Supporting Information

Improving Sampling by Exchanging Hamiltonians with Efficiently Configured Nonequilibrium Simulations

Robert M. Dirks*, Huafeng Xu, and David E. Shaw*

D. E. Shaw Research, 120 W. 45th St., 39th Floor, New York, NY USA 10036

1. Comments on Hybrid Monte Carlo

Hybrid Monte Carlo[1] is one way to generate new conformations for a many-atom system by using plain molecular dynamics to propose trial moves. To justify this, note that a repeated application of standard velocity verlet integration steps,

$$
\begin{aligned}
x(t+\Delta t) &= x(t)+v(t)\Delta t +0.5a(t)\Delta t^2 \\
v(t+\Delta t) &= v(t)+0.5\big[a(t)+a(t+\Delta t)\big]\Delta t , \\
ma(t) &= -\frac{dH\big[x(t)\big]}{dt}
\end{aligned}
\tag{S1}
$$

is simply a linear transformation of coordinates, with Δt a constant time step. Given infinite precision, the above transformation is time-reversible and volume preserving (i.e. it has a Jacobian determinant of one). Furthermore, since velocity verlet integration nearly conserves total energy for small Δt, trial moves can have an average acceptance rate close to one.

In a fashion similar to Hybrid Monte Carlo, nonequilibrium molecular dynamics can be used as a trial move generator for exchanging Hamiltonians. Essentially, the time-independent Hamiltonian in Equation (S1) is replaced with a time-dependent form, $H_{i\to j}(\hat{x};t')$, that smoothly links Hamiltonians H_i and H_j. To preserve reversibility, the time-dependent Hamiltonian is chosen to satisfy

$$H_{i \to j}(\hat{x};0) = H_i(\hat{x})$$
$$H_{i \to j}(\hat{x};t_{i,j}) = H_j(\hat{x})$$
$$H_{i \to j}(\hat{x};t') = H_{j \to i}(\hat{x};t_{i,j} - t'), \forall t' \in [0, t_{i,j}]$$
$$t_{i,j} = t_{j,i},$$

(S2)

where $t_{i,j}$ is the simulation length of a nonequilibrium trial exchange used to transform the coordinates.

Such a transformation is also time-reversible and unitary, but it does not conserve energy[2].

2. Derivation of Equation (2) from Main Text.

To quickly estimate mean round trip time (mrtt) for a given set of intermediates and nonequilibrium trial exchanges, the transitions between Hamiltonian levels can be roughly modeled as a Markov process with an equilibrium distribution that visits all levels equally. In replica exchange, the nature of the algorithm guarantees that equal time is spent simulating at every level, while in serial exchange, the average number of moves spent at each Hamiltonian depends on the values of tunable parameters.

Consider a Markov Chain on a system with N states, numbered 1 through N, obeying the special property that an attempted transition from state i to $i+1$ and from state $i+1$ to i are accepted with the same probability: $a_{i,i+1} = a_{i+1,i}$. From state i, the system attempts to transition into state $i-1$ or into state $i+1$, each with probability 1/2. Overall, the system will transition from state i to state $i-1$ with probability $\frac{1}{2}a_{i-1,i} = \frac{1}{2}a_{i,i-1}$, and to state $i+1$ with probability $\frac{1}{2}a_{i,i+1} = \frac{1}{2}a_{i+1,i}$, and will otherwise remain in state i. By construction, $a_{0,1} = a_{N,N+1} = 0$, and the symmetry of the acceptance probabilities guarantees that the equilibrium distribution of the system is uniform across all states. Next, we associate a cost to any given trajectory of this system. Each step spent in state i incurs a cost of s_i. Each transition attempt between i and $i+1$ incurs a cost of $t_{i,i+1} = t_{i+1,i}$. The mean round trip time is defined as the average cost required to travel from state 1 to state N, and back to state 1.

To prove,

$$mrtt = 2\left(\sum_{i=1}^{N} s_i + \sum_{i=1}^{N-1} t_{i,i+1}\right)\left(\sum_{i=1}^{N-1} 1/a_{i,i+1}\right), \tag{S3}$$

we first consider the simpler case where all $s_i = 1$ and all $t_{i,i+1} = 0$. The mean round trip time for this case is equivalent to the mean number of moves required to execute a round trip, and Equation (S3) simplifies to

$$mrtt = 2N\sum_{i=1}^{N-1} 1/a_{i,i+1}. \tag{S4}$$

We proceed by induction. Given $N = 2$, the average number of attempts required to transition from state 1 to N is the reciprocal of the transition probability, yielding $2/a_{0,1}$. The return trip behaves similarly, resulting in an average number of steps (and mrtt) of $4/a_{0,1}$.

Next, we show if (S4) is true for $N = k$, it must also hold for $N = k+1$. This is equivalent to showing the marginal cost of adding a state $k+1$ is given by

$$2k/a_{k,k+1} + \sum_{i=1}^{k} 2/a_{i,i+1}. \tag{S5}$$

We will show that the marginal cost of going forward from 1 to $k+1$ is given by the first term, and marginal cost of returning from $k+1$ to 1 is given by the second.

The expected number of moves to go from 1 to $k+1$, denoted as $m_{1,k+1}$, can be decomposed as $m_{1,k+1} = m_{1,k} + m_{k,k+1}$. This holds since all paths from 1 to $k+1$ must pass through state k. Note that $m_{1,k}$ is the same as it was for the smaller system with only k states, so the marginal cost of adding a state is simply $m_{k,k+1}$. To derive $m_{k,k+1}$, we use the well known Markov Chain result that the average number of moves it takes to return to a given state is the reciprocal of the equilibrium probability of that state. By construction, every state has equal equilibrium probability $1/(k+1)$. Thus, the mean return time for any state is simply $k+1$. Alternatively, the mean return time for state $k+1$ can be written explicitly as

$$\frac{1}{2}1 + \frac{1}{2}a_{k,k+1}\left(m_{k,k+1} + 1\right) + \frac{1}{2}\left(1 - a_{k,k+1}\right), \tag{S6}$$

where the first term represents a rejected transition attempt to non-existent state $k+2$, the middle term

represents a successful move to state k, and the final term represents a rejected move to state k.

Setting (S6) equal to $k+1$ and solving for $m_{k,k+1}$ gives the marginal forward cost of adding state $k+1$:

$$m_{k,k+1} = 2k / a_{k,k+1}. \qquad (S7).$$

Similarly, the cost of returning can be written as $m_{k+1,1} = \sum_{i=2}^{k+1} m_{i,i-1}$. To derive $m_{i,i-1}$, we consider the sub-problem of adding state $i-1$ to the sub-system comprising states i through $k+1$. The marginal cost of adding this state is also given by (S7), but with the k in the numerator replaced by the number of states in our sub-problem, $k+2-i$, and the denominator replaced by the acceptance probability of a transition from i to $i-1$:

$$m_{i,i-1} = 2(k-i+2) / a_{i-1,i}. \qquad (S8)$$

The total return trip cost, in the system of $k+1$ states, is thus $m_{k+1,1} = \sum_{i=2}^{k+1} 2(k-i+2)/a_{i-1,i}$; the corresponding total return trip cost from state k to state 1 in a system of only k states is obtained by replacing $k+1$ with k. The marginal cost of the return trip is the difference of the two, which simplifies to the second term in (S4). This completes the proof of (S3).

Next we sketch how to generalize to arbitrary cost functions. Since round trips generate the same limiting distribution of states as the Markov Chain, the average total cost of a round-trip is given by

$$mrtt = \sum_i \langle n_i \rangle s_i + \sum_i \langle n_{i,i+1} \rangle t_{i,i+1}, \qquad (S9)$$

where $\langle n_i \rangle$ is the average number of visits to state i during a round trip, and $\langle n_{i,i+1} \rangle$ is the average number of attempted transitions between states i and $i+1$. Since our Markov Chain visits all states equally, on average, and since transitions to either neighbor are attempted with probability 1/2, $\langle n_i \rangle = \langle n_{i,i+1} \rangle \equiv \bar{n}$ for all i, where $\bar{n}$ does not depend on s_i or $t_{i,i+1}$. Setting $s_i = 1, t_{i,i+1} = 0$ and equating (S9) with (S4), we get $\bar{n} = 2\sum_i 1/a_{i,i+1}$. Substituting $\bar{n}$ into Equation (S9) gives us (S3).

3. Outline of Parameter Selection Algorithm

Inputs to the parameter selection algorithm are a list of L possible levels (ranging from 1 to L) with associated costs s, and acceptance rate estimates of the form (l_0, l_1, t, a) where l_0 and l_1 are two levels, t is the length of nonequilibrium simulations connecting the levels, and a is the associated acceptance probability. The output is the subset of the inputs that form the lowest cost path connecting levels 1 and L, where cost is defined by the mean round trip time. The mean round trip time for such a path of N nodes is given by Equation (S3). In dynamic programming, a solution is found by first optimizing sub-problems. A simple, but not necessarily optimal strategy is given by the pseudocode in the main text. At the conclusion of the pseudocode, a fast backtracking algorithm can recover the precise path.

Although the heuristic described by the pseudocode is likely good enough for the problem at hand, optimality is not guaranteed. The multiplicative nature of (S3) means the best solution need not be an extension of an optimal path for a smaller problem. To exactly solve the problem, one approach is to store some suboptimal paths as well. One way to do this is to save the lowest cost to intermediate j for each obtainable value of $\left(\sum s + \sum t \right)$. If the s values are all the same, and t values take on only a few possibilities, the extra work is not prohibitive. This is the case for the problem studied here, and no further enhancements are needed.

If desired, various dominance criteria can be used to reduce the algorithm cost. As an example, if both $\left(\sum s + \sum t \right)$ and $\left(\sum 1/a \right)$ for one path to i are smaller than the corresponding values for a different path to i, then the latter path can be eliminated from contention. Also, one could execute the cheaper heuristic to get a reasonable solution, and use this answer to truncate the search of any sub-paths that exceed this known cost (since all cost parameters are non-negative).

4. Reaction Coordinate Details and the Softcore Potential

The full interaction Hamiltonian, H_{on} was prepared using ANTECHAMBER[3], and includes a simulation box with 1,022 TIP3P water molecules. To connect H_{on} and H_{off}, a fixed reaction

coordinate[4] is chosen that first linearly scales the charges to zero, yielding an intermediate Hamiltonian, $H_{uncharged}$. Next, the ligand–solvent van der Waals interactions are turned off with a softcore potential[4]

$$f_{ij}(r, \lambda_{vdw}) = 4(1 - \lambda_{vdw})\epsilon \left(\left[\frac{2\sigma^6}{\lambda_{vdw}^2 \sigma^6 + 2r^6} \right]^2 - \frac{2\sigma^6}{\lambda_{vdw}^2 \sigma^6 + 2r^6} \right), \tag{S10}$$

where r is the distance between atoms i and j, σ and ϵ are the standard van der Waals parameters, and λ_{vdw} is a softcore parameter ranging from zero to one. When λ_{vdw} is zero, this is the standard van der Waals potential, and when λ_{vdw} is one, this potential is zero everywhere. The reaction coordinate from $H_{uncharged}$ to H_{off} linearly scales λ_{vdw} from zero to one. To help overcome energy barriers, the reaction coordinate is extended further by turning off all ligand–ligand van der Waals interactions using the same softcore formula shown above. During initialization, the reaction coordinate is subdivided into a large number of possible intermediates. In the segment between H_{on} and $H_{uncharged}$, the ligand charge scaling parameter has candidate intermediates ranging from 0.0 to 1.0 separated by increments of 0.05. From $H_{uncharged}$ to H_{off}, the softcore parameter λ_{vdw} for ligand–solvent interactions has candidate intermediates ranging from 0.0 (full strength) to 0.75 in increments of 0.025, and from 0.75 to 1.0 (completely off) in increments of 0.0125. In switching from H_{off} to $H_{pseudo1D}$, ligand–ligand interactions are disabled using candidate λ_{vdw} values ranging from 0.0 to 1.0, in increments of 0.05.

5. Free Energies of Full Simulations

To estimate the free energy difference between Hamiltonians H_{on} and H_{off}, Bennett's method is used. After a simple serial exchange run, trajectory frames are sorted by their Hamiltonian level, and Bennett analysis is used to estimate the relative free energies between adjacent levels. Summing the free energy differences connecting H_{on} and H_{off} gives the desired value. This is repeated for the 19 serial exchange trajectories to give 19 independent estimates. To analyze serial exchange runs with nonequilibrium trials, a similar analysis, but with work values[5] replacing potential energy differences, gives 19 more estimates. In the case of replica exchange, the same approach yields a single estimate of

the free energy difference. To produce multiple estimates for comparison with the serial exchange methods, the 19-level replica exchange simulation is also approximated as 19 independent serial exchange runs[6], and the first 100 ns of these 19 runs are analyzed independently.

6. Molecular Dynamics Details

As described in the main text, simulations use M-SHAKE to constrain bonds involving hydrogen. Technically, using constraints that are not completely converged will disrupt the exact reversibility or phase-preserving properties of numerical integration schemes[7], but these deviations are considered small. Following the extended ensemble approach of Andersen[8], simulations included an isotropic scaling variable to control the size of the box, along with an associated momentum, η, and barostat mass, w. In these simulations, the conserved quantity is

$$H(\hat{x}) = E_{\text{pot}}(\hat{x}) + E_{\text{kin}}(\hat{x}) + p_{\text{ext}}V(\hat{x}) + \frac{\eta^2}{2w}, \qquad (\text{S11})$$

where p_{ext} is the constant external pressure, V is the volume of the simulation box, and E_{pot} and E_{kin} are the potential and kinetic energies, respectively. Andersen showed that such a simulation will have the same equilibrium properties as a constant pressure, constant enthalpy ensemble (NPH). Here, all momenta were periodically resampled from a target Gaussian distribution, thereby converting the simulation ensemble to constant pressure, constant temperature (NPT).

7. Hamiltonian Exchange Details

The Hamiltonian exchange performed here can be described as an alternating sequence of two types of moves. First, the system is sampled using the current Hamiltonian, and second, an attempt is made to change or swap the Hamiltonian. Here the focus is on serial exchange, but an analogous protocol for replica exchange is also possible.

In almost all cases (see main text for exceptions), the current Hamiltonian is sampled by randomizing the velocities according to the target Boltzmann distribution, followed by 1 ps of molecular dynamics with a constant enthalpy integrator. In a true Monte Carlo process, the end results of this molecular dynamics simulation should be subjected to an acceptance criterion to test whether this short simulation

sufficiently preserves the equilibrium distribution. However, as is implicitly done in most applications of Hamiltonian exchange, the end results of these simulations are always accepted. This introduces the same possible errors inherent in standard molecular dynamics applications, but these errors are considered negligible for a small enough time step. To be more rigorous, a Hybrid Monte Carlo move in the NPT ensemble[9] could be performed instead, using the conserved quantity from Equation (S11) to test for detailed balance.

As the second component of Hamiltonian Exchange, nonequilibrium simulations are introduced. Given a system at Hamiltonian level i, a target level $j = i \pm 1$ is chosen randomly. If j is outside the range of defined Hamiltonians, the move is automatically rejected. Otherwise, velocities are resampled, and a nonequilibrium simulation of length $t_{i,j}$ is performed. During the course of the run, either a charge scaling factor or a softcore λ_{vdw} is linearly interpolated between the endpoint values, with an update every 24 fs. When $t_{i,j} = 0$, this is essentially a standard implementation of Hamiltonian exchange, and atomic positions do not change. Acceptance or rejection of a trial move is determined by a Metropolis criterion[10], obtained by plugging the conserved quantity (S11) into the detailed balance equation, giving

$$a(\hat{x} \to T_{i,j}[\hat{x}]) = \min \left[\exp\left(-\beta \left[H_j(T_{i,j}[\hat{x}]) - H_i(\hat{x}) + f_j - f_i \right]\right), 1 \right]. \tag{S12}$$

Here, $T_{i,j}[\hat{x}]$ are the system coordinates after the nonequilibrium simulation, and the f_i are the extra parameters associated with serial exchange. In cases when the trial move is rejected, the system coordinates revert back to $\hat{x}$.

In the case of replica exchange, the Metropolis criterion depends on the total energies of the replicas after the nonequilibrium simulations, minus the original energies.

$$a\left[\left(\hat{x}_i, \hat{x}_j\right) \rightarrow \left(T_{i,j}[\hat{x}_i], T_{i,j}^{-1}[\hat{x}_j]\right)\right] =$$
$$\min\left[\exp\left(-\beta\left[H_j(T_{i,j}[\hat{x}_i]) - H_j(\hat{x}_j) + H_i(T_{i,j}^{-1}[\hat{x}_j]) - H_i(\hat{x}_i)\right]\right), 1\right] \tag{S13}$$

Here, $\hat{x}_i$ and , $\hat{x}_j$. refer to the subset of system coordinates associated with replicas i and j, respectively. Unlike serial exchange, no extra parameters are needed.

8. Method for Combining Conformer Data from Initial Estimates

The main text describes Hamiltonian, H_{pseudo1D}, where all ligand charges, ligand–solvent interactions, and ligand–ligand van der Waals interactions are zero. If bond angles and bond lengths are temporarily fixed (for resampling purposes only, and not during the actual molecular dynamics), the only remaining degrees of freedom are dihedral angles. In this case, the energies of rotating about bonds are independent (except for rotation about bonds within rings). After minimizing bond angles and lengths, each rotatable bond is partitioned into 1 or more rotamers. The relative free energy of a collection of these rotamers can be approximated as

$$-k_B T \sum_{r \in \{\text{rotatable bonds}\}} \log \int_{\theta_{\min}(r)}^{\theta_{\max}(r)} \exp[-U(\theta)/k_B T] d\theta, \tag{S14}$$

where $\theta_{min}(r), \theta_{max}(r)$ are the empirically determined boundaries of a specific rotamer of r, and $U(\theta)$ is the empirically determined potential energy function. The six conformers with the lowest energies are selected, and the initialization procedures described in the main text are applied to each conformer. This gives relative free energy estimates of a single conformer across all the tested Hamiltonians. These free energy approximations are used to estimate the overall mean acceptance rates of Hamiltonian exchange attempts.

It is worth noting that the treatment of a given Hamiltonian level as the combination of distinct conformers will violate the Markov approximation if these conformers do not readily exchange with one another. While it is possible to construct a more complicated Markov model that subdivides each level into conformers, this would complicate the mean round trip time equation, and is found to be unnecessary for the current problem. Since this Markov model is merely a tool for guessing good initial

parameters, errors such as these may make the simulations less efficient, but will not change the asymptotic correctness of the computed free energies.

Level	Ligand Charge Coefficient	Softcore λ_{vdw} (Ligand–Solvent)	Softcore λ_{vdw} (Ligand–Ligand)
0 (H_{on})	1.00	0	0
1	0.80	0	0
2	0.55	0	0
3	0.30	0	0
4 ($H_{uncharged}$)	0	0	0
5	0	0.3000	0
6	0	0.4750	0
7	0	0.5700	0
8	0	0.6500	0
9	0	0.7250	0
10	0	0.7750	0
11	0	0.8125	0
12	0	0.8375	0
13	0	0.8750	0
14 (H_{off})	0	1.0000	0
15	0	1	0.3
16	0	1	0.5
17	0	1	0.7
18 ($H_{pseudo1D}$)	0	1	1.0

Table S1. Production Parameters for Replica Exchange

Level	Ligand Charge Coefficient	Softcore λ_{vdw} (Ligand–Solvent)	Softcore λ_{vdw} (Ligand–Ligand)
0 (H_{on})	1.00	0	0
1	0.75	0	0
2	0.45	0	0
3 ($H_{uncharged}$)	0.00	0	0
4	0	0.400	0
5	0	0.575	0
6	0	0.700	0
7	0	0.775	0
8	0	0.825	0
9	0	0.875	0
10 (H_{off})	0	1.000	0
11	0	1	0.40
12	0	1	0.65
13	0	1	1.00
14 ($H_{pseudo1D}$)	0	1.0000	0

Table S2. Production Parameters for Serial Exchange

Type	λ_0	λ_1	Length (ps)	Estimated Acceptance Rate
Ligand Charge Scaling	1.0	0.0	0.5	0.03
Ligand Charge Scaling	1.0	0.0	1.0	0.07
Ligand Charge Scaling	1.0	0.0	2.0	0.16
Ligand Charge Scaling	1.0	0.0	5.0	0.29
Ligand Charge Scaling	1.0	0.0	10.0	0.38
Ligand–Solvent Softcore	0.0	0.775	1.0	0.01
Ligand–Solvent Softcore	0.0	0.775	2.0	0.04
Ligand–Solvent Softcore	0.0	0.775	5.0	0.12
Ligand–Solvent Softcore	0.0	0.775	10.0	0.22
Ligand–Solvent Softcore	0.0	0.825	5.0	0.02
Ligand–Solvent Softcore	0.0	0.825	10.0	0.06
Ligand–Solvent Softcore	0.0	0.825	20.0	0.13
Ligand–Solvent Softcore	0.0	0.825	50.0	0.27
Ligand–Solvent Softcore	0.825	1.000	1.0	0.04
Ligand–Solvent Softcore	0.825	1.000	2.0	0.07
Ligand–Solvent Softcore	0.825	1.000	5.0	0.16
Ligand–Solvent Softcore	0.825	1.000	10.0	0.25
Ligand–Ligand Softcore	0.0	0.65	0.2	0.23
Ligand–Ligand Softcore	0.0	0.65	0.5	0.31
Ligand–Ligand Softcore	0.0	0.65	1.0	0.38
Ligand–Ligand Softcore	0.0	0.65	2.0	0.46
Ligand–Ligand Softcore	0.0	0.65	5.0	0.48
Ligand–Ligand Softcore	0.0	0.65	10.0	0.54
Ligand–Ligand Softcore	0.0	0.70	0.2	0.17
Ligand–Ligand Softcore	0.0	0.70	0.5	0.25
Ligand–Ligand Softcore	0.0	0.70	1.0	0.32
Ligand–Ligand Softcore	0.0	0.70	2.0	0.37
Ligand–Ligand Softcore	0.0	0.70	5.0	0.41
Ligand–Ligand Softcore	0.0	0.70	10.0	0.46

Table S3. Nonequilibrium Trial Exchange Data Used for Level Selection

References

(1) Brass, A.; Pendleton, B. J.; Chen, Y.; Robson, B. Hybrid Monte Carlo Simulations Theory and Initial Comparison with Molecular Dynamics. *Biopolymers* **1993,** *33,* 1307–1315.

(2) Crooks, G. E. Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems. *J. Stat. Phys.* **1998,** *90* (5/6), 1481–1487.

(3) Wang, J.; Romain, M. W.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. of Comput. Chem.* **2004,** *25* (9), 1157–1174.

(4) Shirts, M. R.; Pande, V. S. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J. Chem. Phys.* **2005,** *122,* 134508.

(5) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Phys. Rev. Lett.* **2003,** *91* (14), 140601.

(6) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *J. Chem. Theory Comput.* **2007,** *3* (1), 26–41.

(7) Leimkuhler, B. J.; Skeel, R. D. Symplectic numerical integrators in constrained Hamiltonian systems. *J. Comput. Phys.* **1994,** *112,* 117–125.

(8) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **1980,** *72* (4), 2384–2393.

(9) Faller, R.; de Pablo, J. J. Constant pressure hybrid Molecular Dynamics-Monte Carlo simulations. *J. Chem. Phys.* **2002,** *116* (1), 55–59.

(10) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equations of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953,** *21* (6), 1087–1092.