THE JOURNAL OF PHYSICAL CHEMISTRY B LETTERS

2008, *112,* 10386–10387
Published on Web 07/31/2008

Optimized Explicit-Solvent Replica Exchange Molecular Dynamics from Scratch

Walter Nadler*,† and Ulrich H. E. Hansmann*,†,‡

John-von-Neumann Institute for Computing, Forschungszentrum Jülich, D-52425 Jülich, Germany, and Department of Physics, Michigan Technological University, Houghton, Michigan

Received: June 10, 2008; Revised Manuscript Received: July 19, 2008

Replica exchange molecular dynamics (REMD) simulations have become an important tool to study proteins and other biological molecules in silico. However, such investigations require considerable, and often prohibitive, numerical effort when the molecules are simulated in explicit solvents. In this communication we show that in this case the cost can be minimized by choosing the number of replicas as $N^{(\mathrm{opt})} \approx 1 + 0.594\sqrt{C}\,\ln(T_{\max}/T_{\min})$, where C is the specific heat, and the temperatures distributed according to $T_i^{(\mathrm{opt})} \approx T_{\min}(T_{\max}/T_{\min})^{(i-1)/(N-1)}$.

Downloaded via MUENSTER LIBRARIES on March 5, 2020 at 16:23:02 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

Replica exchange molecular dynamics (REMD) simulations have become an important tool to study proteins and other biological molecules in silico. However, such investigations require considerable, and often prohibitive, numerical effort when the molecules are simulated in explicit solvents. In this communication we show that in this case the cost can be minimized by choosing the number of replicas as

$$N^{(\mathrm{opt})} \approx 1 + 0.594\sqrt{C}\ln(T_{\max}/T_{\min}) \tag{1}$$

with C the specific heat, and the temperatures distributed according to

$$T_i^{(\mathrm{opt})} = T_{\min}\left(\frac{T_{\max}}{T_{\min}}\right)^{\frac{i-1}{N-1}} \tag{2}$$

In the replica exchange method[1−3] regular thermal Monte Carlo (MC) or molecular dynamics (MD) simulations are performed in parallel on a ladder of temperatures. $T_{\min}$ is usually the temperature of interest while $T_{\max}$ is chosen so that the largest relevant barriers in the system can be overcome within a time $t_{\min}$. At certain times ($\Delta t > t_{\min}$) the current conformations of replicas at neighboring temperatures are exchanged according to a generalized Metropolis rule.[4] As a consequence, individual replicas perform a random walk in temperature space leading to a faster convergence than observed by spending the entire computer time on regular low temperature simulations. Note that the number of round trips from low temperatures to high temperatures and back is a lower bound for the number of independent configurations observed at $T_{\min}$ and, therefore, measures the efficiency of the algorithms.

Depending on the system under study, the required numerical costs can still be daunting. A prime example is simulations of biological molecules, such as proteins, in explicit solvent: For realiable studies, the solvent part has to be much larger than the biomolecule investigated. There have been a number of attempts in recent years to speed up simulations by optimizing the temperature discretization.[5−12] The most promising are based on the analysis of the flow of replicas across temperature space,[9−11] which allows for an iterative improvement of the discretization. The advantage of these schemes is that they are very general. However, they are also often very costly. For this reason, in the present letter, we choose a different approach. Utilizing a priori knowledge on the system, for the special case of proteins in an explicit solvent we derive simple formulas for optimizing the temperature distribution and the number of replicas. For this we assume that the behavior of the full system is dominated by the solvent part, and most solvent models employed do not show a phase transition in the temperature range of interest.[13] The problem of broken ergodicity,[11] i.e., the state space of the full system partitioning into disjoint subsets, is less severe in such cases. Moreover, such systems exhibit a heat capacity that is practically constant over a wide range of temperatures,[13] a feature that allows a simple analytical estimate of the acceptance probabilities. Constant heat capacity and lack of broken ergodicity are the two assumptions required in the following derivation of our formulas.

In the following, we consider REMD with N replicas and N temperatures T_i numbered $i = 1, \ldots, N$. As ergodicity is not broken, the flow transition probabilities between neighboring temperatures are given by the average acceptance probability.[11] The second ingredient is that for systems with a constant heat capacity C (in units of k_B) over a wide range of temperatures the density of states can be approximated by that of a $d = 2C$ dimensional harmonic oscillator.[5−7] The resulting average acceptance probability for replica exchange between temperatures T and T' is given by

$$p_{\mathrm{acc}}(T,T') = \frac{2B\left(\frac{1}{1+R};C,C\right)}{B(C,C)} \tag{3}$$

with $R = \mathrm{Max}(T'/T, T/T')$ and the incomplete Beta-function, $B(x;a,b) = \int_0^x t^{a-1}\,(1-t)^{b-1}\,\mathrm{d}t$ in the numerator, while the complete Beta function, $B(a,b) = B(1;a,b)$, is used in the denominator. Optimal flow across temperature space requires

* To whom correspondence should be addressed. E-mail: wnadler@fz-juelich.de; hansmann@mtu.edu.
† John-von-Neumann Institute for Computing.
‡ Michigan Technological University.

10.1021/jp805085y CCC: $40.75 © 2008 American Chemical Society

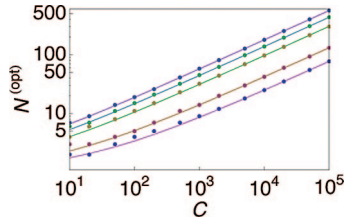

Figure 1. Optimal number of replicas, N^{opt}, vs total heat capacity C for various values of T_{max}/T_{min}), (from bottom 1.5, 2,5, 10, and 20), determined from minimizing the round trip time, eq 4, and using eq 3 as acceptance probability; the lines denote the approximation eq 1.

constant acceptance probabilities.[11] As a consequence, R is constant, and $R^{opt} = (T_{max}/T_{min})^{1/(N-1)}$ leads to the optimal temperature distribution of eq 2. Discretization along these lines has been proposed before,[5−8] but see also the discussion in ref 12. We emphasize here that it is indeed the optimal one under the above assumptions.

Further optimization is possible by minimizing the average round trip time τ as a function of the number of replicas.[14] As all transition probabilities are equal, τ is given by[14]

$$\tau \propto \frac{N(N-1)}{p_{acc}(T_n^{(opt)}, T_{n+1}^{(opt)})} = \frac{N(N-1)B(C,C)}{2B\left(\frac{1}{1+R^{(opt)}};C,C\right)} \quad (4)$$

Minimizing this quantity for particular values of C and (T_{max}/T_{min}) leads to the data points shown in Figure 1. Their asymptotic behavior can be determined analytically. Using the complementary error function, eq 3 can be approximated for large values of C by

$$p_{acc}(T,T') \approx \mathrm{erfc}\left(\frac{R-1}{R+1}\sqrt{C}\right) \quad (5)$$

Defining a scaled number of replicas,

$$\nu = \frac{N}{\sqrt{C}\ln(T_{max}/T_{min})} \quad (6)$$

the average round trip time is given for large values of N in the parameter-free form

$$\tau \propto \frac{\nu^2}{\mathrm{erfc}(1/2\nu)} \quad (7)$$

Minimizing the above equation numerically leads to $\nu^{opt} = 0.594$. Inserting ν^{opt} in eq 6 and including corrections for small C, we obtain our expression for the optimal number of replicas in eq 1. Figure 1 demonstrates the quality of that approximation.

Equations 2 and 1 are the central results of this contribution. They provide a simple cooking recipe for optimizing REMD simulations of proteins with explicit solvent (and other systems with unbroken ergodicity and constant heat capacity) from scratch, without the need for any iteration. Required input are only the values of two extremal temperatures, T_{min} and T_{max}, and the total heat capacity C. The latter can be determined either by a canonical simulation run at T_{max}, or, if the specific heat of the solvent model, c_{solv}, is known beforehand, via the approximation $C \approx Mc_{solv}$, with M the number of solvent molecules. Given their ease of use, we expect that our formulas[15] will increase further the usefulness of explicit solvent REMD for studies of proteins and other biological molecules.[16−19]

Acknowledgment. This work is supported, in part, by research Grant CHE-0809002 of the National Science Foundation (U.S.A.).

References and Notes

(1) Geyer, C. J.; Thompson, A. *J. Am. Stat. Assoc.* **1995**, *90*, 909–920.
(2) Hukushima, K.; Nemoto, K. *J. Phys. Soc. (Jpn)* **1996**, *65*, 1604–1608.
(3) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.
(4) Metropolis, N.; Rosenbluth, A.; Rosenbluth, M.; Teller, A.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
(5) Kofke, D. A. *J. Chem. Phys.* **2002**, *117*, 6911–6914. Erratum: *J. Chem. Phys.* **2004**, *120*, 10852.
(6) Predescu, C.; Predescu, M.; Ciabanu, C. *J. Chem. Phys.* **2004**, *120*, 4119–4128.
(7) Kofke, D. A. *J. Chem. Phys.* **2004**, *121*, 1167.
(8) Kone, A.; Kofke, D. A. *J. Chem. Phys.* **2005**, *122*, 206101.
(9) Katzgraber, H. G.; Trebst, S.; Huse, D. A.; Troyer, M. *J. Stat. Mech.* **2006**, *6*, P03018.
(10) Trebst, S.; Troyer, M.; Hansmann, U. H. E. *J. Chem. Phys.* **2006**, *124*, 174903.
(11) Nadler, W.; Hansmann, U. H. E. *Phys.Rev. E* **2007**, *75*, 026109.
(12) Garcia, A. E.; Herce, H.; Paschek, D. *Ann. Rep. Comp. Chem.* **2006**, *2*, 83–95.
(13) Paschek, D.; Nymeyer, H.; Garcia, A. E. *J. Struct. Biol.* **2007**, *157*, 524–533.
(14) Nadler, W.; Hansmann, U. H. E. *Phys. Rev. E* **2007**, *76*, 065701R.
(15) We note that violation of our assumptions could easily be detected by replica flow analysis.[9−11] Further iterative improvement of the temperature discretization and of the value of N is then readily possible along the lines sketched in refs 9−11 and 14. Nevertheless, even in such a situation our results provide the optimal preoptimization given the knowledge about the system.
(16) Nanias, M.; Czaplewski, C.; Scheraga, H. A. *J. Chem. Theor. Comput.* **2006**, *5*, 513–528.
(17) Kmiecik, S.; Gront, D.; Kolinski, A. *BMC Struct. Biol.* **2007**, *7*, 43.
(18) Li, H.; Li, G.; Berg, B. A.; Yang, W. *J. Chem. Phys.* **2006**, *125*, 144902.
(19) Simmerling, C.; Strockbine, B.; Roitberg, A. *J. Am. Chem. Soc.* **2006**, *124*, 11258.

JP805085Y