.5×10^5 steps to accumulate average energy, and estimate the parameter f_2. Once the first transition $T_1 \rightarrow T_2$ was successful, all weights were updated and reached quickly to the exact values (Fig. 1(b)). The system was able to explore all the temperatures (Fig. 1(c)) by performing an efficient random walk in temperature space as indicated by the uniform distribution of temperatures (Fig. 1(d)). At the lowest temperature $T_1 = 0.2$, Fig. 1(e) shows that particles were able to move between minima, resulting in a distribution of positions, identical to the analytical distribution (Fig. 1(f)). Note that at this temperature, the particles were trapped indefinitely in one of the two minima using a conventional MD (data not shown). To explain in detail the working mechanism of our on-the-fly weight updating scheme, Fig. 2 displays a short time window of the ST trajectory. Clearly, the trajectory starting at $T_1 = 0.2$ got trapped at $x \approx -1.5$ for up to 8.7×10^5 steps (Fig. 2(a)). During this time, the weight f_2 (corresponding to $T_2 = 0.3$) was continuously updated via the accumulated average potential energy at $T_1 = 0.2$, and the transitions $T_1 \rightarrow T_2$ were repeatedly attempted but all failed (Fig. 2(c)). At the step 7.65×10^5, a large positive energy fluctuation occurred (Fig. 2(b)) and this induced successfully the $T_1 \rightarrow T_2$ transition (Fig. 2(c)). Immediately, the average energy
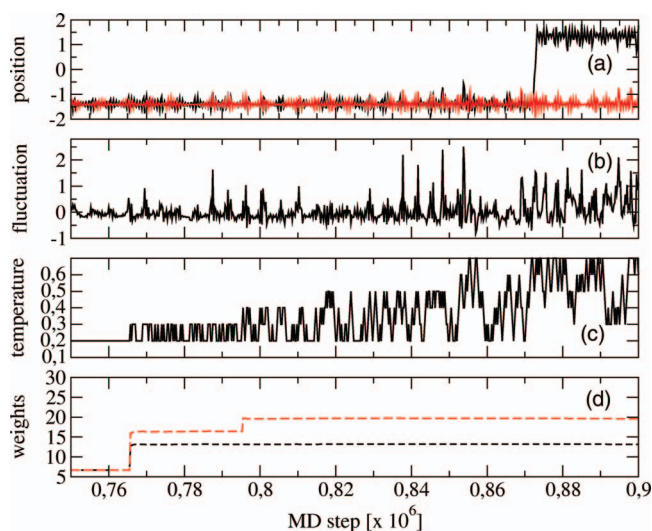
FIG. 2. Short time window of the ST trajectory of the 1D system. (a) Time evolution of the coordinates of two particles, and (b) fluctuation of the instantaneous energy around the average energy at $T_1 = 0.2$. (c) Change of temperatures during ST. (d) Time evolution of two selected weights corresponding to $T_2 = 0.3$ (black) and $T_3 = 0.4$ (red).
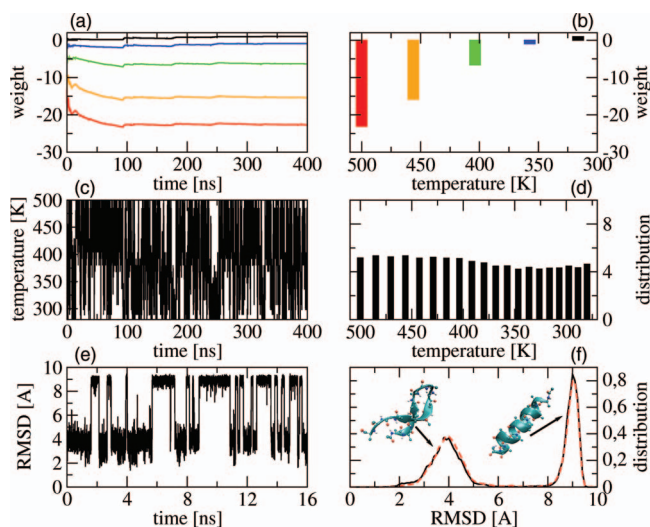

FIG. 3. ST simulation of Ala20 with on-the-fly weights determination. (a) Time evolution of the weights during ST. (b) Weights obtained from 100 ns REMD. (c) Change of temperatures during ST. (d) Temperature distribution. Time evolution (e) and distribution (f) of the RMSD at $T = 300$ K. (f) Results obtained from ST (black) and REMD (red) simulations. The structures corresponding to the two dominant peaks are shown.

$\bar{E}_2$ at T_2 was populated and consequently f_2 was updated from $f_2 \approx 6$ to $f_2 \approx 13$. Of course, f_3 ($\equiv f_4 \equiv \cdots \equiv f_7$) was also updated via the new values of f_2 and $\bar{E}_2$ (Fig. 2(d)). After this step, the simulation ran with three weight parameters f_1, f_2, and f_3. As seen, while the $T_1 \leftrightarrow T_2$ transitions took place more often because the parameter f_2 was getting better, the $T_2 \rightarrow T_3$ transitions were also attempted but all failed. At the step 7.95 $\times 10^5$, another large positive energy fluctuation occurred and the attempt $T_2 \rightarrow T_3$ was successful, and this resulted in the new update for f_3 (Fig. 2(d)). This update process was continued as the simulation progressed, and finally all weights converged and correct sampling was realized (Fig. 1). The analysis also shows that the weights were able to be self-updated even if the system was trapped in one state and instantaneous energy fluctuation drove the temperature transitions.

Next, we applied the ST simulations to two more realistic models, namely, a 20-residue alanine peptide (Ala20) and a trimer formed by the 7-residue Alzheimer's peptide fragment $A\beta_{16-22}$ ($(A\beta_{16-22})_3$) blocked by Ace and NH_2. Both systems were modeled using the coarse-grained Optimized Potential for Efficient peptide structure Prediction (OPEP) force field, found appropriate for studying the folding and self-assembly of non-amyloid and amyloid peptides.[21] Oligomers of amyloid peptides are of therapeutic interest since they are the most toxic species in human neurodegenerative diseases.[22] For both systems, we chose the T range of 280–500 K spaced exponentially into 20 values. Starting from fully extended conformations for both systems and well separated peptides for the trimer, the ST simulations were performed for 400 ns using an integration step of 1.5 fs and the Langevin thermostat.[23] Temperature transitions were attempted every 250 steps. All weight parameters were initially set to zero. Data were collected every 1000 steps. As a reference, we also performed a 100 ns REMD simulation[24] for each system using 20 replicas and the same other technical parameters. The

first 100 ns and 20 ns of the ST and RE simulations, respectively, were excluded in the analyses.

For the Ala20 system, the time evolution of five selected weight parameters is shown in Fig. 3(a). Starting from zero values, all weight parameters were updated to non-zero values within a few picoseconds. Then, the weights were continuously updated and reached plateau values after 100 ns, which are essentially identical with those obtained from 100 ns REMD with a difference of less than 1% (Fig. 3(b)). The system was able to explore all the temperatures (Fig. 3(c)) with a nearly uniform distribution of temperatures (Fig. 3(d)). The slight non-flatness, which can result from a relatively short simulation, can also be controlled by the selection of temperatures (optimization of the T values or slight increase in the number of temperatures). To characterize the sampling, Fig. 3(e) shows the time evolution of the root-mean-square deviation (RMSD) of the configurations at $T = 300$ K from the initial fully extended state. As seen, the peptide underwent many transitions between the β-hairpin (RMSD ≈ 4 Å) and α-helical (RMSD ≈ 9 Å) states. Figure 3(f) displays the distribution of the RMSDs at $T = 300$ K obtained using ST and REMD simulations. There is an excellent agreement between the two simulations.

For the $(A\beta_{16-22})_3$ system, as seen from Fig. 4(a), the weight parameters were quickly updated within 10 ns and then essentially converged after 100 ns to the values obtained from 100 ns REMD with a difference of less than 1% (Fig. 4(b)). The system performed a random walk and visited all temperatures (Fig. 4(c)). To characterize the aggregates, we calculated the nematic order parameter P_2,[25] and its distribution is shown in Fig. 4(d). As seen, the system mainly underwent the transition between two peptides forming an antiparallel β-sheet with the third peptide moving around the sheet ($P_2 \approx 0.3$), and three peptides forming a fully antiparallel β-sheet ($P_2 \approx 0.7$) (Fig. 4(d)). Such a conformational
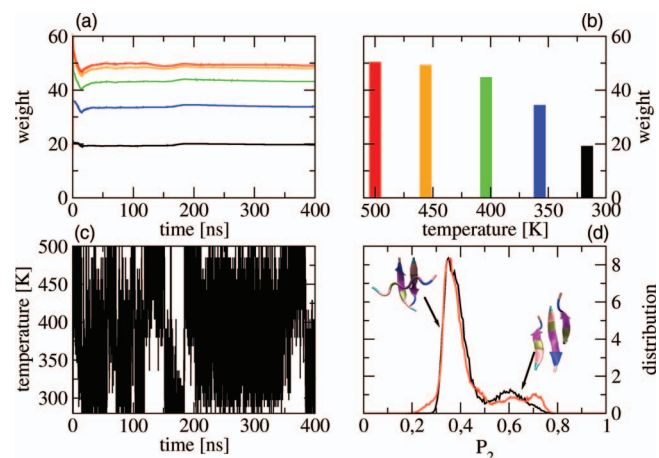
FIG. 4. ST simulation of the $(A\beta_{16-22})_3$ system with on-the-fly weights determination. (a) Time evolution of the weights during ST. (b) Weights obtained from 100 ns REMD. (c) Change of temperatures during ST. (d) Distribution of the order parameter P_2 at $T = 300$ K. (d) Results obtained from ST (black) and REMD (red). The structures corresponding to the two dominant peaks are shown.

ensemble has already been discussed using atomistic simulations in explicit solvent.[25] Again, the ST-derived configurational distribution matches that obtained from 100 ns REMD (Fig. 4(d)).

We proposed a scheme for the on-the-fly ST weight determination. Testing on two complex systems has shown that the method is not only simple but also accurate and very fast as indicated by the excellent agreement of results obtained from 400 ns ST simulation and 2000 ns (100 ns/replica × 20 replicas) REMD simulation. This efficient adaptive scheme coupled to the ST algorithm, which is five times faster than REMD in the two systems studied, should be very useful in the computer simulation community. Its applications to protein folding/aggregation problems in explicit solvent are under way.

[1] A. Mitsutake, Y. Sugita, and Y. Okamoto, Biopolymers **60**, 96 (2001).
[2] J. Kim, T. Keyes, and J. Straub, J. Chem. Phys. **132**, 224107 (2010).
[3] E. Rosta and G. Hummer, J. Chem. Phys. **132**, 034102 (2010).
[4] D. Sindhikara, D. Emerson, and A. Roitberg, J. Chem. Theory Comput. **6**, 2804 (2010).
[5] A. P. Lyubartsev, A. A. Martinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, J. Chem. Phys. **96**, 1776 (1992).
[6] E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992).
[7] C. J. Geyer, in *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface* (Interface Foundation, Fairfax Station, 1991), pp. 156–163.
[8] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
[9] A. Mitsutake and Y. Okamoto, Chem. Phys. Lett. **332**, 131 (2000).
[10] C. Zhang and J. Ma, J. Chem. Phys. **129**, 134112 (2008).
[11] S. Rauscher, S. Rauscher, C. Neale, and R. Pomes, J. Chem. Theory Comput. **5**, 2640 (2009).
[12] A. Irbäck and F. Potthast, J. Chem. Phys. **103**, 10298 (1995).
[13] U. Hansmann and Y. Okamoto, J. Comput. Chem. **18**, 920 (1997).
[14] S. Park and V. Pande, Phys. Rev. E **76**, 016703 (2007).
[15] A. Mitsutake and Y. Okamoto, J. Chem. Phys. **121**, 2491 (2004).
[16] A. Mitsutake and Y. Okamoto, J. Chem. Phys. **130**, 214105 (2009).
[17] R. Denschlag, M. Lingenheil, P. Tavan, and G. Mathias, J. Chem. Theory Comput. **5**, 2847 (2009).
[18] R. Chelli, J. Chem. Theory Comput. **6**, 1935 (2010).
[19] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[20] S. Park, Phys. Rev. E **77**, 016709 (2008).
[21] Y. Chebaro, S. Pasquali, and P. Derreumaux, J. Phys. Chem. B **116**, 8741 (2012).
[22] Y. Chebaro and P. Derreumaux, Proteins **75**, 442 (2009).
[23] Y. G. Spill, S. Pasquali, and P. Derreumaux, J. Chem. Theory. Comput. **7**, 1502 (2011).
[24] Y. Chebaro, X. Dong, R. Laghaei, P. Derreumaux, and N. Mousseau, J. Phys. Chem. B **113**, 267 (2009).
[25] P. H. Nguyen, M. S. Li, J. E. Straub, and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **104**, 111 (2007).