

Vendi sampling for molecular simulations: Diversity as a force for faster convergence and better exploration

Cite as: J. Chem. Phys. 159, 144108 (2023); doi: 10.1063/5.0166172

Submitted: 3 July 2023 • Accepted: 25 September 2023 •

Published Online: 12 October 2023



Amey P. Pasarkar,¹ Gianluca M. Bencomo,² Simon Olsson,^{3,a)} and Adji Bouso Dieng^{1,a)}

AFFILIATIONS

¹Vertaix, Department of Computer Science, Princeton University, 35 Olden Street, Princeton, New Jersey 08544, USA

²Department of Computer Science, Princeton University, 35 Olden Street, Princeton, New Jersey 08544, USA

³Department of Computer Science and Engineering, Chalmers University of Technology, Rännvägen 6, 41258 Gothenburg, Sweden

Note: This paper is part of the JCP Special Topic on Machine Learning Hits Molecular Simulations.

a) Authors to whom correspondence should be addressed: simonols@chalmers.se and adjidi@princeton.edu

ABSTRACT

Molecular dynamics (MD) is the method of choice for understanding the structure, function, and interactions of molecules. However, MD simulations are limited by the strong metastability of many molecules, which traps them in a single conformation basin for an extended amount of time. Enhanced sampling techniques, such as *metadynamics* and *replica exchange*, have been developed to overcome this limitation and accelerate the exploration of complex free energy landscapes. In this paper, we propose *Vendi Sampling*, a replica-based algorithm for increasing the efficiency and efficacy of the exploration of molecular conformation spaces. In Vendi sampling, replicas are simulated in parallel and coupled via a global statistical measure, the Vendi Score, to enhance diversity. Vendi sampling allows for the recovery of unbiased sampling statistics and dramatically improves sampling efficiency. We demonstrate the effectiveness of Vendi sampling in improving molecular dynamics simulations by showing significant improvements in coverage and mixing between metastable states and convergence of free energy estimates for four common benchmarks, including Alanine Dipeptide and Chignolin.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0166172>

I. INTRODUCTION

The exchange between metastable configurations of proteins and nucleic acids is essential to their biological function.¹ Molecular dynamics (MD) simulations are widely adopted for simulating the transitions between metastable states because they offer full spatial and temporal resolution of molecular systems. These transitions can range from simple localized changes, such as aromatic ring-flips,² to complex global structural rearrangements, including protein folding^{3–5} and protein-ligand binding.^{6–9} Experimental techniques that enable the measurement of signals sensitive to the exchange between these states^{10–13} have greatly advanced the study of these metastable states. However, low spatial resolution and ensemble averaging are still challenges that complicate data analyses.¹⁴

MD simulations can predict stationary and dynamic correlations, allowing for direct comparison to experimental data.

However, a significant limitation of MD arises due to the strong metastability of many molecular systems, which can trap them in a single conformational basin for an extended amount of time.^{15–17} This limitation undermines quantitative comparisons to experiments because statistical sampling comes at a high computational cost. To address this issue, there has been intense research aimed at developing faster algorithms for sampling highly metastable systems at a lower computational cost.¹⁸

These enhanced sampling or extended ensemble methods rely on two broad strategies: finding a surrogate of the Boltzmann ensemble that rapidly mixes between metastable states or exchanging the state variable with one or multiple ensembles that mix faster.^{19–26} To ensure that the modified ensembles sufficiently overlap with the true Boltzmann ensemble, statistical reweighting techniques such as importance sampling, the weighted histogram analysis method,²⁷ the multistate Bennett acceptance ratio,²⁸ or transition-based methods^{29–31} are used. These techniques allow for

an effective reweighting of the equilibrium statistics but are disadvantaged by the intrinsic need to identify collective variables, perturb macroscopic thermodynamic variables, or alter Hamiltonians, which often rely on manual trial-and-error of numerous potential candidates.¹⁸ The identification of useful collective variables is a field of its own, with multiple application-specific approaches, including for ligand binding³² and slow dynamics.^{33–35} In particular, collective variable estimation has seen a surge in interest in recent years due to the broad accessibility of deep representation learning methods.^{33–37}

Machine learning-based approaches, such as Boltzmann generators,^{38–41} learn surrogate models of the intractable Boltzmann distribution. If these surrogate models allow for exact sample likelihood estimation, highly effective sampling of unbiased equilibrium statistics can be achieved via reweighting or importance sampling. While a promising strategy, current architectural limitations in the deep neural networks used to learn such surrogate models have prevented their broad adoption.

In this paper, we propose a replica-based method called *Vendi Sampling* where multiple copies of a molecular system are simulated in parallel. To enhance the conformational space exploration, the replicas are coupled via a global statistical measure, the *Vendi score*.⁴² The Vendi score reflects the instantaneous and internal diversity of the replicas and is a function of the eigenvalues of a Gram matrix computed via a pre-specified kernel. The Gram matrix is constructed by evaluating the kernel between all pairs of simulation replicas. As such, the proposed *Vendi sampling* method does not rely on modulating thermodynamic variables, Hamiltonians, or defining collective variables to enhance sampling. Instead, an extended ensemble is defined, which can drive sampling in infinite dimensional feature spaces, without explicit calculation of collective variables.

Our investigation of Vendi sampling in several challenging systems reveals its ability to rapidly detect free energy minima, particularly in cases where there are large free energy barriers between states. Moreover, Vendi sampling enables the rapid convergence to the unbiased equilibrium statistics, facilitating the computation of observables such as free energy differences. We accomplished these findings using generic kernels, which do not encode slowly relaxing degrees of freedom. However, the current implementation incurs some computational overhead. We anticipate that further research into the development of more efficient kernels and the use of multiple time-stepping schemes⁴³ can further boost the performance of Vendi sampling.

II. THEORY

Here we provide some background, describe Vendi sampling, and explain how to recover unbiased observable statistics.

A. Boltzmann distribution, observables, and free energies

At equilibrium, a single molecular system denoted by $\mathbf{x} \in \Omega \subset \mathbb{R}^{3N}$, where N is the number of atoms, has a distribution of conformational states which is equal to the Boltzmann distribution,

$$p(\mathbf{x}) = \mathcal{Z}^{-1} \exp(-\beta U(\mathbf{x})), \quad (1)$$

with inverse temperature β , and $\mathcal{Z} = \int \exp(-\beta U(\mathbf{x})) d\mathbf{x}$, is the partition function. In MD simulations, we asymptotically generate samples from this distribution.

Given a sample $\{\mathbf{x}_i\}_{i=0}^M$ drawn independently and identically from $p(\mathbf{x})$ (i.i.d.), a state function—or forward-model—corresponding to an experimental observable denoted by $f(\mathbf{x})$, we can compute bulk ensemble averaged observables using the empirical average,

$$o_f = \int_{\Omega} f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \approx \frac{1}{M} \sum_{i=1}^M f(\mathbf{x}_i). \quad (2)$$

However, due to slow mixing between meta-stable states in MD simulations, samples are not i.i.d., and consequently, Eq. (2), is biased.

We can define the free energy of a state $A \subset \Omega$ as a special case of a stationary observable,

$$F_A = -\log \int_{\Omega} p(\mathbf{x}) \delta(x \in A) d\mathbf{x} = -\log \mathbb{P}(x \in A) \quad (3)$$

where $\mathbb{P}(\cdot)$ denotes the probability of an event. We can similarly express the free energy difference between states $A, B \subset \Omega$ as

$$\begin{aligned} F_{AB} &= -\log \int_{\Omega} p(\mathbf{x}) \delta(x \in A) d\mathbf{x} + \log \int_{\Omega} p(\mathbf{x}) \delta(x \in B) d\mathbf{x} \\ &= -\log \frac{\mathbb{P}(x \in A)}{\mathbb{P}(x \in B)}. \end{aligned} \quad (4)$$

1. Replicated systems and expanded ensembles

In replicated systems, N copies of the same system, $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, are simulated simultaneously. Each copy is referred to as a “replica.” In the simplest case, when the replicas are uncoupled, they can be simulated independently as their joint distribution factorizes as the product of the different marginal distributions over each replica,

$$p_{\text{uncoupled}}(\mathcal{X}) = \prod_{i=1}^N p_i(\mathbf{x}_i). \quad (5)$$

Here the densities $p_i(\cdot)$ can either be identical or different. One common choice for rendering the densities different is by using different temperatures β_i for each of the replicas, as is done in replica-exchange and parallel tempering methods.^{22,23,44}

In the case where the replicas are not independent, we call this the *coupled case*, the replicas exchange information via an additional function, $\pi(\cdot)$, which depends on all or parts of the replicas. The corresponding joint distribution can be written as

$$p_{\text{coupled}}(\mathcal{X}) = \pi(\mathcal{X}) \prod_{i=1}^N p_i(\mathbf{x}_i). \quad (6)$$

This approach is extensively used to impose experimental constraints on molecular simulations, e.g. by enforcing an averaged experimental observable to match an average computed across multiple replicas.^{19,45–51}

B. Vendi sampling

Here we introduce *Vendi sampling* a replica approach to enhanced sampling. In Vendi sampling replicas are coupled using a statistical measure of diversity, the Vendi Score.⁴² The goal is to encourage the replicas to cover different regions of conformation space, thereby rapidly discovering and mixing between metastable states and allowing faster convergence of stationary ensemble properties, such as free energy differences.

1. The Vendi score

The Vendi score⁴² is an interpretable diversity metric that quantitatively describes how diverse a collection of items is. It counts the effective number of dissimilar elements in the collection being evaluated for diversity. Given a collection of N items, the maximum possible value of the Vendi Score is N , when all items in the collection are uniquely distinct from each other. The measure of similarity between the items in the collection is an input to the Vendi Score, which allows for a flexible specification of any form of similarity. Mathematically, the similarities are computed using kernels, $k(\cdot, \cdot)$.

The Vendi Score of a collection of samples $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$ is defined as the Shannon entropy of the eigenvalues of a similarity matrix,

$$\text{VS}(\{\mathbf{x}_1, \dots, \mathbf{x}_M\}) = \exp\left(-\sum_{i=1}^M \lambda_i \log \lambda_i\right). \quad (7)$$

Here λ_i is the i 'th eigenvalue of $\mathbf{K}$, the matrix induced by a user-defined similarity function $k(\cdot, \cdot)$. Technically, the similarity matrix $\mathbf{K}$ is a *Kernel matrix* with elements

$$\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j), \quad 0 < i < M, \quad 0 < j < M. \quad (8)$$

In Vendi sampling, the M samples correspond to M replicas of a molecular system, and the resulting Vendi Score is a coupling term. We combine Eqs. (6) and (7),

$$p_{\text{vendi-coupled}}(\mathcal{X}) \propto \text{VS}(\mathcal{X}) \prod_{i=1}^N p_i(\mathbf{x}_i). \quad (9)$$

Note that the joint distribution of the replicas is no longer normalized as the Vendi Score is not a probability distribution function. However, Vendi sampling doesn't require normalization of the joint distribution of the replicas. More specifically, we use the *Vendi energy function* implied by the density $p_{\text{vendi-coupled}}$

$$U_{\text{vendi-coupled}}(\mathcal{X}) = \sum_{i=1}^N u_i(\mathbf{x}_i) + \lambda_i \log \lambda_i, \quad (10)$$

where $u_i(\cdot) \triangleq \beta_i U_i(\cdot)$ is a unitless energy and λ_i is as described earlier. We use the Vendi energy function to simulate the system using the overdamped Langevin (Brownian) dynamics,⁵²

$$\frac{d\mathcal{X}(t)}{dt} = -\nabla U_{\text{vendi-coupled}}(\mathcal{X}(t))/\gamma + \sqrt{2D} dW \quad (11)$$

where dW denotes a Weiner process, D is a diffusion constant and γ is a friction coefficient. Each replica evolves over time according to the stochastic differential equation

$$\frac{d\mathbf{x}_i(t)}{dt} = -[\nabla u_i(\mathbf{x}_i(t)) + \nu f_{\text{vendi}}(\mathbf{x}_i)]/\gamma + \sqrt{2D} dW \quad (12)$$

where $f_{\text{vendi}}(\mathbf{x}_i) = -\frac{\partial}{\partial \mathbf{x}_i} \log \text{VS}(\mathcal{X})$ is the “Vendi force” for $\mathbf{x}_i$ —a repulsive force applied to $\mathbf{x}_i$ to drive it away from the other replicas—and ν is an empirical weight factor. $f_{\text{vendi}}(\mathbf{x}_i)$ is computed using automatic differentiation through PyTorch and Torch Force for OpenMM,⁵³ and involves computation of the Kernel matrix [Eq. (8)], its eigenvalues λ_i , and the log Vendi score [Eq. (7)].

In our experiments, we explore two types of Langevin sampling to simulate the combined system [Eq. (9)]: we use overdamped Langevin dynamics (Brownian dynamics) for the model systems (the Prinz potential and the double well) and use the Langevin thermostat⁵⁴ for the molecular systems.

The weight ν is a hyperparameter which we can vary during simulation, through for instance, annealing schemes. We describe the different schemes used in the experiments.

2. Recovering unbiased samples

After simulating an ensemble of coupled replicas for T time steps $\{\mathcal{X}(0), \dots, \mathcal{X}(T)\}$ we can reweight these trajectories into an unbiased ensemble by resampling using the state-dependent weights

$$w_t = \text{VS}(\mathcal{X}(t))^{-\nu}, \quad (13)$$

which will yield asymptotically unbiased samples from the joint distribution Eq. (5).

In practice, $\nu = 0$ for the vast majority of the simulation. So, we collect statistics when $\nu = 0$, in between intermittent phases where $\nu > 0$. In this case, samples do not need to be reweighted.

C. Kernels

The *Vendi force* applied to each replica at every simulation step is contingent on the kernel matrix $\mathbf{K}$, as it encodes the diversity across replicas.

Kernels are inner-products in a feature space, $\mathcal{V}$, defined by the feature function $\psi: \mathbb{R}^{3N} \rightarrow \mathcal{V}$, for a three dimensional system of N particles,

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \psi(\mathbf{x}_i), \psi(\mathbf{x}_j) \rangle_{\mathcal{V}} = \psi_i^\top \psi_j \quad (14)$$

where the latter equality holds if $\mathcal{V}$ is a vector space. In general, the dimension of $\mathcal{V}$ is infinite. This gives rise to the celebrated “Kernel trick” in machine learning,⁵⁵ where data is embedded in to a high-dimensional space to allow for linear separation. In other words, instead of specifying $\psi(\cdot)$ we can specify a kernel function $k(\cdot, \cdot)$, and avoid featurization into high-dimensional spaces. In the context of molecular simulations, the kernel can be interpreted as a way of comparing high-dimensional collective variables simultaneously, side-stepping the need to engineer the collective variables. Below we detail the kernel used in the molecular applications we studied.

1. Invariant kernel for molecular systems

To ensure, we do not maximize diversity by rigid body transformations, e.g. translation of the center-of-mass, or the rotation of

the reference frame, we wish to define a kernel which is invariant to these transformations. We note that these constraints are valid when we consider MD simulations at equilibrium without exchange of matter or energy with the environment, and without interactions with an external field. Consequently, we require a kernel $k(\cdot, \cdot)$ that is invariant to those symmetries, e.g. does not change due when the molecule is translated or rotated. Since we would like to quantify some metric for distance within the space of possible molecular conformations, removing these symmetries in the space of all possible molecular configurations is a necessary requirement for a diversity metric that is accurate and efficient.

We achieve an invariant kernel, by using a Gaussian kernel,

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right) \quad (15)$$

where $\mathbf{x}$ and $\mathbf{x}'$ are input coordinates, σ is a hyperparameter that determines the width of the kernel, and $\|\cdot\|$ denotes the Euclidean norm. The input coordinates are transformed to make $k(\mathbf{x}, \mathbf{x}')$ constant under rigid body transformations. Concretely, given some replicas $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, we can ensure translational invariance by forming a new set $\tilde{\mathbf{X}}$ that is mean-centered,

$$\tilde{\mathbf{X}} = \mathbf{X} - \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i. \quad (16)$$

To ensure invariance of $k(\mathbf{x}, \mathbf{x}')$ under rotation we use a procedure outlined previously to find rotation matrices $\mathbf{U}_i$ which align all replicas onto a common frame.⁵⁶ These transformed coordinates are in turn used to evaluate the kernel.

III. RESULTS

We compare Vendi sampling against an uncoupled replica method with overdamped Langevin dynamics to showcase the added benefit of enforcing a repulsive force through the Vendi Score. We refer to this baseline as *Replica sampling* in the rest of the paper. We report free energy profiles on two model systems, the Prinz potential and the double well, and two molecular systems, Alanine dipeptide in vacuum and Chignolin in implicit solvent. In all cases we found Vendi sampling converges faster and explores energy surfaces better than Replica sampling.

A. Model systems

We first test Vendi sampling on two low-dimensional model systems that have exact expressions for their free energy functions. For these systems we use the simple kernel function:

$$k(x, x') = 1 - \frac{|x - x'|}{|x| + |x'|}. \quad (17)$$

1. Prinz potential

The “Prinz potential”¹⁵ is a four-well 1D potential that features an energy barrier at $x = 0$. We test Vendi sampling’s ability to both discover these wells and provide unbiased equilibrium sampling statistics. We initialized the replicas uniformly from one side of the central energy barrier ($\sim U[0, 1]$). We performed this experiment ten times for each choice of replica size, comparing against

Replica sampling with an integration step of $\eta = 10^{-4}$ for both samplers. For all replica sizes, we used a linear annealing schedule for the vendi force coefficient given by

$$v_t = \max\left(0, 1 - \frac{t}{1000}\right) * 100 \quad \forall t = 0, 1, 2, \dots \quad (18)$$

We observe that Vendi sampling rapidly identifies modes on the opposite side of the energy barrier [Figs. 1(a) and 1(b)]. After 10 000 time steps, Replica sampling poorly characterizes the left-most mode, while Vendi sampling is already able to approximate it relatively well.

We compare the convergence of Vendi sampling and Replica sampling, by observing the free energy difference [Eq. (4)] between the segments $A : \{x \in [-1, 0]\}$ and $B : \{x \in [0, 1]\}$ for both methods, as a function of simulation step. Vendi sampling consistently outperforms the Replica sampling baseline [Fig. 1(c)] by reaching the true free energy difference within 0.1 kT in fewer steps. Specifically, for four replicas Vendi sampling converged at 150 000 steps, whereas for 8 and 16 replicas Vendi Sampling reached similar values by the 50 000th step. In contrast, Replica sampling required at least twice as many steps to achieve similar convergence.

Vendi sampling discovers all high probability (low energy) within 10 000 time steps while Replica sampling requires an order of magnitude more steps to discover all states.

To additionally demonstrate the role of the Vendi Force, we vary the barrier height of the Prinz Potential function and measure how long it takes for samples initialized on one side of the barrier to cross over (Fig. 2). In this setting, we find that replica sampling’s time to cross the barrier grows exponentially as the barrier height increases. For large values of the Vendi force weight v , Vendi Sampling scales well and does not require a noticeable increase in time to cross.

2. Double well potential

To gauge the performance of Vendi sampling on systems with a higher free energy barrier (>10 kT), akin to those observed in protein folding-unfolding or in chemical reactions, we considered a previously studied two-dimensional double-well benchmark.³⁸

We limited our experiments to 32 replicas for both Vendi sampling and Replica sampling, with initial starting positions drawn randomly in the interval $x \in [-2.5, 2.5]$ and $y \in [-2.5, 2.5]$. The free energy difference was calculate as before with Eq. (4), using $A = \{x \in [-2.5, 0], y \in [-4, 4]\}$ and $B = \{x \in [0, 2.5], y \in [-4, 4]\}$. We used step-size $\eta = 10^{-2}$ for both samplers. We again use a linear annealing schedule over the course of 50 000 steps for the Vendi force weight v .

$$v_t = \max\left(0, 1 - \frac{t}{50\,000}\right) * 100 \quad \forall t = 0, 1, 2, \dots \quad (19)$$

Vendi sampling converged within 0.2 kT of the true free energy difference by the 1 000 000th step, whereas Replica sampling did not reach this over the course of the entire simulation [Fig. 3(c)]. We note that we observe ~ 1200 transitions over the energy barrier in the time period while the Vendi Force is active, and ~ 500 for the rest of the duration of the simulation. Replica sampling also observes 500 transitions over the course of the entire simulation.

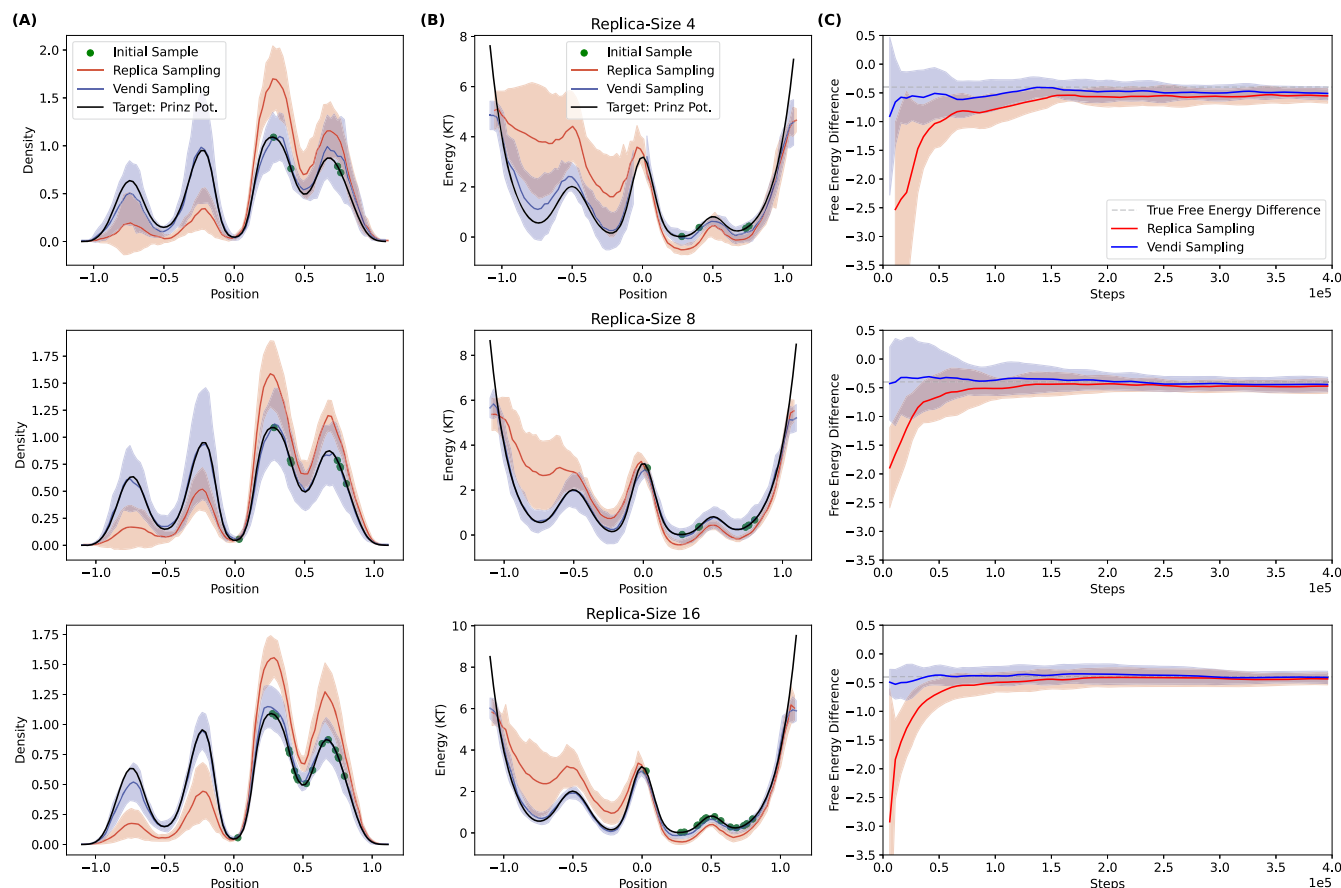


FIG. 1. Vendi sampler rapidly identifies states and quickly recovers ground-truth equilibrium statistics. (a) Probability density functions for Vendi sampling and Replica sampling compared against the ground truth Prinz Potential density after 10^4 steps of each sampler for various replica sizes. (b) Energy functions for Vendi sampling and Replica sampling after 10^4 steps for various replica sizes. (c) Free energy difference across boundary over time for both samplers. In all cases, the shaded region reflects uncertainty as estimated using the standard deviation across $n = 10$ simulations after 10^3 burn-in steps.

B. Molecular systems

Here we further investigate two molecular systems: capped alanine (alanine di-peptide) and the miniprotein chignolin, in vacuum and implicit solvent respectively. These systems are established benchmark systems to evaluate sampling methods in molecular dynamics.^{57,58} We perform similar benchmarks as above, however, we use a kernel which is invariant to rotations and translation as described in Sec. II C using kernel bandwidth parameter $\sigma = \sqrt{\frac{1}{2}}$. This kernel ensures only internal degrees of freedom are encouraged to diversify, and avoids that the center of mass and orientation of the molecular system influences the Vendi score.

1. Alanine dipeptide in vacuum

Alanine dipeptide (Ala2) or capped alanine is a small, yet meaningful benchmark system to test enhanced sampling methods and the impact of solvation.^{58–60}

In vacuum, Ala2 exhibits a three-basin free-energy landscape in the Ramachandran plot of the back-bone dihedral angles

ϕ (C-N-C α -C) and ψ (N-C α -C-N). Two of the basins occupy the β -strand like structures and are disconnected by only a small free energy barrier. The third, minor basin occupies the region of the Ramachandran plot associated with left-handed helices, and is disconnected from the β -strand states by a large free-free energy barrier $>10k_B T$ in the Amber99SB forcefield at 300 K.⁶¹ All simulations were carried out using the Langevin integrator implemented in OpenMM (v. 8.0)⁵⁴ with a time step of 2.0 fs and a collision frequency of 1.0 ps^{-1} . A free energy baseline was established by running ten simulations of 100 ns each, as depicted in supplementary material Fig. 1. To analyze the conformational distribution, we recorded the ψ and ϕ backbone dihedral angles of each sample. The free energy difference was calculated via Eq. (4) with $A = \{\phi \in [-\pi, 0], \psi \in [-\pi, \pi]\}$ and $B = \{\phi \in [0, \pi], \psi \in [-\pi, \pi]\}$.

For Vendi sampling we computed the Vendi Force that was applied to the Langevin dynamics at every timestep for the first 10 ps, with a learning rate of $\nu = 1000$. After 10 ps, the Vendi Sampler reverted to using standard Langevin dynamics.

We experimented with 32-replica systems for both Vendi sampling and Replica sampling, initializing all replicas in the β -stranded

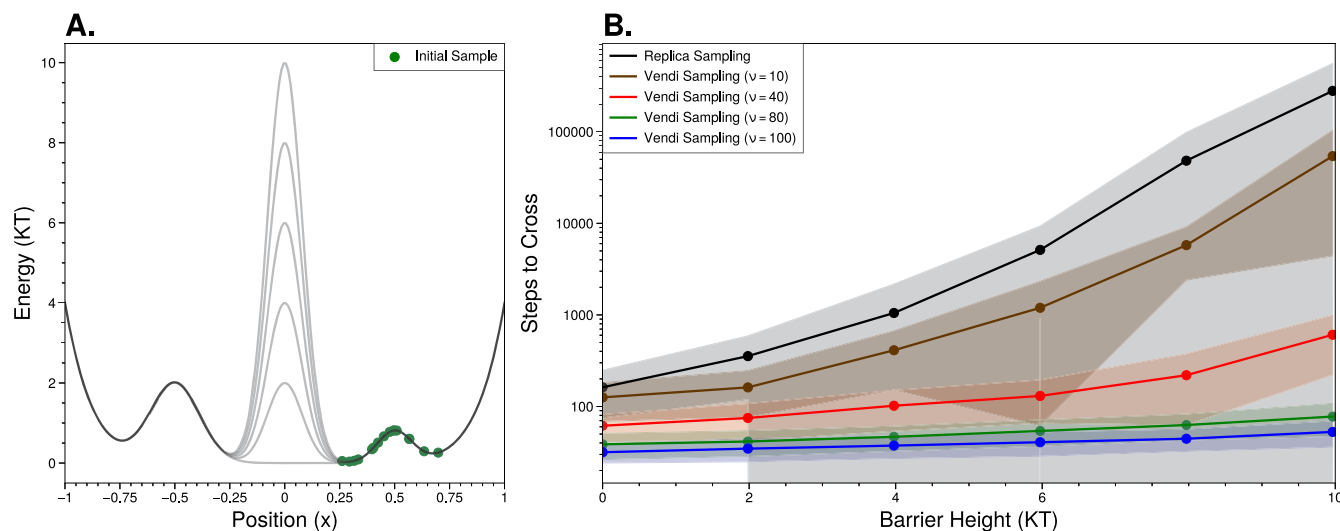


FIG. 2. The Vendi sampling can overcome Free Energy Barriers of varying heights. (a) Six Different Prinz Potentials, each with a varied height of the central barrier at $x = 0$, as well as the initial 16 sample positions. (b) The average number of steps required for Replica Sampling and Vendi Sampling to have at least one sample cross the energy barrier. Vendi force coefficient v is varied. Shaded regions represent uncertainty as estimated using the standard deviation across $n = 20$ simulations.

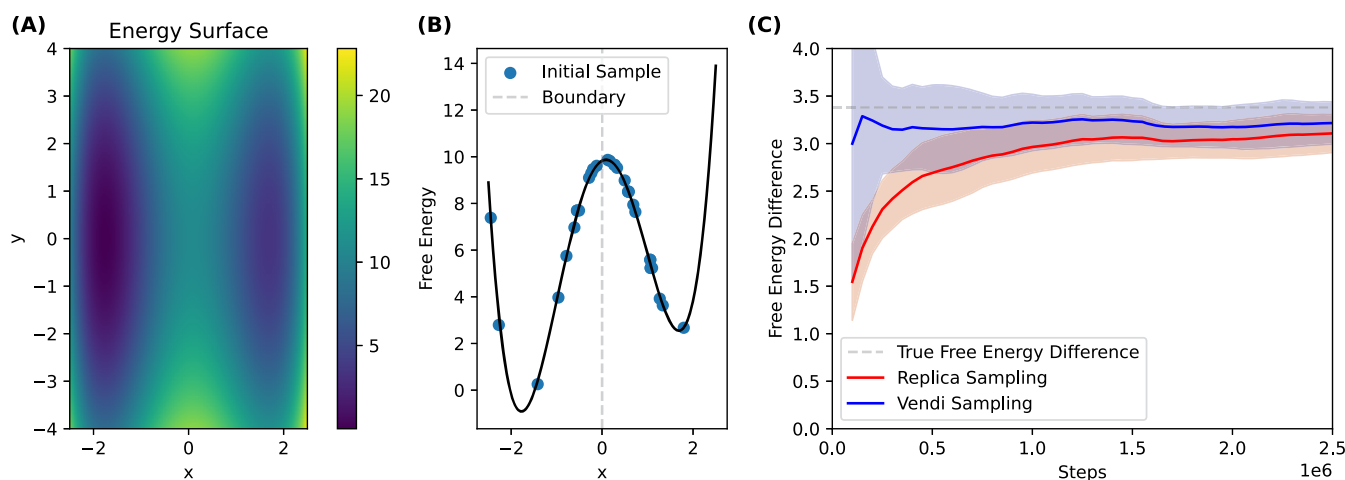


FIG. 3. The Vendi sampler rapidly converges in the double well potential. (a) The underlying 2D free energy surface. (b) The marginal free energy along the x -axis. The initial conditions are shown as dots on the curve and the free energy boundary for free energy calculations is shown with a dashed line. (c) Comparison of convergence over 2.5×10^6 steps for Vendi sampling and Replica sampling. The shaded area represents uncertainty as measured by the standard deviation over $n = 10$ experiments and after 50 000 steps.

state. We find that the Vendi sampler effectively used the Vendi force to break the hydrogen bonds between the amide nitrogen and carbonyl oxygen atoms stabilizing the β -stranded states, resulting in several replicas transitioning to the left-handed state within the first 1.0 ns [Figs. 4(a) and 4(b)]. In contrast, the Replica sampler relies on thermal fluctuations to cross this free energy barrier. Replica sampling, further requires an approximate three-fold larger simulation effort to equilibrate the average relative state populations within the accuracy of the baseline estimate. This demonstrates Vendi sampling's enhanced capacity for climbing the large energy

barrier and discovering the left-handed minima quickly. The increased mode coverage during the early stages of the simulation facilitated faster convergence to the true free energy difference, as illustrated in Fig. 4(d). It also important that our sampler encourages mixing across modes in the MD trajectories. We observe that while the Vendi force is applied over the first 10 ps of the simulation, the Vendi Sampler provides much better mixing in the MD simulation [Figs. 5(a)–5(c)]. The Replica sampler would mostly transition to the left-handed state later in the simulation [Figs. 5(d)–5(f)], by which time the Vendi sampler had already converged.

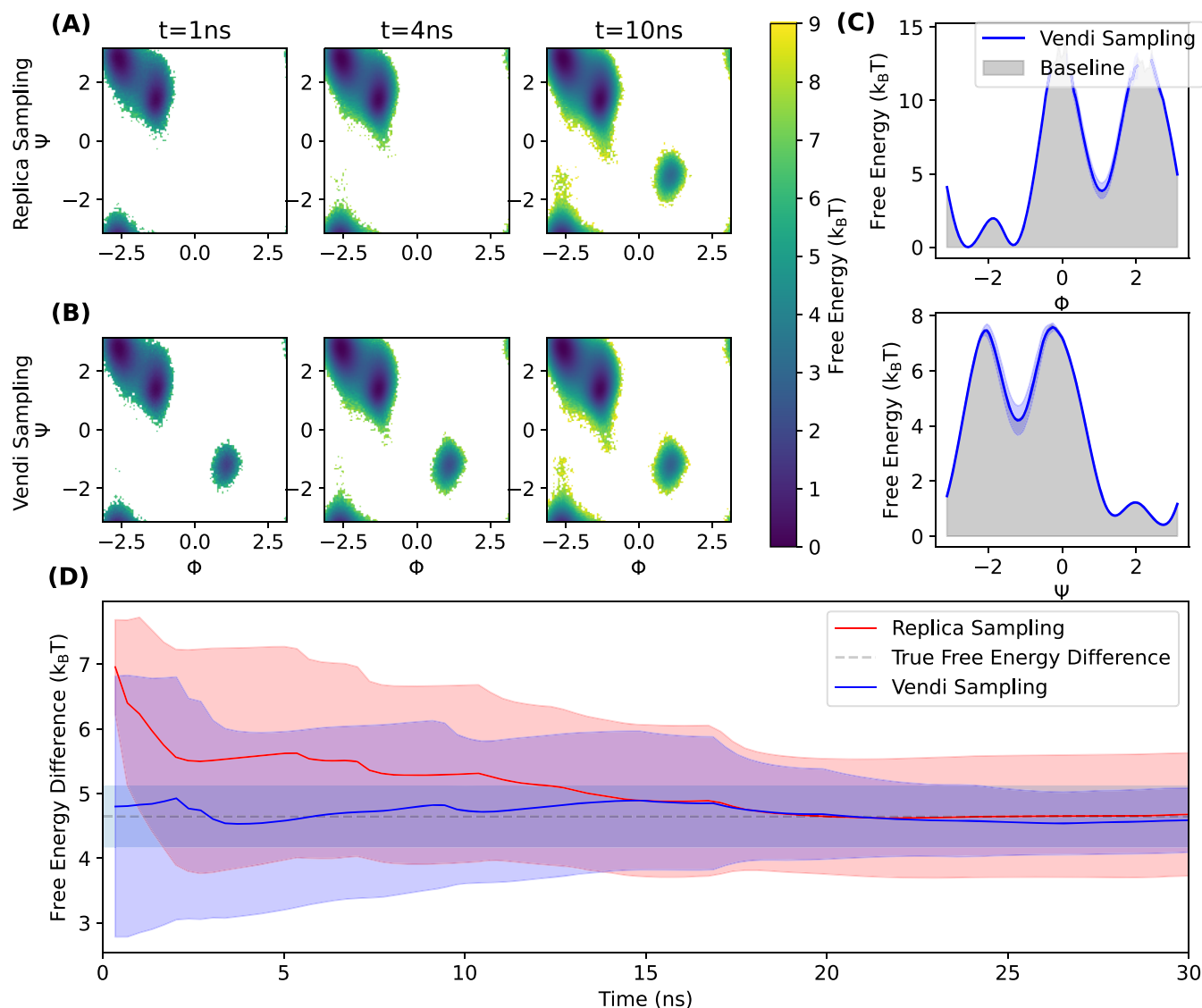


FIG. 4. Vendi Sampler rapidly identifies modes, converges fast, and recovers accurate free energy difference in Ala2 in vacuum. (a) Free energy maps of dihedral angles of replica sampler at various time points. (b) Density maps of dihedral angles of the Vendi sampler at various time points show noticeable improvement in the time taken to discover all modes. (c) Comparison of the average learned 1D Free Energy functions along each angle compared against one baseline run. Bars show standard deviation across trials. (d) Free Energy Difference over time for each sampler across trials. Throughout, the shaded region for samplers represents uncertainty estimated as the standard deviation with $n = 10$ after 5000 burn-in steps. The black bar also represents uncertainty estimated as the standard deviation in the final free energy difference across ten long runs.

C. Protein folding chignolin in implicit solvent (CLN025)

Chignolin is a small artificial mini-protein with fast-folding kinetics, is a widely used model system for computational and experimental study of protein folding.^{57,62,63} Its folded state conformation is characterized by a strong hydrogen bonding network between the backbone amide and carbonyl groups forming a stable β -hairpin conformation.

We used the CHARMM22 protein force field from MacKerell *et al.*⁶⁴ with the OBC2 (Onufriev-Bashford-Case) implicit solvent model⁶⁵ implemented by OpenMM to simulate Chignolin at a temperature of 350 K. Vendi sampling and Replica sampling both used the standard Langevin integrator from OpenMM with constraints on bonds to hydrogens and hydrogen mass repartitioning, time step of 4.0 fs and a collision frequency of 0.1 ps^{-1} . The Vendi sampler additionally incorporated a Vendi Force that was applied to the

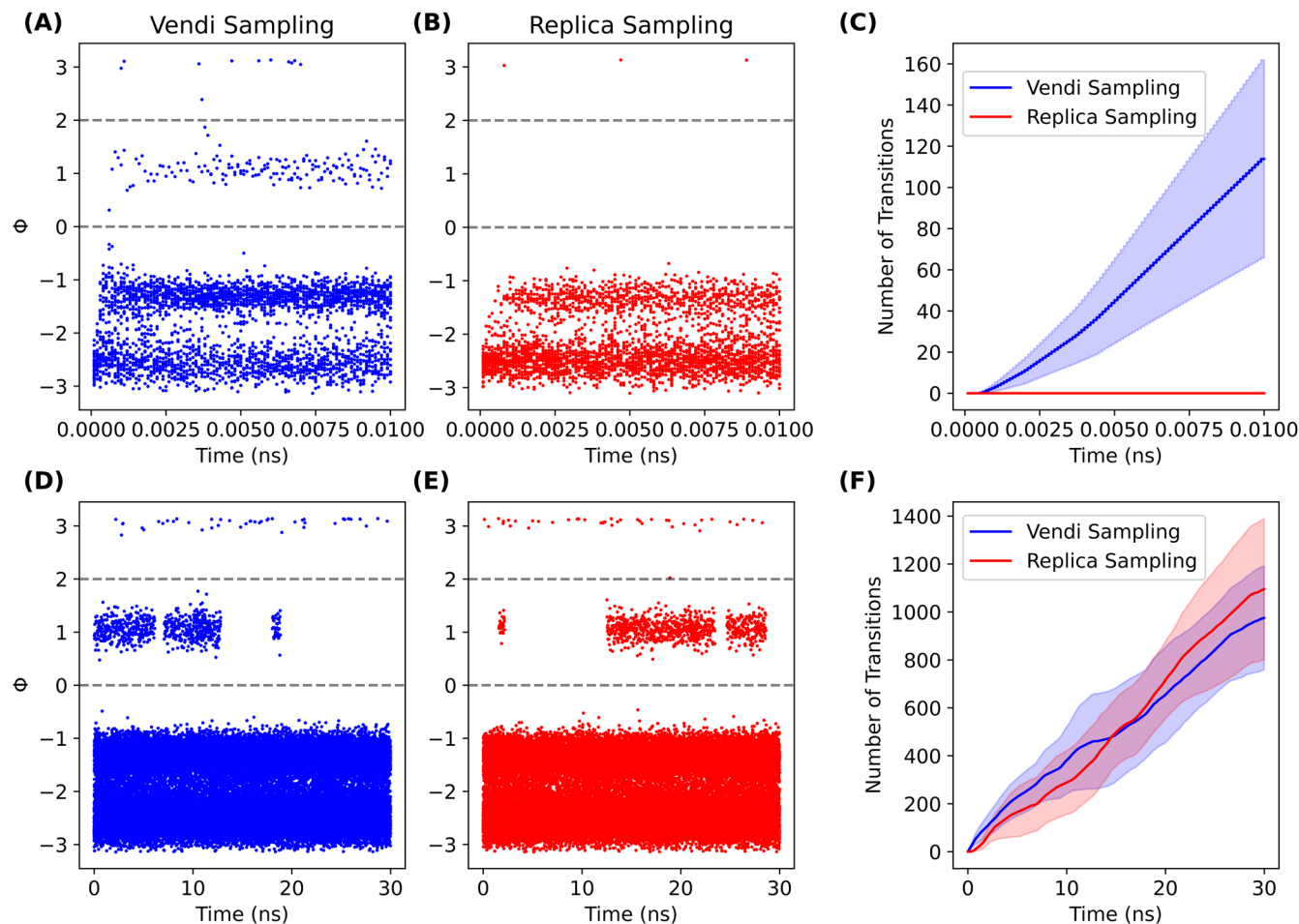


FIG. 5. Vendi sampling provides fast mixing in Ala2 system by rapidly crossing the energy barrier, performing multiple transitions. (a) and (b): Scatter plot of dihedral angle ϕ over time of samples from each method (first 10 ps are shown). The dashed line represents the region where specific Ala2 conformation is sampled. (c) Number of transitions in and out of the boxed region. (d) and (e): Scatter plot of the dihedral angle ϕ over time of samples from each method for an entire 30 ns-length simulation. (f) Number of transitions in and out of the boxed region. In all plots, the shaded region depicts uncertainty represented as an 80% confidence interval.

Langevin dynamics at every timestep for the first 40 ps, with a weight coefficient of $\nu = 1000$. After 40 ps, the Vendi sampler reverted to using standard Langevin dynamics.

As for Ala2, we used a 32-replica setup to compare Vendi sampling and Replica sampling. We initializing all replicas in the misfolded state shown in supplementary material Fig. 2(a). A free energy baseline was established by running ten simulations of 150 ns each. Using previously published simulation data on Chignolin in explicit water⁴ to identify collective variables via TICA (time-lagged independent component analysis).^{66,67} We used all pair-wise C α distances as features and a lag-time of 5 ns. The first two TICs separate folded and unfolded states, and are used to compute free energies. The two primary TICs are shown in supplementary material Fig. 2(a). The free energy difference was calculated via Eq. (4) with $A = \{ \text{TICA1} < 0 \}$ and $B = \{ \text{TICA1} > 0 \}$ (Fig. 6).

We observe that the Vendi sampling average across trials converges to the estimated true free energy difference in the first 25 ns

of the simulation. We see from the marginal free energy show in supplementary material Fig. 2(b) and the reference free energy surface in supplementary material Fig. 2 that the Vendi sampler is observing folding events earlier in the simulation.

We define folding events using TICA 1 features, measuring if at a given timestep the replica has had $\text{TICA1} < -0.2$ at least five times within a 100 step window. If so, then we mark the replica as in a folded state and then count the number of total transitions into the folded state we observe. We use a similar method for counting the number of transitions into the unfolded state in supplementary material Fig. 2(c), instead looking at $\text{TICA1} \in [0.5, 2.5]$ and $\text{TICA2} \in [-0.5, -3.5]$.

Supplementary material Table S1 shows that the Vendi Force calculation is quite expensive, but the force is only applied over a small fraction (0.3%) of steps. It's application over even a small number of steps is enough to achieve a noticeable improvement in performance.

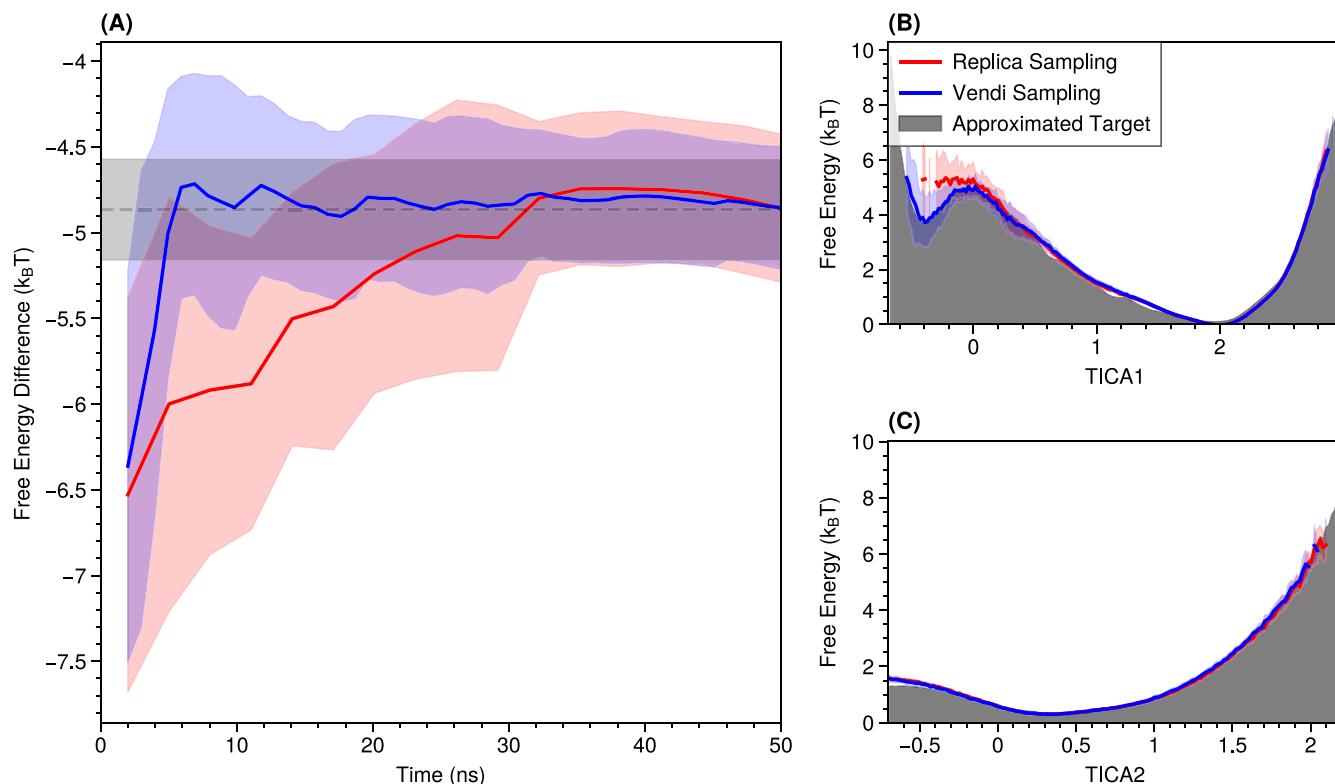


FIG. 6. Vendi Sampling rapidly converges in Chignolin system. (a) Free energy Difference across TICA1 = 0, over time for the MD simulation of each sampler. Shaded region represents standard deviation across ten trials after 10 000 burn-in steps. (b) and (c) Marginal free energy of each sampler after 10 ns along TICA axis compared against the average of ten long unbiased runs.

IV. DISCUSSION AND CONCLUSION

We introduced Vendi sampling, a replica-based sampling method for molecular simulations. Vendi sampling overcomes large free energy barriers and enhances sampling. While there is a wide variety of enhanced sampling techniques available in the literature,¹⁸ Vendi sampling, to our best knowledge, takes a unique approach to the problem. It is driven by a diversity metric, the Vendi score, which is computed according to a kernel matrix that performs pairwise comparisons of the states of all replicas. Consequently, the system of coupled replicas defined by Vendi sampling forms an extended ensemble that is neither driven by the choice or modulation of a macroscopic thermodynamic parameter (e.g. temperature or Hamiltonian) nor by the definition of a collective variable. Instead, the extended ensemble is defined by a kernel function. The kernel function computes inner products in feature spaces, e.g. high-dimensional collective variables. However, since we do not need to explicitly compute the features in the kernel formalism, the underlying features can, in principle, be infinite-dimensional. We can thereby avoid costly collective variable identification, without suffering from “the curse of dimensionality.”

It is important to note that the use of kernels is not new in enhanced sampling, however, previous use has focused on estimating an adaptive biasing potential.^{58,68–70} While these approaches can

yield low-variance free energy estimates, they intrinsically have to balance exploration and accuracy,⁷¹ e.g. by slowly tampering off the estimation of the adaptive biasing potential.⁷² On the contrary, Vendi sampling yields high-variance sample estimates, yet avoids problems associated with adaptive biasing potential estimation.

Through our analyses of tractable model systems and benchmark molecular systems, we find that Vendi sampling indeed excels in cases that are defined by high free energy barriers in two ways: it rapidly covers the local minima of the free energy landscape and enables fast mixing across free energy barriers to allow for fast convergence of free energy estimates. However, while not observed in any systems considered in this manuscript, in certain extreme cases, where very high free energy barriers separate meta-stable states and populations between the states are highly skewed, Vendi sampling may require an impractical number of replicas to ensure accurate free energy estimates.

Future work will explore whether Vendi sampling can benefit from more advanced kernel functions, e.g. in the context of sampling interfacial water molecules, which may be important for molecular processes such as ligand binding. Finally, exploring alternative annealing strategies of the Vendi force during sampling may lead to further efficiency gains, and speeding up Vendi calculations through more efficient implementations and approximations, may further extend the practical scope of the approach.

SUPPLEMENTARY MATERIAL

Additional figures and one table is available in the supplementary material.

ACKNOWLEDGMENTS

Adji Bousso Dieng is supported by the National Science Foundation, Office of Advanced Cyberinfrastructure (OAC): Grant No. 2118201, and Schmidt Futures' AI2050 Early Career Fellowship. Simon Olsson is supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. Simon Olsson thanks NAISS for providing part of the computational infrastructure used to generate preliminary results for this work (Project No. SNIC 2022/22-616). Amey Pasarkar is supported by an NSF-GRFP fellowship. We thank Michele Invernizzi (FU Berlin) for insightful discussions.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Amey P. Pasarkar: Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – review & editing (equal). **Gianluca M. Bencomo:** Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – review & editing (equal). **Simon Olsson:** Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Adji Bousso Dieng:** Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES

- ¹K. Henzler-Wildman and D. Kern, "Dynamic personalities of proteins," *Nature* **450**, 964–972 (2007).
- ²D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Grigera, "Atomic-level characterization of the structural dynamics of proteins," *Science* **330**, 341–346 (2010).
- ³F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikel, "Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations," *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19011–19016 (2009).
- ⁴K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, "How fast-folding proteins fold," *Science* **334**, 517–520 (2011).
- ⁵S. Piana, K. Lindorff-Larsen, and D. E. Shaw, "Atomic-level description of ubiquitin folding," *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5915–5920 (2013).
- ⁶K. S. Chakrabarti, S. Olsson, S. Pratihari, K. Giller, K. Overkamp, K. O. Lee, V. Gapsys, K.-S. Ryu, B. L. de Groot, F. Noé, S. Becker, D. Lee, T. R. Weikel, and C. Griesinger, "A litmus test for classifying recognition mechanisms of transiently binding proteins," *Nat. Commun.* **13**, 3792 (2022).
- ⁷N. Plattner and F. Noé, "Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models," *Nat. Commun.* **6**, 7653 (2015).
- ⁸I. Buch, T. Giorgino, and G. De Fabritiis, "Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations," *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10184–10189 (2011).
- ⁹N. Plattner, S. Doerr, G. De Fabritiis, and F. Noé, "Complete protein-protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling," *Nat. Chem.* **9**, 1005–1011 (2017).
- ¹⁰S. Olsson and F. Noé, "Mechanistic models of chemical exchange induced relaxation in protein NMR," *J. Am. Chem. Soc.* **139**, 200–210 (2016).
- ¹¹F. Noé, S. Doose, I. Daidone, M. Löllmann, M. Sauer, J. D. Chodera, and J. C. Smith, "Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments," *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4822–4827 (2011).
- ¹²N.-V. Buchete and G. Hummer, "Coarse master equations for peptide folding dynamics," *J. Phys. Chem. B* **112**, 6057–6069 (2008).
- ¹³J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, and N. J. Skelton, *Protein NMR Spectroscopy: Principles and Practice* (Academic Press, 1996).
- ¹⁴O. Opanasyuk, A. Barth, T.-O. Peulen, S. Felekyan, S. Kalinin, H. Sanabria, and C. A. M. Seidel, "Unraveling multi-state molecular dynamics in single-molecule FRET experiments. II. Quantitative analysis of multi-state kinetic networks," *J. Chem. Phys.* **157**, 031501 (2022).
- ¹⁵J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, "Markov models of molecular kinetics: Generation and validation," *J. Chem. Phys.* **134**, 174105 (2011).
- ¹⁶T. Hempel, S. Olsson, and F. Noé, "Markov field models: Scaling molecular kinetics approaches to large molecular machines," *Curr. Opin. Struct. Biol.* **77**, 102458 (2022).
- ¹⁷S. Olsson and F. Noé, "Dynamic graphical models of molecular kinetics," *Proc. Natl. Acad. Sci. U. S. A.* **116**, 15001–15006 (2019).
- ¹⁸J. Hénin, T. Lelièvre, M. R. Shirts, O. Valsion, and L. Delemotte, "Enhanced sampling methods for molecular dynamics simulations [article v1.0]," *Living J. Comput. Mol. Sci.* **4**, 1583 (2022).
- ¹⁹C. Camilloni, A. Cavalli, and M. Vendruscolo, "Replica-averaged metadynamics," *J. Chem. Theory Comput.* **9**, 5610–5617 (2013).
- ²⁰C. Abrams and G. Bussi, "Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration," *Entropy* **16**, 163–199 (2013).
- ²¹H. Grubmüller, "Predicting slow structural transitions in macromolecular systems: Conformational flooding," *Phys. Rev. E* **52**, 2893–2906 (1995).
- ²²Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chem. Phys. Lett.* **314**, 141–151 (1999).
- ²³R. H. Swendsen and J.-S. Wang, "Replica Monte Carlo simulation of spin-glasses," *Phys. Rev. Lett.* **57**, 2607–2609 (1986).
- ²⁴A. Laio and M. Parrinello, "Escaping free-energy minima," *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12562–12566 (2002).
- ²⁵P. Kríž, Z. Šučur, and V. Spiwok, "Free-energy surface prediction by flying Gaussian method: Multisystem representation," *J. Phys. Chem. B* **121**, 10479–10483 (2017).
- ²⁶Z. Šučur and V. Spiwok, "Sampling enhancement and free energy prediction by the flying Gaussian method," *J. Chem. Theory Comput.* **12**, 4644–4650 (2016).
- ²⁷A. M. Ferrenberg and R. H. Swendsen, "Optimized Monte Carlo data analysis," *Comput. Phys.* **3**, 101–104 (1989).
- ²⁸M. R. Shirts and J. D. Chodera, "Statistically optimal analysis of samples from multiple equilibrium states," *J. Chem. Phys.* **129**, 124105 (2008).
- ²⁹H. Wu, F. Paul, C. Wehmeyer, and F. Noé, "Multiensemble Markov models of molecular thermodynamics and kinetics," *Proc. Natl. Acad. Sci. U. S. A.* **113**, E3221–E3230 (2016).
- ³⁰L. S. Stelzl and G. Hummer, "Kinetics from replica exchange molecular dynamics simulations," *J. Chem. Theory Comput.* **13**, 3927–3935 (2017).

- ³¹M. M. Galama, H. Wu, A. Krämer, M. Sadeghi, and F. Noé, "Stochastic approximation to MBAR and TRAM: Batchwise free energy estimation," *J. Chem. Theory Comput.* **19**, 758–766 (2023).
- ³²V. Limongelli, M. Bonomi, and M. Parrinello, "Funnel metadynamics as accurate binding free-energy method," *Proc. Natl. Acad. Sci. U. S. A.* **110**, 6358–6363 (2013).
- ³³J. M. L. Ribeiro, P. Bravo, Y. Wang, and P. Tiwary, "Reweighted autoencoded variational Bayes for enhanced sampling (RAVE)," *J. Chem. Phys.* **149**, 072301 (2018).
- ³⁴P. Tiwary and B. Berne, "Spectral gap optimization of order parameters for sampling complex molecular systems," *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2839–2844 (2016).
- ³⁵M. M. Sultan and V. S. Pande, "tICA-metadynamics: Accelerating metadynamics by using kinetically selected collective variables," *J. Chem. Theory Comput.* **13**, 2440–2447 (2017).
- ³⁶W. Chen and A. L. Ferguson, "Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration," *J. Comput. Chem.* **39**, 2079–2102 (2018).
- ³⁷L. Bonati, G. Piccini, and M. Parrinello, "Deep learning the slow modes for rare events sampling," *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2113533118 (2021).
- ³⁸F. Noé, S. Olsson, J. Köhler, and H. Wu, "Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning," *Science* **365**, eaaw1147 (2019).
- ³⁹J. Köhler, L. Klein, and F. Noé, "Equivariant flows: Exact likelihood generative learning for symmetric densities," in *International Conference on Machine Learning* (PMLR, 2020), pp. 5361–5370.
- ⁴⁰J. Köhler, A. Krämer, and F. Noé, "Smooth normalizing flows," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2021), Vol. 34, 2796–2809.
- ⁴¹B. Jing, G. Corso, J. Chang, R. Barzilay, and T. Jaakkola, "Torsional diffusion for molecular conformer generation," in *Neural Information Processing Systems 35 (NeurIPS 2022)* (Curran Associates, Inc., 2022).
- ⁴²D. Friedman and A. B. Dieng, "The Vendi Score: A diversity evaluation metric for machine learning," *arXiv:2210.02410* (2022).
- ⁴³M. J. Ferrarotti, S. Bottaro, A. Pérez-Villa, and G. Bussi, "Accurate multiple time step in biased molecular simulations," *J. Chem. Theory Comput.* **11**, 139–146 (2014).
- ⁴⁴E. Marinari and G. Parisi, "Simulated tempering: A new Monte Carlo scheme," *Europhys. Lett.* **19**, 451–458 (1992).
- ⁴⁵G. Hummer and J. Köfinger, "Bayesian ensemble refinement by replica simulations and reweighting," *J. Chem. Phys.* **143**, 243150 (2015).
- ⁴⁶A. Cavalli, C. Camilloni, and M. Vendruscolo, "Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle," *J. Chem. Phys.* **138**, 094112 (2013).
- ⁴⁷S. Olsson and A. Cavalli, "Quantification of entropy-loss in replica-averaged modeling," *J. Chem. Theory Comput.* **11**, 3973–3977 (2015).
- ⁴⁸K. Lindorff-Larsen, R. B. Best, M. A. DePristo, C. M. Dobson, and M. Vendruscolo, "Simultaneous determination of protein structure and dynamics," *Nature* **433**, 128–132 (2005).
- ⁴⁹R. B. Best and M. Vendruscolo, "Determination of protein structures consistent with NMR order parameters," *J. Am. Chem. Soc.* **126**, 8090–8091 (2004).
- ⁵⁰J. W. Pitera and J. D. Chodera, "On the use of experimental observations to bias simulated ensembles," *J. Chem. Theory Comput.* **8**, 3445–3451 (2012).
- ⁵¹B. Roux and J. Weare, "On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method," *J. Chem. Phys.* **138**, 084107 (2013).
- ⁵²M. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation* (Oxford University Press, 2010).
- ⁵³P. Eastman, R. Galvelis, J. Rodriguez-Guerra, Y. Chen, and J. D. Chodera, *Openmm pytorch plugin*, 2023.
- ⁵⁴P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, and V. S. Pande, "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics," *PLoS Comput. Biol.* **13**, e1005659 (2017).
- ⁵⁵B. Schölkopf, A. J. Smola, F. Bach *et al.*, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and beyond* (MIT Press, 2002).
- ⁵⁶P. Jaini, L. Holdijk, and M. Welling, "Learning equivariant energy based models with equivariant stein variational gradient descent," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2021), Vol. 34, pp. 16727–16737.
- ⁵⁷Y. Chen, A. Krämer, N. E. Charron, B. E. Husic, C. Clementi, and F. Noé, "Machine learning implicit solvation for molecular dynamics," *J. Chem. Phys.* **155**, 084101 (2021).
- ⁵⁸M. Invernizzi and M. Parrinello, "Rethinking metadynamics: From bias potentials to probability distributions," *J. Phys. Chem. Lett.* **11**, 2731–2736 (2020).
- ⁵⁹J. Brady and M. Karplus, "Configuration entropy of the alanine dipeptide in vacuum and in solution: A molecular dynamics study," *J. Am. Chem. Soc.* **107**, 6103–6105 (1985).
- ⁶⁰X. Wu and S. Wang, "Self-guided molecular dynamics simulation for efficient conformational search," *J. Phys. Chem. B* **102**, 7238–7250 (1998).
- ⁶¹K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, "Improved side-chain torsion potentials for the Amber ff99sb protein force field," *Proteins: Struct., Funct., Bioinf.* **78**, 1950–1958 (2010).
- ⁶²M. Sobieraj and P. Setny, "Granger causality analysis of chignolin folding," *J. Chem. Theory Comput.* **18**, 1936–1944 (2022).
- ⁶³H. Okumura, "Temperature and pressure denaturation of chignolin: Folding and unfolding simulation by multibaric-multithermal molecular dynamics method," *Proteins: Struct., Funct., Bioinf.* **80**, 2397–2416 (2012).
- ⁶⁴A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczyra, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *J. Phys. Chem. B* **102**, 3586–3616 (1998).
- ⁶⁵A. Onufriev, D. Bashford, and D. A. Case, "Exploring protein native states and large-scale conformational changes with a modified generalized born model," *Proteins: Struct., Funct., Bioinf.* **55**, 383–394 (2004).
- ⁶⁶G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, and F. Noé, "Identification of slow molecular order parameters for Markov model construction," *J. Chem. Phys.* **139**, 015102 (2013).
- ⁶⁷C. R. Schwantes and V. S. Pande, "Modeling molecular kinetics with tICA and the kernel trick," *J. Chem. Theory Comput.* **11**, 600–608 (2015).
- ⁶⁸L. Mones, N. Bernstein, and G. Csányi, "Exploration, sampling, and reconstruction of free energy surfaces with Gaussian process regression," *J. Chem. Theory Comput.* **12**, 5100–5110 (2016).
- ⁶⁹J. Debnath and M. Parrinello, "Gaussian mixture-based enhanced sampling for statics and dynamics," *J. Phys. Chem. Lett.* **11**, 5076–5080 (2020).
- ⁷⁰P. Maragakis, A. van der Vaart, and M. Karplus, "Gaussian-mixture umbrella sampling," *J. Phys. Chem. B* **113**, 4664–4673 (2009).
- ⁷¹M. Invernizzi and M. Parrinello, "Exploration vs convergence speed in adaptive-bias enhanced sampling," *J. Chem. Theory Comput.* **18**, 3988–3996 (2022).
- ⁷²A. Barducci, G. Bussi, and M. Parrinello, "Well-tempered metadynamics: A smoothly converging and tunable free-energy method," *Phys. Rev. Lett.* **100**, 020603 (2008).