Title
How hot? Systematic convergence of the replica exchange method using multiple reservoirs

Journal
Journal of Computational Chemistry, 31(3)

ISSN
0192-8651

Authors
Ruscio, Jory Z
Fawzi, Nicolas L
Head-Gordon, Teresa

Publication Date
2009

DOI
10.1002/jcc.21355

Peer reviewed

How Hot? Systematic Convergence of the Replica Exchange Method using Multiple Reservoirs

Jory Z. Ruscio[1], Nicolas L. Fawzi[2], and Teresa Head-Gordon[1,2*]
[1]*Department of Bioengineering, University of California, Berkeley*
[2]*UCSF/UCB Joint Graduate Group in Bioengineering*
Berkeley, California 94720 USA

We have devised a systematic approach to converge a replica exchange molecular dynamics simulation by dividing the full temperature range into a series of higher temperature reservoirs and a finite number of lower temperature sub-replicas. A defined highest temperature reservoir of equilibrium conformations is used to help converge a lower but still hot temperature sub-replica, which in turn serves as the high temperature reservoir for the next set of lower temperature sub-replicas. The process is continued until an optimal temperature reservoir is reached to converge the simulation at the target temperature. This gradual convergence of sub-replicas allows for better and faster convergence at the temperature of interest and all intermediate temperatures for thermodynamic analysis, as well as optimizing the use of multiple processors. We illustrate the overall effectiveness of our multiple reservoir replica exchange strategy by comparing sampling and computational efficiency with respect to replica exchange, as well as comparing methods when converging the structural ensemble of the disordered $A\beta_{21-30}$ peptide simulated with explicit water by comparing calculated NMR ROESY intensities to experimentally measured values.

**Corresponding author*

INTRODUCTION

Replica exchange (RE) molecular dynamics[1-5] for protein simulations has become a standard enhanced sampling tool used by the biomolecular simulation community. However in the most standard implementation of RE, various sub-optimal traits are becoming apparent, curtailing its sampling efficiency and practicality. Most relevant to the sampling efficiency is the recognition that too many replicas[6,7], use of replicas at very high temperature[8,9], system size[10,11], and frequency of configurational exchange among replicas[12], can actually impede the convergence of the (lower) target temperature replica. In addition, there is also a practical question in regards the best way to use RE on parallel hardware platforms[13,14]- one of its perceived strengths- that must consider the optimal use of fine-grained parallelization of energy and forces vs. the coarse-grained parallelization of multiple temperatures, and the realities of supercomputer center queues, contention for processors among multiple projects on a research group cluster, and turn-around times for results.

While very high temperature replicas are effective when broken ergodicity is a genuine problem[15-18], the sampling efficiency in biomolecular simulations is more limited by the large number of degrees of freedom (protein and water), rather than insurmountable energy barriers[19-22]. In practice this means that the maximum temperature replica in biomolecular simulation need not be as high as ~500-600K, as is commonly used in the literature, and should be lowered in order to better sense the relevant distinct free energy basins (and thus their relative probabilities) when folding small proteins in molecular solvent[8,9]. The number of intermediate temperature replicas can also affect sampling efficiency[16,19,6.] First because any given replica that is not at equilibrium will keep all other replicas out of equilibrium as well- so a great deal of computational time is wasted on bringing all replicas into equilibrium at their respective temperatures. Second, if the transition rates between temperatures is infrequent so that RE barrier hopping is slower or much not much faster than the primary relaxation time for sampling distinct energy basins within the target replica, then increasing the number of replicas will impede sampling performance. The finite length of typical 10-100ns trajectories for fully atomistic simulations of large system size, combined with conservative exchange frequencies between replicas, can reduce the system to diffusive barrier crossing. This is likely why some increase in exchange frequency actually improved sampling performance[12] to more quickly draw down configurations that sample different target temperature basins. However, too frequent generation of trial moves from the high temperature replica can pose

the problem that accepted configurations are highly correlated, overemphasizing the importance of some free energy basin, thereby degrading the sampling of the true limiting distribution at the lower temperature.

In the original formulation of the J-walking technique[17], Frantz and co-workers were the first to suggest that a well-converged high temperature reservoir that is accessed randomly for trial moves can break up unwanted correlations to aid convergence of a target temperature replica. Enhanced sampling methods such as Cool-walking[16], Smart-walking[18], Smart-darting[15], Annealed Swapping[23], and more recent variations[24] all use an extended high temperature reservoir to draw down configurations, but differ primarily in how those configurations are manipulated to improve acceptance, how many intermediate temperature replicas are used, and whether the method does or does not satisfy detailed balance. Recently, in the context of a multicanonical simulation performed on two model peptides in a generalized Born model of solvent, Simmerling et al. have shown that trial moves generated from a high temperature reservoir of 400K can in principle ameliorate many of the convergence problems of a standard RE simulation[25,26].

In this paper we describe an expanded use of reservoir RE that allows for improved convergence over standard RE, decreased total computational time, and more practical parallelization, by dividing the full temperature range into a series of higher temperature reservoirs and a finite number of associated lower temperature sub-replicas. Starting with a defined highest temperature reservoir of equilibrium conformations, it is used to help converge a lower but still hot temperature sub-replica. The converged low temperature sub-replica in turn serves as the high temperature reservoir for the next set of lower temperature sub-replicas, and the process is continued until an optimal higher temperature reservoir is reached to converge the simulation at the lower target temperature. This gradual convergence of sub-replicas allows for better and faster convergence at the temperature of interest and all intermediate temperatures for thermodynamic analysis, as well as optimizing the use of multiple processors for cluster platforms[13,27]. We illustrate the overall effectiveness of our multiple reservoir replica exchange (MRRE) strategy in three ways: (1) with a direct internal comparison of sampling similarity between the RE reference simulation with a single reservoir replica exchange (SRRE) and two different manifestations of the MRRE approach, (2) by evaluating both cpu and wall time savings of the MRRE methods over SRRE and standard RE, and (3) by external validation by comparing the prediction accuracy among the methods for calculating NMR ROESY intensities for the disordered $A\beta_{2130}$ peptide in explicit water. This system and its calculated observables are an exceptionally challenging test for REMD and its

variants, and if successful gives hope for continued use of the REMD method on larger unstructured peptides and proteins. We find that a high temperature reservoir of 320K actually provides overall improved performance in both computational and sampling efficiency relative to higher temperature reservoirs or standard replica exchange for the $A\beta_{21-30}$ peptide system.

METHODS

Standard Replica Exchange and measurement of sampling efficiency. We have recently completed a simulation study of the conformational ensemble of $A\beta_{21-30}$ in explicit water to calculate NMR observables at 284K to compare to experimental data taken on the same system. To converge the ensemble, we used a standard replica exchange (RE) molecular dynamics method in which 64 temperature replicas were exponentially spaced between 270-508K. In that work we used an estimate of ensemble convergence originally defined by Thirumalai and co-workers to measure ergodicity[33].

$$d_{12} = \sum_{i<j}^{N} \left\langle r_{ij} \right\rangle_1 - \left\langle r_{ij} \right\rangle_2 \tag{1}$$

where subscripts 1 and 2 label two independent trajectories, $<r_{ij}>$ is the hydrogen-to-hydrogen linearly averaged distance for that trajectory, and the sum is over all *i,j* hydrogen-hydrogen pairs in the peptide system. In addition to estimating sampling convergence within a method using Eq. (1), we will also consider a variant of that metric

$$d_{XRE,RE} = \sum_{i<j}^{N} \left\langle r_{ij} \right\rangle_{XRE} - \left\langle r_{ij} \right\rangle_{RE} \tag{2}$$

in which we compare SRRE, MRRE-12, or MRRE-8 (XRE) to the RE simulation where $<r_{ij}>$ is averaged over both independent trajectories for a given method, in order to compare sampling efficiency against the RE reference.

Single Reservoir Replica Exchange. Based on the published reference RE simulation, we found that the 400K replica was the lowest “high-temperature” replica in which linear averaged distances between all hydrogens in the system (Eq. (1)) were found to be less than 5%. Therefore we initiated the SRRE method by conducting an NVT simulation at 400K at the corresponding density (found from an NPT simulation) to fill a single 400K reservoir with 12,225 structures chosen at regular time intervals over a 14ns trajectory. It serves as the high temperature reservoir that is combined with 20 additional lower temperature replicas down to the target temperature of

284K. Using the AMBER program (see details below), we conduct two SRRE simulations from two independent start states to check for convergence as per Eq. (1), in which both simulations use the same 400K reservoir. This corresponds to the implementation of the SRRE in reference [25] which we use to understand the drawbacks, if any, of using a small, shared reservoir.

Multiple Reservoir Replica Exchange. In our multiple reservoir replica exchange (MRRE) approach, a temperature range is divided into multiple single reservoir groups that are simulated in a serial fashion. First we define a highest temperature reservoir and four associated lower temperature sub-replicas (Figure 1). In the work presented here, we choose 400K as the highest temperature in which we create a single reservoir of 12,225 configurations (same as the SRRE reservoir creation) from a single long molecular dynamics trajectory. We then define four sub-replicas at temperatures of 393K, 386K, 379K, and 372K that use the 400K reservoir as per the usual SRRE method implemented in AMBER. We conduct two simulations starting from two independent start states for each replica to check for convergence over the 372K-400K temperature range, using the ergodic sampling measure defined in Eq. (1). Once the first set of sub-replicas is found to be converged, the equilibrated structures from the lowest 372K temperature replica of the two independent simulations are then used to populate two reservoirs of 20,000 configurations each for the next set of associated sub-replicas (365K, 358K, 351K, and 344K). We again conduct two simulations starting from two independent start states to check for convergence over the 344K-372K temperature range using Eq. (1).

At 344K, we branch the simulation up to this point into two MRRE implementations. One starts with the equilibrated structures from the lowest 344K temperature replica of the two independent simulations to populate two reservoirs of 20,000 configurations, that are used for the set of 12 sub-replicas starting at 337K down to 284K (MRRE-12). The second does yet another serial round of SRRE simulations, using the two 20,000 configuration reservoirs from the equilibrated 344K ensemble to converge four more replicas down to 320K (338K, 332K, 326K, 320K). The converged 320K replica then serves to populate the final two independent reservoirs of 20,000 structures each for a SRRE simulation down to 284K with 8 replicas (MRRE-8). This branching is done to examine how aggressive we can be in lowering the temperature of the final reservoir without affecting the sampling quality.

Simulation Protocols and Models. We use the AMBER (version 10) package[28] for all of the molecular dynamics simulations reported here. The temperature spacing between replicas was determined by aiming for an exchange probability of 20-30%, resulting in replica temperature

separations of 4-8K. Exchanges are attempted every 1.0ps, in which the odd time values exchange attempts occur between a given replica and its adjacent lower temperature replica and the even time exchange attempts occur between a given replica and its adjacent higher temperature replica. Distinct from the standard RE method in AMBER which satisfies detailed balance, AMBER's implementation of a practical Reservoir RE strategy does not formally satisfy detailed balance. This arises in the case if an exchange attempt between the reservoir and its adjacent lower temperature is successful, there is no exchange move from the replica to the reservoir, only from the reservoir to the replica. We attempt to mitigate this detailed balance violation by ensuring that any individual structure will almost never be sampled twice during a ~40ns simulation due to the large reservoir size of 20,000 structures and a number of trial move attempts that does not deplete the reservoir.

We use the ff99SB[29] protein force field to model the 10-residue Aβ peptide, and the TIP4P-Ew[30] model to describe the molecular water solvent. We determined in previous work that this combination of protein and water force fields yielded excellent agreement between simulated and experimental NMR observables, especially for dynamic observables (e.g. heteronuclear backbone nuclei relaxation "T_1" and "T_2") and dynamic-based observables (nuclear Overhauser effect)[31]. We simulate one Aβ peptide neutralized with one sodium ion and immersed in a box of 1579 TIP4P-Ew water molecules. The system reached the ambient thermodynamic state by first equilibrating at constant volume (1 g/cc) for 20ps with harmonic constraints on the peptide while the temperature was brought to 300K using the Anderson thermostat. This is followed by a 1.0ns NPT simulation with weak barostat coupling at 1 bar (default parameters for Berendsen barostat in AMBER10) to determine the density, and all production runs are fixed at that density in all subsequent NVT simulations. The equations of motion are integrated with 1fs timesteps, the long-range electrostatic interactions are calculated using Particle Mesh Ewald method[32], and a cutoff of 9.0Å is used for real space electrostatics and LJ interactions. We removed the first 5ns, thereby evaluating production statistics over the last ~35ns for each trajectory for each method.

RESULTS

Table 1 shows that the sampling characteristics of the various reservoir methods. The two SRRE simulations combined made 38,000 exchange attempts with the single 12,225 structure reservoir. As the number of attempts is three times greater than the reservoir size, we see that 9,967 of these structures were sampled more than once; in fact, 1,150 structures were sampled more than 5 times. Of the 7,909 reservoir structures that were successfully exchanged, 1,914 reservoir

structures were exchanged more than once, hence only 76% of the exchanged structures were unique. MRRE-12 and MRRE-8 do not suffer from this multiple sampling problem as there are two reservoirs of 20,000 structures each at 344K and 320K- much larger than the number of total exchange attempts, giving ~90% unique exchanged structures. Thus the internal convergence of the SRRE implementation as estimated by Eq. (1) may (or may not) be artificial, and needs to be vetted by other convergence metrics.

To compare the various replica exchange approaches, we first calculate the convergence of hydrogen-to-hydrogen pair distances of the disordered $A\beta_{21\text{-}30}$ peptide simulated with explicit water. In Figure 2 we show the average distance difference between the two independent simulations at 284K for all 1830 hydrogen-hydrogen pairs for each replica exchange method (Eq. (1)). Out of all hydrogen-hydrogen pair distances, the number of distances with a difference of less than 5% is 1820 (99.5%), 1596 (87.2%), 1804 (98.6%), and 1729 (94.5%), while the number of distances with a difference of less than 7% is 1830 (100%), 1785 (97.5%), 1828 (99.9%), and 1799 (98.3%), for standard RE, SRRE, MRRE-12, and MRRE-8, respectively. If only the "interesting" distances are considered, i.e. hydrogen pairs between residues i to i+3 or greater (1022 total pairs), the number of distances with differences less than 5% in standard RE, SRRE, MRRE-12, and MRRE-8 are 1016 (99.4%), 831 (81.3%), 1009 (98.7%) and 956 (93.5%) respectively. It is interesting to note that the methods that use higher temperature replicas tend to have greater differences for large hydrogen-hydrogen distances, and thus are likely populating a more diverse set of extended peptide configurations than what is seen from MRRE-8. Figure 3 shows an example of this for the $C\alpha$ proton distance between Ala21 and Ala30, where the MRRE approach shows tighter convergence of the long-ranged distance.

However, since Eq. (1) only reports on convergence within a method, it is also important to ask if the four methods are probing similar conformational ensembles. To answer this we next calculated all 1830 hydrogen-to-hydrogen pair distances averaged over two independent simulations for SRRE, MRRE-12, and MRRE-8, and then compared them to the same averaged quantity calculated from the standard RE method using Eq. (2) and shown in Figure 4. We found that the number of distances with a difference of less than 5% to the RE distance is 1747 (95.5%), 1787 (97.6%), and 1796 (98.1%) for SRRE, MRRE-12, and MRRE-8, respectively. The remaining 2-4% of distances measured by the SRRE and MRRE-12 are dissimilar with respect to RE at the same distances in which they are most dissimilar between their independent simulations. However, Figure 2b shows that the MRRE-8 method is most similar to the RE conformational ensemble overall. In

summary, the SRRE reservoir is more poorly converged with respect to the more expanded and independent reservoirs of MRRE-12 and MRRE-8, and the higher temperature replicas used in the RE approach- which we attribute to the use of a single small reservoir.

We next compare the replica exchange methods and their generated structural ensembles by calculating the $^{3}J_{HNH\alpha}$ scalar coupling constants and ROESY cross-peaks at 284K[31], and comparing them to an external benchmark of the same experimentally determined NMR data. Table 2 shows the calculated $^{3}J_{HNH\alpha}$ scalar coupling constants for all the simulation methods. The three reservoir methods (SRRE, MRRE-12 and MRRE-8) are seen to perform slightly better than traditional RE, as they are within error for all of the values measured from the COSY spectra, whereas standard RE is only within error for eight of the nine experimental J-coupling values.

As we found in our previous work, a more stringent assessment of simulation convergence is comparison of the peak intensities back-calculated from simulation to the experimentally determined ROESY cross-peaks[31]. Although the r^{1} (linear) pair distance averages vary less than 5-7% and suggest good convergence, this difference can translate into much larger variations in $\langle r_{ij}^{-6} \rangle$ and noticeable changes in the corresponding peak volumes, and thus peak ranking. Predicted ROESY cross peaks from simulation were calculated for a 300ms mixing time and compared to the medium- and long-range ROE cross peaks observed in the H_2O and D_2O experimental spectra.

Tables 3 and 4 show that the RE, SRRE, MRRE-12 and MRRE-8 simulations assign 9, 10, 10 and 11, respectively, of the 19 experimentally assigned cross peaks from the 900 MHz H_2O spectra, and 15, 15, 14 and 15, respectively, of the 21 assigned cross peaks in the 800 MHz D_2O experiment (note that there is redundancy of cross peaks between the two experiments so that there are only 28 distinct cross-peaks in total). There is a recovery of an additional 6, 4, 5, and 5 peaks that are just below the noise for RE, SRRE, MRRE-12, and MRRE-8, respectively. Thus all RE methods pick out the majority of the 28 experimentally observed medium-range ROE interactions from the 600 possible distinguishable medium and long-range interactions, with a small net advantage realized by the MRRE-8 simulation. While the reservoir methods show an increase in false positives, the SRRE method shows the largest number of false positives in both the H_2O and D_2O simulated spectra, which again shows that the implementation of a single small reservoir is degrading overall sampling.

Finally, we compare the computational expense of RE, SRRE, MRRE-12 and MRRE-8 in Table 5, in which each replica is computed on 16 processors and takes 4.2 hours to simulate 1.0ns of $A\beta_{21-30}$ peptide in explicit water. Using MRRE-8 cuts the cpu time of RE by 70%, and of SRRE by

38%. Additionally, the number of concurrently used processors is much fewer in MRRE, with a maximum of 128 processors for MRRE-8 versus 512 processors for RE and 320 processors for SRRE. Of course, trivial parallelization and excellent scalability are one of the primary advantages of any replica exchange method that exploits a large number of temperature replicas, and thus it seems counter-intuitive to tout the smaller number of processors used with MRRE, and its serial progression through the multiple reservoirs would seem to expand the wall time of the net calculation compared to standard RE. However, there are practicalities that must be recognized in regards how computational research groups utilize in-house computers or a shared external computational platform such as a supercomputer center. Within our own research group we have ~500 processors to divide among ~10 projects at any given time, and hence standard RE on any significant sampling problem can be a burden on these other projects. At supercomputer centers there can be long waits in the queues, and although we bundled multiple simulation runs into one job to exploit the largest number of processors, there are queue time limitations of anywhere between 24- or 48-hours per job in these larger queues. In reality, overall wall time actually increased for the standard RE and SRRE methods compared to our more practical MRRE approach.

CONCLUSION

The multi-reservoir replica exchange method provides both improved sampling efficiency and computational efficiency when compared to standard RE simulation or an SRRE implementation that uses a single, small reservoir. We found that overall internal measures of sampling efficiency based on Eqs. (1) and (2) were most poor for the SRRE method, and while all reservoir methods showed marginally better agreement than standard RE with the experimental $^{3}J_{HNH\alpha}$ scalar coupling constants, SRRE performs worse in regards to a smaller number of identified ROEs and far greater numbers of false positive peaks in predicting the experimental ROESY spectra. We believe the less optimal performance of the SRRE method is due to the use of a single small reservoir for the two independent trajectories, where we found that the number of exchange attempts far exceeded the reservoir size, and many of the same structures were therefore drawn down to the target temperature simulation of the two trajectories. Thus in any reservoir approach, as demonstrated by MRRE-12 and MRRE-8 implementations, two independent reservoirs whose sizes are greater than the number of exchange attempts is absolutely necessary.

It appears that MRRE-8 performs somewhat better than MRRE-12. This is likely due to the fact that 320K reservoir has a more optimal overlap with the distribution at the target temperature,

so that the MRRE-8 simulation is the best compromise between basin hopping while measuring the relative importance of basins. This is however a drawback of the MRRE approach since the optimal reservoir temperature is not clear *a priori* for any system, but can only be estimated after some experience. Nonetheless, while the 344K reservoir is still too hot to be optimal in this particular case, MRRE-12 is still a clear improvement with respect to SRRE and aspects of the RE method.

We emphasize that the MRRE approach would be a poor choice of an enhanced sampling method when broken ergodicity is a genuine problem. It should only be applied to systems that are not inherently glassy, but are simply hard to sample because they possess a large number of degrees of freedom. As such, the MRRE approach is certainly appropriate in characterizing the structural ensembles for most aqueous peptides as shown here, and for the self-assembly process in transitioning between unfolded and folded ensembles. In protein folding the conformational sampling within basins is as important as basin hopping[34], and the MRRE approach would be particularly suitable since the lower temperature reservoir better senses the relative importance of basins but is still hot enough to basin hop. Finally, any reservoir approach can handle variable temperature spacing in order to concentrate replicas near the mean of the folding and unfolding temperatures, if that is required[34], and the temperature range explored is completely adjustable depending on the problem at hand. In summary, we have developed and validated a practical approach to replica exchange that systematically converges a more complete equilibrated reservoir at an optimal high temperature to enhance biomolecular comformational sampling.

ACKNOWLEDGEMENTS. We thank the NSF Cyberinfrastructure program for support of the work presented here. We acknowledge the National Energy Research Scientific Computing Center (supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098) for the needed computational resources.

REFERENCES

1. Hukushima, K.; Nemoto, K. *J Phys Soc Jpn* 1996, 65, 1604-1608.
2. Hansmann, U. H. E. *Chem. Phys. Letters* 1997, 281(1-3), 140-150.
3. Sugita, Y.; Okamoto, Y. *Chemical Physics Letters* 1999, 314(1-2), 141-151.
4. Wang, J.-S.; Swendsen, R. H. *Phys. Rev. B* 1988, 37(13), 7745.
5. Marinari, E.; Parisi, G. *EPL (Europhysics Letters)* 1992, 19(6), 451-458.
6. Rathore, N.; Chopra, M.; de Pablo, J. *J. Chem. Phys.* 2005, 122(2), -.
7. Nadler, W.; Hansmann, U. H. E. *Phys. Rev. E* 2007, 76(6), -.
8. Nymeyer, H. *J. Chem. Theory Comp.* 2008, 4(4), 626-636.

9. Zheng, W. H.; Andrec, M.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. USA* 2007, 104(39), 15340-15345.
10. Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* 2002, 116(20), 9058-9067.
11. Rick, S. W. *J. Chem. Phys.* 2007, 126(5), 054102-054108.
12. Sindhikara, D.; Meng, Y. L.; Roitberg, A. E. *J. Chem. Phys.* 2008, 128(2), 024103.
13. Hagen, M.; Kim, B.; Liu, P.; Friesner, R. A.; Berne, B *J. Phys. Chem. B* 2007, 111(6), 1416-1423.
14. Gallicchio, E.; Levy, R. M.; Parashar, M. *J. Comp. Chem.* 2008, 29(5), 788-794.
15. Andricioaei, I.; Straub, J. E.; Voter, A. F. *J. Chem. Phys.* 2001, 114(16), 6994-7000.
16. Brown, S.; Head-Gordon, T. *J. Comp. Chem.* 2003, 24(1), 68-76.
17. Frantz, D. D.; Freeman, D. L.; Doll, J. D. *J. Chem. Phys.* 1990, 93(4), 2769-2784.
18. Zhou, R.; Berne, B. J. *J. Chem. Phys.* 1997, 107(21), 9185-9196.
19. Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. USA* 2005, 102(39), 13749-13754.
20. Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comp.* 2006, 2(2), 420-433.
21. Periole, X.; Mark, A. E. *J. Chem. Phys.* 2007, 126(1), 014903-014911.
22. Zuckerman, D. M.; Lyman, E. *J. Chem. Theory Comp.* 2006, 2(4), 12001202-12001202.
23. Opps, S. B.; Schofield, J. *Phys. Rev. E* 2001, 63(5), 056701.
24. Ballard, A. J.; Jarzynski, C. Manuscript submitted for publication 2009.
25. Okur, A.; Roe, D. R.; Cui, G.; Hornak, V.; Simmerling, C *J. Chem. Theory Comput.* 2007, 3(2), 557-568.
26. Roitberg, A. E.; Okur, A.; Simmerling, C. *J. Phys. Chem. B* 2007, 111(10), 2415-2418.
27. Huang, X.; Bowman, G. R.; Pande, V. S. *J. Chem. Phys.* 2008, 128(20), 205106-205115.
28. D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, Ross C. Walker,W. Zhang, K.M. Merz, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F.Wong, F. Paesani, J. Vanicek, X.Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, and P.A. Kollman. University of California, San Francisco 2008.
29. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* 2006, 65(3), 712-725.
30. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* 2004, 120(20), 9665-9678.
31. Fawzi, N. L.; Phillips, A. H.; Ruscio, J. Z.; Doucleff, M.; Wemmer, D. E.; Head-Gordon, T. *J. Am. Chem. Soc.* 2008, 130(19), 6145-6158.
32. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* 1993, 98(12), 10089-10092.
33. Thirumalai, D.; Mountain, R. D.; Kirkpatrick, T. R. *Phys. Rev. A* 1989, 39, 3563-3574.
34. Zheng, W.; Andrec, M.; Gallicchio, E.; Levy, R. M. *J. Phys. Chem. B* 2008, 112, 6083-6093.

Table 1. Comparison of reservoir sampling characteristics between SRRE, MRRE-12, and MRRE-8. The use of one small reservoir, as in SRRE, may artificially give better convergence between simulations.

Reservoir Type	Number of Structures in Reservoir	Exchange Attempts	Reservoir Structures Attempted > 1	Structures exchanged > 1 per # successful exchanges
SRRE	12,225	38,000	9967	1914/7909 (24%)
MRRE-12	20,000 20,000	16,000 19,000	3838 4895	207/2171 (10%) 325/2564 (13%)
MRRE-8	20,000 20,000	16,176 18,000	3,483 4,008	219/2448 (9%) 246/2696 (9%)

Table 2. $^{3}J_{H}N_{H}\alpha$ scalar coupling constants calculated from the four different simulations. The mean and standard deviation for each residue is shown. Shaded boxes highlight the simulation values that are outside error of the experimental values.

NMR	RE	SRRE	MRRE-12	MRRE-8
6.0±0.5 (Glu)	5.90±0.4	5.97±0.7	5.75±0.8	5.67±0.6
5.5±0.5 (Asp)	6.62±0.5	6.68±0.8	6.71±0.9	6.68±0.8
6.75±0.5 (Val)	5.83±0.5	6.49±0.7	5.85±0.8	6.12±0.7
10.0±0.5 (Gly)	8.92±0.9	9.07±1.3	9.0±1.4	9.18±1.1
5.75±0.5 (Ser)	5.93±0.6	6.48±0.9	6.1±0.9	6.16±0.7
6.4±0.5 (Asn)	6.69±0.5	6.8±0.9	6.5±0.9	6.66±0.8
6.0±0.5 (Lys)	5.51±0.4	5.43±0.8	5.6±0.7	5.52±0.7
9.75±0.5 (Gly)	9.16±0.8	8.96±1.1	9.01±1.3	9.09±1.2
5.5±0.5 (Ala)	5.27±0.4	5.31±0.5	5.27±0.6	5.14±0.5

Table 3. Proton ROESY crosspeak predictions for the various replica exchange experiment compared to the 900MHz spectra in H_2O. The 19 experimental intensities (Iexp) are normalized to the intensity of the weakest assigned peak, and labeled as "<1.0" if some evidence of a peak is present but is too weak to be assigned. Simulation intensities (Isim) are normalized to experimental intensity as described in [31], and are labeled as a dash if the simulation peak is absent.

I_RE	I_SRRE	I_MRRE-12	I_MRRE-8	I_EXP	Proton1	Proton2
3.3	4.9	2.9	2.6	15.7	HBALA21	HGVAL24
1.3	1.1	1.7	1.2	7.2	HG2GLU22	HGVAL24
2.0	1.3	1.9	2.0	4.9	HG3GLU22	HGVAL24
---	---	---	---	3.8	HB3ASN27	HBALA30
0.5	0.4	0.4	0.3	3.8	HAGLY25	HD2LYS28
0.4	0.4	0.2	0.9	3.3	HB2ASN27	HBALA30
2.1	2.1	2.1	2.0	2.9	HBALA21	HASP23
0.9	0.4	1.0	0.5	2.8	HGVAL24	HB3ASN27
2.7	1.6	2.1	2.1	2.6	HGVAL24	HSER26
1.0	1.6	1.1	1.5	2.6	HAVAL24	HSER26
0.6	0.8	0.6	0.6	2.6	HGVAL24	HB2ASN27
0.2	0.1	0.2	0.3	2.5	HB3LYS28	HALA30
0.4	0.4	0.4	0.6	1.7	HALYS28	HALA30
2.1	2.0	1.9	2.0	1.3	HB3ASP23	HGLY25
1.3	1.1	1.3	1.1	1.2	HAGLY25	HASN27
0.9	0.8	0.9	1.0	1.1	HB2ASN27	HGLY29
0.1	0.3	0.3	0.2	1.1	HASER26	HLYS28
2.5	1.1	1.3	2.0	<1.0	HG3GLU22	HVAL24
0.9	1.4	0.5	1.3	<1.0	HAVAL24	HB3ASN27
9	**10**	**10**	**11**	**# crosspeaks ≥1.0**		
14	**12**	**13**	**15**	**# crosspeaks ≥0.5**		
5	**12**	**9**	**9**	**# false positives ≥1.0 (data not shown)**		

Table 4. Proton ROESY crosspeak predictions for the various replica exchange methods compared to the experimental 800MHz spectra in D_2O. There are 21 experimental intensities; see Table 3 for details.

I_RE	I_SRRE	I_MRRE-12	I_MRRE-8	I_EXP	Proton1	Proton2
7.2	10.4	6.3	5.5	7.4	HBALA21	HGVAL24
4.7	2.9	4.3	4.5	5.1	HG3GLU22	HGVAL24
2.9	2.3	3.9	2.7	4.3	HG2GLU22	HGVAL24
---	---	---	---	3.9	HB3ASN27	HBALA30
3.6	4.5	3.9	4.3	3.8	HB3GLU22	HGVAL24
1.6	3.0	1.2	1.7	3.1	HGVAL24	HE2LYS28
1.0	1.0	0.5	2.0	2.5	HB2ASN27	HBALA30
3.7	4.8	3.4	5.1	2.4	HB2GLU22	HGVAL24
---	0.5	0.4	---	2.2	HB3ASP23	HB3SER26
2.5	5.4	2.1	4.9	1.8	HB3ASP23	HB2SER26
1.3	1.0	0.8	0.8	1.8	HAGLY35	HD2LYS28
2.2	1.0	2.3	1.1	1.8	HGVAL24	HB3ASN27
---	---	---	---	1.7	HB2ASP23	HB2SER26
1.6	1.9	1.6	1.5	1.6	HGVAL24	HB2ASN27
2.7	8.5	4.3	5.1	1.4	HAVAL24	HB2ASN27
1.7	0.7	1.5	1.2	1.3	HGVAL24	HB2SER26
0.4	0.3	0.2	0.4	1.0	HB3SER26	HG3LYS28
0.9	2.4	1.2	1.0	0.9	HGVAL24	HALYS28
---	---	---	---	0.8	HB2ASP23	HB3SER26
2.1	2.7	1.9	1.7	0.8	HAGLU22	HGVAL24
2.2	3.2	1.3	3.2	<1.0	HAVAL24	HB3ASN27
15	**15**	**14**	**15**	**# crosspeaks ≥1.0**		
16	**17**	**16**	**16**	**# crosspeaks ≥0.5**		
22	**35**	**26**	**30**	**# false positives ≥1.0 (data not shown)**		

Table 5. Comparison of wall clock timings between standard replica exchange and the reservoir methods: SRRE, MRRE-12 and MRRE-8.

Simulation Type	# Replicas	Simulation Time (ns)	# Processors	Total Processor Time (hrs)
RE	32	50	512	**107,530**
SRRE	20	50	320	**67,200**
MRRE-12	4	20	64	5,376
	4	20	64	5,376
	12	38	192	30,643
				41,395
MRRE-8	4	20	64	5,376
	4	20	64	5,376
	4	20	64	5,376
	8	38	128	20,429
				36,557

FIGURE CAPTIONS

Figure 1. *Schematic representation of MRRE.* The reservoirs are depicted by a dashed red line and the individual replicas are depicted by black solid lines. Each subreplica is run until the lowest temperature replicas (solid red line) between two independent simulations converge.

Figure 2. *Difference between all hydrogen-hydrogen averaged pair distances at 284K from two independent simulations based on Eq. (1).* For the RE (top, left), SRRE (top, right), MRRE-12 (bottom, left), and MRRE-8 (bottom, right) simulations.

Figure 3. *Convergence profile of the time averaged distance between Cα protons between Ala21 and Ala30 between two independent simulations.* For the RE (top, left), SRRE (top, right), MRRE-12 (bottom, left), and MRRE-8 (bottom, right) simulations.

Figure 4. *Difference between all hydrogen-hydrogen averaged pair distances at 284K between a given reservoir method and the standard RE method based on Eq. (2).* For the SRRE (left), MRRE-12 (middle), and MRRE-8 (right) simulations.

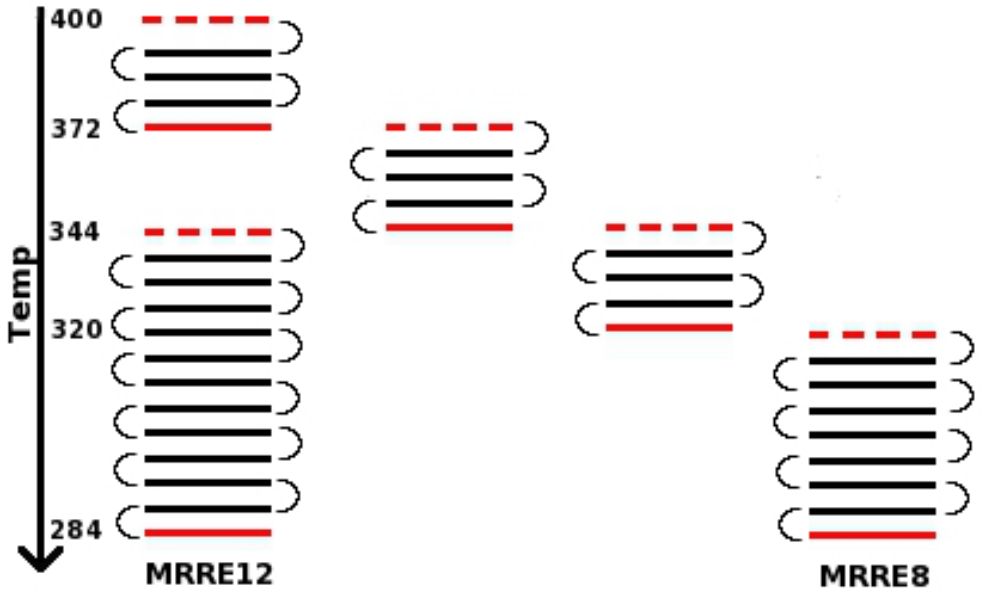

Figure 1. Ruscio, Fawzi and HeadGordon

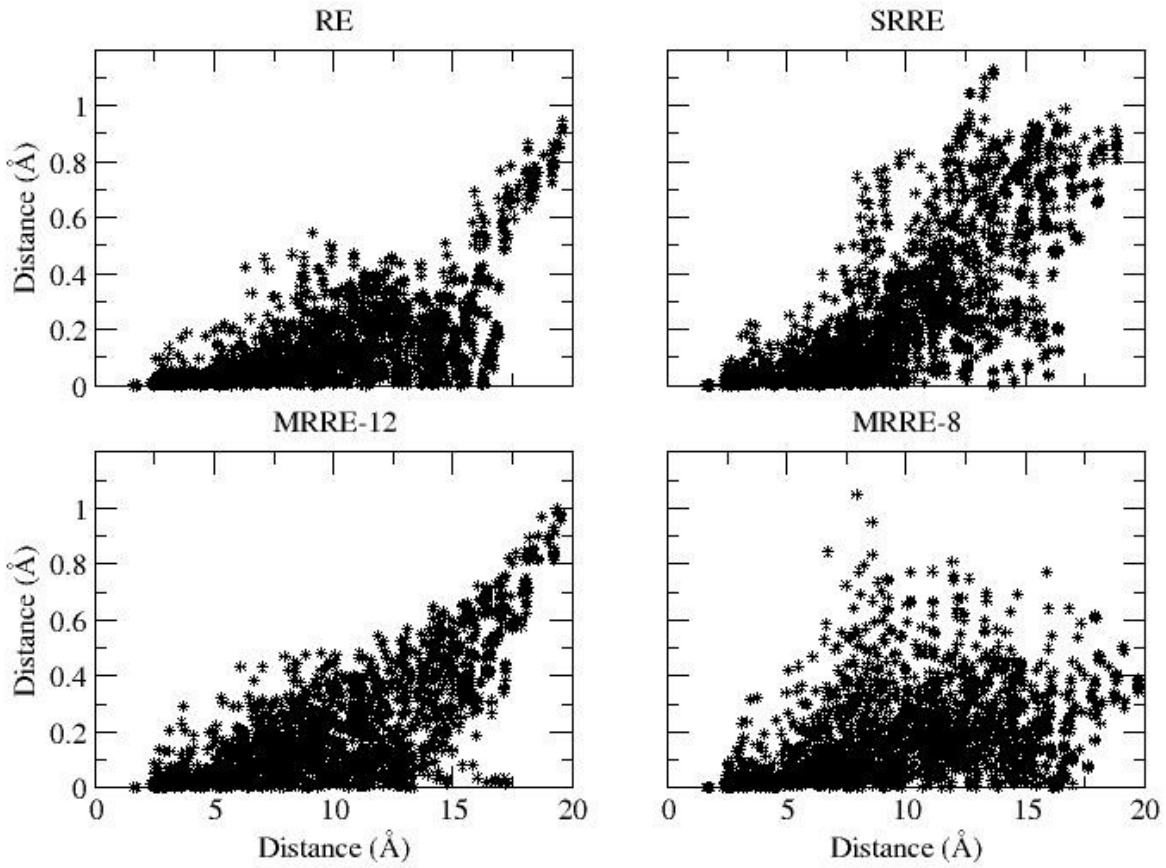

Figure 2. Ruscio, Fawzi and Head-Gordon

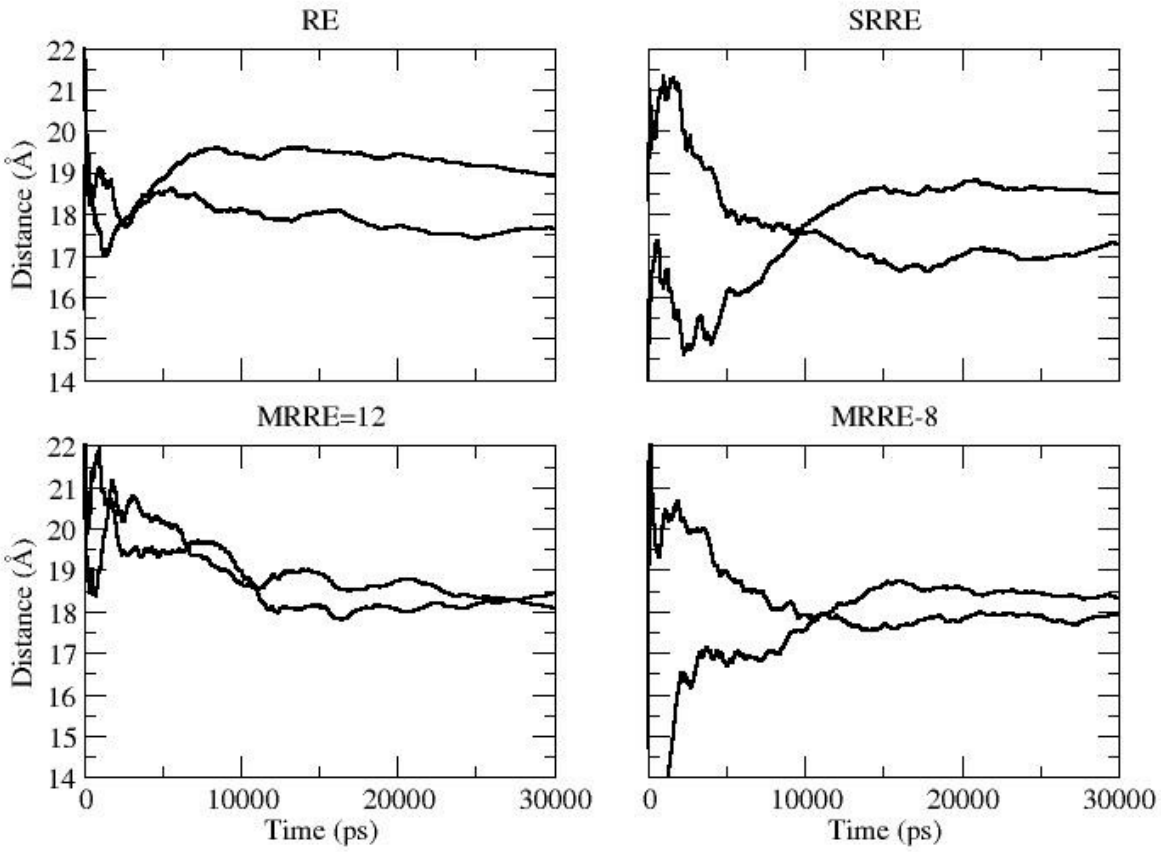

Figure 3. Ruscio, Fawzi and Head-Gordon

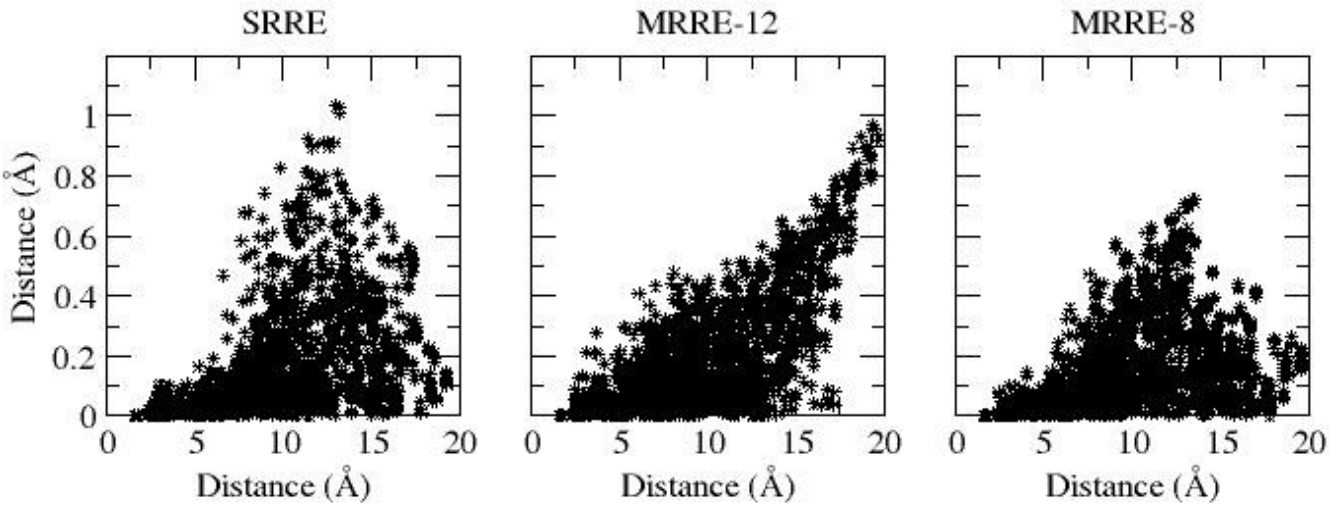

Figure 4. Ruscio, Fawzi and Head-Gordon