

## Using Novel Variable Transformations to Enhance Conformational Sampling in Molecular Dynamics

Zhongwei Zhu

*Department of Chemistry, New York University, New York, New York 10003*

Mark E. Tuckerman

*Department of Chemistry and Courant Institute of Mathematical Sciences, New York University, New York, New York 10003*

Shane O. Samuelson and Glenn J. Martyna\*

*Department of Chemistry, Indiana University, Bloomington, Indiana 47405*

(Received 28 March 2001; published 25 February 2002)

One of the computational “grand challenges” is to develop methodology capable of sampling conformational equilibria in systems with rough energy landscapes. Here, a significant step forward is made by combining molecular dynamics with a novel variable transformation designed to enhance sampling by reducing barriers without introducing bias and, thus, to preserve, perfectly, equilibrium properties.

DOI: 10.1103/PhysRevLett.88.100201

PACS numbers: 02.70.Ns, 05.10.-a, 36.20.-r, 87.15.-v

One of the premier challenges in molecular simulation is to determine accurately and efficiently conformational equilibria of systems described by complex potential surfaces that possess high energy barriers separating important minima or basins of attraction. Indeed, rough energy landscapes are ubiquitous in physical, chemical, and biological problems, including protein folding. Despite recent advances [1,2], this important class of problems cannot, yet, be adequately addressed by molecular simulation.

Current approaches include umbrella sampling, guiding potentials, parallel tempering algorithms, and rate-enhancing schemes. The first two techniques reweight the phase space distribution in order to accelerate sampling at the price of requiring an *a posteriori* correction. Thus, the efficiency scales exponentially poorly with the number of degrees of freedom involved in the reweighting factor. Parallel tempering methods introduce a series of  $M$  simulations at different temperatures which are permitted to interchange. Thus, the cost of the calculation is  $M$ -fold larger, and the number of temperature interchanges accepted scales exponentially poorly with the system size. Other methods that rely on harmonic transition state approximations do not, in general, preserve equilibrium properties.

In this Letter, a novel approach to the conformational sampling problem that is free of the above difficulties is presented. The method is based on an *exact* reformulation of the classical statistical mechanical partition function via a nonlinear transformation of the integration variables. The transformation is constructed so as to effect a warping of the configuration space wherein barrier regions are shrunk, barrier heights are lowered, and attractive basins stretched. The result is a smoother surface, which allows barriers to be easily traversed without (i) altering equilibrium properties of the system, (ii) requiring an *a posteriori* reweighting, (iii) relying on harmonic approximations, or

(iv) scaling exponentially poorly with the number of transformed degrees of freedom. The technique is combined, here, with molecular dynamics (MD) to form a novel conformational sampling scheme, although it could also be used with Monte Carlo. Moreover, the new approach could be combined with complementary transformations [3] that help overcome entropic barriers as well as employed to enhance the efficiency of other techniques such as parallel tempering. Applications of the new method to model and complex systems as large as 400-mer alkane chains are considered in order to demonstrate its favorable scaling with system size and its enhancement in sampling efficiency (a factor of  $10^6$  over standard MD for the long chains, as will be seen shortly).

Consider, first, a classical particle with momentum,  $p$ , mass,  $m$ , and coordinate,  $x$ , in a one-dimensional potential,  $V(x)$ , which possesses two stable minima, separated by a large energy barrier. The canonical partition function is

$$Q(\beta) = \frac{1}{h} \int dx dp \exp\left\{-\beta \left[ \frac{p^2}{2m} + V(x) \right]\right\}, \quad (1)$$

where  $h$  is Planck's constant, and  $\beta = 1/k_B T$ . The spatial probability distribution function,  $P(x) \propto \exp[-\beta V(x)]$ , appearing in Eq. (1), could be sampled using a classical Hamiltonian,  $H = p^2/2m + V(x)$ , in conjunction with thermostated MD [4]. However, if the potential barrier is too high, barrier crossing events will be rare, and extremely long trajectories will be needed to generate an adequate sampling of the configuration space.

Since the variable,  $x$ , is just an integration variable, a transformation to a new variable,  $u = f(x)$ , can be performed arbitrarily without altering the partition function in Eq. (1), assuming that the inverse  $x = f^{-1}(u) \equiv g(u)$  exists and is single valued. Upon changing variables and exponentiating the Jacobian factor  $g'(u)$ , the partition function becomes

$$Q(\beta) = \frac{1}{h} \int dp du \exp\left[-\beta\left[\frac{p^2}{2m} + V_{\text{eff}}(u)\right]\right]. \quad (2)$$

The key feature of Eq. (2) is the appearance of an effective potential,  $V_{\text{eff}}(u) \equiv V(g(u)) - (1/\beta) \ln g'(u)$  that can be adjusted via the transformation to be smoother than  $V(x)$ . It is important to note that this approach completely eliminates the need for *a posteriori* reweighting factors required by the umbrella and guiding potential methods and does not introduce harmonic approximations. It is, therefore, not equivalent to any other existing schemes.

A spatial warping transformation is achieved via the following change of variables:

$$u = f(x) = cx_0 + c \int_{x_0}^x dy \exp[-\beta V_r(y)]. \quad (3)$$

Here,  $V_r(x)$  is an arbitrary reference potential,  $x_0$  is an arbitrary point, and  $c$  is a constant. Since the integrand in Eq. (3) is nonnegative and  $u$  is a monotonically increasing function of  $x$ , a unique inverse  $x = g(u)$  exists. The Jacobian of the transformation is  $du/dx = c \exp[-\beta V_r(x)]$ . Substituting Eq. (3) and the Jacobian into Eq. (2) gives an effective potential  $V_{\text{eff}}(u) \equiv \{V(g(u)) - V_r(g(u))\}$ . If  $V_r(x)$  is taken to be equal to the bare potential,  $V(x)$ , in the barrier region and zero outside and  $x_0$  is taken to be the left minimum ( $x_0 = -a$ ), then  $V_{\text{eff}}(u)$  is barrier free [see Fig. 1(a)]. Thus, the Jacobian weights the “ $u$ ” space so as to reduce the barrier region to a negligibly small volume without altering the partition function. We refer to the new approach as REPSWA (reference potential spatial warping algorithm).

The reformulated partition function, Eq. (2), can be sampled by a canonical MD approach based on a Hamiltonian,  $\hat{H} = p^2/2m + V_{\text{eff}}(u)$ , forming the REPSWA-MD method. Since the transformation is not canonical,  $\hat{H}$  generates trajectories which sample phase space more effectively than those of  $H = p^2/2m + V(x)$ . That is, the true dynamics, in which barrier crossing is, by definition, a rare event, is not, nor is intended to be, preserved.

Since the integral in Eq. (3) cannot generally be evaluated analytically, the function,  $\exp[-\beta V_r(x)]$ , is expressed in a finite basis of integrable functions, thereby defining a transformation  $u = \hat{f}(x)$ , so that the integrals can be performed. The function,  $\hat{f}(x)$ , will be very close to  $f(x)$  in Eq. (3) for a large enough basis set. This is a key step that allows the REPSWA-MD to be applied beyond the harmonic approximation. Clearly, the transformation,  $u = \hat{f}(x)$ , also exactly preserves the partition function.

The REPSWA-MD technique is first applied to a one-dimensional analytically solvable system in order to demonstrate its validity and performance. A double-well potential of the form  $V(x) = (V_b/a^4)(x^2 - a^2)^2$  is selected, and canonical MD simulations are performed for a high barrier,  $V_b = 10k_B T$ . The equations of motion for  $\{p, u\}$  are coupled to a Nosé-Hoover chain thermostat [4] to effect the canonical sampling. Runs of length of

$10^6$  steps were performed using a time step of  $\Delta t = 0.01$  ( $k_B T = 1$ ,  $m = 1$ , and  $a = 1$ ,  $c = 1$ ). Figure 1 shows trajectories  $x(t) = g(u(t))$  with and without the REPSWA transformation as well as the probability distribution function,  $P(x)$ . The ability of the REPSWA technique to reproduce the correct distribution,  $P(x)$ , demonstrates that the partition function is perfectly preserved. In contrast, ordinary canonical MD based on  $H$  is unable to generate the correct distribution within  $10^6$  steps indicating, quantitatively, an increase in efficiency of over 6 orders of magnitude using REPSWA.

Next, the nontrivial extension of the REPSWA method to enhance sampling of multidimensional potential surfaces is discussed. In particular, chain molecules possess many dihedral barriers in the range of  $(5-10)k_B T$ , which, as the previous example has demonstrated, are traversed infrequently, hindering conformational sampling. It will first be shown how to treat the commonly used “united atom” (UA) molecular models (e.g.,  $\text{CH}_n$ -like moieties are merged into pseudoatoms) [5], and then the additional steps needed to treat all-atom models [6] will be discussed.

Dihedral barriers can be removed in united atom models by applying the REPSWA transformation [cf. Eq. (3)] directly to the dihedral angle. However, since dihedral angles are not explicit coordinates in MD, additional steps are needed in order to apply the technique. Consider a UA representation of butane, and assume the connectivity is  $C_1-C_2-C_3-C_4$ , where  $C_i$ , the  $i$ th pseudoatom, has Cartesian position  $\mathbf{r}_i$ . The transformation scheme is (i) the vector  $\mathbf{r}_4$  is rotated/translated into a frame in which  $\mathbf{r}_3$  is at the origin and  $\mathbf{r}_3 - \mathbf{r}_2$  lies along the  $z$  axis; (ii) the new  $\mathbf{r}'_4$  is then resolved into spherical polar coordinates,  $(r, \theta, \phi)$ ;

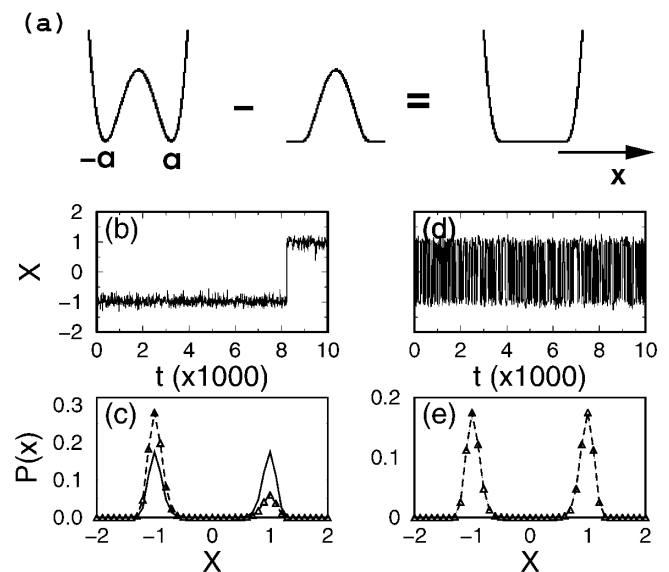


FIG. 1. (a) Schematic of REPSWA for a double-well potential. (b) and (d)  $x(t)$  in the double-well potential ( $V_b = 10k_B T$ ) without(b)/with(d) REPSWA. (c) and (e) Probability distribution without(c)/with(e) REPSWA (dashed lines) and analytical result (solid line).

(iii) REPSWA is applied to the azimuthal angle to generate  $\phi_u$ ; (iv) a Cartesian  $\mathbf{u}'_4$  is then created using  $(r, \theta, \phi_u)$ ; (v)  $\mathbf{u}'_4$  is placed back into the original frame by inverting the transformations (i) and (ii) to obtain a transformed set of Cartesian coordinates. The constant  $c$  is adjusted to ensure that  $\phi_u \in [0, 2\pi)$ . The reference potential in step (iii) is chosen to be equal to the dihedral potential between the two gauche conformations and zero otherwise. The forces on the  $\mathbf{u}$  are obtained by the chain rule. This procedure allows for facile transformation of all the dihedral angles in an  $N$ -pseudoatom chain by moving sequentially down the chain. This procedure forms an upper triangular Jacobian matrix whose determinant is the product of the exponentials of the reference potentials, which cancel the desired barriers, while preserving the partition function. Both the coordinate and force transformations can be carried out in  $\mathcal{O}(N)$  operations.

The behavior of the dihedral REPSWA scheme is studied using the Ryckaert-Bellemans model [5] for a 400-mer chain at  $T = 300$  K (ignoring intermolecular Lennard-Jones interactions to avoid freezing the chain). It should be noted that guiding potentials/umbrella sampling could not be used to enhance the sampling of *all 397 dihedral angles*. The molecule was simulated using canonical molecular dynamics in the gas phase with a time step of 0.5 fs for  $10^6$  steps with and without the REPSWA dihedral angle transformation. Figure 2 demonstrates the large improvement in sampling efficiency given by REPSWA. The evolution of the end-to-end distance,  $R^2(t)/Nd_{C-C}^2$  ( $d_{C-C} = 1.53$  Å,  $N = 400$ ) [Figs. 2(c) and 2(f)] indicates that the fluctuations in this quantity are virtually nonexistent without REPSWA. *Quantitatively*, the number of

dihedral angle flips per torsion increases by a factor of 30, and the correlation time for the relaxation of  $R^2(t)$  decreases by 6 orders of magnitude. Note, the REPSWA calculation was found to be in good agreement with the prediction  $(\langle R^2 \rangle)/Nd_{C-C}^2 \approx 10$  of a three state rotational-isomeric state (RIS) model [7], parametrized using our potential.

Last, it is shown how REPSWA can be used to simulate “all-atom” models of chain molecules with high efficiency [see Fig. 3(a)]. In all-atom models, the backbone  $C_1-C_2-C_3-C_4$  dihedral angle is strongly coupled to the  $H_1-C_1-C_2-C_3$ ,  $H_2-C_1-C_2-C_3$ , and  $H_3-C_1-C_2-C_3$  dihedrals due to the  $sp^3$  hybridization of the carbon atoms (represented by harmonic bond and bend angle potentials). The rigidity caused by the hybridization requires groups of atoms to be transformed as rigid units. Groups of three atoms [e.g.,  $C_3, H_4$ , and  $H_5$  form such a group in Fig. 3(a)] are selected, and a “primary” dihedral angle (e.g.,  $H_1-C_1-C_2-C_3$ ), defined. A set of “secondary” dihedrals is formed by collecting dihedral angles involving atoms both in and to the left of the group (e.g., eight). The reference potential is then taken to be  $V_r(\phi_p) = [V_p^{(d)}(\phi_p) + \sum_{i=1}^8 V_{s,i}^{(d)}(\phi_p + \delta_i)]S(\phi_p)$ , where  $V_p^{(d)}(\phi)$  and  $V_{s,i}^{(d)}(\phi)$  are the primary and secondary dihedral potentials, respectively,  $S(\phi_p)$  is a function that switches off the potential outside the region between the two gauche conformations, and  $\{\delta_i\}$  is a set of constants chosen based on the ideal differences between the secondary and primary dihedral angles for  $sp^3$  geometry. In a simulation, the actual secondary dihedrals will generally not differ much from  $\phi_{\text{prim}} + \delta_i$ . The full REPSWA

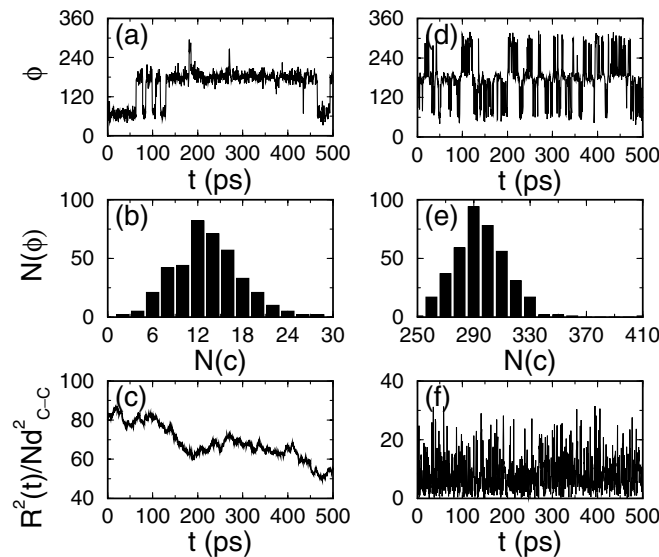


FIG. 2. (a) and (d) Dihedral angle,  $\phi_{C_{198}-C_{199}-C_{200}-C_{201}}(t)$ , in the UA 400-mer without(a)/with(b) REPSWA. (b) and (e) Number of dihedrals,  $N(\phi)$ , that have crossed barriers  $N(c)$  times without(b)/with(e) REPSWA. (c) and (f) End-to-end distance [ $R^2(t)/Nd_{C-C}^2$ ,  $d_{C-C} = 1.52$  Å] without(c)/with(f) REPSWA.

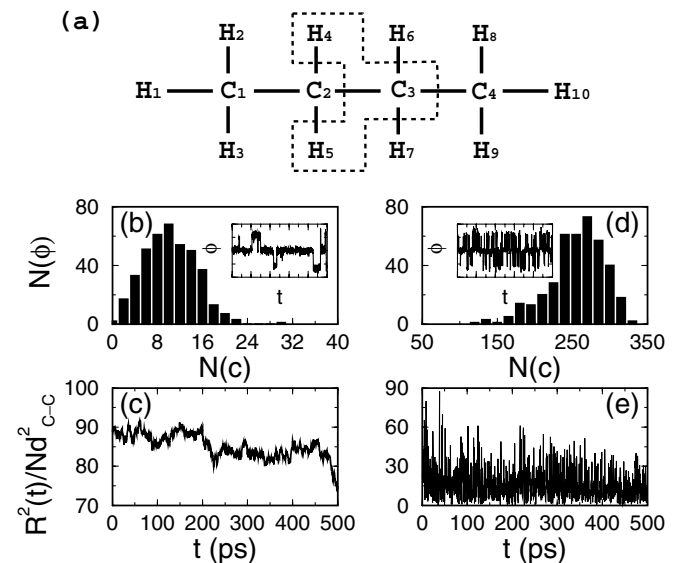


FIG. 3. (a) Schematic of group REPSWA for  $C_4H_{10}$ . (b) and (d) Number of dihedrals,  $N(\phi)$ , that have crossed barriers  $N(c)$  times without(b)/with(d) REPSWA; inset:  $\phi_{C_{198}-C_{199}-C_{200}-C_{201}}(t)$ . (c) and (e) End-to-end distance [ $R^2(t)/Nd_{C-C}^2$ ,  $d_{C-C} = 1.53$  Å] without(c)/with(e) REPSWA.

transformation procedure is then the same as that for the united atom case, except that three vectors  $\mathbf{r}_{C_2-C_3}$ ,  $\mathbf{r}_{C_2-H_4}$ , and  $\mathbf{r}_{C_2-H_5}$  are rotated about the  $C_1-C_2$  axis.

The performance of the group REPSWA transformation scheme is tested on an all-atom 400-mer using the CHARMM22 force field [6]. The simulation details are the same as in the united atom case (see above). Figures 3(b) and 3(d) show the histogram of backbone dihedral angle “flips,” again, without and with REPSWA, respectively; the insets show the evolution of a typical backbone dihedral for the two cases. It can be seen that the use of REPSWA leads to a much greater dihedral flipping rate by a factor of 30, and, hence, an improved sampling of conformational space by a factor of  $10^6$  based on  $R^2(t)/Nd_{C-C}^2$  [Figs. 3(d) and 3(e)]. Again, the REPSWA calculation was found to be in good agreement with the RIS model [7] prediction ( $\langle R^2 \rangle / Nd_{C-C}^2 \approx 15$ ).

In conclusion, a novel method for enhancing conformational sampling simulations of complex molecular systems, REPSWA-MD, has been introduced. REPSWA-MD smooths a rough potential energy surface via a nonlinear variable transformation and, thus, does not alter the equilibrium properties of a system. Unlike guiding potentials or umbrella sampling, the method does not require *a posteriori* reweighting and can be used to enhance sampling in a very large number of degrees of freedom without exponentially poor scaling. A dramatic improvement in the rate of barrier crossing events and the number of statistically independent conformations sampled has been demonstrated (e.g., 6 orders of magnitude for a 400-mer chain). The new technique can easily be combined with complementary transformations [3] that help overcome entropic barriers as well as with parallel tempering schemes. Our current efforts are focused on applying REPSWA-MD to proteins and other large biomolecules.

This research was supported by PRF 32139-AC and NSF CHE-96-5015 (G. J. M.), by Research Corp. RI0218, NSF CHE-98-75824, and by an NYU Whitehead Fellowship in Biomedical and Biological Sciences (M. E. T.) and NSF-EIA-0081307 (G. J. M and M. E. T.).

\*Present address: T.J. Watson Research Center, International Business Corporation, P.O. Box 218, Yorktown Heights, New York 10598.

- [1] J. I. Siepmann and D. Frenkel, *Mol. Phys.* **75**, 59 (1992); P. Amara and J. E. Straub, *J. Phys. Chem.* **99**, 14 840 (1995); R. Olender and R. Elber, *J. Chem. Phys.* **105**, 9299 (1996); N. Nakajima, H. Nakamura, and A. Kidera, *J. Phys. Chem. B* **101**, 817 (1997); R. H. Zhou and B. J. Berne, *J. Chem. Phys.* **107**, 9185 (1997); H. Xu and B. J. Berne, *J. Chem. Phys.* **110**, 10 229 (1999); Y. H. Lee and B. J. Berne, *J. Phys. Chem. A* **104**, 86 (2000); S. Melchionna, *Phys. Rev. E* **62**, 8762 (2000); P. Eastman and N. Gronbech-Jensen, *J. Chem. Phys.* **114**, 3823 (2001).
- [2] D. D. Frantz, D. L. Freeman, and J. D. Doll, *J. Chem. Phys.* **93**, 2769 (1990); E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451 (1992); B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992); B. M. Forest and U. W. Suter, *Mol. Phys.* **82**, 393 (1994); A. Monge *et al.*, *J. Mol. Biol.* **247**, 995 (1995); H. Jönsson, *Surf. Sci.* **324**, 305 (1995); T. Lazaridis and M. Karplus, *Science* **278**, 1928 (1997); J. E. Shea *et al.*, *J. Chem. Phys.* **109**, 2895 (1998); C. Dellago *et al.*, *J. Chem. Phys.* **108**, 1964 (1998); S. S. Plotkin and P. G. Wolynes, *Phys. Rev. Lett.* **80**, 5015 (1998); J. Pillardy, A. Liwo, and H. A. Scheraga, *J. Phys. Chem. A* **103**, 9370 (1999); H. Lu and K. Schulten, *Proteins* **35**, 453 (1999); E. Shakhnovich and L. Mirny, *Biophys. J.* **78**, 761 (2000); R. J. Petrella and M. Karplus, *J. Phys. Chem. B* **104**, 11 370 (2000); M. R. Sorensen and A. F. Voter, *J. Chem. Phys.* **112**, 9599 (2000).
- [3] M. E. Tuckerman, B. J. Berne, G. J. Martyna, and M. L. Klein, *J. Chem. Phys.* **99**, 2796 (1993); M. E. Tuckerman and G. J. Martyna, *J. Phys. Chem. B* **104**, 159 (2000); S. Samuelson, A. Hughes, and G. J. Martyna, *J. Chim. Phys.* **94**, 1503 (1997); S. Samuelson, K. Pihakari, and G. J. Martyna (to be published).
- [4] G. J. Martyna, M. E. Tuckerman, and M. L. Klein, *J. Chem. Phys.* **97**, 2635 (1992).
- [5] J. P. Ryckaert and A. Bellemans, *Faraday Discuss* **66**, 95 (1978); M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B* **102**, 2569 (1998).
- [6] A. MacKerell, Jr. *et al.*, *J. Phys. Chem. B* **102**, 3586 (1998).
- [7] P. J. Flory, *Statistical Mechanics of Chain Molecules* (Wiley-Interscience, New York, 1969).