

Accelerating the Replica Exchange Method Through an Efficient All-Pairs Exchange

Paul Brenner, Christopher R. Sweet, Dustin VonHandorf, Jesús A. Izaguirre*

University of Notre Dame, Notre Dame, Indiana 46556. USA

(Dated: December 29, 2006)

Abstract

We accelerate the replica exchange method through an efficient all-pairs replica exchange. A proof of detailed balance is shown along with an analytical estimate of the enhanced exchange efficiency. The new method provides asymptotically 4-fold speedup of conformation traversal for replica counts of eight and larger with typical exchange rates. Experimental tests using the blocked alanine dipeptide demonstrate the method's correctness and show an approximate sampling efficiency improvement of 100% according to potential energy cumulative averages and an ergodic measure. An explicitly solvated PIN1 WW domain system of 4958 atoms is sampled using our new method, yielding a cluster sampling rate almost twice that of the single exchange near neighbor implementation. Computational software and scripts along with input and output data sets are available at <http://www.nd.edu/~lcls/APEREM>.

I. INTRODUCTION

Sampling conformations of biomolecules through computational techniques is a key step toward understanding molecular function¹⁻⁵. Computational methods based on molecular dynamics (MD) can now routinely identify energetically predominant protein structures evident over nanosecond (ns) to microsecond (μ s) trajectories. Unfortunately representative sets of structural conformations for large biomolecules are not sampled completely by these trajectories due to the complexity of the configuration space and energetic barriers that localize sampling³. Despite numerous advances in non-bonded force computation and extended time steps⁵, routine MD trajectories for large solvated proteins on μ s and millisecond (ms) timescales are computationally inaccessible. Elevated temperature simulations provide a higher probability to surpass the energy barriers, however the resulting energy surface is not necessarily representative of the biologically relevant lower temperature regime.

The replica exchange method (REM)⁶, also known as parallel tempering^{7,8}, has become a popular method for using the configurational sampling of elevated temperature simulations with simulations run at a lower target temperature. REM uses multiple independent MD (or Monte Carlo) simulations running in parallel, each at a different temperature. Each simulation, also known as a replica, runs independently for set durations between which a transfer of positions and velocities is attempted with another replica, typically at neighboring temperatures. A random walk is performed on temperature, or another system variable, which in turn, induces a walk on configuration space. Replica coordinates are swapped according to a probability function, with an intent to enable sampling at low temperature to overcome barriers between local energy minima. The sampling efficiency has been shown greater than that of standard canonical techniques⁹⁻¹¹ and the method has found favor because the weighting factors are known a priori.

Recent novel methods have been published to efficiently utilize heterogeneous grid resources to simulate with large (>100) temperature sets¹², or multiplex additional MD simulations at each temperature in order to more effectively explore the temperature space¹³. As the number of replicas grows, the effective contribution of each replica with respect to the low target temperature is limited by the random walk, constrained within a finite simulation time.

We introduce a new method to accelerate the random walk with negligible computational

overhead. This is done through a novel implementation of the replica switch method which maintains detailed balance. In the next section we present the foundational mathematical basis and a detailed balance proof for our method. We follow the derivations with an analytical study of the method’s efficiency, and experimental results demonstrating sampling speed and accuracy.

II. DETAILED BALANCE FOR SINGLE SWITCH METHODS

A. Definitions

We present our new method in reference to the established and current methods using the following notation: A replica is denoted as microstate \mathbf{x}_i at a temperature T_l , where $1 \leq i, l \leq K$ and K represents the number of replicas, equal to the number of temperatures. We define a macro state A , comprised of K replicas \mathbf{x}_i which are ordered by temperature, as:

$$A = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_j, \dots, \mathbf{x}_{K-1}, \mathbf{x}_K\}. \quad (1)$$

There are N macro states which can be reached from A by the exchange of two replicas, where N is dependent on the exchange scheme, and we denote this set of macro states as Φ . Without loss of generality we define macro state $B \in \Phi$ as:

$$B = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_j, \dots, \mathbf{x}_i, \dots, \mathbf{x}_{K-1}, \mathbf{x}_K\}, \quad (2)$$

where B is reached from A by the exchange of \mathbf{x}_i and \mathbf{x}_j for some $1 \leq i, j \leq K$, $i \neq j$, where i and j may have other method dependent constraints such as being nearest neighbors. By similar reasoning there is a set of macro states Ψ which can be reached from B by the exchange of two replicas, dependent on the exchange scheme, and furthermore $A \in \Psi$.

The probability of generating a possible exchange between macro states A and B is denoted $P_{\text{gen}}(A \rightarrow B)$ and the probability of accepting the exchange is $P_{\text{acc}}(A \rightarrow B)$. The probability of the system being in macro state A is denoted $P(A)$.

B. Near Neighbor Pairs Method

Early implementations of REM, such as Ref. [6], are based on the random selection of a pair of near-neighbor replicas (closest neighbors in temperature space) for which an exchange

is made according to a Metropolis acceptance criterion meeting detailed balance. Here the number of macro states which can be reached from a given state is given by $K - 1$. We consider the detailed balance equation for arbitrary states A and B

$$P(A)P_{\text{gen}}(A \rightarrow B)P_{\text{acc}}(A \rightarrow B) = P(B)P_{\text{gen}}(B \rightarrow A)P_{\text{acc}}(B \rightarrow A). \quad (3)$$

Since the exchange is selected randomly and there are $K - 1$ possible exchanges, we have

$$P_{\text{gen}}(A \rightarrow B) = P_{\text{gen}}(B \rightarrow A) = \frac{1}{(K - 1)}, \quad (4)$$

and

$$\frac{P_{\text{acc}}(A \rightarrow B)}{P_{\text{acc}}(B \rightarrow A)} = \frac{P(B)}{P(A)}. \quad (5)$$

For the canonical ensemble where $\rho(\mathbf{x}_i, \mathbf{p}_i) \propto \exp(-\beta H(\mathbf{x}_i, \mathbf{p}_i))$ for Hamiltonian H and momenta \mathbf{p}_i , this is satisfied by the following Metropolis criterion,

$$\begin{aligned} P_{\text{acc}}(A \rightarrow B) &= \min \{1, \exp(-(\beta_j - \beta_i)(U(\mathbf{x}_i) - U(\mathbf{x}_j)))\}, \\ P_{\text{acc}}(B \rightarrow A) &= \min \{1, \exp(-(\beta_j - \beta_i)(U(\mathbf{x}_j) - U(\mathbf{x}_i)))\}. \end{aligned} \quad (6)$$

where the potential energy U of the configuration is used in place of the Hamiltonian energy as a rescaling of temperatures following exchange allows the momentum term to cancel⁹, and the $\beta_m = 1/kT_m$ are the inverse temperatures with k the Boltzmann constant. We note that either $P_{\text{acc}}(A \rightarrow B) = 1$ or $P_{\text{acc}}(B \rightarrow A) = 1$ with this scheme.

This implementation has been widely accepted and is in current use. However the method's sampling efficiency scales poorly with the number of replicas.

C. Calvo's All-Pairs Exchange Method

Recent work demonstrating improved sampling efficiency over near-neighbor exchange REM has been reported by Calvo¹⁴. He suggests an alternative method which includes all replica pairs and modifies the probability of generation to enhance exchange, here the number of possible exchanges is $\frac{K^2}{2} - \frac{K}{2}$. We show that this method does not satisfy detailed balance.

Calvo seeks to improve the efficiency of the near-neighbor method by considering the probabilities of all exchanges simultaneously. Clearly, if the sum of all exchange probabilities was less than 1 it would be possible to use a kinetic MC method¹⁵ to select exchanges with the correct probability (Algorithm 1).

Algorithm 1. *Kinetic MC Updating*

Given a set of n events with probabilities $\{P_i\}$ and $\sum_i P_i \leq 1$

1. A random number $0 \leq R \leq 1$ is drawn and
2. Event j is selected if $R \in [PT_j, PT_j + P_j)$ where $PT_j = \sum_{i < j} P_i$.

This condition is not met in general, Calvo uses

$$P_{\text{gen}}(A \rightarrow B) = \left(\sum_{M \in \Phi} P_{\text{acc}}(A \rightarrow M) + 1 \right)^{-1}, \quad (7)$$

and

$$P_{\text{gen}}(B \rightarrow A) = \left(\sum_{L \in \Psi} P_{\text{acc}}(B \rightarrow L) + 1 \right)^{-1}. \quad (8)$$

Then Algorithm 1 is used to select the exchanges with the correct probabilities.

Note that $\Psi \neq \Phi$, i.e., the set of macro states accessible from A is different to the set of macro states accessible from B in one exchange, since there are $K!$ possible macro states but only $\frac{K^2}{2} - \frac{K}{2}$ can be reached from a given macro state. Then we have

$$P_{\text{gen}}(A \rightarrow B) \neq P_{\text{gen}}(B \rightarrow A), \quad (9)$$

and as a consequence (3) is not satisfied; i.e. detailed balance is not met.

D. A Balanced and Efficient All Pairs Exchange Method

We present a revised method, similar to Calvo's, which meets detailed balance while using the extended Metropolis criterion (6) for the probability of acceptance. We reference our all pairs exchange method with the acronym APE. Again the number of possible exchanges is given by $\frac{K^2}{2} - \frac{K}{2}$. The revised probabilities of generation are:

$$P_{\text{gen}}(A \rightarrow B) = \max \left\{ \sum_{M \in \Phi} P_{\text{acc}}(A \rightarrow M), \sum_{L \in \Psi} P_{\text{acc}}(B \rightarrow L) \right\}^{-1}, \quad (10)$$

and

$$P_{\text{gen}}(B \rightarrow A) = \max \left\{ \sum_{L \in \Psi} P_{\text{acc}}(B \rightarrow L), \sum_{M \in \Phi} P_{\text{acc}}(A \rightarrow M) \right\}^{-1}. \quad (11)$$

We see that $P_{\text{gen}}(A \rightarrow B) = P_{\text{gen}}(B \rightarrow A)$ and the detailed balance equation reduces to

$$\frac{P_{\text{acc}}(A \rightarrow B)}{P_{\text{acc}}(B \rightarrow A)} = \frac{P(B)}{P(A)}, \quad (12)$$

which has solution for the extended Metropolis criterion (6). Algorithm 1 is then used to select the exchanges with the correct probabilities meeting $\sum_i P_i \leq 1$ as the P_{acc} values are properly normalized P_{gen} given in (10) or (11). Note that because P_{gen} is a maximum, there will always be a probability for ‘no transition’ in the direction of the less favorable Boltzmann exchange.

We note that if the system is in a macro state A then we can calculate the probabilities for exchange between any macro state $M \in \Phi$ and the respective set of macro states that can be reached from M . This leads to explicit solutions for (10)-(11). From the property of the extended Metropolis criterion (6) we have that one of the $P_{\text{acc}}(A \rightarrow M)$ or $P_{\text{acc}}(B \rightarrow L)$ must be equal to 1 from our observation in Section II B, which guarantees $P_{\text{gen}} \leq 1$ as required.

We review the degree of improvement in exchange, conformational transfer, and sampling metrics in the following section.

III. ANALYSIS OF EXCHANGE EFFICIENCY

The motivation of replica exchange is to move conformations from a high temperature regime, where the probability of overcoming large potential energy barriers is greater, into a lower temperature regime. Our method accelerates this movement through an overall increase in the exchange rate resulting from the larger P_{gen} value in our method.

The improvement in efficiency of APE over the near-neighbor method can be characterized by the ratio of the probabilities of generation for some average acceptance probability $\langle P_{\text{acc}} \rangle$. Given that these probabilities of generation are \tilde{P}_{gen} and \hat{P}_{gen} for APE and nearest-neighbor respectively we show that the relative efficiency E_{APE} is given by

Proposition 1. *Assume that the probabilities of some macrostate A exchanging with all*

the macro states in Φ can be approximated by ignoring the exchanges which are not nearest-neighbor. Then

$$E_{\text{APE}} = \frac{\tilde{P}_{\text{gen}}}{\hat{P}_{\text{gen}}} \approx \frac{(K-1)}{1 + (K-2)\langle P_{\text{acc}} \rangle} \text{ and } \lim_{K \rightarrow \infty} E_{\text{APE}} \approx \frac{1}{\langle P_{\text{acc}} \rangle}. \quad (13)$$

Proof. Using our assumption,

$$\sum_{M \in \Phi} P_{\text{acc}}(A \rightarrow M) \approx P_{\text{acc}}(A \rightarrow B) + (K-2)\langle P_{\text{acc}} \rangle, \quad (14)$$

and

$$\sum_{L \in \Psi} P_{\text{acc}}(B \rightarrow L) \approx P_{\text{acc}}(B \rightarrow A) + (K-2)\langle P_{\text{acc}} \rangle. \quad (15)$$

From Section II B we have that $P_{\text{acc}}(A \rightarrow B) = 1$ or $P_{\text{acc}}(B \rightarrow A) = 1$ and hence, from (10),

$$\tilde{P}_{\text{gen}} \approx (1 + (K-2)\langle P_{\text{acc}} \rangle)^{-1}. \quad (16)$$

We can calculate E_{APE} using (16) and (4) to yield (13) as required. \square

The higher efficiency allows more rapid traversal over temperature space. The average number of exchange events $\langle S \rangle$ required for conformational transfer from the highest temperature replica to the lowest is given by

$$\langle S \rangle = \sum_{S=(K-1)}^{\infty} S P_S. \quad (17)$$

Where P_S represents the probability for conformational transfer in S exchange events. As an example, the traversal probabilities for four replica near neighbor exchange can be written as follows

$$\begin{aligned} P_3 &= (P_{\text{gen}})^3 \langle P_{\text{acc}} \rangle^3, \\ P_4 &= (P_{\text{gen}})^4 (3 \langle P_{\text{acc}} \rangle^3 (1 - \langle P_{\text{acc}} \rangle)), \\ P_5 &= (P_{\text{gen}})^5 (6 \langle P_{\text{acc}} \rangle^3 (1 - \langle P_{\text{acc}} \rangle)^2 + 2 \langle P_{\text{acc}} \rangle^5), \\ &\vdots \end{aligned} \quad (18)$$

The terms in (18) are based on the multiple of exchange probabilities required to transfer a conformation in S exchange events. For example the terms in P_3 account for three accepted exchanges in order between replicas (4,3), (3,2), and (2,1). The terms in P_4 include the three accepted exchanges along with a rejected exchange and accounts for the three possible orderings of events. We implemented a numerical model on this basis and calculated the average number of REM switch attempts required to traverse a conformation from the highest temperature ensemble to the lowest. Figure 1 demonstrates the average ‘traversal efficiency’ of the single switch methods with respect to the number of replicas. This model system utilized normal distributions for each replica spaced such that the average P_{acc} between near neighbors was 23% as advocated by Kone and Kofke¹⁶. Our APE method scales much more efficiently than the standard near neighbor method, traversing 3 times faster with 6 replicas and almost 5 times faster with 8 replicas.

IV. EXPERIMENTAL RESULTS

The REM algorithms were implemented and tested in the open source molecular simulation package PROTOMOL^{17,18}. The exchange logic occurs internally via MPI removing the requirement for an external application to manage the file I/O and reducing overhead associated with individual application startup/shutdown between switch cycles. For simplicity, we select a replica temperature distribution that maintains a uniform exchange rate near 23%¹⁶, even though alternative temperature distributions exist, cf.¹⁹⁻²⁴. Our method to generate a temperature set which closely maintains the uniform exchange rate is shown in the Appendix A. All trajectories were obtained by canonical MD using the Langevin Impulse (LI) integrator²⁵ with a 1 fs timestep. The total simulation time, number of switch attempts, temperature range, and observed P_{acc} values are reported for each test.

A. Convergence to Canonical Averages

We selected the well documented blocked alanine dipeptide molecular model for validation of sampling correctness and speed²⁶⁻²⁸. In addition to the cumulative average potential energy, and associated ergodic measure^{29,30}, Ramachandran plots were generated to validate the degree of conformational sampling. Ten simulations for each method, Single Near Neigh-

bor (SNN) and our All Pairs Exchange (APE), were performed. Five of the simulations were started with a molecule in the C5 axial conformation while the remaining 5 started in the C7 axial conformation. This variation in the initial configurations allows for a more accurate ergodic measure. The simulations utilized 6 replicas with the temperature distribution (300,410,539,707,935,1245K). The damping coefficient γ for LI was 1000 ps^{-1} . Each replica was run for 20 ns with an exchange attempted every 10 ps. Observed P_{acc} values between near neighbor replicas were (22, 22.5, 22.5, 22, 21.5; Avg = 22.1; Std = 0.4) for SNN and (23.5, 23, 23.5, 21.5, 22; Avg = 22.7 Std = 0.9) for APE.

Cumulative potential energy averages are shown in Fig. 2 (a) and the corresponding ergodic measure is shown in Fig. 2 (b). The SNN and APE cumulative potential energy averages converge providing an experimental validation of correctness. APE most quickly approaches the stable potential energy average and smallest ergodic measure. For reference APE reaches a target cumulative PE value of 0.7kcal/mol in less than half the simulation time (1500 vs 3250 ps) of the standard SNN REM implementation. In Figure 3 the SNN and APE Ramachandran plots are shown for the simulation segment up to timestep 3250 ps. In comparison with the SNN Ramachandran at 20 ns, SNN over-samples the C7 axial while APE more closely aligns with the 20 ns results and expectations based on reports in^{26-28,31,32}.

B. Efficient Sampling of Conformation Space

A conformational geometry based clustering measure was computed for the WW domain of a PIN1 enzyme (pdb 1I6C)³³ a 551 atom (33 Residue) biomolecule explicitly solvated with 1469 TIP3P water molecules. The simulations utilized 10 replicas with the temperature distribution (278, 284, 289, 294, 299, 305, 311, 317, 323, 330K). The damping coefficient γ for LI was 100 ps^{-1} . Each replica was run for 1 ns with an exchange attempted every 1 ps. Observed P_{acc} values between near neighbor replicas were (13, 20, 25, 24, 19, 18, 22, 20, 16; Avg = 19.7 Std = 3.8) for SNN and (14, 23, 23, 27, 16, 18, 22, 22, 16; Avg = 20.1; Std = 4.3) for APE. We note that the actual average P_{acc} was closer to 20% than the target 23%.

The degree of geometric differentiation between trajectory snapshots, one DCD frame taken every 100 fs in this case, can be utilized to measure the degree of sampling. We use a clustering method based on that reported by Daura *et al.*^{34,35}. RMSD values were calculated based on all protein atoms (minus hydrogens) following a best fit alignment of the protein

backbone. Clustering was performed with a 1 Å neighbor cutoff. In Fig. 4 the cluster count is calculated every 200 ps over a 1 ns simulation. This quantitative measure demonstrates that the APE method explores conformation space almost 2 times faster than SNN, as the average ratio of APE to SNN cluster counts is 1.97 for 200 ps increments plotted.

V. CONCLUSIONS

We have introduced APE, an REM algorithm which considers all replica pairs and implements an efficient exchange criterion shown to meet detailed balance. The enhanced efficiency of the method comes from the novel designation of the P_{gen} term. Our APE method shows 4-fold asymptotic improvement in traversal efficiency over standard single switch implementations for replica counts of 8 and larger, using typical exchange rates of about 23%. A two fold sampling speedup and experimental correctness validation was shown for an unsolvated alanine dipeptide using 6 replicas. A nearly two fold sampling speedup was shown with cluster counts for a solvated 33 residue WW domain using 10 replicas.

The improved efficiency enabled by APE is complemented by the following three characteristics: the computation time for calculation of all pairs exchange is negligible compared to the molecular simulation segments, the implementation is system independent, and no new parameters are introduced. Computational software and scripts along with input and output data sets are available at <http://www.nd.edu/~lcls/APEREM>.

Acknowledgments

This work was partially supported by NSF grants DBI-0450067 and CCF-0135195 to JAI. The authors are thankful for valuable insights obtained through conversations with Florent Calvo at Paul Sabatier University, France and Radford Neal at the University of Toronto, Canada. Computational resources were made available by the Interdisciplinary Center for the Study of Biocomplexity, High Performance Computing Center, and Department of Computer Science and Engineering at Notre Dame.

-
- * Corresponding author: izaguirr@nd.edu
- ¹ W. F. van Gunsteren, T. Huber, and A. E. Torda (AIP, 1995), vol. 330, pp. 253–268.
 - ² R. Elber, *Curr. Opin. Struct. Biol.* **6**, 232 (1996).
 - ³ B. J. Berne and J. E. Straub, *Curr. Opin. Struct. Biol.* **7**, 181 (1997).
 - ⁴ F. Nardi and R. Wade, *Molecular Dynamics. From Classical to Quantum Methods* (Elsevier Science B.V., 1999), chap. 21, pp. 859–898, 1st ed.
 - ⁵ A. R. Leach, *Molecular Modelling: Principles and Applications* (Prentice Hall, 2001), 2nd ed.
 - ⁶ K. Hukushima and K. Nemoto, *J. Phys. Soc. of Japan* **65**, 1604 (1996).
 - ⁷ U. H. Hansmann, *Chem. Phys. Lett.* **281**, 140 (1997).
 - ⁸ D. J. Earl and M. W. Deem, *Phys. Chem. Chem. Phys.* **7**, 3910 (2005).
 - ⁹ Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999).
 - ¹⁰ C.-Y. Lin, C.-K. Hu, and U. H. E. Hansmann, *PROTEINS:Structure,Function, and Genetics* **52**, 436 (2003).
 - ¹¹ S. S. Hampton, P. Brenner, A. Wenger, S. Chatterjee, and J. A. Izaguirre, in *New Algorithms for Macromolecular Simulation*, edited by B. Leimkuhler, C. Chipot, R. Elber, A. Laaksonen, A. Mark, T. Schlick, C. Schütte, and R. Skeel (Springer-Verlag New York, Secaucus, NJ, USA, 2006), vol. 49 of *LNCS*, pp. 103–121.
 - ¹² C. J. Woods, M. H. Ng, S. Johnston, S. E. Murdock, B. Wu, K. Tai, H. Fangohr, P. Jeffreys, S. Cox, J. G. Frey, et al., *Phil. Trans. of The Royal Society A* **363**, 2017 (2005).
 - ¹³ Y. M. Rhee and V. S. Pande, *Biophysical Journal* **84**, 775 (2003).
 - ¹⁴ F. Calvo, *J. Chem. Phys.* **123**, 124106 (2005).
 - ¹⁵ A. F. Voter, *Proceedings of the NATO Advanced Study Institute on Radiation Effects in Solids*, Edited by Kurt E. Sickafus, Eugene A. Kotomin, and Blas P. Uberuaga **235**, 1 (2006).
 - ¹⁶ A. Kone and D. A. Kofke, *J. Chem. Phys.* **122**, 206 (2005).
 - ¹⁷ T. Matthey, Ph.D. thesis, University of Bergen, Bergen, Norway (2002).
 - ¹⁸ T. Matthey, T. Cickovski, S. S. Hampton, A. Ko, Q. Ma, M. Nyerges, T. Raeder, T. Slabach, and J. A. Izaguirre, *ACM Trans. Math. Softw.* **30**, 237 (2004).
 - ¹⁹ Y. Sugita, A. Kitao, and Y. Okamoto, *J. Chem. Phys.* **113**, 6042 (2000).
 - ²⁰ D. A. Kofke, *J. Chem. Phys.* **117**, 6911 (2002).

- ²¹ K. Sanbonmatsu and A. E. Garcia, *Proteins:Struct.,Funct,Genet.* **46**, 225 (2002).
- ²² C. Predescu, M. Predescu, and C. V. CioBanu, *J. Phys. Chem. B* **109**, 4189 (2005).
- ²³ N. Rathore, M. Chopra, and J. J. de Pablo, *J. Chem. Phys.* **122**, 024111 (2005).
- ²⁴ S. Trebst, M. Troyer, and U. H. E. Hansmann, *J. Chem. Phys.* **124**, 174903 (2006).
- ²⁵ R. D. Skeel and J. A. Izaguirre, *Mol. Phys.* **100**, 3885 (2002).
- ²⁶ C. L. Brooks and D. A. Case, *Chem. Rev.* **93**, 2487 (1993).
- ²⁷ P. E. Smith, *J. Chem. Phys.* **111**, 5568 (1993).
- ²⁸ E. Barth, B. Leimkuhler, and S. Reich, in *Computational Methods for Macromolecules: Challenges and Applications*, edited by T. Barth, M. Griebel, D. Keyes, R. Nieminen, D. Roose, and T. Schlick (Springer, 2002), vol. 24 of *Lecture Notes in Computational Science and Engineering*, pp. 73–103.
- ²⁹ D. Thirumalai, R. D. Mountain, and T. R. Kirkpatrick, *Phys. Rev. A* **39**, 3563 (1989).
- ³⁰ R. H. Zhou and B. J. Berne, *J. Chem. Phys.* **107**, 9185 (1997).
- ³¹ M. K. Gilson, J. A. McCammon, and J. D. Madura, *J. Comp. Chem.* **16**, 1081 (1995).
- ³² M. B. Gillies, Ph.D. thesis, Utrecht University (2002).
- ³³ R. Wintjens, J. Wieruszeski, H. Drobecq, P. Rousselot-Pailley, L. Buee, G. Lippens, and I. Landrieu, *J. Biol. Chem.* **276**, 25150 (2001).
- ³⁴ X. Daura, W. F. van Gunsteren, and A. E. Mark, *PROTEINS: Struc., Func., and Genetics* **34**, 269 (1999).
- ³⁵ L. J. Smith, X. Daura, and W. F. van Gunsteren, *PROTEINS: Struc., Func., and Genetics* **48**, 487 (2002).

APPENDIX A: TEMPERATURE DISTRIBUTION FOR UNIFORM EXCHANGE RATE

The probability of exchange acceptance between two replicas P_{acc} can be mathematically estimated based on the overlap of normal potential energy distributions according to

$$P_{\text{acc}} = \min(1, \exp(-\Delta\beta(U_2 - U_1))), \quad (\text{A1})$$

and

$$\begin{aligned} \langle P_{\text{acc}} \rangle &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \min(1, \exp(-\Delta\beta(U_2 - U_1))) P(U_1) P(U_2) dU_1 dU_2 \\ &= \int_{-\infty}^{\infty} \int_{U_1}^{\infty} \exp(-\Delta\beta(U_2 - U_1)) P(U_1) P(U_2) dU_1 dU_2 \\ &\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{U_1} P(U_1) P(U_2) dU_1 dU_2, \end{aligned} \quad (\text{A2})$$

where

$$P(U) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(U - \mu)^2}{2\sigma^2}\right).$$

A reasonably short MD sampling segment over each temperature ensemble is taken to obtain an estimate of the potential energy distribution. When the distributions for each ensemble are approximately normal and the change in distribution mean and variance scale linearly with respect to increasing temperature we can estimate the distribution mean and variance along with exchange rates for any temperatures near or within the initial range selected for the MD sampling segments. We then determine a new temperature set for a given uniform exchange rate. This procedure is repeated until the average and deviation in observed exchange rate is acceptably close to the uniform target.

LIST OF FIGURES

1	Traversal efficiency model. Replicas modeled with normal distributions uniformly spaced for 23% near neighbor exchange	16
2	(a) Potential energy cumulative average and (b) Ergodic measure at 300K for 10 (6 Replica) simulations of 22 atom unsolvated alanine dipeptide.	17
3	3D Ramachandran Distributions (a) SNN at 3250 ps (b) APE at 3250 ps (c) SNN at 20 ns	18
4	Conformational cluster sampling measure for explicitly solvated PIN1 WW domain. Clusters based on a 1 Å RMSD separation where the RMSD is calculated over all protein atoms minus hydrogens.	19

FIGURES

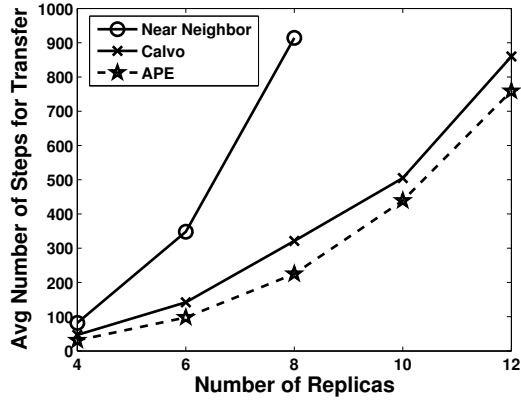
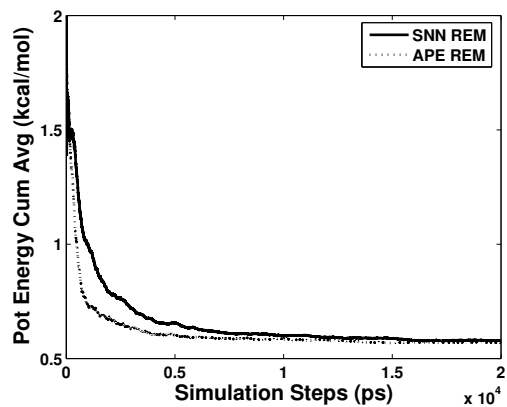
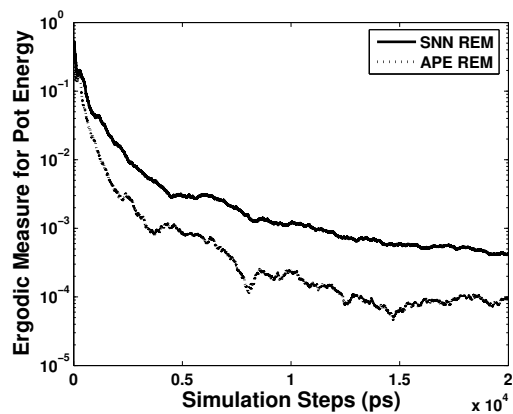


FIG. 1: Traversal efficiency model. Replicas modeled with normal distributions uniformly spaced for 23% near neighbor exchange



(a)



(b)

FIG. 2: (a) Potential energy cumulative average and (b) Ergodic measure at 300K for 10 (6 Replica) simulations of 22 atom unsolvated alanine dipeptide.

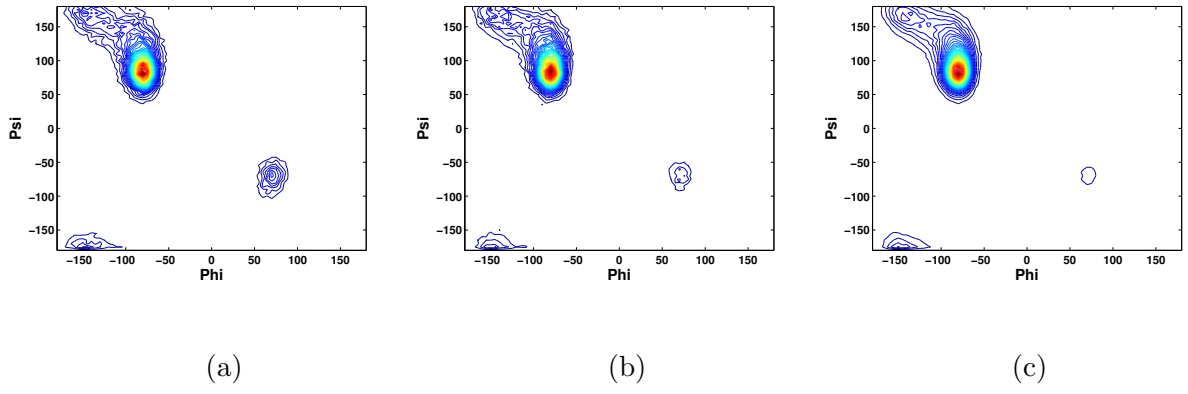


FIG. 3: 3D Ramachandran Distributions (a) SNN at 3250 ps (b) APE at 3250 ps (c) SNN at 20 ns

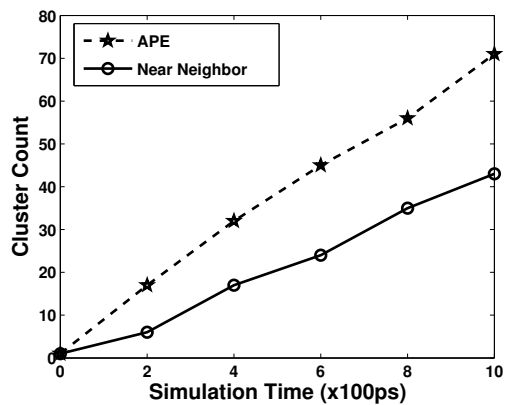


FIG. 4: Conformational cluster sampling measure for explicitly solvated PIN1 WW domain. Clusters based on a 1 \AA RMSD separation where the RMSD is calculated over all protein atoms minus hydrogens.