JCTC Journal of Chemical Theory and Computation

LETTER

pubs.acs.org/JCTC

Replica Temperatures for Uniform Exchange and Efficient Roundtrip Times in Explicit Solvent Parallel Tempering Simulations

Meher K. Prakash,* Alessandro Barducci, and Michele Parrinello

Department of Chemistry and Applied Biosciences, ETH Zurich USI Campus Via Giuseppe Buffi 13, CH 6900 Lugano, Switzerland

ABSTRACT: The efficiency of parallel tempering simulations is greatly influenced by the distribution of replica temperatures. In explicit solvent biomolecular simulations, where the total energy is dominated by the solvent, specific heat is usually assumed to be constant. From this, it follows that a geometric distribution of temperatures is optimal. We observe that for commonly used water models (TIP3P, SPC/E) under constant volume conditions and in the range of temperatures normally used, the specific heat is not a constant, consistent with experimental observations. Using this fact, we derive an improved temperature distribution which substantially reduces the round-trip times, especially when working with a small number of replicas.

1. INTRODUCTION

Parallel tempering (PT) is a popular choice for obtaining enhanced sampling in molecular simulations.[1−5] In standard PT, multiple NVT simulations are performed in parallel on the same system at different temperatures T_i, $i = 1-N_T$. At regular intervals, attempts are made to exchange the replica pairs $i \leftrightarrow i+1$, and the exchanges are accepted with a probability that conserves the detailed balance. One of the advantages of PT is that barriers and bottlenecks can be overcome in the high temperature replicas. In addition, these multiple replica simulations can be carried out in an embarassingly parallel mode, thus offsetting the added computational costs.

Recently, there have been several studies aimed at estimating[6−8] and improving the PT efficiencies either by the choice of replica temperature[8−19] or by enhancing the energy fluctuations of the replicas.[20] In a PT simulation, assuming that there are no bottlenecks, the individual replicas perform a random walk in the temperature space. The efficiency of the simulation is measured by the replica round-trip time (τ) across the temperatures, and this optimal value is obtained when the probability of exchange between any two neighboring replicas $P^{ex}_{i,i+1}$ is constant.[17]

$$P^{\mathrm{ex,\,opt}}_{i,\,i+1} = \text{const} \tag{1}$$

Thus, for improving the PT efficiency, one usually tries to achieve probabilities of exchange ($P^{ex}_{i,i+1}$) between neighboring replicas as uniform as possible. One of the assumptions in these theoretical analyses is a constant specific heat (C_V), which leads to the choice of geometrically distributed replica temperatures.[17,18]

However, when it comes to constant volume biomolecular simulations in explicit solvent, the assumption of a constant C_V is not valid. In fact, the specific heat in these systems is dominated by water, and it has been shown experimentally[21] and theoretically[22,23] that for water C_V is far from being a constant in the temperature interval normally used in biomolecular PT simulations. In fact, C_V decreases with temperature. Thus, in order to optimize the distribution of replica temperatures, the actual behavior of C_V under these conditions needs to be considered.

We have here recalculated the dependence of the average potential energy ($\overline{E}$) and C_V on T for the commonly used TIP3P[24] and SPC/E[25] models of water. It can be seen from Figure 1 that for both models the sublinear temperature dependence of $\overline{E}$ in the interval (280−650 K) of relevance in biomolecular PT is well approximated by the expression

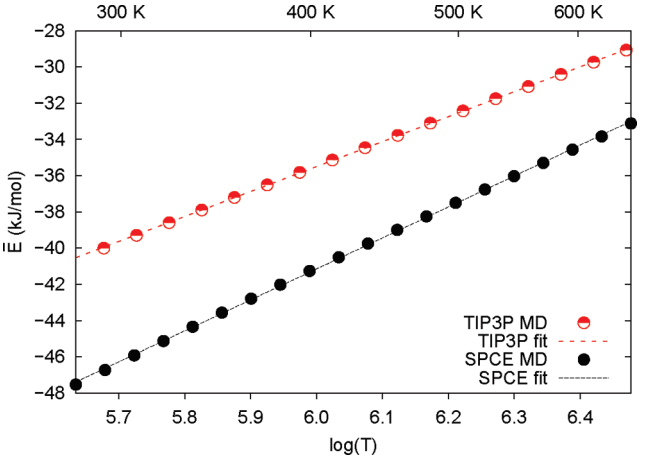

Figure 1. Average potential energy vs log(T) from NVT simulations of TIP3P and SPC/E waters over the temperature range 280−650 K. Starting with a box of 15 500 water molecules, box size was initially equilibrated at 300 K and 1 bar. NVT simulations at all temperatures were performed using a Nose−Hoover thermostat and this box size.

$$\overline{E} = a\log(T) + E_0 \tag{2}$$

where the constants (a, E_0) are 14 and −118 kJ/mol and 17 and −143 kJ/mol for TIP3P and SPC/E, respectively. From this dependence, and the relation $C_V = \partial\overline{E}/\partial T$, it follows, as shown in Figure 2, that $C_V = a/T$ in qualitative agreement with the experimental decrease of C_V. The origins of this surprising behavior are not entirely clear, but we take this as our empirical observation. On the basis of this, we reconsider here the issue of optimal choice of temperature distribution for explicit solvent calculation.

Published: June 06, 2011

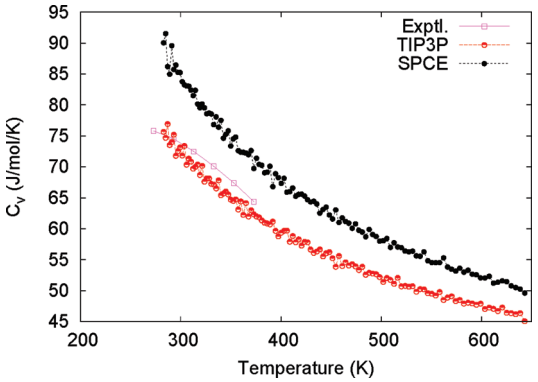

Figure 2. C_V dependence on temperature from experiments[21] and our MD simulations under NVT conditions. C_V from TIP3P and SPC/E water simulations was computed as a numerical derivative of the average energies with respect to T.

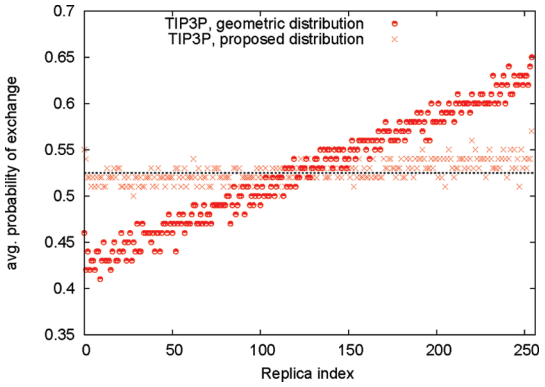

Figure 3. Probability of exchange between replicas i and $i+1$ ($P^{ex}_{i,i+1}$) for 2 ns simulations using geometric series distribution and the proposed distribution of temperatures in the range of 280−650 K. $P^{ex}_{i,i+1}$ are calculated by default by Gromacs. The solid line is a guide for the eye, set at $P^{ex} = 0.52$, and shows that $P^{ex}_{i,i+1}$ is uniform for all i's.

We use the following arguments. The exchange probability $P^{ex}_{i,i+1}$ in a replica exchange simulation is $\min\{1,\exp(-\Delta\beta_{i,i+1}\Delta E_{i+1,i})\}$, where $\Delta E_{i+1,i} = E_{i+1} - E_i$, $\Delta\beta_{i,i+1} = \beta_{i+1} - \beta_i$, and $\beta_i = 1/k_B T_i$. To obtain a uniform P^{ex}, as in eq 1, we consider a temperature distribution where the neighboring replicas satisfy the first-order condition with the mean energies[16,26]

$$\exp(-\Delta\beta_{i-1,i}\Delta\overline{E}_{i,i-1}) = \exp(-\Delta\beta_{i,i+1}\Delta\overline{E}_{i+1,i}) \quad (3a)$$

$$\text{i.e.,} \left(\frac{1}{T_{i-1}} - \frac{1}{T_i}\right)(\overline{E}_i - \overline{E}_{i-1}) = c \quad (3b)$$

where c is a constant related to the exchange probability. Using eq 2, one has

$$\frac{c}{a} = \left(\frac{1}{T_i} - \frac{1}{T_{i-1}}\right)(\log T_{i-1} - \log T_i)$$
$$= \left(\frac{1}{T_i} - \frac{1}{T_{i-1}}\right)\left[\log\left(1 + \frac{T_{i-1} - T_i}{T_i}\right)\right] \quad (4)$$

This equation is satisfied at all i if the temperature distribution follows the relationship

$$\frac{1}{T_i} = \frac{1}{T_{i-1}} - \sqrt{\frac{c/a}{T_{i-1}}} \quad (5)$$

The above equation is a simple analytical expression for generating the replica temperature distribution starting from an initial temperature T_1 and T_2 whose choice is determined by $P^{ex}_{1,2}$.

As a check of this derivation, we consider a hypothetical system with a constant C_V

$$\overline{E}(T) = bT + E_0 \quad (6)$$

Using the same arguments as before, we find

$$\frac{c}{b} = \left(\frac{1}{T_{i-1}} - \frac{1}{T_i}\right)(T_i - T_{i-1}) = \frac{(T_i/T_{i-1} - 1)^2}{T_i/T_{i-1}} \quad (7)$$

which is satisfied by the usual geometric distribution:

$$\frac{T_i}{T_{i-1}} = \text{const} \quad (8)$$

We have numerically checked the behavior of P^{ex} in PT molecular dynamics simulations using eqs 5 and 8 for 256 replicas in the temperature range 280−650 K. Tests were performed using Gromacs[27] on a system of 15 500 waters, which is the size required in protein folding and large scale conformational changes.[28] Figure 3 shows the uniform distribution of $P^{ex}_{i,i+1}$ that was obtained using the present T_i distribution; in contrast, a geometric distribution of T_i leads to an increasing $P^{ex}_{i,i+1}$ with a replica index.

Performing the simulations to obtain a large number of roundtrips required for an estimate of the converged average round-trip time (τ) as a function of N_T is an expensive proposition. Thus, we resort to a simplified model. We assumed in each replica the energy fluctuations are Gaussian distributed with mean $\overline{E}$ and width σ and obtain these parameters using eq 2. Assuming that the exchange attempts between i and $i+1$ are made at an interval larger than the energy correlation time, the energies E_i and E_{i+1} at every swap attempt will be uncorrelated and Gaussian distributed. Thus, we simulate swapping attempts comparing two random energies extracted from their respective distributions. In such a way, it is a simple exercise to evaluate τ. We checked the validity of this model by comparing the τ evaluated with those obtained in the explicit solvent simulations of a smaller system with 215 water molecules. The results are shown in Figure 4.

Using this model validated in a small system, we computed the τ with the simplified model for a box of 15 500 waters. As a function of N_T, one can identify three regions (Figure 5). In the low N_T region, the geometric distribution τ is dominated by the low $P^{ex}_{i,i+1}$ in the cold replicas. In the same regime, this shortcoming is avoided by the use of eq 5, which places colder replicas at closer intervals and results in a higher P^{ex} in the colder replicas (Figure 3). The computational gain obtained using eq 5 is particularly advantageous in the simulations of large systems, where one is constrained to work with lesser than optimal number of replicas.

In the mid N_T range, τ is low and N_T is optimal. Assuming a well-defined minimum in τ and for an alternative odd, even replica-pair exchange scheme, theoretical estimates of the optimal number of replicas ($N^*_T = 1 + (0.594(mC_V)^{1/2} - 1/2)\log(T_{N_T}/T_1)$) for simulating m water molecules have been discussed.[19] This estimate for our system with a 15 500 water molecules and using the C_V at room temperature is $N^*_T \approx 185$. The optimal τ in our model however is obtained at 120 replicas. In

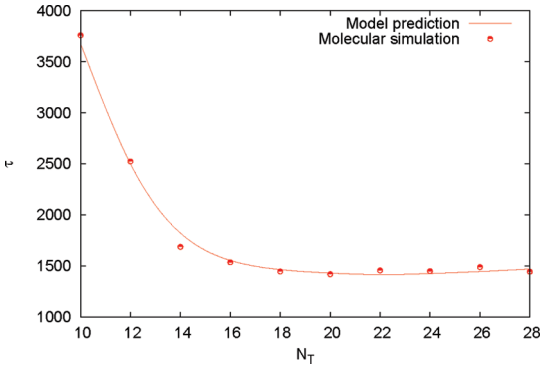

Figure 4. τ obtained from molecular simulations of a box of 215 waters is compared with that from the model. To obtain the τ corresponding to N_T, 100 ns simulations were performed on N_T replicas. Converged round-trip times were obtained from these simulations.

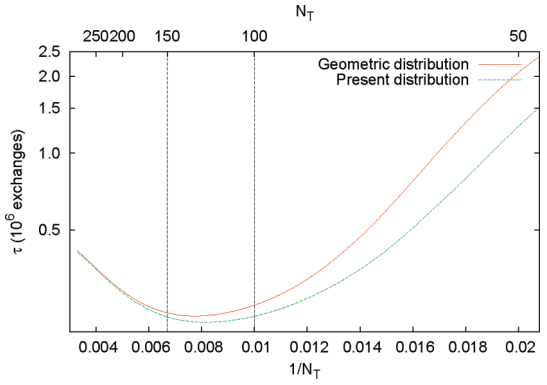

Figure 5. τ in units of millions of exchanges is shown in log scale for a box of 15 500 waters. τ was computed using the model validated in Figure 4. The three zones of decreasing, intermediate, and increasing τ with N_T are marked with dotted lines.

addition, τ has a weak dependence on N_T in this mid-N_T region. Because of this weak dependence, the simulations can be performed with a lower number of replicas, say 100, only marginally comprimising efficiency. In the high N_T region, which is suboptimal, τ increases with N_T, and the difference between the two distributions becomes small, as it should.

To conclude, eq 5 offers a simple and practical way of generating replica temperatures for obtaining a uniform exchange and improving the computational efficiency, especially when performing the calculations on a smaller number of replicas. The temperature distribution proposed in the present work is based on the realistic dependence of C_V in the context of explicit solvent NVT, thus becoming directly relevant for improving the efficiency of biomolecular PT simulations.

AUTHOR INFORMATION

Corresponding Author

*Phone: +41-58-666-4811. E-mail: meher@phys.chem.ethz.ch.

ACKNOWLEDGMENT

We thank Max Bonomi for the round-trip simulation code and a critical reading of the manuscript and Hagai Eshet for helpful discussions. We thank Swiss National Supercomputing Center (CSCS), Manno and High Performance Computing Group of ETH Zurich for computational resources. Financial support from the European Union grant (ERC-2009-AdG-247075) and Canton Ticino ("CSCS in rete") are gratefully acknowledged.

REFERENCES

(1) Geyer, C. J.; Thompson, A. *J. Am. Stat. Assoc.* **1995**, *90*, 909–920.
(2) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.
(3) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.
(4) Sugita, Y; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
(5) Earl, D. J.; Deem, M. W. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910–3916.
(6) Rosta, E.; Hummer, G. *J. Chem. Phys.* **2009**, *131*, 165102.
(7) Lingenheil, M.; Denschlag, R.; Mathias, G.; Tavan, P. *Chem. Phys. Lett.* **2009**, *478*, 80–84.
(8) Nymeyer, H. *J. Chem. Theory Comput.* **2008**, *4*, 626–636.
(9) Kofke, D. A. *J. Chem. Phys.* **2002**, *117*, 6911–6914.
(10) Predescu, C.; Predescu, M.; Ciabanu, C. *J. Chem. Phys.* **2004**, *120*, 4119–4128.
(11) Kofke, D. A. *J. Chem. Phys.* **2004**, *121*, 1167.
(12) Kone, A.; Kofke, D. A. *J. Chem. Phys.* **2005**, *122*, 206101.
(13) Rathore, N.; Chopra, M.; de Pablo, J. J. *J. Chem. Phys.* **2005**, *122*, 024111.
(14) Katzgraber, H. G.; Trebst, S.; Huse, D. A.; Troyer, M. *J. Stat. Mech.* **2006**, *6*, P03018.
(15) Trebst, S.; Troyer, M.; Hansmann, U. H. E *J. Chem. Phys.* **2006**, *124*, 174903.
(16) Garcia, A. E.; Herce, H.; Paschek, D. *Ann. Rep. Comp. Chem.* **2006**, *2*, 83–95.
(17) Nadler, W.; Hansmann, U. H. E. *Phys. Rev. E* **2007**, *75*, 026109.
(18) Nadler, W.; Hansmann, U. H. E. *J. Phys. Chem. B* **2008**, *112*, 10386–10387.
(19) Denschlag, R.; Lingenheil, M.; Tavan, P. *Chem. Phys. Lett.* **2009**, *473*, 193–195.
(20) Bonomi, M.; Parrinello, M. *Phys. Rev. Lett.* **2010**, *104*, 190601.
(21) Dass, N.; Varshneya, N. C. *J. Phys. Soc. Jpn.* **1968**, *25*, 1452–1454.
(22) Efimov, Y. Y.; Naberukhin, Y. I. *J. Struct. Chem.* **2000**, *41*, 433–439.
(23) Efimov, Y. Y.; Naberukhin, Y. I. *Spectrochim. Acta, Part A* **2005**, *42*, 1789–1794.
(24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(25) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
(26) Sanbonmatsu, K. Y.; Garcia, A. E. *Proteins* **2002**, *46*, 225.
(27) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
(28) Pitera, J. W.; Swope, W. C.; Abraham, F. F. *Biophys. J.* **2008**, *94*, 4837–4846.