

Overcoming potential energy distortions in constrained internal coordinate molecular dynamics simulations

Saugat Kandel, Romelia Salomon-Ferrer, Adrien B. Larsen, Abhinandan Jain, and Nagarajan Vaidehi

Citation: *The Journal of Chemical Physics* **144**, 044112 (2016); doi: 10.1063/1.4939532

View online: <http://dx.doi.org/10.1063/1.4939532>

View Table of Contents: <http://scitation.aip.org/content/aip/journal/jcp/144/4?ver=pdfcov>

Published by the **AIP Publishing**

Articles you may be interested in

[Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates](#)
J. Chem. Phys. **141**, 014111 (2014); 10.1063/1.4885338

[A density functional theory based estimation of the anharmonic contributions to the free energy of a polypeptide helix](#)

J. Chem. Phys. **135**, 084122 (2011); 10.1063/1.3629451

[Relating kinetic rates and local energetic roughness by accelerated molecular-dynamics simulations](#)

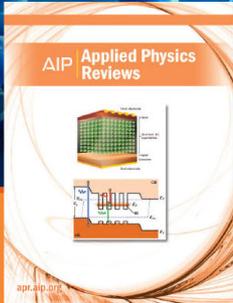
J. Chem. Phys. **122**, 241103 (2005); 10.1063/1.1942487

[Molecular Dynamics Study of Protein Folding: Potentials and Mechanisms](#)

AIP Conf. Proc. **661**, 195 (2003); 10.1063/1.1571312

[An energy function for dynamics simulations of polypeptides in torsion angle space](#)

J. Chem. Phys. **108**, 8264 (1998); 10.1063/1.476181



NEW Special Topic Sections

NOW ONLINE
Lithium Niobate Properties and Applications:
Reviews of Emerging Trends

AIP | Applied Physics
Reviews

Overcoming potential energy distortions in constrained internal coordinate molecular dynamics simulations

Saugat Kandel,¹ Romelia Salomon-Ferrer,¹ Adrien B. Larsen,¹ Abhinandan Jain,^{2,a)} and Nagarajan Vaidehi^{1,b)}

¹Division of Immunology, Beckman Research Institute of the City of Hope, Duarte, California 91010, USA

²Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA

(Received 21 September 2015; accepted 21 December 2015; published online 29 January 2016)

The Internal Coordinate Molecular Dynamics (ICMD) method is an attractive molecular dynamics (MD) method for studying the dynamics of bonded systems such as proteins and polymers. It offers a simple venue for coarsening the dynamics model of a system at multiple hierarchical levels. For example, large scale protein dynamics can be studied using torsional dynamics, where large domains or helical structures can be treated as rigid bodies and the loops connecting them as flexible torsions. ICMD with such a dynamic model of the protein, combined with enhanced conformational sampling method such as temperature replica exchange, allows the sampling of large scale domain motion involving high energy barrier transitions. Once these large scale conformational transitions are sampled, all-torsion, or even all-atom, MD simulations can be carried out for the low energy conformations sampled via coarse grained ICMD to calculate the energetics of distinct conformations. Such hierarchical MD simulations can be carried out with standard all-atom forcefields without the need for compromising on the accuracy of the forces. Using constraints to treat bond lengths and bond angles as rigid can, however, distort the potential energy landscape of the system and reduce the number of dihedral transitions as well as conformational sampling. We present here a two-part solution to overcome such distortions of the potential energy landscape with ICMD models. To alleviate the *intrinsic distortion* that stems from the reduced phase space in torsional MD, we use the Fixman compensating potential. To additionally alleviate the *extrinsic distortion* that arises from the coupling between the dihedral angles and bond angles within a force field, we propose a *hybrid ICMD method* that allows the selective relaxing of bond angles. This hybrid ICMD method bridges the gap between all-atom MD and torsional MD. We demonstrate with examples that these methods together offer a solution to eliminate the potential energy distortions encountered in constrained ICMD simulations of peptide molecules. © 2016 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4939532>]

I. INTRODUCTION

The all-atom Molecular dynamics (MD) simulations involve solving the classical Newton equations of motion in absolute Cartesian coordinates with all degrees of freedom movable. The all-atom model is attractive for its simplicity but is not suitable for coarsening the dynamic model of the simulation system to efficiently simulate long time scale processes such as domain motion in proteins. Attempts to focus the conformational search in the internal coordinate degrees of freedom have been made by eliminating the high-frequency degrees of freedom through the use of holonomic constraints in Cartesian coordinates. However, solving these equations requires differential-algebraic equation solvers even in the absence of loops, thus adversely affecting the robustness and increasing the complexity of the simulations.

The internal coordinates provide a more natural approach to apply holonomic constraints directly to the dynamics model of a protein system. Thus, the Internal Coordinate Molecular Dynamics (ICMD) method that describes protein structure in

terms of relative bond, angle, and torsion (BAT) coordinates is highly suited for coarsening the dynamics model of a protein. In ICMD methods, the holonomic constraints can be applied by simply excluding selected degrees of freedom from the dynamics model of the simulation system, while still retaining the ordinary differential equations form for the equations of motion for non-loop systems. For example, the ICMD model with all the bonds and angles frozen, and only the torsions free is the well-known *torsional* molecular dynamics (TMD) model.

A longstanding challenge with the application of holonomic constraints is that their use in MD simulations alters the equilibrium statistical properties from those observed for classical all-atom models without constraints.¹⁻⁷ The statistical differences in the constrained simulations affect the probability density functions (pdf's) as well as the transition barrier crossing rates between the various microstates of the system. Such statistical distortions stem from the following sources:

1. A mass-matrix dependent *intrinsic* distortion: This intrinsic distortion is inherent to the constrained model and arises from the reduced dimension of the phase space

^{a)}Electronic mail: Abhi.Jain@jpl.nasa.gov

^{b)}Electronic mail: nvaidehi@coh.org

in constrained ICMD models. In these models, the configuration of the degrees of freedom treated as rigid affects the mass matrix determinant and thus the partition function calculated from the constrained dynamics. The intrinsic distortion does not depend on the force field potentials.

2. A forcefield dependent *extrinsic* distortion: The extrinsic distortion is a consequence of the dynamic coupling among the BAT degrees of freedom from the external force field. In proteins, since the bond lengths are well separated in frequency space from the bond angles and torsions, the coupling between the bond lengths and the other degrees of freedom is negligible. Bond angles, however, can have overlapping frequencies with torsional angles, thus causing their motion to be dynamically coupled. The freezing of bond angle degrees of freedom in a constrained model distorts the torsional angle distributions and alters the heights of the transition barriers.

Correcting for the distortions introduced in the constrained model has been a topic of significant interest in the literature, with a range of proposed solutions.^{1,8-12} In particular, a configuration dependent correction potential proposed by Fixman,¹ the *Fixman potential*, has been the subject of several investigations to quantify the extent to which the Fixman potential corrects for the bias introduced by the constraints. These investigations have been limited to simulations of small and idealized serial chains without any extrinsic effects from forcefields.¹³⁻¹⁸

In previous works, we have developed a computationally tractable algorithm to calculate the Fixman potential using spatial operator algebra methods^{19,20} for idealized serial chains as well as general branched systems. We have shown that, in simulations without any extrinsic coupling, the constrained model with the Fixman potential recovers the torsional pdf's obtained from the unconstrained model even for complex branched chains such as peptides.²⁰

In the general case, the pdf that is used to calculate the statistical averages of the configuration and velocity dependent properties of a system is simply the partition function integrand modulo a normalizing constant. In constrained systems, the intrinsic distortion introduces additional mass matrix dependent terms in this integrand, thus introducing biases in the statistical behavior of the system. For quantities that do not depend upon the velocity coordinates, we can integrate the partition function over the velocity coordinates to obtain the simpler *configuration pdf* and use it instead to compute the statistical averages. Quantities such as the barrier crossing rates, however, depend on the velocity coordinates as well as the configuration coordinates. In this paper, we show that while the Fixman potential completely removes the intrinsic bias from the configuration pdf's, it only does so partially for the velocity-dependent barrier crossing rates. In what we believe to be a first, we also apply the Fixman potential to simulations of general branched molecules with all-atom forcefields, establishing that the Fixman potential alone does not compensate for the extrinsic distortions in general constrained simulations.

To reduce the extrinsic distortions in the configurational probability density function, we have developed the "hybrid ICMD method." The hybrid ICMD simulation method allows the user to treat any desired bond angle degree of freedom as flexible, while keeping the other bond angle degrees of freedom rigid. The hybrid ICMD method, with the Fixman potential applied by default, effectively bridges the gap between the entirely flexible all-atom model and the TMD model. In this work, we apply the hybrid ICMD method for the conformational sampling of several dipeptides to identify the key bond angles for each dipeptide model. We demonstrate that opening this small subset of angles is sufficient to retrieve the configuration pdf's observed in all-atom Cartesian models with no adverse impact on the time step size. Additionally, we demonstrate the use of hybrid ICMD simulations in sampling the NMR structures of a ten amino acid peptide. The advantages of the hybrid ICMD simulation method are as follows:

1. It allows the user to freeze any torsional degree of freedom in the rigid parts of the protein while allowing movement of bond angles in other flexible parts of the protein. This provides a simulation capability for coarse graining of the dynamics model without compromising the accuracy of the force fields.
2. Using the hybrid ICMD method, one can perform a fully flexible internal coordinate molecular dynamics simulation. Such a method is not available to date and may be useful for studying the dynamics of bonded systems.

II. METHODS

The hybrid ICMD simulations carried out in this work are build upon our previously developed robust long time scale ICMD simulation method called the Generalized Newton-Euler Inverse Mass Operator (GNEIMO) method.²⁰⁻²⁵ Here, we describe the GNEIMO method briefly since it is the basis for the hybrid ICMD method. Since some of the bond lengths and bond angles can be treated as rigid in the hybrid ICMD GNEIMO method, the degrees of freedom in the equations of motion in ICMD method become coupled and have the form

$$\mathcal{M}(\alpha)\ddot{\alpha} + \mathcal{C}(\alpha, \dot{\alpha}) = \mathcal{J}(\alpha), \quad (1)$$

where α is the vector of the generalized coordinates (e.g., torsional angles), \mathcal{J} denotes the vector of generalized forces (e.g., torques), $\mathcal{M}(\alpha)$ denotes the mass matrix (moment of inertia tensor), and $\mathcal{C}(\alpha, \dot{\alpha})$ includes the velocity dependent Coriolis forces. The dynamics of motion is obtained by solving Eq. (1) for the $\ddot{\alpha}$ acceleration and integrating them to obtain new velocities and coordinates. The GNEIMO method uses a spatial operator algebra based method to derive an analytical expression for the inverse of the mass matrix followed by the following expression for $\ddot{\alpha}$:

$$\ddot{\alpha} = [\mathbb{I} - \mathcal{H}\psi\mathcal{K}]^* \mathcal{D}^{-1} [\mathcal{J} - \mathcal{H}\psi(\mathcal{K}\mathcal{J} + \mathcal{P}\mathbf{a} + \mathbf{b})] - \mathcal{K}^*\psi^*\mathbf{a}. \quad (2)$$

The \mathcal{H} , ψ , \mathcal{K} , etc., terms in the above expression are associated with mass matrix related factorizations and are described in detail in Refs. 21 and 26. The expression on the right can

be evaluated using computationally cost-effective recursive algorithms. Importantly, these recursive equations are generic and remain the same even when some or all of the bond angles are open as in the hybrid ICMD method. Additionally, the GNEIMO-Fixman method^{19,20} for the calculation of the Fixman torque continues to apply even when the bond angles are open.

The GNEIMO method forms the basis of the *GneimoSim* ICMD software package²⁵ which has been used for a variety of biomolecular applications such as protein folding,^{23,24,27} domain motion,^{28,29} and refinement of protein homology models.^{22,23,30,31}

A. Fixman potential to correct for the intrinsic distortion

In TMD simulations, all bond lengths and bond angles are treated as rigid, leading to changes in the statistics calculated from the simulation trajectories. To compensate for this distortion, Fixman proposed a correction potential that depends on the mass matrix determinant.¹ We have previously shown that the Fixman potential corrects for these differences and recovers the correct configuration pdf's for simple and complex branched molecules.^{14,15,20} In this section, we demonstrate that the Fixman potential only partially removes the biases in the velocity dependent barrier crossing rates. We further show that, as expected, the Fixman potential does not correct for statistical distortions arising from force field induced extrinsic distortion.

1. Derivation of the Fixman potential

A polymer model with n atoms and $3n$ Cartesian coordinates can also be equivalently described in terms of $3n$ BAT coordinates. For an unconstrained system, all $3n$ coordinates are allowed to vary. For a constrained system, however, we can partition these $3n$ BAT coordinates into N unconstrained coordinates, denoted as α , and $(3n - N)$ constrained coordinates, denoted as q .

Using the α and q terminologies for the unconstrained system, where both α and q can vary, we can represent the canonical momentum coordinates as p to express the unconstrained Hamiltonian $\mathcal{H}_u(\alpha, q, p)$ in the form

$$\mathcal{H}_u(\alpha, q, p) = \frac{1}{2} p^* \mathcal{M}_B^{-1}(\alpha, q) p + \mathcal{U}(\alpha, q), \quad (3)$$

where \mathcal{M}_B denotes the $3n$ -dimensional mass matrix and \mathcal{U} the forcefield potential function. At a temperature T , the ensemble partition function $\mathcal{Z}_u(T)$ takes the form

$$\mathcal{Z}_u(T) = c_1 \int dp d\alpha dq e^{-\mathcal{H}_u(\alpha, q, p)/kT}, \quad (4)$$

where k and c_1 are the Boltzmann and scaling constants, respectively. Modulo a normalizing constant, the integrand in the partition function is the general pdf that can be used to compute the statistical averages of configuration and velocity dependent quantities.

Substituting for the Hamiltonian from Eq. (3) in Eq. (4) and integrating over the momentum coordinates, we get the

following expression:¹⁹

$$\mathcal{Z}_u(T) = c_2 \int d\alpha dq \det \{ \mathcal{M}_B^{\frac{1}{2}}(\alpha, q) \} e^{-\mathcal{U}(\alpha, q)/kT}, \quad (5)$$

where we use the shorthand $\det \{ \mathcal{M}_B^{\frac{1}{2}}(\alpha, q) \} = (\det \{ \mathcal{M}_B(\alpha, q) \})^{1/2}$. The *configuration pdf*, $\rho_u(\alpha, q)$, thus has the form

$$\rho_u(\alpha, q) \propto \det \{ \mathcal{M}_B^{\frac{1}{2}}(\alpha, q) \} e^{-\mathcal{U}(\alpha, q)/kT}. \quad (6)$$

The configuration pdf is sufficient for computing the statistical averages of quantities that do not depend on the velocity coordinates. Remarkably, $\det \{ \mathcal{M}_B \}$ does not depend on the torsions and can be expressed as the following simple product of functions of the individual BAT bond lengths and bond angles:¹⁹

$$\det \{ \mathcal{M}_B \} = \sin^2 \gamma_2 d_2^4 \prod_{i=3}^n d_i^4 \sin^2 \theta_i \prod_{i=1}^n m_i^3, \quad (7)$$

where d_i are the $(n - 1)$ bond lengths, θ_i are the $(n - 2)$ bond angles, m_i are the masses, and $(\gamma_1, \gamma_2, \gamma_3)$ are the ZXZ Euler angles for the overall orientation of the molecule. Thus, $\det \{ \mathcal{M}_B(\alpha, q) \}$ can be decomposed into factors that depend on the individual α and q BAT coordinates as follows:

$$\det \{ \mathcal{M}_B(\alpha, q) \} = f_1(\alpha) f_2(q). \quad (8)$$

For a constrained model, with the $(3n - N)$ coordinates q frozen at constant values q_0 , the partition function is

$$\mathcal{Z}_c(T) = c_3 \int d\alpha \det \{ \mathcal{M}^{\frac{1}{2}}(\alpha, q_0) \} e^{-\mathcal{U}(\alpha, q_0)/kT}, \quad (9)$$

where $\mathcal{M}(\alpha) \in \mathcal{R}^{N \times N}$ is the mass matrix for the constrained model. This yields the pdf

$$\rho_c(\alpha, q_0) \propto \det \{ \mathcal{M}^{\frac{1}{2}}(\alpha, q_0) \} e^{-\mathcal{U}(\alpha, q_0)/kT}. \quad (10)$$

In contrast with $\det \{ \mathcal{M}_B \}$ (Eq. (7)), $\det \{ \mathcal{M} \}$ cannot be decomposed into a simple product of functions of the individual bond lengths and bond angles. As a result, this $\det \{ \mathcal{M} \}$ term effectively introduces a coupling between the torsions, bond lengths, and bond angles in the ρ_c pdf. A consequence of this coupling is the introduction of a systematic bias in the statistical behavior of the constrained model. We refer to this bias as an intrinsic distortion since its source is the constrained dynamic model's mass matrix and not the external forcefield.

To compensate for this intrinsic distortion, Fixman¹ proposed the use of a modified potential function,

$$\mathcal{U}'(\alpha) = \mathcal{U}(\alpha, q_0) + \mathcal{U}_f(\alpha),$$

where

$$\mathcal{U}_f(\alpha) = \frac{1}{2} kT \ln \frac{\det \{ \mathcal{M}(\alpha, q_0) \}}{\det \{ \mathcal{M}_B(\alpha, q_0) \}} \quad (11)$$

instead of $\mathcal{U}(\alpha, q_0)$ in constrained dynamics simulations. Using Eq. (11) in Eq. (10), we get a compensated pdf of the form

$$\rho_f(\alpha, q_0) \propto \det \{ \mathcal{M}_B^{\frac{1}{2}}(\alpha, q_0) \} e^{-\mathcal{U}(\alpha, q_0)/kT}. \quad (12)$$

The \mathcal{U}_f compensating potential is known as the *Fixman potential*.

Comparing Eqs. (6) and (12), we can see that the compensated constrained pdf $\rho_f(\alpha, q_0)$ is equivalent to the unconstrained pdf $\rho_u(\alpha, q)$ under any of the following conditions:

1. There is no force potential, i.e., $\mathcal{U}(\alpha, q) = 0$. For this case, the force term drops out of Eqs. (6) and (12) so that they fully agree. Prior studies^{13–18} relating to the Fixman potential have focused on this case by ignoring the force potential.
2. The constrained coordinates q and the unconstrained coordinates α are *separable*, i.e., the total potential can be decomposed as $\mathcal{U}(\alpha, q) = \mathcal{U}_1(\alpha) + \mathcal{U}_2(q)$. For this case, $e^{-\mathcal{U}(\alpha, q)/kT}$ can be decomposed into a product of terms that depend individually on α and q , respectively. Using Eq. (8), it follows then that the right hand sides of Eqs. (6) and (12) can also be decomposed into such factors, and thus that the α and q coordinates are statistically independent so that the compensated constrained pdf $\rho_f(\alpha, q_0)$ in Eq. (12) is equivalent to the $\rho_u(\alpha, q)$ unconstrained one for the α coordinates.
3. The potential $\mathcal{U}(\alpha, q)$ is very steep for q around q_0 , such that q does not vary appreciably around q_0 in the unconstrained model. In this case, $\mathcal{U}(\alpha, q) \approx \mathcal{U}(\alpha, q_0)$ and $\det \{\mathcal{M}_B(\alpha, q)\} \approx \det \{\mathcal{M}_B(\alpha, q_0)\}$ in Eq. (12).

In general, the above conditions may not hold for $\mathcal{U}(\alpha, q)$, i.e., $\mathcal{U}(\alpha, q)$ may be non-zero, it may not be separable, and it may not be steep around a specific q_0 . In this case, the Fixman compensated constrained pdf will not agree with the unconstrained pdf. In other words, the Fixman potential alone is unable to overcome the statistical biases for such force field potentials.

We now carry out numerical simulations to verify these observations using idealized four-carbon serial chains, referred to as C4 chains,²⁰ with constrained bond lengths and angles and unconstrained torsions. We use two separate models. In Section II A 2, the force field potential is separable across the bond angle and dihedral degrees of freedom so that the ICMD model only has intrinsic and no extrinsic distortion. We verify that the Fixman potential exactly corrects for the mass matrix dependent bias introduced by the constraints in the ICMD model in the torsional probability density function and only partially in the barrier crossing rates. In Section II A 3, we use a non-separable force field potential that introduces both intrinsic and extrinsic distortions and confirm that the Fixman potential is unable to compensate for the statistical bias introduced, motivating our development of the hybrid ICMD method described later in the paper.

2. Linear C4 chain with separable degrees of freedom

Note that $\det \{\mathcal{M}_B\}$ in Eq. (7) does not depend on the torsional coordinates. Thus, when the unconstrained coordinates α are just the torsions in the molecule, as in the TMD model, we can simplify Eqs. (11) and (12) into the

forms

$$\mathcal{U}_f(\alpha) = c_f + \frac{1}{2}kT \ln \det \{\mathcal{M}(\alpha)\}$$

and

$$\rho_f(\alpha) \propto e^{-\mathcal{U}(\alpha, q_0)/kT}, \quad (13)$$

where α represents the unconstrained torsions, and c_f is a constant that consists of the bond length and bond angle contributions in $\det \{\mathcal{M}_B\}$. Under the condition that the potential $\mathcal{U}(\alpha, q)$ is separable across the torsion, and the bond length and bond angles, we expect from our earlier discussion that the Fixman potential will exactly compensate for the difference in the configuration pdf's between the unconstrained and constrained models of polymer systems.

In prior work,²⁰ we have examined the specific separable case where $\mathcal{U}(\alpha) = 0$, i.e., when the external forcefield does not depend on the torsional coordinates. For this case, the pdf for any torsion angle α_i is uniform,

$$\rho_u(\alpha_i) = \frac{1}{2\pi}. \quad (14)$$

We have demonstrated²⁰ that the application of the Fixman potential correctly compensates for any bias introduced in the torsional pdf's when the bond lengths and bond angles are treated as rigid for both serial and branched polymer systems.

a. Calculation of transition barrier crossing rates. We now look at the inability of the Fixman compensating potential to fully remove statistical biases when the function of interest also depends on velocity coordinates. For illustration, we study the use of the Fixman potential in recovering the transition barrier crossing rate statistics for a C4 system with a single barrier torsional potential $\mathcal{U}(\alpha)$, when the torsional coordinates are *separable* from the bond length and bond angle coordinates. We find that while the application of the Fixman potential corrects for the bias in the torsional pdf's, it only does so partially for the barrier-crossing rates. The results for the barrier crossing rates extend the results previously reported by Pear and Weiner.¹⁴

In the simulations for an idealized C4 serial chain in this section, we apply a single-barrier harmonic torsional potential of the form

$$\mathcal{U}(\alpha) = k_\alpha(1 + \cos(\alpha - \alpha_0)), \quad (15)$$

where $k_\alpha = 0.30$ kcal/mol is the amplitude of the potential barrier peak, and α_0 denotes the location of the peak. This potential is graphically represented for $\alpha_0 = 0$ in Figure 1. It is clearly separable since it only depends on the α torsional coordinates.

To measure the barrier-crossing rates, we apply the transition-state rate theory, which assumes that once the torsion crosses the barrier peak, it will not go backwards. For a generic one-dimensional well, this transition-state barrier-crossing rate is defined by the expression³²

$$f_{TS}(x_0) = \int_0^\infty \dot{x} \rho(x = x_0, \dot{x}) d\dot{x}, \quad (16)$$

where x is a reaction coordinate with a barrier center at $x = x_0$, and $\rho(x, \dot{x})$ is the probability density function. This expression assumes that once the reaction coordinate crosses the barrier

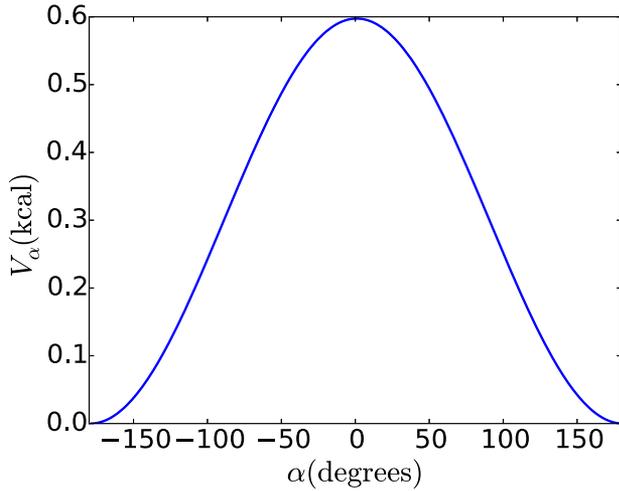


FIG. 1. Harmonic dihedral potential with $k_\alpha = 0.30$ kcal and $\alpha_0 = 0$.

peak at $x = x_0$ with positive velocity, it goes without fail to the adjacent potential well and remains there until the next transition occurs. This ignores the possibility that the reaction coordinate reverses its direction just after crossing the barrier at $\alpha = 0$, as can happen due to collisions in the system. The transition-state rate serves as a reasonable approximation for the effective barrier crossing rate in Langevin simulations with low viscosity.

For our torsional potential with the barrier peak at $\alpha = \alpha_0$, the barrier-crossing rate takes the form¹⁴

$$f_{\text{TS}}(\alpha_0) = \int_0^\infty \dot{\alpha} d\dot{\alpha} \iiint_{-\infty}^\infty d\dot{\gamma}_0 d\dot{\gamma}_1 d\dot{\gamma}_2 \rho(\alpha = \alpha_0, \dot{\gamma}, \dot{\alpha}), \quad (17)$$

where $\gamma = (\gamma_0, \gamma_1, \gamma_2)$ are the Euler angles, α is the torsion angle, and $\rho(\alpha, \dot{\gamma}, \dot{\alpha})$ denotes the probability density function.¹⁴ For the constrained model, this expression simplifies to (the Appendix)

$$f_{\text{TS,cons}}(\alpha_0) = C e^{-2k_\alpha/kT} \times [(2\pi)^3 (kT)^5 \det \{\mathcal{M}(\alpha_0)\} S^{-1}(\alpha_0)]^{1/2},$$

where

$$S^{-1}(\alpha_0) = [\mathcal{M}^{-1}(\alpha_0)]_\alpha, \quad (18)$$

where C is a normalization constant, $\mathcal{M}(\alpha)$ is the mass matrix, and $[\mathcal{M}^{-1}(\alpha_0)]_\alpha$ is the (α, α) sub-block of the $\mathcal{M}^{-1}(\alpha_0)$ matrix. We observe that the transition-state barrier crossing rate for the constrained model is a function of the mass matrix determinant. When the Fixman potential is applied, the mass matrix determinant drops out so that

$$f_{\text{TS,fix}}(\alpha_0) = C e^{-2k_\alpha/kT} [(2\pi)^3 (kT)^5 S^{-1}(\alpha_0)]^{1/2}. \quad (19)$$

While this corrected barrier crossing rate no longer dependent on the mass matrix determinant, it retains the term $S^{-1}(\alpha_0)$, which implies a remnant dependency on the location of the barrier peak—a dependency that the unconstrained model does not exhibit. Thus, while the Fixman potential fully compensates for the distortions in the pdf over configuration variables, the presence of the $S^{-1}(\alpha_0)$ term shows that such

full recovery does not occur when the averaged quantity depends on velocity coordinates.

b. Calculation of barrier crossing rates from simulations. To verify this observation from theory, we performed three types of MD simulations as follows:

- FLEXIBLE simulation with all-atom Cartesian model without any constraints;
- TMD simulation with ICMD model with bonds and angles frozen; and
- FIXMAN simulation with the TMD ICMD model with frozen bonds and angles, but with the additional Fixman potential applied.

For our simulations of the C4 system, we set the bond angles at 90° , bond lengths at 1.54 \AA , and masses at 14.01 amu . We performed simulations with the location of the barrier at $0^\circ, 45^\circ, 90^\circ, 135^\circ$, and 180° , and measured the barrier crossing rate for varying temperatures T . We performed three 20 ns Langevin dynamics simulations, each with a time step of 1 fs and with a damping constant of $0.1/\text{fs}$. For the FLEXIBLE simulations, the bond and angle spring constants were set to 303.1 kcal/\AA^2 and 63.21 kcal , respectively.

Figure 2(a) shows the results for the torsional pdf's for the C4 system with $\alpha_0 = 90^\circ$ and $T = 800 \text{ K}$, where the TMD simulation shows a bias in the torsional pdf, sufficient to alter the location of the effective barrier peak from that set by the application of the barrier potential. Only with the application of the Fixman potential did we recover the expected probability density function. Additionally, in Figure 2(b), we plot the RMS deviation from the FLEXIBLE pdf for the TMD and FIXMAN simulations, showing that the pdf's for the TMD simulation are not only different from pdf's for the FLEXIBLE simulation, but that the difference even exhibits a dependence on the location of the applied barrier peak. Again, the application of the Fixman potential removed both the introduced bias as well as the dependence on the location of the barrier peak.

Figure 2(c), however, shows that for the sample cases with barrier centers at $\alpha_0 = 90^\circ$ and $\alpha_0 = 0^\circ$, applying the Fixman potential does not sufficiently compensate for the bias introduced in the barrier crossing rates. Figure 2(d) further illustrates that, for both the TMD and FIXMAN simulations, the barrier crossing rates exhibit a dependence on the location of the barrier peak, a dependence that is independent of the applied simulation temperature. We should note that while the application of the Fixman potential did not recover the barrier-crossing rates, it did reduce the error therein, also allowing us to correctly identify the location as well as the magnitude of the potential barrier.

3. Linear chain C4 with non-separable degrees of freedom

When the torsions are cross-coupled to the bond lengths and bond angles by a non-separable potential $\mathcal{U}(\alpha, q)$, we expect that the unconstrained pdf (Eq. (6)) is equivalent to the corrected constrained pdf (Eq. (13)) only when $\mathcal{U}(\alpha, q)$ forms a very steep potential well around a value q_0 . In such a stiff system, $\mathcal{U}(\alpha, q)$ is effectively $\mathcal{U}(\alpha, q_0)$, and $\det \{\mathcal{M}_B(\alpha, q)\}$ is effectively $\det \{\mathcal{M}_B(\alpha, q_0)\}$. In the absence of such steep

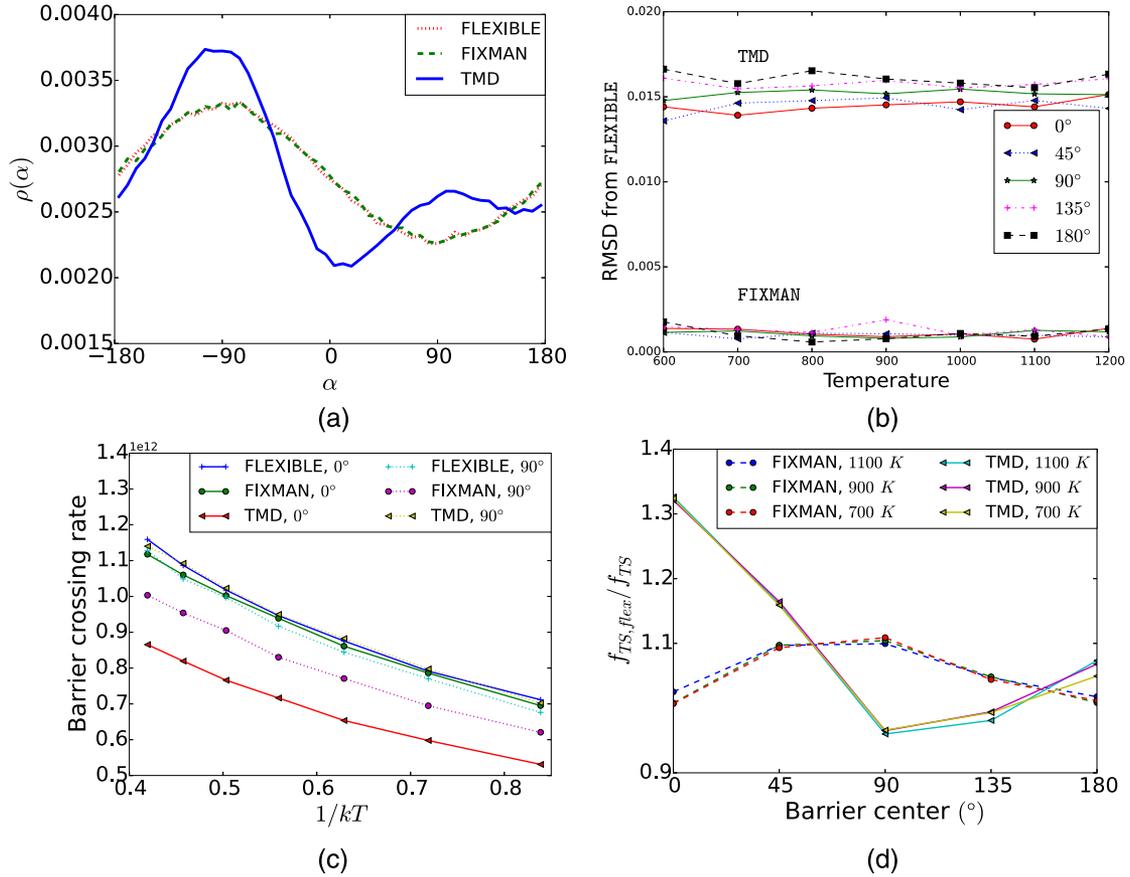


FIG. 2. (a) Torsional probability density function for C4 for a harmonic torsional potential (Eq. (15)) with the barrier center at $\alpha_0 = 90^\circ$ and $T = 800$ K. (b) RMS difference from the FLEXIBLE pdf for the TMD and FIXMAN simulations for various temperatures. (c) Transition state barrier crossing rates for C4 at $\alpha_0 = 90^\circ$ and $\alpha_0 = 180^\circ$. (d) Ratio of the FLEXIBLE barrier crossing rates to the TMD and FIXMAN barrier crossing rates as a function of the location of the barrier peak.

form of the potential function, q varies significantly around q_0 , and the Fixman compensated constrained pdf will differ from the unconstrained pdf. The differences between these pdf's arise from the intrinsic and extrinsic distortions introduced by the use of constraints.

In this section, we apply a Coulombic potential that couples the torsions and the bond angles to an idealized C4 serial chain and show that only when $\mathcal{U}(\alpha, q)$ is very steep is the Fixman potential alone sufficient to eliminate the bias introduced by freezing the bond angles. We verify that, for realistic values of $\mathcal{U}(\alpha, q)$, the extrinsic distortion dominates, and the cross-coupling significantly alters the equilibrium statistics for the torsions.

a. Simulation results. To couple the torsion to the bond angles in an idealized C4 serial chain, we added Coulombic charges q_1 and q_2 at the ends of the chain, giving us a potential of the form

$$\mathcal{U}_{\text{coul}} = \frac{1}{2} k_{\text{coul}} \frac{q_1 q_2}{r}, \quad (20)$$

where r is the distance between the two beads, and $k_{\text{coul}} = 332.06 \text{ kcal}/\text{\AA}^2$ is the force constant. The distance r is a function of both the torsion and the bond angles in such a way that the potential is non-separable. When the charges q_1 and q_2 are of opposite signs, $\mathcal{U}_{\text{coul}}$ is an attractive potential, and $\rho(\alpha)$ has the maxima at $\alpha = 0^\circ$.

As before, in the unconstrained model, we also apply harmonic spring potentials of the form

$$\mathcal{U}_\theta = k_\theta (\theta - \theta_0)^2, \quad (21)$$

where θ are the bond angles and k_θ is the angle spring constant. The k_θ spring constant can be used to vary the stiffness of this potential. If the potential is sufficiently steep, we expect the Fixman corrected torsional pdf for the constrained model to recover the pdf for the unconstrained model.

By varying the value of k_θ , we have analyzed the effect of stiffness on the extent to which the Fixman potential compensates for the bias in the torsional pdf. To quantify the difference between the constrained and unconstrained pdf's, we have calculated the root mean square deviation between these pdf's. We denote the RMS deviation between the unconstrained and the constrained pdf's as R_{TMD} , and the RMS difference between the unconstrained and the corrected (Fixman) constrained pdf's as R_{FIX} .

We applied Coulombic charges of magnitudes $0.2e$ and $-0.2e$ to the terminal atoms and used bond lengths of 1.54 \AA , bond angles at 90° , and 14.01 amu atom masses. We varied the value of the spring constant k_θ and performed three unconstrained, constrained, and corrected (Fixman) simulations for each value of k_θ , obtaining the final results by averaging from each set of simulations.

With a Coulombic potential applied, we observed that the Fixman potential removed the bias from the torsional pdf of the C4 system only for high (i.e., very stiff) values of k_θ . For values of k_θ from typical all-atom forcefields such as AMBER (30-100 kcal for AMBER99SB³³), the torsional pdf in FLEXIBLE simulations differed from that in the TMD simulations, even when the Fixman potential is included (Figure 3(a)). The RMS difference between the FLEXIBLE and FIXMAN simulations, R_{FIX} , is shown in Figure 3(b), where we see that the FLEXIBLE and FIXMAN simulations agree only for large values of k_θ . The extent to which the FLEXIBLE pdf differs from the FIXMAN simulations for realistic values of k_θ demonstrates that, for these systems, the cross-coupling between the torsional and angle degrees of freedom is not negligible. In such systems, for the configuration pdf's, applying the Fixman potential compensates for the intrinsic distortion, but not for the extrinsic distortion that the constraints introduce in the potential energy landscape.

B. Hybrid ICMD to correct for extrinsic distortions

All-atom force-fields used widely in MD simulations typically take the form

$$U_{\text{ff}} = \frac{1}{2} \sum_{i=1}^{N_{\text{bonds}}} K_{r_i} (r_i - r_i^0)^2 + \frac{1}{2} \sum_{i=1}^{N_{\text{angles}}} K_{\theta_i} (\theta_i - \theta_i^0)^2 + U_{\text{ff}}^{\text{tors}}(\alpha_i) + U_{\text{ff}}^{\text{long-range}}, \quad (22)$$

where K_{r_i} and K_{θ_i} are the harmonic potentials, also referred to as the *restraining* potentials for the bonds and angles, respectively. $U_{\text{ff}}^{\text{tors}}$ comprises the dihedral potentials, and $U_{\text{ff}}^{\text{long-range}}$ comprises the long-range Coulombic and van der Waal's interactions and the implicit solvent potential for implicit solvent simulations. Here, each long-range potential term is a function of the distance between a given pair of atoms, with the distance itself dependent on the bond lengths, bond angles, and the torsional angles. This introduces coupling among the degrees of freedom.

It is customary to assume that the harmonic restraining terms on the bonds and the angles dominate over the other interactions, and hence, the potential at any point in the torsional subspace is independent of the bond and angle

values.⁴ When this assumption is valid, the application of the Fixman potential is sufficient to eliminate the bias introduced by constraining the bond and angle degrees of freedom. Echenique *et al.*,^{34,35} however, showed that for realistic force fields, there is significant coupling between the torsion and bond angle degrees of freedom and that the harmonic restraining terms do not dominate over the non-harmonic interactions. In this event, the forcefield is non-separable and the statistics for the torsions are not a function of the torsional angles alone, but also depend on the bond angle degrees of freedom. Constraining the bond angles introduces not only an intrinsic distortion but also an extrinsic distortion, an effect which the Fixman potential does not eliminate.

One of the approaches for reducing the barrier heights arising from the extrinsic distortion has been to alter the torsional and nonbonded potential functions in the forcefield.⁹⁻¹² Additionally, Arnautova *et al.* have altered the forcefield terms while also keeping some backbone bond angles open.⁸ These approaches are system and forcefield dependent and not easy to generalize across polymer applications.

We now present the “hybrid internal coordinates molecular dynamics” method, where any desired subset of bond angles can be treated as flexible (instead of rigid) at any point during the simulation, while keeping the bonds and remaining bond angles rigid. Such a hybrid method bridges the gap between all-atom flexible simulations and the torsional MD simulations. It allows the user to keep all bond angles open or selectively keep only some of the bond angles open. We have implemented the hybrid ICMD method in the *GneimoSim* software.

In a previous analysis of protein structures in the protein data bank, Berkholtz *et al.*³⁶ have shown that the protein backbone covalent geometry is a function of the backbone torsions. Additionally, Hinsen *et al.*³⁷ have shown that the inclusion of a subset of the backbone angles in the unconstrained coordinate set is sufficient to accurately represent the protein structure. In developing the hybrid ICMD method, we postulate that opening only specific backbone bond angles sufficiently compensates for the extrinsic distortion and retrieves potential energy landscapes similar to that in unconstrained simulations.

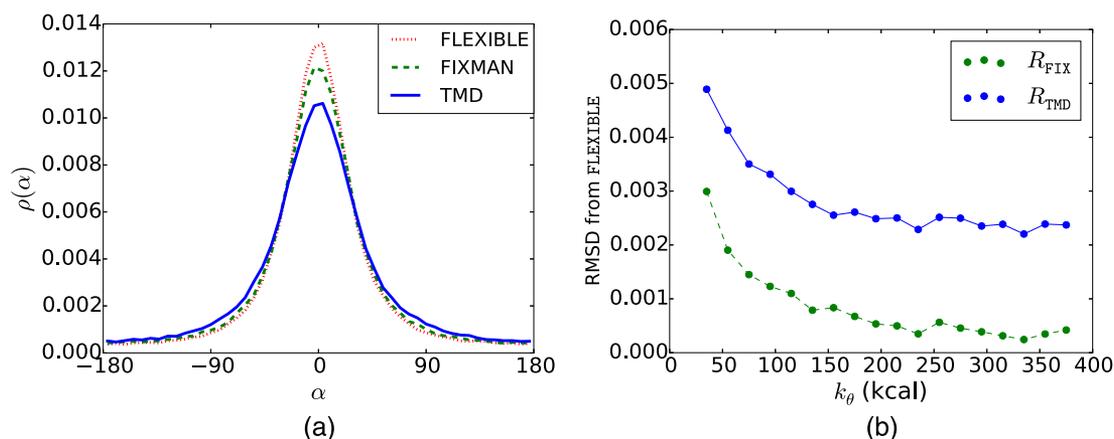


FIG. 3. (a) Torsional pdf's for C4, with charges 0.2e and $-0.2e$ at the terminal atoms, for $k_\theta = 95$ kcal. (b) RMS deviation from the FLEXIBLE pdf as a function of k_θ . The FIXMAN and FLEXIBLE simulations agree only at very high values of k_θ . The TMD simulation remains divergent throughout.

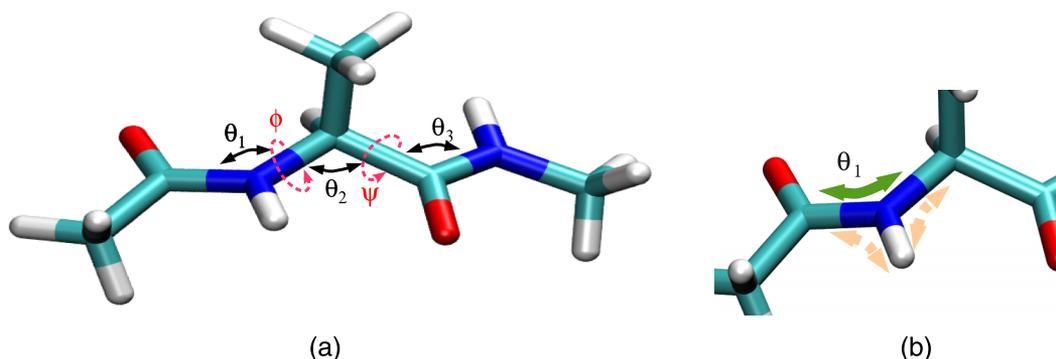


FIG. 4. (a) Structure of alanine dipeptide. The torsions $\phi = \text{C}-\text{N}-\text{C}_\alpha-\text{C}$ and $\psi = \text{N}-\text{C}_\alpha-\text{C}-\text{N}$ effectively represent the conformational space explored. (b) Schema for opening a bond angle θ_1 —we open the two terminal angles (dashed) to effectively open the non-terminal (solid) angle.

III. SIMULATION RESULTS FOR THE HYBRID ICMD METHOD

In this section, we apply the hybrid ICMD method to simulations of small peptide systems with all-atom forcefields and present the probability density functions thus calculated. Our analyses here provide guidelines as to which backbone angles should be free in ICMD simulations to facilitate dihedral transitions. We demonstrate that, for non-separable potential functions, the introduction of constraints leads to distortions in the free energy surface (FES). In particular, the long-range forces couple the torsions to the bond lengths and the bond angles, introducing a distortion that cannot be compensated for by applying the Fixman potential alone. However, when we use the hybrid ICMD method in conjunction with the Fixman potential, we find that opening just a few key bond angles is sufficient to remove these distortions from the FES. In addition, in Section III D, we examine the size of the time steps possible for hybrid ICMD applications and find that, even with these key bond angles open, we continue to get stable simulations with time steps of up to 5 fs.

A. Application of hybrid ICMD to alanine dipeptide

To comprehensively investigate how the introduction of the Fixman potential affects the conformational landscape of a dipeptide molecule, we performed TMD, FIXMAN, and FLEXIBLE simulations of alanine dipeptide under a variety of different conditions. For the TMD and FIXMAN simulations, we modeled the alanine dipeptide molecule as a collection of rigid body clusters, connected by hinges. In this scheme, a general peptide molecule is divided into clusters such that each cluster is composed of a non-terminal atom along with all of its terminal neighbors. Aromatic ring moieties in the side chains are treated as rigid clusters. Proline rings in the main chain are broken into tree structures while using the stiff harmonic forcefield valence bond parameters to keep the ring together. Disulfide bonds are also broken while again using the harmonic forces to keep the atoms together.

We used the *GneimoSim*²⁵ software with forcefield parameters from the AMBER99SB forcefield³³ and the Generalized Born/Surface Area (GBSA) solvation method.

The conformations of an alanine dipeptide molecule can be described by the two main chain dihedrals, namely,

the ϕ and ψ angles, defined as $\phi = \text{C}-\text{N}-\text{C}_\alpha-\text{C}$ and $\psi = \text{N}-\text{C}_\alpha-\text{C}-\text{N}$, shown in Figure 4(a). The FES of alanine dipeptide can be conveniently projected onto these two dihedrals, giving us an effective method to analyze the conformational ensembles in the equilibrium sampling of this molecule. To study the effect of constraints on this FES, we performed 20 FLEXIBLE and TMD simulations each, at 300 K and 800 K, with each simulation 20 ns long.

Additionally, to study how to alleviate the intrinsic and extrinsic distortions introduced in this system, we first performed FIXMAN simulations at 300 K, then systematically opened the various bond angles shown in Figure 4(a) and performed 20 simulations of 20 ns length each for each of the backbone angles and their various combinations. All the simulations were performed starting from an initial alpha-helical conformation ($(\phi, \psi) = (-60^\circ, -40^\circ)$), using Langevin dynamics with a damping constant of 0.1/fs and a time step of 2 fs.

1. Results and discussion

Figure 5 shows the FES for the FLEXIBLE and TMD simulations at 300 K and 800 K. It is evident from the figure

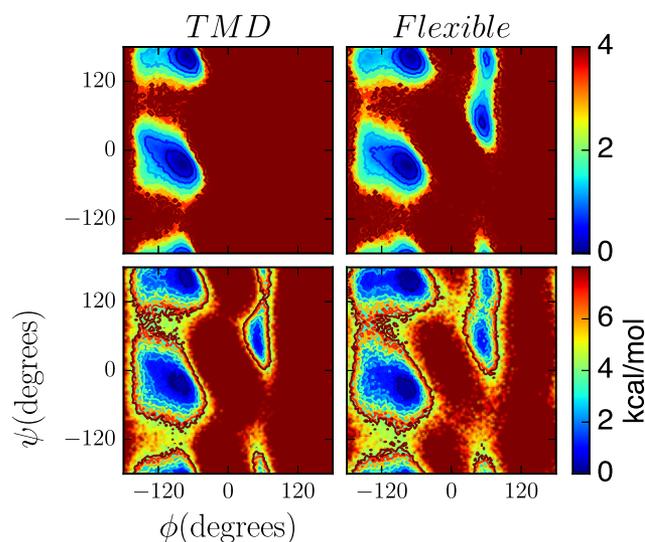


FIG. 5. Free energy surface calculated for the backbone dihedral angles of alanine dipeptide: FLEXIBLE and TMD simulations at 300 K (top) and 800 K (bottom).

that, at 300 K, the TMD simulations lead to barriers that limit the sampling of the minima in the first and fourth quadrants of the FES, with the quadrants numbered counterclockwise with the first quadrant in the upper right. These barriers were overcome somewhat at 800 K, but the sampling in TMD simulations is still limited in comparison to the FLEXIBLE simulations. When we additionally applied the Fixman potential, we found that the alanine dipeptide simulation again does not sample the minima in the first and fourth quadrants of the FES (Figure 6). This indicates that the cross-coupling between the torsional and angle degrees of freedom significantly affects the FES of an alanine dipeptide molecule, an extrinsic bias that the Fixman potential is unable to compensate for.

Applying the hybrid ICMD model to alanine dipeptide, we found that opening the backbone angles $C-N-C_\alpha$ and $N-C_\alpha-C$ together recovers the features of the FLEXIBLE simulations as shown in Figure 6. Opening these angles together was sufficient to remove the barriers introduced by freezing the bond angle degrees of freedom. Opening additional angles had little effect on reducing the extrinsic distortion in the system. Opening just one of the two backbone bond angles shown in Figure 4(a) did not alleviate the transition barriers stemming from the extrinsic distortion in the system either. Thus, allowing the backbone angles to change with the hybrid ICMD method overcomes the effect of the cross-coupling in the FES.

B. Application of hybrid ICMD to other dipeptides

We have so far demonstrated that the extrinsic bias in the FES resulting from the cross-coupling between the torsional and the bond angle degrees of freedom in the ICMD simulations of alanine-dipeptide is eliminated by opening up backbone bond angles during the ICMD simulations. In this section, we apply the hybrid ICMD method to valine, leucine, isoleucine, methionine, phenylalanine, tryptophan, proline, and tyrosine dipeptides (Fig. 7). We study how the opening of various bond angles, individually and in combination, affects the FES. We thereby identify the bond angles that are most coupled to the torsions for each of these dipeptide molecules. This allows us to construct a minimal subset of angles that we need to open to sufficiently remove the barriers introduced by

the extrinsic distortion and to allow constrained simulations of these dipeptide molecules to sample the same conformational space as do unconstrained simulations.

To establish how each bond angle in a dipeptide affects its conformational sampling, we performed separate simulations with each backbone bond angle open in the dipeptide. In addition, we also tested how the various combinations of these bond angles each affected the conformational sampling of the dipeptides. We performed 20 simulations of length 20 ns for each such system, using the clustering and simulation methods as described in Section III A. All of the constrained simulations were performed from an initial alpha-helical conformation ($(\phi, \psi) = (-60^\circ, -40^\circ)$), with the Fixman potential enabled.

1. Results and discussion

The free energy surfaces calculated from the dipeptide simulations are shown in Figure 8(a). It is seen that applying the Fixman potential with hybrid ICMD model with open backbone angles $C-N-C_\alpha$ and $N-C_\alpha-C$ allows the valine, leucine, methionine, phenylalanine, tryptophan, and tyrosine simulations to sample the conformations in the first quadrant.

To quantify how the FES obtained for the FIXMAN and hybrid ICMD models differs from that of the FLEXIBLE model, we calculated the Hellinger distance between the FIXMAN and FLEXIBLE probability density functions and also between the hybrid ICMD and FLEXIBLE probability density functions (Table S1 of the supplementary material³⁸). The Hellinger distances measuring the quantitative differences in the conformational distributions for various dipeptides are shown in Figure 8(b). Using this metric, we found that opening any additional backbone angles does not significantly improve the probability density functions for the PHE, ALA, LEU, and TRP dipeptides. For ILE, VAL, MET, and TYR, however, we needed to additionally open the $C_\alpha-C-N$ backbone angle to obtain maximal improvement in the hybrid ICMD probability density function (Figure 9). Proline dipeptide (Figure 7(g)) presents a special case since it is a ring structure and hence is a loop structure. As described previously, we break the ring structure to obtain a tree topology for the FIXMAN and hybrid ICMD simulations. For proline, we found that we need to open up the backbone $N-C_\alpha-C$ bond angle as well as the sidechain

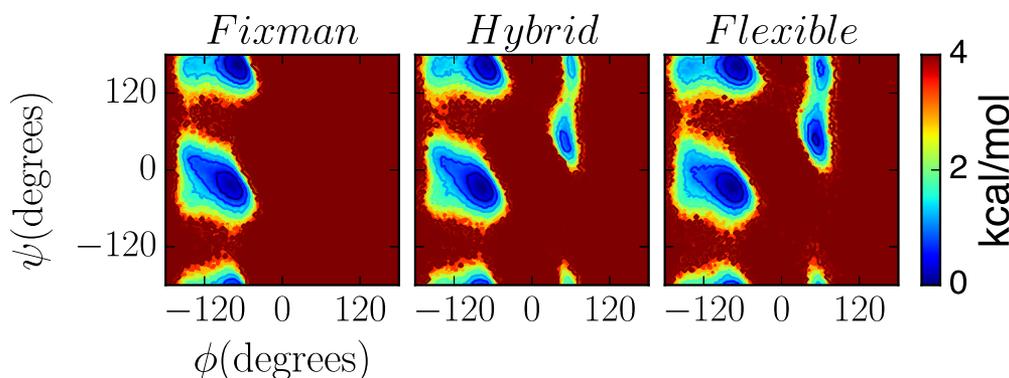


FIG. 6. FES for FIXMAN, hybrid ICMD, and FLEXIBLE simulations of alanine dipeptide. The hybrid ICMD simulation has the bond angles $C-N-C_\alpha$ and $N-C_\alpha-C$ open.

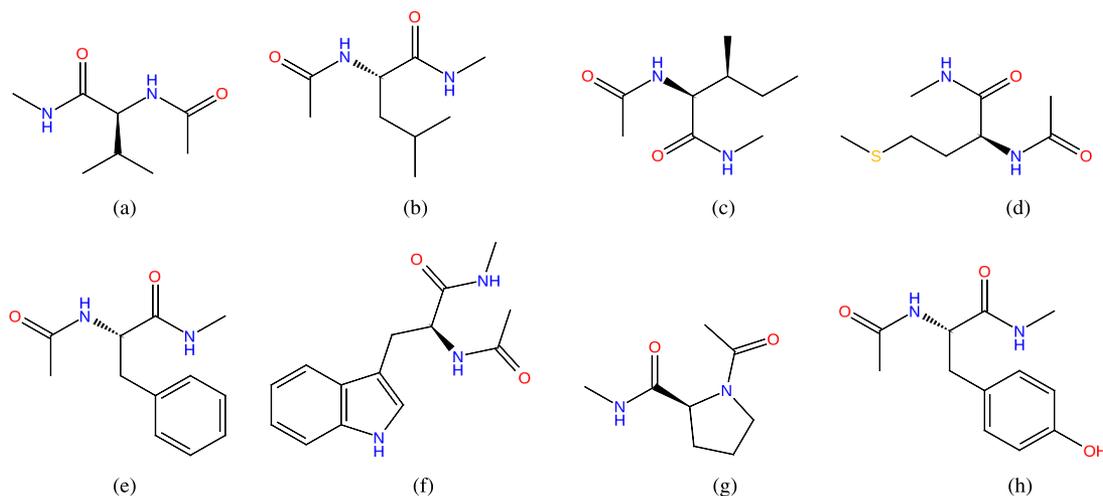
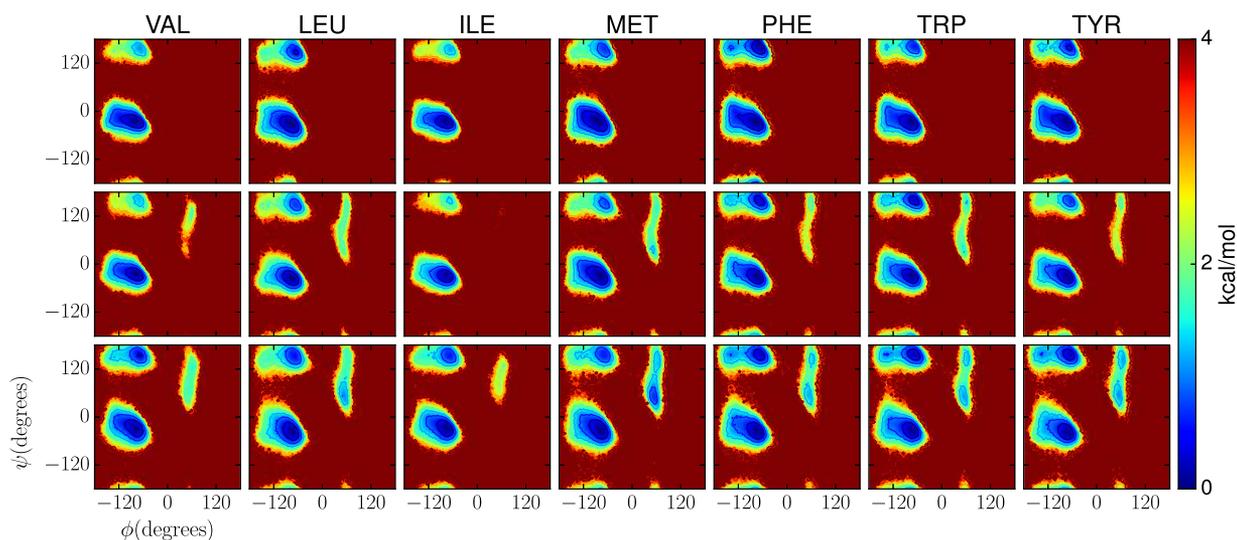


FIG. 7. Structures of (a) valine (VAL) dipeptide, (b) leucine (LEU) dipeptide, (c) isoleucine (ILE) dipeptide, (d) methionine (MET) dipeptide, (e) phenylalanine (PHE) dipeptide, (f) tryptophan (TRP) dipeptide, (g) proline (PRO) dipeptide, and (h) tyrosine (TYR) dipeptide.

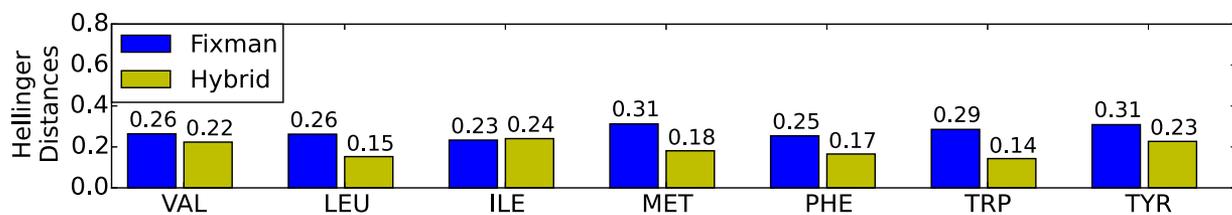
C_{α} - C_{β} - C_{γ} bond angle to get the minimal Hellinger distance of 0.26 as shown in Figure 10. For all the dipeptides, opening bond angles other than the specified angles had little effect on the barrier heights and conformational sampling.

In the literature, Arnautova *et al.*⁸ have used loop modeling simulations with a modified forcefield to argue that it is sufficient to open just the $N-C_{\alpha}-C$ angle to

effectively sample the conformational space of a protein system. Similarly, Hinsen *et al.*³⁷ have analyzed reduced coordinate sets for proteins, concluding that including only an additional angle centered at C_{α} is sufficient to accurately represent the protein conformations. In our simulations, we found that opening of only the $N-C_{\alpha}-C$ angle produces simulations with probability density functions quite distant



(a)



(b)

FIG. 8. (a) Free energy surfaces for FIXMAN simulations (top), hybrid ICMD simulations (mid), and FLEXIBLE simulations (bottom) for dipeptide molecules. (b) Hellinger distances from the FLEXIBLE probability density functions for the FIXMAN and hybrid ICMD simulations. For VAL, LEU, MET, PHE, TRP, and TYR, the FIXMAN simulations are unable to sample the conformations in the first quadrant, the sampling of which requires that the $C-N-C_{\alpha}$ and $N-C_{\alpha}-C$ angles be kept open in the hybrid ICMD simulations. For ILE, opening the $C-N-C_{\alpha}$ and $N-C_{\alpha}-C$ angles does not improve the free energy surface.

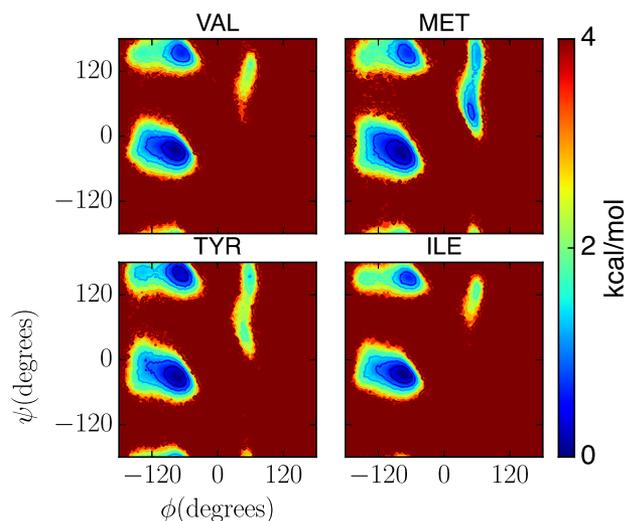


FIG. 9. Free energy surfaces for hybrid ICMD simulations with the open backbone angles $C-N-C_{\alpha}$, $N-C_{\alpha}-C$, and $C_{\alpha}-C-N$ for valine (VAL), methionine (MET), tyrosine (TYR), and isoleucine (ILE) dipeptides. With these angles open, the Hellinger distances to the FLEXIBLE probability density function are now 0.14, 0.13, 0.17, and 0.16, respectively.

from that for the FLEXIBLE simulations (Table S1 of the supplementary material³⁸). In all cases but proline dipeptide, we needed to open at least the additional angle $C-N-C_{\alpha}$ to significantly reduce the observed extrinsic bias. For valine, isoleucine, methionine, and tyrosine dipeptides, we needed to additionally open the third backbone angle $C_{\alpha}-C-N$ to maximally compensate for the extrinsic bias. For proline dipeptide, we needed to open the backbone $N-C_{\alpha}-C$ angle as well as the sidechain $C_{\alpha}-C_{\beta}-C_{\gamma}$ angle for optimum compensation.

C. Application of hybrid ICMD to longer peptide chains

While we have demonstrated that opening select bond angles alleviates the extrinsic bias in small dipeptide simulations, it is not clear whether the same effect holds for simulations of longer peptides which can have long-range couplings between atoms far apart in the chain. In this section, we apply the hybrid ICMD method to simulate a

ten amino acid peptide CLN025 (Figure 11) for which there is an ensemble of NMR structures available.³⁹ Additionally, there are NMR data on the possible hydrogen bonds present in this system. We compared the conformation sampling thus obtained in FIXMAN and hybrid ICMD simulations to experimental data obtained from NMR studies as well as to the conformational sampling in FLEXIBLE simulations. Through this study, we establish that using the hybrid ICMD method allows for better agreement with both the experimental data and the FLEXIBLE conformational sampling.

To examine whether the backbone bond angles introduce extrinsic bias in simulations of the CLN025 peptide, we performed eight simulations of 10 ns length each for FIXMAN, hybrid ICMD, and FLEXIBLE models. For the hybrid ICMD simulations, we opened the $N-C_{\alpha}-C$ angle for the proline residue, and the $C-N-C_{\alpha}-C$ and $N-C_{\alpha}-C$ backbone angles for the other residues. The simulations were performed at a temperature of 300 K maintained using the Nose-Hoover thermostat, using a time step of 1 fs. We used the GBSA solvation method with an internal dielectric constant of 4.0 and external dielectric constant of 78.0. All the simulations used the identical starting conformation shown in Figure 11, with the conformation obtained from the NMR structure with the PDB ID 2RVD.³⁹

1. Results and discussion

In a previous experiment, Honda *et al.*³⁹ have studied the NMR spectra of CLN025 in solution and have identified the hydrogen bonds observed in the NMR structures. In Table I, we evaluated the conformational sampling efficiency of the hybrid ICMD method by calculating the percentage of the snapshots in the simulations that have the hydrogen bond distances observed in the NMR structures within a standard deviation of the mean distance reported. We found that the only a very small percentage of the FIXMAN conformations agree with observed NMR distances. When we opened the backbone bond angles in the hybrid ICMD simulations, we observed that a much larger proportion of the conformations showed hydrogen bond distances similar to those in the NMR structural ensemble.

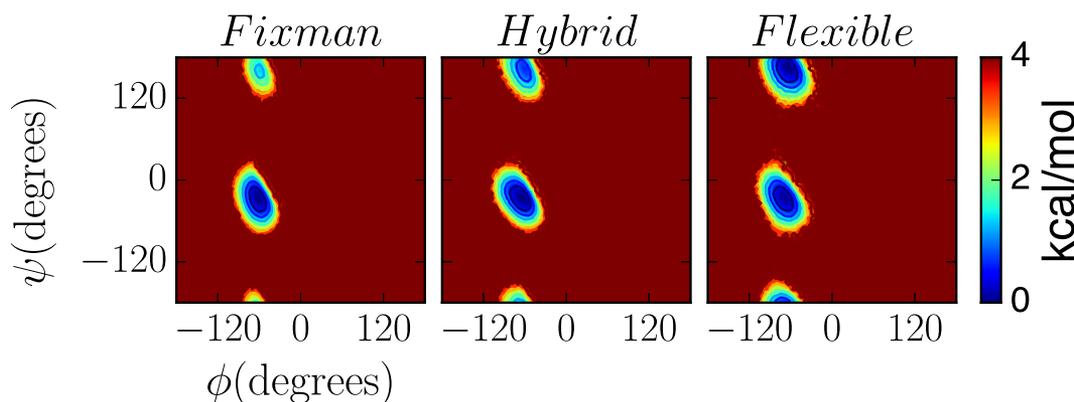


FIG. 10. Free energy surfaces for FIXMAN, hybrid ICMD, and FLEXIBLE simulations for proline dipeptide (PRO). The FIXMAN probability density function has a Hellinger distance of 0.45 from the FLEXIBLE density function. The hybrid ICMD simulation with the backbone angle $N-C_{\alpha}-C$ and the sidechain angle $C_{\alpha}-C_{\beta}-C_{\gamma}$ relaxed yields a Hellinger distance of 0.26 from the FLEXIBLE simulation.

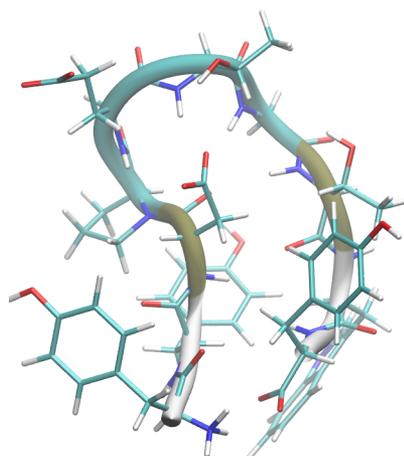


FIG. 11. Starting structure for the simulations of the 10-residue peptide CLN025. The PDB ID for the NMR structure of CLN025 is 2RVD.

Additionally, we also calculated the backbone Ramachandran (ϕ, ψ) probability density function for the FIXMAN, hybrid ICMD, and FLEXIBLE models (Figure S1 of the supplementary material³⁸), where we observed that the hybrid ICMD simulations show greater conformational sampling than the FIXMAN simulations. We used these distributions to calculate the per residue Hellinger distances between the FIXMAN and FLEXIBLE, and the hybrid ICMD and FLEXIBLE simulations. Figure 12 shows the Hellinger distances thus obtained—the hybrid ICMD simulations generally have Hellinger distances much lower than do the FIXMAN simulations.

The observed agreement of hybrid ICMD conformations with NMR data, which the FIXMAN conformations do not demonstrate, indicates that when all the bond angles are fixed, there is an extrinsic distortion introduced that can limit the conformational sampling in the FIXMAN simulation. By opening only the backbone bond angles, we can significantly ameliorate this extrinsic distortion, leading to better agreement with the NMR spectra. This also has the effect of increasing the conformational sampling, producing Ramachandran distributions closer to that observed in the FLEXIBLE simulations.

D. Time step size for the hybrid ICMD method

We have demonstrated that opening select bond angles can help alleviate the extrinsic rigidity introduced in TMD

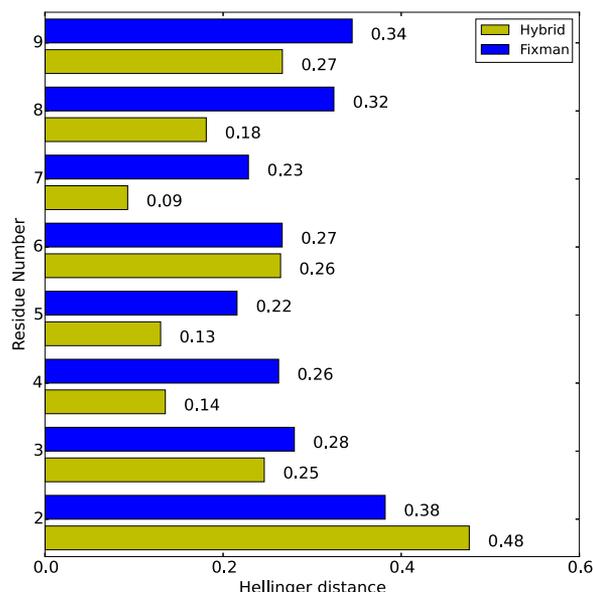


FIG. 12. Hellinger distances from the FLEXIBLE (ϕ, ψ) probability density functions for the FIXMAN and hybrid ICMD simulations for each of the residues in CLN025.

simulations. However, opening the angle degrees of freedom can potentially require smaller integration time step size. To examine this, we performed hybrid ICMD simulations of the dipeptide molecules with the angles $N-C_{\alpha}-C$ and $C-N-C_{\alpha}$ open for alanine, phenylalanine, leucine, and tryptophan dipeptides, and with the additional $C_{\alpha}-C-N$ angle open for isoleucine, valine, methionine, and tyrosine dipeptides. For proline dipeptide, we opened the $N-C_{\alpha}-C$ backbone angle and the $C_{\alpha}-C_{\beta}-C_{\gamma}$ sidechain angle. We used time steps of 5 fs, with 20 simulations of length 50 ns for each angle.

Our results in Figure 13 show that opening up the backbone bond angles in a protein even while performing coarse grain ICMD simulations can effectively alleviate the distortion in the free energy surface, while still allowing the use of the significantly higher 5 fs time step than is possible with all-atom Cartesian simulations. Since the ICMD model can be easily coarsened by treating different parts of the protein as rigid while treating the other parts as flexible with backbone dihedral angles open, the ICMD method provides a wide range of dynamic models that can be used for simulating proteins.

TABLE I. Percentage of conformations in the FIXMAN, hybrid ICMD, and FLEXIBLE simulations that are found to be within a standard deviation of the mean NMR values for hydrogen bonds observed in NMR experiments.

Hydrogen bond	NMR mean	NMR std	% of FIXMAN confs.	% of hybrid ICMD confs.	% of FLEXIBLE confs.
Thr8.O–Tyr10.N	3.62	0.25	58.46	22.46	8.32
Asp3.OD1–Thr6.N	3.38	0.14	0.55	7.27	10.25
Asp3.O–Thr8.N	2.89	0.16	9.78	65.51	57.60
Asp3.O–Gly7.N	3.56	0.20	0.18	31.20	25.29
Asp3.O–Thr6.N	5.58	1.14	0.19	45.98	43.62
Asp3.O–Thr8.O	3.51	0.42	1.16	0.41	0.49

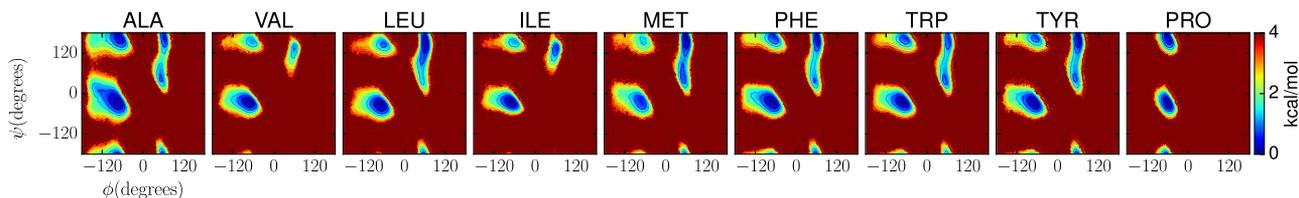


FIG. 13. Free energy surfaces for hybrid ICMD simulations with open backbone angles $C-N-C_\alpha$ and $N-C_\alpha-C$ for alanine, leucine, phenylalanine, and tryptophan, with the additional angle $C_\alpha-C-N$ open for valine, isoleucine, methionine, and tyrosine dipeptides. For the proline dipeptide, the $N-C_\alpha-C$ and $C_\alpha-C_\beta-C_\gamma$ are kept open. The simulations use time steps of 5 fs each, with each simulation 50 ns long.

IV. CONCLUSIONS AND FUTURE WORK

Distortions in the potential energy surface for constrained ICMD models stem from both the dependence of the mass matrix on the BAT coordinates and the forcefield dependent coupling between the unconstrained and constrained degrees of freedom. The distortion introduced by the mass matrix is intrinsic to the constrained model. In this study, we have shown that the application of the Fixman potential alleviates the intrinsic distortion in the configuration probability density function but is only able to partially overcome the biases for quantities such as the barrier crossing rate that depends on the velocity coordinates.

Additionally, we have tested the effect of the Fixman potential in constrained ICMD simulations with all-atom forcefields, where we have shown that the long range non-bonded forces couple the constrained and unconstrained degrees of freedom, introducing an extrinsic distortion to the potential energy landscape, one that cannot be compensated for by only the Fixman potential. To alleviate the extrinsic distortion, we have developed the hybrid ICMD simulation method that allows the user to open up any bond angle degree of freedom within the ICMD model. We have demonstrated that for short peptide chains, the use of the Fixman potential in conjunction with the opening a small subset of the backbone angles via the hybrid ICMD method is sufficient to remove the extrinsic distortions from the potential energy landscape, while still allowing for the use of large time step sizes. We have also shown the effectiveness of the hybrid ICMD method in recovering the free energy surface of a complex polypeptide section. We believe that the hybrid ICMD method represents a robust and accurate ICMD simulation method that bridges the gap between all-atom and torsional MD models. The hybrid ICMD method has been implemented within the GneimoSim software. With the hybrid ICMD model in the toolkit, the GNEIMO method now supports dynamics models at multiple levels of coarsening—from all open angles to arbitrarily large rigid domains, all within the same simulation. For example, it is possible to have simulations in which the mostly static helical secondary structures are kept rigid while the very dynamic loops have open backbone angles, allowing for fast and accurate sampling of the protein conformational space.

Much remains to be done in validating and testing the performance of hybrid ICMD for larger proteins in explicit solvent simulations and also in extending the applicability of this method to other systems such as nucleic acids. Also, since the standard Fixman potential only fully corrects statistical biases stemming from the intrinsic distortions for configuration dependent quantities, we plan to investigate

extensions of the Fixman potential that will overcome such biases for velocity dependent quantities such as the barrier crossing rates.

ACKNOWLEDGMENTS

This work was supported by Grant No. RO1GM082896 from the National Institutes of Health. The research described in this paper was also performed in part at the Jet Propulsion Laboratory (JPL), California Institute of Technology, under contract with the National Aeronautics and Space Administration. Government sponsorship is acknowledged.

APPENDIX: CALCULATING THE BARRIER-CROSSING RATE

For a C4 system, from Eq. (17), the transition-state barrier-crossing rate takes the form¹⁴

$$f_{TS}(\alpha_0) = \int_0^\infty \dot{\alpha} d\dot{\alpha} \iiint_{-\infty}^\infty d\dot{\gamma}_0 d\dot{\gamma}_1 d\dot{\gamma}_2 \rho(\alpha = \alpha_0, \dot{\alpha}, \dot{\gamma}), \quad (A1)$$

with the torsion angle α , the Euler angles $\gamma = (\gamma_0, \gamma_1, \gamma_2)$, and with the probability density function $\rho(\alpha, \dot{\alpha}, \dot{\gamma})$. For the constrained model, the pdf takes the form,¹⁴

$$\rho(\alpha, \dot{\alpha}, \dot{\gamma}) = C \det \{ \mathcal{M}(\alpha) \} e^{-[E_k(\alpha, \dot{\alpha}, \dot{\gamma}) + \mathcal{U}(\alpha)]/kT}, \quad (A2)$$

with the normalization constant C , the potential energy $\mathcal{U}(\alpha)$ (Eq. (15)), and the kinetic energy $E_k(\alpha, \dot{\alpha}, \dot{\gamma})$ which is given by

$$E_k(\alpha, \dot{\alpha}, \dot{\gamma}) = \frac{1}{2} [\dot{\alpha}^*, \dot{\gamma}^*] \mathcal{M}(\alpha) \begin{bmatrix} \dot{\alpha} \\ \dot{\gamma} \end{bmatrix}, \quad (A3)$$

where $\mathcal{M}(\alpha)$ is the mass matrix. The $\det \{ \mathcal{M}(\alpha) \}$ term in Eq. (A2) arises from the change from the momentum to velocity coordinates in the pdf.

In general, for α and γ coordinates of dimensions m and n , respectively, we can partition the matrix $\mathcal{M}(\alpha)$ as

$$\mathcal{M}(\alpha) = \begin{bmatrix} \overbrace{S_0}^m & \overbrace{V}^n \\ \overbrace{V^*}^m & \overbrace{W_0}^n \end{bmatrix}. \quad (A4)$$

With $S = S_0 - VW_0^{-1}V^*$ and $W = W_0 - V^*S_0^{-1}V$ Schur complement matrices, we have

$$\mathcal{M}^{-1}(\alpha) = \begin{bmatrix} S^{-1} & -S_0^{-1}VW^{-1} \\ -W_0^{-1}V^*S^{-1} & W^{-1} \end{bmatrix}$$

and

$$\det \{\mathcal{M}\} = \det \{W_0\} \det \{S\}. \quad (\text{A5})$$

Note that the $S^{-1}(\alpha')$ is the α coordinates square sub-block of the $\mathcal{M}^{-1}(\alpha')$ matrix, i.e.,

$$S^{-1}(\alpha') = [\mathcal{M}^{-1}(\alpha')]_{\alpha}. \quad (\text{A6})$$

Eq. (A3) can be re-expressed in the following decomposed form:

$$E_k(\alpha, \dot{\alpha}, \dot{\gamma}) = \frac{1}{2} \dot{\alpha}^* S \dot{\alpha} + \frac{1}{2} (\dot{\gamma} - \beta)^* W_0 (\dot{\gamma} - \beta),$$

where

$$\beta = -W_0^{-1} V^* \dot{\alpha}. \quad (\text{A7})$$

For the C4 case, α consists of only the torsion angle and is a scalar, and γ consists of the three Euler angles. In this case, S is a scalar, and $\det \{S\} = S$. Substituting Eqs. (A2)

and (A7) in Eq. (A1), with $\alpha = \alpha_0$, we get the expression

$$f_{\text{TS}}(\alpha_0) = C \det \{\mathcal{M}(\alpha_0)\} e^{-\mathcal{U}(\alpha_0)/kT} \times \int_0^{\infty} \dot{\alpha} e^{S(\alpha_0) \dot{\alpha}^2/2kT} d\dot{\alpha} \times \iiint_{-\infty}^{\infty} e^{-(\dot{\gamma}^* - \beta)^* W_0(\alpha_0) (\dot{\gamma} - \beta)/2kT} d\dot{\gamma}. \quad (\text{A8})$$

In general, for a p -dimensional vector x and a p -dimensional square invertible matrix $A(x)$, and scalars y and s , we have

$$\int_{-\infty}^{\infty} e^{-x^* A x/2} dx = \left[\frac{(2\pi)^p}{\det \{A\}} \right]^{1/2}$$

and

$$\int_0^{\infty} y e^{-s y^2/2} dy = 1/s. \quad (\text{A9})$$

Using these in Eq. (A8) with $A = W_0(\alpha_0)/(kT)$, $p = 3$, and $s = S(\alpha_0)/(kT)$, we obtain

$$\begin{aligned} f_{\text{TS}}(\alpha_0) &= C \det \{\mathcal{M}(\alpha_0)\} e^{-\mathcal{U}(\alpha_0)/kT} \frac{kT}{S} \left[\frac{(2\pi kT)^3}{\det \{W_0(\alpha_0)\}} \right]^{1/2} \\ &= C \det \{\mathcal{M}(\alpha_0)\} e^{-\mathcal{U}(\alpha_0)/kT} \left[\frac{(2\pi)^3 (kT)^5}{\det \{\mathcal{M}(\alpha_0)\} S(\alpha_0)} \right]^{1/2} \quad (\text{using Eq. (A5)}) \\ &= C e^{-\mathcal{U}(\alpha_0)/kT} \left[(2\pi)^3 (kT)^5 \det \{\mathcal{M}(\alpha_0)\} S^{-1}(\alpha_0) \right]^{1/2}. \end{aligned} \quad (\text{A10})$$

When the Fixman potential from Eq. (13) is also applied by including it in the forcefield \mathcal{U} , the $\det \{\mathcal{M}(\alpha_0)\}$ term drops out from Eq. (A10) expression for the f_{TS} barrier transition rates. However, S remains as a term dependent on the α_0 location of the barrier peak. Thus, while the Fixman potential fully corrects for the distortions in the averages of configuration dependent quantities, it is only able to partially correct for distortions in the averages of velocity dependent functions.

- ¹M. Fixman, *Proc. Natl. Acad. Sci. U. S. A.* **71**, 3050 (1974).
²M. Fixman, *J. Chem. Phys.* **69**, 1538 (1978).
³N. Gō and H. A. Scheraga, *J. Chem. Phys.* **51**, 4751 (1969).
⁴N. Gō and H. A. Scheraga, *Macromolecules* **9**, 535 (1976).
⁵W. Van Gunsteren and H. Berendsen, *Mol. Phys.* **34**, 1311 (1977).
⁶W. Van Gunsteren, H. Berendsen, and J. Rullmann, *Mol. Phys.* **44**, 69 (1981).
⁷G. König and B. R. Brooks, *Biochim. Biophys. Acta* **1850**, 932 (2015).
⁸Y. a. Arnavtova, R. a. Abagyan, and M. Totrov, *Proteins* **79**, 477 (2011).
⁹J. Chen, W. Im, and C. L. Brooks, *J. Comput. Chem.* **26**, 1565 (2005).
¹⁰F. A. Bornemann and C. Schütte, *A Mathematical Approach to Smoothed Molecular Dynamics: Correcting Potentials for Freezing Bond Angles* (ZIB, 1995).
¹¹R. D. Skeel and S. Reich, *Eur. Phys. J.: Spec. Top.* **200**, 55 (2011).
¹²V. Katritch, M. Totrov, and R. Abagyan, *J. Comput. Chem.* **24**, 254 (2003).
¹³W. K. den Otter, *J. Chem. Theory Comput.* **9**, 3861 (2013).
¹⁴M. Pear and J. Weiner, *J. Chem. Phys.* **71**, 212 (1979).
¹⁵M. Pear and J. Weiner, *J. Chem. Phys.* **72**, 3939 (1980).
¹⁶D. Perchak, J. Skolnick, and R. Yaris, *Macromolecules* **18**, 519 (1985).
¹⁷E. Helfand, *J. Chem. Phys.* **71**, 5000 (1979).
¹⁸W. Van Gunsteren, *Mol. Phys.* **40**, 1015 (1980).
¹⁹A. Jain, *J. Comput. Phys.* **136**, 289 (1997).
²⁰A. Jain, S. Kandel, J. Wagner, A. Larsen, and N. Vaidehi, *J. Chem. Phys.* **139**, 244103 (2013).
²¹A. Jain, N. Vaidehi, and G. Rodriguez, *J. Comput. Phys.* **106**, 258 (1993).

- ²²N. Vaidehi, A. Jain, and W. A. Goddard, *J. Phys. Chem.* **100**, 10508 (1996).
²³G. S. Balaraman, I.-H. Park, A. Jain, and N. Vaidehi, *J. Phys. Chem. B* **115**, 7588 (2011).
²⁴J. R. Wagner, G. S. Balaraman, M. J. Niesen, A. B. Larsen, A. Jain, and N. Vaidehi, *J. Comput. Chem.* **34**, 904 (2013).
²⁵A. B. Larsen, J. R. Wagner, S. Kandel, R. Salomon-Ferrer, N. Vaidehi, and A. Jain, *J. Comput. Chem.* **35**, 2245 (2014).
²⁶A. Jain, *Robot and Multibody Dynamics: Analysis and Algorithms* (Springer, 2010).
²⁷R. A. Bertsch, N. Vaidehi, S. I. Chan, and W. Goddard, *Proteins* **33**, 343 (1998).
²⁸N. Vaidehi and W. A. Goddard, *J. Phys. Chem. A* **104**, 2375 (2000).
²⁹V. K. Gangupomu, J. R. Wagner, I.-H. Park, A. Jain, and N. Vaidehi, *Biophys. J.* **104**, 1999 (2013).
³⁰I.-H. Park, V. Gangupomu, J. Wagner, A. Jain, and N. Vaidehi, *J. Phys. Chem. B* **116**, 2365 (2012).
³¹A. B. Larsen, J. R. Wagner, A. Jain, and N. Vaidehi, *J. Chem. Inf. Model.* **54**, 508 (2014).
³²M. Pear and J. Weiner, *J. Chem. Phys.* **69**, 785 (1978).
³³W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.* **117**, 5179 (1995).
³⁴P. Echenique, I. Calvo, and J. L. Alonso, *J. Comput. Chem.* **27**, 1733 (2006).
³⁵P. Echenique, C. N. Cavasotto, and P. García-Risueño, *Eur. Phys. J.: Spec. Top.* **200**, 5 (2011).
³⁶D. S. Berkholz, M. V. Shapovalov, R. L. Dunbrack, Jr., and P. A. Karplus, *Structure* **17**, 1316 (2009).
³⁷K. Hinsen, S. Hu, G. R. Kneller, and A. J. Niemi, *J. Chem. Phys.* **139**, 124115 (2013).
³⁸See supplementary material at <http://dx.doi.org/10.1063/1.4939532> for Hellinger distances for the dipeptide simulations and for the per-residue Ramachandran distributions for CLN025.
³⁹S. Honda, T. Akiba, Y. S. Kato, Y. Sawada, M. Sekijima, M. Ishimura, A. Ooishi, H. Watanabe, T. Odahara, and K. Harata, *J. Am. Chem. Soc.* **130**, 15327 (2008).