University of New Orleans

ScholarWorks@UNO

Chemistry Faculty Publications   Department of Chemistry

2-2007

Replica exchange with dynamical scaling

Steven W. Rick
University of New Orleans, srick@uno.edu

Follow this and additional works at: https://scholarworks.uno.edu/chem_facpubs

Part of the Chemistry Commons

Replica exchange with dynamical scaling

Steven W. Rick[a)]

Department of Chemistry, University of New Orleans, New Orleans, Louisiana 70148

A replica exchange method is presented which requires fewer replicas and is designed to be used for large systems. In this method, dynamically scaled replicas are placed between conventional replicas at broadly spaced temperatures. The potential of the scaled replicas is linearly scaled by a dynamical variable which varies between 0 and 1. When the variable is near either end point the replica can undergo exchanges with one of its neighboring replicas. Two different versions of the method are presented for a model system of a small peptide in water. The scaled replica can replace many replicas and the method can be up to ten times more efficient than conventional replica exchange. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2431807]

I. INTRODUCTION

Molecular dynamics and Monte Carlo simulations often encounter energetic barriers leading to incomplete sampling of configuration space, or broken ergodicity. Replica exchange, or parallel tempering, is a powerful method for overcoming ergodicity problems (for a recent review, see Ref. 1). At a high enough temperature, the system is assumed to have enough energy to overcome barriers and the simulations will be ergodic. In replica exchange (RE), a high temperature replica is simulated along with the desired temperature and exchanges between replicas are accepted with the appropriate Boltzmann weighting, thereby introducing configurations from larger regions of phase space into the low temperature simulation. In order for exchanges to be accepted, there must be sufficient overlap between the energy distributions of the different replicas. A number of replicas can then be required to span the range of temperatures. As the number of degrees of freedom, f_s, increases, the number of replicas needed increases as approximately $f_s^{1/2}$.[2]

The system size dependence is made more severe by the requirement that all replicas cycle through the span of temperatures because implicitly only through reaching the highest temperature can the simulation achieve ergodicity.[3,4] For a larger number of replicas, more swaps need to be made to reach the highest temperature replica and the time required to cycle through all replicas increases. This dependence on system size place some practical limits on applications of RE. Some of the largest RE simulations are of the order of 10–20 amino acid peptides with 1000–3000 water molecules.[5,6] These are at least 5–10 times smaller then most biologically relevant proteins, indicating there is a need for improvement of RE for many applications.

Several methods have been proposed to improve the efficiency of RE. Two such methods involve the multicanonical algorithm (MUCA)[7,8] and simulated tempering (ST).[9,10] Both these methods involve a one-dimensional Monte Carlo walk in potential energy space, in the case of MUCA, and in temperature, in the case of ST. In methods combining MUCA and RE, replicas are in the multicanonical ensemble, which have broader energy distributions than canonical ensemble replicas.[11–13] Fewer replicas are then required [4 rather than 10 (Ref. 11) or 6 rather than 18 (Ref. 13)] relative to conventional RE. In methods combining ST and RE, each replica uses ST to sample a range of temperatures which has some overlap with the temperature spanned by neighboring replicas.[14,15] The MUCA and ST methods both require the determination of weight factors to ensure uniform sampling over the energy or temperature variable. The weights are not known initially and have to be determined through an iterative process using simulations preceding those for data collecting. Other methods, including smart walking, smart darting, and cool walking, work by increasing the acceptance probability of a swap with a high temperature replica, which may have no overlap in energy with the low temperature walker, by quenching the energy.[16–18] In order for the replica swaps between the low temperature replica and the quenched high temperature replica to satisfy detailed balance, the weights associated with the quenching must be determined.

An alternate approach called Hamiltonian RE scales all or part of the potential.[2,14,19–22] This may involve modification of the entire potential through a Tsallis transformation[19,20] or linear scaling of parts of the potential.[2,14,21,22] Replica exchanges are then made between a replica with the original potential and replicas with the modified potential at the same or a different temperature. The modified potential is presumed either to have smaller barriers, to otherwise promote sampling by stabilizing certain conformations, or to have a higher acceptance ratio for a given temperature difference. The method can lead to a reduction in the number of replicas [2 rather than 5 (Ref. 20) or 5 rather than 22 (Ref. 21)] with little or no added computational cost. On the other hand, Hamiltonian RE which just scales part of the potential is not so generalizable to systems which do not separate into obvious parts. In addition, because only one replica samples the original potential, high temperature data is not generated like it is in conventional RE or ST.

In this paper, a Hamiltonian RE approach, replica ex-

a)Electronic address: srick@uno.edu

change with dynamical scaling (REDS), is presented which can generate data over the full range of temperatures and requires less replicas than conventional RE.

II. METHODS

Exchanges between replicas i and j, with potentials E_i and E_j, are accepted with a probability

$$P = \min[1, e^{-\Delta_{ij}}], \tag{1}$$

where

$$\Delta_{ij} = \beta_i[E_i(r_i) - E_i(r_j)] + \beta_j[E_j(r_j) - E_j(r_i)], \tag{2}$$

where r_k denotes the configuration of replica k, β_k is $1/k_B T_k$, and T_k is the temperature of replica k. Consider two replicas, A and B, with the potential energy, $E(r)$, and temperatures T_A and T_B far enough apart so that there is no energy overlap between the two. A third replica, m, with a temperature, T_m, intermediate between T_A and T_B has the potential

$$E_m(r,\lambda) = [\lambda\beta_A + (1-\lambda)\beta_B]/\beta_m \, E(r) + E_{\text{bias}}^{(m)}(\lambda), \tag{3}$$

where λ is an additional variable of the system, constrained to be in the interval 0 to 1, and $E_{\text{bias}}^{(m)}(\lambda)$ is a biasing potential acting only on λ. Because swaps will be attempted with replicas A and B, the other replicas need to have a λ variable, which will be uncoupled to the system, and subject only to a biasing potential, or $E_k(r,\lambda) = E(r) + E_{\text{bias}}^{(k)}(\lambda)$ for $k=A$ and B. Swaps between replicas k and m will depend on

$$\Delta_{km} = E(r_m)[\lambda_m\beta_A + (1-\lambda_m)\beta_B - \beta_k] - E(r_k)[\lambda_k\beta_A$$
$$+ (1-\lambda_k)\beta_B - \beta_k] + \beta_k[E_{\text{bias}}^{(k)}(\lambda_k) - E_{\text{bias}}^{(k)}(\lambda_m)]$$
$$+ \beta_m[E_{\text{bias}}^{(m)}(\lambda_m) - E_{\text{bias}}^{(m)}(\lambda_k)]. \tag{4}$$

If λ_A and λ_m are about equal, so that the terms involving in E_{bias} will cancel, and both equal to 1, then Δ_{1m} will be 0. In this case, all swaps between replicas A and m will be accepted. If λ_B and λ_m both equal 0, then Δ_{2m} will be 0 and all swaps between replicas B and m will be accepted. Replica m can then exchange with both its neighboring replicas as long as λ samples evenly from 0 to 1. For the other replicas, λ_A and λ_B should be kept near 1 and 0, respectively. Replica exchanges will include exchanges of λ, as well as the positions of the atoms.

The variable λ is treated as a dynamical variable, given a mass and subject to equations of motion, just as in the λ-dynamics applications used for free energy calculations[23–26] and constant pH simulations[27–29] and similar to other extended Lagrangian methods, including constant pressure,[30] Car–Parinello,[31] fluctuating charge,[32] and Nosé–Hoover simulations.[33,34] To keep λ between 0 and 1, a new dynamical variable is used, $\lambda = \sin^2(\theta)$, and θ is propagated[29] with the equation of motion

$$M_\theta \ddot{\theta} = -\frac{\partial E_m}{\partial\theta} = -2\cos(\theta)\sin(\theta)\frac{\partial E_m}{\partial\lambda}$$
$$= -2\cos(\theta)\sin(\theta)$$
$$\times \left[\frac{\beta_A - \beta_B}{\beta_m}E(r) - \frac{\partial E_{\text{bias}}^{(m)}}{\partial\lambda} \right]. \tag{5}$$

The θ variables for the unscaled replicas 1 and 2 will just feel a force from $E_{\text{bias}}^{(k)}$.

In the REDS method, not only will the unscaled replicas generate data for the original potential, but the scaled replicas will as well. For the scaled replicas, the average of property A for a particular value of λ is

$$\langle A(\lambda = \lambda_n) \rangle$$
$$= \frac{\int dr d\lambda A(r)\delta(\lambda - \lambda_n)e^{-[\lambda_n\beta_A - (1-\lambda_n)\beta_B]E(r)}e^{-\beta_m E_{\text{bias}}^{(m)}(\lambda_n)}}{\int dr d\lambda e^{-[\lambda\beta_A - (1-\lambda)\beta_B]E(r)}e^{-\beta_m E_{\text{bias}}^{(m)}(\lambda)}}$$
$$= \frac{e^{-\beta_m E_{\text{bias}}^{(m)}(\lambda_n)}\int dr A(r)e^{-\beta_n E(r)}}{Z}, \tag{6}$$

where $\beta_n = [\lambda_n\beta_A - (1-\lambda_n)\beta_B]$. The probability distribution function for λ is given by

$$\langle P(1-\lambda_n) \rangle = \frac{\int dr d\lambda \delta(\lambda - \lambda_n)e^{-[\lambda_n\beta_A - (1-\lambda_n)\beta_B]E(r)}e^{-\beta_m E_{\text{bias}}^{(m)}(\lambda_n)}}{\int dr d\lambda e^{-[\lambda\beta_A - (1-\lambda)\beta_B]E(r)}e^{-\beta_m E_{\text{bias}}^{(m)}(\lambda)}}$$
$$= \frac{e^{-\beta_m E_{\text{bias}}^{(m)}(\lambda_n)}\int dr e^{-\beta_n E(r)}}{Z}. \tag{7}$$

Then

$$\langle A \rangle_n = \frac{\int dr A e^{-\beta_n E(r)}}{\int dr e^{-\beta_n E(r)}} = \frac{\langle A(\lambda = \lambda_n) \rangle}{\langle P(\lambda = \lambda_n) \rangle} \tag{8}$$

and the scaled replica gives averages with the original potential, $\langle A \rangle_n$ over the range of temperatures, T_n, from T_A to T_B. Any average calculated from Eq. (8) will not depend on the biasing potential, so accurate values of E_{bias} are not required. The biasing potential is only needed to help sample over the λ variable.

The connection between ensemble averages from the unscaled and scaled replicas provides a simple method for determining the biasing potential. The purpose of the biasing potential is to ensure that λ evenly samples the entire the entire region from 0 to 1. The biasing potential will cause λ to evenly sample range if the average force on λ (or θ) is zero. From Eq. (3)

$$\left\langle \frac{\partial E(\lambda = \lambda_n)}{\partial\lambda} \right\rangle = \frac{\beta_A - \beta_B}{\beta_m}\langle E \rangle_n - \frac{\partial E_{\text{bias}}^{(m)}(\lambda = \lambda_n)}{\partial\lambda} \tag{9}$$

and the derivative of $E_{\text{bias}}^{(m)}$ should cancel the average of the force from the potential energy so that $\langle \partial E/\partial\lambda \rangle$ equals zero for all λ. Standard simulations can find $\langle E \rangle$ at the endpoints ($\lambda=0$ and 1, corresponding to $T_n=T_A$ and T_B) as well as intermediate points. A good estimate of $E_{\text{bias}}^{(m)}$ can be made prior to the RE simulation, just from knowing the potential energy at a few temperatures. For the present applications, a cubic form for the biasing potential works well, $E_{\text{bias}}^{(m)}=A\lambda + B\lambda^2 + C\lambda^3$. The coefficients can be fitted to $\langle E \rangle$ calculated at T_A, T_m, and T_B. For replica A, $E_{\text{bias}}^{(A)}$ is chosen to keep λ_A near 1 and $E_{\text{bias}}^{(A)}(\lambda=1) = E_{\text{bias}}^{(m)}(\lambda=1)$ to minimize Δ_{1m} when λ_m is near 1. For replica B, $E_{\text{bias}}^{(B)}$ is chosen to keep λ_B near 0 and $E_{\text{bias}}^{(B)}(\lambda=0) = E_{\text{bias}}^{(m)}(\lambda=0)$. The set-up time for REDS is comparable to that of conventional RE, in which choosing the number and temperatures of the replicas has to be done, for

example by choosing them to get 20% acceptance ratios or through a more involved criteria.[3,4,35,36] It is also possible to find the parameters for the biasing potential while the simulation is running by periodically fitting to an average force, similar to what is done in self-guided molecular dynamics.[37]

Other scalings can be used as well. For example, for a solvent/solute system the solute could be unscaled so it is utilizes the full potential and the solvent scaled, so that the solvent parts, which can contain most of the degrees of freedom, do not contribute to the acceptance ratio in the $\lambda=0$ and 1 limits. This scaled potential is given by

$$E_m(r,\lambda) = E_{\text{solute}} + [\lambda\beta_A + (1-\lambda)\beta_B]/\beta_m\, E_{\text{solvent}}(r)$$
$$+ [\lambda\beta_A + (1-\lambda)\beta_B + \beta_m]/(2\beta_m)$$
$$\times E_{\text{solvent-solute}}(r) + E_{\text{bias}}^{(m)}(\lambda). \quad (10)$$

The solute-solvent interactions are scaled by an intermediate amount. This form, in the limit where λ is fixed at 1, is what was proposed by Liu et al.[21] Another possibility is to use the Tsallis scaling of the potential[38]

$$E_m(r) = q(\lambda)/\{\beta[q(\lambda) - 1]\}\ln\{1 - [1 - q(\lambda)]\beta E(r)\}, \quad (11)$$

where $q(\lambda)$ equals $1+\lambda\delta q$, taking the potential from the original potential ($q=1$) to some maximum value, $1+\delta q$. In these scaling, Eq. (8) is not valid and the expression for the force on λ is different. The biasing potential would have to be fit not to $\langle E \rangle$ but the λ derivative of the scaled potential. Values for the fitting could be found by running a single replica with the potential given by Eq. (10) with λ fixed at 0, 1/2, and 1.

While the scaling given by Eq. (3) suggests that the temperature is a variable, reminiscent of ST, the temperature is a constant, and molecular dynamics are run at a constant temperature. In addition, as has been pointed out elsewhere,[21] the partial scaling indicated by Eq. (10) does not mean that the various parts of the system are at different temperatures. The fact that the potential rather than the temperature is being changed also results in a simpler form for the biasing potential.

A. Constant pressure

For the isothermal-isobaric ensemble, the probability density is proportional to $\exp[-\beta(E+PV)]$, where P is the external pressure and V is volume. The REDS method in this ensemble scales the enthalpy, $H=E+PV$, according to

$$H_m(r,\lambda) = [\lambda\beta_A + (1-\lambda)\beta_B]/\beta_m[E(r) + PV] + E_{\text{bias}}^{(m)}(\lambda). \quad (12)$$

Both the energy and the pressure are scaled, so the pressure is not constant during the simulation but depends on the value of λ. Swaps between the scaled and an unscaled replica, at β_k and pressure P, will depend on

$$\Delta_{km} = [E(r_m) + PV_m][\lambda_m\beta_A + (1-\lambda_m)\beta_B - \beta_k] - [E(r_k)$$
$$+ PV_k][\lambda_k\beta_A + (1-\lambda_k)\beta_B - \beta_k] + \beta_k[E_{\text{bias}}^{(k)}(\lambda_k)$$
$$- E_{\text{bias}}^{(k)}(\lambda_m)] + \beta_m[E_{\text{bias}}^{(m)}(\lambda_m) - E_{\text{bias}}^{(m)}(\lambda_k)]. \quad (13)$$

where replica swaps exchange volumes as well coordinates.[39] The method is similar to the canonical version and exchanges with the neighboring replicas will be accepted with high probability when λ equals 0 or 1.

Averages of the property A for a particular value of λ in this ensemble is

$$\langle A(\lambda = \lambda_n) \rangle = \frac{e^{-\beta_m E_{\text{bias}}^{(m)}(\lambda_n)}\int dr dV d\lambda A(r)\delta(\lambda - \lambda_n)e^{-[\lambda\beta_A - (1-\lambda)\beta_B][E(r)+PV]}}{\int dr dV d\lambda e^{-[\lambda\beta_A - (1-\lambda)\beta_B][E(r)+PV]}e^{-\beta_m E_{\text{bias}}^{(m)}(\lambda)}} \quad (14)$$

and the probability distribution function for λ is

$$\langle P(\lambda = \lambda_n) \rangle = \frac{e^{-\beta_m E_{\text{bias}}^{(m)}(\lambda_n)}\int dr dV d\lambda\, \delta(\lambda - \lambda_n)e^{-[\lambda\beta_A - (1-\lambda)\beta_B][E(r)+PV]}}{\int dr dV d\lambda e^{-[\lambda\beta_A - (1-\lambda)\beta_B][E(r)+PV]}e^{-\beta_m E_{\text{bias}}^{(m)}(\lambda)}}. \quad (15)$$

The ratio of these two will give averages for the isothermal-isobaric ensemble over the temperature range T_A to T_B and a pressure P,

$$\langle A \rangle_n = \frac{\int dr dV A e^{-\beta_n[E(r)+PV]}}{\int dr dV e^{-\beta_n[E(r)+PV]}} = \frac{\langle A(\lambda = \lambda_n) \rangle}{\langle P(\lambda = \lambda_n) \rangle}. \quad (16)$$

By scaling the pressure while keeping the temperature constant, the scaled replicas retain the important features of canonical REDS: good acceptance ratios with neighboring replicas and the generation of ensemble averages over a range of temperatures. The scaling given by Eq. (12) is optimal for linking different constant pressure replicas, as can be demonstrated by looking at the λ dependence of the pressure. The instantaneous pressure, P_i, is the sum of an ideal gas, kinetic energy part and a virial part, which, at atmospheric pressure, have about the same magnitude and opposite signs. The instantaneous pressure for the scaled replica is given by

$$P_i = \frac{1}{3V}\sum m_i\dot{r}_i^2 - [\lambda\beta_A - (1-\lambda)\beta_B]/\beta_m\frac{\partial E}{\partial V} \quad (17)$$

from which it can be seen that the kinetic energy part is not scaled by λ (because the temperature is fixed at T_m) and the virial part is scaled. For a given value of λ, λ_n, the volume

will adjust so that on average the instantaneous pressure is equal to the external pressure, which is itself scaled by λ,

$$\langle P_i \rangle_n = \left\langle \frac{1}{3V} \sum m_i \dot{r}_i^2 \right\rangle_n - [\lambda_n \beta_A - (1$$

$$- \lambda_n)\beta_B]/\beta_m \left\langle \frac{\partial E}{\partial V} \right\rangle_n$$

$$= [\lambda_n \beta_A - (1 - \lambda_n)\beta_B]/\beta_m P. \quad (18)$$

Using the fact that the volume and the velocities are independent and $\langle 1/3 \sum m_i \dot{r}_i^2 \rangle_n = NkT_m$ gives

$$\left\langle \frac{1}{V} \right\rangle_n NkT_m - [\lambda_n \beta_A - (1 - \lambda_n)\beta_B]/\beta_m \left\langle \frac{\partial E}{\partial V} \right\rangle_n$$

$$= [\lambda_n \beta_A - (1 - \lambda_n)\beta_B]/\beta_m P \quad (19)$$

or

$$\left\langle \frac{1}{V} \right\rangle_n NkT_m /\{[\lambda_n \beta_A - (1 - \lambda_n)\beta_B]/\beta_m\} - \left\langle \frac{\partial E}{\partial V} \right\rangle_n = P. \quad (20)$$

From

$$[\lambda_n \beta_A - (1 - \lambda_n)\beta_B]/\beta_m = T_m[\lambda_n/T_A + (1 - \lambda_n)/T_B]$$

$$= T_m/T_n, \quad (21)$$

where $T_n = 1/[\lambda_n/T_A + (1 - \lambda_n)/T_B]$, we get

$$\left\langle \frac{1}{V} \right\rangle_n NkT_m /(T_m/T_n) - \left\langle \frac{\partial E}{\partial V} \right\rangle_n = P \quad (22)$$

or

$$\left\langle \frac{1}{V} \right\rangle_n NkT_n - \left\langle \frac{\partial E}{\partial V} \right\rangle_n = P. \quad (23)$$

Equation (23) is identical to standard T, P, N simulations at a temperature T_n and a pressure P, so REDS will give the same volume on average with λ_n value as conventional T, P, N simulations at T_n. This also must be true from Eq. (16) for the volume or any equilibrium average.

The simulations described in this paper are all in the canonical ensemble, but, as demonstrated here, the REDS method will also work in the isothermal, isobaric ensemble. To our knowledge, this is the first time that a Hamiltonian RE approach has been described for constant pressure simulations. Prior Hamiltonian RE applications have all been constant volume.[2,14,19–22] Hamiltonian RE will change the virial contribution to the pressure and so the volume will be different from that of the original Hamiltonian. The REDS method gets around this by a suitable scaling of the pressure.

B. Simulation details

The alanine dipeptide, using the OPLS-AA/L potential,[40,41] with 512 TIP4P (Ref. 42) water molecules is used to test the method. The molecular dynamics simulations are done in the canonical (constant T, V, N) ensemble with Nosé–Hoover chains for thermostating,[43] SHAKE for constraining bond lengths and Ewald sums for the long-ranged electrostatic interactions.[44] A value of 2.0(kcal/mol)ps² was
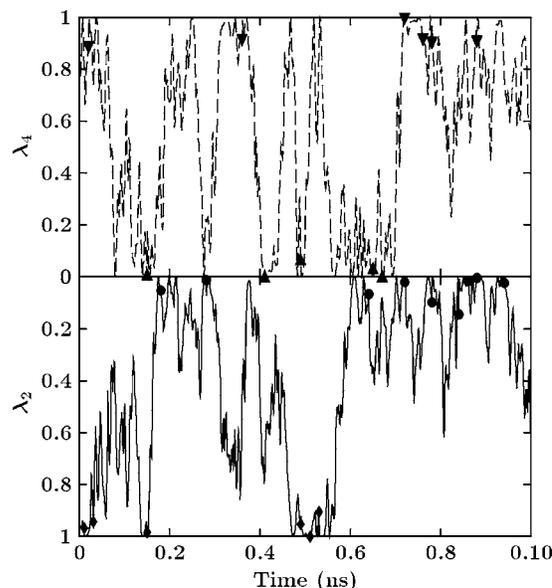

FIG. 1. The variable λ as a function of time for REDS with five replicas for the two scaled replicas, replica 4 (dashed line, top) and 2 (solid line, bottom). The symbols indicate successful replica exchanges (◆, 1↔2; ●, 2↔3; ▲, 3↔4; ▼, 4↔5).

used for M_θ. Three different RE methods are used, all spanning the same range of temperature. The first method is conventional RE with 22 replicas spread so that the acceptance ratio between adjacent replicas is about 20%. The second is REDS with five replicas: three unscaled replicas at 300, 420, and 600 K and two scaled replicas at 350 K (with $T_A = 300$ K and $T_B = 420$ K) and 494 K ($T_A = 600$ K and $T_B = 420$ K). For the second scaled replica, $T_A > T_B$ so that both scaled replicas can make exchanges with the $T = 420$ K replica with $\lambda = 0$. The third method is RE with partial dynamical scaling (REPDS), as defined by Eq. (10), with five replicas just like the REDS setup. For all three methods, replica exchanges are attempted between adjacent replicas once every 1 ps. With the REDS method, the coefficients for the biasing potentials are $1632 + 60(\lambda - 1)^2$, $1512\lambda + 131\lambda^2 - 11.3 \lambda^3$, $60 \lambda^2$, $-1601\lambda + 99\lambda^2 + 11.3 \lambda^3$, and $-1491 + 60(\lambda - 1)^2$ for replicas 1–5, respectively (all in units of kcal/mol). With the REPDS method, the biasing potentials are $1571.5 + 60(\lambda - 1)^2$, $1452\lambda + 129.5\lambda^2 - 10 \lambda^3$, $60 \lambda^2$, $-1536\lambda + 96\lambda^2 + 12 \lambda^3$, and $-1428 + 60(\lambda - 1)^2$ for replicas 1–5, respectively. For all three methods, simulations were repeated twice, starting with different initial conditions. The conventional RE simulations were simulated for 1 ns and the REDS and REPDS were simulated for 2 ns.

III. RESULTS

Figure 1 shows a short trajectory for λ_2 and λ_4 (the scaled replicas) from the REDS simulation. The trajectories are plotted so that they meet in the middle at $\lambda = 0$, where they both can make transitions with replica 3 at $T = 420$ K. Both undergo many transitions between 0 and 1 during this period, indicating that in a relatively short time the scaled replicas can make swaps with both adjacent replicas. The symbols indicate the points where successful replica swaps have occurred, which are all near, but not exactly at, 0 or 1.
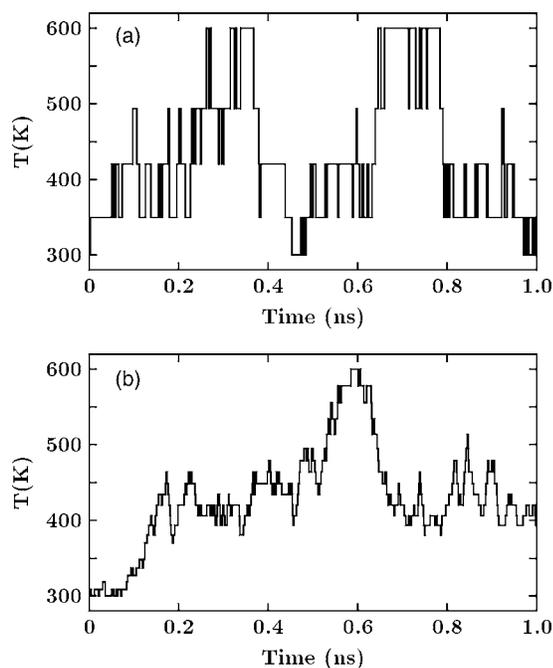
FIG. 2. The temperature of a given replica as a function of time for (a) REDS, with five replicas and (b) RE, with 22 replicas.

Transitions from 0 to 1 occur with a similar frequency for the REPDS method, but the acceptance ratios are less because the unscaled part of the potential will increase Δ_{ij}. For REDS, the acceptance ratio averaged 0.13 and for REPDS it averaged 0.09. The acceptance ratio of RE was 0.26.

The movement of a replica through each temperature is shown in Fig. 2 for REDS (*A*) and RE (*B*). The REDS replica moves much quicker from the lowest to the highest temperature and in 1 ns, cycles through all temperatures twice. The RE replica has to move through 22 temperatures and the process is much slower. The time for transitions across all the replicas can be quantified by calculating the time it takes to go from the lowest to higher temperatures (Fig. 3). The time required to go from the lowest to the highest tempera-
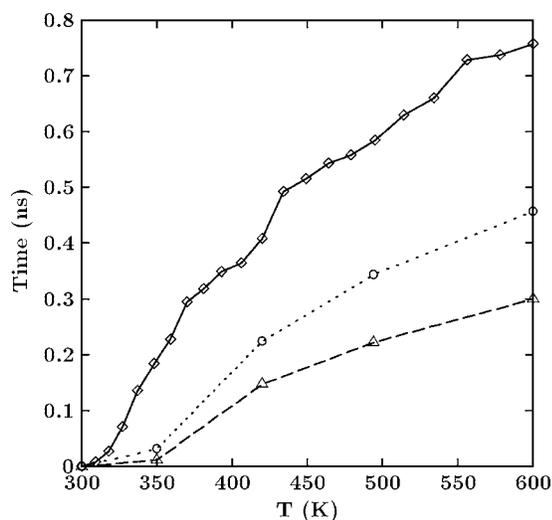
FIG. 3. The average time for a replica to move from the lowest temperature, 300 K, to a higher temperature for RE (solid line, ◇), REDS (dashed line, △), and REPDS (dotted lined, ○).
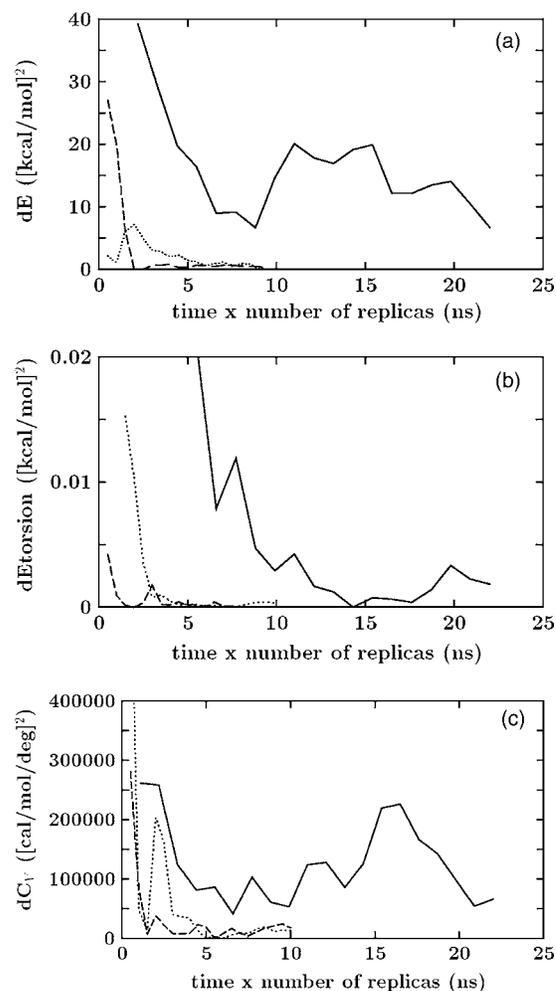
FIG. 4. The ergodic metric for the total potential energy (a), the torsional energy (b), and the constant volume heat capacity (c) for RE (solid line), REDS (dashed line) and REPDS (dotted line) at $T=298$ K as a function of total simulation time.

ture will depend on the acceptance ratio, the number of replicas, and the time scale for energy fluctuations. For REDS, the time scale for the fluctuations required to have successful exchanges with one neighboring replica then another can be seen in Fig. 1. RE will have its own time scale for fluctuations from low energy, where exchanges will the lower temperature replica are likely to occur, to high energy, where exchanges with the higher temperature replica are more likely to occur. For RE, the average time to move over the range of temperatures is 0.76 ± 0.1 ns. The time is over twice as fast for REDS (0.30 ± 0.08 ns). For REDS, the time is 0.45 ± 0.2 ns.

The efficiency of sampling can be measured using ergodic metrics, which give the rate that averages from independent simulations become equal.[45,46] One metric is defined as

$$dX(t) = [\langle X(t) \rangle_A - \langle X(t) \rangle_B], \qquad (24)$$

where $\langle X(t) \rangle$ is the average of a property X after a time t and A and B represent the two independent simulations. The metrics for the total energy, the torsional energy, and constant volume heat capacity, C_v, are shown in Fig. 4. The heat capacity is calculated through

$$C_V = \frac{1}{kT^2}(\langle E^2 \rangle - \langle E \rangle^2) + \frac{1}{2}f_s R, \qquad (25)$$

where E is the potential energy, f_s is the total number of degrees of freedom, and R is the ideal gas constant. The results are plotted versus the simulation time for each replica times the number of replicas for the various methods, since this represents the total CPU cost. Both REDS and REPDS the averages of the energies and C_V from the independent simulations approach each other much more rapidly than RE. Even with over twice the total simulation time, the metrics are still greater for RE. By a total simulation time of 5 ns for REDS and REPDS, which corresponds to simulating each of the five replicas for 1 ns, the metrics are much less than the metrics for RE at a total simulation time of 22 ns, which also corresponds to a simulation time of 1 ns for each replica. That means that comparing simulation time per replica, and not considering the extra time necessary to simulate more replicas, the dynamical scaling methods have a greater rate of self-averaging, because it takes less time to cycle through all temperatures.

Error estimates representing two standard deviations can be found from

$$\delta x = \frac{2}{\sqrt{N-1}} \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2}, \qquad (26)$$

where the data has been split up into N 0.1 ns intervals. The error estimates for the total energy, the torsional energy, and C_v are shown in Fig. 5 as a function of total simulation time. The data represent the average of the error bars from the two independent simulations. For the energy, the error estimates are considerably smaller for REDS and REPDS than for RE, consistent with quicker self-averaging. Even comparing the same amount of time per replica (1 ns) corresponding to 5 ns for REDS and REPDS and 22 ns for RE in Fig. 5, the error estimates are smaller. This again must be due to faster rate at which replicas cycle through all temperatures. The total energy and the heat capacity are largely determined by the solvent, which represents the majority of the degrees-of-freedom. The torsional energy only involves the peptide solute, including the ϕ, ψ backbone torsional angles. The torsional energy error bars are smallest for the REDS method. The REPDS method has error bars closer to the RE method, unlike the other properties shown in Fig. 5 for which they are essentially the same as REDS. The torsional energy represents the highest energy barriers of the system, so for this property it is likely to be most critical to reach the highest temperature replica, which is done quickest for the REDS method. For the heat capacity, the decrease in the error bars is not as pronounced, but there is still a noticeable reduction for REDS and REPDS over RE. Taken together with Fig. 4, both dE and δE are smaller for REDS and REPDS after 2 ns than RE is after the complete 22 ns. For the torsional energy, the ergodic measure and the error estimate show about the same improvement as the total energy for the REDS method. The REPDS method shows an improvement over RE for the torsional energy similar to that of the heat capacity. For C_V, dC_V is less for REDS and REPDS after 3 ns and for δC_V
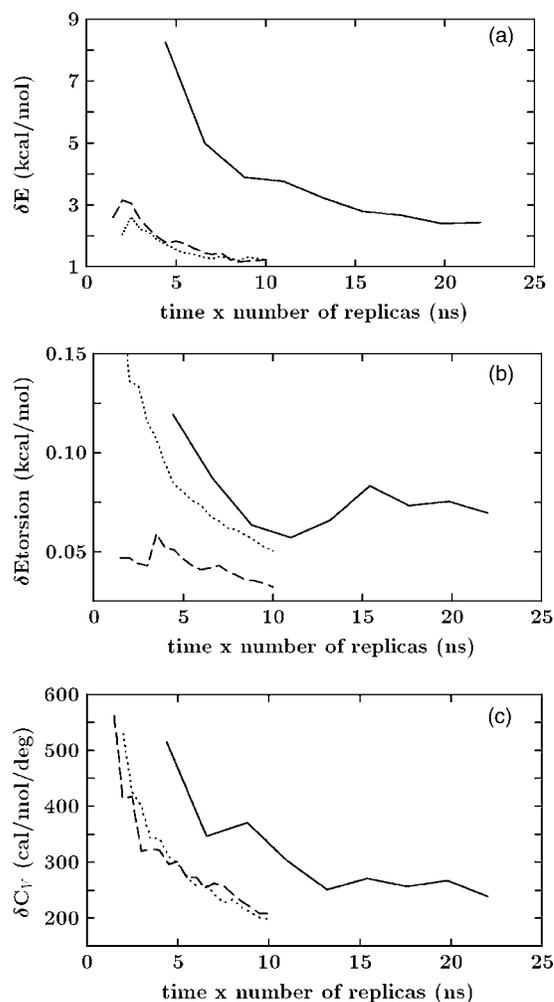

FIG. 5. Error estimates for the total energy (a), the torsional energy (b), and the constant volume heat capacity (c) for RE (solid line), REDS (dashed line) and REPDS (dotted line) at $T=298$ K as a function of total simulation time.

after 8 ns. So, depending on what property is looked at, REDS and REPDS is 3–10 times faster than RE.

One advantage of REDS over some other Hamiltonian RE methods is that REDS, like conventional RE, gives ensemble averages for other temperatures. This is true for the unscaled replicas (3 and 5 in this implementation) and also for the scaled replicas, which can each give ensemble averages over a large range of temperature from Eq. (8). The temperature dependences of E and C_V from the 22 replicas and from the two scaled REDS replicas are shown in Fig. 6. The solid line shows the data from replica 2 (from 300 to 420 K) and the dotted lines shows the data from replica 4 (from 420 to 600 K). The temperature range of the data determined from $1/kT = [\lambda\beta_A - (1-\lambda)\beta_B]$, with β_A and β_B being the inverse temperatures neighboring each scaled replica. The RE values use 2 ns of data for each replica (from the total of the two independent simulations) and the REDS curve uses 2 ns of data for both replicas as well. This means that the RE points represent about ten times more data than the REDS curves (22×2 ns versus 2×2 ns) For the energy, there is very close agreement between the two methods. The error bars for the energy from the RE are about the size of
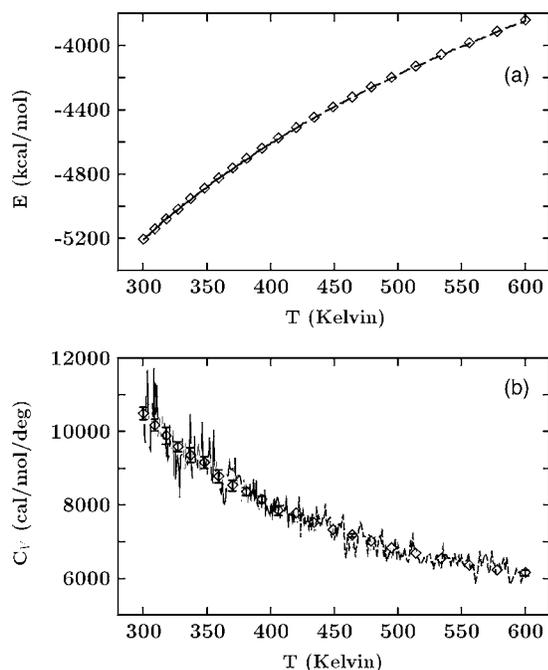
FIG. 6. Potential energy (a) and constant volume heat capacity (b) as a function of temperature from RE with 22 replicas ($\diamond$) and REDS from the two scaled replicas (solid and dashed lines).

the diamond in Fig. 6(a), averaging 1.5 kcal/mol for the 22 replicas. For the REDS replicas, the error bars are about three times greater than this, averaging 3.3 kcal/mol. This is consistent with the fact that they represent ten times less data as the error should decrease as the square root of the simulation time. The heat capacity error bars for RE are shown in Fig. 6(b) and for REDS the errors can be judged by the oscillations in the lines. Even though there is some noise the two REDS replicas give the same overall shape of the temperature dependence of C_V as the 22 replicas. The unscaled replicas in REDS (at 300, 420, and 600 K) will give smaller error estimates than the RE replicas, as was demonstrated for the 300 K data.

IV. CONCLUSIONS

In this paper, a replica exchange method is presented which combines conventional replicas at a set of temperatures (here 300, 420, and 600 K) with dynamically scaled replicas at temperatures in between the conventional replicas (350 and 494 K). In the dynamically scaled replicas, the potential is linearly scaled by a dynamical parameter λ which ranges from 0 to 1. When the parameter is near 0 or near 1, the scaled replica will have a good probability of a successful replica exchange with one or the other of its neighboring replicas. A scaled replica was shown to be able to replace ten conventional replicas. The method, replica exchange with dynamical scaling (REDS) and a variant, replica exchange with partial dynamical scaling (REPDS), increases the sampling efficiency by about a factor of 10 for some properties. This is more than simply the gain in time from having less replicas (22 versus 5, giving a factor of 4.4) and represents a combination of two factors: the reduction in the number of replicas and the increased speed at which the replicas cycle

through all temperatures. The scaled replicas can move from replica 1 (at 300 K) to replica 3 (at 420 K) or from replica 3 to replica 5 (at 600 K) much faster than conventional RE can go from replica 1 to replica 12 (420 K) or replica 12 to replica 22 (at 600 K). The overall time to go from the lowest temperature to the highest replica is 2.5 times faster for REDS than for RE and 1.7 times faster for REPDS (Fig. 3). This factor of about two times the factor of 4.4 for having less replicas gives the overall increase in efficiency.

The scaled replicas require a biasing potential to ensure that the λ variable evenly samples the range from 0 to 1. As discussed in the Introduction, other methods developed to increase the efficiency of RE (multicanonical,[11–13] simulated tempering,[14,15] smart darting,[17] smart walking[18]) also require some sort of biasing, or weight factor. An advantage of the REDS approach is that the biasing potential is relatively straightforward to determine. The biasing potential can be estimated from knowing the potential energy at a few different temperatures. This is much simpler than calculating the simulated tempering weights, which are the Helmholtz free energies at the different temperatures, or the multicanonical weight, which is the microcanonical entropy. In addition, the scaled replicas require no more computational expense than standard replicas and are easy to implement.

Another advantage of the method is that it does not require that the potential separate into different parts, like solvent/solute[21,22] or hydrophobic/hydrophilic.[2] The method can be applied to arbitrary systems, but it is flexible enough that the system can be separated into scaled and unscaled parts if that is advantageous [see Eq. (9) and (10)]. It can be used in both the canonical ensemble or isothermal isobaric ensemble. In the present study, done in the canonical ensemble, only one setup of scaled and unscaled replicas was examined, envisioning a need for precise ensemble averages at certain temperatures, which would have unscaled replicas, as well as an unscaled replica at the highest temperature to help with sampling. As demonstrated in Fig. 6 all replicas, scaled and unscaled, can give the correct ensemble averages, but the unscaled replicas give more precise averages. There are many other combinations of scaled and unscaled replicas, depending on the needs of the study. It is possible to have only one single scaled replica, for example. Different implementations of the model could be explored in future studies.

[1] D. J. Earl and M. W. Deem, Phys. Chem. Chem. Phys. **7**, 3910 (2005).
[2] H. Fukunishi, O. Watanabe, and S. Takada, J. Chem. Phys. **116**, 9058 (2002).
[3] C. Predescu, M. Predescu, and C. V. Ciobanu, J. Phys. Chem. B **109**, 4189 (2005).
[4] S. Trebst, M. Troyer, and U. H. E. Hansmann, J. Chem. Phys. **124**, 174903 (2006).
[5] R. Zhou, B. Berne, and R. Germain, Proc. Natl. Acad. Sci. U.S.A. **98**, 14931 (2001).

[6] D. Paschek, S. Gnanakaran, and A. E. Garcia, Proc. Natl. Acad. Sci. U.S.A. **102**, 6765 (2005).

[7] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).

[8] B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).

[9] A. P. Lyubartsev, A. A. Martinsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, J. Chem. Phys. **96**, 1776 (1992).

[10] E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992).

[11] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **329**, 261 (2000).

[12] F. Calvo and J. P. K. Doye, Phys. Rev. E **63**, 010902 (2000).

[13] R. Faller, Q. Yan, and J. J. de Pablo, J. Chem. Phys. **116**, 5419 (2002).

[14] M. K. Fenwick and F. A. Escobedo, J. Chem. Phys. **119**, 11998 (2003).

[15] A. Mitsutake and Y. Okamoto, J. Chem. Phys. **121**, 2491 (2004).

[16] R. Zhou and B. J. Berne, J. Chem. Phys. **107**, 9185 (1997).

[17] I. Andricioaei, J. E. Straub, and A. F. Voter, J. Chem. Phys. **114**, 6994 (2001).

[18] S. Brown and T. Head-Gordon, J. Comput. Chem. **24**, 68 (2003).

[19] T. W. Whitfield, L. Bu, and J. E. Straub, Physica A (Amsterdam) **305**, 157 (2002).

[20] S. Jang, S. Shin, and Y. Pak, Phys. Rev. Lett. **91**, 058305 (2003).

[21] P. Liu, B. Kim, R. A. Friesner, and B. J. Berne, Proc. Natl. Acad. Sci. U.S.A. **102**, 13749 (2005).

[22] P. Liu, X. Huang, R. Zhou, and B. J. Berne, J. Phys. Chem. B **110**, 19018 (2006).

[23] X. Kong and C. L. Brooks III, J. Chem. Phys. **105**, 2414 (1996).

[24] R. M. Lynden-Bell and J. C. Rasaiah, J. Chem. Phys. **107**, 1981 (1997).

[25] Z. Guo, C. L. Brooks III, and X. Kong, J. Phys. Chem. B **102**, 2032 (1998).

[26] N. F. A. van der Vegt and W. J. Briels, J. Chem. Phys. **109**, 7578 (1998).

[27] U. Börjesson and P. H. Hünenberger, J. Chem. Phys. **114**, 9706 (2001).

[28] M. S. Lee, F. R. Salsbury, and C. L. Brooks III, Proteins **56**, 738 (2004).

[29] J. Khandogin and C. L. Brooks III, Biophys. J. **89**, 141 (2005).

[30] H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).

[31] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[32] S. W. Rick, S. J. Stuart, and B. J. Berne, J. Chem. Phys. **101**, 6141 (1994).

[33] S. Nosé, Mol. Phys. **52**, 255 (1984).

[34] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).

[35] N. Rathore, M. Chopra, and J. J. de Pablo, J. Chem. Phys. **122**, 024111 (2005).

[36] A. Kone and D. A. Kofke, J. Chem. Phys. **122**, 206101 (2005).

[37] X. Wu and S. Wang, J. Chem. Phys. **110**, 9401 (1999).

[38] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).

[39] M. Doxastakis, V. G. Mavrantzas, and D. N. Theodorou, J. Chem. Phys. **115**, 11352 (2001).

[40] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).

[41] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, J. Phys. Chem. B **105**, 6474 (2001).

[42] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[43] G. Martyna, M. Klein, and M. Tuckerman, J. Chem. Phys. **97**, 2635 (1992).

[44] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).

[45] D. Thirumalai and R. Mountion, J. Stat. Phys. **57**, 789 (1989).

[46] D. Thirumalai and R. Mountion, Phys. Rev. A **42**, 4574 (1990).