Chemical Physics Letters 460 (2008) 375–381

Contents lists available at ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett

Equilibrium exchange enhances the convergence rate of umbrella sampling

Chris Neale, Tomas Rodinger [1], Régis Pomès *

Molecular Structure and Function, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8
Department of Biochemistry, University of Toronto, Canada

ARTICLE INFO

Article history:
Received 9 May 2008
In final form 30 May 2008
Available online 5 June 2008

ABSTRACT

Umbrella sampling, a technique for sampling energetically unfavorable states by computer simulation, is subject to systematic sampling errors arising from quasi-nonergodicity. In order to address this problem, we compare free energy profiles generated along the backbone dihedral angles φ and ψ of alanine dipeptide in water using umbrella sampling successively with and without the provision of equilibrium exchange between umbrellas. The results show that the addition of equilibrium exchange to umbrella sampling alleviates these errors by restoring the utilization of transition paths that leave and reenter the specified sampling region.

1. Introduction

The convergence of free energy profiles derived from molecular simulation depends not only on complete sampling of the relevant conformational space, but also on determination of the relative probabilities of the sampled states. This information may be obtained from unrestrained simulation only when a sufficiently large number of transitions have occurred between local minima on the energy landscape. The rugged energy landscape of a biomolecule contains energetic barriers that are small enough to be traversed rapidly on biological and experimental timescales but large enough to be traversed infrequently, if at all, on timescales currently available to computer simulation. Unrestrained simulations are therefore limited in their ability to provide converged free energy profiles along many biologically interesting reaction coordinates.

A variety of computational methods can be used to accelerate the rate of barrier crossing by artificially enhancing the sampling of high energy conformations [1]. Methods that execute a random walk in temperature space [2–7] utilize the greater kinetic energy available at high temperatures to help the system surmount energetic barriers and thereby provide more complete probability distributions at biological temperatures. High temperatures, however, may permit the system to leave the desired sampling region of phase space. For example, high temperatures may lead to the denaturation of a folded protein, a process that is entropically difficult to reverse. An alternative is to apply a biasing potential that reduces the magnitude of known barriers at biological temperatures [8]. In the general case, where the location and magnitude of the barriers are not known a priori, the biasing potential can be constructed concurrently with simulation [9,10], or a series of simulations can be run using different restraints such that the ensemble of simulations blankets the selected reaction coordinate, a technique known as umbrella sampling (US) [11,12].

During US, the independence of each sampling region may increase systematic sampling errors. Transitions between energy basins orthogonal to the reaction coordinate are hindered because restraining potentials block any transition path that leaves and reenters a given section of the reaction coordinate. This unintended side-effect of US can be reduced by restoring the ability of individual trajectories to sample the entire length of the reaction coordinate. To this end, an exchange-capable version of umbrella sampling has previously been introduced [13] and used to determine the free energies associated with DNA base stacking [14] and conformational fluctuations in folded proteins [15,16]. However, the efficiency of US with equilibrium exchange has been questioned [17]. Controlled comparisons to non-exchanging US are lacking in the literature and the extent to which equilibrium exchange enhances the efficiency and accuracy of US, if at all, has not been evaluated directly. In this Letter, we systematically evaluate the addition of equilibrium exchange to US and demonstrate how equilibrium exchange may enhance the convergence rate of US. As our exchange methodology, we utilize distributed replica (DR) sampling, a generalized-ensemble method recently introduced to couple an ensemble of simulations running on inhomogeneous or distributed computing platforms [18]. As a test case, we have chosen the well-studied alanine dipeptide (*N*-acetylalanine-*N*′-methylamide). The simplicity of this system allows us to generate converged reference free energy profiles along the backbone dihedral angles φ and ψ to which comparisons are made for the purpose of methodological evaluation.

* Corresponding author. Address: Molecular Structure and Function, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8. Fax: +1 416 813 5022.
E-mail address: pomes@sickkids.ca (R. Pomès).
[1] Present address: Zymeworks, 201-1401 W Broadway, Vancouver, BC, Canada V6H 1H6.

0009-2614/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cplett.2008.05.099

2. Theory

Let us consider a collection of N simulations, each of which is restricted to sample a specific region of conformational space by adding to the Hamiltonian of the system, H, a biasing function U

$$H'(\boldsymbol{R_i}, \xi_{i[0]}) = H(\boldsymbol{R_i}) + U(\mathbf{R_i}, \xi_{i[0]}), \tag{1}$$

where i is the index of the simulation, $\boldsymbol{R_i}$ defines a point in conformational space, and $\xi_{i[0]}$ defines the region of enhanced sampling. One approach for U is a harmonic restraining potential of the form

$$U(\boldsymbol{R_i}, \xi_{i[0]}) = \frac{1}{2} k_i (\xi_i - \xi_{i[0]})^2 \tag{2}$$

that biases sampling from the current value of the reaction coordinate, ξ_i, toward the center of the restraint, $\xi_{i[0]}$, in proportion to a constant, k_i. Further, it is possible to increase the amount of sampling along the entire range of a selected reaction coordinate ξ by independently carrying out many such simulations, each with a different value of $\xi_{i[0]}$. This is the US approach [11,12]. The drawback to US is that only a portion of the reaction coordinate is accessible to each simulation, possibly increasing systematic sampling errors. In order to address this problem, each simulation may be permitted to change $\xi_{i[0]}$ to a new value, $\xi_{j[0]}$, according to the Metropolis criterion [19] with probability

$$P_{i\to j} = \min(1, \exp\{-\beta\sigma_{i\to j}\}), \tag{3}$$

where β^{-1} is the Boltzmann constant times the absolute temperature, i and j index the available centers of restraint (umbrellas, using Eq. (2)), and

$$\sigma_{i\to j} = H'(\boldsymbol{R_i}, \xi_{j[0]}) - H'(\boldsymbol{R_i}, \xi_{i[0]}). \tag{4}$$

This strategy is similar to hybrid Monte Carlo (MC) methods [20] in which simulations undergo occasional MC moves. When moves are accepted based on Boltzmann probabilities (Eqs. (3) and (4)), the reaction coordinate is sampled according to the Boltzmann distribution even though restraining potentials are applied during intervening segments. In this case, the simulation is no more efficient than an unrestrained simulation; both methods preponderantly sample low energy conformations. One way to increase sampling uniformity along the exchanging reaction coordinate is to flatten the free energy profile that underlies MC move attempts by subtracting the average free energy difference between sampling at $\xi_{i[0]}$ and $\xi_{j[0]}$. In this case, Eq. (4) is replaced by

$$\sigma_{i\to j} = H'(\boldsymbol{R_i}, \xi_{j[0]}) - H'(\boldsymbol{R_i}, \xi_{i[0]}) - (A_j - A_i), \tag{5}$$

where A_i is the Boltzmann-weighted average of the free energy sampled by umbrella i. This approach is similar to simulated tempering (ST) [2,3]. A major drawback to all ST-like methods is that determination of the free energies for use in Eq. (5) is non-trivial. If these estimates vary significantly from the actual free energies, some values of $P_{i\to j}$ are greater than others and the sampling rate varies significantly along the reaction coordinate, reducing the efficiency of ST-like US. In order to further increase the sampling uniformity across the reaction coordinate, one can utilize the distributed replica potential energy function (DRPE) introduced by Rodinger et al. [18]. During DR sampling, many simulations (or replicas) are distributed along a reaction coordinate and occasional MC move attempts of individual replica $\xi_{i[0]}$ values are coupled by the DRPE. The DRPE is constructed such that the probability of accepting a $\xi_{i[0]}$ move increases with the uniformity of the instantaneous distribution of the ensemble of simulations. The probability of changing $\xi_{i[0]}$ is then based on

$$\sigma_{i\to j} = H'(\boldsymbol{R_i}, \xi_{j[0]}) - H'(\boldsymbol{R_i}, \xi_{i[0]}) - (A_j - A_i) + D_j - D_i, \tag{6}$$

where D_i and D_j represent the value of the DRPE before and after the proposed MC move, respectively. We refer to this approach as distributed replica umbrella sampling (DRUS). In DRUS, the free energy estimates applied during MC move attempts in Eq. (6) need not be exact, as the DRPE enforces near-uniform sampling.

Since free energy is a measure of reversible work, the free energy estimates in Eq. (6) may be obtained by integrating the mean force exerted by restraining potential i, F_i. Specifically, A_i is calculated for N umbrellas according to

$$A_i = \left\{ \begin{array}{ll} 0 & \text{if } i = 1 \\ \Delta\xi_{i-1,i}(F_{i-1} + F_i)/2 + A_{i-1} & \text{if } 1 < i \leqslant N \end{array} \right\}, \tag{7}$$

where

$$\Delta\xi_{j,i} = \xi_{i[0]} - \xi_{j[0]}. \tag{8}$$

Errors in F_i are accumulated during integration and, generally, the absolute error in A_i increases with i. This is not problematic for a non-periodic reaction coordinate because the values A_i and A_j in every evaluation of Eq. (6) are from neighboring umbrellas. The accumulated error therefore largely cancels out. A periodic reaction coordinate, however, requires evaluating Eq. (6) for the case when $i = N$ and $j = 1$. It is therefore preferable to generate preliminary free energy estimates, A'_i, for which A'_{N+1} estimates the error accumulated through integration for a periodic reaction coordinate. Specifically,

$$A'_i = \left\{ \begin{array}{ll} 0 & \text{if } i = 1 \\ \Delta\xi_{i-1,i}(F_{i-1} + F_i)/2 + A'_{i-1} & \text{if } 1 < i \leqslant N \\ \Delta\xi_{N,1}(F_N + F_1)/2 + A'_N & \text{if } i = N+1 \end{array} \right\}. \tag{9}$$

The error accumulated through integration is then distributed across all umbrellas to generate a final estimate of the free energy at each umbrella according to

$$A_i = A'_i - i(A'_{N+1})/N \quad \text{for } 1 \leqslant i \leqslant N. \tag{10}$$

3. Methods

Simulations of alanine dipeptide in water were conducted with version 3.3.1 of the GROMACS simulation package [21], using the Amber94 force field [22,23] and solvating with 632 TIP3P water molecules [24]. Periodic boundary conditions were enforced via a rhombic dodecahedron unit cell with an initial minimum distance of 1.1 nm between the solute and the boundary. Lennard–Jones interactions were evaluated using a group-based twin-range cutoff [25] calculated every step for separation distances less than 0.9 nm and every ten steps for distances between 0.9 and 1.4 nm, when the nonbonded list was updated. Coulomb interactions were calculated using the smooth particle-mesh Ewald method [26,27] with a real-space cutoff of 0.9 nm. Simulation in the NpT ensemble was achieved by isotropic coupling to a Berendsen barostat [28] at 1 bar with a coupling constant of 4 ps and separate coupling of the solute and the solvent to Berendsen thermostats [28] at 300 K with coupling constants of 0.1 ps. Bonds involving hydrogen were constrained with SETTLE [29] and LINCS [30] for solvent and solute, respectively. The integration time step was 2 fs. Coordinates were saved every 0.1 ps.

From the unrestrained simulation, the free energy dependence on backbone torsion angles φ and ψ, $W(\varphi, \psi)$, was determined based on sampling probabilities, $P(\varphi, \psi)$, according to

$$W(\phi, \psi) = -\beta^{-1} \ln[P(\phi, \psi)/P_{\max}], \tag{11}$$

where $P_{\max}$ is the maximum value of $P(\varphi, \psi)$.

A sequential method was used to construct the 2D free energy surface via 2D umbrella sampling (2DUS). The initial conformation

was simulated for 13.325 ns while restraining the φ and ψ dihedral angles via harmonic restraints with a force constant of 0.0364 kcal mol^{-1} deg^{-2}. New simulations were initiated using conformations from simulations at neighboring φ and ψ restraints, 10° apart. The dihedral angles from the resulting 17.3 μs of simulation were used to compute the free energy surface using Alan Grossfield's periodic implementation of the 2D weighted histogram analysis method (WHAM) [12,31–33] using a tolerance of 10^{-5} kcal mol^{-1}.

Initial structures for 1D US and DRUS simulations were taken from snapshots of 2DUS simulations along $\varphi = -70°$ or $\varphi = +60°$. A total of 1.044 μs was then simulated via US or DRUS, using ψ as the reaction coordinate, yielding an average of 29 ns at each umbrella. Changes of $\psi_{i[0]}$ were attempted every 2 ps during DRUS. The umbrella centered at −170° was capable of changing directly to the umbrella centered at +180° (and vice versa) in order to take advantage of the periodic nature of the reaction coordinate. The DRPE was taken from Eqs. (15)–(17) of reference [18] using a constant of 0.004 kcal mol^{-1}.

In order to construct 1D free energy profiles, 2D free energy surfaces were first constructed using WHAM during which the orthogonal reaction coordinate was taken to contain a single umbrella with a force constant of zero. These 2D free energy surfaces were then projected onto the desired reaction or orthogonal coordinate according to

$$W(\phi_p) = -\beta^{-1}\ln\sum_q \exp\{-\beta W(\phi_p,\psi_q)\}, \tag{12}$$

where p and q index the umbrella potentials available to φ and ψ, respectively, and then shifted so that the absolute minimum of $W(\varphi)$ is zero, and similarly for $W(\psi)$.

In order to evaluate the similarity between two free energy profiles, we first constructed a representation of each profile that is independent of the unknown constant introduced by WHAM. Specifically, each free energy profile containing M values of W was used to construct the upper triangle of a matrix $\boldsymbol{W}$, indexed by c and d, according to

$$\boldsymbol{W} = (W_c - W_d)_{c=1,\ldots,M;\ d=c+1,\ldots,M}. \tag{13}$$

We then compare two matrices by calculating a root mean squared difference (RMSD) between corresponding elements. For two profiles, A and B, the $\Delta\boldsymbol{W}$ RMSD is given by

$$\Delta\boldsymbol{W} = \sqrt{\frac{2}{M(M-1)}\sum_{c=1}^{M}\sum_{d=c+1}^{M}\left(\boldsymbol{W}^A_{c,d} - \boldsymbol{W}^B_{c,d}\right)^2}. \tag{14}$$

4. Results and discussion

In this Letter we compare the sampling efficiency of US and DRUS in 1D. This comparison hinges on our ability to provide a reference free energy profile that closely approximates the exact solution. Fig. 1 shows the free energy surface of the backbone dihedral angles of the alanine dipeptide. The global minimum of the free energy surface is located in the right-handed α-helical region (α_R) of the Ramachandran plot [34] at (φ, ψ)=(−67.5°, −16.5°). Two local minima are located in the β and left-handed α-helical region (α_L) of the Ramachandran plot at (−64.5°, 163.5°) and (55°, 13.5°) with relative free energies of 1.25 kcal mol^{-1} and 2.86 kcal mol^{-1}, respectively. A fourth, shallow local minimum exists at (60°, 180°). The unrestrained free energy surface is similar to the reference 2DUS free energy surface for dihedral combinations whose free energy is within 5 kcal mol^{-1} of the global minimum, but becomes inaccurate above that value and altogether fails to sample conformations above 7 kcal mol^{-1}. Comparison of the free energy

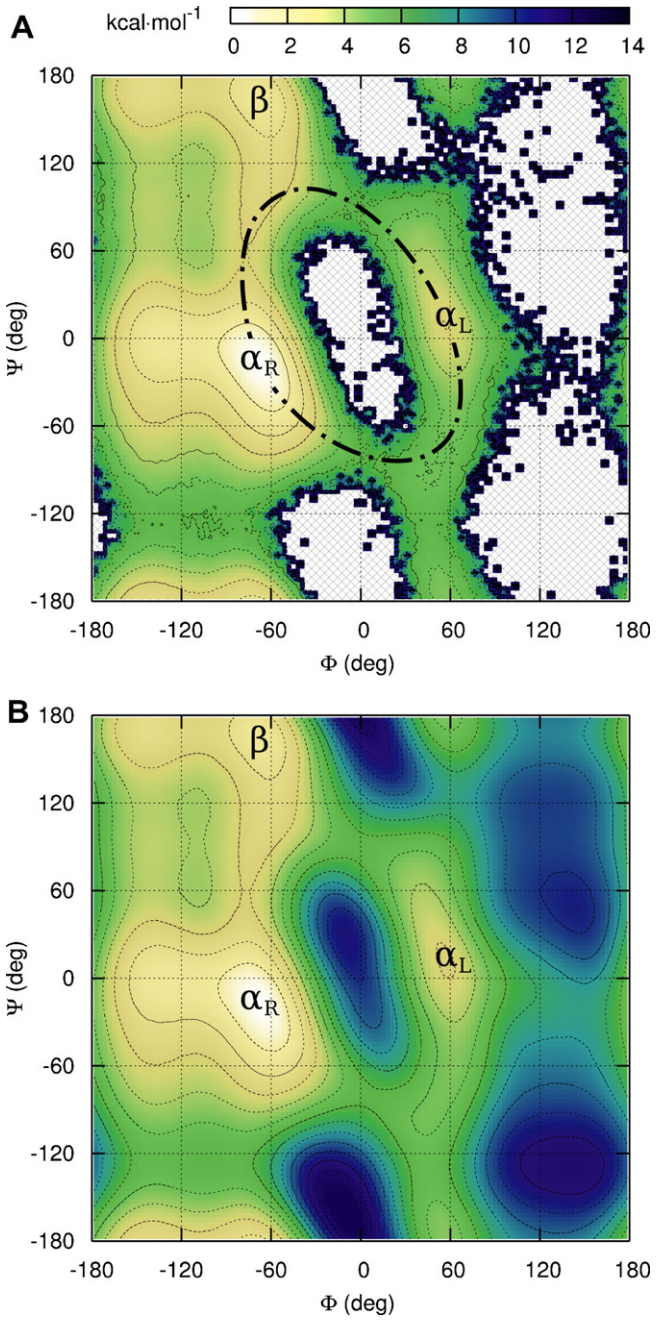

Fig. 1. Free energy surfaces of the alanine dipeptide at 300 K along backbone torsions φ and ψ. Contours are drawn at 1 kcal mol^{-1} intervals. Hatched areas indicate a complete absence of sampling. The global minimum is marked α_R and two local minima are marked β and α_L. (A) Generated by 3 μs of unrestrained simulation. A low energy pathway connecting α_R and α_L basins is shown as a dash-dot line. (B) Generated by 17.3 μs of 2DUS.

surfaces generated by 3 μs of 2DUS and 17.3 μs of 2DUS yields a $\Delta\boldsymbol{W}$ RMSD less than 0.02 kcal mol^{-1} (Fig. 2) and a maximum $\Delta\boldsymbol{W}$ difference of 0.1 kcal mol^{-1} (not shown). The ability of 2DUS to provide more complete information about the free energy surface than the unrestrained simulation is therefore predominantly methodological and is not due to the extra sampling time. Previously-published reports using the same force field are in agreement with the 2D free energy surface shown in Fig. 1B [10,35,36]. The only exception in the literature [37] is quantitatively incorrect (Jay Ponder, personal communication). In this study, we utilize 1D projections of the 17.3 μs 2DUS free energy

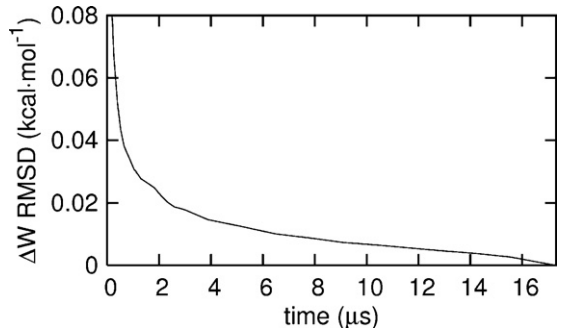

Fig. 2. RMSD between corresponding elements of final and partial-time $W(\varphi, \psi)$ difference matrices based on 2DUS (see Eq. (14)).

surface as reference profiles against which to compare the results of 1D US and DRUS. Because our interest is methodological, we make no comment on the accuracy of the force field.

Our first evaluation of the addition of exchange to US utilizes ψ as a reaction coordinate and is based on simulations initiated in the lowest free energy valley, near $\varphi = -70°$. Fig. 3A compares 1D free energy profiles along ψ generated by US and DRUS to the reference profile along ψ. Although DRUS converges more quickly to the reference profile than does US (Fig. 3B), the difference is trivial and both methods reproduce the 1D free energy profile along ψ very well.

The same simulations are also used to evaluate the ability of each method to sample a degree of freedom orthogonal to the specified reaction coordinate. Free energy profiles along φ generated using ψ as a reaction coordinate are compared to the reference profile along φ in Fig. 4. Significant discrepancies exist between the US and reference profiles. The inability of US to reproduce the local minimum at $\varphi = 55.5°$ (Fig. 4) is due to the inability of simulations restrained near $\psi_{i[0]} = 10°$ to overcome energy barriers that lie between α_R and α_L (not shown). These energy barriers are 7.8 kcal mol^{-1} at $\varphi = 148.5°$ and 9.5 kcal mol^{-1} at $\varphi = -4.5°$ (Fig. 1B). As a result, α_L is never visited by these US simulations. By contrast, DRUS trajectories may explore the reaction coordinate and undergo transitions between α_R and α_L via saddle points at (0°,−75°) and (0°, 90°) (dash-dot line in Fig. 1A).

Neither US nor DRUS along ψ were capable of reproducing the reference profile for $100° < \varphi < 160°$ (Fig. 4). This sampling defect is due to the high free energy ridge, which lies above 6 kcal mol^{-1} (Figs. 4 and 1B) and is seldom visited in the absence of any bias along φ (Fig. 1A).

Next, we investigate the implications of starting each simulation from a high energy conformation in order to assess US and DRUS under sub-optimal initial conditions. To this end, we repeated US and DRUS simulations along ψ starting from conformations in the secondary free energy valley, near $\varphi = +60°$. DRUS reproduces the reference profile along ψ almost exactly, whereas US deviates significantly (Fig. 5A). The DRUS simulations quickly migrated to the α_R and β basins in the lowest free energy valley (not shown). However, during the entire US initiated in the secondary free energy valley, the simulations restrained to $\psi_{i[0]} = -160°$, −150°, +10°, and +40° remained in the α_L and (60°, 180°) free energy basins. These systematic sampling errors resulted in artificial inflection points in the US free energy profile at corresponding values of ψ (Fig. 5A). The effect of exclusive sampling of high energy φ values on the free energy profile along ψ is somewhat unintuitive. One might expect that free energies along the selected reaction coordinate ψ would increase when sampling along the orthogonal coordinate φ was poorly converged. However, the local free energy profiles along ψ sampled by the segments trapped in the α_L and (60°, 180°) basins during US are relatively flat in comparison to the respective free energy profiles along ψ in the α_R and β basins, where the free energy is more favorable (Fig. 1B). Linear WHAM reproduces the local minimum at $\psi = 166.5°$ for DRUS, but not for US (not shown).

The underestimate of the free energy at $\varphi = 55.5°$ from US along ψ (Fig. 5B) is due to the same energetic barriers that caused the overestimate of this region of the free energy profile in Fig. 4. In this case, however, simulations restrained near $\psi_{i[0]} = 10°$ never visited α_R, which led to oversampling φ at α_L.

To further illustrate the systematic sampling errors introduced by US, the 2D free energy surfaces used to generate Figs. 4 and 5

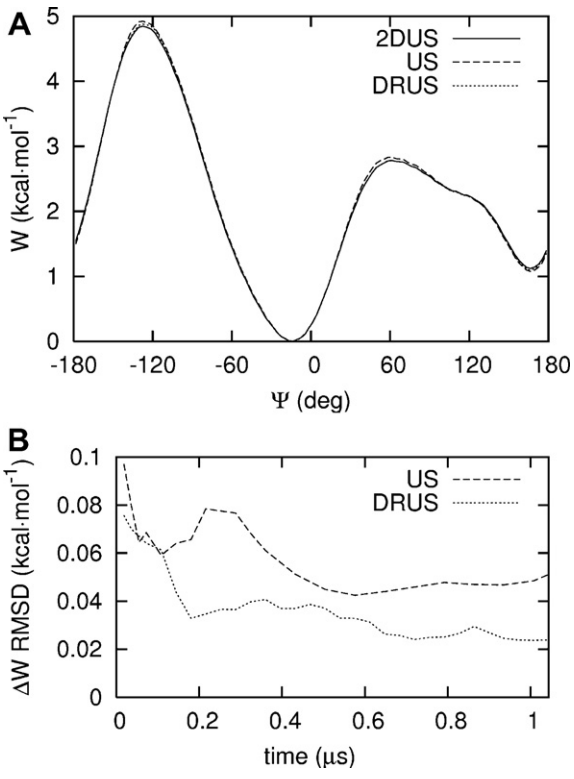

Fig. 3. (A) Free energy profile along ψ generated by (solid line) projection of the 2DUS free energy surface, (dashed line) US along ψ initiated near $\varphi = -70°$, and (dotted line) DRUS along ψ initiated near $\varphi = -70°$. (B) RMSD between corresponding elements of the reference and partial-time US or DRUS $W(\psi)$ difference matrices based on simulations initiated near $\varphi = -70°$; (dashed line) US and (dotted line) DRUS along ψ.

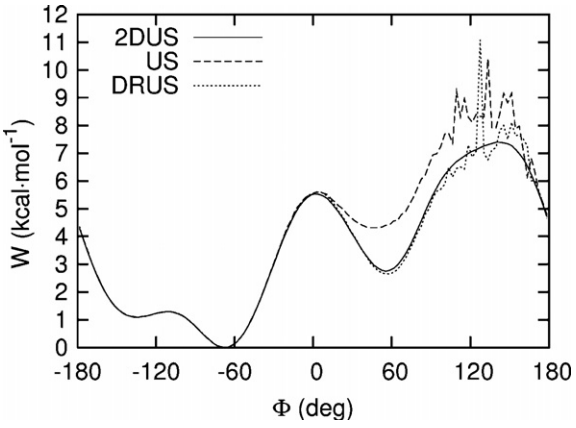

Fig. 4. Free energy profile along φ generated by (solid line) projection of the 2DUS free energy surface, (dashed line) US along ψ initiated near $\varphi = -70°$, and (dotted line) DRUS along ψ initiated near $\varphi = -70°$.

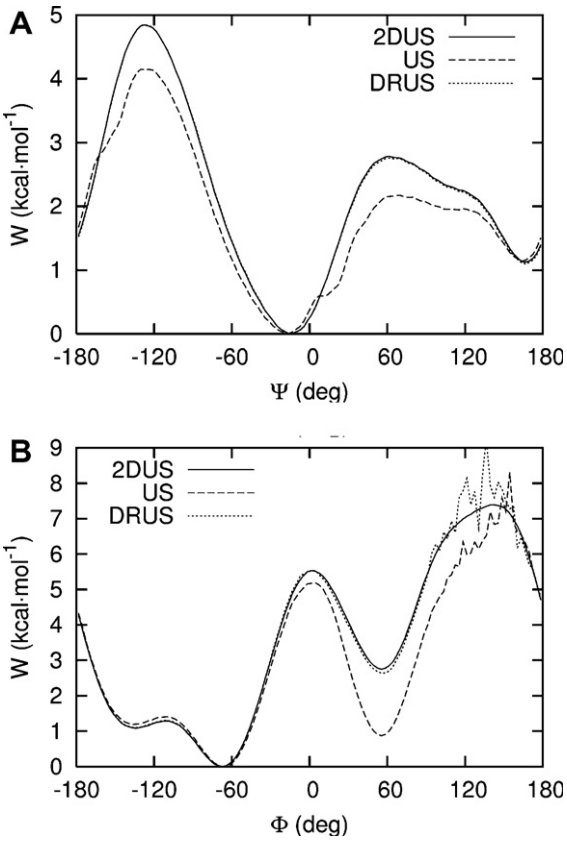

Fig. 5. Free energy profile along (A) ψ, and (B) φ generated by (solid line) projection of the 2DUS free energy surface; and (dashed line) US and (dotted line) DRUS along ψ initiated near $\varphi = +60°$.

are shown in Fig. 6. US does not sample the α_L and (60°, 180°) basins when initiated in the lowest free energy valley (Fig. 6B). US traps some simulations in α_L and (60°, 180°) when initiated in the secondary free energy valley (Fig. 6D). The main sampling defect in DRUS is the absence of sampling near (60°, −20°) (Fig. 6A and C). There is a minimum along ψ at −16.5° (Fig. 3A) and the subtractive free energy estimate in Eq. (6), $-(A_j - A_i)$, therefore comes to a maximum at the corresponding $\psi_{i[0]}$. While this maximum enhances sampling uniformity along ψ, it also reduces the sampling near (60°, −20°), where the underlying free energy surface is relatively flat (Fig. 1).

Fig. 7 shows the amount of simulated time at each umbrella for US and DRUS. Efficiency is maximized when each umbrella is sampled for an equal amount of time. This type of efficiency is trivially accomplished by US due to the independence of each umbrella. DRUS simulations, however, rely on free energy estimates and the DRPE to achieve sampling uniformity. The DRUS simulation initiated in the secondary free energy valley deviates significantly from sampling uniformity because the free energy estimates applied in Eq. (6) were calculated based on an initial US simulation that was plagued with systematic sampling errors. In systems of greater complexity, a preliminary DRUS simulation may be necessary to generate more accurate free energy estimates for the production DRUS simulation. At any rate, the DRPE significantly reduces the number of iterations of such adaptation required to achieve approximate sampling uniformity along the reaction coordinate (not shown).

It has already been stated [17] that the enhancement of sampling accompanying the introduction of equilibrium exchange to US depends on the size of energetic barriers orthogonal to the selected reaction coordinate. While we agree with this claim, we emphasize that, in general, the effective size of energetic barriers orthogonal to the selected reaction coordinate depends on

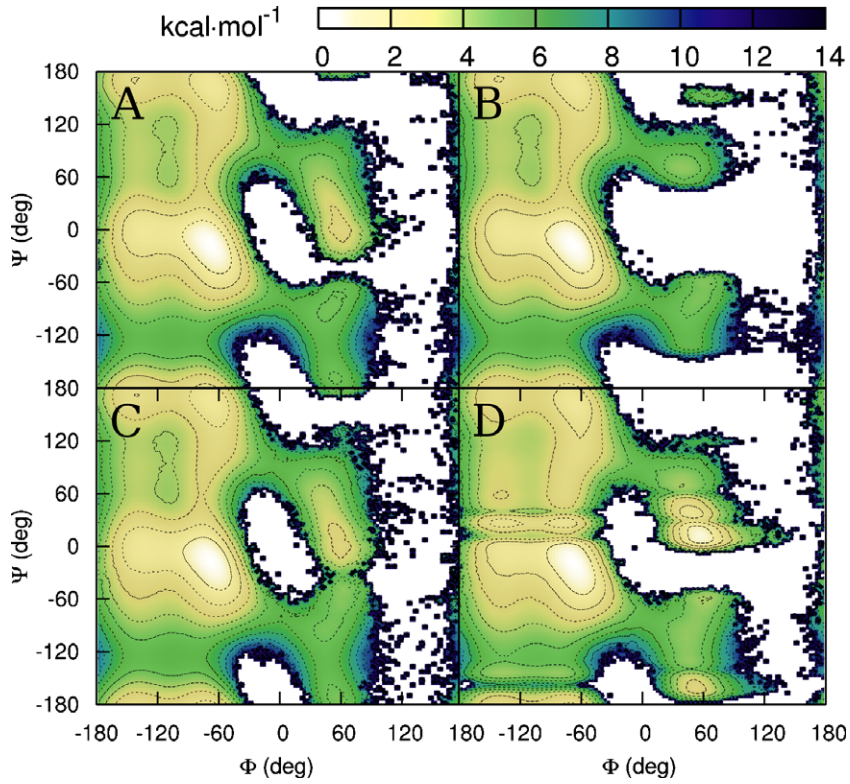

Fig. 6. 2D free energy profiles generated by 1D US or DRUS along ψ. Contours are drawn at 1 kcal mol^{-1} intervals. (A) DRUS and (B) US initiated near $\varphi = -70°$, (C) DRUS and (D) US initiated near $\varphi = +60°$.

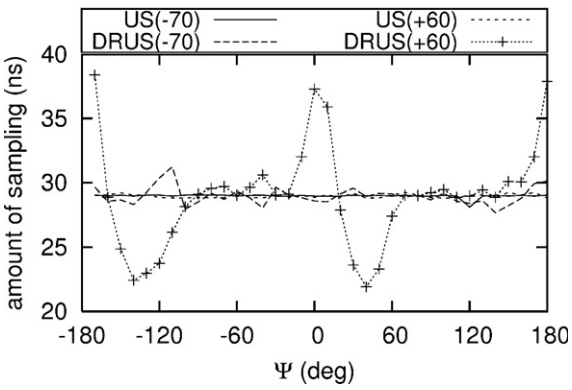

Fig. 7. Sampling density across the reaction coordinate ψ during (solid line) US initiated near $\varphi = -70°$, (long dash line) DRUS initiated near $\varphi = -70°$, (short dash line) US initiated near $\varphi = +60°$, and (dotted line) DRUS initiated near $\varphi = +60°$.

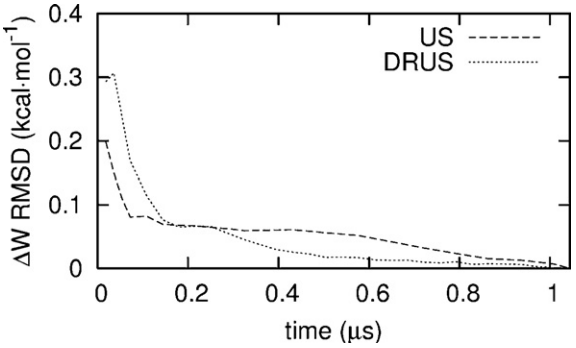

Fig. 8. RMSD between corresponding elements of final and partial-time $W(\psi)$ difference matrices based on (dashed line) US, and (dotted line) DRUS along ψ initiated near $\varphi = +60°$.

sampling confinement along the reaction coordinate. This is because restraint potentials, without exchange, may require that trajectories pass over, and not around, a local maximum on the free energy surface in order to transit between local minima. Conversely, exchange-capable US facilitates the occurrence of transitions between local minima orthogonal to the reaction coordinate by allowing trajectories to explore the reaction coordinate and find lower energy pathways. The adverse influence of static umbrellas on the sampling of orthogonal coordinates is illustrated by the inability of US along ψ to reproduce the reference $W(\psi)$ within 1 μs when initial conformations are separated from energy minima by energetic barriers lying orthogonal to the selected reaction coordinate (Fig. 5A). Specifically, US along ψ forbids transitions between α_R and α_L that utilize the saddle points at (0°, −75°) and (0°, 90°).

In principle, free energy calculations should target the relevant degree (s) of freedom with the highest energetic barriers and sample long enough to generate statistical convergence along all other degrees of freedom. In practice, however, these degrees of freedom are not necessarily known a priori, underscoring the need for sampling methods that, at the very least, do not impede sampling in degrees of freedom orthogonal to the reaction coordinate. Although adding equilibrium exchange to US improves convergence, additional reaction coordinates may still be required when attempting to obtain information from multiple energetically unfavorable states for which a 1D reaction coordinate is not easily developed. This requirement is illustrated by the inability to precisely determine the free energy profile along φ between $\varphi = 100°$ and $\varphi = 160°$ when using ψ as a reaction coordinate in simulations lasting 1 μs (Fig. 4). A similar argument has been made for increasing the dimensionality of the reaction coordinate, even when only a single degree of freedom is of interest [38,39].

The generation of a highly-converged reference free energy surface by 2DUS was possible in this study due to the few degrees of freedom available to alanine dipeptide. We were therefore able to evaluate the convergence of US and DRUS free energy profiles using nearly exact reference solutions. However, studies undertaken with complex systems, for which external reference solutions are, in general, unavailable, must evaluate convergence internally. Fig. 8 shows the convergence of the free energy profiles along ψ generated by US and DRUS along ψ, starting from conformations in the secondary free energy valley, to their own final free energy profiles. Here, projections of 2DUS results were not used. Importantly, while the free energy profile along ψ from the US simulation is relatively constant throughout the last 0.9 μs (Fig. 8), it is systematically incorrect (Fig. 5A), illustrating how barriers to conformational rearrangement arising from artificial restriction of movement along the reaction coordinate can cause systematic sampling errors on simulation timescales. Although the sampling from umbrellas that remained trapped in the α_L and (60°, 180°) free energy basins during the 1 μs simulation would eventually undergo transitions to the α_R and β basins, it is likely that the simulation would be judged complete and stopped prematurely.

5. Conclusion

We have presented a detailed analysis of inherent sampling efficiencies that are suppressed by independent simulations along a predefined reaction coordinate during US, but that can be restored by the addition of equilibrium exchange criteria. While US effectively enhances sampling of energetically unfavorable states along the reaction coordinate, artificial restriction of sampling to distinct regions of the reaction coordinate forbids some types of conformational rearrangement. As a result, transition pathways that utilize both reaction and orthogonal coordinates to connect energetically favorable conformations are suppressed by US. Our main result is that equilibrium exchange partially restores these transition paths. We therefore expect that the addition of equilibrium exchange to any approach that restricts motion along a predefined reaction coordinate will help alleviate systematic sampling errors and increase the rate of convergence in many disparate systems of biological interest.

Acknowledgements

The authors thank Sarah Rauscher for stimulating discussions regarding A-values in ST-like methods. This work was made possible by the facilities of the Centre for Computational Biology High Performance Facility (CCBHPF) at the Hospital for Sick Children and the Shared Hierarchical Academic Research Computing Network (SHARCNET). C.N. is funded by the Research Training Center at the Hospital for Sick Children and by the University of Toronto. This work was funded in part by CIHR Operating Grant MOP-43998. R.P. is a CRCP chair holder.

References

[1] B. Ensing, M. De Vivo, Z. Liu, P. Moore, M.L. Klein, Acc. Chem. Res. 39 (2006) 73.
[2] A.P. Lyubartsev, A.A. Martsinovski, S.V. Shevkunov, P.N. Vorontsov-Velyaminov, J. Chem. Phys. 96 (1992) 1776.
[3] E. Marinari, G. Parisi, Europhys. Lett. 19 (1992) 451.
[4] M. Hagen, B. Kim, P. Liu, R.A. Friesner, B.J. Berne, J. Phys. Chem. B 111 (2007) 1416.
[5] U.H.E. Hansmann, Y. Okamoto, Curr. Opin. Struct. Biol. 9 (1999) 177.
[6] A. Mitsutake, Y. Sugita, Y. Okamoto, Biopolymers 60 (2001) 96.
[7] Y. Sugita, Y. Okamoto, Chem. Phys. Lett. 314 (1999) 141.

[8] D.W. Rebertus, B.J. Berne, D. Chandler, J. Chem. Phys. 70 (1979) 3395.
[9] M. Mezei, J. Comput. Phys. 68 (1987) 237.
[10] A. Laio, M. Parrinello, Proc. Nat. Acad. Sci. USA 99 (2002) 12562.
[11] G.M. Torrie, J.P. Valleau, J. Comput. Phys. 23 (1977) 187.
[12] B. Roux, Comput. Phys. Commun. 91 (1995) 275.
[13] Y. Sugita, A. Kitao, Y. Okamoto, J. Chem. Phys. 113 (2000) 6042.
[14] K. Murata, Y. Sugita, Y. Okamoto, Chem. Phys. Lett. 385 (2004) 1.
[15] H. Lou, R.I. Cukier, J. Phys. Chem. B 110 (2006) 24121.
[16] L. Su, R.I. Cukier, J. Phys. Chem. B 111 (2007) 12310.
[17] J. Norberg, L. Nilsson, Chem. Phys. Lett. 393 (2004) 282.
[18] T. Rodinger, P.L. Howell, R. Pomès, J. Chem. Theory Comput. 2 (2006) 725.
[19] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, J. Chem. Phys. 21 (1953) 1087.
[20] S. Duane, A.D. Kennedy, B.J. Pendleton, D. Roweth, Phys. Lett. B 195 (1987) 216.
[21] E. Lindahl, B. Hess, D. van der Spoel, J. Mol. Model. 7 (2001) 306.
[22] W.D. Cornell et al., J. Am. Chem. Soc. 117 (1995) 5179.
[23] E.J. Sorin, Y.M. Rhee, V.S. Pande, Biophys. J. 88 (2005) 2516.
[24] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys. 79 (1983) 926.
[25] W.F. van Gunsteren, H.J.C. Berendsen, Agnew. Chem. Int. Ed. Engl. 29 (1990) 992.
[26] T. Darden, D. York, L. Pedersen, J. Chem. Phys. 98 (1993) 10089.
[27] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, J. Chem. Phys. 103 (1995) 8577.
[28] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, J. Chem. Phys. 81 (1984) 3684.
[29] S. Miyamoto, P.A. Kollman, J. Comput. Chem. 13 (1992) 952.
[30] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, J. Comput. Chem. 18 (1997) 1463.
[31] S. Kumar, D. Bouzida, R.H. Swendsen, P.A. Kollman, J.M. Rosenberg, J. Comput. Chem. 13 (1992) 1011.
[32] S. Kumar, J.M. Rosenberg, D. Bouzida, R.H. Swendsen, P.A. Kollman, J. Comput. Chem. 16 (1995) 1339.
[33] M. Souaille, B. Roux, Comput. Phys. Commun. 135 (2001) 40.
[34] G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, J. Mol. Biol. 7 (1963) 95.
[35] V. Spiwok, P. Lipovová, B. Králová, J. Phys. Chem. B 111 (2007) 3073.
[36] P.G. Bolhuis, C. Dellago, D. Chandler, Proc. Nat. Acad. Sci. USA 97 (2000) 5877.
[37] J.W. Ponder, D.A. Case, Adv. Protein Chem. 66 (2003) 27.
[38] D.L. Mobley, J.D. Chodera, K.A. Dill, J. Chem. Phys. 125 (2006) 084902.
[39] S. Crouzy, T.B. Woolf, B. Roux, Biophys. J. 67 (1994) 1370.