Conformational Populations of Ligand-Sized Molecules by Replica Exchange Molecular Dynamics and Temperature Reweighting

HISASHI OKUMURA,* EMILIO GALLICCHIO, RONALD M. LEVY

Department of Chemistry and Chemical Biology and BioMaPS Institute for Quantitative Biology, Rutgers University, Piscataway, New Jersey 08854

Abstract: The use of the replica exchange (RE) molecular dynamics (MD) method for the efficient estimation of conformational populations of ligand-sized molecules in solution is investigated. We compare the computational efficiency of the traditional constant temperature MD technique with that of the parallel RE molecular dynamics method for a series of alkanes and rilpivirine (TMC278), an inhibitor against HIV-1 reverse transcriptase, with implicit solvation. We show that conformational populations are accurately estimated by both methods; however, replica exchange estimates converge at a faster rate, especially for rilpivirine, which is characterized by multiple stable states separated by high-free energy barriers. Furthermore, convergence is enhanced when the weighted histogram analysis method (WHAM) is used to estimate populations from the data collected from multiple RE temperature replicas. For small drug-like molecules with energetic barriers separating the stable states, the use of RE with WHAM is an efficient computational approach for estimating the contribution of ligand conformational reorganization to binding affinities.

© 2009 Wiley Periodicals, Inc.   J Comput Chem 31: 1357–1367, 2010

Key words: replica-exchange method; molecular dynamics simulation; conformational population; ligand-sized molecule; TMC278

Introduction

Constant temperature molecular dynamics (MD) conformational sampling of systems with rugged potential energy landscapes is problematic due to trapping in local minimum free-energy states. This is often manifested in poor coverage of conformational space and slow rates of convergence of thermodynamic quantities. Generalized ensemble conformational sampling algorithms[1–17] can circumvent these difficulties by adopting sampling distributions with more favorable characteristics than traditional methods. One of the best known generalized ensemble algorithms is the temperature replica-exchange (RE) method.[1,2] The RE method is a well-established and powerful technique to sample the rough energy landscapes of biomolecules, allowing rapid interconversion between conformations separated by high energy barriers.

In the most common implementations,[18–20] several replicas of the system are run in parallel over a series of temperatures using constant temperature MD. Periodically, exchanges of conformations are attempted between pairs of replicas, using a scheme designed to preserve canonical sampling at each temperature. Unlike conformational search techniques, temperature RE can be used to calculate thermodynamic quantities, such as conformational free energies, at each simulated temperature. Finally, the RE method is inherently a parallel algorithm well suited to take advantage of the large number of processors in modern computing clusters.[21]

Although RE is a parallel simulation technique that is relatively straightforward to implement, there is still much to learn about how best to use RE simulations in computational biophysics. In some circumstances, questions have been raised about the efficiency of the algorithm relative to traditional constant temperature MD.[22–24] The efficiency of temperature replica exchange (RE) simulations hinges on their ability to enhance conformational sampling at physiological temperature by taking advantage of more rapid conformational interconversions at higher temperatures. Studies from our laboratory have pointed out that processes characterized by anti-Arrhenius kinetics behavior, such as protein folding, represent a challenge for temperature RE.[25,26] Contrary to normal Arrhenius behavior, whereby the rates of kinetic processes increase with increasing

temperature, anti-Arrhenius kinetics is characteristic of processes with smaller transition rates at higher temperatures. In these cases, RE cannot take advantage of the normally faster kinetics at high temperatures to help convergence at low temperatures.

Additionally, RE incurs a higher computational cost than constant temperature MD if only data at one temperature is used to estimate populations at that temperature. This is because all of the samples from a constant temperature simulation can be directly used to calculate physical quantities at the set temperature. On the other hand, a large fraction of the computational effort of RE is devoted to simulating the system at temperatures other than the temperature of interest, thereby producing mainly data, which may or may not be useful in the estimation of physical quantities at the temperature of interest. Consequently, it is often unclear whether the enhancements of conformational sampling obtainable with RE are sufficient to offset the higher computational costs in order for RE to yield higher computationally efficiency than MD.

In this article, we present an analysis of the relative efficiency of temperature RE and constant temperature MD with respect to the computation of the conformational populations of a series of small molecules with implicit solvation. The study of the distribution of conformations of ligand-sized molecules in water is important for the estimation of the reorganization free energy component of the binding free energies of protein–ligand complexes. The phenomenon of conformational reorganization, sometimes referred to as conformational selection or induced fit,[27] is recognized as a key factor to understand drug binding affinity, specificity, and resistance. It generally refers to the free energy penalty incurred by the binding partners for reorganizing the distribution of conformations present in their unbound forms to those compatible with complexation.

The ligand component ΔF of the binding reorganization free energy is directly related to the population p in solution of the conformational state corresponding to the bound conformation by means of the equation

$$\Delta F = -k_{\mathrm{B}} T \log p, \tag{1}$$

where k_{B} is the Boltzmann constant. Thus, the estimation of ligand conformational free energies rests on the ability to obtain converged values of ligand conformational populations.

The systems we study in this work are a series of alkanes and rilpivirine[28] (also known as TMC278, the name that will be used in the following), a potent inhibitor of reverse transcriptase enzyme of the human HIV virus. We perform MD and temperature RE simulations in vacuum and implicit solvation with the same total number of steps for these molecules and compare their computational efficiency defined in terms of the reduction of the statistical uncertainty of conformational populations at equal computational cost. We also test the applicability of the temperature Weighted Histogram Analysis Method (T-WHAM),[29] a method that allows the utilization of data from multiple temperature replicas to improve the estimates of conformational populations at the temperature of interest. We find that RE, and especially RE in conjunction with WHAM, is more computationally efficient than constant temperature MD in converging the conformational populations of these small molecular
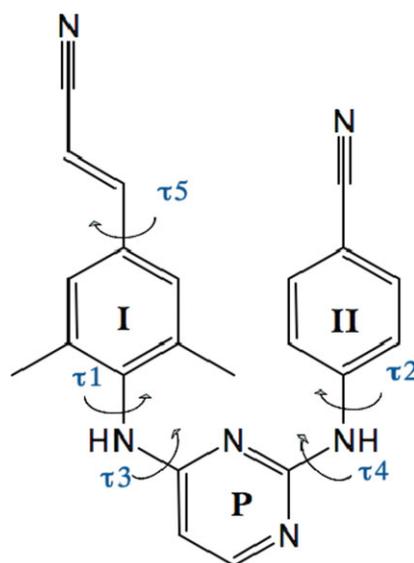

Figure 1. Chemical structure of TMC278. Wing I, wing II, and pyrimidine rings labeled as I, II, and P, respectively. Black arrows correspond to rotations associated with torsion angles $\tau_1 - \tau_5$.

systems. We rationalize this finding based on the relatively small number of degrees of freedom of these molecules limiting the rate of expansion of conformational space with increasing temperature which, for larger systems such as peptides and proteins, causes difficulties for RE sampling. In addition, the energetic, rather than entropic, nature of the free energy barriers between the stable states of these molecules allows more rapid conformational sampling at higher temperatures further enhancing the convergence rate. The WHAM method is found to be particularly helpful for estimating accurately small populations corresponding to high-energy states which become more probable with increasing temperature.

In Methods section, computational details of the serial canonical MD simulations and the RE simulations are presented.

Methods

We performed MD simulations of butane, pentane, and hexane in vacuum and TMC278 with AGBNP implicit solvation.[30] The numbers of dihedral angles excluding methyl groups rotations are one for butane, two for pentane, and three for hexane. The chemical structure of TMC278, which contains five rotatable bonds is shown in Figure 1. The OPLS 2005 force field was used.[31] We used the Nosé-Hoover thermostat,[32–34] whose equations of motion are as follows:

$$\dot{\boldsymbol{r}}_i = \frac{\boldsymbol{p}_i}{m_i}, \tag{2}$$

$$\dot{\boldsymbol{p}}_i = \boldsymbol{F}_i - \zeta \boldsymbol{p}_i, \tag{3}$$

$$\dot{\zeta} = \frac{g k_{\mathrm{B}}}{Q} \left(\mathscr{T}(t) - T_0 \right), \tag{4}$$

where $\boldsymbol{r}_i$, $\boldsymbol{p}_i$, m_i, and $\boldsymbol{F}_i$ are coordinates, momentum, mass, and force of atom i, respectively. The variable ζ is a "viscosity" parameter for temperature control. The constant g is related to the number of atoms N as $g = 3N - 6$. The constant Q is the artificial "mass" for the thermostat. $\mathcal{T}(t)$ and T_0 are the instantaneous temperature and the set temperature, respectively. We integrated the equations of motion in eqs. (2)–(4) by the time-reversible algorithm by Martyna *et al.*[35] We used the relation $Q = gk_BT_0\tau^2$ for the Nosé-Hoover thermostat[36] with a relaxation time of $\tau = 10$ fs. The MD time step was set to $\Delta t = 0.5$ fs.

Serial molecular dynamics simulations were conducted at $T_0 = 300$ K. We performed equilibration runs for 20 ns and then sampled conformations for 80 ns for the alkanes. For TMC 278, we performed an equilibration run for 50 ns and sampled conformations for 200 ns. Parallel replica exchange MD simulations were performed with four replicas at 300 K, 350 K, 400 K, and 450 K for the alkanes and with eight replicas at 300 K, 350 K, 400 K, 450 K, 500 K, 550 K, 600 K, and 650 K for TMC278. We performed REMD for 20 ns after 5 ns equilibration for the alkanes and for 25 ns after 6.25 ns equilibration runs for TMC278. The total sampling times are 80 ns for each alkane and 200 ns for TMC278. These sampling times are the same as for the corresponding serial MD simulations. Temperature exchanges were attempted 1 ps between adjacent temperatures and were accepted with probability

$$w = \min\{1, \exp[-(\beta_i - \beta_j)(E_i - E_j)]\}, \quad (5)$$

where $\beta_i = 1/k_BT_i$ and $\beta_j = 1/k_BT_i$ are the inverse temperature before the exchange and E_i and E_j are the potential energies of replica i and j, respectively.

Dihedral angle distributions at a particular temperature can be calculated by simply binning the data from only the corresponding replica. We also calculated the dihedral angle distributions using the temperature WHAM method[29,37]:

$$P_\beta(\phi) = \sum_E \frac{\sum_{i=1}^{M} N_i(E, \phi)e^{-\beta E}}{\sum_{i=1}^{M} n_i e^{f_i - \beta_i E}}, \quad (6)$$

$$e^{-f_i} = \sum_\phi P_{\beta_i}(\phi), \quad (7)$$

where $P_\beta(\phi)$ is the unnormalized angle distribution at inverse temperature β, M is the number of replicas, $N_i(E, \phi)$ is the histogram of potential energy, E, and dihedral angles $\phi = (\phi_1, \phi_2, \ldots)$ at temperature T_i, and n_i is the total number of samples at temperature T_i. eqs. (6) and (7) are solved iteratively.[29]

Dihedral angle distributions for the alkanes were also obtained by numerical grid integration of the corresponding partition functions. Let us suppose a model of an alkane with fixed bond lengths. This model has only internal degrees of freedom of dihedral angles $\phi_0, \phi_1, \ldots, \phi_{n+1}$ and bond angles $\theta_1, \theta_2, \ldots \theta_{n+1}$, where the number of dihedral angles n is $n = 1$ for butane, $n = 2$ for pentane, and $n = 3$ for hexane. ϕ_0 and ϕ_{n+1} are

the dihedral angles of the ends of the alkane chain, which include hydrogen atoms. Under these assumptions, the partition function Z of the molecule is given by

$$Z = \int d\phi_0 \cdots d\phi_{n+1} \int d\theta_1 \cdots d\theta_{n+1} \prod_{i=1}^{n+1} \sin \theta_i \exp\{-\beta E(\boldsymbol{\phi}, \boldsymbol{\theta})\}, \quad (8)$$

where the bold letters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ denote the set of dihedral angles and bond angles, respectively. $\prod_{i=1}^{n+1} \sin \theta_i$ is the Jacobian for the $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ internal coordinates. The dihedral angle distribution $P(\phi_1, \ldots, \phi_n)$ is calculated from the partition function Z by

$$P(\phi_1, \ldots, \phi_n) = \frac{\int d\phi_0\, d\phi_{n+1} \int d\theta_1 \cdots d\theta_{n+1} \prod_{i=1}^{n+1} \sin \theta_i \exp\{-\beta E(\boldsymbol{\phi}, \boldsymbol{\theta})\}}{\int d\phi_0 \cdots d\phi_{n+1} \int d\theta_1 \cdots d\theta_{n+1} \prod_{i=1}^{n+1} \sin \theta_i \exp\{-\beta E(\boldsymbol{\phi}, \boldsymbol{\theta})\}}. \quad (9)$$

The grid spacing of dihedral angles was set to $10°$. Two grid points θ_{eq} and $\theta_{eq} + 4°$, where θ_{eq} is the minimum energy bond angle, were used for the bond angles integration (the potential energy of the *cis* conformation of butane is smallest for $\theta = \theta_{eq} + 4°$).

Results and Discussion

Alkanes

The torsional potential of mean force of butane

$$F(\phi_1) = -k_BT_0 \log\{P(\phi_1)/P(180°)\}, \quad (10)$$

derived from the probability distribution $P(\phi_1)$ computed from the REMD data and WHAM is shown in Figure 2. Note that as defined $F(\phi_1)$ is set to zero at $\phi_1 = 180°$. The torsional potential of mean force of butane shows (Fig. 2) the characteristic minima in correspondence with the trans ($\phi_1 = 180°$) and gauche ($\phi_1 = -60, 60°$) conformations.

The probability of each state p was calculated by integrating the dihedral-angle distribution function. For example, the population p_T of the trans state for butane is given by:

$$p_T = \int_{\phi_1 \in \text{Trans}} d\phi_1 P(\phi_1), \quad (11)$$

where this integral has been performed in the range of $0° \leq \phi_1 < 120°$ for the G− state, in the range of $120° \leq \phi_1 < 240°$ for the T state, and in the range of $240° \leq \phi_1 < 360°$ for the G+ state. The computed populations for the stable states of butane, pentane, and hexane are listed in Tables 1–3, respectively. The overall probability of the symmetry-related conformations are
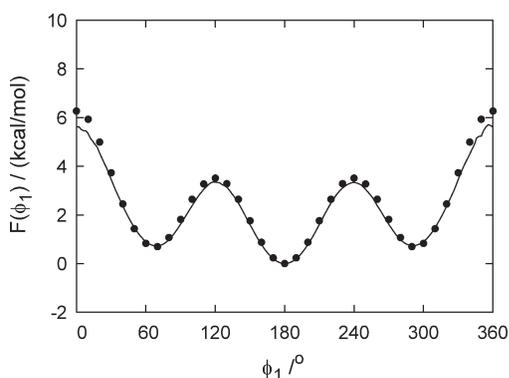
Figure 2. Potential of mean force along the dihedral angle of butane obtained by replica exchange MD simulation with WHAM (line). The filled circles correspond to numerical integration results.

given in these tables, that is, the probability of the gauche structure of butane, for example, is the sum of the populations of the $G+$ and $G-$ conformations.

The distributions obtained from the simulations of the alkanes and the populations derived from them were found to be in agreement with numerical integration calculations to within 1% indicating that the Nosé-Hoover thermostat yields correct canonical conformational sampling for these small molecules. We found instead that the Berendsen thermostat[38,39] (also known as the weak coupling thermostat) can introduce noticeable artifacts for these small molecules. As shown in Figure 3 for butane, the Berendsen thermostat produces distributions which differ substantially from those obtained with the Nosé-Hoover thermostat and by numerical integration, especially for temperature relaxation times shorter than 100 fs. The same phenomenon was observed for the other alkanes and with both MD and REMD sampling. We observed that for these systems, the Berendsen thermostat tends to cause trapping in minimum energy conformations accumulating kinetic energy in the vibration of bonds involving hydrogen atoms. Breakdown of the equipartition of kinetic energy with the Berendsen thermostat has been previously observed.[40] These problems with the Berendsen thermostat are alleviated by using longer temperature relaxation times (Fig. 3) at the expense of higher temperature fluctuations. These results indicate that the Berendsen thermostat is not suitable for the present application, which requires accurate canonical conformational sampling of molecular systems with relatively few degrees of freedom.

The computed population of the conformational states of butane consistently agree among serial MD, REMD without

Table 1. Populations of the Conformational States of Butane at $T = 300$ K.

ϕ_1	MD	RE	RE+WHAM
Trans	0.66(7)	0.63(3)	0.63(3)
Gauche	0.34(7)	0.37(3)	0.37(3)

The values in parenthesis indicate the uncertainty on the last digit.

Table 2. Populations of the Conformational States of Pentane at $T = 300$ K.

ϕ_1	ϕ_2	MD	RE	RE+WHAM
Trans	Trans	0.49(5)	0.47(3)	0.47(3)
Trans	Gauche+	0.45(5)	0.47(3)	0.48(2)
Gauche+	Gauche+	0.050(13)	0.050(6)	0.051(4)
Gauche+	Gauche−	0.0073(16)	0.0081(9)	0.0084(6)

The values in parenthesis indicate the uncertainty on the last digit.

WHAM (denoted as RE in the following), and REMD with WHAM (RE+WHAM). However, differences in the values of the statistical uncertainties, which we take as a measure of computational efficiency, are noted. Figure 4 shows the running averages of the population for each conformational state of butane with the corresponding error bars. To estimate statistical uncertainties, we divide the production run into five blocks, and the uncertainties are estimated from the standard deviation of the five averages from each block. For example, the uncertainties of population at 10 ns were estimated by dividing the first 10 ns into five blocks of 2 ns each. The data listed in all Tables 1–3 were obtained by dividing all 80 ns data into five segments of 16 ns each. Figure 4 shows that the magnitude of the uncertainties follows the order (from smallest to largest): RE+WHAM, RE, and serial MD. The relative uncertainties ($\delta p/p$) of the G state, for example, are 7.1% for RE+WHAM, 8.1% for RE, and 21% for serial MD. In this case, the RE uncertainty is approximately two times smaller than serial MD, despite the fact that the number of data points from the RE simulation (without WHAM) is only one fourth of the data from the serial canonical MD simulation (the length of the RE simulation with four replicas was set to one quarter of that of the MD simulation to compare the two at the same overall computational cost). The faster convergence observed with RE resides in its ability to overcome energy barriers more frequently than MD thereby reaching faster equilibrium between the conformational basins. The higher statistical efficiency of RE outweighs in this case the availability of fewer samples. The further enhancement of con-

Table 3. Populations of the Conformational States of Hexane at $T = 300$ K.

ϕ_1	ϕ_2	ϕ_3	Serial MD	RE	RE+WHAM
Trans	Trans	Trans	0.350(43)	0.360(28)	0.356(22)
Trans	Trans	Gauche+	0.320(22)	0.328(21)	0.330(15)
Trans	Gauche+	Trans	0.158(26)	0.139(17)	0.145(13)
Trans	Gauche+	Gauche+	0.072(16)	0.071(9)	0.067(6)
Gauche+	Trans	Gauche+	0.042(17)	0.041(9)	0.041(8)
Gauche+	Trans	Gauche−	0.037(11)	0.038(4)	0.039(3)
Trans	Gauche+	Gauche−	0.0121(13)	0.0115(20)	0.0113(9)
Gauche+	Gauche+	Gauche+	0.0060(25)	0.0100(17)	0.0087(9)
Gauche+	Gauche+	Gauche−	0.0023(7)	0.0021(9)	0.0022(3)
Gauche+	Gauche−	Gauche+	0.00008(3)	0.00010(10)	0.000069(18)

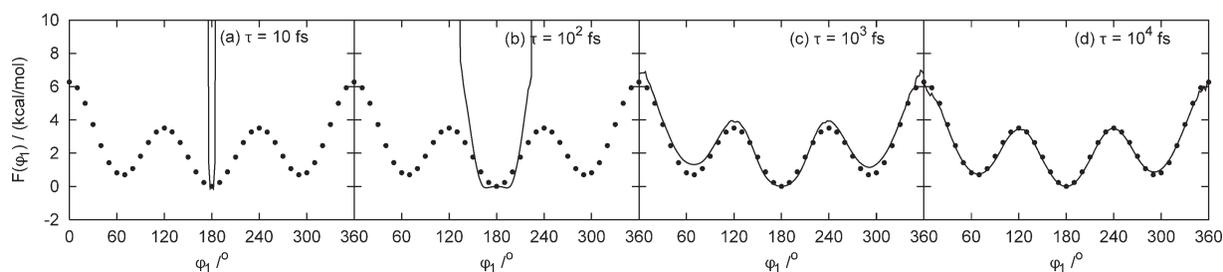The values in parenthesis indicate the uncertainty on the last digit.

Figure 3. Potential of mean force of the dihedral angle of butane by serial MD simulations with the Berendsen thermostat (line). The relaxation time τ is (a) 10 fs, (b) 10^2 fs, (c) 10^3 fs, and (d) 10^4 fs. The filled circles denote numerical integration results.

vergence achieved with RE+WHAM (the uncertainty is roughly three times smaller than with MD with the same computational cost) is interpreted as a consequence of the utilization, by means of WHAM, of additional data from the high temperature replicas.

The dihedral angle distribution $P(\phi_1, \phi_2)$ for pentane is shown in Figure 5. Pentane has a total of nine stable rotamers, but only four states are symmetrically independent: trans-trans (TT), trans-gauche+ (TG+), gauche+-gauche+ (G+G+), and gauche+-gauche− (G+G−). The populations of these states are listed in Table 2. Figure 6 shows that, similar to butane, RE yields faster convergence rates of the populations for all conformations. For example, the relative population uncertainties for the G+G− state are 7.4% for RE+WHAM, 12% for RE, and 22% for serial MD.

The distribution of dihedral angles $P(\phi_1,\phi_2,\phi_3)$ for hexane is shown in Figure 7. Because hexane has three dihedral angles, three 2-D distributions are shown at $\phi_2 = 60°$ (G−), $\phi_2 = 180°$ (T), and $\phi_2 = 300°$ (G+). Hexane has a total of 27 stable rotamers of which 10 are symmetrically independent, as listed in Table 3. Running averages of the populations for the trans-trans-trans (TTT), trans-gauche+-gauche+ (TG+G+), gauche+

gauche+gauche+ (G+G+G+), and gauche+-gauche−-gauche+ (G+G−G+) states out of the 10 states are shown in Figure 8. This set of states includes the most populated state (TTT) and the least populated state (G+G−G+). For the more populated states (TTT, TG+G+, and G+G+G+), the convergence rate of the populations follows the same order as for butane and pentane. However, for the least populated state (G+G−G+), the convergence rate of RE is worse than serial MD. The relative uncertainty of the population of the G+G−G+ state are 25% for RE+WHAM, 31% for serial MD, and 100% for RE. This indicates that for this particular conformational state, the conformational sampling benefit provided by RE is outweighed by the fewer data points available when WHAM is not used. WHAM in this case yields a significant (fourfold) reduction in uncertainty. This is because, as the data listed in Table 4 indicate, that the population of this minor state increases with increasing temperature. At $T = 450$ K, the computed (RE+WHAM) probability of the G+G−G+ state is more than 10 times larger than at 300 K. This suggests that the higher temperature replicas provide substantial statistics for this state that, when combined with WHAM, yields a significantly more precise estimate of the population. As has been previously dis-
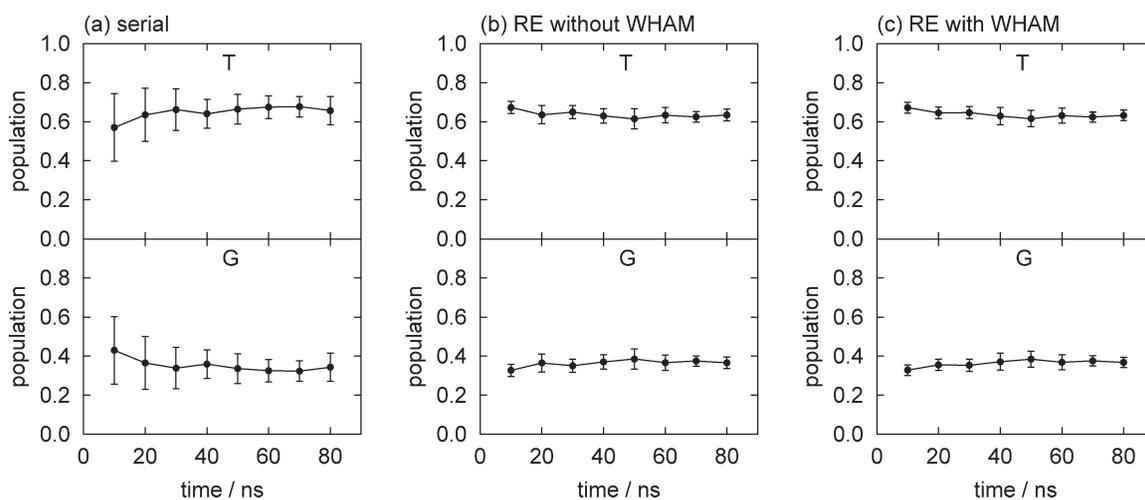

Figure 4. Running averages of the populations for the trans (T) and gauche (G) states of butane obtained by (a) serial MD, (b) replica-exchange MD without WHAM, and (c) replica-exchange MD with WHAM.
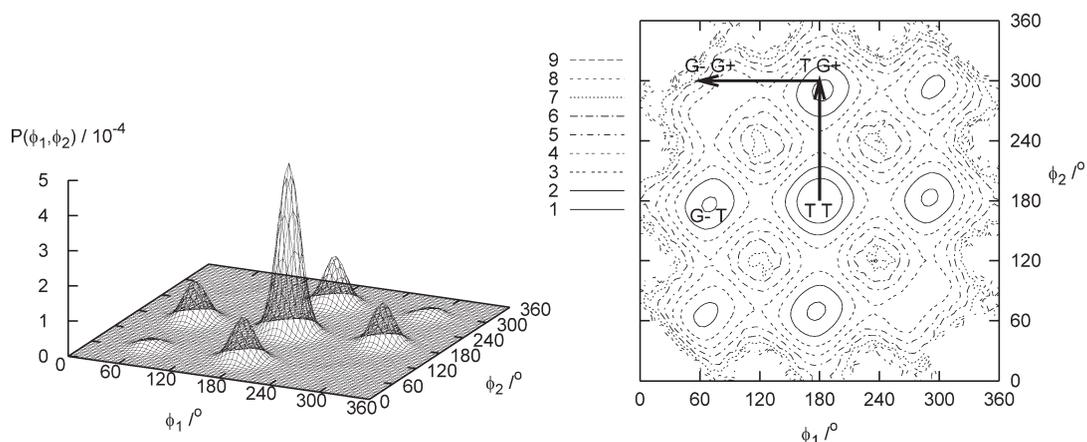
Figure 5. Probability distribution $P(\phi_1, \phi_2)$ of the dihedral angles of pentane (left) and the contour map (right) of the corresponding potential of mean force $F(\phi_1, \phi_2)$ obtained by RE+WHAM. Arrows on the contour map indicate the route for the cut shown Fig. 12.

cussed,[29] temperature-WHAM is generally most useful for the estimation of the free energies of states that are sparsely populated at the temperature of interest and that become more populated at higher temperatures and at lower temperatures.[29] This potential benefit of WHAM is in addition to the enhanced conformational sampling provided by RE, making the RE+WHAM a particularly suitable protocol to study the conformational flexibility of small molecules of this kind.
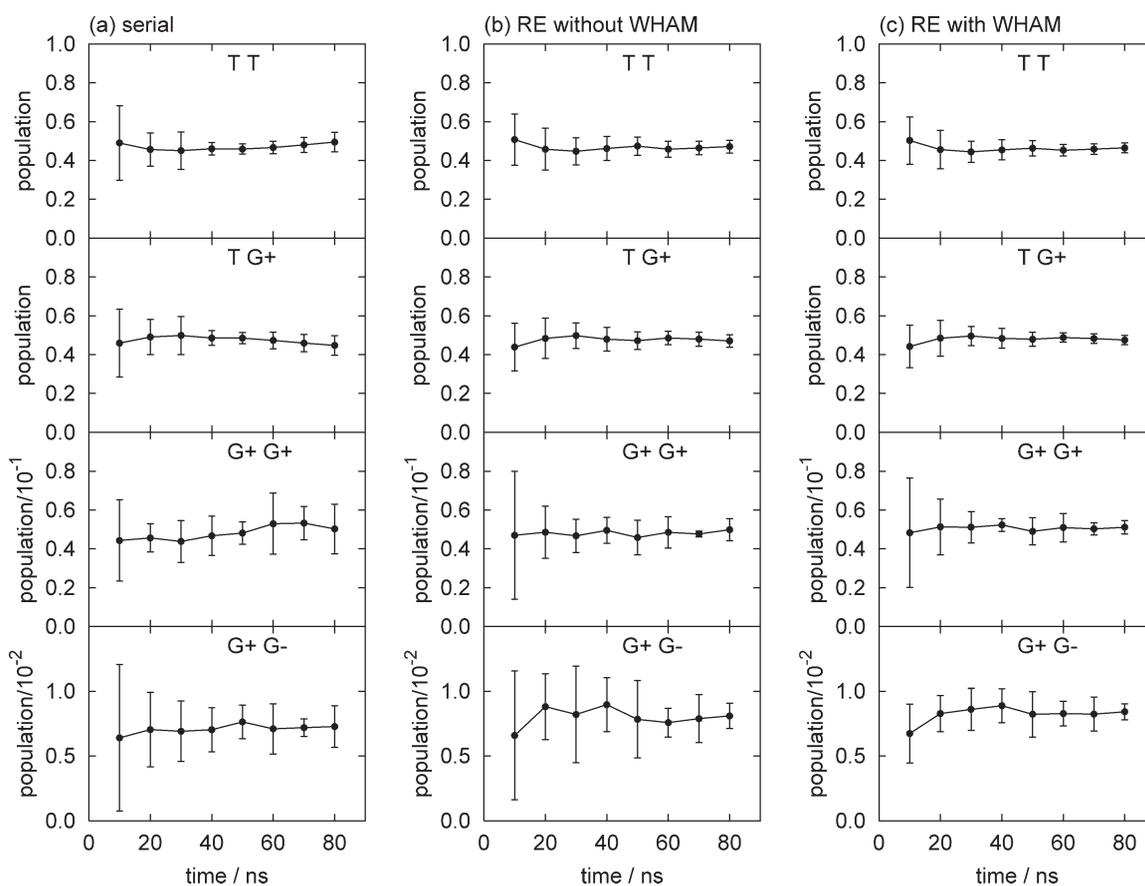

Figure 6. Running averages of the populations for the trans-trans (TT), trans-gauche+ (TG+), gauche+-gauche+ (G+G+), and gauche+-gauche− (G+G−) states of pentane obtained by (a) serial MD, (b) replica-exchange MD without WHAM, and (c) replica-exchange MD with WHAM.
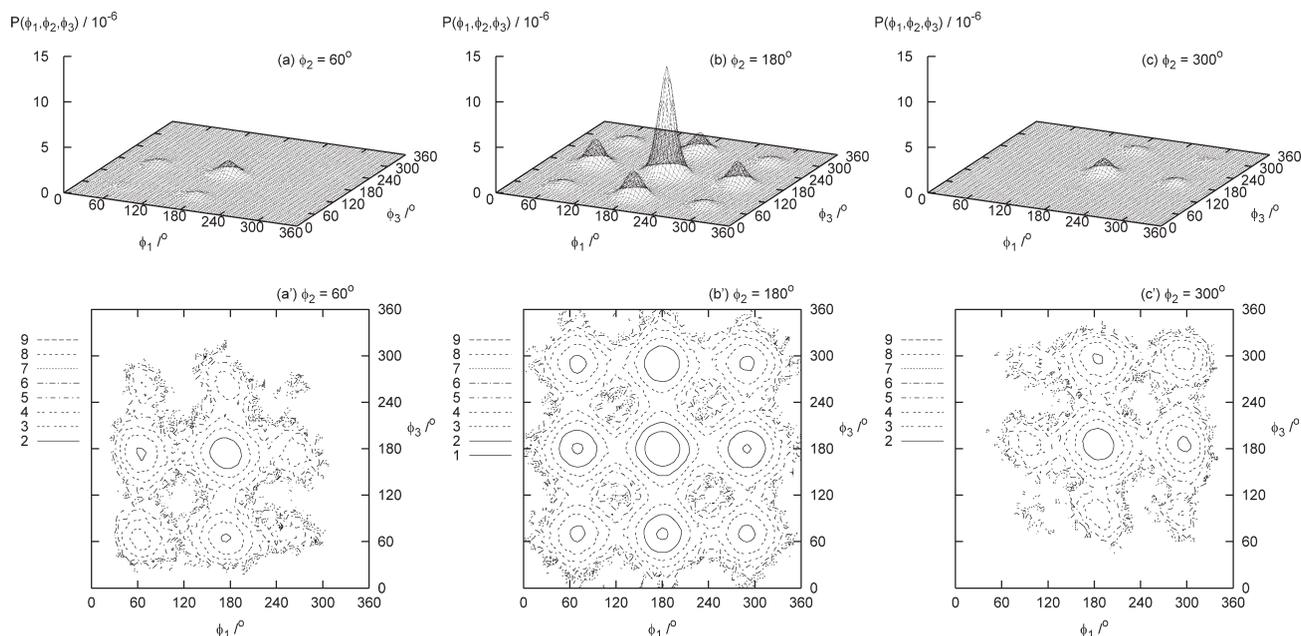
Figure 7. Cuts of the probability distribution $P(\phi_1, \phi_2, \phi_3)$ of dihedral angles of hexane (above) and the contour map (below) of the corresponding potential of mean force $F(\phi_1, \phi_2, \phi_3)$ obtained by RE+WHAM.
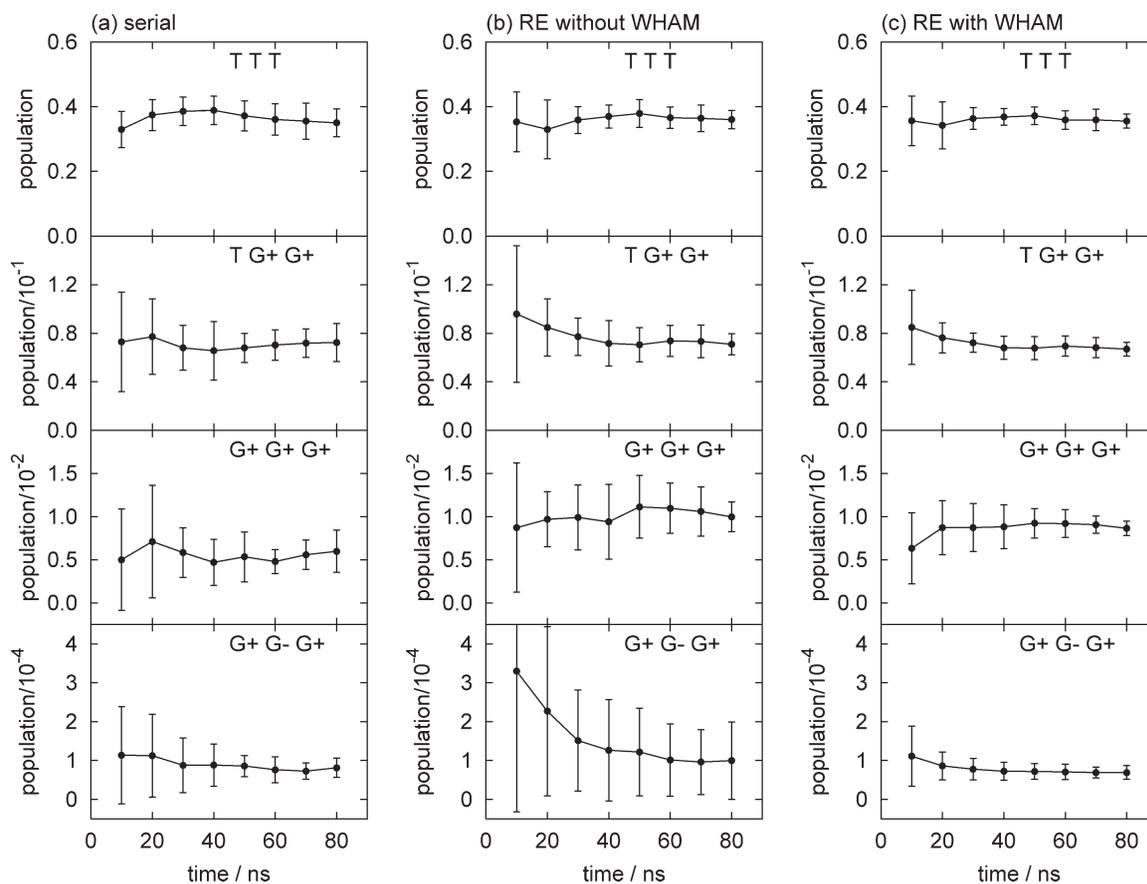

Figure 8. Running averages of the populations for the trans-trans-trans (TTT), trans-gauche+-gauche+ (TG+G+), gauche+-gauche+-gauche+ (G+G+G+), and gauche+-gauche--gauche+ (G+G−G+) states of hexane obtained by (a) serial MD, (b) replica-exchange MD without WHAM, and (c) replica-exchange MD with WHAM.

Table 4. Populations of the Conformational States of Hexane at $T = 450$ K.

ϕ_1	ϕ_2	ϕ_3	RE+WHAM
Trans	Trans	Trans	0.232(17)
Trans	Trans	Gauche+	0.328(11)
Trans	Gauche+	Trans	0.154(14)
Trans	Gauche+	Gauche+	0.100(6)
Gauche+	Trans	Gauche+	0.062(11)
Gauche+	Trans	Gauche−	0.060(7)
Trans	Gauche+	Gauche−	0.0341(12)
Gauche+	Gauche+	Gauche+	0.0190(24)
Gauche+	Gauche+	Gauche−	0.0102(15)
Gauche+	Gauche−	Gauche+	0.00075(16)

The values in parenthesis indicate the uncertainty on the last digit.

TMC278

TMC278, shown in Figure 1, has five rotatable bonds denoted by $\tau_1 - \tau_5$. However τ_1, τ_2, and τ_5 do not appreciably affect the overall shape of the molecule, and it is sufficient in this case to take into account only the τ_3 and τ_4 dihedral angles.[28] The dihedral angle distribution $P(\tau_3, \tau_4)$ of TMC278 in implicit water is shown in Figure 9. There are four symmetrically independent states shown on Figure 10 denoted by the labels E, L_1, L_2, and U.[41] The representative angle values for each state are $(\tau_3, \tau_4) = (180°, 180°)$ for the E, $(\tau_3, \tau_4) = (180°, 0°)$ for the L_1, $(\tau_3, \tau_4) = (0°, 180°)$ for the L_2, and $(\tau_3, \tau_4) = (0°, 0°)$ for the U state. The labels are mnemonics for the shape of the molecule: E for Extended, L_1 and L_2 for L-shaped, and U for U-shaped (Fig. 10). The U state corresponds to the conformation of TMC28 bound to HIV-RT.[28]

The computed populations of the E, L_1, L_2, and U states at 300 K are listed in Table 5; Figure 11 shows the running averages of the corresponding populations. The magnitude of the uncertainties of the populations decreases in the following order:
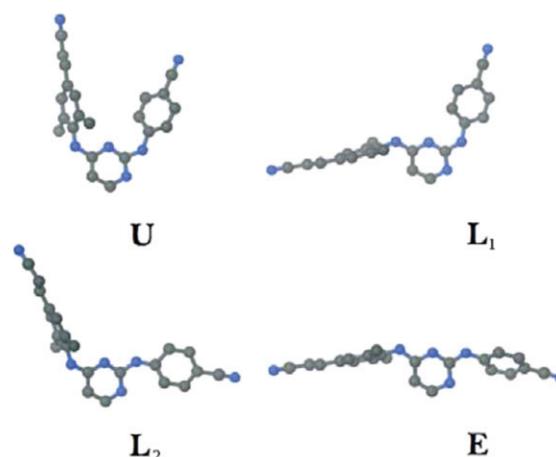

Figure 10. Representative conformations of TMC278 analyzed in this work. U conformation (τ_3 and τ_4 values between 270° and 90°); L_1 conformation ($270° < \tau_4 < 90°$ and $90° < \tau_3 < 270°$); L_2 conformation ($270° < \tau_3 < 90°$ and $90° < \tau_4 < 270°$); E conformation ($90° < \tau_3$ and $\tau_4 < 270°$).

serial MD (largest), RE, and RE+WHAM (smallest). The relative uncertainty of the U state is 27% with RE+WHAM, 29% with RE, and 136% with serial MD. As for butane and pentane, RE (both with WHAM and without WHAM) converges much faster than serial MD, but the effect is quantitatively larger than for the alkanes; the uncertainties for the alkanes with RE+WHAM are one to three times smaller than serial MD at the same computational cost, whereas RE+WHAM population uncertainties are at least five times smaller in the case of TMC278.

The much smaller uncertainties of TMC278 populations with RE sampling relative to serial MD can be understood in terms of the large free energy barriers between stable states which are crossed more frequently at higher temperatures than room tem-
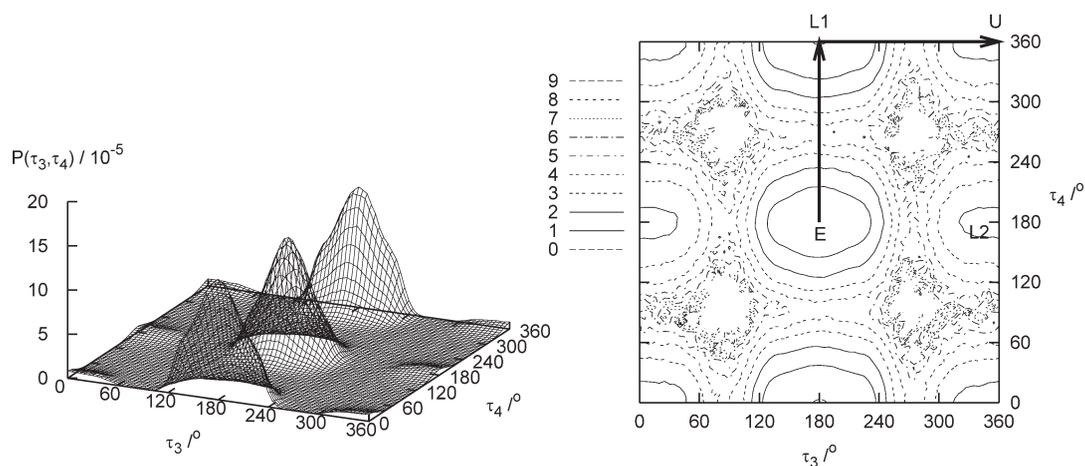

Figure 9. Probability distribution $P(\tau_3, \tau_4)$ of the τ_3 and τ_4 dihedral angles of TMC278 in implicit water (left) and the contour map (right) of the corresponding potential of mean force $F(\tau_3, \tau_4)$ obtained by RE+WHAM. Arrows on the contour map indicate the route of the cut shown in Fig. 12.

Table 5. Populations of the Conformational States of TMC278 in Solution at T = 300 K.

State	MD	RE	RE+WHAM
E	0.41(17)	0.42(3)	0.42(2)
L_1	0.51(12)	0.51(4)	0.52(3)
L_2	0.040(79)	0.029(8)	0.030(6)
U	0.037(50)	0.037(11)	0.037(10)

The values in parenthesis indicate the uncertainty on the last digit.

perature. The high-free energy barriers are illustrated in Figure 12. Figure 12a shows the potential of mean force of pentane along a route from the TT state to TG+ state (from $\phi_2 = 180°$ to $300°$ at fixed $\phi_1 = 180°$) plus that from the TG+ state to G−G+ state (from $\phi_1 = 180°$ to $60°$ at fixed $\phi_2 = 300°$). Figure 12b shows the potential of mean force for TMC278 along a route from the E state to the L_1 state (from $\tau_4 = 180°$ to $360°$ at fixed $\tau_3 = 180°$) plus that from the L_1 state to the U state (from $\tau_3 = 180°$ to $360°$ at fixed $\tau_4 = 360°$). These routes are sche-

matically shown in Figures 5 and 9, respectively. The free energy barriers along these routes are 3.37 kcal/mol from the TT state to the TG+ state and 3.39 kcal/mol from the TG+ state to the G−G+ state of pentane. As shown in Figure 12, the free energy barriers of TMC278 are significantly higher; 4.42 kcal/mol from the E state to the L_1 state and 4.38 kcal/mol from the L_1 state to the U state. The higher free energy barriers cause roughly a 10-fold reduction relative to the alkanes of the rate of interconversions between stable states at room temperature. By taking advantage of the more rapid interconversions at higher temperatures, RE is able to yield much faster convergence than MD for this system.

The reorganization free energy ΔF at 300 K for each conformational state of TMC278 is listed in Table 6 as computed from eq. (1) and the populations listed in Table 5. The reorganization free energy of a state is defined as the free energy required to constrain the molecule in that state, thereby removing it from the thermodynamic equilibrium in which each state assumes its natural population in solution. One of the goals of the present study is to evaluate the ability of RE sampling to compute ligand binding reorganization free energies, that is, the free
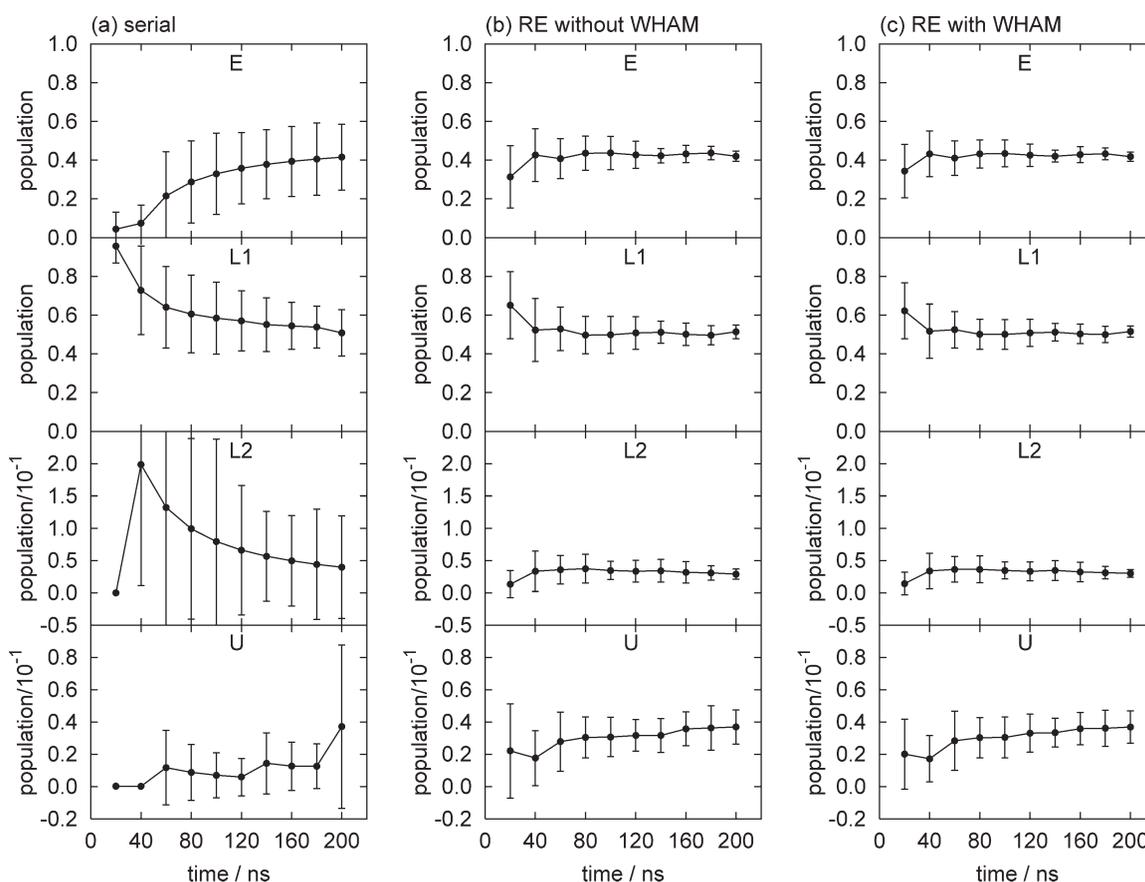

Figure 11. Running averages of the populations for the E, L_1, L_2, and U states of TMC278 in solution obtained by (a) serial MD, (b) replica-exchange MD without WHAM, and (c) replica-exchange MD with WHAM.
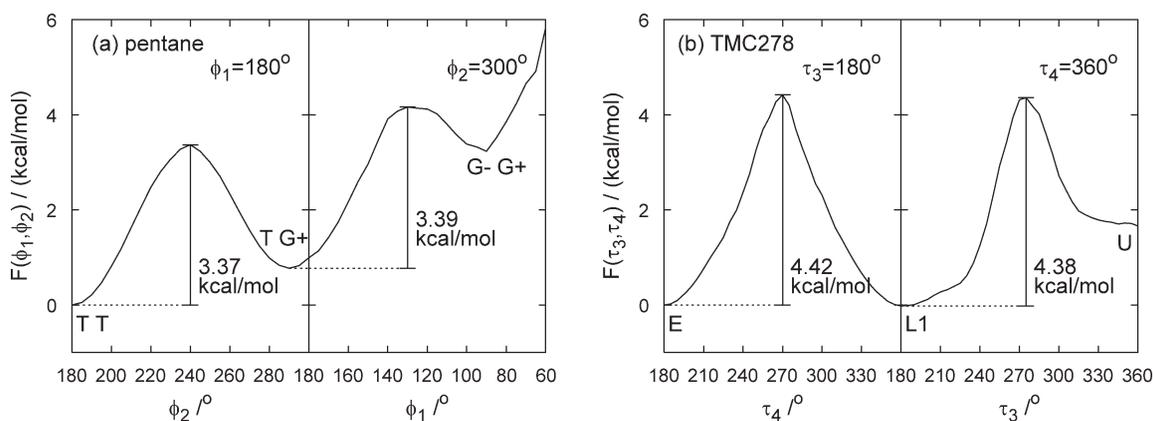
Figure 12. Potential of mean force F obtained by RE+WHAM (a) between TT, TG+, and G−G+ states of pentane, and (b) between E, L_1, and U states of TMC278 in implicit water.

energy penalties incurred by limiting the ligand conformational distribution in solution to the subset of conformations compatible with complexation. TMC278 binds HIV-RT only in the U conformation[28]; therefore, the ligand binding reorganization free energy for TMC278 binding to HIV-RT is predicted to be $\Delta F = 1.97 \pm 0.16$ kcal/mol, as computed by the population (3.7%) of the U state estimated by RE+WHAM. This is a significant free energy contribution to the measured binding free energy (of the order of −10 kcal/mol) of this compound for HIV-RT. As shown in Table 6, the uncertainty of the RE+WHAM estimate of the binding reorganization free energy (0.16 kcal/mol) is significantly smaller than that obtained from serial MD sampling (0.81 kcal/mol) at the same computational cost. This suggests RE sampling is a generally suitable method for computing ligand binding reorganization free energies of drug-like molecules with molecular weights similar to that of TMC278.

Conclusions

We have analyzed the relative efficiency of serial constant temperature MD and parallel replica exchange (RE) sampling for the calculation of conformational populations of a series of small molecules in vacuum and implicit solvent. The analysis focused on computational efficiency (as opposed to real time efficiency) measured as the level of convergence at equal CPU cost. Because RE uses multiple replicas of the system, we conducted proportionally shorter parallel RE simulations compared with se-

rial MD to maintain the same computational cost. Conformational populations by RE sampling have been computed exclusively from the data of the room temperature RE replica and from the data of all replicas using the temperature version of the WHAM. Efficiency was measured in terms of the size of the statistical uncertainties of the populations estimated by block averaging.

We find that based on this measure, RE is generally more computationally efficient than constant temperature MD despite the fact that, for the same computational cost, substantially fewer samples are directly available from RE at the temperature of interest. This finding is rationalized in terms of the higher statistical efficiency of RE trajectories characterized by more frequent interconversions between stable states compared to serial MD. This is a consequence of the relatively small number of degrees of freedom of these molecules, which limit the increase in conformational space with increasing temperature which, for larger systems such as peptides and proteins, causes difficulties for RE sampling.[25,26] Furthermore, the energetic, rather than entropic, nature of the free energy barriers between the stable states of these molecules, allows higher interconversion rates at higher temperatures thereby enhancing convergence. The WHAM method is found to help in the estimation of small populations corresponding to high energy states whose occupancies increase with increasing temperature. Noticeable benefits are observed for alkanes, which have relatively low energy barriers, and especially for TMC278, which is characterized by higher free energy barriers.

The RE method is a parallel algorithm, which takes advantage of computational clusters that are nowadays commonly used in biocomputing research. In this work, we have compared different sampling protocols in terms of their aggregate computational cost, which for RE is given by the length of the run multiplied by the number of replicas used in the simulation. We observe that based on this measure RE is three to five times more efficient than serial MD. The benefits of RE for this particular application would be further magnified if we had used for comparison real time efficiency, that is, the actual time required to get the result regardless of the number of CPU's involved in

Table 6. Reorganization Free Energies of the Conformational States of TMC278 in Solution at $T = 300$ K.

State	MD	RE	RE+WHAM
E	0.52 ± 0.25	0.52 ± 0.04	0.52 ± 0.03
L_1	0.40 ± 0.14	0.40 ± 0.04	0.40 ± 0.03
L_2	1.92 ± 1.19	2.11 ± 0.16	2.09 ± 0.12
U	1.96 ± 0.81	1.97 ± 0.17	1.97 ± 0.16

the calculation. This could be the case for example in situations in which low cost computational resources are readily available and the main concern is the wall clock time required to obtain the answer. On the basis of this measure, we observe that the efficiency for RE is on the order of 10–20 times greater than that of serial MD.

This work paves the way for large-scale automated computations of the ligand binding reorganization free energies as done here for TMC278, an important inhibitor of HIV-1 reverse transcriptase (RT). For TMC278 we measure with the RE+WHAM protocol a binding reorganization free energy of approximately 2 kcal/mol (a substantial contribution to the total binding free energy) with less than 10% relative uncertainty. This constitutes a substantial contribution to the total binding free energy. Ligand reorganization free energies potentially play a significant role in modulating the affinities of inhibitors to their receptors. We plan to investigate with this method the binding reorganization free energies of a series of nonnucleoside inhibitors of HIV-1 RT and other pharmaceutically relevant systems. This work will be presented in future publications.

Acknowledgments

References

1. Hukushima, K.; Nemoto, K. J Phys Soc Jpn 1996, 65, 1604.
2. Sugita, Y.; Okamoto, Y. Chem Phys Lett 1999, 314, 141.
3. Mitsutake, A.; Sugita, Y.; Okamoto, Y. Biopolymers (Pept Sci) 2001, 60, 96.
4. Berg, B. A. Comp Phys Commun 2002, 104, 52.
5. Berg, B. A.; Neuhaus, T. Phys Lett B 1991, 267, 249.
6. Berg, B. A.; Neuhaus, T. Phys Rev Lett 1992, 68, 9.
7. Hansmann, U. H. E.; Okamoto, Y.; Eisenmenger, F. Chem Phys Lett 1996, 259, 321.
8. Nakajima, N.; Nakamura, H.; Kidera, A. J Phys Chem B 1997, 101, 817.
9. Okumura, H.; Okamoto, Y. Chem Phys Lett 2004, 383, 391.
10. Okumura, H.; Okamoto, Y. Phys Rev E 2004, 70, 026702.
11. Okumura, H.; Okamoto, Y. Chem Phys Lett 2004, 391, 248.
12. Okumura, H.; Okamoto, Y. J Phys Soc Jpn 2004, 73, 3304.
13. Okumura, H.; Okamoto, Y. J Comput Chem 2006, 27, 379.
14. Okumura, H.; Okamoto, Y. Bull Chem Soc Jpn 2007, 80, 1114.
15. Morishita, T.; Mikami, M. J Chem Phys 2007, 127, 034104.
16. Okumura, H.; Okamoto, Y. J Phys Chem B 2008, 112, 12038.
17. Okumura, H. J Chem Phys 2008, 129, 124116.
18. Feig, M.; Karanicolas, J.; Brooks C. L., III. J Mol Graphics Modelling 2004, 22, 377.
19. Case, D. A.; Cheatham T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. J Comput Chem 2005, 26, 1668.
20. Banks, J. L.; Beard, J. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. J Comput Chem 2005, 26, 1752.
21. Gallicchio, E.; Levy, R. M.; Parashar M. J Comput Chem 2008, 29, 788.
22. Beck, D. A. C.; George W. N.; White, G. W. N.; Valerie Daggett, V. J Struct Biol 2007, 157, 514.
23. Zuckerman, D. M.; Lyman, E. J Chem Theory Comput 2006, 2, 1200.
24. Denschlag, R.; Lingenheil, M.; Tavan, P. Chem Phys Lett 2008, 458, 244.
25. Zheng, W.; Andrec, M.; Gallicchio, E.; Levy, R. M. Proc Natl Acad Sci USA 2007, 104, 15340.
26. Zheng, W.; Andrec, M.; Gallicchio, E.; Levy, R. M. J Phys Chem B 2008, 112, 6083.
27. Ma, B.; Shatsky, M.; Wolfson, H. J.; Nussinov, R. Protein Sci 2002, 11, 184.
28. Das, K.; Bauman, J. D.; Clark, A. D.; Frenkel, Y. V.; Lewi, P. J.; Shatkin, A. J.; Hughes, S. H.; Arnold, E. Proc Natl Acad Sci USA 2008, 105, 1466.
29. Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. J Phys Chem B 2005, 109, 6722.
30. Gallicchio, E.; Levy, R. M. J Comp Chem 2004, 79, 479.
31. Banks, J. L.; Beard, J. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. J Comput Chem 2005, 26, 1752.
32. Nosé, S. Mol Phys 1984, 52, 255.
33. Nosé, S. J Chem Phys 1984, 81, 511.
34. Hoover, W. G. Phys Rev A 1985, 31, 1695.
35. Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Mol Phys 1996, 87, 1117.
36. Evans, D. J.; Holian B. L. J Chem Phys 1985, 83, 4069.
37. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. J Comp Chem 1992, 13, 1011.
38. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. J Chem Phys 1984, 81, 3684.
39. Morishita, T. J Chem Phys 2000, 113, 2976.
40. Harvey, S. C.; Tan, R. K.-Z.; Cheatham, T. M., III. J Comp Chem 1998, 19, 726.
41. Frenkel, Y. V. The Roles of Structural Variability and Amphiphilicity of TMC278/Rilpivirine and Mechanisms of Resistance Avoidance and Enhanced Oral Bioavailability. PhD thesis, Rutgers University, Graduate Program in Biochemistry, 2009.