Coupling of Replica Exchange Simulations to a Non-Boltzmann Structure Reservoir

3 AUTHORS, INCLUDING:

Adrian Roitberg
University of Florida

164 PUBLICATIONS **7,515** CITATIONS

Asim Okur
National Heart, Lung, and Blood Institute

19 PUBLICATIONS **3,108** CITATIONS

Coupling of Replica Exchange Simulations to a Non-Boltzmann Structure Reservoir

Adrian E. Roitberg,*,† Asim Okur,‡ and Carlos Simmerling*,‡,§

Quantum Theory Project, Department of Chemistry, University of Florida, Gainesville, Florida 32611, Department of Chemistry, and Center for Structural Biology, Stony Brook University, Stony Brook, New York 11794

Received: December 4, 2006; In Final Form: January 23, 2007

Computing converged ensemble properties remains challenging for large biomolecules. Replica exchange molecular dynamics (REMD) can significantly increase the efficiency of conformational sampling by using high temperatures to escape kinetic traps. Several groups, including ours, introduced the idea of coupling replica exchange to a pre-converged, Boltzmann-populated reservoir, usually at a temperature higher than that of the highest temperature replica. This procedure reduces computational cost because the long simulation times needed for extensive sampling are only carried out for a single temperature. However, a weakness of the approach is that the Boltzmann-weighted reservoir can still be difficult to generate. We now present the idea of employing a non-Boltzmann reservoir, whose structures can be generated through more efficient conformational sampling methods. We demonstrate that the approach is rigorous and derive a correct statistical mechanical exchange criterion between the reservoir and the replicas that drives Boltzmann-weighted probabilities for the replicas. We test this approach on the trpzip2 peptide and demonstrate that the resulting thermal stability profile is essentially indistinguishable from that obtained using very long (>100 ns) standard REMD simulations. The convergence of this reservoir-aided REMD is significantly faster than for regular REMD. Furthermore, we demonstrate that modification of the exchange criterion is essential; REMD simulations using a standard exchange function with the non-Boltzmann reservoir produced incorrect results.

Conformational sampling is crucial to simulating biologically relevant events in atomic detail, particularly when correct Boltzmann-weighted populations are required. Trapping in local energy minima often impairs complete exploration of conformational space. The challenges and improvements in conformational sampling have been discussed in several reviews.[1,2]

One popular approach to overcoming poor sampling in biomolecular simulation is the replica exchange method (also known as parallel tempering).[3−5] In replica exchange molecular dynamics (REMD),[6] a series of molecular dynamics simulations (replicas) are performed for the system of interest. In the original form of REMD, each replica is an independent realization of the system, coupled to a thermostat at a different temperature. The temperatures of the replicas span a range from low values of interest up to high values at which the system is expected to rapidly overcome potential energy barriers. At intervals during the simulations, exchanges are attempted between conformations of the system being sampled by different replicas. These exchanges are based on a Metropolis-type criterion[7] that considers the probability of sampling each conformation at the alternate temperature. In this manner, REMD allows the conformations sampled at low temperatures to escape kinetic traps by "jumping" directly to alternate minima being sampled

at higher temperatures. Importantly, the transition probability is constructed such that the canonical ensemble properties are maintained for each replica, providing useful information about conformational probabilities as a function of temperature. Thus, REMD employs high-temperature simulations for improved sampling and a series of coupled simulations over a range of temperatures to reweight the ensembles to the lower temperatures of interest. REMD has been widely applied to studies of peptide and small protein folding.[3,6,8−17]

For large systems, REMD becomes intractable because the number of replicas needed to span a given temperature range increases with the square root of the number of degrees of freedom in the system.[18−21] Several promising techniques have been proposed to deal with this apparent disadvantage of REMD.[18,22−27] When starting from non-native conformations, high-temperature replicas give limited advantage for finding native states because more minima on the free energy landscape become accessible, further complicating the search. When a high-temperature REMD replica locates a favorable low-energy basin (such as the native structure), this conformation is exchanged to a lower temperature with very high probability, and the high-temperature replica needs to repeat the search process. This search process happens in a time scale related to the folding time (which is often not strongly temperature dependent). As a result, the entire set of replicas is simulated for timescales on the order of the folding time even though the conformational search improvement of REMD is generally due only to the highest temperatures.

---

* Corresponding author. E-mail: roitberg@qtp.ufl.edu (A.E.R.); carlos.simmerling@stonybrook.edu (C.S.).
† Quantum Theory Project.
‡ Department of Chemistry, Stony Brook University.
§ Center for Structural Biology, Stony Brook University.

2416 *J. Phys. Chem. B, Vol. 111, No. 10, 2007*

Letters

Recognizing this weakness, we and others[26−28] have introduced a method that separates the conformational sampling and reweighting aspects of the REMD replicas by using extensive simulations at a single high temperature to generate an ensemble of structures. Because the probabilities of alternate conformations at this high temperature are not the same as at the low temperature of interest, the ensemble is reweighted by subsequent coupling to REMD. Periodic exchanges are made between randomly chosen conformations from the reservoir set and the highest temperature replica, in a manner analogous to J-walking.[29] This process formally provides correct ensembles at lower temperature with free energies that reflect the proper relative populations of minima. We called this method Reservoir REMD (R-REMD) because REMD is coupled to a high-temperature reservoir.[28]

One major advantage of the reservoir approach with REMD is that a converged ensemble of conformations has to be generated for only one temperature, as compared to standard time-correlated REMD where all replicas are simulated during the entire time that the highest temperatures perform efficient searches. In R-REMD, after extensive conformational searching at one temperature, the remaining temperatures feed from this ensemble and anneal these structures to rapidly construct equilibrium distributions consistent with their own temperatures. As a result of coupling to a pre-converged ensemble, the R-REMD simulation converges much more rapidly than standard temperature REMD.

Although existing reservoir REMD schemes provide substantial improvement over the original REMD, the problem of generating a converged, Boltzmann-populated reservoir is by no means trivial, especially when an explicit water model is employed. This is an important point because the cost of reservoir generation must be included in the total effort required for the simulation. We recognize that it is very likely to be much less challenging to generate a set of low-energy structures than to obtain an ensemble in which they are all present in correct relative populations. It is in this spirit that we demonstrate that a reservoir with a non-Boltzmann population can nonetheless be used to rapidly and efficiently converge REMD simulations. We start by re-deriving the REMD exchange probability equation, with emphasis on the steps that involve reservoir driven REMD. We demonstrate that *any* complete reservoir can be employed as long as the probability distribution is known, and correct Boltzmann-weighted ensembles can be obtained for the replicas.

We consider N replicas (R_1 to R_N) associated with N temperatures (T_1 to T_N). At time t, each replica is found with coordinate X_i, and temperature T_m, which we will write as X_i^m. We denote the set of coordinates for all replicas at time t as the vector $\mathbf{Q} = \{X_i^m\}$ with both i and m indices bounded between 1 and N. The existence of a *well-defined limiting probability distribution* of structures at each temperature is assumed, written as $P(X_i^m)$. *Importantly, the distributions for each temperature need not obey the same rules.*[30] Because the replicas are non-interacting, the probability of finding coordinate set $\mathbf{Q}$ at the current distribution of temperatures can be written as a direct product.

We now consider the exchange of a pair of replicas, between coordinates X_j at temperature s (X_j^s) and coordinate X_k at temperature t (X_k^t). The rest of the replicas are not exchanged and hence stay at their previous temperatures. Ensuring microscopic reversibility, we require that the forward and backward transition rates are equal to each other

$$W(X_1^1,...,X_j^s,X_k^t,...,X_N^N \rightarrow X_1^1,...,X_j^t,X_k^s,...,X_N^N)$$
$$P(X_1^1,...,X_j^s,X_k^t,...,X_N^N) =$$
$$W(X_1^1,...,X_k^s,X_j^t,...,X_N^N \rightarrow X_1^1,...,X_j^s,X_k^t,...,X_N^N)\, P(X_1^1,...,X_k^s,X_j^t,...,X_N^N)$$

$$(1)$$

which, by renaming the transition probabilities and exploiting the ability to express the populations as products, yields

$$\frac{W(X_j^s,X_k^t \rightarrow X_j^t,X_k^s)}{W(X_k^s,X_j^t \rightarrow X_j^s,X_k^t)} = \frac{P(X_k^s)\,P(X_j^t)}{P(X_j^s)\,P(X_k^t)} \qquad (2)$$

The next step, which is key to the present algorithm, is the insertion of the limiting probability distributions for each temperature. In the particular case of Boltzmann populations in the canonical ensemble, with $P(X_k^s) \propto e^{-\beta_s E(X_k)}$ and $\beta_s = (k_B T_s)^{-1}$, the exchange criteria equation becomes the well-known form of standard temperature REMD:

$$\frac{W(X_j^s,X_k^t \rightarrow X_j^t,X_k^s)}{W(X_k^s,X_j^t \rightarrow X_j^s,X_k^t)} = e^{(\beta_t-\beta_s)(E_k-E_j)} \qquad (3)$$

In the case of reservoir REMD, the equations for exchange resemble those of regular REMD, with the only difference being that the highest replica has been pre-obtained and it is not time-correlated with the others. This means that at every exchange point, the $(N-1)$-th replica (or others also if desired) attempts to exchange with one of the members of the reservoir. Our present focus is to permit a different probability distribution for the reservoir.

Suppose that for replicas 1 to $N-1$, we desire a Boltzmann population of structures. Let us assume also that the reservoir has the limiting distribution $P_R(X)$. It is clear that the derivation above would produce exchange probabilities exactly like those in REMD if none of the two replicas involved is the reservoir, but it would create a different condition when attempting to exchange between the reservoir and any regular replica.

Let us assume that one has a reservoir with M members, where each structure is present exactly one time, as would be obtained using a method such as a grid-based search. In this case, using the nomenclature above, the probability of finding structure i in the reservoir of size M is $P(X_i^R) = 1/M$. Under these conditions, the exchange probability between replicas *not* involving the reservoir obeys the standard $e^{(\beta_t-\beta_s)(E_k-E_j)}$ expression, and exchanges between any real replica and the reservoir must obey

$$\frac{W(X_j^R,X_k^t \rightarrow X_j^t,X_k^R)}{W(X_k^R,X_j^t \rightarrow X_k^t,X_j^R)} = \frac{P(X_k^R)\,P(X_j^t)}{P(X_j^R)\,P(X_k^t)} = \frac{(1/M)e^{-\beta_t E_j}}{(1/M)e^{-\beta_t E_k}} = e^{-\beta_t(E_j-E_k)}$$

$$(4)$$

This is essentially the same as standard REMD except for the weighting factor for ΔE. The exchange between the reservoir and any temperature replica looks like a regular Metropolis MC step. In REMD terms, this is also equivalent to having the highest replica being at an infinite temperature ($\beta_R = 0$), consistent with a flat probability distribution. In practical terms, reservoir structures generated at very high temperatures would likely have a significant thermal component to the potential energy, and thus the acceptance probability would be negligible.

Although the exchange probability equations can formally be derived for any reservoir of known probability distribution, we focus in this report on the straightforward case described
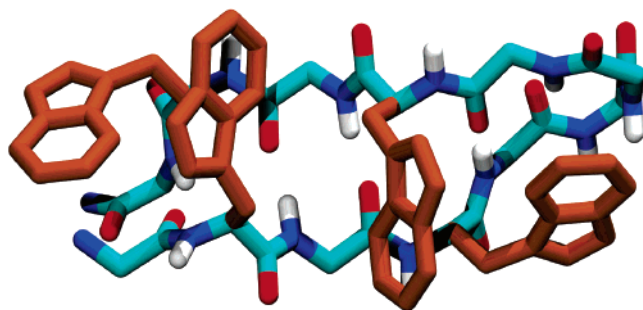
Letters

J. Phys. Chem. B, Vol. 111, No. 10, 2007 **2417**

Figure 1. The trpzip2 model peptide, in the native β-hairpin conformation. Backbone atoms are shown colored by element; for clarity, side chain heavy atoms are only shown for Trp residues (orange).

above in which structures are present in the reservoir in equal probabilities (eq 4). We implemented the resulting algorithm into the widely used Amber program,[31] and we describe below the results obtained for a non-trivial model peptide that we have previously studied using standard and reservoir REMD.[28]

System and Results

In order to test the results of the modified exchange criterion (eq 4), we studied the tryptophan zipper peptide that we have previously sampled using extensive standard REMD simulations[28] as well as REMD with a Boltzmann-weighted reservoir.[28] Developed by Starovasnik and co-workers,[32] trpzip2 (1LE1) has a high propensity to form a β-hairpin in solution at room temperature (Figure 1). We employed the same force field and simulation methodology as we used for those studies (see Supporting Information). We showed that this combination is able to reproduce not only the experimentally observed structure of trpzip2 but also the temperature-dependent stability. Thus, the peptide provides an excellent model for testing sampling algorithms under conditions that are directly relevant to experimental data.

Our previous work using a Boltzmann-weighted reservoir for trpzip2 REMD employed a reservoir of 10 000 structures that were extracted from standard MD simulations at 400 K. This was shown to be sufficient to properly seed the REMD exchange simulations at lower temperatures when the standard exchange criterion (eq 3) was employed. We demonstrated that coupling to this reservoir provided significantly more rapid convergence, with the same thermal stability profile as standard REMD. To test whether similar results could be obtained with the non-Boltzmann reservoir, we performed a cluster analysis on the Boltzmann-weighted reservoir. We generated a new, non-Boltzmann weighted ensemble by extracting *only the representative structures* from each of the 700 clusters. This new "flat" reservoir has by definition a probability of 1/700 for each of the members, because the additional population of each cluster was discarded. Although the original Boltzmann-weighted reservoir had approximately 3% native structures, the native hairpin is adopted by only 1 of the 700 structures in the non-Boltzmann reservoir. We note that clustering only approximates the flat distribution implied in eq 4, because individual conformational basins are represented by single structures rather than a continuous probability density.

The main result of this letter is presented in Figure 2. Melting curves were generated for several simulations by calculating the average population of native structures at each temperature. The values are computed over the entire (post-equilibration) REMD ensemble. The simulations include standard REMD (black line, from our previously published data)[28] along with two simulations in which the REMD replicas were coupled to
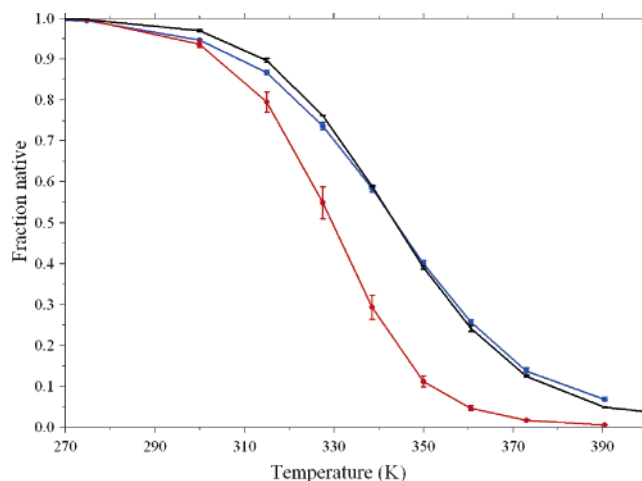

Figure 2. Thermal stability profiles for trpzip2 obtained from REMD simulations: black, standard REMD (150 ns); red, REMD with non-Boltzmann reservoir and standard (eq 3) exchange probability (20 ns); blue, REMD with non-Boltzmann reservoir and modified (eq 4) exchange probability (25 ns). Error bars indicate lower bounds to uncertainty, obtained from the difference between data calculated from first and last half of the ensemble.

the non-Boltzmann ensemble. In one simulation, the standard exchange criterion (eq 3) was used (red line) to couple the reservoir and the replicas, and the other (blue line) employed an exchange criterion that accounted for the non-Boltzmann probability distribution in the reservoir (eq 4). We observe that the thermal stability profile obtained using the modified exchange criterion is essentially indistinguishable from that obtained using standard REMD simulations without a reservoir.

Our main point is to demonstrate that the REMD exchange criterion can be successfully modified to rigorously incorporate non-Boltzmann-weighted reservoirs without any detrimental effect on the equilibrium populations sampled by the temperature replicas. The reservoir that we generated has the same structure types as our Boltzmann-weighted reservoir (only the relative populations changed from the original values to the $1/M$ "flat" reservoir). One might therefore ask if the change in exchange acceptance criterion is truly needed. To address this point, we repeated the reservoir REMD calculation with the 700 member non-Boltzmann-weighted reservoir but used the standard REMD exchange acceptance criterion (eq 3 instead of the modified version (eq 4)). As shown in Figure 2, this procedure resulted in significant errors in the thermal profile over the entire range simulated due to the different scaling factor applied to the potential energy difference in the calculation of exchange probability. The $\Delta\beta$ scaling factor in standard REMD (eq 3) is smaller than the β-term in the criterion modified for the "flat" ensemble (eq 4); thus, the low-energy native conformation is accepted too infrequently with the standard criterion. It is interesting to note that this strongly affects not only the ensemble sampled at 390 K, where exchanges with the reservoir occur, but the ensembles at all other temperatures also show a depression of native stability when the incorrect exchange criterion is used. As shown in Figure 2, the use of an exchange criterion that directly accounts for the probability distribution in the reservoir essentially eliminates this error (despite the approximate nature of the clustering approach to "flat" reservoir generation, as noted above).

Although standard REMD converges to the proper native population, the new method also does so but at a much faster rate. Figure 3 shows the population of the native state at the melting temperature versus simulation time. The colors are the
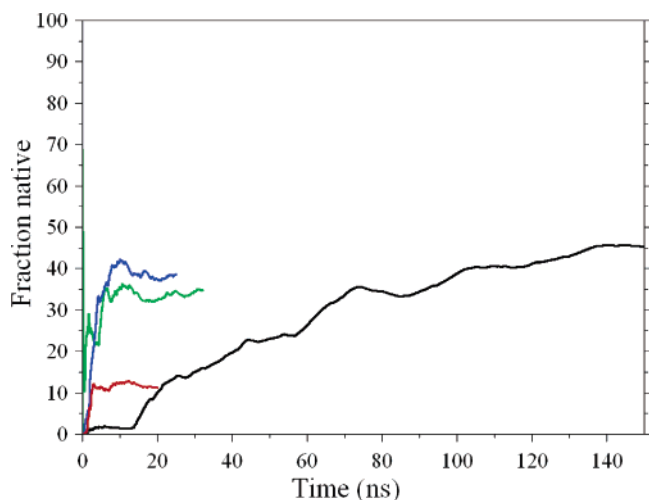
Figure 3. Fraction of native hairpin conformation as a function of time at 350 K for four REMD simulations: black, standard REMD; red, REMD with non-Boltzmann reservoir and standard exchange probability (eq 3); blue, REMD with non-Boltzmann reservoir and modified exchange probability (eq 4); green, REMD with Boltzmann reservoir and standard exchange probability (eq 3). It is apparent that the simulations employing the reservoir reach a plateau value much more rapidly than standard REMD and that the non-Boltzmann reservoir requires modified exchange probability.

same as those in Figure 2. It is clear that a speeding up of the order of $10\times$ is obtained for this particular system by using the reservoir. A similar speeding up is obtained with the unmodified exchange criterion (eq 3), but the final population has significant error. The speeding up from using Boltzmann- or non-Boltzmann-weighted reservoirs is comparable (Figure 3).

In summary, we have demonstrated that REMD simulations can be rigorously coupled to a reservoir with an arbitrary, known probability distribution, while still efficiently driving the REMD replicas toward Boltzmann-weighted ensembles. We employed a cluster-based approach to reservoir generation but point out that the procedure we used to derive eq 4 will also enable the use of more general methods for generation of structure reservoirs, at significantly reduced cost than previously reported Boltzmann-weighted reservoirs. We expect that this procedure will dramatically extend the range of problems that are amenable to study through REMD simulation.

Supporting Information Available: Additional experimental information. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Tai, K. *Biophys. Chem.* **2004**, *107*, 213.

(2) Roitberg, A.; Simmerling, C. *J. Mol. Graphics Modell.* **2004**, *22*, 317.

(3) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140.

(4) Swendsen, R. H.; Wang, J. S. *Phys. Rev. Lett.* **1986**, *57*, 2607.

(5) Tesi, M. C.; vanRensburg, E. J. J.; Orlandini, E.; Whittington, S. G. *J. Stat. Phys.* **1996**, *82*, 155.

(6) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.

(7) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.

(8) Feig, M.; Karanicolas, J.; Brooks, C. L. *J. Mol. Graphics Modell.* **2004**, *22*, 377.

(9) Garcia, A. E.; Sanbonmatsu, K. Y. *Proteins: Struct., Funct., Genet.* **2001**, *42*, 345.

(10) Garcia, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782.

(11) Karanicolas, J.; Brooks, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3954.

(12) Kinnear, B. S.; Jarrold, M. F.; Hansmann, U. H. E. *J. Mol. Graphics Modell.* **2004**, *22*, 397.

(13) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587.

(14) Roe, D. R.; Hornak, V.; Simmerling, C. *J. Mol. Biol.* **2005**, *352*, 370.

(15) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042.

(16) Zhou, R. H.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931.

(17) Wickstrom, L.; Okur, A.; Song, K.; Hornak, V.; Raleigh, D. P.; Simmerling, C. L. *J. Mol. Biol.* **2006**, *360*, 1094.

(18) Cheng, X.; Cui, G.; Hornak, V.; Simmerling, C. *J. Phys. Chem. B* **2005**, *109*, 8220.

(19) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058.

(20) Kofke, D. A. *J. Chem. Phys.* **2002**, *117*, 6911.

(21) Rathore, N.; Chopra, M.; de Pablo, J. J. *J. Chem. Phys.* **2005**, *122*, 024111.

(22) Jang, S. M.; Shin, S.; Pak, Y. *Phys. Rev. Lett.* **2003**, *91*, 058305.

(23) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *J. Chem. Phys.* **2003**, *118*, 6664.

(24) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **2000**, *329*, 261.

(25) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comput.* **2006**, *2*, 420.

(26) Li, H. Z.; Li, G. H.; Berg, B. A.; Yang, W. *J. Chem. Phys.* **2006**, *125*.

(27) Lyman, E.; Ytreberg, F. M.; Zuckerman, D. M. *Phys. Rev. Lett.* **2006**, *96*.

(28) Okur, A.; Roe, D.; Cui, G.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comput.* In Press.

(29) Frantz, D. D.; Freeman, D. L.; Doll, J. D. *J. Chem. Phys.* **1990**, *93*, 2769.

(30) Geyer, C. J. Markov Chain Monte Carlo Maximum Likelihood. In *Proceedings of the 23rd Symposium on the Interface.* Keramidas, E. M., Ed.; 1991; p 156.

(31) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.

(32) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5578.