THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Equilibrium Molecular Thermodynamics from Kirkwood Sampling

Sandeep Somani,*[,†,#] Yuko Okamoto,[‡,§,∥,⊥] Andrew J. Ballard,[†] and David J. Wales[†]

[†]University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, United Kingdom

[‡]Department of Physics, Graduate School of Science, Nagoya University, Nagoya, Aichi 464-8602, Japan

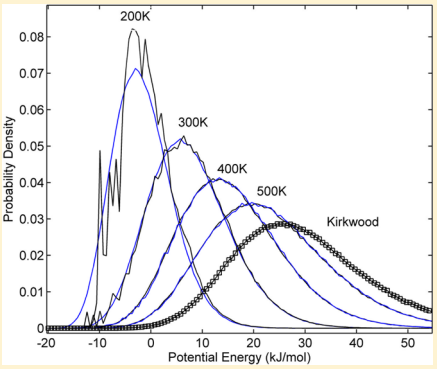
[§]Structural Biology Research Center, Graduate School of Science, Nagoya University, Nagoya, Aichi 464-8602, Japan

[∥]Center for Computational Science, Graduate School of Engineering, Nagoya University, Nagoya, Aichi 464-8603, Japan

[⊥]Information Technology Center, Nagoya University, Nagoya, Aichi 464-8601, Japan

Supporting Information

ABSTRACT: We present two methods for barrierless equilibrium sampling of molecular systems based on the recently proposed Kirkwood method (*J. Chem. Phys.* **2009**, *130*, 134102). Kirkwood sampling employs low-order correlations among internal coordinates of a molecule for random (or non-Markovian) sampling of the high dimensional conformational space. This is a geometrical sampling method independent of the potential energy surface. The first method is a variant of biased Monte Carlo, where Kirkwood sampling is used for generating trial Monte Carlo moves. Using this method, equilibrium distributions corresponding to different temperatures and potential energy functions can be generated from a given set of low-order correlations. Since Kirkwood samples are generated independently, this method is ideally suited for massively parallel distributed computing. The second approach is a variant of reservoir replica exchange, where Kirkwood sampling is used to construct a reservoir of conformations, which exchanges conformations with the replicas performing equilibrium sampling corresponding to different thermodynamic states. Coupling with the Kirkwood reservoir enhances sampling by facilitating global jumps in the conformational space. The efficiency of both methods depends on the overlap of the Kirkwood distribution with the target equilibrium distribution. We present proof-of-concept results for a model nine-atom linear molecule and alanine dipeptide.



1. INTRODUCTION

Equilibrium simulation of molecular systems entails sampling of conformations from the appropriate distribution corresponding to a thermodynamic ensemble, such as the Boltzmann distribution for the isothermal canonical ensemble. Given a potential energy function that describes the interatomic interactions, equilibrium sampling is required to compute the thermodynamic observables, such as internal energy and heat capacity. Development of efficient sampling methods is an important and highly active field of research in biomolecular simulation and molecular science in general.[1−4]

Most thermodynamic sampling methods are based on either the molecular dynamics (MD) or Monte Carlo (MC) approaches. MD simulations are prone to trapping in local minima of the potential energy surface (PES) if there exist barriers that are larger than the available thermal energy. Consequently, long simulations may be required for equilibration on the high dimensional energy landscapes corresponding to biomolecules of practical interest. MC simulation involves random perturbations, or moves, in the conformational space, designed to preserve the canonical distribution. Due to the random perturbations, the system may directly jump between local minima of the PES. However, due to the bonded topology and compact structures of biomolecules, large jumps in the conformational space typically require cooperative motion of multiple atoms. Designing such cooperative moves, which satisfy detailed balance and are computationally efficient, is the primary difficulty in the MC simulation of biomolecules.[5,6]

We recently developed a method,[7] namely, Kirkwood sampling, for surveying the $N = 3M - 6$ dimensional conformational space of a molecule containing M atoms. The key challenge in random conformational sampling of compact biomolecules is to avoid steric clashes. Kirkwood sampling addresses this problem by incorporating correlations among internal coordinates, as captured by the joint probability distribution between them. The joint distributions may be obtained from relatively short high temperature MD or MC simulation trajectories, or even using an informatics approach by exploiting statistics from structural databases such as the Protein Data Bank[8] or the Cambridge Structural Database.[9] Results for small molecules[7] and small peptides[10] suggest that incorporating low order (pairwise and/or 3-fold) correlations may be sufficient to greatly reduce the occurrence of steric clashes in comparison with sampling that ignores all correlations. This is fortuitous, since, in practice, sufficient data is likely to be available for populating only the low order

Received: February 23, 2015
Revised: April 23, 2015
Published: April 27, 2015

DOI: 10.1021/acs.jpcb.5b01800
J. Phys. Chem. B 2015, 119, 6155−6169

pdf's. Furthermore, neglect of the higher order correlations leads to greater coverage of the conformational space compared to the data used to populate the joint distribution functions.

Kirkwood sampling is a geometrical conformational sampling method, independent of the potential energy surface. In the present contribution, we describe two new algorithms, which combine Kirkwood sampling with existing Monte Carlo and replica exchange based approaches. The idea in each case is to improve convergence by taking advantage of the greater conformational space coverage and random (or non-Markovian) sampling properties of the Kirkwood procedure. This paper is organized as follows. Section 2 describes the various aspects of the underlying theory, including the Kirkwood sampling algorithm, application of Kirkwood sampling for conformational sampling of molecules, and the description of the new equilibrium sampling algorithms introduced in this work. Section 3 defines a model nine-atom molecule with an unbranched chain topology, characterizes its energy landscape, and presents a detailed analysis of the convergence of the new algorithms. Section 3 concludes with results for alanine dipeptide, a popular molecule[11,12] for benchmarking simulation approaches. Section 4 summarizes this work and discusses directions for further study.

2. THEORY

2.1. Kirkwood Sampling Algorithm. Let X_1, ..., X_N be N discrete random variables such that the ith variable takes D_i discrete values, $X_i \in \{v_{i,1}, ..., v_{i,D_i}\}$. We denote the kth order probability distribution function (pdf) of a set of k such variables (corresponding to coordinates in the present work) $\{X_1, ..., X_k\}$ as $p_k(X_1, ..., X_k)$. Using lower case letters for specific values for a random variable, $p_k(x_1, ..., x_k)$, denotes $p_k(X_1 = x_1, ..., X_k = x_k)$. The pdf's are assumed to be normalized.

Kirkwood sampling refers to a family of algorithms for sampling points in the full N-dimensional space consistent with select joint distributions among different subsets of the variables, as described previously.[7] The doublet level Kirkwood sampling employs only the 1-D, or singlet, $p_1(X_i)$, and 2-D, or doublet, $p_2(X_i,X_j)$, pdf's. In this work, we employ the doublet level sampling algorithm, but extensions using different sets of pdf's are feasible.[13] Algorithm 1 (main text) presents the pseudo code for generating a point, $\vec{x} = (x_1, ..., x_N)$, in the N-dimensional space using the doublet level sampling algorithm.

Algorithm 1 Doublet level Kirkwood sampling to generate a single point in the N-dimensional space, given all singlet and doublet joint distributions. The symbol "$\sim$" means "sampled from" and upper case denotes the variable to be sampled.

$x_1 \sim p_1(X_1)$

$x_2 \sim \frac{p_2(x_1, X_2)}{p_1(x_1)}$

for $k \leftarrow 3, N$ **do**

$$x_k \sim p_1^{(2)}(X_k|x_1, .., x_{k-1}) = \frac{1}{n_k(x_1, \ldots, x_{k-1}, X_k)} \frac{\prod_{1\le j\le k-1} p_2(x_j, X_k)}{p_1(X_k)^{k-2}}$$

end for

The variables are sampled sequentially from their corresponding one-dimensional conditional pdf, $p_1^{(2)}(X_k|x_1, .., x_{k-1})$. The normalization factor, n_k, of the conditional pdf in Algorithm 1 is obtained numerically by summing the probability for all possible values of X_k, the variable to be sampled. The various Kirkwood sampling algorithms differ in the expression for the conditional probability distribution. The doublet level algorithm samples points from the N-dimensional probability distribution

$$\tilde{p}_N^{(2)}(\vec{x}) = p_2(x_1, x_2) \prod_{3\le k\le N} p_1^{(2)}(x_k|x_1, x_2, .., x_{k-1}) \quad (1)$$

Following previous notation,[7] the superscript "(2)" in the above equations denotes doublet level and will henceforth be dropped. The sampling probability, $\tilde{p}_N^{(2)}$, can be readily computed for any given point. The computational cost is proportional to the number of pdf's used. The computational complexity for doublet level Kirkwood sampling (Algorithm 1) is $O(N^2)$, since there are $N(N-1)/2$ doublet pdf's. The Kirkwood sampling distributions are normalized by construction via normalization of the conditional probability distribution of each variable.

In general, the Kirkwood sampling distribution involves the product of all pdf's employed. As a result, if certain cells in the pdf's have zero probability, then the corresponding combinations of coordinates will never be generated. Thus, if the input pdf's contain such zero probability cells, or "holes", then, strictly speaking, Kirkwood sampling will not be ergodic, since it eliminates certain regions of the conformational space. In principle, this problem can be remedied by filling the holes with a small but finite probability. However, if the pdf's employed are representative of the true distributions, then the conformational regions eliminated by the holes are likely to correspond to high potential energy due to unfavorable interactions such as steric clashes. For example, the unpopulated regions of the Ramachandran plots for the backbone ϕ and ψ torsion angles of a protein correspond to conformations with steric clashes. Hence, one can think of the input pdf's as constraints on the conformational space, and the accessible conformational space shrinks as more pdf's are included.

One approach for obtaining the input pdf's is to populate them using conformations generated through an alternative sampling approach. In the present work, the input pdf's among the coordinates were populated using a representative set of conformations of the molecule obtained by molecular dynamics simulations. In practice, since the order of the input pdf's is likely to be much smaller than the dimensionality of the system, the conformational space accessible to Kirkwood sampling will invariably be much larger than that represented by the conformations used to populate the pdf's.[13] This effect is illustrated for a three-dimensional system in the Supporting Information. Note that the representative set of conformations need not be exhaustive, just large enough to ensure the absence of spurious holes. In practice, even for high dimensional systems, sufficient data is likely to be available for avoiding spurious holes in the low order (1- and 2-D) pdf's.

2.2. Conformational Sampling. In the present context of molecular conformational sampling, the random variables, X_i, denote discretized internal coordinates of a molecule. Following previous work, we employ the bond-angle-torsion (BAT) internal coordinate system.[14] The singlet and doublet pdf's are obtained as histograms of the internal coordinate values observed using MD. Individual binning of each coordinate provides the N singlet (or 1-D) pdf's, and joint binning of all pairs of coordinates provides the $N(N-1)/2$ doublet pdf's. Each coordinate is discretized into B equally spaced bins between the minimum and maximum values observed in MD. The random variables X_i now denote a bin number and can take values in the discrete set {1, ..., B}, that is, for all variables D_i = B and $v_{i,j} = j$. The bin width for discretizing the ith coordinate is set to $\delta_i = \Delta_i/B$, where Δ_i is the range of values observed in the simulation. To visualize a conformation

DOI: 10.1021/acs.jpcb.5b01800
J. Phys. Chem. B 2015, 119, 6155−6169

and for computing energies, the bin numbers for the discrete BAT coordinates need to be mapped into real values, and subsequently to Cartesian coordinates. A discrete BAT coordinate, $x_i \in \{1, ..., B\}$, is mapped to the continuous space value, $\chi_i \in \mathbb{R}$, by picking a random point in the corresponding bin

$$\chi_i = \chi_{i,\min} + (x_i - 1/2)\delta_i + r\delta_i \tag{2}$$

where $\chi_{i,\min}$ is the minimum value for the ith coordinate and $r \in \mathbb{R}$ is a uniformly distributed random number in $[0, 1]$. The left edge of the bin is given by the first two terms of eq 2.

2.3. Canonical Distribution. The conformational distribution for a molecule in contact with a heat bath is given by the Boltzmann distribution. Assuming translational and rotational invariance of the potential energy, the Boltzmann distribution in terms of internal coordinates is given by

$$p_N^{\mathrm{b}}(\vec{\chi}; \beta) \propto \mathrm{e}^{-\beta E(\vec{\chi})} J(\vec{\chi}) \tag{3}$$

where $\beta^{-1} = k_{\mathrm{B}}T$ is the inverse temperature and k_{B} is the Boltzmann constant. In eq 3, $E(\vec{\chi})$ is the potential energy and $J(\vec{\chi})$ is the Jacobian for the transformation from internal to Cartesian coordinates. To present the equations in the following sections using a more familiar notation, we define

$$U(\vec{\chi}; \beta) \equiv E(\vec{\chi}) - \frac{1}{\beta} \ln J(\vec{\chi}) \tag{4}$$

so that the Boltzmann distribution can be written as

$$p_N^{\mathrm{b}}(\vec{\chi}; \beta) \propto \mathrm{e}^{-\beta U(\vec{\chi};\beta)} \tag{5}$$

2.4. Monte Carlo Sampling. Monte Carlo (MC) sampling is a general method for generating points from a multidimensional probability distribution function. MC sampling involves generation of trial points or moves, which are then subjected to an acceptance criterion designed to impose detailed balance with respect to the distribution of interest, here the Boltzmann distribution. The two MC algorithms used in this work are described below.

The first MC algorithm, referred to here as perturbation MC (pMC), is a standard implementation of MC for molecular systems, where a trial conformation, $\vec{\chi}_t$, is generated by random perturbation of each coordinate of the current conformation, $\vec{\chi}_c$,

$$\vec{\chi}_t \leftarrow \vec{\chi}_c + r\vec{s} \tag{6}$$

with r being a uniform random number in $[-0.5, 0.5]$ and $\vec{s}$ a vector of step sizes for the N coordinates. The trial conformation is accepted or rejected using the Metropolis acceptance function

$$f_{\mathrm{pMC}} = \min\left(1, \frac{\mathrm{e}^{-\beta U(\vec{\chi}_t;\beta)}}{\mathrm{e}^{-\beta U(\vec{\chi}_c;\beta)}}\right) \tag{7}$$

The trial conformation is accepted and added to the Markov chain if f_{pMC} is greater than or equal to a random number uniformly generated in $[0, 1]$; otherwise, the move is rejected and the current conformation is added to the chain. During equilibration, components of the step size vector $\vec{s}$ can be adjusted to achieve a specified acceptance ratio. After an optimal $\vec{s}$ has been found, it must be fixed for production runs to guarantee detailed balance.[15] Algorithm 2 gives the pseudo code for perturbation MC.

The second MC algorithm is a variant of standard biased MC[1] (bMC) where the trial conformations are drawn from a

Algorithm 2 Perturbation Monte Carlo (pMC). N_{MC} is the total number of MC steps and $\vec{\chi}_i$ are the sampled conformations. $Y([L, H])$ denotes a uniform random number in $[L, H]$.

```
set χ⃗_0 to a random conformation
set step size vector s⃗
set i=0
while i < N_MC do
    i ← i + 1
    # Perturbation move
    χ⃗_t ← χ⃗_{i−1} + Y([−0.5, 0.5]) × s⃗
    # Metropolis check
    f_pMC ← min(1, e^{−βU(χ⃗_t;β)} / e^{−βU(χ⃗_{i−1};β)})
    if f_pMC ≥ Y([0, 1]) then
        χ⃗_i ← χ⃗_t                                   ▷ accept
    else
        χ⃗_i ← χ⃗_{i−1}                               ▷ reject
    end if
end while
```

probability distribution, termed the biasing distribution, $b_N(\vec{\chi})$, which is defined over the full conformational space

$$\vec{\chi}_t \sim b_N(\vec{\chi}) \tag{8}$$

In contrast to pMC, which is sequential, the moves in biased MC are independent of the current conformation. Figure 1

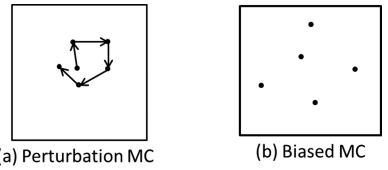


Figure 1. Schematics for the exploration of conformational space by (a) perturbation Monte Carlo (Algorithm 2) and (b) the biased Monte Carlo algorithm (Algorithm 3). The square box represents the conformational space, arrows represent perturbation moves, and filled circles represent conformations. In the case of biased MC, conformations are sampled independently from a biasing distribution.

schematically compares how the conformational space is explored by the two schemes. For biased MC, the acceptance function of pMC is modified to reweight the current and the trial conformation according to the Boltzmann distribution

$$f_{\mathrm{bMC}} = \min\left(1, \frac{\mathrm{e}^{-\beta U(\vec{\chi}_t;\beta)}/b_N(\vec{\chi}_t)}{\mathrm{e}^{-\beta U(\vec{\chi}_c;\beta)}/b_N(\vec{\chi}_c)}\right) \tag{9}$$

Algorithm 3 provides the pseudo code for biased MC. Given a temperature, the acceptance ratio in a biased MC simulation is

Algorithm 3 Biased move Monte Carlo

```
set biasing distribution b_N(χ⃗) to doublet level Kirkwood distribution (eq 1)
set χ⃗_0 ∼ b_N(χ⃗)
set i = 0
while i < N_MC do
    i ← i + 1
    # biased move
    χ⃗_t ∼ b_N(χ⃗)
    # Metropolis check
    f_bMC ← min(1, [e^{−βU(χ⃗_t;β)}/b_N(χ⃗_t)] / [e^{−βU(χ⃗_{i−1};β)}/b_N(χ⃗_{i−1})])
    if f_bMC ≥ Y([0, 1]) then
        χ⃗_i ← χ⃗_t                                   ▷ accept
    else
        χ⃗_i ← χ⃗_{i−1}                               ▷ reject
    end if
end while
```

completely determined by the biasing distribution. MC moves are more likely to be accepted if the overlap between the biasing and the Boltzmann distribution is high. Indeed, all moves will be accepted if the biasing distribution matches the Boltzmann distribution. Biased move MC can be trivially parallelized, since the trial conformations are generated independent of the current conformation in the Markov

DOI: 10.1021/acs.jpcb.5b01800
J. Phys. Chem. B 2015, 119, 6155−6169

chain. As a result, unlike perturbation MC, no equilibration is required for a biased MC simulation. Here, we apply the biased MC algorithm with the trial moves generated by Kirkwood sampling. The key objective of this work is to obtain a Boltzmann distributed set of conformations corresponding to a given potential energy function and temperature, using conformations generated by Kirkwood sampling.

2.5. Replica Exchange Using a Kirkwood Reservoir. Replica exchange (REX) is an enhanced sampling method designed to overcome the trapping problem in canonical simulations.[1,16−21] In REX, multiple canonical simulations, or replicas, are run in parallel and occasionally exchanges between one or more pairs of replicas are attempted. In temperature replica exchange (T-REX), all replicas are run using the same potential energy function but at different temperatures. The lowest replica temperature is usually the temperature of interest where the trapping problem is most severe. Exchanges with the high temperature replicas help the low temperature replicas to escape traps, thereby enhancing sampling in the low temperature replicas. In the present work, MC simulation is used for canonical sampling for each replica so that the exchanges depend only on the potential energy, and not the kinetic energy.

Let β_1, $\vec{\chi}_1$ and β_2, $\vec{\chi}_2$ be the inverse temperatures and instantaneous conformations of two replicas between which an exchange is being attempted. The exchange probability function in T-REX is given by

$$f_{\text{T-REX}} = \min(1, e^{[U(\vec{\chi}_1;\beta_1)-U(\vec{\chi}_2;\beta_2)](\beta_1-\beta_2)}) \quad (10)$$

The conformations are exchanged if $f_{\text{T-REX}}$ is greater than or equal to a uniform random number in $[0, 1]$. Usually, exchanges are attempted between replicas that are adjacent in the temperature ladder, though other schemes have also been suggested.[21−23] Algorithm 4 provides the pseudo code for the present implementation of T-REX.

Algorithm 4 Temperature Replica Exchange (T-REX).

set N_R, the number of replicas
set N_{pMC}, the number of perturbation MC steps between exchange attempts
set N_{MC}, the total number of MC steps
set $\beta^r (r = 1, \ldots, N_R)$, inverse temperature of replica r, s.t. $1/\beta^{r+1} > 1/\beta^r$
set $\vec{\chi}_0^r (r = 1, \ldots, N_R)$, initial conformation of replica r
set $i = 0$, MC step counter
while $i < N_{\text{MC}}$ **do**
  # canonical simulation for each replica
  for $r = 1 : N_R$ **do**
    For replica r do N_{pMC} steps of pMC (Algorithm 2)
  end for
  $i \leftarrow i + N_{\text{pMC}}$
  # exchange between replica p and $p+1$
  p = random integer in $\{1, 2, \ldots, N_R - 1\}$
  # Metropolis check
  $f_{\text{T-REX}} \leftarrow \min\left[1, \exp\left(\left[U\left(\vec{\chi}_i^{p+1}; \beta^{p+1}\right) - U\left(\vec{\chi}_i^{p}; \beta^{p}\right)\right]\left(\beta^{p+1} - \beta^{p}\right)\right)\right]$
  if $f_{\text{T-REX}} \geq Y([0, 1])$ **then**
    $\vec{\chi}_t \leftarrow \vec{\chi}_i^p$; $\vec{\chi}_i^p \leftarrow \vec{\chi}_i^{p+1}$; $\vec{\chi}_i^{p+1} \leftarrow \vec{\chi}_t$ ▷ exchange
  end if
end while

In T-REX, each replica may be considered as a reservoir of Boltzmann distributed conformations for other replicas, albeit at different temperatures. Exchanges may also be performed with a reservoir of non-Boltzmann distributed conformations, as long as the probability distribution of the reservoir is known. Kirkwood sampling satisfies this condition, as the probability of sampling a given conformation can be computed (here, using eq 2). A temperature replica exchange where one (or more) temperature replica(s) are coupled to a Kirkwood sampling scheme will be referred to as Kirkwood reservoir replica exchange (KR-REX). The probability for exchanges between a temperature replica at inverse temperature β and a reservoir with distribution $b_N(\vec{\chi})$ is given by

$$f_{\text{KR-REX}} = \min\left(1, \frac{b_N(\vec{\chi}_\beta)}{b_N(\vec{\chi}_R)} e^{\beta(U(\vec{\chi}_\beta;\beta)-U(\vec{\chi}_R;\beta))}\right) \quad (11)$$

where $\vec{\chi}_\beta$ is the conformation from the temperature replica and $\vec{\chi}_R$ is a conformation drawn from the reservoir, that is, $\vec{\chi}_R \sim b_N(\vec{\chi})$. Note that the acceptance function involves the reservoir probability of the conformation from the temperature replica, $b_N(\vec{\chi}_\beta)$, as well as the potential energy of the reservoir conformation, $U(\vec{\chi}_R;\beta)$. The likelihood of acceptance from the reservoir depends on the conformational overlap between the reservoir and the actual Boltzmann distributions.

The acceptance ratio for biased MC simulation at a given temperature gives the probability of exchange with a replica at the same temperature. Therefore, given an acceptance ratio for exchanges with the reservoir, short biased MC simulations are performed at multiple temperatures to determine the lowest temperature to which the Kirkwood reservoir can be coupled. Schematics of the exchange protocols and exchange probability function for T-REX and KR-REX are shown in Figure 2. Exchanges with the reservoir are attempted at every alternate exchange cycle, as detailed in Algorithm 5.

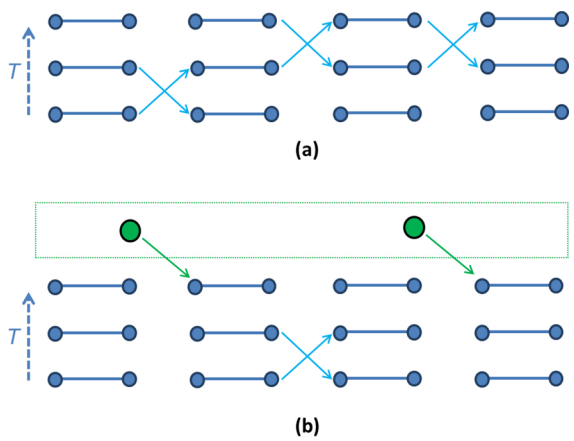


Figure 2. Schematic representation of (a) temperature (T-REX, Algorithm 4) and (b) Kirkwood reservoir (KR-REX, Algorithm 5) replica exchange. The horizontal lines represent different replicas with the lower lines corresponding to lower temperatures. The horizontal line segments represent N_p steps of perturbation MC. Blue arrows represent attempts for exchanging conformations between two replicas. The green dotted box in part b represents the Kirkwood reservoir, filled green circles represent conformations sampled from the reservoir distribution, and green arrows indicate exchange attempts between the reservoir and the highest temperature replica. Note that conformations in the reservoir are not updated with conformations from the temperature replicas.

2.6. Consistency Check on Equilibrium Sampling of Replica Exchange Simulations. The Kirkwood replica exchange strategy can efficiently sample thermodynamic states by coupling a set of temperature replicas to a reservoir of Kirkwood-generated structures. Although the trial configurations are sampled from the Kirkwood distribution, which is different from the underlying Boltzmann distribution, the acceptance criterion, eq 11 above, guarantees that the move satisfies detailed balance, ensuring that each of the temperature replicas samples its correct equilibrium distribution.

DOI: 10.1021/acs.jpcb.5b01800
J. Phys. Chem. B 2015, 119, 6155−6169

Algorithm 5 Kirkwood Reservoir Replica Exchange (KR-REX).

set N_R, the number of replicas
set N_{pMC}, the number of perturbation MC steps between exchange attempts
set N_{MC}, the total number of MC steps
set $\beta^r (r = 1, \ldots, N_R)$, inverse temperature of replica r, s.t. $1/\beta^{r+1} > 1/\beta^r$
set $\vec{\chi}_0^r (r = 1, \ldots, N_R)$, initial conformation of replica r
set b_N, biasing distribution (here, doublet level Kirkwood distribution (Eq. (1))
set $i = 0$, MC step counter
set $i_{cyc} = 0$, number of replica exchange attempts
while $i < N_{\mathrm{MC}}$ **do**
  # canonical simulation in each replica
  for $r = 1 : N_R$ **do**
    N_{pMC} steps of pMC (Algorithm 2) in replica r
  end for
  $i \leftarrow i + N_{\mathrm{pMC}}$
  $i_{cyc} \leftarrow i_{cyc} + 1$
  if i_{cyc} is ODD **then**
    # attempt exchange between a randomly chosen pair of
    # adjacent temperature replicas as described in Algorithm 4
  else
    # attempt exchange between reservoir and the highest temperature replica
    $\vec{\chi}_R \sim b(\vec{\chi})$ ▷ sample from reservoir
    $f_{\mathrm{KR-REX}} \leftarrow \min\left(1, \frac{b_N(\vec{\chi}_i^{N_R})}{b_N(\vec{\chi}_R)} e^{\beta_{N_R}\left(U\left(\vec{\chi}_i^{N_R};\beta_{N_R}\right) - U(\vec{\chi}_R;\beta_R)\right)}\right)$
    if $f_{\mathrm{KR-REX}} \geq Y([0,1])$ **then**
      $\vec{\chi}_i^{N_R} \leftarrow \vec{\chi}_R$ ▷ accept
    end if
  end if
end while

We employ the overlapping distribution method[1] to verify that the Kirkwood exchange moves do not disturb the underlying equilibrium ensembles in the temperature replicas. This approach was originally introduced by Bennett[24] for free energy calculations but can also be used as a consistency check on equilibrium sampling[25,26] for replica exchange simulations.[27] Below, we detail how it was applied in the present work.

We wish to verify that each of our temperature replicas samples its corresponding equilibrium distribution. Let us consider a pair of temperature replicas, A and B, at inverse temperatures β_A and β_B, respectively. In the canonical ensemble, a particular replica, A, for example, samples energy E with distribution

$$p^A(E;\beta_A) = \Omega(E) e^{-\beta_A(E - f^A)} \tag{12}$$

where $\Omega(E)$ is the energy density of states and $f^A = -\beta_A^{-1} \ln Z^A$ is the Helmholtz free energy of the system. (We emphasize that the relevant quantity to analyze with this method is the energy E, not the quantity U which contains the Jacobian factors.) We have a similar expression for p^B corresponding to replica B. Although p^A and p^B are generally unknown (due to the unknown density of states), they are related to each other in a simple way as

$$\frac{p^A(E)}{p^B(E)} = e^{\Delta\beta E - \Delta f} \tag{13}$$

Here, $\Delta\beta = \beta_B - \beta_A$ and $\Delta f = \beta_B f^B - \beta_A f^A$ is the reduced free energy difference between the two replicas. Hence, these two energy distributions are connected via the unknown constant Δf. This relation can be seen as a manifestation of Crooks' fluctuation theorem,[28] and is in general true only when both distributions sample their respective equilibrium states, eq 12 above. Hence, a verification of eq 13 for the temperature replicas in a Kirkwood reservoir replica exchange simulation provides a consistency check on the sampling.

To proceed with this consistency check, we follow Bennett[24] and define two functions:

$$L_A(E) = \ln p^A(E) - \Delta\beta E/2 \tag{14a}$$

$$L_B(E) = \ln p^B(E) + \Delta\beta E/2 \tag{14b}$$

Although L_A and L_B are functions of E, from eq 13, we see that

$$\Delta L(E) \equiv L_B(E) - L_A(E) = \Delta f \tag{15}$$

is a constant, whose value corresponds to the reduced free energy difference. This relationship can be verified from simulation data by binning the sampled energies from our various replicas, constructing the functions L_A and L_B, and plotting their difference in each bin. A plot of ΔL as a function of E should, within statistical errors, provide a horizontal line with a slope of zero. In section 3.1.4, we apply this test to our replica exchange simulations in order to verify equilibrium sampling.

3. RESULTS

The equilibrium sampling algorithms, described in section 2, were implemented in Matlab[29] and Octave.[30] We first applied the methods on a model nine-atom chain molecule (Figure 3)

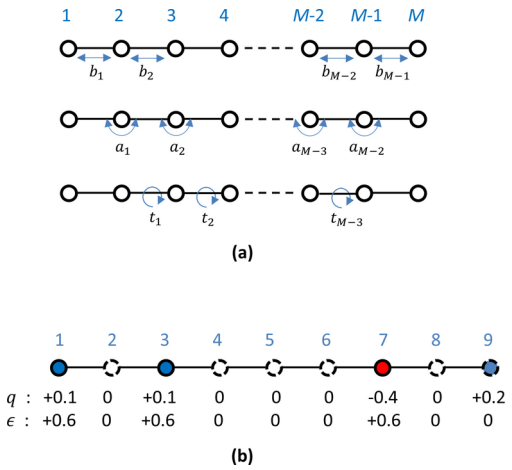


Figure 3. (a) Bond-angle-torsion (BAT) coordinate system for an M atom chain molecule and (b) electrostatic charge and Lennard-Jones well-depth of the test system studied in this work. Atoms without charge and Lennard-Jones interaction are represented by unfilled dashed circles.

with a simplified force field. The model system was constructed for computational efficiency and for ease of implementing internal coordinate Monte Carlo moves. Kirkwood based biased MC and reservoir replica exchange are applied to the model system, and results are compared with conventional alternative approaches. The simulations were designed with the objective of validating the new sampling algorithms and describe the considerations for setting up the simulations. The last section presents biased MC results for alanine dipeptide illustrating the applicability to small biomolecules.

3.1. Model System. This section describes the force field and coordinate system for the model system, characterizes its energy landscape, and generates benchmark results using an independent well-converged MD replica exchange simulation. We then present results for perturbation and biased MC simulations applied to our system at progressively lower temperatures. We will see that both of these methods fail to converge below a certain temperature. Finally, we present results from T-REX and KR-REX replica exchange simulations, which enhance convergence at the lower temperatures. The number of MC steps is 5×10^6 for all MC and replica exchange simulations described in this section.

DOI: 10.1021/acs.jpcb.5b01800
J. Phys. Chem. B 2015, 119, 6155−6169

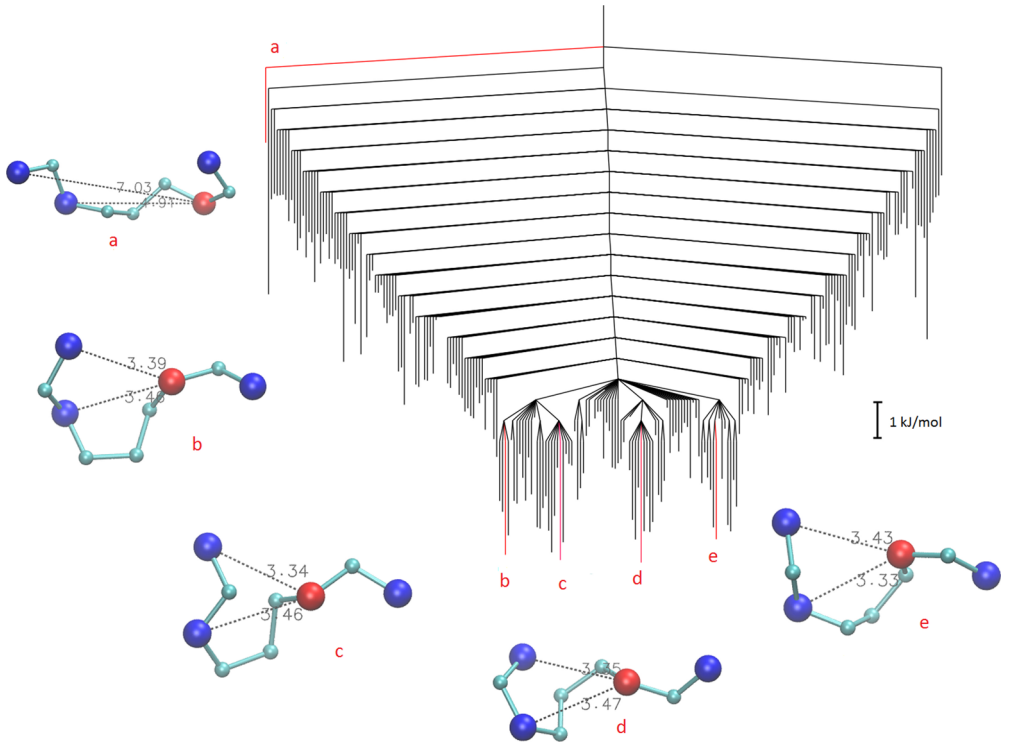


Figure 4. Disconnectivity graph constructed from a database of 236 minima and 1454 transition states. Branches leading to select minima (a−e) are colored red. Conformations of these labeled minima are shown with the positively charged atoms in blue and the negatively charged atom in red (see Figure 3b for atom numbering and the Supporting Information for potential energy definition). Distances between atom pairs (1,7) and (3,7) are also shown. Note that the low energy minima (b−e) are more compact than the high energy minimum (a) and have a similar arrangement of atoms 1, 3, and 7. The energy of the global minimum (d) is −22.6 kJ/mol, and that of minimum (a) is −9.9 kJ/mol. Typical barriers between minima belonging to the three low energy funnels are around 4.8 kJ/mol, corresponding to a temperature of roughly 1000 K.

3.1.1. Energy Landscape and Coordinate System. The conformation of the model system is specified by 21 BAT coordinates, which include eight bond lengths (b_i), seven bond angles (a_i), and six torsion angles (t_i). Figure 3 gives the atom labeling and the definition of the BAT coordinates. The functional form of the potential energy of the molecule employs a molecular mechanics-type force field, which includes bonded and nonbonded terms

$$E(\vec{\chi}) = \sum_{i=1}^{M-1} E_b(b_i) + \sum_{i=1}^{M-2} E_a(a_i) + \sum_{i=1}^{M-3} E_t(t_i) + \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} [E_{LJ}(r_{i,j}) + E_E(r_{i,j})] \quad (16)$$

where $M = 9$ is the number of atoms. The terms associated with the bonded, angular, and torsional degrees of freedom are E_b, E_a, and E_t, respectively. A harmonic functional form was used for the bonded and angular contributions, while the Ryaekart−Bellman functional form was used for the torsional term. In eq 16, E_{LJ} and E_E denote the nonbonded Lennard-Jones (LJ) and electrostatic interactions, respectively, and $r_{i,j}$ denotes the distance between atoms i and j. To compute distances between atoms, the BAT coordinates were transformed to anchored Cartesian coordinates.

In the present study, the potential energy of the test system included all bonded terms but only selected nonbonded contributions (see Figure 3 and the Supporting Information). LJ interactions were considered only for atom pairs (1,7) and (3,7), while electrostatic interactions were included only for atom pairs (1,7), (3,7), (1,9), and (3,9). Atoms 1, 3, and 9 carried positive charges of +0.1, +0.1, and +0.2, respectively. Atom 7 carried a negative charge of −0.4, and the remaining atoms were neutral. Fewer nonbonded interactions were included to reduce the cost of the energy evaluation and obtain accurately converged results for the benchmarking more efficiently. The full energy function and its parameters and the input files for GROMACS[31] and AMBER[32] are provided as Supporting Information.

Although the potential that we have defined is relatively simple, the corresponding landscape can still provide a useful benchmark to compare different sampling schemes. In order to gain some idea of this complexity, we sampled local minima using basin-hopping global optimization,[33−35] and then computed pathways connecting the global minimum with all other minima using the doubly nudged[36,37] elastic band[38−40] method with accurate refinement of transition states by hybrid eigenvector-following.[41] The resulting database of minima and transition states was used to plot the disconnectivity graph[42,43] shown in Figure 4, which suggests a largely funnelled potential energy surface. Figure 4 shows conformations of representative low and high energy minima. The low energy minima adopt a more compact conformation with similar geometry of the oppositely charged atoms. The energy of the global minimum was −22.6 kJ/mol, and there are three prominent funnels at energies below −16 kJ/mol, with barriers between funnels of roughly 4.14 kJ/mol, corresponding to a temperature of 500 K. Thus, at temperatures significantly lower than 500 K, the system is likely to be trapped in one of the low energy funnels.

DOI: 10.1021/acs.jpcb.5b01800
J. Phys. Chem. B 2015, 119, 6155−6169

In this section, we investigate convergence of thermodynamic quantities for temperatures ranging from 20 to 500 K.

3.1.2. Reference Equilibrium Sampling. Reference canonical potential energy distributions were generated at different temperatures using MD replica exchange. Twelve replica temperatures were used: 10, 20, 30, 50, 75, 100, 130, 165, 200, 300, 400, and 500 K. The MD REX simulations were performed with GROMACS 4.6.5,[31] using Langevin dynamics with a time step of 1 fs and a friction coefficient of 5 ps^{-1}. The first 50 ns of the simulation was discarded for system equilibration. Production data was collected in the subsequent run of 500 ns. Replica exchanges were attempted every 500 time steps. Conformations and energies were saved every 0.5 ps (500 time steps), giving 10^6 data points per replica.

Figure 5 shows the potential energy distribution observed during the production run, and Table 1 displays the average

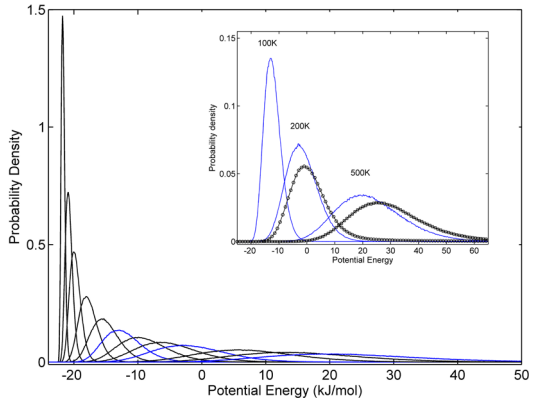


Figure 5. Potential energy distributions obtained from reference MD replica exchange simulation described in section 3.1.2. The distributions correspond to temperatures 10, 20, 30, 50, 75, 100, 130, 165, 200, 300, 400, and 500 K, from left to right. The inset shows the energy distribution of doublet level Kirkwood samples from refpdfsT500 (line marked by boxes) and refpdfsT200 (solid line by circles) overlaid on select canonical energy distributions (unmarked lines). Kirkwood samples generated using refpdfsT200 and refpdfsT500 have significant overlap with the canonical distribution at 200 and 500 K, respectively.

energy and heat capacity computed for the simulation temperatures using the potential energy values observed in each replica. The heat capacity for the replica at temperature T was computed using

$$C(T) = \frac{1}{k_B T^2}(\langle E^2\rangle - \langle E\rangle^2) \quad (17)$$

The average energy and heat capacity were interpolated between each replica exchange simulation temperature by applying the multihistogram method[44−48] to the MD REX potential energy distributions, using an energy bin width of 0.1 kJ/mol. The peak heat capacity from the multihistogram heat capacity curve is 26.6 $k_B/2$ at 104.5 K. Figure 6 shows the two

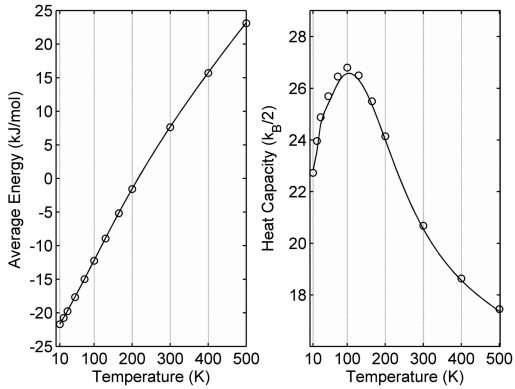


Figure 6. Average energy and configurational heat capacity computed using the MD REX potential energy distributions shown in Figure 5. Circles represent values computed directly from the observed energy values in each replica.

quantities computed at 50 equally spaced temperatures between 10 and 500 K. At the simulation temperatures, the interpolated values are in good agreement with the values directly computed from the simulation data (see Table 1).

BAT coordinates were extracted from all 12 trajectories, giving 12×10^6 data points for each coordinate, with 30 equally spaced bins between the minimum and maximum observed values in each case. The discretized coordinates were used to populate the 1-D (singlet) probability distribution for each coordinate and 2-D (doublet) probability distribution for all pairs of coordinates. In all, 21 singlet and 210 doublet distributions were populated and used to sample conformations using the doublet level Kirkwood algorithm (Algorithm 1). Note that, since data from all replicas were pooled together, the conformations used to populate the pdf's are not Boltzmann distributed for a single temperature. Nevertheless, since more conformational space is covered at higher temperatures, the configurations will be more representative of the highest temperature replica at 500 K. We refer to this set of pdf's as refpdfsT500. Another set of pdf's, referred to as refpdfsT200, was generated using 2×10^6 conformations from a separate 200 ns MD simulation at a lower temperature of 200 K.

Kirkwood sampling was performed to generate 10^6 conformations for each pdf set. The Figure 5 inset shows the distribution of potential energies of the Kirkwood samples overlaid on the canonical potential energy distributions at different temperatures. The energy distribution of Kirkwood samples generated from refpdfsT500 (curve marked by boxes) has significant overlap with the canonical distribution at 500 K, and the distribution corresponding to refpdfsT200 (curve marked by open circles) has significant overlap with the canonical distribution at 200 K. These results are consistent

Table 1. Average Energy (kJ/mol), $\langle E\rangle$, and Heat Capacity ($k_B/2$), C, Computed Using Energy Values from the Reference MD Replica Exchange Simulation Described in Section 3.1.2

T (K)	$\langle E\rangle$	$\langle E^2\rangle$	C
10	−21.70	471.05	22.72
20	−20.75	430.91	23.96
30	−19.75	390.95	24.88
50	−17.66	314.32	25.69
75	−14.98	229.75	26.45
100	−12.24	159.10	26.80
130	−8.93	95.36	26.49
165	−5.18	50.82	25.49
200	−1.57	35.82	24.14
300	7.62	122.37	20.68
400	15.70	349.62	18.64
500	23.11	685.02	17.45

DOI: 10.1021/acs.jpcb.5b01800
J. Phys. Chem. B 2015, 119, 6155−6169

Table 2. Average Energy (kJ/mol), $\langle E \rangle$, and Heat Capacity ($k_B/2$), C, Computed Using Energy Values for Perturbation and Biased Move MC Using refpdfsT500[a]

	reference		perturbation MC				biased MC				
T (K)	$\langle E \rangle$	C	$\langle E \rangle$		C		$\langle E \rangle$		C		Acc ratio
50	−17.66	25.69	−17.48	(0.18)	23.42	(−2.27)					
100	−12.24	26.8	−12.25	(−0.01)	25.59	(−1.21)					
200	−1.57	24.14	−1.63	(−0.06)	23.88	(−0.26)	−0.73	(0.84)	21.52	(−2.62)	0.009
300	7.62	20.68	7.54	(−0.08)	20.52	(−0.16)	7.76	(0.14)	20.73	(0.05)	0.07
400	15.7	18.64	15.87	(0.17)	18.58	(−0.06)	15.79	(0.09)	18.57	(−0.07)	0.25
500	23.11	17.45	23.18	(0.07)	17.36	(−0.09)	23.16	(0.05)	17.28	(−0.17)	0.29

[a]The acceptance ratio observed in the biased MC simulations and reference values from Table 1 are also given. The numbers in parentheses are differences with respect to the reference.

with earlier work,[7,10] where the energy distribution of the Kirkwood samples was found to overlap strongly with the original Boltzmann distribution from which the conformations used to populate the reference pdf's were generated. Indeed, Kirkwood sampling was originally developed for approximating the Boltzmann distribution at a given temperature. Note that the energy distributions of the Kirkwood samples are shifted to higher values relative to the original canonical distribution, due to the greater coverage of the conformational space, and the fact that there are more conformations at higher energies.

3.1.3. Perturbation and Biased Monte Carlo. Perturbation MC (Algorithm 2) and biased MC (Algorithm 3) simulations were performed for successively lower temperatures starting from 500 K. Both simulations were performed at temperatures of 200, 300, 400, and 500 K; perturbation MC was also conducted at the lower temperatures of 100, 50, and 20 K. The number of MC steps for all simulations was 5×10^6. For the biased MC simulations, Kirkwood samples were generated from refpdfsT500. Table 2 gives the average energy and heat capacity computed from the potential energies from different simulations and their difference with respect to the reference values in Table 1.

Considering the biased MC results first, Figure 7 shows the distribution of potential energies for each simulation overlaid on the reference distributions from Figure 5. The potential energy distribution of Kirkwood samples is also shown. The biased MC and reference distributions are in good agreement

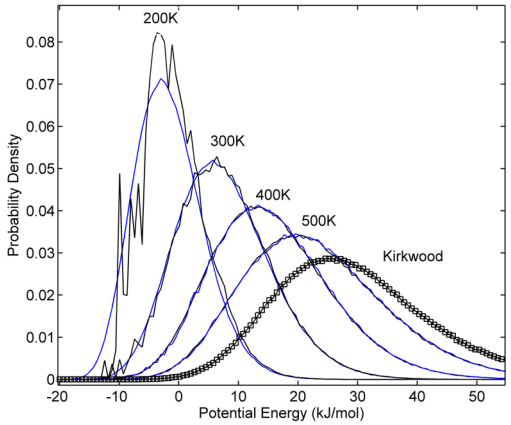


Figure 7. Potential energy distributions (in black) from biased move MC simulations (section 3.1.3) at T = 200, 300, 400, and 500 K using Kirkwood moves with refpdfsT500. The reference distributions from Figure 5 are in blue. The potential energy distribution of the original Kirkwood samples is also shown marked by boxes.

for temperatures ≥300 K, consistent with the average energy and heat capacity results in Table 2. These results show that Kirkwood sampling with a fixed set of input pdf's can be used to obtain Boltzmann distributed conformations at different temperatures. This is the key finding of the present work. The biased MC algorithm may be viewed as a resampling approach[49,50] wherein a set of Kirkwood distributed conformations are resampled to generate a Boltzmann distributed set. In Figure 7, resampling effectively shifts the right-most distribution (line marked by boxes) to match with the different Boltzmann distributions. Note that Boltzmann distributions can be generated for temperatures lower than that of the MD simulation used to populate the input pdf's. Biased MC simulations in Figure 7 used input pdf's effectively populated with 500 K MD simulation data but were able to generate Boltzmann distributions at 300 K.

Acceptance ratios for the biased MC simulations were 0.29, 0.25, 0.07, and 0.009 for T = 500, 400, 300, and 200 K, respectively. The acceptance ratio falls as the temperature is reduced, consistent with the decreasing overlap between the Kirkwood potential energy distribution and the canonical energy distributions (see Figure 5). Note that the conformations obtained by Kirkwood sampling are completely determined by the input pdf's. As a result, in contrast to perturbation MC, the acceptance ratio cannot be adjusted for a given temperature. In other words, a given acceptance ratio would impose a lower limit on the temperatures for which biased MC simulations can be run. Note that the overlap of the Kirkwood distribution with a target canonical distribution can be determined even in the absence of the target distribution. The acceptance ratio of a biased MC simulation is a direct measure of the overlap. Moreover, since no equilibration is required for biased MC simulations, a relatively short run can provide an estimate of the acceptance ratio.

We now characterize the perturbation MC simulations, which are used in the next section for canonical sampling in the replica exchange simulations. In each pMC simulation, a 10 000 step preliminary run was performed to adjust the step size for an acceptance ratio between 0.2 and 0.3. Figure 8 shows the comparison of the pMC potential energy distributions with the reference distributions (see Figure 5). Table 2 and Figure 8 show that the pMC simulations are in good agreement with the reference values for $T \geq 100$ K, with higher temperatures producing better agreement. The deviations at low temperatures are due to the trapping of the Markov chain in local potential energy wells.

The successful convergence for pMC at the low (relative to typical barriers of the PES) temperature of 100 K is likely due to the efficiency of the internal coordinate BAT moves. For

DOI: 10.1021/acs.jpcb.5b01800
J. Phys. Chem. B 2015, 119, 6155−6169

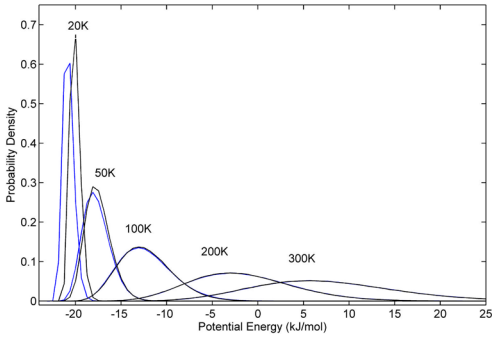


Figure 8. Potential energy distributions (in black) from the perturbation MC simulations (section 3.1.3) overlaid on the reference distributions (in blue).

comparison, we performed a 100 K pMC simulation using random Cartesian moves, which was 10 times longer than the above BAT move pMC simulation. The Cartesian move MC was performed using the GMIN program.[51] Figure 9 shows that

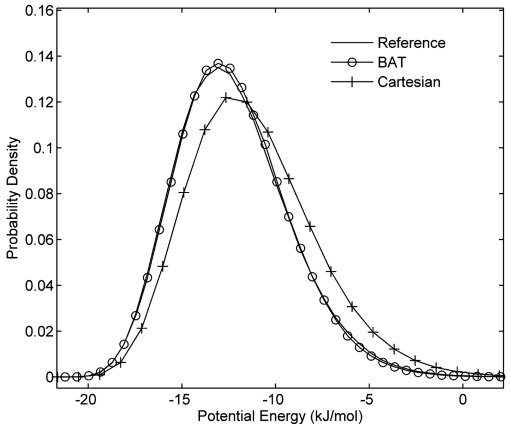


Figure 9. Potential energy distributions from perturbation MC simulations at 100 K using bond-angle-torsion (BAT) and Cartesian coordinate moves. The Cartesian move simulation was run for 5×10^7 steps, while the BAT move simulation was 10-fold shorter at 5×10^6 steps. The distributions show that BAT coordinate moves are move efficient than Cartesian coordinate moves.

the energy distribution from the Cartesian move simulation is in much poorer agreement with the reference distribution. We note that, even though the coordinates are perturbed independently, BAT coordinate moves are particularly effective here because of the unbranched chain topology of the molecule. Absence of side chains greatly reduces the chances of steric clashes, even if a torsion angle in the middle of the chain is perturbed substantially.

3.1.4. Temperature and Kirkwood Reservoir Replica Exchange. We now compare temperature replica exchange (Algorithm 4) and reservoir replica exchange (Algorithm 5), focusing on the low temperatures (≤100 K) for which the MC simulations of the previous section failed to converge. Replica exchange simulations were performed for two sets of temperatures, namely, a high temperature set, 20, 50, 100, and 200 K, and a low temperature set, 20, 30, 50, and 100 K. The high temperature set was chosen to inspect the convergence of the heat capacity peak at 104.5 K, and the low temperature set was used to inspect the impact of the reservoir on the convergence. In the case of reservoir replica exchange simulations, for both sets, the Kirkwood reservoir was coupled to the highest temperature replica. The Kirkwood reservoir for the high temperature set was based on refpdfsT500, and that for the low temperature set was based on refpdfsT200.

Perturbation MC was used for canonical simulation in the temperature replicas, and for each temperature, the step size was adjusted during equilibration to achieve an acceptance ratio between 0.2 and 0.3. The length of the production run was 5×10^6 steps, and exchanges were attempted every 20 steps using the exchange protocols described in Algorithm 4 and Algorithm 5. Energy values were saved at every MC step generating 5×10^6 energy values per replica.

High Temperature pdf Set. We first discuss the REX simulations using the high temperature pdf set. For all REX simulations, we performed the overlap test (section 2.6) to validate the implementation of the algorithms and, in the case of KR-REX simulations, also verify that coupling to the reservoir did not disturb the Boltzmann distribution in the temperature replicas. Figure 10 shows the analysis for replicas 2 and 3, corresponding to 50 and 100 K, respectively, for the KR-REX simulation. The inset in the lower panel of Figure 10 also shows the plot of ΔL for the corresponding replicas in the T-REX simulation. We see that the function ΔL is flat within statistical errors, indicating that the fluctuation theorem, eq 13,

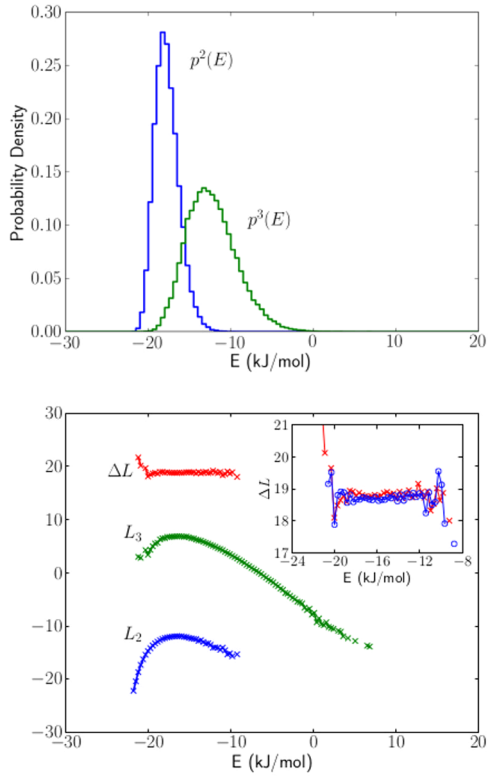


Figure 10. Consistency check for equilibrium sampling. The overlap test is performed on replicas 2 and 3, at temperatures 50 and 100 K, respectively, taken from the KR-REX simulations described in section 3.1.4 using the high temperature pdf set. Upper panel: Energy distributions for replicas 2 and 3. Lower panel: The overlap functions L_2, L_3, and ΔL (eqs 14 and 15). The function ΔL is flat within statistical error, indicating that the replicas are sampling their correct equilibrium distributions (see section 2.6). Inset: The function ΔL expanded (red crosses), compared to ΔL as calculated from the same temperatures in the T-REX simulation (blue circles).

DOI: 10.1021/acs.jpcb.5b01800
J. Phys. Chem. B 2015, 119, 6155−6169

is satisfied, and hence both replicas are sampling their respective canonical distributions correctly. Other replica pairs displayed similar results (data not shown).

We next analyze and compare the convergence of the T-REX and KR-REX simulations using the high temperature set. For both simulations, the exchange acceptance ratios were 0.02, 0.08, and 0.08 for replica pairs (20,50), (50,100), and (100,200), respectively. The lowest exchange acceptance ratio of 0.02 corresponds to 5000 successful exchanges. For KR-REX simulation, the number of successful exchanges with the reservoir was roughly 1100, consistent with the expected acceptance ratio of 0.009 based on the biased MC simulation at 200 K. The convergence of a replica exchange simulation is limited by the time needed for a trajectory to diffuse between its lowest temperature, where it is trapped in a metastable state, to a high temperature one, where it can overcome its energetic barriers via enhanced thermal fluctuations or the Kirkwood reservoir. To analyze this convergence, we investigate the diffusion of replica trajectories among the ladder of temperatures. Figure 11 shows the trajectory that began in the lowest

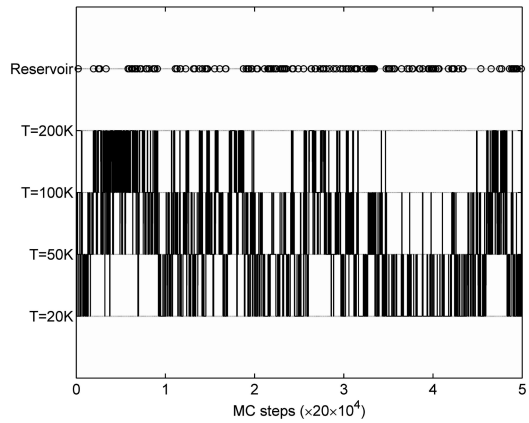


Figure 11. Trajectory of the 20 K replica from the KR-REX simulation described in section 3.1.4 using the high temperature pdf set. Successful exchanges with the reservoir are indicated by open circles. Exchanges were attempted after every 20 MC steps. The trajectory for the first 5×10^4 exchange attempts is depicted here.

temperature replica in the KR-REX simulation. It shows significant diffusion between the different replicas. On average, 30 round trips were observed per million MC steps. The profile was similar for the T-REX simulation.

We now investigate the convergence of the energy distributions and thermodynamic quantities. Figure 12 compares the energy distributions of the T-REX and KR-REX simulations with the reference distribution from MD REX. The three distributions are essentially identical for all temperatures, including 20 and 50 K, which failed to converge in the case of single temperature perturbation MC simulations (see Figure 8). The T-REX and KR-REX simulations were also compared in terms of convergence of the average energy and heat capacity as a function of temperature, which is a more stringent test of convergence than comparison of energy distributions. The average energy and heat capacity were obtained by applying the multihistogram method to the potential energy distributions computed using successively larger fractions of the simulation data. Figure 13 shows the convergence of the average energy and heat capacity overlaid on the reference curves, and Figure 14 shows the difference

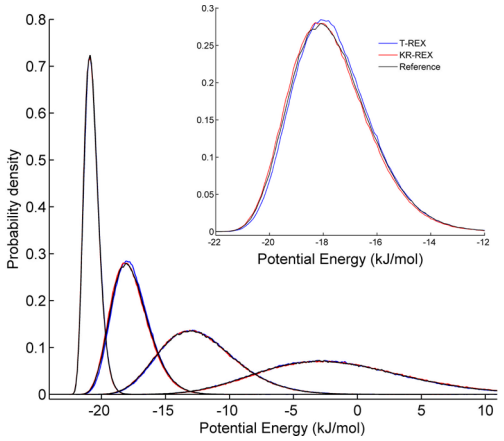


Figure 12. Potential energy distributions for T-REX (blue) and KR-REX (red) simulations using the high temperature set described in section 3.1.4. Distributions are shown for T = 20 (left-most), 50, 100, and 200 K (right-most). The reference distributions are in black. The three distributions are nearly overlapping for all temperatures, with T = 50 K (inset) showing the largest deviations.

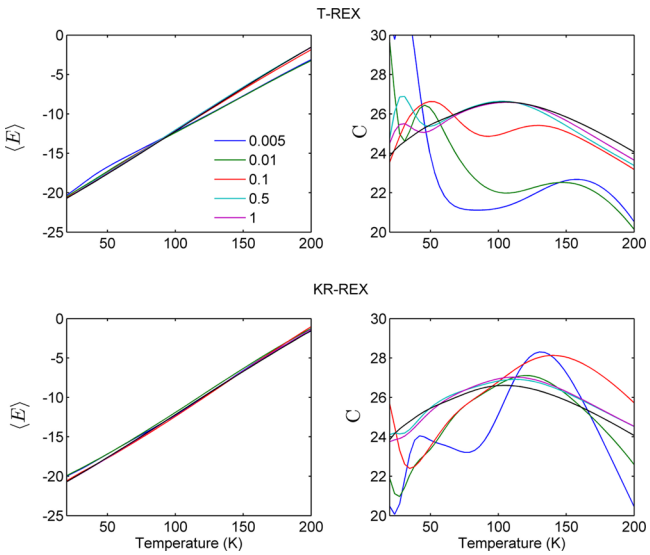


Figure 13. Convergence of average energy (left panels) and heat capacity (right panels) for T-REX (top) and KR-REX (bottom) simulations using the high temperature pdf set (section 3.1.4). The legend indicates the fraction of data used from the 5×10^6 step production run. The reference curves (from Figure 6) are in black.

between each curve and the reference. For both T-REX and KR-REX, the average energy converges to within 0.5 kJ/mol over the full temperature range of 20 and 200 K using only 10% (500 000 steps) of the production data. The convergence of the heat capacity is, as expected, slower than that for the average energy. Nevertheless, both the T-REX and KR-REX curves converge to within 0.25 k_B of the reference with 50% of the production data. These results show that the T-REX and KR-REX simulations were consistent with each other. However, from these simulations, it is difficult to gauge the impact of the reservoir on the convergence.

Low Temperature pdf Set. The effect of the reservoir in the high temperature set REX simulations was not discernible—probably because the highest temperature of 200 K was sufficiently high to avoid trapping, so that additional jumps in the conformation space through exchanges with the reservoir

DOI: 10.1021/acs.jpcb.5b01800
J. Phys. Chem. B 2015, 119, 6155−6169

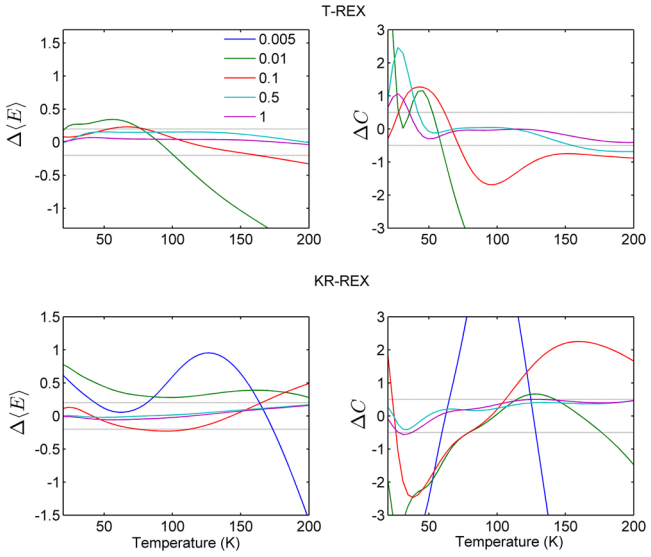


Figure 14. Differences between the average energy and heat capacities in Figure 13 and the reference values from Figure 6. The horizontal dotted lines in the left panels indicate energy differences of ±0.2 kJ/mol, and those in the right panel indicate heat capacity differences of $\pm k_B/4$.

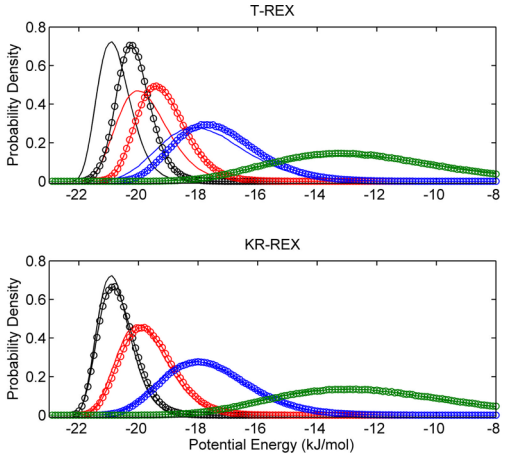


Figure 15. Potential energy distributions (lines marked by circles) from the 5 × 10^6 steps (a) T-REX and (b) KR-REX simulations described in section 3.1.4 using the low temperature pdf set with replica temperatures of 20 K (black), 30 K (red), 50 K (blue), and 100 K (green). The distributions from the KR-REX simulation are much closer to the reference than those from the T-REX simulation, indicating that coupling to the reservoir enhanced convergence.

did not improve convergence. To investigate the effect of the reservoir, we attempted to reduce the highest temperature in the REX simulations. However, for temperatures below 200 K, refpdfsT500 could not be used to set up the reservoir because of a low acceptance ratio. We therefore created the refpdfsT200 set of pdf's (see section 3.1.2) using data from a 200 K MD simulation. Recall that the potential energy distribution of the Kirkwood samples from this set has high overlap with the canonical distribution at temperatures around 100 K (see Figure 5, inset). The acceptance ratio for a preliminary 10^5 step biased MC simulation at 100 K using refpdfsT200 was found to be 0.05. Pdf set refpdfsT200 could, therefore, be used to set up the reservoir for KR-REX simulation using the lower temperature set where the highest replica temperature is 100 K.

Figure 15 shows the energy distributions obtained from the T-REX and KR-REX simulations for the low temperature set. The distributions from the KR-REX simulation (lines marked by circles in the bottom panel of Figure 15) are substantially closer to the reference distribution than those from the T-REX (top panel of Figure 15) simulation. Table 3 gives the average energy and heat capacity at the simulation temperatures computed from the raw simulation data, without multihistogram analysis. For the lowest temperature replica at 20 K, the average energy from the T-REX simulation deviates from the reference by 0.62 kJ/mol, while the deviation of average energy from KR-REX is much smaller at 0.09 kJ/mol. The heat capacity values did not converge for either run. The number of successful exchanges with the reservoir was over 23 000, consistent with the biased MC acceptance ratio of 0.05. The results show that coupling with the reservoir enhances sampling in the low temperature replicas. In other words, the reservoir can mimic a higher temperature replica.

3.2. Alanine Dipeptide. In this section, we apply biased Monte Carlo (Algorithm 3) to alanine dipeptide. Similar to the model system, the BAT internal coordinate system was used for Kirkwood sampling as described elsewhere.[10] The conformation of the 22-atom molecule was defined using 21 bond lengths, 20 bond angles, and 19 bond torsions. The torsions included the backbone ϕ and ψ angles for the alanine residue. Each coordinate was discretized into 72 equally spaced bins. The range for angular coordinates was $[0,2\pi]$ radians, while that for bond coordinates was [0.6,2] Å. MD data for populating the pdf's was generated by a 1000 ns replica exchange simulation with replica temperatures of 300, 400, and 500 K. In each replica, coordinates were saved at regular intervals to generate 10^6 conformations for each temperature. The all-atom AMBER99SB[52] force field was used, and the simulations were performed with GROMACS 4.6.5.[31] Kirkwood sampling was performed in Matlab[29] and Octave,[30] and force field energy was computed using OpenMM 6.2.[53] In terms of computational cost, most (>90%) of the time was spent in generating the Kirkwood samples. The MATLAB implementation generated approximately 10 conformations per second.

MD data were used to populate the 60 singlet pdf's and 1770 doublet pdf's corresponding to all pairwise combinations of the coordinates. Two separate sets of pdf's, namely, refpdfsT400 and refpdfsT500, were constructed using data from the 400 and 500 K replicas of the MD simulation, respectively. Each pdf set was used to generate 5 million conformations using doublet level Kirkwood sampling. Figure 16 shows the distribution of the potential energy of the Kirkwood samples overlaid on the distribution from the MD replica exchange simulation. Kirkwood samples from the two sets of pdf's have good overlap with the reference distributions for the corresponding temperatures, suggesting overlap between the MD and Kirkwood sampled conformational spaces. The energy distribution for Kirkwood samples follows the reference distribution such that the distribution for refpdfsT400 (red dashed line) is shifted to lower energies as compared to the distribution using refpdfsT500 (blue dashed line), consistent with the observations for the model system.

We performed biased MC simulations to obtain a canonical distribution for temperatures lower than that of the original MD simulation used to populate the reference pdf's. Doublet level Kirkwood sampling is employed to set up the biasing distribution, and 5 × 10^6 MC steps are taken for all simulations

DOI: 10.1021/acs.jpcb.5b01800
J. Phys. Chem. B 2015, 119, 6155−6169

Table 3. Average Energy (kJ/mol), $\langle E \rangle$, and Heat Capacity ($k_B/2$), C, Computed from the Replica Exchange Simulation Using the Low Temperature pdf Set (Section 3.1.4)[a]

	reference		T-REX				KR-REX			
T (K)	$\langle E \rangle$	C	$\langle E \rangle$		C		$\langle E \rangle$		C	
20	−20.75	23.96	−20.13	(0.62)	25.98	(2.02)	−20.66	(0.09)	28.52	(4.56)
30	−19.75	24.88	−19.15	(0.6)	23.28	(−1.6)	−19.63	(0.12)	26.87	(1.99)
50	−17.66	25.69	−17.26	(0.4)	23.5	(−2.19)	−17.56	(0.1)	26.07	(0.38)
100	−12.24	26.8	−12.38	(−0.14)	24.15	(−2.65)	−12.12	(0.12)	27.36	(0.56)

[a]The pdf set refpdfsT200 was used to set up the reservoir in KR-REX simulation. Reference values from Table 1 are also given. The numbers in parentheses are differences with respect to the reference values.

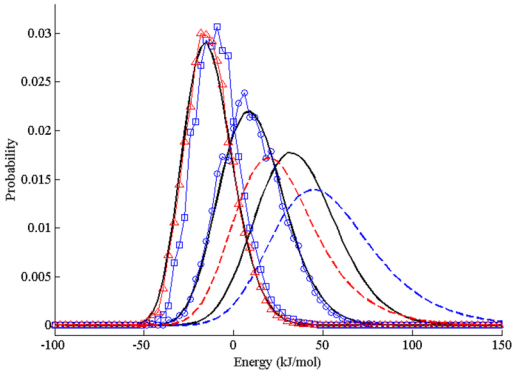


Figure 16. Potential energy distributions for alanine dipeptide. Reference distributions from MD replica exchange simulation are in solid black lines with 300 K being the left-most, followed by 400 and 500 K. The dashed lines show distributions for 5×10^6 Kirkwood samples with red corresponding to refpdfs400 and blue to refpdfs500 pdf sets. The marked lines are distributions from different biased MC simulations. Blue squares and circles correspond to biased MC using refpdfs500 at 300 and 400 K, respectively. Red triangles correspond to 300 K biased MC simulation using refpdfs400.

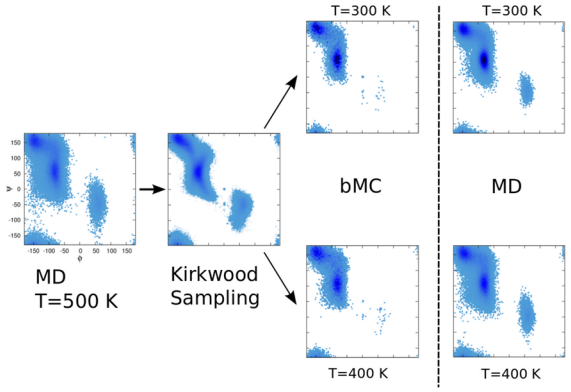


Figure 17. Ramachandran plots for the alanine residue of alanine dipeptide with ϕ torsion on the horizontal axis and ψ on the vertical axis. Both axes range from $-\pi$ to π radians. Each plot shows the normalized distribution computed using 10^5 data points. Darker colors represent higher probability regions. The MD plots employ data from the 1000 ns replica exchange simulation. The other plots employ refpdfsT500 for Kirkwood sampling, which was populated using data from the 500 K MD replica.

in this section. We first present results for 300 and 400 K biased MC simulations with refpdfsT500 used for Kirkwood sampling. Figure 16 shows the energy distribution obtained from the MC simulations overlaid on the reference MD potential energy distributions. The 400 K distribution (blue circles) is in good agreement with the target distribution, unlike the 300 K distribution (blue squares) which is shifted to higher energies. The acceptance ratio for the 400 K simulation was 0.023. In comparison, the acceptance ratio at 300 K was lower (0.007), consistent with the poorer agreement of the 300 K energy distribution.

It is instructive to compare the Ramachandran plots of the alanine residue for ensembles generated by the different sampling methods. In Figure 17, each two-dimensional plot shows normalized joint distribution of ϕ and ψ angles computed using 10^5 conformations. The left-most plot represents the 500 K replica of the 1000 ns MD REX simulation which was used to populate the refpdfsT500 pdf's. The next plot corresponds to Kirkwood sampling using refpdfsT500. The plot for Kirkwood generated conformations is in good agreement with the 500 K MD plot, though the Kirkwood plot has a slightly smaller coverage. Recall that the Kirkwood sampling was performed while accounting for the pairwise correlations of the ϕ and ψ angles with all remaining 59 BAT coordinates. By relaxing some correlations, for example, ones with side chain coordinates, a greater conformational space is likely to become accessible by Kirkwood sampling. Note that the number of conformations (10^5) for the Kirkwood Ramachandran plot is several orders of magnitude smaller than the number of time steps (3×10^9) in the MD REX simulation. The two plots are still comparable because Kirkwood sampling generates uncorrelated samples.

Figure 17 also shows the Ramachandran plot for the biased MC simulations using refpdfsT500 and the reference plots from the MD replica exchange simulation. Biased MC effectively resamples from the Kirkwood distribution to reproduce a target distribution, here the Boltzmann distributions at 300 and 400 K. All regions of the Ramachandran plot that are sampled by the MD simulation are represented in the biased MC simulation. At both temperatures, the basins in the $\phi < 0$ region of the plot, which are highly populated in the MD simulation, are well sampled in the MC simulations. The less populated basins in the $\phi > 0$ region are poorly sampled by the biased MC simulations, and may explain the deviations in the energy distribution (Figure 16). Nevertheless, in spite of the low acceptance ratio, since all regions are sampled in the biased MC simulation, it would be advantageous to combine Kirkwood sampling with MD or MC based local sampling, as demonstrated for the model system above.

Finally, we performed a T = 300 K biased MC simulation using the refpdfsT400 pdf set which was populated using conformations from the 400 K MD replica. An acceptance ratio of 0.024 was obtained in a 5×10^6 step simulation. The energy distribution of the MC samples using refpdfsT400 (marked by red triangles in Figure 16) is in much better agreement with the reference distribution as compared to that using refpdfsT500

DOI: 10.1021/acs.jpcb.5b01800
J. Phys. Chem. B 2015, 119, 6155−6169

(marked by blue boxes). Note that the acceptance ratio for T = 400 K biased MC simulation using refpdfsT500 is comparable to that of T = 300 K simulation using refpdfsT400. One could envision a multistage simulation to enable simulating arbitrarily low temperature where each stage would consist of biased MC simulation followed by repopulation of the pdf's using biased MC sampled conformations. The successive stage would simulate lower temperature with the first stage employing pdf's populated by high temperature MD or other enhanced sampling methods.

4. SUMMARY AND DISCUSSION

In this contribution, we have presented two methods for barrierless equilibrium sampling of molecular systems. Our approach builds upon Kirkwood sampling, which employs low-order correlations among internal coordinates of a molecule for random sampling of the conformation space. Both methods make use of the property of Kirkwood sampling whereby the normalized probability of generating a given conformation can be computed. We have presented proof-of-concept results for the new methods using a model system with nine atoms where the intramolecular force field can be adjusted to highlight particular sampling issues. We also showed results for alanine dipeptide, a commonly used model system for benchmarking sampling algorithms.

The first of the two algorithms is based on biased Monte Carlo where the Kirkwood samples are used as MC moves. In contrast to standard molecular dynamics or Monte Carlo simulations, no equilibration is required in the biased MC simulation, since successive Kirkwood samples are generated independently. As a result, biased MC using Kirkwood moves can be trivially parallelized in a distributed computing environment, with no communication required between the compute nodes. Furthermore, since Kirkwood sampling is a geometrical sampling method, independent of the energy landscape, the same set of samples can be used to generate Boltzmann distributions for different temperatures and potential energy functions.

The convergence of a biased MC simulation depends primarily on the overlap of the Kirkwood sampling distribution and the target Boltzmann distribution. The conformational space covered by Kirkwood sampling is determined by the input set of probability distribution functions. In the present work, the pdf's were populated using conformations obtained from MD simulations. The results showed that Kirkwood samples could be used to generate a Boltzmann distribution, not only for the temperature of the original MD simulation but also for lower temperatures, although the convergence slows as the temperature is reduced. One can imagine an iterative scheme where the initial set of pdf's is constructed in a manner that provides coverage of a wide conformational space. For instance, the pdf's could be populated using high temperature MD, or using a database of conformations from the PDB or a fragment pdf library. Given this initial set of pdf's and a potential energy function, one could then perform successive stages of biased MC simulations and repopulation of the pdf's to reach arbitrarily low temperatures.

The second algorithm introduced in this work is a modification of temperature replica exchange where a temperature replica is coupled with a Kirkwood reservoir of conformations. Exchanges with the reservoir help to enhance sampling for that replica. As for biased MC, the acceptance ratio for exchanges with the reservoir is determined by the overlap of the Kirkwood distribution and the Boltzmann distribution for the coupled replica. Note that in the absence of a reservoir the highest temperature in a replica exchange simulation needs to be high enough to overcome the energy barriers and avoid trapping. By coupling to a Kirkwood reservoir, the highest temperature can be set independent of the energy barriers. In the presence of a Kirkwood reservoir, the criterion for setting the highest temperature becomes the requirement of sufficient overlap to facilitate frequent exchanges with the reservoir. Thus, coupling with a reservoir effectively places a limit on the highest temperature required, which may be lower than the highest temperature dictated by the barriers of the PES. Indeed, if the reservoir has good overlap with the Boltzmann distribution corresponding to the temperature of interest, then just a single replica would suffice. In that case, the temperature replica essentially performs local sampling, while the reservoir facilitates global sampling of the conformational space. Finally, we also note that the sampling for the lowest temperature replicas can be enhanced by coupling to a reservoir constructed from a set of low energy minima from the potential energy landscape.[54] In this scenario, Kirkwood samples could be used for seeding basin-hopping[33−35] simulations in order to generate the low energy minima of the system, or in the MC part of a basin-sampling calculation.[55]

In the present work, canonical sampling for the temperature replicas of a KR-REX simulation was performed by Monte Carlo. The Kirkwood reservoir can also be coupled with temperature replicas evolving via heat bath coupled molecular dynamics. Coupling to an MD replica would require a relatively minor modification of the current algorithm to assign velocities to the reservoir generated conformations. The velocities would be sampled from the Maxwell−Boltzmann[1] distribution corresponding to the temperature of the coupled temperature replica. It should therefore be straightforward to enhance the performance of existing MD based replica exchange codes by coupling them to a Kirkwood sampler.

Kirkwood sampling provides a normalized probability distribution over the full conformational space. Therefore, it can be Boltzmann inverted[56] to construct a reference energy function. We can then define intermediate potentials, which progress from the potential energy of the physical force field to the Kirkwood reference energy. The intermediate potentials could be used to perform Hamiltonian replica exchange[57] with the two boundary replicas corresponding to the reference and physical energy functions. All replicas may now be simulated at the temperature of interest using standard perturbation move Monte Carlo. This approach would facilitate coupling of the Kirkwood sampling with existing Monte Carlo codes.[51,58] The number and spacing of the intermediate replicas will depend on the overlap between the reference and the physical energy surfaces. We note that molecular dynamics cannot be used for canonical simulation with Hamiltonian replicas, since the reference energy surface is piecewise continuous and its gradients are not well-defined. This limitation arises because the Kirkwood distribution is constructed over a discretized conformational space. Note that in this scheme Kirkwood sampling is only used for defining a reference energy function and not for generating MC moves. We are currently working on Hamiltonian replica exchange using a Kirkwood reference energy and will present the results in a separate publication. Finally, nonequilibrium simulations, which switch the Hamiltonian between the physical and reference energy func-

DOI: 10.1021/acs.jpcb.5b01800
J. Phys. Chem. B 2015, 119, 6155−6169

tions,[6,27,59] can also be used to enhance replica overlap and reduce the number of intermediate replicas.

Application of the present methods to systems of practical interest—from small drug-like molecules to large proteins—would require a flexible and scalable implementation of Kirkwood sampling. Some of the strategies that could be pursued in this direction are as follows. In the present work, the low order pdf's were populated using conformations for the full molecule. The pdf's can also be generated for smaller molecular fragments and assembled for an arbitrary molecule. We are currently working on developing such a fragment pdf library for proteins using peptide fragments up to five residues long. The fragment pdf's will be populated using PDB[8] data and/or exhaustive MD simulations. For small fragments, it is computationally feasible to calculate triplet or even quadruplet pdf's, which would help better account for local packing. We note that the Kirkwood framework can be used to sample a subset of coordinates while keeping the other coordinates fixed. In the context of conformational sampling of polymers, local packing can be accounted for by incorporating correlations among coordinates of adjacent subunits. Therefore, Kirkwood sampling is likely to be particularly useful for sampling small molecules, short peptides, or for sampling side-chain conformations for a fixed backbone.

Construction of pdf libraries for proteins is relatively straightforward, since proteins are built from a fixed set of amino acid residues. In the case of small molecules, it may be necessary to enumerate the different possible bonded topologies for a given number of atoms and atom types. Certain simplifications, such as treating rings or other groups as rigid bodies,[60] may also be beneficial. For certain families of molecules, such as macrocycles, it may be necessary to explore alternative internal coordinate systems that better capture the constraints imposed by the cyclic bonded topologies. We note that auxiliary variables could be used in the Kirkwood framework to impose constraints on the samples. For example, for the system studied in this work, by using end-to-end distance as a sampled variable, one could generate samples with a given end-to-end distance.

In principle, Kirkwood sampling can be performed using any number of pdf's of different orders. In practice, however, it is desirable to use as few pdf's as possible, since the computational cost is proportional to the number of pdf's used and also the accessible conformational space reduces with increasing pdf's. Therefore, it would be useful to develop an adaptive scheme that successively adds pdf's for different combinations of variables to achieve a certain overlap with a target distribution. The overlap could be measured in terms of the acceptance ratio in a biased MC simulation. Finally, we note that Kirkwood sampling is a general method for sampling points in high dimensional spaces and methods presented here may be applicable to other areas which deal with high dimensional probability distributions such as machine learning.

■ ASSOCIATED CONTENT

Supporting Information

Input files for the model system and alanine dipeptide. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.5b01800.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ssomani@its.jnj.com.

Present Address

#(S.S.) Janssen Research and Development, LLC, 1400 McKean Road, Spring House, PA 19477, USA.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was funded by the European Research Council and EPSRC grant EP/I001352/1. Y.O. was supported, in part, by the JSPS Grant-in-Aid for Scientific Research on Innovative Areas ("Dynamical Ordering and Integrated Functions").

■ REFERENCES

(1) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*; Academic Press: San Diego, CA, 2002.
(2) Wales, D. J. *Energy Landscapes*; Cambridge University Press: Cambridge, U.K., 2003.
(3) Mitsutake, A.; Sugita, Y.; Okamoto, Y. Generalized-Ensemble Algorithms for Molecular Simulations of Biopolymers. *Biopolymers* **2001**, *60*, 96−123.
(4) Orozco, M. A Theoretical View of Protein Dynamics. *Chem. Soc. Rev.* **2014**, *43*, 5051−5066.
(5) Betancourt, M. R. Optimization of Monte Carlo Trial Moves for Protein Simulations. *J. Chem. Phys.* **2011**, *134*, 014104+.
(6) Nilmeier, J. P.; Crooks, G. E.; Minh, D. D. L.; Chodera, J. D. Nonequilibrium Candidate Monte Carlo is an Efficient Tool for Equilibrium Simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, E1009−E1018.
(7) Somani, S.; Killian, B. J.; Gilson, M. K. Sampling Conformations in High Dimensions Using Low-dimensional Distribution Functions. *J. Chem. Phys.* **2009**, *130*, 134102.
(8) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(9) Allen, F. H. The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. *Acta Crystallogr.* **2002**, *58*, 380−388.
(10) Somani, S.; Gilson, M. K. Accelerated Convergence of Molecular Free Energy via Superposition Approximation-based Reference States. *J. Chem. Phys.* **2011**, *134*, 134107.
(11) Hermans, J. The Amino Acid Dipeptide Small But Still Influential After 50 Years. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3095−3096.
(12) Tobias, D. J.; Brooks, C. L. Conformational Equilibrium in the Alanine Dipeptide in the Gas Phase and Aqueous Solution: A Comparison of Theoretical Results. *J. Phys. Chem.* **1992**, *96*, 3864−3870.
(13) Somani, S. Conformational Sampling and Calculation of Molecular Free Energy Using Superposition Approximations. Ph.D. Thesis, University of Maryland, College Park, 2011.
(14) Chang, C.-E.; Potter, M. J.; Gilson, M. K. Calculation of Molecular Configuration Integrals. *J. Phys. Chem. B* **2003**, *107*, 1048−1055.
(15) Miller, M. A.; Amon, L. M.; Reinhardt, W. P. Should One Adjust the Maximum Step Size in a Metropolis Monte Carlo Simulation? *Chem. Phys. Lett.* **2000**, *331*, 278−284.
(16) Swendsen, R.; Wang, J.-S. Replica Monte-Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607−2609.
(17) Hukushima, K.; Nemoto, K. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604−1608.
(18) Sugita, Y.; Okamoto, Y. Replica-exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(19) Earl, D.; Deem, M. W. Parallel Tempering: Theory, Applications, and New Perspectives. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910−3916.

DOI: 10.1021/acs.jpcb.5b01800
J. Phys. Chem. B 2015, 119, 6155−6169

(20) Hansmann, U. Parallel Tempering Algorithm for Conformational Studies of Biological Molecules. *Chem. Phys. Lett.* **1997**, *281*, 140−150.
(21) Nadler, W.; Hansmann, U. H. E. Optimizing Replica Exchange Moves for Molecular Dynamics. *Phys. Rev. E* **2007**, *76*, 057102.
(22) Brenner, P.; Sweet, C. R.; VonHandorf, D.; Izaguirre, J. A. Accelerating the Replica Exchange Method Through an Efficient All-pairs Exchange. *J. Chem. Phys.* **2007**, *126*, 074103.
(23) Chodera, J. D.; Shirts, M. R. Replica Exchange and Expanded Ensemble Simulations As Gibbs Sampling: Simple Improvements for Enhanced Mixing. *J. Chem. Phys.* **2011**, *135*, 194110.
(24) Bennett, C. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(25) Pohorille, A.; Jarzynski, C.; Chipot, C. Good Practices in Free-Energy Calculations. *J. Phys. Chem. B* **2010**, *114*, 10235−10253.
(26) Shirts, M. R. Simple Quantitative Tests to Validate Sampling from Thermodynamic Ensembles. *J. Chem. Theory Comput.* **2012**, *9*, 909−926.
(27) Ballard, A.; Jarzynski, C. Replica Exchange with Nonequilibrium Switches: Enhancing Equilibrium Sampling by Increasing Replica Overlap. *J. Chem. Phys.* **2012**, *136*, 194101.
(28) Crooks, G. E. Entropy Production Fuctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences. *Phys. Rev. E* **1999**, *60*, 2721−2726.
(29) *Matlab*, version 8.4 (R2014a); The MathWorks Inc.: 2014.
(30) Eaton, J. W.; Bateman, D.; Hauberg, S. GNU Octave: a high-level interactive language for numerical computations, 2008.
(31) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. GROMACS 4.5: A High-throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(32) Case, D.; et al. *Amber 12*; University of California: San Francisco, CA, 2010.
(33) Li, Z.; Scheraga, H. A. Monte Carlo-minimization Approach to the Multiple-minima Problem In Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611−6615.
(34) Wales, D. J.; Doye, J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A* **1997**, *101*, 5111−5116.
(35) Wales, D. J.; Scheraga, H. A. Global Optimization of Clusters, Crystals, and Biomolecules. *Science* **1999**, *285*, 1368−1372.
(36) Trygubenko, S. A.; Wales, D. J. A Doubly Nudged Elastic Band Method for Finding Transition States. *J. Chem. Phys.* **2004**, *120*, 2082−2094.
(37) Trygubenko, S. A.; Wales, D. J. Erratum: A Doubly Nudged Elastic Band Method for Finding Transition States [J. Chem. Phys. 120, 2082 (2004)]. *J. Chem. Phys.* **2004**, *120*, 7820−7820.
(38) Henkelman, G.; Jónsson, H. A Dimer Method for Finding Saddle Points on High Dimensional Potential Surfaces Using Only First Derivatives. *J. Chem. Phys.* **1999**, *111*, 7010−7022.
(39) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.* **2000**, *113*, 9901−9904.
(40) Henkelman, G.; Jónsson, H. Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points. *J. Chem. Phys.* **2000**, *113*, 9978−9985.
(41) Munro, L. J.; Wales, D. J. Defect Migration in Crystalline Silicon. *Phys. Rev. B* **1999**, *59*, 3969−3980.
(42) Becker, O. M.; Karplus, M. The Topology of Multidimensional Potential Energy Surfaces: Theory and Application to Peptide Structure and Kinetics. *J. Chem. Phys.* **1997**, *106*, 1495−1517.
(43) Wales, D. J.; Miller, M. A.; Walsh, T. R. Archetypal Energy Landscapes. *Nature* **1998**, *394*, 758−760.
(44) Ferrenberg, A. M.; Swendsen, R. H. New Monte Carlo Technique for Studying Phase Transitions. *Phys. Rev. Lett.* **1988**, *61*, 2635−2638.
(45) Ferrenberg, A. M.; Swendsen, R. H. Optimized Monte Carlo Data Analysis. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.
(46) Labastie, P.; Whetten, R. L. Statistical Mechanics of the Cluster Solid-liquid Transition. *Phys. Rev. Lett.* **1990**, *65*, 1567−1570.
(47) Weerasinghe, S.; Amar, F. G. Absolute Classical Densities of States for Very Anharmonic Systems and Applications to the Evaporation of Rare Gas Clusters. *J. Chem. Phys.* **1993**, *98*, 4967.
(48) Calvo, F.; Labastie, P. Configurational Density of States from Molecular Dynamics Simulations. *Chem. Phys. Lett.* **1995**, *247*, 395−400.
(49) Ytreberg, F. M.; Zuckerman, D. M. A Black-box Re-weighting Analysis Can Correct Flawed Simulation Data. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 7982−7987.
(50) Bishop, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 1st ed.; Springer: 2007.
(51) Wales, D. J. GMIN: A Program for Basin-hopping Global Optimisation, Basin-sampling, and Parallel Tempering.
(52) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(53) Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; et al. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2013**, *9*, 461−469.
(54) Sharapov, V. A.; Meluzzi, D.; Mandelshtam, V. A. Low-temperature Structural Transitions: Circumventing the Broken-ergodicity Problem. *Phys. Rev. Lett.* **2007**, *98*, 105701.
(55) Wales, D. J. Surveying a Complex Potential Energy Landscape: Overcoming Broken Ergodicity Using Basin-sampling. *Chem. Phys. Lett.* **2013**, *584*, 1−9.
(56) Ytreberg, F. M.; Zuckerman, D. M. Simple Estimation of Absolute Free Energies for Biomolecules. *J. Chem. Phys.* **2006**, *124*, 104105+.
(57) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional Replica-exchange Method for Free-energy Calculations. *J. Chem. Phys.* **2000**, *113*, 6042−6051.
(58) Jorgensen, W. L.; Tirado-Rives, J. Molecular Modeling of Organic and Biomolecular Systems Using boss and mcpro. *J. Comput. Chem.* **2005**, *26*, 1689−1700.
(59) Ballard, A.; Jarzynski, C. Replica Exchange with Nonequilibrium Switches. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 12224−12229.
(60) Kusumaatmaja, H.; Whittleston, C. S.; Wales, D. J. A Local Rigid Body Framework for Global Optimization of Biomolecules. *J. Chem. Theory Comput.* **2012**, *8*, 5159−5165.

DOI: 10.1021/acs.jpcb.5b01800
J. Phys. Chem. B 2015, 119, 6155−6169