

Monte Carlo Sampling with Hierarchical Move Sets: POSH Monte Carlo

Jerome Nilmeier* and Matthew P. Jacobson

*Graduate Group in Biophysics, University of California, San Francisco,
California 94158*

Received November 30, 2008

Abstract: We present a new Monte Carlo method for sampling rugged energy landscapes that allows for efficient transitions across sparsely distributed local basins. The trial move consists of two steps. The first step is a large initial trial move, and the second step is a Monte Carlo trajectory generated using smaller trial moves. To maintain detailed balance, a reverse transition probability is estimated along a path that differs from the forward path. Since the forward and reverse transitions are different, we name the algorithm POSH (port out, starboard home) Monte Carlo. The process obeys detailed balance to the extent that the transition probabilities are correctly estimated. There is an optimal range of performance for a given energy landscape, which depends on how sparsely the low energy states of the system are distributed. For simple model systems, adequate precision is obtained over a large range of inner steps settings. Side chain sampling of residues in the binding region of progesterone antibody 1dba are studied, and show that significant improvement over a comparable standard protocol can be obtained using POSH sampling. To compare with experimental data, the phosphopeptide Ace-Gly-Ser-Ser-Ser-Nma is also studied, and the resulting NMR observables compare well with experiment. For the biomolecular systems studied, we show that POSH sampling generates precise distributions using the number of inner steps set up to 20.

Introduction

The Metropolis algorithm¹ has been in use for over 50 years, with generalizations of the idea applied to fields far beyond the field of molecular modeling, for which it was initially developed. For systems with densely packed geometries, the generation of good trial moves can be a substantial challenge. A variety of methods in the field of liquid simulation have been developed to address this type of problem.^{2–4} For polymeric systems, chain growth methods,^{5,6} pivot moves,⁷ and other methods have been developed using the idea of biased sampling in polymeric^{8–10} and continuum contexts.^{11,12}

Many sampling algorithms have been designed expressly for crossing large energetic barriers and obeying, or nearly obeying, detailed balance. Most of these involve generating a Markov chain of states (a walker) in an expanded state space,¹³ which can be accepted to the configuration space of interest based on a modified acceptance probability.^{14–18}

The most widely adopted of these types is the replica exchange¹⁹ method,²⁰ and other closely related methods,^{21–24} which are particularly well suited to parallel processing.

The landscapes of biomolecular systems are rugged, which has motivated development of methods designed to locate global minima. For biomolecular systems, molecular dynamics is widely used, but it has been shown that all atom MC approaches can provide significant improvements in sampling.²⁵ Methods such as Monte Carlo Minimization²⁶ have emerged as a practical method for sampling landscapes that have multiple, sparsely distributed minima. This produces an approximately correct distribution while sampling a broader range of configurations. The approximation results in an underestimate of the entropy of local basins. More rigorous estimates of basin entropy can be made, as is done with the Mining Minima approach.^{27,28} In many cases, particularly in difficult biomolecular optimization problems, such as protein folding,^{29,30} drug design, and homology modeling,^{31–34} the use of minimization and other optimiza-

* To whom correspondence should be addressed. E-mail: jerome.nilmeier@ucsf.edu.

tion approaches have become the mainstay of the field, with entropic considerations added as a secondary effect, if at all.

In this work, we propose a simple, general method for sampling rugged landscapes that obeys detailed balance to the extent that the transition probabilities are correctly estimated. The estimate of the transition probabilities is discussed in the theory section. The basic idea is to generate a large initial trial perturbation, followed by a series of small Monte Carlo moves to anneal the initial trial move to a lower energy, and accept the resulting trial move with a modified acceptance probability.

Motivation

Partitioning of Configuration Space. A Monte Carlo sampling strategy seeks to evaluate integrals of the type

$$\begin{aligned}\langle O \rangle &= \frac{1}{Z} \int d\mathbf{q} O(\mathbf{q}) \exp(-\beta U(\mathbf{q})) \\ Z &= \int d\mathbf{q} \exp(-\beta U(\mathbf{q}))\end{aligned}\quad (1)$$

where \mathbf{q} is the set of all coordinates of the system of interest, $\beta = (k_B T)^{-1}$ is the inverse temperature, and $U(\mathbf{q})$ is the potential energy. The observed quantity $\langle O \rangle$ is averaged over many instances of $O(\mathbf{q})$. Z is the normalization constant or configuration integral.

The Boltzmann factor in the integrand is the unnormalized probability distribution:

$$p(\mathbf{q}) = \exp(-\beta U(\mathbf{q})) \quad (2)$$

Often, a natural partition of the entire space becomes convenient. A common example is the Born–Oppenheimer approximation, where the nuclear degrees of freedom are considered to be uncoupled or adiabatic relative to the electronic degrees of freedom.⁶ This type of approximation also appears in the formulation of implicit solvation models, where the solvent degrees of freedom are integrated out, and an approximate model for the interaction between a macromolecule and the solvent is introduced.³⁵ Propagation along adiabatic degrees of freedom has been introduced in both Monte Carlo^{11,36,37} and dynamical^{36–38} contexts. For the applications motivating this work, the coordinate decomposition is between protein backbone and side chain coordinates.

The motivation for sampling separate subspaces is often guided by the definition of a configuration space partitioning where the covariant fluctuations between the partitioned subspaces are small. This can often be justified by a dynamical argument, as is the case for the examples given above. The practical motivation for decomposition of subspaces is computational efficiency. In proteins, for example, different geometric algorithms are appropriate for sampling backbone^{39,40} and side chain degrees of freedom,⁴¹ and the challenge lies in combining these trial moves in a way that preserves ergodicity, while generating a high acceptance ratio and, of course, the expected distribution of states.

Figure 1 shows a schematic to motivate the development of a sampling protocol. The complete configurational space is partitioned into torsional coordinates of the backbone (ϕ) and torsional coordinates of the side chains (χ). Consider a trial move that consists of randomly selecting a subspace to

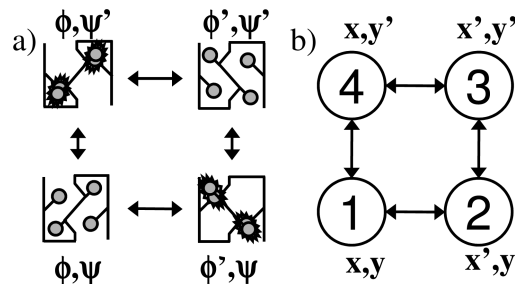


Figure 1. (a) Schematic of the multiple subspace sampling problem for proteins. Backbone and side chain coordinates are ϕ and χ , respectively. (b) Generalization of the 2 subspace sampling problem. \mathbf{x} and \mathbf{y} represent subspaces to be sampled.

perturb (ϕ or χ), and accepting with the Metropolis criterion. A perturbation in backbone space may generate a configuration with the side chains in a high energy state. Likewise, a perturbation of side chains only may also lead to a high energy state. Although a series of samples from the trial state could generate a lower energy state, the initial move would be rejected outright with a standard scheme. If both degrees of freedom are perturbed simultaneously, however, the likelihood of generating a reasonable trial can become vanishingly small. A system that remains in a macrostate for long times relative to local correlations is often referred to as quasi-ergodic,¹⁵ frustrated/glassy,⁴² or kinetically trapped.

For the remainder of the work, the *original* coordinate state will be labeled as state 1, the *initial trial state* will be state 2, and the *final trial state* will be labeled state 3. The *reflected trial state*, labeled as state 4, will be defined shortly. These states are illustrated in Figure 2. In the cases shown in Figure 1, the most intuitive solution is to generate a series of Monte Carlo steps in the alternate subspace to locate a low energy coordinate and then accept the final trial coordinate with some reasonable probability. In this sense, the transition from the initial trial to the final trial can be thought of as an annealing step.

Theory

Detailed Balance. To determine the proper acceptance criterion, the condition of detailed balance is:

$$p_i T_{ij} = p_j T_{ji} \quad (3)$$

where i and j are two arbitrary coordinate states with probabilities $p_i = p(\mathbf{q}_i)$ and $p_j = p(\mathbf{q}_j)$, as given by eq 2, respectively. The transition probability T_{ij} is the probability of transitioning to coordinate state j from state i , and T_{ji} is the reverse transition probability. The condition of detailed balance will be applied to the states 1 and 3, as shown in Figure 2

$$p_1 T_{13} = p_3 T_{31} \quad (4)$$

where the forward transition consists of a trial move followed by a chain of moves. In the cases described by Figure 2a and b, for example, the initial trial could be also notated as $p_2 = P(\mathbf{x}', \mathbf{y})$, where \mathbf{x}' is a trial move, and so on, as labeled.

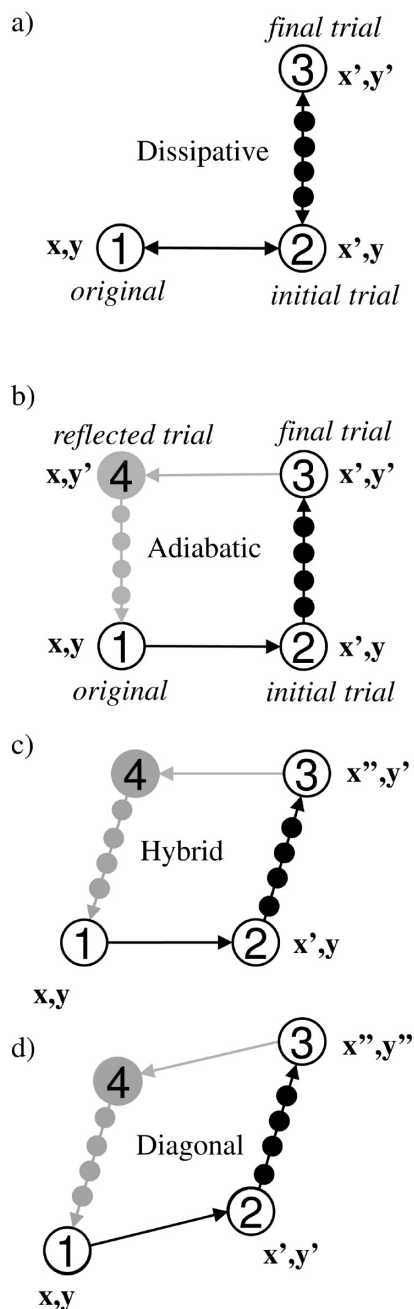


Figure 2. Topological representation of detailed balance condition (eq 4): (a) dissipative pathway (eq 5) and (b–d) posh pathways (eq 12), (b) adiabatic pathway, (c) hybrid pathway, and (d) diagonal pathway. Calculation of coordinate 4 is described in the text and in Figure 3. Coordinate labels correspond to state numbers.

The condition of detailed balance is met only if we choose to enforce the flow rates between states 1 and 3. If all states were accounted for (1–4), then only the condition of balance would be satisfied.⁴³ In either case, however, the satisfaction of eq 4 will ensure a proper distribution of states.

Hierarchical Perturbations. The general scheme is a hierarchical decomposition of move sets. The initial perturbation is designed to be a large move which will cover large regions of configurational space, while the series of trial moves in the “inner loop” are much smaller and designed to be an annealing move. The initial trial move, labeled $\xi^{(1,2)}$, is, for all cases here, a vector of uniform variates over some

domain $[-\mathbf{d}^{(1,2)}/2, \mathbf{d}^{(1,2)}/2]$, where $\mathbf{d}^{(1,2)}$ is a vector of the same dimension as the complete space (with zero entries for the degrees of freedom that are not sampled). The series of Monte Carlo steps in the inner loop use a different perturbation type, $\xi_n^{(2,3)}$, where n is the inner step number, and the domain is $[-\mathbf{d}^{(2,3)}/2, \mathbf{d}^{(2,3)}/2]$.

Figure 2 shows schematically the ways that the perturbations can differ. The first pathway, labeled as the dissipative pathway, is used frequently in nonequilibrium studies. For the adiabatic pathway, the initial perturbation is only within one subspace, \mathbf{x} , while the annealing steps are within the \mathbf{y} subspace only. For the hybrid pathway, the initial perturbations are in \mathbf{x} , while the remaining perturbations are in either \mathbf{x} or \mathbf{y} (or both). Finally, the diagonal path allows for perturbations in both \mathbf{x} and \mathbf{y} in both the initial and final step, with the differences being in the magnitude of the perturbations.

The essential feature of the algorithm presented is that the perturbation domains $\mathbf{d}^{(1,2)}$ and $\mathbf{d}^{(2,3)}$ perturbations differ in some hierarchical way. Typically, the initial perturbations will be larger, such that local basins can be traversed. The annealing step uses smaller perturbations to search for nearby low energy states from the initial trial. As long as the total space is covered by the combination of perturbations, meaning that there are no zero values in the vector sum $\mathbf{d}^{(1,2)} + \mathbf{d}^{(2,3)}$, the complete sampling of space is possible.

This approach is not, in and of itself, guaranteed to solve the quasi-ergodic problem, however. The connectivity between states that is generated by perturbation domain $\mathbf{d}^{(1,2)}$ can still limit the accessibility of alternative macrostates. The use of the annealing steps, however, permits larger move sets in the initial trial step that might not otherwise be practical. Table 1 describes the types of move sets used for all simulations generated for the present work.

Acceptance Criteria. The forward transition can be defined as a combination of moves described diagrammatically in Figure 2. Consider first the forward transition probability, which for all cases follows the pathway (1 \rightarrow 2 \rightarrow 3), which is a combination of the (1 \rightarrow 2) transition and the (2 \rightarrow 3) transition. The difference in the nature of the reverse pathway determines the two acceptance criteria presented.

Dissipative Pathway and Acceptance Criterion. The standard approach from this point would be to require that the reverse transition be along the pathway (3 \rightarrow 2 \rightarrow 1). See Figure 2a. For this case, the condition for detailed balance is

$$p_1 \alpha_{12} \alpha_{23} \text{acc}_{13} = p_3 \alpha_{12} \alpha_{32} \text{acc}_{31} \quad (5)$$

where α_{ij} and acc_{ij} are the selection and acceptance probabilities, respectively, of state j from state i . This particular condition of detailed balance has been invoked in the study of nonequilibrium systems^{43,44} and is labeled here as the dissipative pathway. The trial coordinate \mathbf{q}_2 is generated as

$$\mathbf{q}_2 = \mathbf{q}_1 + \xi^{(1,2)} \quad (6)$$

where $\xi^{(1,2)}$ is a uniform deviate vector that perturbs along the (1 \rightarrow 2) portion of the pathway. Figure 3a shows the

Table 1. Simulation Settings for Simulations of Model Systems^a

figure	B	$\mathbf{d}_x^{(1,2)}$	$\mathbf{d}_y^{(1,2)}$	$\mathbf{d}_x^{(2,3)}$	$\mathbf{d}_y^{(2,3)}$	δb	path type	reverse pathway
5	45	22.5	22.5	9	9	1.0	diagonal	true
6	60	60	60	12	12	0.25	diagonal	true
11	45	22.5	0	0	9	0.9	adiabatic	true
11	45	22.5	0	9	9	0.9	hybrid	true
11	45	22.5	22.5	9	9	0.9	diagonal	true
12	45	22.5	22.5	9	9	0.9	diagonal	both

^a Parameters of landscapes as described in Appendix A. B represents the lower and upper bound of the landscape in both the domains of both x and y . The values in columns 3 and 4 are the lower and upper bounds of the initial ($1 \rightarrow 2$) perturbation for x and y , respectively, and columns 5 and 6 are the lower and upper bounds for the ($2 \rightarrow 3$) perturbation for x and y , respectively. δb is the square bin width of the 2D histograms collected.

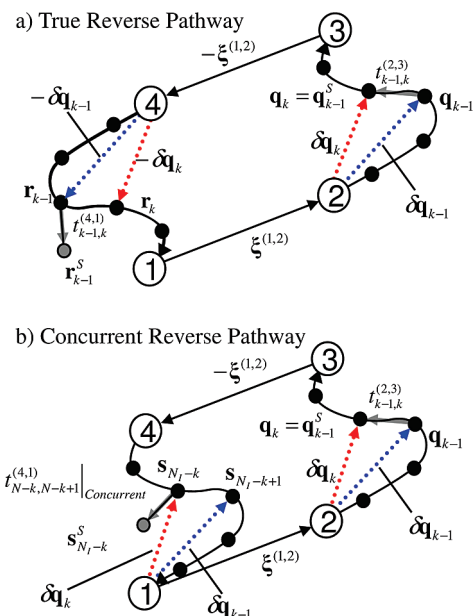


Figure 3. Notation and indexing for transition matrix construction. Coordinate states 1–4 are described in the text. The perturbation from 1 to 2 is given by the vector $\xi^{(1,2)}$, which also connects states 3 and 4. The forward trajectory along the ($2 \rightarrow 3$) pathway is connected by a line with black dots, which represent coordinates along the annealed path. A set of trials at step ($k-1 \rightarrow k$) is shown for reference. For clarity, the case of an accepted trial move in both forward and reverse transitions are shown. For the forward trajectory, the accepted move is a black dot connected with a gray arrow, given by eqs 17 or 19. The gray arrow represents the transition for which a probability is computed. The red and blue arrows show the difference vectors for which the reverse trajectory is constructed (eq 14). The reverse transition probability connects the $k-1$ reverse coordinate to an accepted trial coordinate, which is not necessarily the same as the next step in the reverse pathway. The only difference between a and b is in the way that the reverse coordinate is constructed using the difference vectors (shown in red and blue): (a) true reverse pathway (eq 22) and (b) concurrent reverse pathway (eq 24).

forward path construction in detail. The final trial coordinate \mathbf{q}_3 is generated using a series of Monte Carlo transitions.

The selection probabilities along the ($2 \rightarrow 3$) and ($3 \rightarrow 2$) are given by the transition probabilities⁴⁵

$$\begin{aligned} \alpha_{23} &= T_{23}^{(N_i)} \\ \alpha_{32} &= T_{32}^{(N_i)} \end{aligned} \quad (7)$$

where $T_{23}^{(N_i)}$ and $T_{32}^{(N_i)}$ are the transition probabilities along the respective Markov chains of length N_i . Returning to eq 5, we can see that the probability of selecting state 3 from state 1 is the probability of selecting the trial move ($1 \rightarrow 2$) and selecting the final trial ($2 \rightarrow 3$). The joint probability of this selection process is represented by the product of the selection probabilities. The final trial ($2 \rightarrow 3$) selection probability is determined by the transition probability of the inner loop sampling procedure. Equation 7 simply states that the probability of selecting state j from state i is the transition probability of the Markov Chain. The transition probability of a multistep stochastic walk is described by the Chapman–Kolmogorov equation

$$T_{ij}^{(N_i)} = T^{(N_i)}(\mathbf{q}_j|\mathbf{q}_i) = \int d\mathbf{q}_1 \dots d\mathbf{q}_{N_i-1} \prod_{k=1}^{N_i} t^{(ij)}_{k-1,k}(\mathbf{q}_{k-1}|\mathbf{q}_k) \quad (8)$$

where $k=0$ corresponds to initial state i and $k=N_i$ corresponds to state j , and $t^{(ij)}_{k-1,k}(\mathbf{q}_{k-1}|\mathbf{q}_k) = t^{(ij)}_{k-1,k}$ is the transition probability from state $k-1$ to state k (at step k)

$$T_{ij}^{(N_i)} = \prod_{k=1}^{N_i} t^{(ij)}_{k-1,k} \quad (9)$$

Equations 8 and 9 form an exact description of the transition probability, which incorporates all possible paths connecting states i and j , if the transition elements $t^{(ij)}_{k-1,k}$ are known rigorously.

If each inner step obeys detailed balance, the resulting detailed balance condition is satisfied⁴⁶

$$p_2 T_{23}^{(N_i)} = p_3 T_{32}^{(N_i)} \quad (10)$$

Solving eq 27 gives

$$\frac{\text{acc}_{13}}{\text{acc}_{31}} = \frac{p_3 T_{32}^{(N_i)}}{p_1 T_{23}^{(N_i)}} = \frac{p_2}{p_1} \quad (11)$$

This criterion, while not originally derived as a Monte Carlo acceptance criterion, is an important criterion because it provides a foundation for thinking of these systems in broader terms. This criterion can be thought of as the work required to place the system in a trial state. This represents a rigorously correct acceptance criterion that does not require the estimation of a transition probability, which can give insight into limiting behavior.

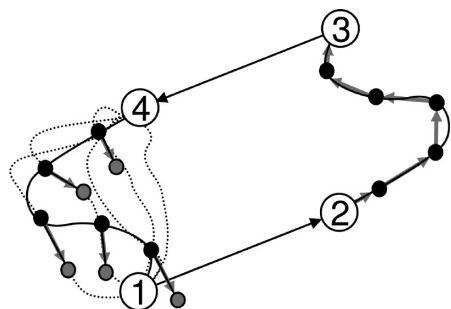


Figure 4. Illustration of forward and reverse transition probability calculations. Dotted lines represent putative trajectories that exhibit similar transitions in various stages of the trajectory.

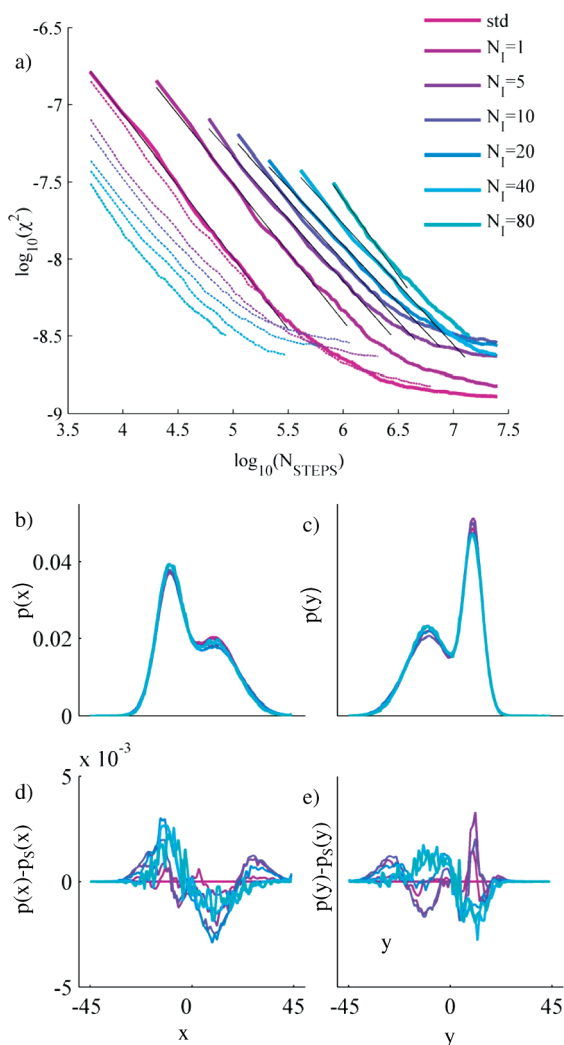


Figure 5. Convergence behavior of simple model system. Model system is landscape D. (a) log–log plot of convergence rates. Solid lines are shown for $N_{\text{STEPS}} = N_0(2N_i + 1)$, where N_0 is the number of outer steps, and dotted lines are plotted against $N_{\text{STEPS}} = N_0$. (b) PMFs along x coordinate in normalized probability units (c) PMFS along y (d, e) residual probability from standard simulation PMFs.

posh Pathway and Acceptance Criterion. Consider the possibility of the reverse pathway passing through a different region of coordinate space. This type of consideration has

been made in studying alternative transitions of trajectories in phase space.⁴⁷ To emphasize this difference in forward and reverse paths, we describe this as the posh pathway (port out, starboard home). Here, state 4 consists of the reverse trial perturbation, followed by a trajectory, which arrives at state 1. This can be described as

$$p_1\alpha_{12}\alpha_{23}\text{acc}_{13} = p_3\alpha_{34}\alpha_{41}\text{acc}_{31} \quad (12)$$

This condition for detailed balance as shown in Figure 2b was originally stated by Siepmann^{11,12} in the development of the adiabatic nuclear and electronic sampling MC method. In his derivation, the degrees of freedom for the initial perturbation were orthogonal to the annealed degrees of freedom, under the adiabatic assumption. Defining the topology of the states is sufficient to derive a slightly more general form, with the assignment of coordinate states added only for clarity. The 3 cases of interest are described in Figure 2b–d.

Coordinate state 4, the reflected trial coordinate \mathbf{q}_4 , can be defined in terms of the final trial coordinate and the uniform deviate $\xi^{(1,2)}$ in eq 6 used to generate the trial coordinate

$$\begin{aligned} \mathbf{q}_4 &= \mathbf{q}_3 - \xi^{(1,2)} \\ \mathbf{q}_4 &= \mathbf{q}_1 + \delta\mathbf{q}_{N_i} \end{aligned} \quad (13)$$

where

$$\delta\mathbf{q}_{N_i} = \mathbf{q}_3 - \mathbf{q}_2 \quad (14)$$

is the change in position from the trial position to the final trial position after a series of inner steps N_i .

In principle, the coordinate state \mathbf{q}_4 need not be defined in terms of the forward random deviate to preserve the property $\alpha_{12} = \alpha_{34}$. The choice of the reflected trial coordinate given by eq 13 has the convenient property of allowing the transition pathway ($4 \rightarrow 1$) to be defined using the information from the forward trajectory, forming a closed loop. The ($2 \rightarrow 3$) and ($4 \rightarrow 1$) selection probabilities are given by the following transition probabilities⁴⁶

$$\begin{aligned} \alpha_{23} &= T_{23}^{(N_i)} \\ \alpha_{41} &= T_{41}^{(N_i)} \end{aligned} \quad (15)$$

where $T_{23}^{N_i}$ and $T_{41}^{N_i}$ are the transition probabilities along the respective Markov chains of length N_i . Solving for the ratio of acceptance probabilities gives

$$\frac{\text{acc}_{13}}{\text{acc}_{31}} = \frac{p_3 T_{41}^{(N_i)}}{p_1 T_{23}^{(N_i)}} \quad (16)$$

Equation 16 obeys detailed balance to the extent that the transition probabilities are correctly estimated. The estimates of the transition probabilities are described in the next sections.

Forward Transition Probability. The forward transition probability is estimated using the record of the selection and acceptance probabilities of the single trajectory generated. While this is only an estimate of eq 8, this type of approach

is commonly used in other biased sampling methods^{6,13,14,47} in varying contexts. For the purposes of this work, it is assumed that the selection probability is uniform for all trial moves, so that only acceptance probabilities need to be computed. The acceptance probabilities presented are either the Barker or the Metropolis–Hastings acceptance probability.^{8,46} For the Barker acceptance probability, uniform variate is generated over the domain $[0, p(\mathbf{q}_{k-1}) + p(\mathbf{q}'_{k-1})]$. If the random number is greater than or equal to $p(\mathbf{q}'_{k-1})$, the trial move is accepted, and $\mathbf{q}_{k-1}^S = \mathbf{q}'_{k-1}$. Otherwise, the trial move is rejected, and $\mathbf{q}_{k-1}^S = \mathbf{q}_{k-1}$. If a trajectory in the inner loop is generated using the Barker acceptance criterion, the transition probability at the k th inner step is

$$t_{k-1,k}^{(2,3)} = \frac{p(\mathbf{q}_k = \mathbf{q}_{k-1}^S)}{p(\mathbf{q}_{k-1}) + p(\mathbf{q}'_{k-1})} \quad (17)$$

where the trial coordinate \mathbf{q}'_{k-1} is generated using a uniform deviate vector

$$\mathbf{q}'_{k-1} = \mathbf{q}_{k-1} + \xi_k^{(2,3)} \quad (18)$$

The selected coordinate \mathbf{q}_{k-1}^S is the coordinate resulting from the application of the acceptance probability (shown here as the Barker criterion), which is \mathbf{q}'_{k-1} if the trial is accepted and \mathbf{q}_{k-1} if the trial is rejected.

Rewriting the transitions of eq 17 with the Metropolis transition probability gives:

$$(t_{k-1,k}^{(2,3)})_{\text{Metropolis}} = \delta(\mathbf{q}_{k-1}^S - \mathbf{q}'_{k-1}) \text{acc}_{k-1,k}^{(2,3)} + \delta(\mathbf{q}_{k-1}^S - \mathbf{q}_{k-1}) (1 - \text{acc}_{k-1,k}^{(2,3)}) \quad (19)$$

Dellago uses a similar relation to define the Metropolis action.⁴⁸ The usual Metropolis criterion is applied to obtain the selected coordinate \mathbf{q}_{k-1}^S

$$(\text{acc}_{k-1,k}^{(2,3)})_{\text{Metropolis}} = \min(1, p(\mathbf{q}'_{k-1})/p(\mathbf{q}_{k-1})) \quad (20)$$

The notation and generation of the trial coordinate of eq 19 are identical to that of eq 17. In principle, any record of transitions can be used in lieu of either of the two expressions presented. In practice, however, the relations which obey detailed balance have given good results for the model systems studied. Using Metropolis transitions is the generally preferable approach for most types of simulations,⁵⁰ and this is true for the present case as well. For the remainder of the paper, however, the transition elements will be described using the Barker acceptance probability, since the notation is easier to read.

Reverse Transition Estimates. It should be noted that the estimate of the $(4 \rightarrow 1)$ transition probability represents a challenging class of problems whereby the end points are known, and the calculation of all paths connecting them as described by eq 8 cannot be enumerated practically. In general, the estimate of this probability is accomplished through importance sampling. Techniques such as Transition Path Sampling, have gained wide use in generating such estimates,^{49,51,52} whereby macrostate end points are defined, rather than fixed coordinate states. This is a conceptually helpful way to consider the transition probability estimates,

especially since the forward $(2 \rightarrow 3)$ trajectory can be thought of as an importance-sampled transition path.

Reverse Transition Pathway (True). Figure 3a describes the reverse transition path construction. Since it most closely mirrors the forward path, we label this as the true reverse transition pathway. To construct the reverse pathway, we first keep track of the displacement from initial trial state at step k

$$\delta \mathbf{q}_k = \mathbf{q}_k - \mathbf{q}_2 \quad (21)$$

The reverse path is defined using the displacements from the forward path. The true reverse transition pathway can be estimated using the forward trajectory information in the following way:

$$\mathbf{r}_k = \mathbf{q}_4 - \delta \mathbf{q}_k \quad (22)$$

where \mathbf{r}_k is the k th coordinate state in the reverse pathway as constructed in Figure 3a. The transition probability from state $k-1$ to k is recorded as

$$t_{k-1,k}^{(4,1)} = \frac{p(\mathbf{r}_{k-1}^S)}{p(\mathbf{r}_{k-1}) + p(\mathbf{r}'_{k-1})} \quad (23)$$

where $\mathbf{r}'_{k-1} = \mathbf{r}_{k-1} + \xi_k^{(4,1)}$ is constructed using the same perturbation strategy (using domain $\mathbf{d}^{(2,3)}$) as the forward case. The selection of coordinate \mathbf{r}_{k-1}^S follows the same procedure as in the forward pathway. It is important to notice that the selected coordinate $\mathbf{r}_{k-1}^S \neq \mathbf{r}_k$, since $p(\mathbf{r}_k)$ can easily become vanishingly small relative to $p(\mathbf{r}_{k-1})$ in the reverse trajectory for complex landscapes, whereas $p(\mathbf{r}_{k-1}^S)$ is selected according to its probability weight. The coordinate $\mathbf{r}_{k-1} = \mathbf{q}_4 - \delta \mathbf{q}_{k-1}$ forms the anchor point at each step along the reverse pathway, from which a trial coordinate is generated.

One practical consideration when using eqs 21–23 to generate the reverse trajectory is that it requires the storage of the complete forward trajectory $\delta \mathbf{q} = \{\delta \mathbf{q}_1, \delta \mathbf{q}_2, \dots, \delta \mathbf{q}_{N_1}\}$ prior to generating the reverse trajectory. Since these trajectories are generated using random deviates, this can be accomplished by maintaining a list of the random seeds, rather than an exhaustive storage of coordinate states. Even with this approach, however, it is can be cumbersome to reconstruct the entire reverse trajectory only after the entire forward trajectory has completed.

Concurrent Reverse Transition Pathway. To simplify the storage requirements, an alternative path for the reverse coordinate as a set of vectors $\mathbf{s} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{N_1}\}$ can be defined

$$\mathbf{s}_{N_1-k} = \mathbf{q}_1 + \delta \mathbf{q}_k \quad (24)$$

which provides the same overall connectivity between state 4 and state 1 as the set of vectors \mathbf{r} , given by eq 22 (see also Figure 3b). The difference is that the reverse pathway can be generated concurrently with the forward trajectory in this case. Using this alternative definition of the reverse coordinate path, the following transition can be defined:

$$t_{N_1-k, N_1-k+1}^{(4,1)}|_{\text{Concurrent}} = \frac{p(\mathbf{s}_{N_1-k}^S)}{p(\mathbf{s}_{N_1-k}) + p(\mathbf{s}_{N_1-k}')} \quad (25)$$

This pathway can be generated as the forward trajectory is being generated (since it does not require knowledge of the final trial state). The storage requirements are much smaller for this pathway, and it is slightly easier to implement. It also appears to provide slightly better precision of sampling. A discussion of the errors introduced by using the concurrent versus the reverse pathway is in Appendix B.

Qualitative Justification for the Reverse Pathway Estimation. The primary motivation for using accepted trial moves (using either method) at each step in the reverse pathway is to maintain numerical stability. Since the reverse pathway is constructed in a region of space that has not been located using importance sampling, as is the case with the forward pathway, reconstructing the path exactly will generate vanishingly small probabilities for even the simplest of landscapes, such as those studied for this work. The fact that the states are no longer connected contiguously may in fact improve the estimate, as a collection of transitions along the reverse pathway is estimating an ensemble of reverse pathways (see Figure 4). In fact, the key challenge to improving this sampling strategy is an understanding of how to efficiently and accurately estimate these transition probabilities.

Error Metric. As a general measure of the quality of the sampled distribution versus the true distribution, we can define the following ergodicity metric which is commonly used in assessing sampling quality:^{14,52,53}

$$\chi^2 = \frac{1}{N_B} \sum_{i=1}^{\sqrt{N_B}} \sum_{j=1}^{\sqrt{N_B}} (G(\mathbf{x}_i, \mathbf{y}_j) - H(\mathbf{x}_i, \mathbf{y}_j))^2 \quad (26)$$

where χ^2 is the mean squared error (MSE) over the course of the entire simulation, N_B is the number of bins, and $G(\mathbf{x}_i, \mathbf{y}_j)$ is the normalized distribution as described in Appendix A. $H(\mathbf{x}_i, \mathbf{y}_j)$ is the normalized histogram at the square bin centered about $(\mathbf{x}_i, \mathbf{y}_j)$. The length and width of the 2D histogram bin are $\delta b \times \delta b$.

Results and Discussion

We first explore the precision and accuracy of posh sampling protocols using simple two-dimensional model systems. We next present two initial molecular applications: several side chains in a flexible antibody binding sites and a phosphorylated tetrapeptide.

Precision of 2D Model System. A variety of parameter settings are studied in detail, and presented in Appendix B. Here, we study a single landscape (landscape D of Table 6 and Figure 10). This landscape was chosen as a metric for precision, as it has a saddle point with interesting asymmetry that can often reveal systematic biases in sampling. It is also a landscape that can be easily sampled with a standard method, providing a control.

Figure 5 shows the convergence behavior of χ^2 versus the number of steps on a log–log plot. Parameter settings for this system appear in Table 1. The dotted lines are plotted

against the number of outer steps (N_O) required to generate a desired χ^2 value, while the solid lines are plotted against total the number of steps, $2N_O(N_I + 1)$, which is more representative of the rate of convergence, as it accounts for the total number of energy evaluations. For the standard simulation, the total number of energy evaluations was 25M, and the remaining simulations were scaled to be of comparable length. The first thing to notice is that the dotted lines appear to converge faster than the standard simulation but in fact do not converge faster when accounting for the total number of energy evaluations. This is not surprising for this particular model system. A more careful discussion of regimes efficiency in sampling is discussed in the following section. The next thing to notice is that the (log–log) convergence rates are initially roughly linear with the same slope for all simulations. The standard simulation, which should rigorously reproduce the distribution, deviates from a steady rate of convergence at around $N_O = 10^5$, leveling off at roughly 1.3×10^{-9} , as it appears to be reaching a limit of precision for this metric.

The remaining simulations to not converge as rapidly, but do appear to converge to reasonable values of χ^2 , with a maximum value of 2.7×10^{-9} . The $N_I = 1$ simulation appears to be approaching the limit achieved by the standard simulation, while values of $N_I = 5, 10$, and 20 appear to converge to a higher value, suggesting a slight bias in the sampling for these regions. Values of $N_I = 40$ and 80 appear to be converging to a lower value. A plot of the pmfs of landscape D are given in Figures 5b and 5c. Close inspection reveals some systematic bias in the sampling that is consistent with the values reported in Figure 5a. The errors that are apparent appear to be within a reasonable margin, however, as the residuals here reach a maximum value of 3×10^{-3} . The interesting change in behavior of the precision as the number of inner steps increases is due to a simple effect. The initial bias introduced comes from the overestimate of the value of the reverse transition probability, which results in distributions that have slightly higher entropy than the true distribution. (This effect is seen in the biomolecular systems as well, and is discussed in later sections.) As the number of inner steps become very large, such that each trial move is essentially fully equilibrated, the anomalously high reverse transition probability is offset by the fact that the trial move is annealed to an equilibrium state, and thus will not have an artificially high entropy. While the error is detectable here, it appears to be sufficiently small to assert that this approach obeys detailed balance.

Efficiency Considerations. To compare efficiency of different sampling methods, we compute the MSE generated using the same number of energy evaluations throughout the entire trajectory, including those energy evaluations in the inner loop. For an inner loop of length N_I , the number of energy evaluations required is $2(N_I + 1)$ per outer step. We can define a simple improvement metric to compare the efficiency of posh MC to a “standard” MC trajectory (no inner steps) as

$$I = \frac{\chi_s^2}{\chi^2(N_I)} \quad (27)$$

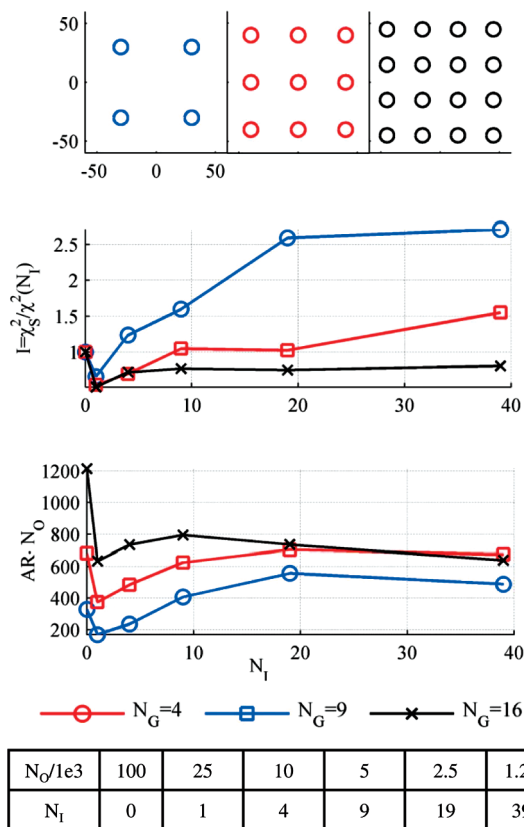


Figure 6. Efficiency of posh sampling depends on the sparsity of minima. (a) Schematic of landscapes sampled. Colors correspond to legend. Each circle represents a single Gaussian basin (Table 1, landscape G). (b) MSE versus number of inner steps. (c) Number of steps accepted. The number of outer steps in each case is adjusted such that the total number of energy evaluations is the same (100k) for each setting. See Table 6 for simulation settings.

where the MSE of a reference simulation, χ_s^2 , is computed using the same number of total energy evaluations as the posh MC simulation. In the posh MC simulations, the number of outer steps is thus $N_0 = 2N_{STD}/(N_i + 1)$, where N_{STD} is the number of energy evaluations in a standard Monte Carlo simulation. Thus, while a larger number of inner steps will almost always improve the acceptance ratio, it will not always improve the efficiency of sampling, because of the cost of generating the trial move.

As a first evaluation of efficiency, we use very simple energy landscapes, represented schematically in Figure 8a. The 3 landscapes each have evenly distributed arrangements of equivalent basins in the same domain. We denote the number of basins as N_G . The basins are identical, each with parameters given as landscape G in Table 6.

For $N_G = 4$ (the energy landscape least densely populated with energy minima), the improvement in sampling efficiency increases with increasing number of inner steps, N_i . For $N_i = 4$, the posh sampling is actually less efficient than that of the standard simulation. There are simply not enough inner step moves to reliably locate nearby basins because the large initial trial move in the sparse space will rarely land in a favorable region of space. At $N_i = 9$, however, the annealing steps begin to locate basins and a noticeable improvement

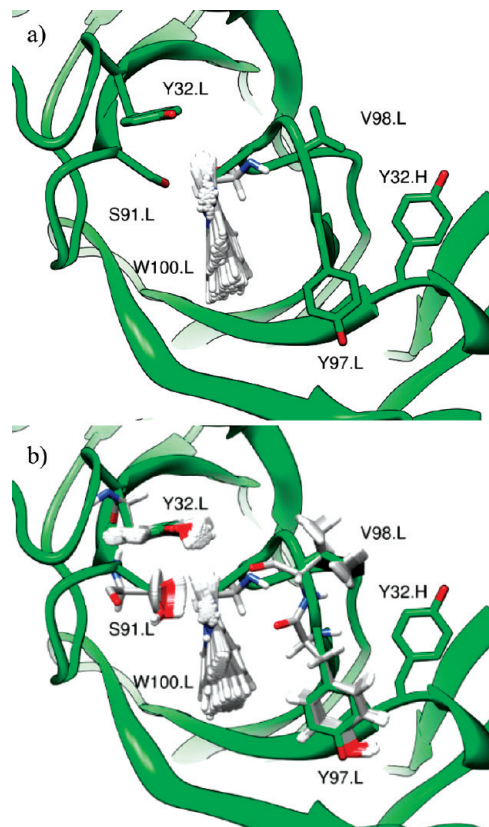


Figure 7. (a) Side chains of 1dba used for trajectories.. All remaining side chains and backbone configurations are held at the native state. (a) For single side chain case in Table 2, only TRP 100.L (in green) is sampled. (b) Multiple side chains sampled as reported in Table 3.

is observed, which tapers off at $N_i = 39$, since the cost of a trial move is 80 times that of a standard simulation. The total number of accepted outer steps is a good measure of the effectiveness of the sampling, and is shown in Figure 8c. The slight improvement in error at $N_i = 39$ occurs even though there are fewer newly accepted configurations because the trial moves are less correlated, which will also serve to improve the statistics.

For $N_G = 16$ (the energy landscape most densely populated with energy minima), the standard sampling approach is more effective. Put simply, the basins are sufficiently densely packed that a posh scheme is no longer needed. Note also that the number of newly accepted outer steps is much higher for the standard setting ($N_i = 0$). It is clear from these initial results that, as expected, the posh MC scheme will be most effective for energy landscapes with relatively few low energy basins separated by large energy barriers, as is frequently the case in macromolecular applications.

These results also suggest that the initial perturbation should be large relative to the annealing step, and large enough to cross barriers, while the size of the annealing step is chosen to give good acceptance in sampling the local basin. It is usually straightforward to estimate the size of the annealing step. Often, the length scales emerge quite naturally from knowledge of the system. For example, a typical range for a single dihedral perturbation in a proteins or small molecules is typically less than $2\pi/3$, which is roughly the width of a single χ well.

Table 2. Energy Statistics of Single Side Chain Trajectory^a

	$N_I = 1$			$N_I = 5$	
	std	50% posh	100% posh	50%	100% posh
$N_O(\text{total}) / 10^4$	100	50	35	20	15
$\beta(\langle E \rangle - \langle E \rangle_0)$	0.07(0.06)	0.06(0.06)	0.09(0.12)	0.07(0.11)	0.00(0.11)
	-0.02(0.06)	0.00(0.09)	0.01(0.14)	0.05(0.10)	0.00(0.08)
σ	1.05	1.05	1.10	1.10	1.00
	1.05	1.08	1.20	1.12	1.18
$(\langle E \rangle - \langle E \rangle_0) / \sigma_0$	-0.01	0.00	-0.01	-0.04	0.00
	-0.01	0.00	0.01	-0.04	0.00
	$N_I = 10$			$N_I = 20$	
		50%	100% posh	50%	100% posh
$N_O(\text{total}) / 10^4$		15	8	6	4
$\beta(\langle E \rangle - \langle E \rangle_0)$		-0.01(0.12)	-0.01(0.12)	0.15(0.23)	0.17(0.33)
		-0.10(0.08)	0.12(0.08)	0.43(0.17)	0.46(0.18)
σ		1.11	1.03	1.31	1.39
		1.13	1.34	1.69	1.78
$(\langle E \rangle - \langle E \rangle_0) / \sigma_0$		-0.01	-0.01	0.15	0.17
		0.08	0.10	0.37	0.39

^a TRP 100.L of protein 1dba is simulated at 350K, as described in text (see Figure 7a). Values in gray are from a simulation using the full reverse transition calculation, and values in white are from a simulation using the stationary transition approximation (eq 28). Statistics are computed using 10 trajectories started from an equilibrated structure as an initial condition. Values in parentheses are the variance of the mean energy over the 10 trajectories, σ is the time averaged standard deviation of the energy trajectory. $\langle E \rangle_0$ and σ_0 are chosen as a reference state from the standard simulations at $N_I = 5$ (highlighted in yellow and rose). The total number of outer steps N_O used are summed over each of the 10 trajectories of equal length.

Table 3. Energy Statistics of Multiple Side Chain Trajectory at 350 K^a

a)		$N_I = 1$		$N_I = 5$	
		50% posh	100%	50% posh	100%
	$N_O(\text{total}) / 10^4$	125	90	50	20
	$\beta(\langle E \rangle - \langle E \rangle_0)$	0.58(0.69)	0.43(0.56)	0.14(0.24)	0.00(0.43)
	σ	2.40	2.27	2.34	2.24
	$(\langle E \rangle - \langle E \rangle_0)/\sigma_0$	0.26	0.19	0.06	0.00
		$N_I = 10$		$N_I = 20$	
		50% posh	100%	50% posh	100%
	$N_O(\text{total})/ 10^4$	20	15	10	8
	$\beta(\langle E \rangle - \langle E \rangle_0)$	0.01(0.29)	0.36(0.33)	0.85(0.66)	0.66(0.27)
	σ	2.28	2.47	2.57	2.76
	$(\langle E \rangle - \langle E \rangle_0)/\sigma_0$	0.00	0.16	0.38	0.30

b)

Average Values for Individual Standard Trajectories						
$(\langle E \rangle - \langle E \rangle_0)/\sigma_0$	1.03	0.13	0.67	0.40	0.62	
	0.71	0.36	0.18	1.03	0.67	

^a Side chains of 1dba as shown in Figure 7b. (a) Statistics are reported as in Table 2. All trajectories are generated with the stationary transition assumption. (b) Average values of individual standard trajectories are reported for reference.

For the cases studied here, the acceptance ratios for the standard protocols for $N_G = 4$ and 9 (those showing improvement) were less than 1%. The use of posh sampling can be thought of as a method for refining trial moves that would otherwise be inefficient. For a given landscape and perturbation protocol, there exists an optimal N_I , which decreases to 1 as the landscapes become less sparse.

Empirical Corrections: Stationary Transitions. A practical consequence of the reverse pathway estimation in complex systems is that the reverse transitions corresponding to a rejected trial have been observed to result in an error in the estimate the ratio of forward and reverse transitions. In particular, the reverse transition calculations

are most susceptible to error, as the construction of the reverse pathway will often pass through sterically hindered portions of configuration space, resulting in anomalously high energies. This region can often lead to numerical error in the estimate of the transition probabilities. In these cases, it has proven to be useful to introduce the following empirical correction to the reverse transition probability:

$$t_{N-k,N-k+1}^{(4,1)} = \delta(\mathbf{q}_{N-k}^S - \mathbf{q}_{N-k+1}^T) \text{acc}_{N-k,N-k+1}^{(4,1)} + \delta(\mathbf{q}_{N-k}^S - \mathbf{q}_{N-k}^S) t_{N-k,N-k+1}^{(2,3)} \quad (28)$$

where $\text{acc}_{N-k,N-k+1}^{(4,1)}$ is the Metropolis acceptance probability for the reverse transition. This relation simply uses the

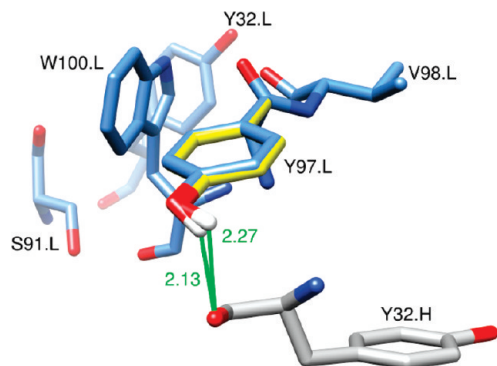


Figure 8. Anecdotal case of a large change in energy resulting from a small change in position. This change in position results in a $6.0k_B T$ change in energy during the course of a simulation.

forward transition probabilities in the reverse trial move when the forward trial is rejected, which prevents the overestimate of the ratio $T_{41}^{N_1}/T_{23}^{N_1}$, especially in longer inner loop settings. The motivation for the use of this correction comes from the notion of an ideal transition. If we consider the ideal forward transition move, it would consist of a series of purely downhill moves. Likewise, the reverse transition pathway would also consist of purely downhill moves. For these ideal transitions, the estimate of the ratio $T_{41}^{N_1}/T_{23}^{N_1}$ would be unity, if Metropolis transitions are used. The practical effect of estimating these ratios is to generate numbers that are slight corrections to this ratio. The introduction of the empirical correction helps to facilitate this assumption, and improved the statistics such that acceptable precision could be observed in the $N_1 = 20$ range. We have nominally assigned a tolerance of $\pm 0.1\sigma_0$ in the estimate of the average energy as an indicator for acceptable precision, where σ_0 is the standard deviation of the energy of a trajectory generated with a standard protocol. Table 2 compares simulations generated with and without the stationary transition assumption, and is discussed in the following section.

Molecular Applications

While a detailed study of model systems is worthwhile to understand the elements of the sampling, the ultimate goal is to apply this sampling to complex biomolecular systems. Here, we present two initial applications, specifically using posh Monte Carlo to generate ensembles for (a) side chains in an antibody binding pocket and (b) a phosphorylated tetrapeptide.

Side Chain Sampling: Precision Studies. Previous work has established the utility of side chain Monte Carlo sampling approaches to study binding pocket dynamics.⁴⁵ The published method perturbs a subset of protein side chain torsions while keeping the remaining backbone and side chain coordinates fixed. Here, we perform a similar simulation to study the effectiveness of the current algorithm in such a regime. The protein studied here is the Fab fragment of antibody protein 1dba,^{55,56} which was also considered in the previous study.⁴⁵

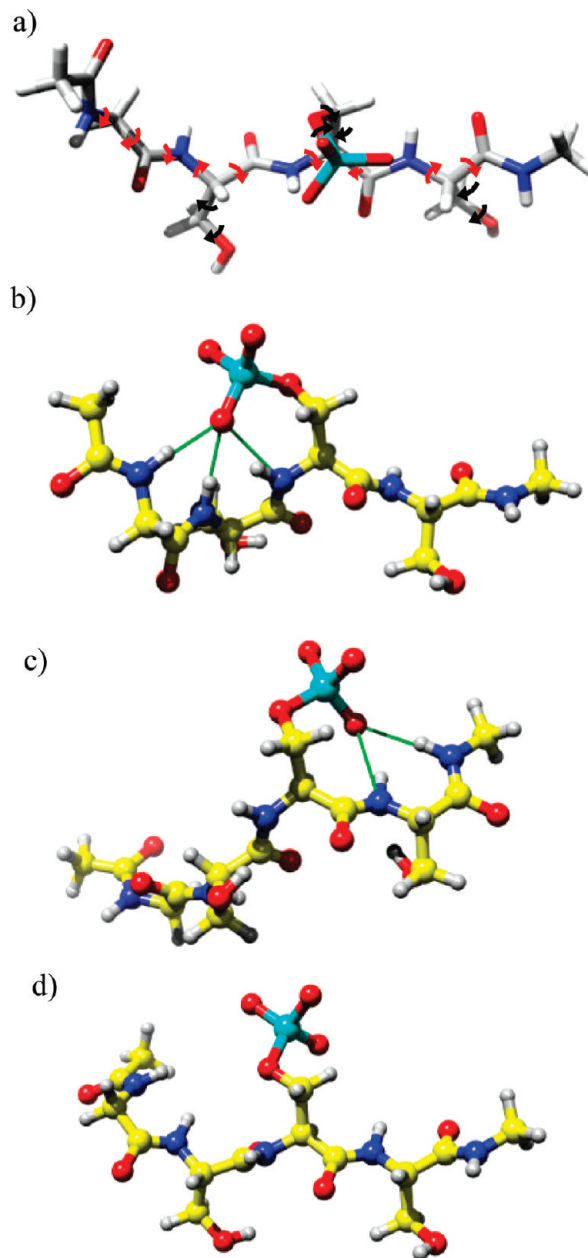


Figure 9. Phosphopeptide Ace-Gly-Ser-pSer-Ser-Nma studied. a) Dihedral angles sampled. Red arrows indicate all backbone dihedral angles allowed to fluctuate, while black arrows indicate the side chain dihedrals allowed to fluctuate. All other degrees of freedom are held fixed. (b, c, d) Typical hydrogen bonding configurations: (b) N-terminal hydrogen bonding network, (c) C-terminal hydrogen bonding network, and (d) solvated phosphate configurations.

To account for implicit solvation, the SGB model⁵⁷ is used, but for these cases, the Born Radii are held fixed at that which was calculated for the initial conformation. This constraint avoids introducing the additional bookkeeping of the Born Radii updates and is justifiable in this case because the solvent exposure of the side chains does not change dramatically as the side chains explore different conformations. This bookkeeping, however, will be reintroduced for the phosphopeptide case, where solvent exposure of various portions of the peptide change dramatically among sampled conformations, with the appropriate modified acceptance

Table 4. Energy Statistics of Phosphopeptide at 300 K^a

	$N_I = 1$			$N_I = 5$		
	std	50% posh	100%	std	50% posh	100%
$N_O/10^4$	1000	1000	1000	400	400	400
$\beta(\langle E \rangle - \langle E \rangle_0)$	0.51(0.70)	-0.15(0.60)	0.24(0.59)	0.00(0.41)	0.12(0.79)	0.26(0.67)
$\beta\sigma$	3.65	3.55	3.55	3.66	3.61	3.61
$(\langle E \rangle - \langle E \rangle_0)/\sigma_0$	0.14	-0.04	0.06	0.00	0.03	0.07
	$N_I = 10$			$N_I = 20$		
	std	50% posh	100%	std	50% posh	100%
$N_O/10^4$	200	200	200	100	100	100
$\beta(\langle E \rangle - \langle E \rangle_0)$	-0.64(0.38)	-0.58(0.40)	0.11(0.97)	-0.10(0.56)	-0.05(0.83)	1.62(0.64)
$\beta\sigma$	3.61	3.64	3.88	3.66	3.77	4.26
$(\langle E \rangle - \langle E \rangle_0)/\sigma_0$	-0.17	-0.16	0.03	-0.03	-0.01	0.44

^a Statistics are reported as described in Tables 2 and 3. All trajectories are generated with the stationary transition assumption.

Table 5. J Coupling Data^a

	exptl	Wong et al. simulation	$N_I = 1$			$N_I = 5$		
			std	posh	50%	std	posh	50%
Gly1	5.63(0.12)	4.12(0.03)	3.24(0.78)	3.01(0.22)	3.30(0.35)	3.25(0.19)	3.19(0.20)	3.39(0.25)
Ser2	6.65(0.12)	6.62(0.11)	7.63(0.21)	7.76(0.16)	7.68(0.20)	7.70(0.12)	7.57(0.21)	7.65(0.09)
PSer3	5.48(0.12)	5.73(0.10)	5.62(0.34)	5.62(0.45)	5.46(0.39)	5.64(0.23)	5.76(0.48)	5.60(0.25)
Ser4	6.93(0.12)	7.37(0.03)	7.46(0.06)	7.43(0.05)	7.42(0.05)	7.42(0.05)	7.45(0.07)	7.42(0.03)
			$N_I = 10$			$N_I = 20$		
			std	posh	50%	std	posh	50%
Gly1			3.11(0.15)	3.50(0.44)	3.19(0.19)	3.32(0.19)	3.88(0.23)	3.18(0.12)
Ser2			7.67(0.13)	7.53(0.14)	7.68(0.10)	7.71(0.10)	7.40(0.10)	7.67(0.12)
pSer3			5.58(0.24)	5.54(0.40)	5.56(0.35)	5.81(0.30)	5.76(0.27)	5.82(0.34)
Ser4			7.43(0.03)	7.42(0.07)	7.43(0.03)	7.44(0.05)	7.42(0.07)	7.44(0.07)

^a Data set is the same as that used for Table 4.

probability. The nonpolar term of the solvation model is from the Levy and Gallicchio⁵⁸ model, and the force field is OPLS-AA 2005.^{59,60} Figure 9a shows the set of side chains sampled, which were chosen because they form direct interactions with the ligand in the holo structure and because they occupied somewhat different rotameric conformations in holo and apo structures of the antibody, suggesting that they may be flexible. The remainder of the protein is held at its native state. The initial perturbation consists of selecting a side chain randomly, assigning a uniform deviate of the torsion angles over the domain $[-\pi/2, \pi/2]$. The inner step trial moves follow the same procedure, but over the smaller domain $[-\pi/36, \pi/36]$. The reverse trajectory is generated using the concurrent pathway approach, where each reverse trial is generated using eqs 24 and 25.

To demonstrate that the sampling obeys detailed balance, a simple trajectory using only the Trp100 side chain is first generated (see Figure 7a). Table 2 provides energy statistics for a range of inner loop settings, $N_I = 1, 5, 10$, and 20. For each parameter setting, 10 trajectories were generated at 350 K. A standard trajectory was generated, where each trial move follows the initial perturbation, or (1→2) recipe. Both 100% posh trajectories and 50% posh trajectories were generated. For this series, we also studied the effect of introducing the stationary transition assumption (eq 28).

The metric used to evaluate precision is the difference in mean energies divided by the standard deviation from a

nominal baseline trajectory. The trajectory generated using 100% posh ($N_I = 5$) was used as the baseline. A cutoff of ~ 0.1 to 0.15 was assigned to determine if the trajectories were converging to the correct expectation value. Any value larger than this is deemed to be indicative of a systematic bias in the sampling. Table 2 shows that nearly all parameter settings presented here satisfy this requirement, and that the system clearly obeys detailed balance for all settings up to $N_I = 10$. Clearly, the 50% posh sampling provides higher precision than 100% posh sampling, which is to be expected. The upper bound for maintaining reasonable precision appears to be roughly around $N_I = 20$. This upper bound appears to be consistent across all systems studied in this work, with the exception of the model systems, which do not show this limitation as prominently. We believe that the limitation appears in the biomolecular systems because the reverse trajectories often pass through extremely high energy regions of space. Statistics in this region are more likely to be error prone, whereas the model systems have better defined values throughout the domain of interest.

The use of eq 28 for the reverse transitions appears to reduce the error in the higher N_I regimes and not to affect the statistics in any noticeable way in the smaller regimes. Based on the observed success of this approximation, it is used for the remaining biomolecular systems studied. We believe that the origin of the error is the systematic overestimation of the reverse transition probability by

generating a trial move, as described in eq 25 (or eq 23 had the true reverse pathway been calculated). This error will become more apparent in the longer inner loop settings and thus will benefit more from the stationary transition approximation.

Side Chain Sampling: Efficiency. With the demonstration of precision established by the single side chain data of in Table 2, we study a slightly larger system to demonstrate the efficiency of the algorithm in a more complex system. The side chains sampled are shown in both yellow and green in Figure 7b, and all other settings are the same as in the single side chain case discussed above. Table 3 shows that most of the normalized differences in mean energies fall within an acceptable region, with the similar exception of $N_1 = 20$, as above.

The most striking result is in the anomalously high value (large error) of the standard trajectory (0.57), that is, not using posh. Notice also that the error in the measurement, 0.70, is also quite large relative to the other trajectories. Table 3b shows that of the 10 standard trajectories generated, only 2 are within 0.2 of the baseline. This does not indicate incorrect sampling of the posh trajectories. Rather, it indicates highly nonconverged sampling of the standard trajectories. The size of the initial trial move ($1 \rightarrow 2$) in both the posh and standard cases determines the accessibility of states, such that certain local minima would be inaccessible if a smaller trial move were used in either the standard or the posh case. The fair comparison to be made (as was done in the model system study of Figure 6), therefore, is a standard trajectory with a trial move comparable to the initial trial move of the posh trajectory.

There is, however, a similar, quasiergodic limitation that occurs if *only* large trial moves are used. The much finer detail of the landscape is not as well sampled, and yet the data set will appear to be converged. Consider an anecdotal case from a trajectory presented in Figure 8. This particular side chain displacement, consisting primarily of the movement of the polar hydrogen, results in a $6k_B T$ change in energy. While many of the hydrophobic fluctuations are generally of longer range in nature, the polar interactions tend to vary quite rapidly with position when in close proximity. This effect has been described in protein folding as the champagne landscape problem, and points to the necessity of combining large and small trial move sets. Here, we see that posh sampling is able to capture this effect, giving it a substantial advantage over the standard approach.

Phosphopeptide Sampling. As a second application, we applied posh to a phosphopeptide: Ace-Gly-Ser-pSer-Ser-Nma, which has been studied previously experimentally and computationally.^{61,62} This system is small but challenging for sampling. The phosphate group forms hydrogen bonds with different combinations of backbone amide NH groups, and transitions between these conformations are relatively slow, but accessible with standard (single replica) MD simulations.

For this study, only dihedral angles are allowed to fluctuate, excluding the peptide ω angles and capping methyl group torsions. Figure 9 shows all of the dihedral angles that are sampled. The SGB⁵⁷ implicit solvent model was also

used here, with an external dielectric of 80 and an internal dielectric of 2 (to crudely account for electronic polarizability). The phosphate partial charges are those used by Groban et al.,⁶³ based on a study conducted by Wong et al.⁶¹ An additional protocol, developed in previous work⁴¹ is added to optimize sampling, and is described below.

Implementation Details: Modular Perturbations. Perturbations are classified as either backbone moves or side chain moves. For the peptide, a backbone move consists of randomly selecting a single ϕ or Ψ dihedral angle, and assigning a uniform variate perturbation to that angle over a defined domain. For the initial trial perturbation, the domain is $[-\pi, \pi]$, while the inner step trial moves are over a smaller perturbation domain $[-\pi/10, \pi/10]$.

The side chain perturbation recipe here is slightly different than that for the antibody case discussed above. It follows the same prescription regardless of whether it is considered to be an initial trial or an inner step move. Here, each uniform variate is over the domain of either $[-\pi, \pi]$ or $[-\pi/20, \pi/20]$, determined randomly, with the selection probability of either domain given equal probability. This is the same protocol developed in previous work which studies only side chain fluctuations.⁴¹ The polar hydrogens are sampled over the domain $[-\pi, \pi]$ at every step for which that residue is selected.

At both the initial trial step generation and the inner step moves, either a backbone move or side chain move is selected with equal probability. This corresponds to the diagonal pathway, since all degrees of freedom are allowed to fluctuate in the ($1 \rightarrow 2$), as well as the ($2 \rightarrow 3$) transitions. To generate the ($4 \rightarrow 1$) transition, the concurrent reverse pathway is used (eqs 20 and 21). At each forward transition, a reverse trial move is generated using the same type of perturbation for the forward perturbation (side chain or backbone). Both forward and reverse trajectories use Metropolis transitions.

Multiple Time Step Monte Carlo Acceptance Probability. For this system, since we wish to compare to experimental NMR data, we introduce a Born radii correction term. To improve computational efficiency, some of the energy parameters (such as the Born radii) \mathbf{P}_L are held at the latent state of the original coordinate, giving the parameter set $\mathbf{P}_L(\mathbf{r}_I)$. Most notably, the long-range interactions and Born radii are held fixed during the inner loop sampling. At the end of each cycle of posh sampling, the resulting coordinate state is taken to be a trial move and subjected to the Multiple Time-Step Monte Carlo⁶⁴ acceptance criterion

$$\text{acc}_{I,F} = \min \left[1, \frac{p(\mathbf{r}_F | \mathbf{P}_L(\mathbf{r}_F))}{p(\mathbf{r}_F | \mathbf{P}_L(\mathbf{r}_I))} \right] \quad (29)$$

The probability $p(\mathbf{r}_F | \mathbf{P}_L(\mathbf{r}_I))$ is the Boltzmann factor of the energy evaluated at current coordinate state F with latent parameters from coordinate state I . For all cases studied here, a single posh cycle is followed by a parameter update. This criterion rigorously corrects for the error introduced by holding the latent parameters at a fixed state, such that the resulting distributions are equivalent to that of a sampling

protocol which updated these parameters at every state. This method has been developed and tested in a previous work.⁴¹ Here we refer to state *I* as the initial coordinate state and state *F* is the final coordinate. As in previous work,⁴¹ there is an optimal range of “inner” steps for this sampling protocol as well. Our interest in the present work, however, is only to maintain the rigorous description of the solvation parameters. As such, the number of inner steps before updating the latent parameters is N_I , where N_I is either the inner step setting during a posh cycle, or it is the length of a standard trajectory between latent parameter updates.

To generate comparable trajectories without posh sampling, a set of standard trajectories was also generated, which maintained the MTSMC sampling. For these trajectories, the same number of inner steps are used in between latent parameter updates. The trial moves for these are those which would be used for the (1 \rightarrow 2) moves in the equivalent posh scheme.

Validation Results: Energy Statistics. To generate equilibrium data for validation and comparison to experiments, 10 trajectories were generated for each parameter setting, starting each from a different equilibrated initial condition. The parameter settings were inner steps settings of 1, 5, 10, and 20, with outer steps set to 1×10^6 , 4×10^5 , 2×10^5 , and 1×10^5 , respectively (to maintain the same number of energy evaluations among the simulations). Three sampling protocols were explored: (1) a standard procedure, which uses no posh sampling but uses the MTSMC procedure; (2) a 100% posh procedure, where a single posh cycle using the inner step setting is used, followed by the MTSMC procedure; and (3) a 50% mixture of these two procedures, where a random selection of either of the two protocols is made at every outer step. It should be noted that the MTSMC protocol is used here in all cases.

Table 4 shows energy statistics for a range of inner step settings. As in the previous cases, quantitative agreement can be obtained for settings of up to $N_I = 10$, with systematic deviations appearing at $N_I = 20$, suggesting an upper bound for precision for this particular system. The reasons for this deviation are largely caused by the rugged energy landscape, as discussed above for the antibody test case.

It should be noted that the (4 \rightarrow 1) transitions for the peptide are likely to generate far more pathological energy configurations because of the sampling of the backbone degrees of freedom. This is apparent in a slight increase in deviation at $N_I = 20$ (0.44 versus ~ 0.2 to ~ 0.3 for the side chain cases).

Comparison to Experiment: Physical Observables. *J* couplings were computed using the Karplus equation⁶⁴

$$\langle^3J\rangle = \frac{1}{L} \sum_{i=1}^L (A \cos^2 \theta_i + B \cos \theta_i + C) \quad (30)$$

where $\langle J \rangle$ is the average NMR *J* coupling value, *L* is the length of the simulation, and θ_i is the dihedral angle of the *i*th snapshot of the H–N–C $_{\alpha}$ –H bond. We compare the computed values to experiments^{61,62} and previous simulations reported by Wong and Jacobson using molecular dynamics.⁶¹ Of all of the simulations, we expect the standard (MTSMC)

values to be most reliable because it contains data from all inner steps settings. We find good agreement between the standard simulation (with MTSMC) and experiment, most notably in the pSer coupling, which is the most sensitive to the phosphate interactions. We obtain poor agreement with experiment for the Gly coupling, but this anomaly is in part because *J* couplings are not well-defined with regard to glycines.

The agreement between our results and the previous results obtained using the molecular dynamics simulations⁶² is good, considering that a different forcefield (AMBER) and implicit solvation model were used. We obtain good agreement of the *J* couplings with the standard (MTSMC) simulation as a control, and notice also that the observables appear to be reasonable across a broader range of inner step settings than that observed with the control. This is because the comparison of the standard protocol to the posh protocol is, in general, a more sensitive measure of precision than the experimental comparisons. The experimental measures, however, help to evaluate the complete design, which includes the forcefield, the implicit solvation model, and the geometric assumptions used, and establish a precedent for the overall accuracy of the approach. It should be noted here that satisfactory sampling is easily obtained with a standard Monte Carlo protocol, such that we do not see a marked improvement in sampling efficiency here with the introduction of the posh sampling approach. This is not entirely surprising, considering that previous workers were able to sample this molecule with conventional dynamical methods. While the phosphopeptide makes an excellent system for evaluating precision, the side chain trajectories are more indicative of the capacity to improve the quality of sampling in more complex systems.

Conclusions and Future Directions

We have presented a sampling protocol designed to allow efficient sampling of sparsely distributed basins, such as those that are encountered in complex biomolecular energy landscapes. The protocol obeys detailed balance to the extent that the transition probabilities are correctly estimated. Three variants were presented in terms of pathways: adiabatic, hybrid, and diagonal. All of these approaches could conceivably be useful, depending of the system of interest. Each pathway provides adequate performance, but the diagonal and hybrid approaches are more robust for longer inner loop protocols. Two reverse pathway constructions were presented, and appear to be similar in terms of precision, with a slight preference given to the concurrent reverse pathway due to the ease of implementation. A study of efficiency in simple model systems suggests that the algorithm will be most efficient when considering very sparsely distributed local minima.

Sampling of densely packed side chains appears to benefit greatly from the introduction of posh sampling because of the nature of the energy landscape, characterized by sparsely distributed minima with steep basins. An implementation in a more realistic small peptide system shows that comparison to experimental results gives reasonable agreement. The performance of the approach in complex systems is largely

limited by the nature of the reverse pathway construction, which is much more numerically challenging in a system with steric occlusions. It is also a costly calculation, which will place an upper bound on efficiency. Future work in this area should include methods for estimating the reverse transitions without the need to compute a reverse trajectory.

Acknowledgment. We are grateful to Chris McClendon and Pradipta Bhandyopadhyay for many helpful conversations and testing. This work was supported by NSF Grant MCB-0346399, and in part by NIH Grant R01-GM081710. M.P.J. is a consultant to Schrödinger LLC.

Appendix A: Model System

Initial evaluation of the posh sampling is performed using a simple 2 dimensional model. Consider a 2 dimensional Gaussian distribution

$$g_i(x, y) = w_i \exp(-u_i(x, y)) \quad (31)$$

where x and y are scalar variables, and the index i refers to the i th parameter set defining the distribution. Models of this sort have been used to analyze biomolecular systems because of the simple evaluation of the configuration integral and other observables.^{65,66} The potential is simply a harmonic potential^{68,69}

$$u_i(x, y) = \gamma_i [x - (x_0)_i - (y_0)_i \mathbf{R}_{\theta_i}^T \begin{bmatrix} (\sigma_1)_i^{-2} & 0 \\ 0 & (\sigma_2)_i^{-2} \end{bmatrix} \mathbf{R}_{\theta_i} \begin{bmatrix} x - (x_0)_i \\ y - (y_0)_i \end{bmatrix}] \quad (32)$$

where γ is a force constant, x_0 and y_0 are the coordinates of the minimum, and the σ_1 and σ_2 are the standard deviations along the principal axes of the distribution. \mathbf{R}_{θ} is a matrix which rotates the principal axes of the distribution by an angle θ

$$\mathbf{R}_{\theta_i} = \begin{bmatrix} \cos \theta_i & \sin \theta_i \\ -\sin \theta_i & \cos \theta_i \end{bmatrix} \quad (33)$$

The entire set of parameters defining a single Gaussian term, then, are given as

$$\mathbf{p} = (x_0 \ y_0 \ \sigma_1 \ \sigma_2 \ \theta \ w \ \gamma) \quad (34)$$

One advantage to using potentials of this type is that the free energy of each basin can be computed analytically. Also, potentials of mean force can be analytically computed. Potentials can be superimposed to generate nonlinear behavior:

$$\begin{aligned} G(x, y | \{\mathbf{p}\}_{N_F}) &= \sum_{i=1}^{N_F} g_i(x, y) \\ U(x, y | \{\mathbf{p}\}_{N_F}) &= -\ln(G(x, y | \{\mathbf{p}\}_{N_F})) \end{aligned} \quad (35)$$

For all cases presented, the temperature is unity.

Appendix B: Model System Studies

Model System Properties. Since the algorithm allows for the possibility of sampling two subspaces hierarchically, a

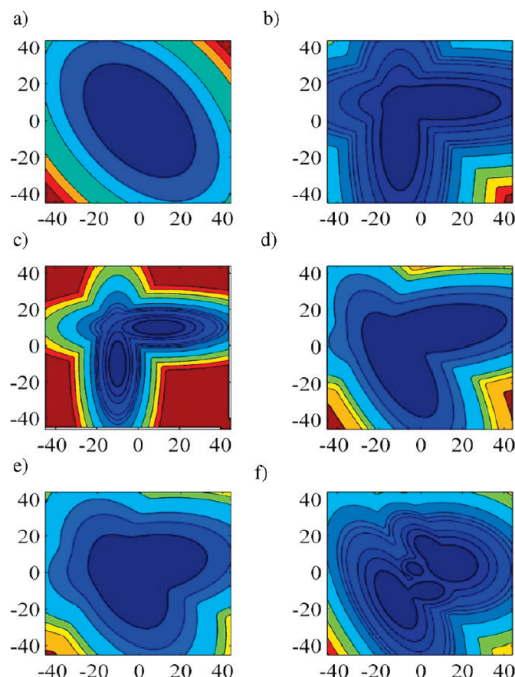


Figure 10. Landscapes studied. 2D contours of landscapes as described in text with parameters given in Table 4. For all plots, x is along horizontal axis. (a) Single basin. (b) Orthogonal disjoint basins. (c) Orthogonal accessible basins. (d) Acute basins. (e) Acute basins. (f) Multiple basins.

two-dimensional surface becomes a convenient model system for studying the precision of sampling (Figure 10). These model systems are superpositions of 2D Gaussian distributions, whose partition functions are known analytically. Table 6 lists the parameters used. These landscapes were designed to have features that are thought to influence the sampling. One of the features is the symmetry/asymmetry of the surface, which can play a role in the precision due to the symmetry in the reverse path construction. Another feature of interest is the distribution of local minima, either as quasi-ergodic or accessible basins. Landscapes with anharmonic saddle points and complex features were also included to mimic rough landscapes in a way that is intended to make errors in sampling more easily detected.

Choice of Perturbation Scheme. In applying the posh sampling to a system of interest, many considerations drive the choice of subspace partitions, but the choice will be limited to the types listed in Figure 2. It is worthwhile to notice the design considerations involved in each of the perturbation types. To explore this issue, we use a variety of simple two-dimensional energy landscapes, shown in Figure 5.

For Figure 11, the MSE of the adiabatic pathway with $N_1 = 1$ is defined as χ_0^2 and calculated for each of the pathways. The MSE, $\chi^2(N_1)$, for each value of N_1 is computed and the normalized value $\chi^2(N_1)/\chi_0^2$ is a measure of the relative error. Table 1 details the simulation conditions of the landscapes studied, as shown in Figure 10. The diagonal sampling scheme emerges as the most accurate of approaches across all landscapes shown here. This is expected because the sampling is unconstrained in both the perturbation and annealing steps. The hybrid approach also performs ex-

Table 6. Parameters of Landscapes as Described in Appendix A^a

landscape	x_0	y_0	σ_1	σ_2	w	θ/π	γ
A	0	0	3	5	1	0.2	0.2
B	-10	-10	4	12	1	0	9
	10	10	12	4	1	0	9
C	-10	-10	4	12	1	0	1.8
	10	10	12	4	1	0	1.8
D	-10	10	4	10	1	0.15	1
	10	10	4	10	1	0.55	1
E	-5	5	5	9	1	0.15	1
	5	5	5	9	1	0.55	1
F	-12	-15	6	14	3	0.15	6
	5	12	5	10	4	0.25	9
	3	-10	5	3	2	0	3
	15	5	10	8	3	-0.1	6
	-3	2	2	3	1	0.3	3
G	XX	XX	3	3	1	0.0	9

^a Letters correspond to those plotted in Figure 6. Landscape G describes a single basin of multiple basin landscapes which is plotted in Figure 5a.

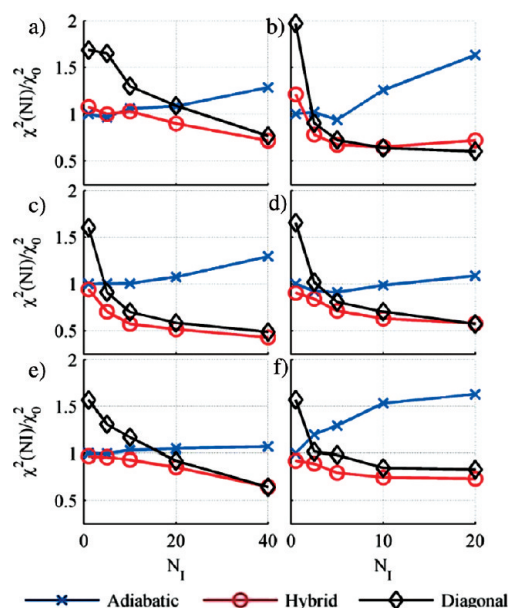


Figure 11. Comparison of errors for different pathway types. Errors are normalized to the Adiabatic case with $N_I = 1$ for each landscape. Simulations were run with $N_0 = 40$ k steps. See Tables 1 and 4 for additional settings.

tremely well, suggesting that allowing all degrees of freedom to fluctuate in the annealing steps is a key to improving precision. The hybrid pathway often outperforms the diagonal pathway for smaller inner step settings, but this may be the result of the the initial perturbation of the hybrid case being effectively much smaller, since it only spans 1 dimension. At larger values of N_I , the hybrid and diagonal pathways perform nearly identically. While the adiabatic pathway performs very well for smaller values of N_I , it exhibits systematic bias as the number of inner steps is increased, due to the constrained sampling of the annealed coordinate. For shorter trajectories ($N_I < 10$ for the cases here), the difference in error is negligible for all pathways. Any of these approaches, for relatively short trajectories, could be applied to more complicated systems with confidence, but in cases where longer inner loop trajectories are needed, the hybrid and diagonal pathways are more robust.

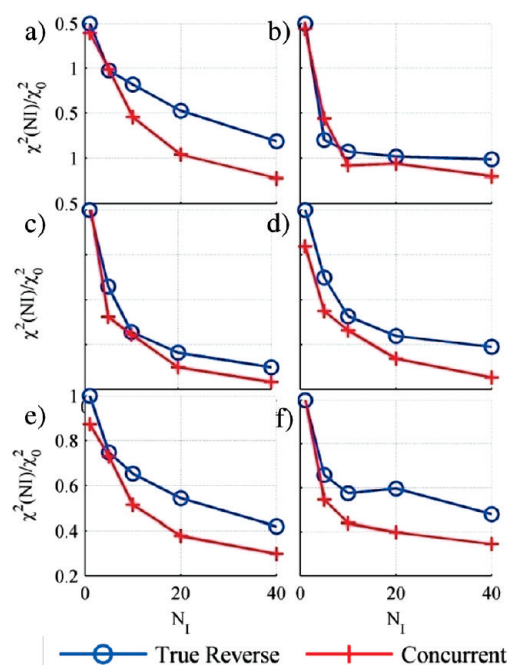


Figure 12. Accuracy of true and concurrent reverse paths are nearly equivalent. Each letter corresponds to the landscapes shown in Figure 5. MSE errors are normalized to the True Reverse pathway with $N_I = 1$ for each landscape. Simulations were run with $N_0 = 40$ k steps. See Tables 1 and 4 for additional settings.

Choice of Reverse Pathway Construction. The estimate of the transition probability using the forward trajectory information is a common and relatively well understood method for estimating the forward transition probability.⁴⁷ Since the reverse pathway construction is the novel feature of this approach, it is worthwhile to compare the reverse pathway construction methods. Figure 12 shows the MSE (eq 22) of the landscapes shown in Figure 10. For Figure 12, the MSE of the true reverse pathway with $N_I = 1$ is defined as χ^2_T and a ratio $\chi^2(N_I)/\chi^2_T$ is calculated for each of the pathways. For all cases, the relative MSE steadily decreases with the number of inner steps. The concurrent reverse pathway appears to generate slightly smaller MSEs. The

performance across this range of landscapes suggests that either choice would be sufficient. This is somewhat surprising, since the shape of the reverse pathway is completely different for each case, and it is tempting to think that the shape of the landscape could have a more profound effect on the error. This does not appear to be the case, however, in these test cases. Since the concurrent reverse pathway is both reliable and easier to implement, it is likely to be used more widely, and is used exclusively for the biomolecular applications presented here.

References

- (1) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
- (2) Panagiotopoulos, A. Z.; Quirke, N.; Stapleton, M.; Tildesley, D. J. *Mol. Phys.* **1987**, *63*, 527.
- (3) Panagiotopoulos, A. Z. *Mol. Phys.* **2002**, *100*, 237.
- (4) De Pablo, J. J.; Prausnitz, J. M. *Fluid Phase Equilib.* **1989**, *53*, 177.
- (5) Rosenbluth, M. N.; Rosenbluth, A. W. *J. Chem. Phys.* **1955**, *23*, 356.
- (6) Frenkel, D.; Siepmann, J. I. *Mol. Phys.* **1992**, *75*, 59.
- (7) Madras, N.; Sokal, A. *J. Stat. Phys.* **1988**, *50*, 109.
- (8) Hastings, W. K. *Biometrika* **1970**, *57*, 97.
- (9) Meirovitch, H. *J. Phys. A* **1982**, *15*, L735.
- (10) Meirovitch, H. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9235.
- (11) Chen, B.; Siepmann, J. I. *Theor. Chem. Acc.* **1999**, *103*, 87.
- (12) Chen, B.; Xing, J.; Siepmann, J. I. *J. Phys. Chem. B* **2000**, *104*, 2391.
- (13) Opps, S.; Schofield, J. *Phys. Rev. E* **2001**, *63*, 56701.
- (14) Brown, S.; Head-Gordon, T. *J. Comput. Chem.* **2003**, *24*, 68.
- (15) Frantz, D. D.; Freeman, D. L.; Doll, J. D. *J. Chem. Phys.* **1990**, *93*, 2769.
- (16) Xu, H.; Berne, B. J. *J. Chem. Phys.* **1999**, *110*, 10299.
- (17) Zhou, R. H. *J. Chem. Phys.* **1997**, *107*, 9185.
- (18) Huber, G.; McCammon, J. *Phys. Rev. E* **1997**, *55*, 4822.
- (19) Swendsen, R. H.; Wang, J. S. *Physical Review Letters* **1986**, *57*, 2607.
- (20) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
- (21) Lyman, E.; Ytreberg, F.; Zuckerman, D. *Phys. Rev. Lett.* **2006**, *96*, 28105.
- (22) Roitberg, A.; Okur, A.; Simmerling, C. *J. Phys. Chem. B* **2007**, *111*, 2415.
- (23) Bandyopadhyay, P. *J. Chem. Phys.* **2008**, *128*, 134103.
- (24) Gordon, M.; Freitag, M.; Bandyopadhyay, P.; Jensen, J.; Kairys, V.; Stevens, W. *J. Phys. Chem. A* **2001**, *105*, 293.
- (25) Ulmschneider, J. P.; Ulmschneider, M. B.; Di Nola, A. *J. Phys. Chem. B* **2006**, *110*, 16733.
- (26) Li, Z. Q.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611.
- (27) David, L.; Luo, R.; Gilson, M. K. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 157.
- (28) Killian, B. J.; Yundenfreund Kravitz, J.; Gilson, M. K. *J. Chem. Phys.* **2007**, *127*, 024107.
- (29) Dill, K.; Chan, H. *Nat. Struct. Biol.* **1997**, *4*, 10.
- (30) Dill, K. *Biochemistry* **1990**, *29*, 7133.
- (31) Fiser, A.; Sali, A. *Methods Enzymol.* **2003**, *374*, 461.
- (32) Rohl, C. A.; Strauss, C. E.; Misura, K. M.; Baker, D. *Methods Enzymol.* **2004**, *383*, 66.
- (33) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269.
- (34) Foreman, K. W.; Phillips, A. T.; Rosen, J. B.; Dill, K. A. *J. Comput. Chem.* **1999**, *20*, 1527.
- (35) Roux, B. *Implicit Solvent Models*; Marcel Dekker: New York City, 2001.
- (36) Rosso, L.; Mináry, P.; Zhu, Z.; Tuckerman, M. *J. Chem. Phys.* **2002**, *116*, 4389.
- (37) Rosso, L.; Tuckerman, M. *Mol. Simul.* **2002**, *28*, 91.
- (38) Darve, E.; Wilson, M.; Pohorille, A. *Mol. Simul.* **2002**, *28*, 113.
- (39) Coutsiar, E. A.; Seok, C. L.; Jacobson, M. P.; Dill, K. A. *J. Comput. Chem.* **2004**, *25*, 510.
- (40) Wedemeyer, W. J.; Scheraga, H. A. *J. Comput. Chem.* **1999**, *20*, 819.
- (41) Nilmeier, J.; Jacobson, M. *J. Chem. Theory Comput.* **2008**, *4*, 835.
- (42) Onuchic, J.; Luthey-Schulten, Z.; Wolynes, P. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545.
- (43) Manousiouthakis, V. I.; Deem, M. W. *J. Chem. Phys.* **1999**, *110*, 2753.
- (44) Crooks, G. *J. Stat. Phys.* **1998**, *90*, 1481.
- (45) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690.
- (46) Hetenyi, B.; Bernacki, K.; Berne, B. J. *J. Chem. Phys.* **2002**, *117*, 8203.
- (47) Jarzynski, C. Arxiv preprint cond-mat/0603185, 2006.
- (48) D Frenkel, B. S. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: Boston, 2002.
- (49) Dellago, C.; Bolhuis, P.; Csajka, F.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 1964.
- (50) Valleau, J.; Whittington, S. *J. Comput. Phys.* **1977**, *24*, 150.
- (51) Bolhuis, P.; Dellago, C.; Geissler, P.; Chandler, D. *J. Phys. Condens. Matt.* **2000**, *12*, 147.
- (52) Singhal, N.; Snow, C.; Pande, V. *J. Chem. Phys.* **2004**, *121*, 415.
- (53) Cao, J.; Berne, B. J. *J. Chem. Phys.* **1990**, *92*, 1980.
- (54) Zhou, R.; Berne, B. J. *J. Chem. Phys.* **1997**, *107*, 9185.
- (55) Arevalo, J. H.; Hassig, C. A.; Stura, E. A.; Sims, M. J.; Taussig, M. J.; Wilson, I. A. *J. Mol. Biol.* **1994**, *241*, 663.
- (56) Arevalo, J. H.; Stura, E. A.; Taussig, M. J.; Wilson, I. A. *J. Mol. Biol.* **1993**, *231*, 103.
- (57) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983.
- (58) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. *J. Comput. Chem.* **2002**, *23*, 517.
- (59) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474.
- (60) Rizzo, R. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1999**, *121*, 4827.

- (61) Wong, S. E.; Bernacki, K.; Jacobson, M. *J. Phys.Chem. B* **2005**, *109*, 5249.
- (62) Smart, J.; McCammon, J. A. *Biopolymers* **1999**, *49*, 225.
- (63) Groban, E. S.; Narayanan, A.; Jacobson, M. P. *PLoS Comput. Biol.* **2006**, *2*, e32.
- (64) Bernacki, K.; Hetenyi, B.; Berne, B. J. *J. Chem. Phys.* **2004**, *121*, 44.
- (65) Karplus, M. *J. Chem. Phys.* **2004**, *30*, 11.
- (66) Haliloglu, T.; Bahar, I.; Erman, B. *Phys. Rev. Lett.* **1997**, *79*, 3090.
- (67) Atilgan, A.; Durell, S.; Jernigan, R.; Demirel, M.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505.
- (68) Stern, H. A. *The Journal of Chemical Physics* **2007**, *126*, 164112.
- (69) Ballard, A. J. Jarzynski, C. *Proceedings of the National Academy of Sciences* **2009**.
CT8005166