Published in final edited form as:

J Chem Theory Comput. 2017 October 10; 13(10): 4649–4659. doi:10.1021/acs.jctc.7b00570.

Hamiltonian Monte Carlo with Constrained Molecular Dynamics as Gibbs Sampling

Laurentiu Spiridon^{*,†,‡} and David D. L. Minh^{*,†}

[†]Department of Chemistry, Illinois Institute of Technology, Chicago, IL 60616, USA

[‡]Department of Bioinformatics and Structural Biochemistry, Institute of Biochemistry of the Romanian Academy, Bucharest, Romania

Abstract

Compared to fully flexible molecular dynamics, simulations of constrained systems can use larger time steps and focus kinetic energy on soft degrees of freedom. Achieving ergodic sampling from the Boltzmann distribution, however, has proven challenging. Using recent generalizations of the equipartition principle and Fixman potential, here we implement Hamiltonian Monte Carlo based on constrained molecular dynamics as a Gibbs sampling move. By mixing Hamiltonian Monte Carlo based on fully flexible and torsional dynamics, we are able to reproduce free energy landscapes of simple model systems and enhance sampling of macrocycles.

Graphical abstract

1 Introduction

Molecular dynamics (MD) simulations are frequently used to provide insight into the behavior of nanoscale systems and to predict macroscopic properties. In many systems, however, the timescales of important conformational transitions far exceeds those accessible with modern computers. Instead of exploring the entirety of thermally accessible configuration space, simulations of these systems often become trapped in local minima. Interest in enhanced sampling continues to drive the development of algorithms that do not

^{*}To whom correspondence should be addressed.

Supporting Information Available

One figure showing Shirts' test for simulations based only on CDHMC moves and another showing comparing cumulative cluster counts for MIXED simulations using constrained dynamics moves with a 1.5 fs time step. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

necessarily preserve kinetics but sample configurations from a specific probability distribution, e.g. the Boltzmann distribution for the canonical ensemble.

One promising way to enhance sampling in MD simulations is to impose holonomic constraints on high-frequency motions that limit the time step used in MD integrators. Constraints allow for simulations to use a larger time step and focus kinetic energy on slower degrees of freedom more relevant for distinguishing molecular structures. MD simulations routinely use constraints on the lengths of molecular mechanics bonds involving hydrogen,^{1,2} enabling a time step boost from 1 to 2 fs. Another set of constraints that have received significant attention are on bond lengths and (some if not all) bond angles; in these simulations, torsion angles are allowed to be flexible. Torsional MD has been shown to be numerically stable with time steps up to 10 fs.³ It has been applied to refinement of NMR⁴ and crystallographic⁵ structures and to protein structure prediction.⁶ Further increases in integrator step size and conformational sampling are possible by constraining even larger subsets of a system. For example, large conformational changes in multimeric proteins have been observed by holding protein domains rigid while allowing connecting hinges to be flexible.⁷

With a number of developments in theory, algorithms, and software, we are now poised to unleash the full potential of constrained dynamics. On the algorithmic side, these advances include constrained MD integrators⁸ and compensating potential calculation methods⁹ that scale linearly, opposed to cubically, with the number of degrees of freedom. An important theoretical advance has been the derivation of an equipartition theorem for internal coordinates,¹⁰ allowing the assignment of reasonable velocities in accordance with a specified temperature. Finally, the GneimoSim¹¹ and the open-source Simbody¹² and Molmodel¹³ software packages facilitate the further development and application of constrained dynamics methods.

A constrained MD integrator that scales linearly opposed to cubically with the number of degrees of freedom has facilitated the use of constrained dynamics in complex systems such as macromolecules. Constrained dynamics can be performed by simulating in Cartesian coordinates and enforcing constraints at every time step^{1,2,14} or by simulating in internal coordinates and not allowing specific degrees of freedom to change.^{15,16} In the former vein, Tao et al.¹⁴ developed an algorithm that can constrain objects of arbitrary size, conserves energy, and is reversible and symplectic. Unfortunately, their algorithm is not efficient when, as in torsional MD, constrained blocks share atoms with other blocks. The latter approach requires fewer degrees of freedom and can eliminate force field terms such as bond length. While the first internal coordinate MD integrator¹⁷ required a mass matrix inversion with $O(n^3)$ complexity, Jain et al.⁸ developed a recursive algorithm with $O(n)$ complexity, eliminating a major bottleneck in internal coordinate MD and enabling its use for conformational search. However, because the probability density sampled by constrained and unconstrained dynamics differ, this was not sufficient to enable sampling from the appropriate Boltzmann distribution and allow the prediction of thermodynamic properties.

Distinctions between probability densities of constrained and unconstrained systems are encapsulated in the mass metric tensor used to calculate kinetic energies. For Cartesian

coordinates q with corresponding linear momenta p_q , the kinetic energy is $\frac{1}{2}p_q^T \mathcal{M}^{-1} p_q$, where $\mathcal{M} \in \mathcal{R}^{3N \times 3N}$ is a *position-independent* diagonal matrix with atomic masses along the diagonal. For N_ϕ generalized coordinates $\phi = \phi(q)$ with momenta p , the kinetic energy, $\frac{1}{2}p^T \mathcal{M}_\phi^{-1} p$, is based on the mass metric tensor, $\mathcal{M}_\phi \in \mathcal{R}^{N_\phi \times N_\phi}$. The mass matrix tensor is defined as $\mathcal{M}_\phi = J^T \mathcal{M} J$, where $J \in \mathcal{R}^{3N \times N_\phi}$ is a Jacobian matrix with elements $J_{kl} = \frac{\partial q_k}{\partial \phi_l}$. Unlike with Cartesian coordinates, $\mathcal{M}_\phi$ is *position-dependent*. To see the connection between the mass metric tensor and probability densities, first note that, in the Boltzmann distribution, the unnormalized joint probability density function for generalized coordinates and momenta is,

$$\rho(\phi, p) \propto \exp \{-\beta [H(\phi, p)]\}, \quad (1)$$

where $\beta = (k_B T)^{-1}$ is the inverse of the product of Boltzmann's constant and the temperature and $H(\phi, p) = \frac{1}{2}p^T \mathcal{M}_\phi^{-1} p + U(\phi)$ is the Hamiltonian, which depends on the potential energy $U(\phi)$. Although molecular dynamics simulations include both position and momentum coordinates, thermodynamic predictions ignore the latter and only consider the former; the marginal probability of coordinates is of greater interest than the joint probability. This marginal may be obtained by a multivariate Gaussian integral of Eq. 1 over generalized momenta,

$$\rho(\phi) \propto |\mathcal{M}_\phi|^{1/2} e^{-\beta U(\phi)}, \quad (2)$$

where $|\mathcal{M}_\phi|$ is the determinant of the mass metric tensor. If the system is unconstrained, then there are $3N$ degrees of freedom and the mass metric tensor is $\mathcal{M}_{3N} \in \mathcal{R}^{3N \times 3N}$. On the other hand, if the generalized coordinates ϕ are partitioned into N_f flexible coordinates ϕ_f and $3N - N_f$ constrained coordinates ϕ_c , then the mass metric tensor, $\mathcal{M}_{N_f} \in \mathcal{R}^{N_f \times N_f}$, is a sub-block of $\mathcal{M}_{3N}$.¹⁸ Because in general $|\mathcal{M}_{3N}| \neq |\mathcal{M}_{N_f}|$, if the same potential energy function is used, the marginal probability density of configurations sampled by constrained and unconstrained dynamics will differ.

To rectify differences between probabilities, Fixman proposed^{19,20} performing simulations with a modified potential $U'(\phi) = U(\phi) + U_F(\phi)$, where,

$$U_F(\phi_f) = \frac{1}{2} \beta^{-1} \ln \frac{|\mathcal{M}_{N_f}|}{|\mathcal{M}_{3N}|}. \quad (3)$$

The marginal probability distribution of generalized coordinates sampled by constrained dynamics on this modified potential matches the marginal for unconstrained dynamics on

the original potential. Patriciu et al.²¹ showed that the Fixman potential increases in importance with the length of a serial chain, and Echenique et al.²² concluded that it should be considered for peptides with more than 2 residues. While general techniques for evaluating determinants such as Cholesky decomposition have $O(n^3)$ complexity, an algorithm that evaluates the Fixman potential (and its derivatives, the Fixman torque) for serial robotic systems and general branched molecules with $O(n)$ complexity has been developed.^{9,23}

In interest of being more complete, we shall mention several alternative ways to circumvent the cubic complexity of the Fixman potential calculation. Both Katritch et al.²⁴ and Chen et al.²⁵ developed specialized internal coordinate force fields that closely approximate a “source” Cartesian force field without the compensating potential. Vitalis and Pappu²⁶ developed a molecular mechanics integrator that propagates dynamics on a modified energy hypersurface that closely approximates the Boltzmann distribution without explicitly evaluating the Fixman potential. These approaches, however, are restricted to a single predetermined set of constraints, e.g. on bond lengths and angles; one is unable to adjust the choice of constrained degrees of freedom in the middle of a simulation. In another strategy, König and Brooks²⁷ explicitly calculated the free energy cost of imposing constraints. This approach requires simulations of additional thermodynamic states between the constrained and unconstrained system.

Here, we employ an approach that only requires periodic evaluations of the Fixman potential and has several other advantages over pure MD: Hamiltonian Monte Carlo²⁸ (HMC). HMC, also known as Hybrid Monte Carlo, is a Markov chain Monte Carlo sampler based on the Metropolis-Hastings algorithm^{29,30} that uses a MD simulation to generate candidate configurations. Given an initial configuration ϕ_i , velocities are randomly initialized and a trajectory is propagated for a specified number of steps to generate a trial configuration ϕ_f . The trial configuration is then accepted or rejected with the probability,

$$\min \left(1, \frac{T(\phi_i|\phi_f)\rho(\phi_o)}{T(\phi_f|\phi_i)\rho(\phi_n)} \right), \quad (4)$$

where $T(\phi_j|\phi_k)$ is the probability of attempting the trial configuration ϕ_j starting from the configuration ϕ_k (j and k are either i or f). This attempt probability $T(\phi_j|\phi_k)$ includes the probability of generating the starting set of velocities. If the trial is accepted, then ϕ_f becomes the next sample in the Markov chain. If it is rejected, then ϕ_i is the next sample. HMC has several advantages over pure MD. First, velocity initialization *does not* need to obey exact Maxwell-Boltzmann statistics; it is only necessary to be able to calculate the probability ratio of generating initial and final velocities. Second, the MD integrator *does not* need to preserve the Boltzmann distribution of interest; it is only necessary to be able to compute the probability ratio of forward and reverse propagation (unity for a reversible, symplectic, and deterministic integrator). This flexibility allows the use of large time steps without concern for disturbing the Boltzmann distribution, constrained MD integrators, and distinct potential energy functions to guide the trajectory and to accept the final configuration. For example, the unmodified potential energy $U(\phi)$ can be used during

dynamics and the modified potential energy $U'(\phi)$ when calculating acceptance probabilities, significantly reducing the number of Fixman potential calculations.

In 1994, Forrest and Suter³¹ pioneered the use of constrained dynamics Hamiltonian Monte Carlo (CDHMC), simulating polymers with HMC based on torsional MD. Subsequently, CDHMC has been generalized to other statistical distributions,^{32,33} but these generalized methods have not been applied to molecular systems. Taking advantage of the lax requirements of HMC, Forrest and Suter³¹ initialized the moves using velocities from a Maxwell-Boltzmann distribution based on fictitious position-independent moments of inertia and excluded the Fixman potential from the dynamical propagator. They did not rigorously sample from the Boltzmann distribution, however, for two reasons. First, the Hamiltonian used in the acceptance criterion did not include the Fixman potential (after all, an efficient method for computing the potential had not yet been developed). Second, they solely used torsional MD and did not allow any bond lengths and angles to change.

To elaborate, *fully rigorous* ergodic sampling using only one set of constraints is impossible. A constrained simulation can be thought of as sampling from a *conditional* probability distribution, $\rho(\phi_f|\phi_c)$. In an ideal case, the conditional probability closely resembles the *marginal* probability integrated over all constrained degrees of freedom, $\rho(\phi_f|\phi_c) \approx \int d\phi_c \rho(\phi_f, \phi_c)$. One strategy to better approximate Boltzmann sampling is to select constraints such that the conditional and marginal probability distributions are similar. For example, Kandel et al.³⁴ recognized that torsional MD leads to distortions of phase space sampling due to coupling between bond and torsion angles. To address this issue, they opted to selectively relax angle constraints, allowing them to closely reproduce free energy surfaces of flexible peptides. This strategy of matching conditional and marginal probabilities, however, limits the efficiency gain made possible by constrained dynamics, requires foreknowledge of correlated degrees of freedom, and is not statistically rigorous unless constrained degrees of freedom are *completely independent* from the flexible ones. Because it does not allow a simulation to access regions of configuration space that have other values for constrained degrees of freedom, sampling cannot be ergodic.

To enable sampling of all configuration space, simulations must alternate constraints such that all degrees of freedom have the opportunity to be flexible. For example, torsional MD may be alternated with fully flexible Cartesian MD. These types of mixed simulations may be thought of as a special case of Gibbs sampling. Gibbs sampling is a Markov chain Monte Carlo method widely used in statistical inference³⁵ that generates samples from an series of different conditional probability distributions. In the realm of molecular simulation, replica exchange and expanded ensemble simulations can be thought of as Gibbs sampling.³⁶ To our knowledge, our present work is the first time constrained dynamics has been used in Gibbs sampling moves for rigorous Boltzmann sampling of molecular systems.

2 Methods

2.1 Software

To implement CDHMC, we developed the Generalized Molmodel (Gmolmodel) API (Figure 1) by modifying the Molmodel 2.3¹³ application program interface (API). Molmodel, in

turn, is based on Simbody 3.0,¹² which implements a multibody dynamics algorithm with $O(n)$ complexity, the generalized Newton-Euler Inverse Mass Operator method.^{8,37} While Molmodel was developed for biopolymers, especially RNA, and employs a specialized force field, Gmolmodel can simulate arbitrary molecules based on the AMBER³⁸ force field and file formats.

In addition to developing a new layer, we also modified Simbody¹² by implementing the linear-scaling Fixman potential calculation^{9,23} and adapting the Velocity Verlet integrator to include acceptance-rejection steps. Results from the functions that calculate the square root of the mass matrix and determinant were tested against the Eigen 3.2 library (<http://eigen.tuxfamily.org/>).

We developed Gmolmodel for use in our software package Alchemical Grid Dock³⁹ (AIGDock), which calculates binding free energies between flexible ligands and rigid protein conformations, a useful quantity in implicit ligand theory.⁴⁰ As such, Gmolmodel implements a shared memory model which allows it to send configurations to and obtain forces from the Molecular Modeling Toolkit⁴¹ simulation library that AIGDock is based on. Boost.Python from the Boost 1.55 libraries (<http://www.boost.org/>) was used to generate a python interface to access CDHMC from within AIGDock. CDHMC is now one of several samplers available for simulating small organic molecules within AIGDock. AIGDock is available under the open-source MIT license and may be downloaded from GitHub at <https://github.com/CCBatIIT/AIGDock>.

2.2 Sampling Algorithms

We used HMC moves based on either constrained or unconstrained MD.

In the former, we used two types of constraints. In the majority of CDHMC simulations, we used torsional dynamics moves: all bonds and angles were constrained and torsions were allowed to be flexible. In simulations of butane, we also considered rigid body moves: only the central torsion angle was allowed to be flexible. For both moves, velocities were drawn in accordance to the equipartition principle for internal coordinate molecular dynamics,¹⁰ such that the probability of initial and final velocities are given by the Boltzmann-weighted kinetic energy. Dynamics were propagated based on the unmodified potential energy function using the Velocity Verlet integrator from Simbody.¹² The integrator ensures that positions and velocities satisfy a set of constraints and computes velocity-dependent Coriolis and gyroscopic forces within a tolerance η of 10^{-4} . Unless otherwise specified, the trajectories consisted of 10 MD time steps of 4 fs. The Fixman compensating potential was included or excluded (for testing purposes) in the Metropolis-Hastings acceptance criterion.

More precisely, given a configuration ϕ_{t-1} , the next element of the Markov chain, ϕ_t is obtained by the following pseudocode:

1. Initialize the trial trajectory.
 - a. Set the initial configuration to $\phi_0^* = \phi_{t-1}$.

- b.** Draw initial momenta according to

$$p_0^* \sim \mathcal{N}(\mu=0, \Sigma=k_B T \cdot \mathcal{M}^{-1}(\phi^{t-1})).$$

- 2.** Propagate the trial trajectory by the Velocity Verlet algorithm with step size ϵ .
While $n < N$:

- a.**

$$\tilde{\phi}_{n+1} = \phi_n^* + \epsilon \nabla_p H(\phi_n^*, p_{n+\frac{1}{2}}^*) + \frac{1}{2} \epsilon^2 \nabla_\phi H(\phi_n^*, p_n^*)$$

- b.**

$$\phi_{n+1}^* = \text{proj}(\tilde{\phi}_{n+1})$$

- c.**

- i.** Compute velocity-dependent forces, f_p

- ii.**

$$\tilde{p}_{n+1} = p_n^* - \frac{1}{2} \epsilon [\nabla_\phi H(\phi_n^*, p_n^*) + \nabla_\phi H(\phi_{n+1}^*, p_n^*) + f_p]$$

- iii.**

$$p_{n+1}^* = \text{proj}(\tilde{p}_{n+1})$$

- iv.**

If $\|\tilde{p}_{n+1} - p_{n+1}^*\| / \|p_{n+1}^*\| > \eta$, repeat.

In this step, $\text{proj}(\cdot)$ signifies a projection that satisfies the constraints.

- 3.** Accept or reject the final configuration.

If $\min \left\{ 1, \exp \left[-\beta (H'(\phi_N^*, p_N^*) - H'(\phi_0^*, p_0^*)) \right] \right\} < (r \sim \mathcal{U})$, then

$$\phi_t = \phi_N^*.$$

Else,

$$\phi_t = \phi_0^*.$$

Here, $H'(\phi, p) = H(\phi, p) + U_F(\phi)$ is the modified Hamiltonian that includes the Fixman potential and $\mathcal{U}$ is the uniform distribution between 0 and 1.

For HMC moves based on unconstrained MD, velocities were drawn from the Maxwell-Boltzmann distribution, trajectories were generated using Velocity Verlet with a 1.5 fs time step, and the final configuration was accepted or rejected based on the the Metropolis-Hastings acceptance criterion. Most moves were based on 100 steps of dynamics, but for fair comparison with CDHMC, simulations to determine average acceptance rates were based on 10 steps of dynamics. Unless otherwise specified, simulations were conducted at 300 K.

We conducted four types of simulations that combined these two moves in different ways:

- 1.** UDHMC - only HMC moves based on unconstrained MD simulations were used.
- 2.** CDHMC-noFixman - HMC moves were based on constrained MD without the Fixman torque and the Fixman potential *was not* used during for calculating the acceptance probability.
- 3.** CDHMC - HMC moves were based on constrained MD without the Fixman torque and the Fixman potential *was* used during for calculating the acceptance probability.

4. **MIXED** - For most simulations, 15 HMC moves based on constrained MD were alternated with 5 HMC moves based on unconstrained MD. For alanine dipeptide simulations, the number of moves were 10 and 1, respectively. The former moves did not use the Fixman torque for dynamics but used the Fixman potential for calculating the acceptance probability.

3 Results and Discussion

Simulations were performed to test the ability of different sampling algorithms to either (1) reproduce the unconstrained Boltzmann distribution or (2) enhance sampling. Sampler correctness was tested on a serial chain, a butane molecule, and alanine dipeptide. Sampling efficiency was tested on five molecules that contain macrocycles.

3.1 Torsion Angle Distribution of an Idealized Serial Chain

To confirm the importance of using the Fixman potential in the Metropolis-Hastings acceptance probability, we conducted CDHMC-noFixman, CDHMC, and MIXED simulations with a 4-bead idealized serial chain, C_4 . The idealized model consists of three bonds with lengths fixed at 1.54 Å, angles fixed at 90°, and harmonic restraints of 83.66 and 43.46 kcal/mol on bonds and angles, respectively. Due to the lack of nonbonded interactions and torsional force field terms, it should have a uniform distribution in torsion space. Its Fixman potential has been derived in an closed-form analytical form^{42,43} and has been calculated by Patriciu et al.²¹ and Jain et al.⁹.

Our results verify that including the Fixman potential is necessary to draw from the appropriate Boltzmann distribution, the uniform distribution (Figure 2). In the CDHMC-noFixman calculations, the histogram has a periodic distortion. On the other hand, both the CDHMC and MIXED calculations recover the appropriate uniform distribution. The former calculations reproduce results from Jain et al.⁹. For this system, it is unnecessary to include the HMC moves based on unconstrained MD because the torsion angle is completely independent from the other degrees of freedom.

In addition to demonstrating the importance of the Fixman potential, these results also confirm that the constrained dynamics integrator does not require the Fixman torque in order to sample from the Boltzmann distribution. The fact that the Fixman torque is unnecessary is consistent with the use of a distinct *guidance* and *acceptance* potential described in Duane et al.²⁸.

3.2 Potential Energies of Butane

As another demonstration of Boltzmann sampling, we performed a test proposed by Shirts⁴⁴, analyzing the ratio of potential energy histograms at different simulation temperatures. In the canonical ensemble, the probability of observing a specific U is,

$$\pi(U|\beta) = Z(\beta)^{-1} \Omega(U) \exp(-\beta U), \quad (5)$$

where $Z(\beta)$ is the partition function at temperature β . Taking the logarithm of the ratio of probabilities of observing U at two different temperatures β_1 and β_2 yields,

$$\ln \left[\frac{\pi(U|\beta_2)}{\pi(U|\beta_1)} \right] = \ln \left[\frac{Z(\beta_1)}{Z(\beta_2)} \right] - (\beta_2 - \beta_1)U. \quad (6)$$

Hence, if one conducts simulations at two temperatures, generates a histogram of the potential energies, and takes the natural logarithm of their ratios, one should observe a linear trend of slope $-(\beta_2 - \beta_1)$ with respect to the potential energy.

To confirm that CDHMC and MIXED simulations pass Shirts' test, we conducted 10 independent simulations of butane for 2×10^5 acceptance-rejection steps each at both 300 and 450 K. Butane was parameterized using the generalized AMBER force field⁴⁵ from AmberTools 12 using AM1BCC partial charges^{46,47}. Simulations were performed with one of two CDHMC moves based on either torsional dynamics or rigid body dynamics in which only the middle torsion angle is flexible. For all four simulations (CDHMC versus MIXED, Torsional versus Rigid Body), the ratio of histograms qualitatively passes Shirts' test (Figure 3 and S1 in the Supporting Information). More quantitatively, the slope of the weighted linear least squares fit (weights were the inverse of the empirical variance) were within a standard deviation of the expected slope, $-(\beta_2 - \beta_1) = 0.13363$ (Table 1), confirming that the simulations pass Shirts' test.

3.3 Free Energy Landscape of Alanine Dipeptide

For our final tests of Boltzmann sampling, 7 independent simulations of alanine dipeptide in vacuum were performed for 10^7 steps at 300 K using the UDHMC and MIXED samplers. Alanine dipeptide was parameterized with the AMBER ff12SB force field. For the two types of simulations, the free energy as a function of backbone dihedrals $F(\Phi, \Psi) = -k_B T \ln[\rho(\Phi, \Psi)]$ are qualitatively consistent with each other (Figure 4). Differences between the estimated landscapes are most evident at high free energies where there is less sampling. In two regions that mark conformations between free energy minima, (1) $-180 < \Phi < -135$ and $-135 < \Psi < -90$ and (2) $-30 < \Phi < 0$ and $-90 < \Psi < 0$, free energy barriers are lower for the MIXED simulations.

For a more quantitative assessment, we compared the free energy minima of four major conformational minima of the landscape. The four regions C5, PPII, C7_{eq}, and α_L were defined as rectangles in backbone dihedral space, as described in Table 2. Comparing estimates based on UDHMC versus MIXED, the free energies of C5, PPII, and α_L relative to C7_{eq}, in units of kJ/mol, are estimated to be 0.18 ± 0.13 versus 0.27 ± 0.27 , 2.15 ± 0.15 versus 2.29 ± 0.30 , and 8.39 ± 0.26 versus 8.29 ± 0.57 , respectively. (Errors reported here are the standard deviations from the 7 simulations.) Free energy estimates for the two methods are within one standard deviation of one another, providing more quantitative evidence that MIXED samples from the same distribution as UDHMC.

3.4 Conformational Transitions in Alanine Dipeptide

In addition to verifying sampling validity, we also considered whether the MIXED approach increases sampling efficiency relative to UDHMC. As sampling was performed by Markov chain Monte Carlo, it is reasonable to consider properties of the Markov chain such as the empirical transition matrix (Table 3). The MIXED empirical transition matrix has diagonal elements that are less than or equal to and off-diagonal elements that are greater than those from CDHMC. This indicates that after the same number of simulation steps, MIXED simulations are more likely to undergo conformational transitions.

Sampling efficiency was also quantified by the mean first passage time (MFPT) between pairs of conformations. The MFPT was calculated from the empirical transition matrix using the PySAL library (<https://github.com/pysal>). For every pair of states, the MIXED approach has a shorter MFPT than UDHMC (Table 4). This is especially evident for transitions to the relatively isolated state α_L . Together, Tables 3 and 4 demonstrate that, per integrator step, MIXED is more efficient at sampling alanine dipeptide than UDHMC.

3.5 Sampling Efficiency of Macrocycle Simulations

To assess our methods on more challenging systems, we performed simulations on a set of macrocycles. Conformations of macrocycles, molecules that contain a cyclic scaffold with at least 12 atoms, are known to be difficult to sample.⁴⁸ We arbitrarily selected five macrocycles from Protein Data Bank Chemical Components Dictionary: 1R6, AA0, AB0, ACZ and ADN. There were parameterized using the generalized AMBER force field⁴⁵ from AmberTools 12 using AM1BCC partial charges.^{46,47}

One aspect of constrained dynamics that has the potential to boost sampling efficiency, but that we have not yet explored in this paper, is the integration time step. As mentioned in the introduction, constrained dynamics is numerically stable with a larger time step than unconstrained dynamics. Although a large time step can disrupt the Boltzmann distribution, the acceptance/rejection step of HMC can correct non-Boltzmann sampling. Moreover, the average Metropolis-Hastings acceptance rate is a reasonable way to assess whether simulated dynamics are numerically stable and capable of generating reasonable trial conformations. For our selected macrocycles, we find that the average acceptance rate for UDHMC is only around 35% at 2 fs and exactly zero at 3 fs (Figure 5). In contrast, with CDHMC, the average acceptance rate monotonically decreases with increasing step size, but is still above 0.1 with a step size of 10 fs. A reasonable (but somewhat arbitrary) compromise between step size and acceptance rate may be at 4 fs, at which all of our systems still have an acceptance probability greater than 0.7.

To compare the ability of constrained and unconstrained dynamics to sample distinct macrocycle conformations, we ran MIXED and UDHMC simulations at 625 K for 13×10^6 MD steps each. In UDHMC simulations, snapshots were saved after every set of 20 trials for a total of 65,000 snapshots per simulation. In MIXED simulations, snapshots were saved after every set of 15 CDHMC and 5 UDHMC trials (every 650 MD steps) for a total of 200,000 snapshots per simulation. One set of MIXED simulations was run with a 4 fs time step for constrained dynamics and another with a 1.5 fs time step; the former explores the

sampling enhancement available with a large time step and the latter tests whether enhanced sampling still occurs with a small time step. For all of our macrocycles, MIXED appears to sample a similar or broader part of configuration space than UDHMC after 6,500,000 MD steps (Figure 6).

To be more quantitative with our comparison, we clustered all simulation snapshots from each macrocycle. Hierarchical clustering using a weighted linkage function and distance cutoff of 0.25 Å was performed using SciPy 0.19.0 (<https://scipy.org/>). With a sufficiently long simulation, both MIXED and UDHMC simulations should access the entirety of conformational clusters available to the molecule. The quality of sampling is quantified by the rate at which clusters are accessed. Verifying the qualitative picture, MIXED samples a similar or greater number of conformational clusters than UDHMC in the same number of MD steps (Figure 7 and S2 in the Supporting Information). In systems where there are a similar number of clusters, the number of clusters is small, implying either that both approaches are locally trapped or that both approaches have explored the majority of thermally relevant configuration space. In all the systems, the intersection of sets (INTERSECTION) is nearly equivalent to the number of clusters sampled by UDHMC. Similarly, the union of sets (UNION) is nearly equivalent to the number of clusters accessed by MIXED. Thus, conformations assessed by MIXED are almost a superset of the conformations assessed by UDHMC, demonstrating the improved sampling efficiency possible through constrained dynamics. Comparing 1.5 and 4 fs time steps, the same qualitative trends are evident but clusters are accessed more quickly with the 4 fs time step.

Constrained dynamics may improve sampling by a similar mechanism as self-guided Langevin dynamics (SGLD).^{49–51} In addition to standard forces from Langevin dynamics, SGLD includes a guiding force based on a local average of the momenta over the recent time window. This effectively is a low-pass filter that suppresses high-frequency fluctuations, focusing kinetic energy on low-frequency degrees of freedom and increasing the number of clusters accessed as a function of time (e.g. Figure 14 in Wu et al.⁵⁰). Constrained dynamics takes the suppression of high-frequency motion to its logical extreme and, furthermore, enables the use of larger time steps.

4 Conclusions and Future Directions

We have shown the feasibility of achieving Boltzmann sampling and improving sampling efficiency by using CDHMC as a Gibbs sampling move. Boltzmann sampling was demonstrated by reproducing (1) the marginal distribution of torsion angles in a simple analytically tractable model, (2) the correct ratio of energy distributions at different temperatures, and (3) a two-dimensional free energy landscape observed with flexible dynamics simulations. Sampling efficiency compared to flexible dynamics was demonstrated by (1) an empirical transition matrix and MFPT between major conformations of alanine dipeptide, and (2) observing a greater number of distinct conformations in the same number of Monte Carlo trials (but fewer energy evaluations.)

Incidentally, we provide evidence that, not only in theory but in practice, constrained dynamics for CDHMC does not require computing the Fixman torque. Omission of the

Fixman torque provides an approximately 25% computational speedup.⁹ As the importance of the Fixman potential grows with system size,²¹ however, neglecting the Fixman torque for larger systems may guide the system to conformations that have low acceptance rates.

While we have demonstrated a proof of principle, there remains significant work to be done to unleash the full potential of constrained dynamics in molecular simulation. Although we showed that torsional dynamics based on an all-atom molecular mechanics force field is a reasonable constraint strategy for macrocycles at high temperature, it may not be as efficient at lower temperatures. At lower temperatures, energy barriers resulting from 1–4 contacts, van der Waals repulsions between atoms separated by three covalent bonds, are more difficult to traverse.^{24,25} To overcome this issue, other developers of torsional angle molecular dynamics methods have modified torsional angle terms^{24,25} and used softened van der Waals and electrostatic interactions²⁵ to increase sampling efficiency. It is possible to exploit these efficiency gains by using modified potential energy functions as the guidance Hamiltonian for generating Monte Carlo trials while retaining the atomistic force field for acceptance/rejection. Another possible issue is that torsional dynamics may not be very efficient when the time step is not limited by bond or angle vibration, but by collisions between nonbonded atoms. This could be the case in complex and compact systems such as biomolecules. We have shown, however, that an efficiency boost (albeit a smaller one) is possible even when using the same time step as in unconstrained dynamics, suggesting that sampling efficiency improvements can come from focusing kinetic energy on soft degrees of freedom. Nonetheless, for such systems, it may also be useful to develop new types of constraints, e.g. making domains in multimeric proteins rigid,⁷ to vastly accelerate rigorous Boltzmann sampling of large conformational transitions. Enhanced sampling of large conformational transitions (compared to unconstrained or torsional dynamics) may ultimately be the biggest benefit of using constrained dynamics as Gibbs sampling.

Another potential benefit of constrained dynamics is to allow higher simulation temperatures, for example, in replica exchange simulations. One limitation to using replica exchange as an enhanced sampling method is that high-temperature replicas may access regions of phase space irrelevant to enhanced sampling of low-temperature replicas. For example, at a high temperature a protein may be completely unfolded and disordered whereas at physiological temperature it is folded and ordered. If high-temperature replicas were restricted to CDHMC moves, then these replicas could enhance the rate of relative motions between rigid domains rather than unfolding them.

Finally, we would like to mention a few other possible ways to improve the efficiency of Gibbs sampling with constrained dynamics. MIXED simulation parameters, such the ratio of CDHMC to UDHMC moves or the time step, can be fine-tuned, perhaps using techniques suggested by Betancourt⁵³. In addition to CDHMD and UDHMC moves, simulations can use other Monte Carlo moves that preserve the Boltzmann distribution. Instead of using a standard velocity Verlet integrator, it may be possible to use the self-guided Langevin dynamics integrator⁵¹ in CDHMC moves. Instead of the original HMC acceptance criteria,²⁸ advanced acceptance criterion⁵² may be used. These advances could go a long way towards addressing sampling limitations in molecular simulation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Bernard Brooks for helpful discussions and Robert “Wes” Ludwig for careful reading of the manuscript. This research was supported by the National Institutes of Health (R15GM114781).

References

1. Ryckaert J-P, Ciccotti G, Berendsen HJ. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J Comput Phys.* 1977; 23:327–341.
2. Andersen HC. Rattle: A Velocity Version of the SHAKE Algorithm for Molecular Dynamics Calculations. *J Comput Phys.* 1983; 52:24–34.
3. Wagner JR, Balaraman GS, Niesen MJM, Larsen AB, Jain A, Vaidehi N. Advanced Techniques for Constrained Internal Coordinate Molecular Dynamics. *J Comput Chem.* 2013; 34:904–914. [PubMed: 23345138]
4. Stein EG, Rice LM, Brünger AT. Torsion-Angle Molecular Dynamics As a New Efficient Tool for NMR Structure Calculation. *J Magn Reson.* 1997; 124:154–164. [PubMed: 9424305]
5. Rice L, Brunger A. Torsion Angle Dynamics: Reduced Variable Conformational Sampling Enhances Crystallographic Structure Refinement. *Proteins.* 1994; 19:277–290. [PubMed: 7984624]
6. Larsen AB, Wagner JR, Jain A, Vaidehi N. Protein Structure Refinement of CASP Target Proteins Using GNEIMO Torsional Dynamics Method. *J Chem Inf Model.* 2013; 54:508–517.
7. Tek A, Korostelev AA, Flores SC. MMB-GUI: A Fast Morphing Method Demonstrates a Possible Ribosomal tRNA Translocation Trajectory. *Nucleic Acids Res.* 2016; 44:95–105. [PubMed: 26673695]
8. Jain A, Vaidehi N, Rodriguez G. A Fast Recursive Algorithm for Molecular Dynamics Simulation. *J Comput Phys.* 1993; 106:258–268.
9. Jain A, Kandel S, Wagner J, Larsen A, Vaidehi N. Fixman Compensating Potential for General Branched Molecules. *J Chem Phys.* 2013; 139:1–11.
10. Jain A, Park I-H, Vaidehi N. Equipartition Principle for Internal Coordinate Molecular Dynamics. *J Chem Theory Comput.* 2012; 8:2581–2587. [PubMed: 23341754]
11. Larsen AB, Wagner JR, Kandel S, Salomon-Ferrer R, Vaidehi N, Jain A. GneimoSim: A Modular Internal Coordinates Molecular Dynamics Simulation Package. *J Comput Chem.* 2014; 35:2245–2255. [PubMed: 25263538]
12. Sherman MA, Seth A, Delp SL. Simbody: Multibody Dynamics for Biomedical Research. *Procedia IUTAM.* 2011; 2:241–261. [PubMed: 25866705]
13. Flores SC, Sherman MA, Bruns CM, Eastman P, Altman RB. Fast Flexible Modeling of RNA Structure Using Internal Coordinates. *IEEE/ACM Trans Comput Biol Bioinf.* 2011; 8:1247–1257.
14. Tao P, Wu X, Brooks BR. Maintain Rigid Structures in Verlet Based Cartesian Molecular Dynamics Simulations. *J Chem Phys.* 2012; 137:134110. [PubMed: 23039588]
15. van Gunsteren W, Berendsen H. Algorithms for Macromolecular Dynamics and Constraint Dynamics. *Mol Phys.* 1977; 34:1311–1327.
16. Vaidehi N, Jain A. Internal Coordinate Molecular Dynamics: A Foundation for Multiscale Dynamics. *J Phys Chem B.* 2015; 119:1233–1242. [PubMed: 25517406]
17. Mazur AK, Dorofeev VE, Abagyan RA. Derivation and Testing of Explicit Equations of Motion for Polymers Described by Internal Coordinates. *J Comput Phys.* 1991; 92:261.
18. Go N, Scheraga H. On the Use of Classical Statistical Mechanics in the Treatment of Polymer Chain Conformation. *Macromolecules.* 1976; 9:535–542.
19. Fixman M. Classical Statistical Mechanics of Constraints: A Theorem and Application to Polymers. *Proc Natl Acad Sci USA.* 1974; 71:3050–3053. [PubMed: 16578718]

20. Fixman M. Simulation of Polymer Dynamics. I. General Theory. *J Chem Phys.* 1978; 69:1527–1537.
21. Patriciu A, Chirikjian GS, Pappu RV. Analysis of the Conformational Dependence of Mass-Metric Tensor Determinants in Serial Polymers with Constraints. *J Chem Phys.* 2004; 121:12708–12720. [PubMed: 15606297]
22. Echenique P, Calvo I, Alonso JL. Quantum Mechanical Calculation of the Effects of Stiff and Rigid Constraints in the Conformational Equilibrium of the Alanine Dipeptide. *J Comput Chem.* 2006; 27:1733–1747. [PubMed: 16900494]
23. Jain A. Compensating Mass Matrix Potential for Constrained Molecular Dynamics. *J Comput Phys.* 1997; 136:289–297.
24. Katritch V, Totrov M, Abagyan R. ICFF: A New Method to Incorporate Implicit Flexibility Into an Internal Coordinate Force Field. *J Comput Chem.* 2003; 24:254–265. [PubMed: 12497604]
25. Chen J, Im W, Brooks CL. Application of Torsion Angle Molecular Dynamics for Efficient Sampling of Protein Conformations. *J Comput Chem.* 2005; 26:1565–1578. [PubMed: 16145655]
26. Vitalis A, Pappu RV. A Simple Molecular Mechanics Integrator in Mixed Rigid Body and Dihedral Angle Space. *J Chem Phys.* 2014; 141:1–18.
27. König G, Brooks BR. Correcting for the Free Energy Costs of Bond or Angle Constraints in Molecular Dynamics Simulations. *Biochim Biophys Acta.* 2015; 1850:932–943. [PubMed: 25218695]
28. Duane S, Kennedy AD, Pendleton BJ, Roweth D. Hybrid Monte Carlo. *Phys Lett B.* 1987; 195:216–222.
29. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller A, Teller E. Equation of State Calculations by Fast Computing Machines. *J Chem Phys.* 1953; 21:1087–1092.
30. Minh DDL, Minh DLP. Understanding the Hastings Algorithm. *Commun Stat Simul Comput.* 2015; 44:332–349.
31. Forrest BM, Suter UW. Generalized Coordinate Hybrid Monte Carlo. *Mol Phys.* 1994; 82:393–410.
32. Girolami M, Calderhead B. Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods. *J R Stat Soc Series B Stat Methodol.* 2011; 73:123–214.
33. Brubaker MA, Salzmann M, Urtasun R. A Family of MCMC Methods on Implicitly Defined Manifolds. *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics (AISTATS-12).* 2012; 22:161–172.
34. Kandel S, Salomon-Ferrer R, Larsen AB, Jain A, Vaidehi N. Overcoming Potential Energy Distortions in Constrained Internal Coordinate Molecular Dynamics Simulations. *J Chem Phys.* 2016; 144:0–14.
35. Liu, JS. *Monte Carlo Strategies in Scientific Computing.* 2nd. Springer; New York: 2002.
36. Chodera JD, Shirts MR. Replica Exchange and Expanded Ensemble Simulations as Gibbs Sampling: Simple Improvements for Enhanced Mixing. *J Chem Phys.* 2011; 135:194110. [PubMed: 22112069]
37. Vaidehi N, Jain A, Goddard WA. Constant Temperature Constrained Molecular Dynamics: The Newton-Euler Inverse Mass Operator Method. *J Phys Chem.* 1996; 100:10508–10517.
38. Case D, Cerutti D, TE Cheatham I, Darden T, Duke R, Giese T, Gohlke H, Goetz A, Greene D, Homeyer N, Izadi S, Kovalenko A, Lee T, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Mermelstein D, Merz K, Monard G, Nguyen H, Omelyan I, Onufriev A, Pan F, Qi R, Roe D, Roitberg A, Sagui C, Simmerling C, Botello-Smith W, Swails J, Walker R, Wang J, Wolf R, Wu X, Xiao L, York D, Kollman P. AMBER. 2017
39. Minh, DDL. Protein-Ligand Binding Potential of Mean Force Calculations with Hamiltonian Replica Exchange on Alchemical Interaction Grids. 2015. arXiv:1507.03703v1 [physics.chem-ph]. arXiv.org ePrint archive. <http://arxiv.org/abs/1507.03703> (accessed Jul 14, 2015).
40. Minh DDL. Implicit Ligand Theory: Rigorous Binding Free Energies and Thermodynamic Expectations From Molecular Docking. *J Chem Phys.* 2012; 137:104106. [PubMed: 22979849]
41. Hinsen K. The Molecular Modeling Toolkit: A New Approach to Molecular Simulations. *J Comput Chem.* 2000; 21:79–85.

42. Pear MR, Weiner JH. Brownian Dynamics Study of a Polymer Chain of Linked Rigid Bodies. II. Results for Longer Chains. *J Chem Phys.* 1980; 72:3939.
43. Pear MR, Weiner JH. Brownian Dynamics Study of a Polymer Chain of Linked Rigid Bodies. *J Chem Phys.* 1979; 71:212.
44. Shirts MR. Simple Quantitative Tests to Validate Sampling From Thermodynamic Ensembles. *J Chem Theory Comput.* 2012; 9:909–926.
45. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and Testing of a General Amber Force Field. *J Comput Chem.* 2004; 25:1157–1174. [PubMed: 15116359]
46. Jakalian A, Bush BL, Jack DB, Bayly CI. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J Comput Chem.* 1999; 21:132–146.
47. Jakalian A, Jack DB, Bayly CI. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J Comput Chem.* 2002; 23:1623–1641. [PubMed: 12395429]
48. Chen JJ, Foloppe N. Tackling the Conformational Sampling of Larger Flexible Compounds and Macrocycles in Pharmacology and Drug Discovery. *Bioorg Med Chem.* 2013; 21:7898–7920. [PubMed: 24184215]
49. Wu X, Brooks BR. Toward Canonical Ensemble Distribution From Self-Guided Langevin Dynamics Simulation. *J Chem Phys.* 2011; 134:134108. [PubMed: 21476744]
50. Wu X, Hodoscek M, Brooks BR. Replica Exchanging Self-Guided Langevin Dynamics for Efficient and Accurate Conformational Sampling. *J Chem Phys.* 2012; 137:044106. [PubMed: 22852596]
51. Wu X, Brooks BR, Vanden-Eijnden E. Self-Guided Langevin Dynamics Via Generalized Langevin Equation. *J Comput Chem.* 2015; 37:595–601. [PubMed: 26183423]
52. Sohl-Dickstein J, Mudigonda M, DeWeese MR. Hamiltonian Monte Carlo Without Detailed Balance. *Proceedings of the 31st International Conference on Machine Learning.* 2014; 32:1–9.
53. Betancourt, M. Identifying the Optimal Integration Time in Hamiltonian Monte Carlo. 2016. arXiv: 1601.00225v1 [stat.ME]. arXiv.org ePrint archive. <http://arxiv.org/abs/1601.00225v1> (accessed Feb 16, 2017)

Figure 1.
The software architecture of GMolmodel

Figure 2.

Histogram of the torsion angle observed in simulations of the idealized serial chain C_4 . We conducted 10 independent simulations of C_4 for 10^6 samples each. Samples were collected after each acceptance-rejection step. Error bars denote the standard deviation.

Figure 3.

Shirts' test for MIXED simulations at 300 and 450 K. The lines are the expected slope (black), observed mean and standard deviation (blue), and the weighted least squares regression fit (red). CDHMC moves were based on torsional dynamics (top) or rigid body dynamics (bottom). A similar plot for CDHMC simulations (only CDHMC moves) is available as Figure S1 in the Supporting Information.

Figure 4.

The free energy landscape of alanine dipeptide, based on histograms of UDHMC (left) and MIXED (right) simulations. Histograms were generated using a 15° spacing. Free energies are displayed in units of kJ/mol. Unsampled regions are shown in red.

Figure 5. Acceptance rates for macrocycle simulations with moves based on ten steps of unconstrained (top) and torsional (bottom) dynamics. The molecules are 1R6 (red), AA0 (black), AB0 (blue), ACZ (magenta), and ADN (green). For each system, 3 independent simulations were run with 2×10^5 trials each. Error bars represent the standard deviations. Data are slightly offset along the x axis so that error bars do not overlap. At 3 fs, the acceptance rate of unconstrained dynamics moves is precisely zero.

Figure 6.
Observed conformations of macrocycles after 6,500,000 MD steps of UDHMC (left) and MIXED (right) simulations. Conformations are aligned to the first snapshot.

Figure 7. Cumulative cluster count

Structures from both UDHMC (circles) and MIXED (squares) simulations were aggregated and clustered. Each curve and error bar represents the mean and standard deviation of the number of clusters accessed up until the specified number of MD steps. The number of clusters obtained from either UDHMC or MIXED (UNION) and from both UDHMC and MIXED (INTERSECTION) are also shown. MIXED simulations used a time step of 4 fs for

torsional dynamics moves. For a similar analysis based on constrained dynamics moves using a 1.5 fs time step, see Figure S2 in the Supporting Information.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1
Slopes fitted to the ratio of histograms

The columns are for type of simulation (only CDHMC or MIXED), DOF for the degrees of freedom flexible in CDHMC moves (Torsional or Rigid Body), m for the slope, $|m - 0.13363|$ for the difference between expected and observed slope, and σ for the standard deviation of the fit.

Type	DOF	m	σ	$ m - 0.13363 $
CDHMC	Torsional	0.13693	3.86×10^{-3}	3.30×10^{-3}
CDHMC	Rigid Body	0.13524	5.03×10^{-3}	1.61×10^{-3}
MIXED	Torsional	0.13357	1.61×10^{-3}	0.06×10^{-3}
MIXED	Rigid Body	0.12259	4.03×10^{-2}	1.10×10^{-2}

Table 2

Boundaries of the four major conformational regions

	ϕ_{min}	ϕ_{max}	ψ_{max}	ψ_{max}
C5	-180	-95	105	180
PPII	-96	-45	105	180
C7 _{eq}	-96	-45	-25	104
α_L	35	85	-180	25

Table 3
Empirical transition matrices

between major conformational regions of alanine dipeptide based on UDHMC (top) and MIXED (bottom). Within each matrix, row i and column j indicate the observed number of transitions from state i to state j over the total number of transitions. Transitions were counted every 2 acceptance-rejection steps for UDHMC or 11 acceptance-rejection steps for MIXED, which use an equal number of MD integrator steps. The quantities in parentheses are standard deviations based on bootstrapping transition events 500 times.

	C5	PPII	C7 _{eq}	α_L
C5	0.831 (4.38×10 ⁻⁴)	0.131 (3.93×10 ⁻⁴)	0.038 (2.16×10 ⁻⁴)	2.5×10 ⁻⁵ (5.59×10 ⁻⁶)
PPII	0.248 (7.00×10 ⁻⁴)	0.560 (8.95×10 ⁻⁴)	0.192 (6.24 ×10 ⁻⁴)	1.0×10 ⁻⁴ (1.46×10 ⁻⁵)
C7 _{eq}	0.031 (1.73×10 ⁻⁴)	0.082 (2.84×10 ⁻⁴)	0.887 (3.22×10 ⁻⁴)	3.7×10 ⁻⁴ (1.94×10 ⁻⁵)
α_L	1.3×10 ⁻⁴ (2.86×10 ⁻⁵)	2.7×10 ⁻⁴ (3.91×10 ⁻⁵)	2.3×10 ⁻³ (1.22×10 ⁻⁴)	0.997 (1.29×10 ⁻⁴)
C5	0.804 (4.71×10 ⁻⁴)	0.145 (4.14×10 ⁻⁴)	0.052 (2.40×10 ⁻⁴)	2.3×10 ⁻⁵ (5.24×10 ⁻⁶)
PPII	0.287 (7.82×10 ⁻⁴)	0.505 (9.44×10 ⁻⁴)	0.207 (6.88×10 ⁻⁴)	1.4×10 ⁻⁴ (1.84×10 ⁻⁵)
C7 _{eq}	0.043 (2.06×10 ⁻⁴)	0.087 (3.08×10 ⁻⁴)	0.869 (3.81×10 ⁻⁴)	5.9×10 ⁻⁴ (2.32×10 ⁻⁵)
α_L	9.5×10 ⁻⁵ (2.19×10 ⁻⁵)	2.8×10 ⁻⁴ (3.87×10 ⁻⁵)	2.9×10 ⁻³ (1.16×10 ⁻⁴)	0.997 (1.24×10 ⁻⁴)

Table 4

MFPTs between major conformational regions of alanine dipeptide based on UDHMC (top) and MIXED (bottom). Within each matrix, row i and column j indicate the MFPT for moving from state i to state j . The quantities in parentheses are standard deviations based on bootstrapping transition events 500 times. Standard deviations in the last rows and columns are distinct from one another beyond the first three digits.

	C5	PPII	C7 _{eq}	α_L
C5		8.79 (2.59×10 ⁻²)	13.91 (4.69×10 ⁻²)	5225.51 (2.51×10 ²)
PPII	9.91 (3.07×10 ⁻²)		10.20 (3.90×10 ⁻²)	5221.93 (2.51×10 ²)
C7 _{eq}	17.27 (5.07×10 ⁻²)	12.43 (4.27×10 ⁻²)		5214.84 (2.51×10 ²)
α_L	386.66 (17.9)	381.97 (17.9)	372.64 (17.9)	
C5		8.20 (2.17×10 ⁻²)	11.66 (3.51×10 ⁻²)	3384.47 (1.27×10 ²)
PPII	8.15 (2.46×10 ⁻²)		8.88 (3.37×10 ⁻²)	3381.53 (1.27×10 ²)
C7 _{eq}	14.47 (4.17×10 ⁻²)	11.73 (3.55×10 ⁻²)		3374.84 (1.27×10 ²)
α_L	312.64 (11.3)	309.75 (11.3)	300.21 (11.3)	