JCTC Journal of Chemical Theory and Computation

Letter

pubs.acs.org/JCTC

Efficient Simulation of Explicitly Solvated Proteins in the Well-Tempered Ensemble

Michael Deighan,[†] Massimiliano Bonomi,[‡,§] and Jim Pfaendtner*[,†]

[†]University of Washington Department of Chemical Engineering, Seattle, Washington, United States

[‡]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and [§]California Institute of Quantitative Biosciences, University of California, San Francisco, California, United States

Supporting Information

ABSTRACT: Herein, we report significant reduction in the cost of combined parallel tempering and metadynamics simulations (PTMetaD). The efficiency boost is achieved using the recently proposed well-tempered ensemble (WTE) algorithm. We studied the convergence of PTMetaD-WTE conformational sampling and free energy reconstruction of an explicitly solvated 20-residue tryptophan-cage protein (trp-cage). A set of PTMetaD-WTE simulations was compared to a corresponding standard PTMetaD simulation. The properties of PTMetaD-WTE and the convergence of the calculations were compared. The roles of the number of replicas, total simulation time, and adjustable WTE parameter γ were studied.

The task of quickly and accurately exploring large regions of phase space (i.e., conformational sampling) continues to be a pressing challenge in biomolecular simulations, especially for large systems. Schemes that address this challenge, frequently referred to as "enhanced sampling" methods, typically fall into two broad categories. In one class of methods, specific degrees of freedom, or collective variables (CVs), are biased in order to traverse conformational space efficiently. A prominent example of this class is the metadynamics method, which has become increasingly popular over the past decade.[1,2] The method is based on the introduction of a bias potential to accelerate sampling and reconstruct the free-energy profile along a set of CVs. Metadynamics suffers from a general problem, however: hidden degrees of freedom, which may not be accurately described by the chosen CVs, can frustrate exploration of phase space, sometimes limiting the extent of convergence and accuracy of results.

An alternate approach is to manipulate some or all degrees of freedom in a more general way (e.g., by increasing a system's temperature). The archetypical example of this class is parallel tempering (PT).[3,4] In PT, a series of replicas of a system are simulated at different temperatures. Periodically, exchanges between adjacent replicas are performed using the Metropolis criterion. A measure of efficiency of PT is the time required for a cold replica to reach the hottest temperature and come back, known as the round-trip time (RTT).[5]

A more efficient approach, incorporating attributes of the two classes mentioned above, is to combine metadynamics with PT (PTMetaD).[6] This method allows a system to overcome hidden energy barriers and comprehensively explore CV space. However, although this approach is very powerful and significantly reduces the required simulation time *per replica*, it is not a panacea. Systems containing more than a few dozen amino acids still remain prohibitively large for the PTMetaD algorithm due to a dramatic increase in the required number of replicas.

This scaling challenge is rooted in the fact that in order to achieve efficient exchange, sufficient overlap between the energy distributions of neighboring temperatures is needed. In the well-tempered ensemble (WTE) this overlap is increased by amplifying energy fluctuations with a tunable factor γ, resulting in a decreased RTT and improved convergence.[7] More importantly, since WTE preserves the average energy of the original ensemble, canonical averages of all properties of interest can be accurately obtained by reweighing techniques.[8,9]

To investigate the performance of combining PTMetaD and WTE (PTMetaD-WTE), we have applied it to an explicitly solvated 20-residue protein and performed over 20 μs of aggregate simulation. We compared PTMetaD-WTE to a full PTMetaD simulation and show that a reduction of the overall computational cost can be achieved by properly tuning γ.

All simulations were carried out using the Gromacs 4.5.3 molecular dynamics engine,[10] the AMBER99SB force field,[11] and the PLUMED 1.2.2 plug-in.[12] Calculations were performed on a 4.9 nm^3 simulation box containing the 20-residue tryptophan-cage protein (PDB entry 1L2Y[13]), 3717 TIP3P water molecules, and a chlorine ion for charge neutrality. To study the properties of PTMetaD-WTE, we performed the following simulations: a 100-replica PTMetaD simulation (I), a 100-replica PTMetaD-WTE (γ = 12; II), a 50-replica PTMetaD-WTE simulation (γ = 12; III), a 10-replica PTMetaD-WTE simulation (γ = 12; IV), and a 10-replica PTMetaD-WTE simulation (γ = 24; V). Simulations I and II were completed with 50 ns/replica, III had 100 ns/replica, and IV and V had 250 ns/replica. A total of 20 μs of sampling is reported (5 μs each for I–III and 2.5 μs each for IV and V). Further details about preparing and running the MD simulations are provided in the Supporting Information. For the well-tempered metadynamics,[14] we used CVs that have often been used for studying protein folding:[2] (1) the

Received: April 10, 2012
Published: June 8, 2012

formation of the secondary structure (i.e., hydrogen bonds) and (2) the hydrophobic core of the protein (see the Supporting Information for more information on the details of the PTMetaD parameters).

The rate of convergence of our calculations as a function of the total computational time is computed using as a reference the FES from the PTMetaD run (simulation I), as shown in Figure 1. Unlike small model systems (e.g., alanine dipeptide)

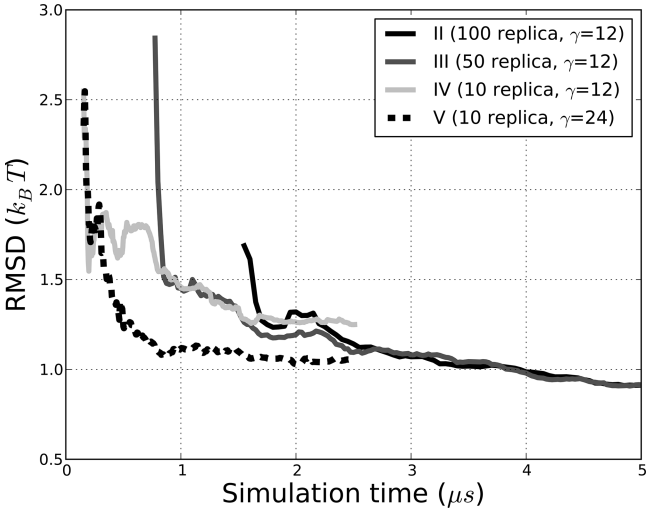

Figure 1. Convergence of the simulations using the RMSD of the FES between PTMetaD-WTE and a reference (details in Supporting Information). For comparison, the reference is taken as the final FES from simulation I, an estimate of the true FES for TRP cage folding. The four simulations shown are labeled in the figure inset. Total simulation time is defined as the simulation time per replica multiplied by the number of replicas.

in which an exhaustively sampled FES from umbrella sampling or very long metadynamics can be obtained,[14] the complexity of the trp-cage system prohibits obtaining such a perfectly sampled reference. However, we believe that the FES from simulation I is a good approximation of the real FES, given that we observed hundreds of recrossings between folded and unfolded states and that we have observed a clear convergence in the free-energy difference between folded and unfolded states (Figure S1, Supporting Information). Furthermore, the quantitative value of folding-free energy we obtained is in reasonable agreement with that determined from experiments (Figure S1), a finding that was previously reported for this same protein and force field.[15] These three criteria all point toward convergence of the reference simulation, however in principle it would be equally appropriate to assess convergence based on the behavior of other experimental observables, whose ensemble averages can be obtained from reweighing PTMetaD-WTE simulations in a straightforward way.[16]

Simulations with equal RTT (within the observed uncertainty, see Table 1) have a similar convergence trend as a function of the total simulation time (run II–IV). In contrast, simulation V shows improved convergence and a reduction of the overall computational cost. This simulation has a smaller RTT value obtained by increasing the adjustable WTE parameter to γ=24. This is especially evident when comparing simulations IV and V, as the only difference between these simulations is the value of γ. As expected, the two factors that most strongly control the convergence of the PTMetaD-WTE

Table 1. Performance Metrics for Enhanced Sampling Simulations

	run				
	I	II	III	IV	V
replicas	100	100	50	10	10
γ	1	12	12	12	24
RTT[a] (ns)	6.3(0.2)	5.1(0.1)	5.4(0.2)	5.7(0.5)	4.0(0.3)
RMSD[b] (Å)	0.34	0.38	0.37	0.38	0.46
t_{FE}[c] (ns)	10.1(0.6)	11.2(0.7)	9.1(0.8)	11.5(2.4)	8.4(1.4)

[a]Average round trip time with standard error of the mean in parentheses. [b]Simulations were initiated from an ensemble of unfolded states as described in the Supporting Information; the best agreement with lowest energy NMR structure is shown. [c]Average time per folding/unfolding event with standard error.

simulations are the RTT and the total simulation time. As we show in Figure 1, similar total simulation times can be achieved by balancing the number of replicas with the simulation time per replica. Over 2.5 μs of total simulation time, we see nearly identical convergence behavior for simulations II–IV. PTMetaD simulations at very small numbers of replicas (e.g., IV, V) have not generally been achievable, so the incorporation of the WTE framework will offer users the ability to tune the number of replicas and RTT to a particular resource availability.

It is important to note that the benefits gained from the walk through temperature correlate with the presence of energy barriers greater than k_BT at the temperature of interest. Even with the metadynamics applied to the two coarse folding CVs, we expect there to be some "hidden variables" preventing diffusive exploration of phase space once the bias potential is converged. To demonstrate that the hidden barriers are indeed energetic and not entropic, we performed a serial simulation of well-tempered metadynamics at 300 K for a simulation time of 500 ns. This simulation had only 10 folding events (~47 ns/event based on total simulation time) compared with ~10 ns/event observed for PTMetaD-WTE. Additionally, a comparison of the FESs from the reference (I), the fastest converging PTMetaD-WTE (V) after 500 ns of total sampling, and the single replica simulation after 500 ns (Figure S2, Supporting Information) shows that although a folded region can be identified, the unfolded region is dramatically overfilled due to the comparative lack of sampling. Therefore, even for studies that do not make use of the higher temperature data, the PTMetaD-WTE algorithm can offer a significant reduction in computational cost.

In spite of the fact that the situation is more complicated due to system size and the additional time-dependent metadynamics bias applied to a subset of the system's degrees of freedom, we have qualitatively reproduced all of the expected scaling behavior expected from many years of experience with standard PT simulations.[17,18] Future work will explore the convergence properties of PTMetaD-WTE calculations with a model system that permits a more systematic exploration of all relevant parameters. In light of recent work that shows a dramatic loss of accuracy when biased simulations depart too far from their original ensemble,[9] future work on such a model system should also provide quantitative guidance into the reasonable upper bound of values of γ appropriate for explicitly solvated systems. A related consideration is the extra computational cost required for completely filling the WTE bias potential at higher values of γ.

Figure 2 shows the FES as a function of the metadynamics CVs obtained in four of the simulations (I–IV). The main features (i.e., a two state system with a tightly defined folded region) are described correctly in all simulations. Additionally, we ensured the additional WTE bias did not affect the FES obtained for the metadynamics CVs. Formally, the FES in Figure 2 is obtained by computing the probability distribution of the HB/HC CVs after removing the effect of the bias on the potential energy of the system. Comparing the FES with and without the removal of the WTE bias showed the results to be equal within a 0.1 kcal/mol. The fact that WTE preserves the canonical averages guarantees a strong overlap of the biased and unbiased ensemble and facilitates the reconstruction of the latter. We additionally verified that the WTE energy distribution is not significantly altered by the bias along other metadynamics CVs. As shown in Figure S3 (Supporting Information), the distribution of the potential energy does not appreciably change from the beginning to the end of the PTMetaD-WTE simulation.

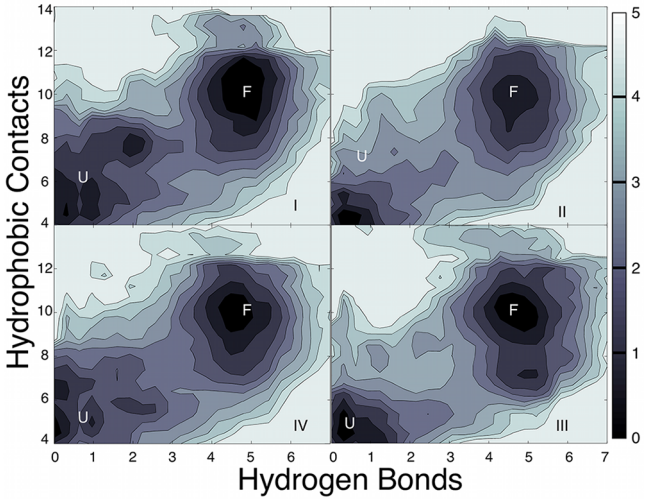

Figure 2. Free energy surfaces for runs I–IV. The folded region is marked in each by the white F. Contour lines are plotted every 0.5 kcal/mol with the energy scale shown at the right side of the figure.

In addition to ensuring the similarity of the obtained FESs, we assessed how PTMetaD-WTE affects the folding/unfolding rate of the protein relative to PTMetaD. As shown in Table 1, the average time per folding event (criterion described in SM) is essentially the same for all of the simulations, with the PTMetaD (simulation I) falling somewhere in the range of observed values from PTMetaD-WTE. This demonstrates that the additional bias on the PE of the system does not introduce an appreciable slowing of the conformational dynamics of the system.

In this letter, we reported initial observations of the convergence and properties of all-atom PTMetaD-WTE simulations. By exploiting the properties of the recently introduced WTE, we were able to maintain amplified potential energy fluctuations in a system containing an explicitly solvated tryptophan-cage protein while actively biasing two additional CVs. The convergence properties of PTMetaD-WTE depend both on the total simulation time and the RTT. Without any significant tuning of the adjustable parameter γ, we observed that the trp-cage simulations can be reduced to 10 total replicas and achieve a similar RTT (and thus equal total computational cost) to 100-replica PTMetaD and PTMetaD-WTE simulations. This is significant in that the 10 replica simulations have no energy overlap (and therefore an infinite RTT) without the WTE framework. A reduction of the overall computational cost can also be achieved by further increasing the value of the WTE parameter γ, thus reducing the RTT. The agreement found in our simulations shows that PTMetaD-WTE in all-atom simulations of biomolecules is a robust improvement to current sampling schemes. Finally, this technique presents a bridge toward the enhanced sampling of systems that are far larger than what has previously been considered, thus greatly extending the applicability of the metadynamics method.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Further details about the setup of the metadynamics calculations, the CVs used, and criteria for our analysis of properties and convergence. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jpfaendt@uw.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the support of NSF award CMMI-1032368. This research was supported in part by the National Science Foundation through TeraGrid resources provided by NICS. This work was facilitated through the use of computational, storage, and networking infrastructure provided by the Hyak supercomputer system, supported in part by the University of Washington eScience Institute. The authors gratefully acknowledge Giovanni Bussi for a careful reading of this manuscript.

■ REFERENCES

(1) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.

(2) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WIREs Comput. Mol. Sci.* **2011**, *1*, 826–843.

(3) Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **1997**, *281*, 140–150.

(4) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(5) Earl, D. J.; Deem, M. W. Parallel tempering: theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910–6.

(6) Bussi, G.; Gervasio, F.; Laio, A.; Parrinello, M. Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435.

(7) Bonomi, M.; Parrinello, M. Enhanced Sampling in the Well-Tempered Ensemble. *Phys. Rev. Lett.* **2010**, *104*, 190601.

(8) Bonomi, M.; Barducci, A.; Parrinello, M. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.* **2009**, *30*, 1615–1621.

(9) Ceriotti, M.; Brain, G. A. R.; Riordan, O.; Manolopoulos, D. E. The inefficiency of re-weighted sampling and the curse of system size in high-order path integration. *Proc. R. Soc. London, Ser. A* **2012**, *468*, 2–17.

(10) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(11) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712−725.
(12) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(13) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **2002**, *9*, 425−30.
(14) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(15) Day, R.; Paschek, D.; Garcia, A. E. Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1889−1899.
(16) Barducci, A.; Bonomi, M.; Parrinello, M. Linking well-tempered metadynamics simulations with experiments. *Biophys. J.* **2010**, *98*, L44−L46.
(17) Nymeyer, H. How Efficient Is Replica Exchange Molecular Dynamics? An Analytic Approach. *J. Chem. Theory Comput.* **2008**, *4*, 626−636.
(18) Rosta, E.; Hummer, G. Error and efficiency of replica exchange molecular dynamics simulations. *J. Chem. Phys.* **2009**, *131*, 165102.