Resolution Exchange Simulation

Edward Lyman,* F. Marty Ytreberg, and Daniel M. Zuckerman†

Department of Computational Biology, School of Medicine and Department of Environmental and Occupational Health, Graduate School of Public Health, Suite 3064 BST3, 3501 Fifth Avenue, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA

(Received 16 June 2005; published 18 January 2006)

We extend replica-exchange simulation in two ways and apply our approaches to biomolecules. The first generalization permits exchange simulation between models of differing resolution—i.e., between detailed and coarse-grained models. Such ''resolution exchange'' can be applied to molecular systems or spin systems. The second extension is to ''pseudoexchange'' simulations, which require little CPU usage for most levels of the exchange ladder and also substantially reduce the need for overlap between levels. Pseudoexchanges can be used in either replica or resolution exchange simulations. We perform efficient, converged simulations of a 50-atom peptide to illustrate the new approaches.

DOI: 10.1103/PhysRevLett.96.028105 PACS numbers: 87.15.Aa, 05.10.−a

The simulation of biomolecular systems with 10^4–10^5 degrees of freedom has become routine because of the accessibility of powerful computing resources. In principle, standard Monte Carlo and molecular dynamics algorithms are perfectly ergodic, and therefore will eventually generate properly distributed conformational ensembles. In practice, the μ sec - sec time scales, which describe biologically relevant fluctuations, are not within reach of computation even for small proteins.

In one approach to the problem, developed early on [1,2], coarse-grained protein representations are adopted [3,4]. A second class of strategies attempts directly to enhance sampling of atomic-resolution models, including multiple time step methods [5], replica exchange [6], parallel tempering [7–9], and generalized ensemble techniques [10]. Parallel tempering (PT), which employs a ladder of replicas simulated at increasing temperatures, is widely used for state-of-the-art molecular dynamics simulations, but presently is limited to small proteins [11], as the resources required increase rapidly with the system size.

This Letter presents two new tools for biomolecular simulation, by extending the PT approach and exploiting the speed of coarse-grained models. The first extension is a ''resolution exchange'' (ResEx) algorithm which—instead of using high-temperature simulation to increase sampling, as does PT—uses inexpensive coarse-grained models to cross barriers. Boltzmann-weighted ensembles are produced. The algorithm is implemented in close analogy to PT and can also be applied to magnetic systems (e.g., the Ising model). The ResEx approach is natural for proteins, and, indeed, the kernel of the idea was suggested in the early days of protein simulation [1]. More recently, the approach has been implemented in an *ad hoc* way, without proper statistical weighting [3]. Also, a rigorous method to calculate free energy differences between all-atom and coarse-grained models was demonstrated [12].

Our ResEx approach is conceptually related to work on Ising systems by Brandt and co-workers (e.g., [13]). The ResEx approach is distinguished, however, by its simplicity, its ready applicability to biomolecules, and the ability to employ arbitrary coarse-grained Hamiltonians—rather than truly renormalized Hamiltonians, as in [13].

We also introduce ''pseudoexchange'' (PsEx) processes, which should significantly improve the efficiency of *any type of exchange simulation,* whether one swaps temperatures (as in PT), Hamiltonians [14], or resolution (ResEx). Pseudoexchanges are performed between a simulation in progress and an existing trajectory. Critically, PsEx permits uneven distribution of CPU time among levels of the exchange ladder. Because all exchange simulations are limited by the sampling obtained at the highest level—i.e., highest temperature (for PT) or lowest resolution (ResEx)—the bulk of CPU time should be devoted to this top level. Although an uneven distribution of CPU time among levels (replicas) would be awkward in a truly parallel implementation, it is natural and highly efficient in a serial PsEx simulation. Furthermore, there is essentially no disadvantage to multiple independent runs, as compared to a single parallel simulation.

Resolution exchange theory.—The key idea behind ResEx is that, in addition to swapping temperature labels (PT) or parameters of the Hamiltonian [14], one can also swap a *subset* of configurational coordinates. A well-chosen subset of coordinates of a detailed model can make up the full set of coordinates for a coarse-grained model, as we demonstrate below.

A general exchange process involves two *independent* simulations of a protein (or a spin system) carried out in parallel, each sampling its own distribution π_1 or π_2. Typical distributions are given by $\pi_i(\Phi_i, \mathbf{x}_i; \mathbf{k}_i; T_i) = \exp[-U(\Phi_i, \mathbf{x}_i; \mathbf{k}_i)/k_B T_i]/Z_i$, with partition function Z_i. A configuration is composed of coordinates $\{\Phi, \mathbf{x}\}$ that include an arbitrarily chosen ''coarse'' subset Φ, and $\mathbf{k}$

denotes the parameters of the potential function U, while k_BT is the product of Boltzmann's constant and the temperature. A general exchange process consists of a swap of a set of coordinates: swapping the full set $\{\Phi, \mathbf{x}\}$ for $\mathbf{k}_1 = \mathbf{k}_2$ leads to PT, while a swap when $\mathbf{k}_1 \neq \mathbf{k}_2$ leads to Hamiltonian exchange. To achieve resolution exchange, one can swap values of the set of coarse coordinates, $\Phi_1 \leftrightarrow \Phi_2$, noting that the corresponding potential parameters $\mathbf{k}_\Phi$ need not match in the two systems. It is, indeed, possible to swap an arbitrary set of coordinates, with any mixture of parameters.

Specializing, for clarity, to resolution exchange, we consider independent simulations governed by a ''high-resolution'' potential function $U_H(\{\Phi, \mathbf{x}\})$ and a coarse-grained (low-resolution) potential $U_L(\{\Phi\})$. Occasionally, we attempt an exchange move by swapping the $\{\Phi\}$ subset. The set $\{\Phi, \mathbf{x}\}$ may be, for example, all the atomic coordinates of a protein, while the subset $\{\Phi\}$ may be only the coordinates of the backbone. For a spin system, $\{\Phi\}$ may correspond to a block spin, and $\mathbf{x}$ to the orientations of the local spins relative to the block spin.

To develop the exchange criterion, assume that at an exchange point the system is characterized by a high-resolution configuration $\{\Phi_a, \mathbf{x}_a\}$ and a low-resolution configuration $\{\Phi_b\}$. Attempting to exchange the $\{\Phi\}$ subset yields the trial conformations $\{\Phi_b, \mathbf{x}_a\}$ and $\{\Phi_a\}$. Because the simulations are independent, the weight of the composite system is given by the simple product $\pi_{\text{tot}} = \pi_1\pi_2 = \pi_H\pi_L$, and detailed balance will be satisfied if we accept such moves with a Metropolis rate $\min[1, R]$, where R is given by

$$R = \frac{\pi_1(\text{new})\pi_2(\text{new})}{\pi_1(\text{old})\pi_2(\text{old})} = \frac{\pi_H(\Phi_b, x_a)\pi_L(\Phi_a)}{\pi_H(\Phi_a, x_a)\pi_L(\Phi_b)}. \quad (1)$$

The analogy to PT and Hamiltonian exchange is clear, but we have now extended the approach.

Naturally, there are limitations on the types of models that can be successfully exchanged, much as PT temperature increments are limited. In the results presented below, we successfully performed exchanges between all-atom and united-atom models of a peptide.

Pseudoexchange simulation.—The ResEx and PT algorithms are motivated by the likelihood that the ''top-level'' simulation (i.e., lowest resolution or highest T) will more rapidly cross barriers and converge to an equilibrium ensemble of conformations. While the associated convergence time is expected to be quite long, even for the top level, it is far from clear that the attainment of an equilibrium ensemble at a lower level requires the same length of simulation. Indeed, given that barriers should be crossed many times at the top level, *significantly less simulation time should be required at the lower levels of the exchange ladder.*

The pseudoexchange process is key to efficiently distributing computing time among ladder levels. One first generates a well-sampled ensemble at the top level (highest temperature or lowest resolution) and randomly reorders this trajectory [Figs. 1(a) and 1(b)]. While the shuffled trajectory preserves the original distribution of states, it exhibits a feature key for exchange simulation: *extremely rapid barrier hops,* as in Fig. 1(b).

One now performs a PsEx simulation with the shuffled trajectory. As with conventional exchange, one runs an independent lower-level simulation [Fig. 1(c)], but now exchanges are attempted with the shuffled top-level trajectory, using (1) or its PT analog. If the exchange attempt is successful, the new lower-level trajectory is continued from the accepted configuration, and the top-level trial configuration is simply discarded. The process is repeated.

Pseudoexchange processes are useful for several reasons: (i) PsEx processes may be used with any exchange simulation; (ii) much lower acceptance ratios are still efficient because frequent pseudoexchange attempts are inexpensive in a serial scheme; and (iii) because of the weaker acceptance ratio requirements, larger gaps among ladder levels (e.g., T increments in PT) can be tolerated.

Simple ResEx demonstration: Butane.—We present simulation results for a single butane molecule, in order to show that the ResEx-PsEx algorithm reproduces the correct ensemble.

Results are presented for two different low-resolution potentials, one of which intentionally breaks the symmetry between the two *gauche* conformers. Comparison data were generated with the TINKER v. 4.2 simulation package [15], using the CHARMM-27 all-atom force field [16]. The all-atom butane molecule was simulated in vacuum for 1 μ sec, at a temperature of 298 K. Langevin dynamics were used, with a friction coefficient of 91 ps^{-1}. The same force field, dynamics, and simulation package were used for the high-resolution portions of the resolution exchange run, except that every 10 fsec a ResEx move was attempted. In all, 10^5 resolution exchanges were attempted, for a total trajectory length of 1 nsec.

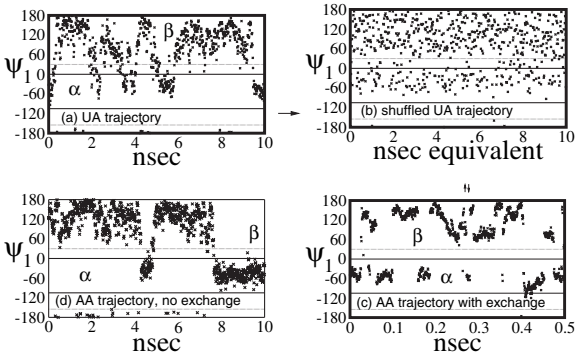

FIG. 1. Coarse-grained simulation accelerates transitions in high-resolution simulation with ResEx and PsEx. (a) United-atom trajectory for dileucine, showing transitions between α and β states, which is randomly reordered to create trajectory (b) with extremely rapid transitions. Via pseudoexchanges with the shuffled trajectory, ResEx generates the all-atom trajectory in (c). The time axes in (c) and (d) differ by a factor of 20.

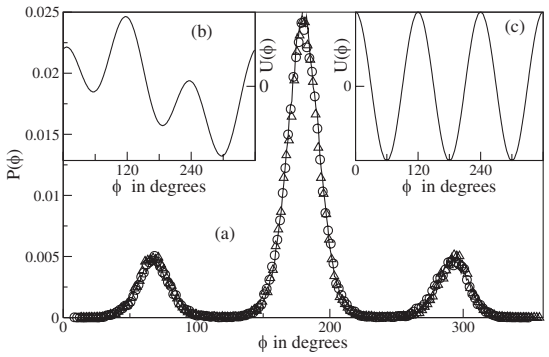

FIG. 2. ResEx produces canonical sampling despite a poor coarse-grained potential. (a) Probability densities $P(\phi)$ for butane. Reference data from standard simulation are indicated by the solid line. The ResEx simulation with the asymmetric potential (b) is plotted with triangles, and ResEx simulation with the symmetric potential (c) is plotted with circles.

Figure 2 compares the results of two different 1 nsec ResEx simulations to a 1 μsec reference trajectory. Plotted is the distribution of the C-C-C-C dihedral angle, ϕ, which measures the conformer populations. The ResEx simulations reproduce the equilibrium distribution, regardless of the potential used for the low-resolution simulation. The low-resolution model was a one-dimensional potential of the form $A\cos(3\phi) + B\sin(\phi)$, where ϕ is the C-C-C-C dihedral angle. Potential (b) has $A = B = 1$, while (c) has $A = 1$ and $B = 0$.

ResEx for a peptide: Dileucine peptide.—We also tested the ResEx-PsEx method on the dileucine peptide [ACE-$(Leu)_2$-NME; ''dileucine'']. Though hardly a full size protein, 50-atom dileucine allows us to address a number of key issues. A united-atom (UA) representation, which omits nonpolar hydrogens, is a natural low-resolution model. Compared to 50 atoms in an all-atom (AA) representation, there are 24 in UA.

The goal is to generate efficiently a converged ensemble of conformations for all-atom dileucine, using the PsEx-ResEx protocol. We focus on the free energy difference between the two dominant conformations, distinguished by rotations about the ψ_1 angle of Leu1 and the ϕ_2 angle of Leu2. Transitions between these two basins are hampered by a significant barrier, and therefore occur rarely (approximately 1/3 nsec at 300 K for AA). We define the ''α'' conformations by $-105 < \psi_1 < 0$ and $-145 < \phi_2 < -25$, and ''β'' by $30 < \psi_1 < -155$ and $-160 < \phi_2 < -40$.

The AA dileucine molecule was modeled with the OPLSaa force field [17]. For UA dileucine, we used a slightly modified version of the OPLSua force field [18], altering a few of the bond length and bond angle parameters to match those in the all-atom force field: these simple changes reduce the likelihood of exchange-induced steric clashes. Both simulations were carried out with TINKER v. 4.2 [15], using Langevin dynamics with a 91 psec^{-1} friction coefficient, a 1 fsec time step, and ''generalized Born'' (GB/SA) implicit solvation [19].

Implementing the ResEx-PsEx strategy, we first carried out a simulation of UA dileucine [Fig. 1(a)]. We then randomly reshuffled this trajectory [Fig. 1(b)] generating a pseudotrajectory with much more frequent $\alpha \leftrightarrow \beta$ transitions than in the original trajectory. The randomized trajectory is then used to generate the AA trajectory [Fig. 1(c)] via the ResEx protocol. Notice the far more frequent $\alpha \leftrightarrow \beta$ transitions in the all-atom trajectory with exchange (about 30/nsec) than without [about 1/3 nsec, Fig. 1(d)].

Assessing convergence and efficiency of a protein simulation is generally a difficult task. Fortunately, the situation here is relatively simple, as we can consider the free energy difference between the α and β substates ($\Delta G_{\alpha\beta}$). In Fig. 3, 8 ResEx simulations are compared to 8 standard stochastic dynamics simulations. Each ResEx data point represents the average and range of 8 independent ResEx trajectories, with ResEx moves attempted every 10 fsec. The ResEx estimates are displaced from the origin to reflect the time invested in the united-atom model. We allotted AA simulation time by attempting to match precision among levels—i.e., generating the same number of interbasin hops as in the UA simulation. Convergence is assessed by considering the spread among the independent simulations.

It is clear that the ResEx simulations reproduce the $\Delta G_{\alpha\beta}$ estimated from the standard simulations. This is accomplished despite the failure of the united-atom model to reflect correctly the populations of the α and β states: $\Delta G_{\alpha\beta}(\text{UA}) = 0.690 k_B T$, with the wrong sign. This is an important point, as it is known that united-atom models do not reproduce all-atom behavior [20]. More important for

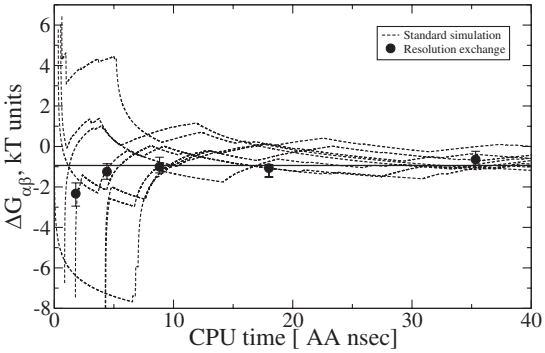

FIG. 3. ResEx simulation accelerates equilibration among dileucine substates. Dashed lines show running estimates of the intersubstate $\Delta G_{\alpha\beta}$ as a function of time for 8 independent, 40 nsec trajectories, and the solid line shows the estimate of $\Delta G_{\alpha\beta}$ from 600 nsec of standard simulation. Each symbol with error bars gives the average and range of 8 independent ResEx simulations, displaced from the origin to reflect total CPU time (measured in all-atom time steps), including investment in the united-atom simulation. The efficiency gain can be estimated by the relative ranges of the $\Delta G_{\alpha\beta}$ estimates.

ResEx simulation is that the coarse-grained model explores conformational space more rapidly, as well as being ''exchangeable'' with the more detailed model.

It is further clear that the ResEx results are generated with significantly higher efficiency. For a given amount of CPU time (nsec in Fig. 3), the ResEx estimates exhibit high accuracy with a greatly reduced uncertainty. For example, 5 nsec of resolution exchange simulation generated an estimate for $\Delta G_{\alpha\beta} = -1.25 \pm 0.40 k_B T$, while 72 nsec of standard simulation are required to reach a comparable level of accuracy and precision, indicating a 15-fold savings in CPU time. We emphasize that our analysis includes the total CPU time, rather than the cost for one of many parallel simulations.

The acceptance ratio of attempted pseudoexchange moves need not be 20%, as conventional wisdom dictates [21]. Indeed, even a very small fraction of accepted exchanges can greatly enhance efficiency, provided those exchanges generate novel conformations. The goal is to optimize diffusion in conformation space, not acceptance ratio. In ResEx trajectories presented here the average acceptance ratio was only 0.156%. Nonetheless, high efficiency is obtained because successful exchanges with a shuffled top-level trajectory are very likely to generate novel conformations, at a fraction of the cost of standard simulation.

Discussion.—We have introduced two extensions of parallel tempering (replica exchange), which show promise for improved efficiency of biomolecular simulations. ''Resolution exchange'' enhances sampling of an expensive, high-resolution model using a cheaper, coarse-grained model. Generalization to a ladder of models is formally trivial. The sampling in the high-resolution model satisfies detailed balance, and therefore generates an equilibrium ensemble. The further introduction of the ''pseudoexchange'' process permits the bulk of computer resources to be invested in sampling and crossing barriers at the top level of the exchange ladder (highest temperature or lowest resolution), and only incremental additional simulation is required at lower levels.

The treatment of even larger, more complex molecules will be the subject of future research. Our efficiency gains here were obtained using only a two-level ladder, implying that much greater efficiency is possible with additional levels—which can be added at a small cost via pseudoexchange. A long-term goal is to develop a full ladder of reduced, exchangeable models, extending up to the ''united residue'' level, because UA computations are themselves slow. Ultimately, resolution and temperature exchange might be combined for high-efficiency simulations of biomolecules.

Two limitations should be mentioned. First, we do not expect the present algorithm to enable exchange between continuum and explicit solvent representations. However, the present degree of undersampling of proteins, when using continuum solvent representations, warrants pursuit of this problem in its own right. A second limitation is that, to be exchangeable, two models must be sufficiently ''similiar'': there should be overlap between low-energy coarse variable conformations. Yet, a process of incremental coarsening—changing part of a molecule at a time, and which we have already implemented for dileucine (data not shown)—will minimize this difficulty for larger systems.

We thank B. J. Berne, C. J. Camacho, M. W. Deem, and R. H. Swendsen for insightful comments. This work was supported by the Department of Environmental and Occupational Health and the Department of Computational Biology at the University of Pittsburgh, and by the NIH (Grants No. ES007318 and No. GM070987).

*Electronic address: elyman@ccbb.pitt.edu
†Electronic address: dmz@ccbb.pitt.edu

[1] M. Levitt and A. Warshel, Nature (London) **253**, 694 (1975).
[2] H. Taketomi, Y. Ueda, and N. Go, Int. J. Peptide Protein Res. **7**, 445 (1975).
[3] G. M. S. De Mori, C. Micheletti, and G. Colombo, J. Phys. Chem. B **108**, 12 267 (2004).
[4] D. M. Zuckerman, J. Phys. Chem. B **108**, 5127 (2004).
[5] D. D. Humphreys, R. A. Friesner, and B. J. Berne, J. Phys. Chem. **98**, 6885 (1994); B. Hetényi, K. Barnacki, and B. J. Berne, J. Chem. Phys. **117**, 8203 (2002).
[6] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **57**, 2607 (1986).
[7] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
[8] U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).
[9] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[10] Q. Yan, T. S. Jain, and J. J. de Pablo, Phys. Rev. Lett. **92**, 235701 (2004).
[11] D. Paschek and A. E. García, Phys. Rev. Lett. **93**, 238105 (2004).
[12] Z. Fan, J. Hwang, and A. Warshel, Theor. Chem. Acc. **103**, 77 (1999).
[13] A. Brandt and D. Ron, J. Stat. Phys. **102**, 231 (2001); D. Ron, R. H. Swendsen, and A. Brandt, Phys. Rev. Lett. **89**, 275701 (2002).
[14] Y. Sugita, A. Kitao, and Y. Okamoto, J. Chem. Phys. **113**, 6042 (2000).
[15] J. W. Ponder and F. M. Richard, J. Comput. Chem. **8**, 1016 (1987); http://dasher.wustl.edu/tinker/.
[16] N. Foloppe and A. D. MacKerell, J. Comput. Chem. **21**, 86 (2000).
[17] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11 225 (1996).
[18] W. L. Jorgensen, J. D. Madura, and C. J. Swenson, J. Am. Chem. Soc. **106**, 6638 (1984).
[19] W. C. Still, A. Tempczyk, and R. C. Hawley, J. Am. Chem. Soc. **112**, 6127 (1990).
[20] M. H. Zaman, M.-Y. Shen, R. S. Berry, and K. F. Freed, J. Phys. Chem. B **107**, 1685 (2003).
[21] A. Kone and D. A. Kofke, J. Chem. Phys. **122**, 206101 (2005).