Computing Absolute Free Energy with Deep Generative Models

Xinqiang Ding and Bin Zhang*

Cite This: *J. Phys. Chem. B* 2020, 124, 10166−10172

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Fast and accurate evaluation of free energy has broad applications from drug design to material engineering. Computing the absolute free energy is of particular interest since it allows the assessment of the relative stability between states without intermediates. Here, we introduce a general framework for calculating the absolute free energy of a state. A key step of the calculation is the definition of a reference state with tractable deep generative models using locally sampled configurations. The absolute free energy of this reference state is zero by design. The free energy for the state of interest can then be determined as the difference from the reference. We applied this approach to both discrete and continuous systems and demonstrated its effectiveness. It was found that the Bennett acceptance ratio method provides more accurate and efficient free energy estimations than approximate expressions based on work. We anticipate the method presented here to be a valuable strategy for computing free energy differences.
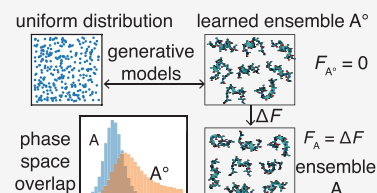
INTRODUCTION

Free energy is of central importance in both statistical physics and computational chemistry. It has important applications in rational drug design[1] and material property prediction.[2] Therefore, methodology development for efficient free energy calculations has attracted great research interest.[3−14] Many existing algorithms have focused on estimating free energy differences between states and originate from the free energy perturbation (FEP) identity:[15]

$$\mathbb{E}_A[e^{-\beta \Delta U}] = e^{-\beta \Delta F} \tag{1}$$

Here, $\Delta F = F_B - F_A$ is the free energy difference between two equilibrium states A and B at temperature T and $\beta = 1/k_B T$. $U_A(\boldsymbol{x})$ and $U_B(\boldsymbol{x})$ are the potential energies for a configuration $\boldsymbol{x}$ in states A and B, respectively, and $\Delta U(\boldsymbol{x}) = U_B(\boldsymbol{x}) - U_A(\boldsymbol{x})$. $\mathbb{E}_A$ represents the expectation with respect to the Boltzmann distribution of $\boldsymbol{x}$ in state A:

$$p_A(\boldsymbol{x}) = \frac{e^{-\beta U_A(\boldsymbol{x})}}{Z_A} \tag{2}$$

where the normalization constant $Z_A = \int e^{-\beta U_A(\boldsymbol{x})} d\boldsymbol{x}$. Computing ΔF with the FEP identity (eq 1) only uses samples from state A. It is more efficient to use samples from both states to compute ΔF by solving the Bennett acceptance ratio (BAR) equation:[16]

$$\sum_{k=1}^{N_A} f(\beta[\Delta U(\boldsymbol{x}_k^A) - M - \Delta F])$$
$$= \sum_{k=1}^{N_B} f(-\beta[\Delta U(\boldsymbol{x}_k^B) - M - \Delta F])) \tag{3}$$

where $f(t) = 1/(1 + e^t)$ and $M = \ln(N_B/N_A)$. Here, $\{\boldsymbol{x}_k^A, k = 1, ..., N_A\}$ and $\{\boldsymbol{x}_k^B, k = 1, ..., N_B\}$ are samples from the two states. Both the FEP and the BAR method converge poorly when the overlap in the configuration space between state A and B is small. In that case, multiple intermediate states along a path with incremental changes in the configuration space can be introduced to bridge the two states.[3] However, sampling from multiple intermediate states greatly increases the computational cost. It is, therefore, useful to develop techniques that can alleviate the convergence issue without the use of intermediate states.[12,13]

The requirement on a significant overlap between the two states' configuration space can be circumvented if we compute their free energy difference from the absolute free energy as $\Delta F = F_B - F_A$. The absolute free energy of a state A/B can be obtained from its difference from a reference state $A°/B°$ as $F_{A/B} = F_{A°/B°} - \Delta F_{A/B \to A°/B°}$. For this strategy to be efficient, however, the reference states must bear significant overlap in configuration space with the states of interest. Their absolute free energy should be available with minimal computational effort. For most systems, designing reference states that satisfy these constraints can be challenging and requires expertise and physical intuition.[17−23] In this work, we demonstrate that reference states can be constructed with tractable generative models for efficient computation of the absolute free energy.[24,25]
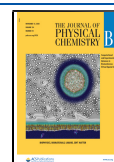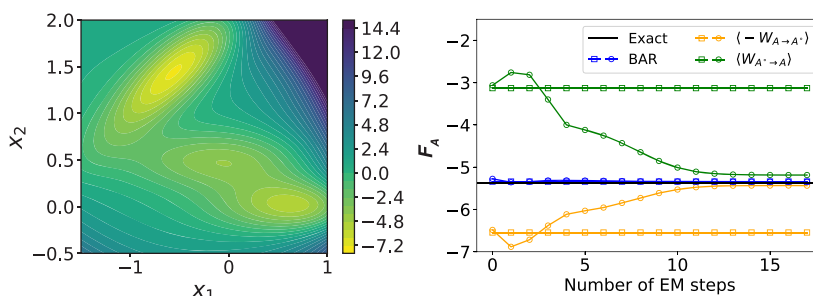
Figure 1. (Left) Contour plot of the Müller potential. Energy is shown in the units of k_BT. (Right) Absolute free energy computed using various estimators with reference states defined as a Gaussian distribution (squares) or a mixture model of two Gaussian distributions (circles). The x axis corresponds to the number of steps for training the Gaussian mixture model with the expectation−maximization (EM) algorithm.[36]

COMPUTATIONAL METHODS

The workflow for calculating the absolute free energy is as follows. State A is used as an example for the discussion, but the same procedure applies to state B. We first draw samples, $\{x_k^A, k = 1, ..., N_A\}$, from the Boltzmann distribution $p_A(x)$. We then learn a tractable generative model, $q_\theta(x)$, that maximizes the likelihood of observing these samples by fine-tuning the set of parameters θ. Here, tractable generative models refer to probabilistic models that have the following two properties: (i) the normalized probability (or probability density), $q_\theta(x)$, can be directly evaluated for a given configuration x without the need of sampling or integration; (ii) independent configurations can be efficiently sampled from the probability distribution. The generative model defines a new equilibrium state $A°$, which serves as an excellent reference to state A. Because it is parameterized from samples of state A, most probable configurations from $A°$ should resemble those from A by design, and the overlap between the two states is guaranteed as long as the generative model has enough flexibility for modeling $p_A(x)$. In addition, since $q_\theta(x)$ is normalized, if we define the potential energy of state $A°$ as $U_{A°}(x) = -(1/\beta)\ln q_\theta(x)$, the partition function $Z_{A°}$ of state $A°$ is equal to 1, i.e., $Z_{A°} = \int q_\theta(x)dx = 1$. The absolute free energy of the reference state $A°$ is $F_{A°} = -(1/\beta)\ln Z_{A°} = 0$. (Strictly speaking, the free energy should be defined as $F_{A°} = -(1/\beta)\ln(Z_{A°}/\int 1dx)$ to normalize the unit in the partition function. This technical detail does not affect any of the conclusions on free energy differences and is not considered for simplicity.) With the reference state defined, the absolute free energy for state A can be determined by solving a similar BAR equation as eq 3. Our use of tractable generative models ensures that sample configurations can be easily produced for the reference state to be combined with those from state A for solving the BAR equation.[26]

We note that a closely related algorithm for computing the absolute free energy has been introduced in variational methods.[27,28] In these prior studies, $q_\theta(x)$ was optimized by minimizing the Kullback−Leibler (KL) divergence[29] from $q_\theta(x)$ to $p_A(x)$:

$$D_{KL}(q_\theta \| p_A) = \int q_\theta(x)\ln\frac{q_\theta(x)}{p_A(x)}\,dx$$
$$= \beta(\langle W_{A°\to A}\rangle - F_A) \tag{4}$$

where $\langle W_{A°\to A}\rangle = \mathbb{E}_{A°}[U_A(x) - U_{A°}(x)]$. Because $D_{KL}(q_\theta\|p_A)$ is non-negative, $\langle W_{A°\to A}\rangle$ is an upper bound of F_A. As

$D_{KL}(q_\theta\|p_A)$ decreases along the optimization, $\langle W_{A°\to A}\rangle$ is assumed to approach closer to the true free energy and was used for its estimation.

Our methodology is different from the variational methods[27,28,30] in two aspects. First, instead of $D_{KL}(q_\theta\|p_A)$, we used

$$D_{KL}(p_A \| q_\theta) = \int p_A(x)\ln\frac{p_A(x)}{q_\theta(x)}\,dx$$
$$= \beta(\langle W_{A\to A°}\rangle + F_A) \tag{5}$$

as the objective function for learning $q_\theta(x)$. $\langle W_{A\to A°}\rangle = \mathbb{E}_A[U_{A°}(x) - U_A(x)]$. We note that minimizing the KL divergence from $p_A(x)$ to $q_\theta(x)$ is equivalent to learning the generative model by maximizing its likelihood on the training data. Moreover, because $D_{KL}(p_A\|q_\theta)$ is also non-negative, $\langle -W_{A\to A°}\rangle$ is a lower bound of F_A. Therefore, minimizing $D_{KL}(p_A\|q_\theta)$ is equivalent to maximizing the lower bound $\langle -W_{A\to A°}\rangle$. At the face value, it may seem that $D_{KL}(q_\theta\|p_A)$ is a better objective function than $D_{KL}(p_A\|q_\theta)$ for model training since its optimization only requires samples from $q_\theta(x)$. As aforementioned, sampling from $q_\theta(x)$ can be made computationally efficient by the use of tractable generative models. On the other hand, training by $D_{KL}(p_A\|q_\theta)$ requires samples from $p_A(x)$, the collection of which often requires costly long timescale simulations with Monte Carlo (MC) or molecular dynamics (MD) techniques. The caveat is that optimization with $D_{KL}(q_\theta\|p_A)$ is more susceptible to traps from local minima due to its more complex dependence on q_θ.[31] When $p_A(x)$ is a high-dimensional distribution and the system exhibits multistability, optimizing $D_{KL}(q_\theta\|p_A)$ often leads to solutions that cover only one of the metastable states.[32,33] Noè and co-workers have recognized the above challenge,[32] and they introduced the Boltzmann generator that uses a combination of both $D_{KL}(q_\theta\|p_A)$ and $D_{KL}(p_A\|q_\theta)$ for model training.

Another significant difference between our methodology and the variational methods[27,28,34] or the Boltzmann generator is the expression used to estimate F_A. In particular, $\langle W_{A°\to A}\rangle$ is an upper bound of the free energy and only becomes exact when the probability distributions from generative models and the state of interest are identical. On the other hand, our use of the BAR equation (eq 3) relaxes this requirement, and F_A can be accurately determined even if the model training is not perfect and there are significant differences between the two distributions. In all but trivial examples, we anticipate that the learning process does not converge exactly to the true
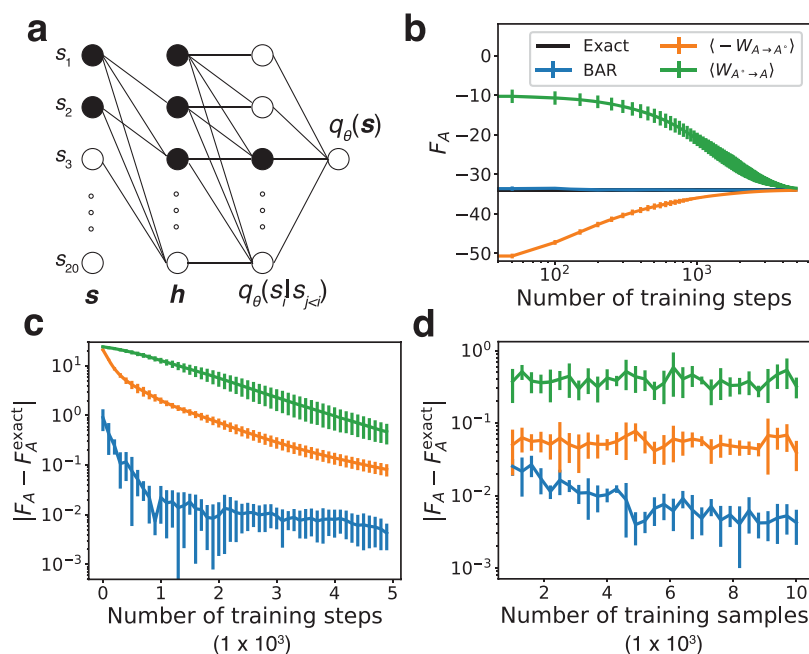
Figure 2. Performance of different free energy estimators on the SK model. Energy is shown in the units of k_BT. (a) Schematic representation of the neural autoregressive model used to parameterize q_θ. The black units illustrate the dependence of the conditional probability $q(s_3|s_1, s_2)$. (b) Absolute free energy of the SK model calculated with three estimators as a function of training steps compared with the exact result obtained from a complete enumeration. (c, d) Errors of the estimated absolute free energy versus the number of training steps and the number of samples used for training q_θ. The coloring scheme is identical to that in part b. Error bars representing one standard deviation are estimated using five independent repeats.

distribution $p_A(x)$ due to its high dimensionality and complexity. The BAR estimation, which is asymptotically unbiased,[35] will be crucial to ensure the accuracy of free energy calculations.

■ RESULTS

The advantage of the BAR estimation is evident when computing the absolute free energy of a two-dimensional system with the Müller potential.[37] When the reference state $q_\theta(x)$ was parameterized with a Gaussian distribution, which fails to capture the multistability inherent to the system, the two bounds based on work deviate significantly from the absolute free energy (Figure 1). The free energy estimated using the BAR equation, on the other hand, is in excellent agreement with the exact value. The work-based bounds begin to approach the exact value when an optimized mixture model of two Gaussian distributions was used to parameterize the reference state. The BAR estimator again converges much faster than the bounds, highlighting its insensitivity to the quality of the reference state. More details about the model training and free energy computation for this simple test system are included in the Supporting Information. We note that an independent study reported similar advantages when using BAR to compute relative free energy with deep generative models.[13]

Encouraged by the results from the above test system, we next computed the absolute free energy of a 20-spin classical Sherrington–Kerkpatrick (SK) model,[38] the value of which can be determined from complete enumeration as well. The discrete configurations of the SK model will be represented using s instead of x. Though we introduced the methodology with continuous variables, all the equations can be trivially extended to s by replacing the integrals with summations over the spin configurations. The potential energy of a configuration $s = (s_1, s_2, ..., s_N)$ is defined as

$$U_A(s) = \frac{1}{\sqrt{N}} \sum_{j>i} J_{ij} s_i s_j \tag{6}$$

where $s_i \in \{-1, +1\}$ and $N = 20$. J_{ij} were chosen randomly from the standard normal distribution. Five thousand samples were drawn from the probability distribution $p(s) = e^{-\beta U_A(s)}/Z_A$ with $\beta = 2.0$. These samples were used to train the reference state $A°$ by minimizing $D_{KL}(p_A \| q_\theta)$ (eq 5). The reference probability $q_\theta(s)$ was defined with a neural autoregressive density estimator (NADE)[24,39−42] as a product of conditional distributions:

$$q_\theta(s) = \prod_{i=1}^{N} q_\theta(s_i|s_1, ..., s_{i-1}) \tag{7}$$

$q_\theta(s_i|s_1, ..., s_{i-1})$ were parameterized using a feed-forward neural network with one hidden layer of 20 hidden units. The neural network's connections are specifically designed such that it maintains the autoregressive property, i.e., $q_\theta(s_i|s_1, ..., s_{i-1})$ only depends on $s_1, ..., s_i$ (Figure 2a). After training $q_\theta(s)$ for some numbers of steps,[43,44] 5000 configurations were independently drawn from $q_\theta(s)$. These configurations, together with the training inputs sampled from $p_A(s)$, were used to determine the absolute free energy of the SK model. In Figure 2b,c, we again compare results from the three estimators with the exact value.

Similar to the results observed for the Müller potential, at early stages of model parameterization with small training step numbers, the work-based estimations deviate significantly from the true value. This deviation is expected and is a direct result of the difference between the two probability distributions
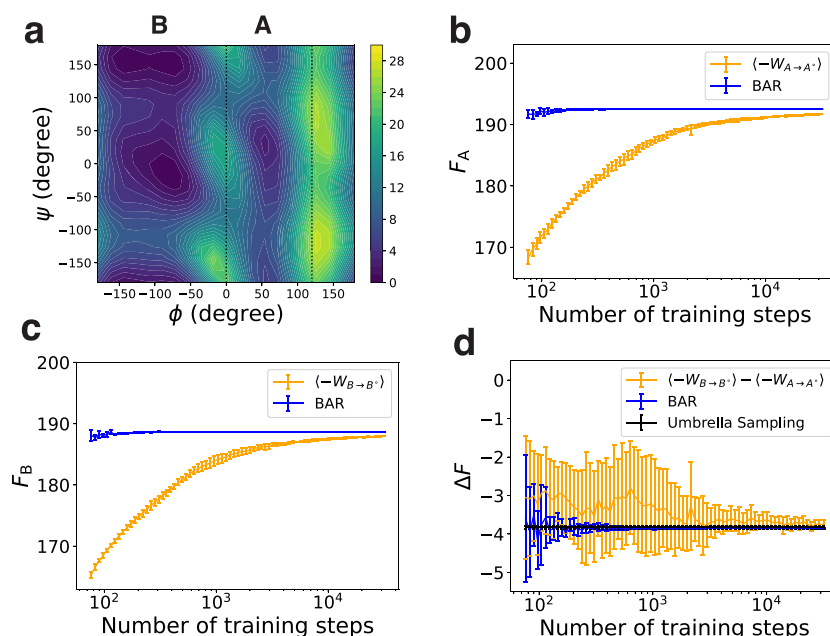
Figure 3. Performance of different free energy estimators on di-alanine. Energy is shown in the units of k_BT. (a) Contour plot of the di-alanine free energy surface as a function of the two torsion angles ϕ and ψ. (b, c) Absolute free energy of state A and B computed with different estimators. (d) Free energy difference between state A and B computed with different estimators. Error bars representing one standard deviation are estimated using five independent repeats.
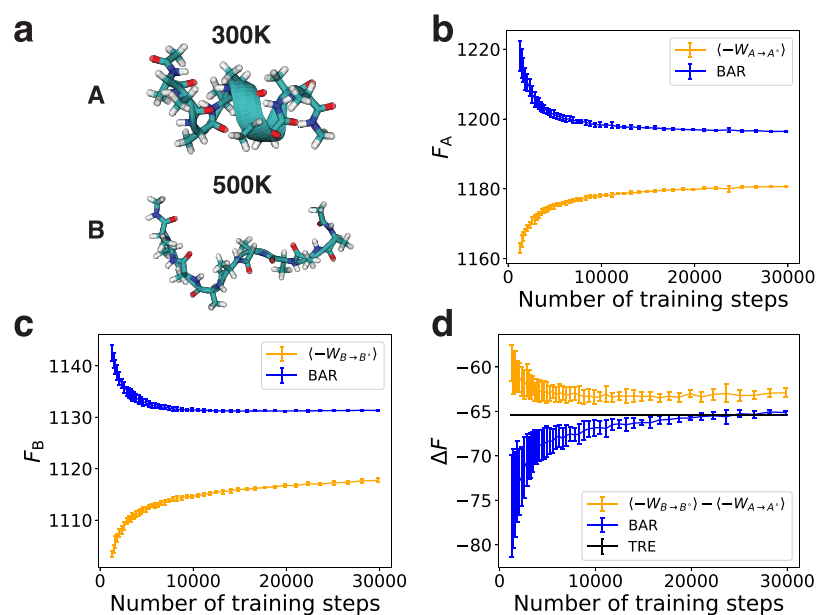

Figure 4. Performance of different free energy estimators on deca-alanine. Energy is shown in the units of k_BT, with T = 300 K. (a) Representative conformations from state A (the ensemble at T = 300 K) and state B (the ensemble at T = 500 K). (b, c) Absolute free energy of states A and B computed with different estimators. (d) Free energy difference between states A and B computed with different estimators. Error bars representing one standard deviation are estimated using five independent repeats.

$q_\theta(s)$ and $p_A(s)$. However, as the training proceeds, the agreement between the distributions improves and $\langle - W_{A \to A^\circ} \rangle$ and $\langle W_{A^\circ \to A} \rangle$ gradually converge to the exact result after 5000 steps (Figure 2b) because the autoregressive model is flexible enough to match the target distribution. On the other hand, the BAR estimator converges much faster to the exact value with a smaller error (Figure 2b,c). In addition, varying the number of samples used for training $q_\theta(s)$ has different effects on the accuracy of converged results for the three approaches (Figure 2d). For both $\langle - W_{A \to A^\circ} \rangle$ and $\langle W_{A^\circ \to A} \rangle$, increasing the

number of training samples from 10^3 to 10^4 does not significantly change the accuracy of their results. In contrast, using more training samples significantly reduces the error of the BAR estimator. This is because solutions of the BAR equation are asymptotically unbiased for estimating F_A, whereas $\langle - W_{A \to A^\circ} \rangle$ and $\langle W_{A^\circ \to A} \rangle$ are not.[35]

Finally, we applied the methodology to two molecular systems, di-alanine and deca-alanine in an implicit solvent. These two systems present features commonly encountered in biomolecular simulations with continuous phase space over a

rugged energy landscape. Their high dimensionality renders a complete enumeration of the configurational space to compute the absolute free energy for benchmarking impractical. Instead, we calculated the free energy difference between two metastable states using their absolute free energy to compare against the value determined from umbrella sampling and temperature replica exchange (TRE) simulations. For di-alanine, the two metastable states were defined using the backbone dihedral angle ϕ (C−CA−N−C), with $0° < \phi \leq 120°$ for state A and $\phi \leq 0°$ or $\phi > 120°$ for state B (Figure 3a). For deca-alanine, states A and B were defined as the configurational ensembles at $T = 300$ K and $T = 500$ K, respectively (Figure 4a).

To compute the absolute free energy, we learned the reference states using normalizing flow based generative models.[39,45] Specifically, $q_\theta(\boldsymbol{x})$ was parameterized with multiple bijective transformations, $T_1, ..., T_K$, to convert a random variable $\boldsymbol{u}$ to a peptide configuration, i.e.

$$\boldsymbol{x} = T(\boldsymbol{u}) = T_K \bigcirc \cdots \bigcirc T_1(\boldsymbol{u}) \tag{8}$$

$\boldsymbol{u}$ shares the same dimension as $\boldsymbol{x}$ and is from a simple base distribution $p_u(\boldsymbol{u})$. Based on the formula of variable change in probability density functions, we have

$$\ln q_\theta(\boldsymbol{x}) = \ln p_u(\boldsymbol{u}) - \sum_{k=1}^{K} \ln |J_{T_k}(\boldsymbol{u}_{k-1})| \tag{9}$$

where $\boldsymbol{u}_k = T_k \circ \cdots T_1(\boldsymbol{u}_0)$ and $\boldsymbol{u}_0 = \boldsymbol{u}$. J_{T_k} is the Jacobian matrix of the transformation T_k, and $|\cdot|$ denotes the absolute value of the determinant. For both molecules, we first transformed $\boldsymbol{u}$ into the internal coordinates $\boldsymbol{z}$ using the neural spline flows[46,47] with coupling layers[25] (Figs. S2 and S3) and then transformed $\boldsymbol{z}$ into the Cartesian coordinates $\boldsymbol{x}$ based on molecular topology (Fig. S1).

The reference models were separately trained using configurations collected for each state from molecular dynamics simulations with the Amber ff99SB force field[48] and the OBC implicit solvent model.[49] As shown in Figures S4−S7, they succeed in generating peptide conformations with reasonable geometry and energy. With the learned reference states, we computed the absolute free energy for states A and B using the three estimators. As shown in Figures 3b,c and 4b,c, the BAR estimator converges much faster than the upper and lower bounds. The results calculated using the upper bound are not shown here because they are much larger than that of the lower bound and the BAR estimator (Figures S8 and S9). Unlike the results for the SK model, the two bounds no longer converge to the same value or the BAR estimator, and their difference can be as large as 6 $k_B T$ for di-alanine (Figure S8) and 60 $k_B T$ for deca-alanine (Figure S9). The large gaps between the two bounds suggest that the generative models are still quite different from the true distributions even after the learning has converged. We expect the numbers from the BAR estimator to be correct because the BAR estimator does not require the generative models to precisely match the original distributions to reproduce the free energy, as shown in both the Müller system and the SK model. Furthermore, the BAR estimations lie in between the two bounds in all four cases (Figures S8 and S9), as expected for the exact values. Therefore, for these two molecular systems, the two bounds cannot be used for reliable estimation of the absolute free energy.

We further evaluated the accuracy of the three estimators in computing the free energy differences between states A and B. For comparison, we also determined the free energy difference using umbrella sampling[3] for di-alanine and TRE simulations for deca-alanine. Results of estimated free energy differences are shown in Figures 3d and 4d. The BAR estimator converges much faster to the results from umbrella sampling or TRE simulations than the two bounds. For di-alanine, the free energy difference estimated using BAR is -3.86 ± 0.01 $k_B T$, which agrees with the result from umbrella sampling (-3.82 ± 0.04 $k_B T$). To our surprise, the difference computed using the lower bound, -3.74 ± 0.10 $k_B T$, is close to the correct result as well. Because the lower bound is biased, we believe that its good performance on the free energy difference is due to error cancellation. For deca-alanine, the free energy difference from TRE is -65.37 ± 0.02 $k_B T$, which deviates from the result obtained from the lower bound (-62.91 ± 0.52 $k_B T$) but agrees well the BAR estimation (-65.13 ± 0.23 $k_B T$).

■ CONCLUSIONS AND DISCUSSION

In summary, we demonstrated that the framework based on deep generative models succeeds at computing the absolute free energy using sample configurations from the state of interest and is applicable for both discrete and continuous systems. Compared with harmonic approximations[18] or histograms[21] used in previous methods to construct reference systems, deep generative models are more flexible and can capture the complex correlation among various degrees of freedom. Their flexibility is crucial for improving the overlap between the learned reference system and the state of interest and for ensuring the convergence of free energy estimations using the BAR method. Although the examples tested here are relatively small in size, the free energy calculation method can be readily applied to larger lattice models and more complex biomolecular systems in the gas phase or implicit solvent.

Additional challenges must be addressed, however, before applying the proposed method to systems with explicit solvation. Similar to other techniques,[50−56] the accuracy of computed free energy relies on the statistical convergence of conformational sampling, which can become more challenging with the presence of solvent molecules. Instead of collecting training data from unbiased molecular dynamics (MD) simulations, enhanced sampling methods such as solute tempering[57] may be necessary to improve the efficiency of configurational space exploration. In addition, including solvent molecules significantly increases the system size and introduces additional symmetry requirements for the generative models. While the issues with a larger system size can in principle be overcome with more powerful computers, efficiently encoding the permutational symmetry of solvent molecules into generative models' architectures is less straightforward. It is worth noting that multiple studies[13,58,59] have introduced approaches for designing deep generative models that are invariant to permutations. Combining these approaches with our framework to compute the absolute free energy of biomolecular systems with explicit solvent would be an exciting direction for future studies. It could greatly facilitate the evaluation of protein−ligand binding affinity and protein conformational stability while accounting for entropic contributions.

■ ASSOCIATED CONTENT

🅢ⓘ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.0c08645.

Architectures of neural autoregressive density estimator and normalizing flow models, Müller potential system, di-alanine system, deca-alanine system, (Figure S1) illustration of the transformation between internal and Cartesian coordinates, (Figure S2) normalizing flows used for learning reference states of di-alanine and deca-alanine, (Figure S3) illustration of the RQ-NSF(C) layer used to derive conditional values of angles and dihedrals from random variables and bond lengths, (Figure S4) the generative model succeeds in producing molecular conformations for di-alanine in state *B* resembling those from molecular dynamics simulations, (Figure S5) the generative model succeeds in producing molecular conformations for di-alanine resembling those from molecular dynamics simulations, (Figure S6) the generative model succeeds in producing molecular conformations for deca-alanine resembling those from molecular dynamics simulations at 300 K, (Figure S7) the generative model succeeds in producing molecular conformations for deca-alanine resembling those from molecular dynamics simulations at 500 K, (Figure S8) performance of different free energy estimators on di-alanine, and (Figure S9) performance of different free energy estimators on deca-alanine (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Bin Zhang — *Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States;* ⓘ orcid.org/0000-0002-3685-7503; Email: binz@mit.edu

Author

Xinqiang Ding — *Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jpcb.0c08645

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813−1818.

(2) Auer, S.; Frenkel, D. Prediction of absolute crystal-nucleation rate in hard-sphere colloids. *Nature* **2001**, *409*, 1020−1023.

(3) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.

(4) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(5) Jorgensen, W. L.; Ravimohan, C. Monte Carlo simulation of differences in free energies of hydration. *J. Chem. Phys* **1985**, *83*, 3050−3054.

(6) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys* **2008**, *129*, 124105.

(7) Schneider, E.; Dai, L.; Topper, R. Q.; Drechsel-Grau, C.; Tuckerman, M. E. Stochastic neural network approach for learning high-dimensional free energy surfaces. *Phys. Rev. Lett.* **2017**, *119*, 150601.

(8) Pohorille, A.; Jarzynski, C.; Chipot, C. Good practices in free-energy calculations. *J. Phys. Chem. B* **2010**, 10235.

(9) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the analysis of free energy calculations. *J. Comput.-Aided Mol. Des.* **2015**, 397.

(10) Kollman, P. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.

(11) Hahn, A. M.; Then, H. Using bijective maps to improve free-energy estimates. *Phys. Rev. E* **2009**, *79*, No. 011113.

(12) Jarzynski, C. Targeted free energy perturbation. *Phys. Rev. E* **2002**, *65*, 5.

(13) Wirnsberger, P.; Ballard, A. J.; Papamakarios, G.; Abercrombie, S.; Racanière, S.; Pritzel, A.; Jimenez Rezende, D.; Blundell, C. Targeted free energy estimation via learned mappings. *J. Chem. Phys* **2020**, *153*, 144112.

(14) Ding, X.; Vilseck, J. Z.; Hayes, R. L.; Brooks, C. L., III Gibbs sampler-based λ-dynamics and Rao−Blackwell estimator for alchemical free energy calculation. *J. Chem. Theory Comput.* **2017**, *13*, 2501−2510.

(15) Zwanzig, R. W. High-temperature equation of state by a perturbation method. i. non polar gases. *J. Chem. Phys* **1954**, *22*, 1420−1426.

(16) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.

(17) Hoover, W. G.; Gray, S. G.; Johnson, K. W. Thermodynamic properties of the fluid and solid phases for inverse power potentials. *J. Chem. Phys* **1971**, *55*, 1128−1136.

(18) Frenkel, D.; Ladd, A. J. C. New Monte Carlo method to compute the free energy of arbitrary solids. Application to the fcc and hcp phases of hard spheres. *J. Chem. Phys* **1984**, *81*, 3188−3193.

(19) Hoover, W. G.; Ree, F. H. Use of computer experiments to locate the melting transition and calculate the entropy in the solid phase. *J. Chem. Phys* **1967**, *47*, 4873−4878.

(20) Amon, L. M.; Reinhardt, W. P. Development of reference states for use in absolute free energy calculations of atomic clusters with application to 55-atom Lennard-Jones clusters in the solid and liquid states. *J. Chem. Phys* **2000**, *113*, 3573−3590.

(21) Ytreberg, F. M.; Zuckerman, D. M. Simple estimation of absolute free energies for biomolecules. *J. Chem. Phys* **2006**, *124*, 104105.

(22) Schilling, T.; Schmid, F. Computing absolute free energies of disordered structures by molecular simulation. *J. Chem. Phys* **2009**, *131*, 231102.

(23) Berryman, J. T.; Schilling, T. Free energies by thermodynamic integration relative to an exact solution, used to find the handedness-switching salt concentration for DNA. *J. Chem. Theory Comput.* **2013**, *9*, 679−686.

(24) Uria, B.; Côtè, M.-A.; Gregor, K.; Murray, I.; Larochelle, H. Neural autoregressive distribution estimation. *J. Mach. Learn. Res* **2016**, *17*, 1−37.

(25) Dinh, L.; Sohl-Dickstein, J.; Bengio, S. *Density Estimation Using Real NVP.* 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24−26, 2017, Conference Track Proceedings. 2017.

(26) Ding, X.; Vilseck, J. Z.; Brooks, C. L., III Fast solver for large scale multistate Bennett acceptance ratio equations. *J. Chem. Theory Comput.* **2019**, *15*, 799−802.

(27) Wu, D.; Wang, L.; Zhang, P. Solving statistical mechanics using variational autoregressive networks. *Phys. Rev. Lett.* **2019**, *122*, 80602.

(28) Li, S.-H.; Wang, L. Neural network renormalization group. *Phys. Rev. Lett.* **2018**, *121*, 260601.

(29) Kullback, S.; Leibler, R. A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79−86.

(30) Nicoli, K. A.; Nakajima, S.; Strodthoff, N.; Samek, W.; Müller, K.-R.; Kessel, P. Asymptotically unbiased estimation of physical observables with neural samplers. *Phys. Rev. E* **2020**, *101*, No. 023304.

(31) Cover, T. M.; Thomas, J. A. *Elements of Information Theory*; Wiley-Interscience: USA, 2006.

(32) Noè, F.; Olsson, S.; Köhler, J.; Wu, H. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* **2019**, *365*, eaaw1147.

(33) Wu, H.; Khler, J.; Noè, F. Stochastic normalizing flows. *arXiv preprint arXiv:2002.06707* **2020**,

(34) Nicoli, K. A.; Anders, C. J.; Funcke, L.; Hartung, T.; Jansen, K.; Kessel, P.; Nakajima, S.; Stornati, P. On estimation of thermodynamic observables in lattice field theories with deep generative models. *arXiv preprint arXiv:2007.07115* **2020**,

(35) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys. Rev. Lett.* **2003**, *91*, 140601.

(36) Dempster, A. P.; Laird, N. M.; Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc.* **1977**, *39*, 1−22.

(37) Müller, K.; Brown, L. D. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theoret. Chim. Acta* **1979**, *53*, 75−93.

(38) Sherrington, D.; Kirkpatrick, S. Solvable model of a spin-glass. *Phys. Rev. Lett.* **1975**, *35*, 1792−1796.

(39) Papamakarios, G.; Nalisnick, E.; Rezende, D. J.; Mohamed, S.; Lakshminarayanan, B. Normalizing flows for probabilistic modeling and inference. *arXiv preprint arXiv:1912.02762* **2019**,

(40) Kingma, D. P.; Salimans, T.; Jozefowicz, R.; Chen, X.; Sutskever, I.; Welling, M. In *Advances in Neural Information Processing Systems 29*; Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., Garnett, R., Eds.; Curran Associates, Inc., 2016; pp. 4743−4751.

(41) Papamakarios, G.; Pavlakou, T.; Murray, I. In *Advances in Neural Information Processing Systems 30*; Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc., 2017; pp. 2338−2347.

(42) Huang, C.; Krueger, D.; Lacoste, A.; Courville, A. C. *Neural Autoregressive Flows*. Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10−15, 2018. 2018; pp. 2083−2092.

(43) Kingma, D. P.; Ba, J. *Adam: A Method for Stochastic Optimization*. 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7−9, 2015, Conference Track Proceedings. 2015.

(44) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. et al. In *Advances in Neural Information Processing Systems 32*; Wallach, H., Larochelle, H., Beygelzimer, A., dAlchBuc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc., 2019; pp. 8026−8037.

(45) Rezende, D. J.; Mohamed, S. *Variational Inference with Normalizing Flows*. Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6−11 July 2015. 2015; pp. 1530−1538.

(46) Durkan, C.; Bekasov, A.; Murray, I.; Papamakarios, G. In *Advances in Neural Information Processing Systems 32*; Wallach, H., Larochelle, H., Beygelzimer, A., dAlchèBuc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc., 2019; pp. 7511−7522.

(47) Rezende, D. J.; Papamakarios, G.; Racanière, S.; Albergo, M. S.; Kanwar, G.; Shanahan, P. E.; Cranmer, K. Normalizing flows on tori and spheres. *arXiv preprint arXiv:2002.02428* **2020**,

(48) Tian, C.; Kasavajhala, K.; Belfon, K. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; et al. ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* **2019**, *16*, 528−552 10.1021/acs.jctc.9b00591..

(49) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55*, 383−394.

(50) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **2015**, *137*, 2695−2703.

(51) Woo, H.-J.; Roux, B. Calculation of absolute protein−ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825−6830.

(52) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150−160.

(53) Shirts, M. R.; Mobley, D. L.; Chodera, J. D. In *Chapter 4 Alchemical Free Energy Calculations: Ready for Prime Time?*; Spellmeyer, D., Wheeler, R., Eds.; Annual Reports in Computational Chemistry; Elsevier, 2007; Vol. 3; pp. 41−59, DOI: 10.1016/S1574-1400(07)03004-6.

(54) Knight, J. L.; Brooks, C. L., III λ-Dynamics free energy simulation methods. *J. Comput. Chem.* **2009**, *30*, 1692−1700.

(55) Hayes, R. L.; Armacost, K. A.; Vilseck, J. Z.; Brooks, C. L., III Adaptive landscape flattening accelerates sampling of alchemical space in multisite λ dynamics. *J. Phys. Chem. B* **2017**, *121*, 3626−3635.

(56) Kong, X.; Brooks, C. L., III λ-dynamics: A new approach to free energy calculations. *J. Chem. Phys* **1996**, *105*, 2414−2423.

(57) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13749−13754.

(58) Bender, C. M.; Garcia, J. J.; O'Connor, K.; Oliva, J. Permutation invariant likelihoods and equivariant transformations. *arXiv preprint arXiv:1902.01967* **2019**,

(59) Köhler, J.; Klein, L.; Noè, F. Equivariant flows: sampling configurations for multi-body systems with symmetric energies. *arXiv preprint arXiv:1910.00753* **2019**,