THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL

Letter

Skipping the Replica Exchange Ladder with Normalizing Flows

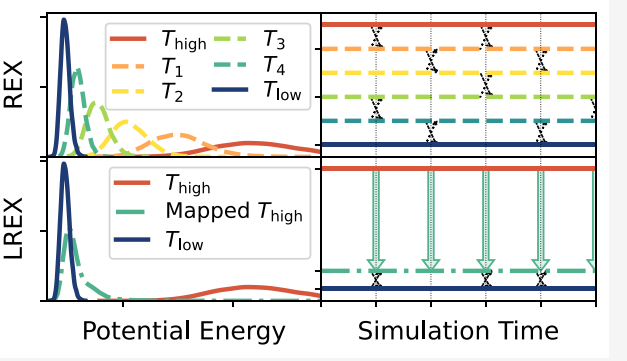
Michele Invernizzi,* Andreas Krämer, Cecilia Clementi, and Frank Noé*

Cite This: J. Phys. Chem. Lett. 2022, 13, 11643−11649

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We combine replica exchange (parallel tempering) with normalizing flows, a class of deep generative models. These two sampling strategies complement each other, resulting in an efficient method for sampling molecular systems characterized by rare events, which we call learned replica exchange (LREX). In LREX, a normalizing flow is trained to map the configurations of the fastest-mixing replica into configurations belonging to the target distribution, allowing direct exchanges between the two without the need to simulate intermediate replicas. This can significantly reduce the computational cost compared to standard replica exchange. The proposed method also offers several advantages with respect to Boltzmann generators that directly use normalizing flows to sample the target distribution. We apply LREX to some prototypical molecular dynamics systems, highlighting the improvements over previous methods.



Molecular simulations are becoming more and more important for studying complex phenomena in physics, chemistry, and biology. One of the long-lasting challenges for molecular simulations is to efficiently sample the equilibrium Boltzmann distribution, which is often characterized by multiple metastable states. In this letter, we propose a novel sampling method that combines the well-established replica exchange method[1,2] with a recent machine-learning technique known as normalizing flows.[3−5] To this end, we first summarize the two original methods and discuss their strengths and weaknesses. We then introduce flow-based exchange moves and show with a few prototypical examples how they help outperform classical replica exchange.

Replica exchange (REX), also known as parallel tempering, is a popular enhanced sampling method.[1,2] It uses a ladder of overlapping distributions to connect the target distribution one wishes to sample with an easier-to-sample probability distribution, for example at higher temperature, that we call here prior distribution. We keep a general notation and indicate with $q(\mathbf{x}) = e^{-u_q(\mathbf{x})}/Z_q$ the prior distribution and with $p(\mathbf{x}) = e^{-u_p(\mathbf{x})}/Z_p$ the target, where u_p and u_q are the respective reduced energies, $Z_q = \int e^{-u_q(\mathbf{x})}d\mathbf{x}$ and $Z_p = \int e^{-u_p(\mathbf{x})}d\mathbf{x}$ are partition functions, and $\mathbf{x}$ is the system configuration. A set of $M + 1$ replicas of the system are chosen to form a ladder of Boltzmann distributions $p_i(\mathbf{x}) \propto e^{-u_i(\mathbf{x})}$ from $p_0(\mathbf{x}) = p(\mathbf{x})$ to $p_M(\mathbf{x}) = q(\mathbf{x})$, such that each $p_i(\mathbf{x})$ overlaps with its neighbors in configuration space. In the typical case of temperature expansion, one has $u_i(\mathbf{x}) = (k_B T_i)^{-1} U(\mathbf{x})$, where $U(\mathbf{x})$ is the potential energy, k_B is the Boltzmann constant, and the temperatures T_i interpolate between the temperatures of the target and prior distributions, $T_0 = T_{low}$ and $T_M = T_{high}$, respectively. Other kinds of expansions are also possible, such as solute tempering[6] or alchemical transformations,[7] generally known as Hamiltonian replica exchange. Each replica is sampled with local moves, such as Markov chain Monte Carlo or molecular dynamics (MD), and at regular time intervals an exchange is proposed between the configurations of different replicas, $\mathbf{x}_i \rightleftarrows \mathbf{x}_j$, and accepted with probability

$$\alpha_{REX} = \min\left\{1, \frac{p_j(\mathbf{x}_i)}{p_i(\mathbf{x}_i)}\frac{p_i(\mathbf{x}_j)}{p_j(\mathbf{x}_j)}\right\} = \min\{1, e^{\Delta u_{ij}(\mathbf{x}_i) - \Delta u_{ij}(\mathbf{x}_j)}\} \quad (1)$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are configurations sampled from the corresponding distribution p_i or p_j and $\Delta u_{ij}(\mathbf{x}) = u_i(\mathbf{x}) - u_j(\mathbf{x})$ is the difference in reduced energy between replica i and replica j. The best choice of intermediate p_i is not trivial even in the well-studied case of temperature REX, and several different approaches have been proposed to optimize it.[2,7] The total number of replicas M required to allow exchanges increases both with the number of degrees of freedom of the system N, $M \propto \sqrt{N}$, and with the distance (e.g., in temperature) between prior and target.[8]

Received: November 2, 2022
Accepted: December 7, 2022
Published: December 9, 2022

One of the main limitations of REX is that a large number of replicas might be necessary to connect the target and the prior distributions, making the method computationally too expensive. There are variants of REX that do not require a fixed number of parallel simulations, such as simulated tempering,[9] or expanded ensemble methods,[10,11] but they share the same $\sqrt{N}$ scaling of sampling efficiency as REX (see Figure S2 of the Supporting Information (SI)). Several approaches have been proposed to mitigate this scaling, such as using nonequilibrium switches,[12,13] but with limited practical success. Two popular strategies to reduce the total number of replicas are to apply the tempering only to part of the system[6] or to broaden the distributions with metadynamics.[14] We will show how to combine REX with normalizing flows to avoid altogether the need to simulate intermediate replicas.

Normalizing flows (NFs) are a class of invertible deep neural networks that can be used to generate samples according to a given target distribution and are at the core of the recently proposed Boltzmann generators.[15] A normalizing flow f is an invertible function that maps a configuration $\mathbf{x}$ drawn from a prior distribution, $q(\mathbf{x})$, into a new configuration $\mathbf{x}' = f(\mathbf{x})$ that samples the output distribution of the flow, $q'(\mathbf{x})$, also called the mapped distribution. Exploiting the invertibility of f, one can compute the output probability density as

$$q'(\mathbf{x}') = q(\mathbf{x})|\det J_f(\mathbf{x})|^{-1} \tag{2}$$

where $\det J_f(\mathbf{x})$ is the determinant of the Jacobian of f and $|\det J_f(\mathbf{x})|^{-1} = |\det J_{f^{-1}}(\mathbf{x}')|$. The aim of NFs is to approximate the ideal map which transforms the prior into the target; thus, $q'(\mathbf{x}) = p(\mathbf{x})$. However, even without a perfect map, one can compute the importance weights $w_f(\mathbf{x})$ needed to reweight from q' to p the ensemble average of any observable $O(\mathbf{x})$, $\langle O(\mathbf{x})\rangle_p = \langle O(f(\mathbf{x}))w_f(\mathbf{x})\rangle_q / \langle w_f(\mathbf{x})\rangle_q$, as discussed, e.g., in ref 15. We choose to define the weights as

$$w_f(\mathbf{x}) = e^{u_q(\mathbf{x}) - u_p(f(\mathbf{x})) + \log|\det J_f(\mathbf{x})|} \tag{3}$$

so that $w_f(\mathbf{x}) \propto p(\mathbf{x})/q'(\mathbf{x})$, where the precise proportionality constant is irrelevant for the purpose of reweighting. In Boltzmann generators, one typically chooses as prior q a normal or uniform distribution, so that independent and identically distributed samples can be used to estimate the ensemble averages $\langle\cdot\rangle_q$. The weights are also useful to assess how effective the f mapping is in increasing the overlap between q and p, for example through the Kish effective sample size,[16]

$$n_{\text{eff}} = \frac{\left[\sum_i^n w_f(\mathbf{x}_i)\right]^2}{\sum_i^n [w_f(\mathbf{x}_i)]^2} \tag{4}$$

In NFs, the mapping f is implemented by an invertible deep neural network that by construction has an easy-to-calculate $\det J_f$. It can be trained to learn the target probability distribution $p(\mathbf{x})$ by minimizing the Kullback−Leibler divergence:

$$\begin{aligned} D_{\text{KL}}(q'\|p) &= -\int q'(\mathbf{x}') \log \frac{p(\mathbf{x}')}{q'(\mathbf{x}')} d\mathbf{x}' \\ &= -\int q(\mathbf{x}) \log w_f(\mathbf{x}) d\mathbf{x} - \Delta F_{pq} \end{aligned} \tag{5}$$

where the second line has been obtained via the change of variables $\mathbf{x}' = f(\mathbf{x})$ and $\Delta F_{\text{pq}} = -\log \frac{Z_q}{Z_p}$ is the free energy difference between q and p that is unknown but independent of f. The loss function then simplifies to $L_f = -\langle \log w_f(\mathbf{x})\rangle_q$, which in ref 15 is referred to as energy-based training, as opposed to maximum-likelihood training that instead requires sampling the target distribution $p(\mathbf{x})$ and minimizing $D_{\text{KL}}(p\|q')$. The loss function is an upper bound to the free energy difference $L_f \geq \Delta F_{pq}$, with the identity holding only in case of the ideal map.[17]

The main limitations of NFs are that they might not be expressive enough to represent the complex map from the prior to the target or that they can be hard to train in practice. In particular, energy-based training is characterized by mode-seeking behavior and struggles to converge reliably when the target distribution is multimodal.[18]

In this letter, we propose to use normalizing flows in a replica exchange setting to map the fast-mixing prior distribution to the target Boltzmann distribution, giving rise to the phase space overlap necessary for direct exchange, without the need to sample intermediate distributions. The idea of using a map to improve the overlap between distributions was first proposed by Jarzynski under the name of *targeted free energy perturbation*[19] and has recently been combined with NFs in the *learned free energy perturbation* (LFEP) method.[17,20−22] By analogy, we will refer to our method as *learned replica exchange* (LREX). Contrary to LFEP, in LREX we combine the NF mapping with local moves, which can greatly improve the sampling, as shown by several recent works.[18,23−26] Our approach can also be seen as a type of Boltzmann generator,[15] with two main differences: (i) the prior is a nontrivial distribution sampled via MD, and (ii) the target is directly sampled in a replica exchange setting, rather than only reconstructed via reweighting.

We now present the proposed LREX method in detail. To perform LREX, we first run a relatively short MD simulation to gather samples from the prior distribution $q(\mathbf{x})$ and use them to train the NF. Following ref 17, the NF is initialized to be the identity; thus, $\mathbf{x}' = \mathbf{x}$ and $q'(\mathbf{x}) = q(\mathbf{x})$. Ideally, the prior distribution should be easy to sample while still exhibiting the main features of the target, such as all the relevant metastable basins. This choice of prior allows us to avoid most of the problems related to energy-based training of normalizing flows[15,27] and significantly reduces computational cost. Training in the LREX setup typically requires only a few epochs and does not suffer from mode collapse, nor from numerical instabilities linked to extremely high energies that can arise from atom clashes and other nonphysical configurations in the prior. While the NF is training, it is possible to estimate its efficiency in increasing the phase space overlap between prior and target by computing n_{eff}, eq 4, over the training set and/or a validation set of n prior samples. The sampling efficiency n_{eff}/n also provides an estimate of the frequency with which exchanges will be accepted in the final LREX run. Once the NF is trained, the prior and the target systems are simulated in parallel with MD and, at fixed time intervals, an exchange between mapped configurations is attempted according to the following probability:

https://doi.org/10.1021/acs.jpclett.2c03327
J. Phys. Chem. Lett. 2022, 13, 11643−11649

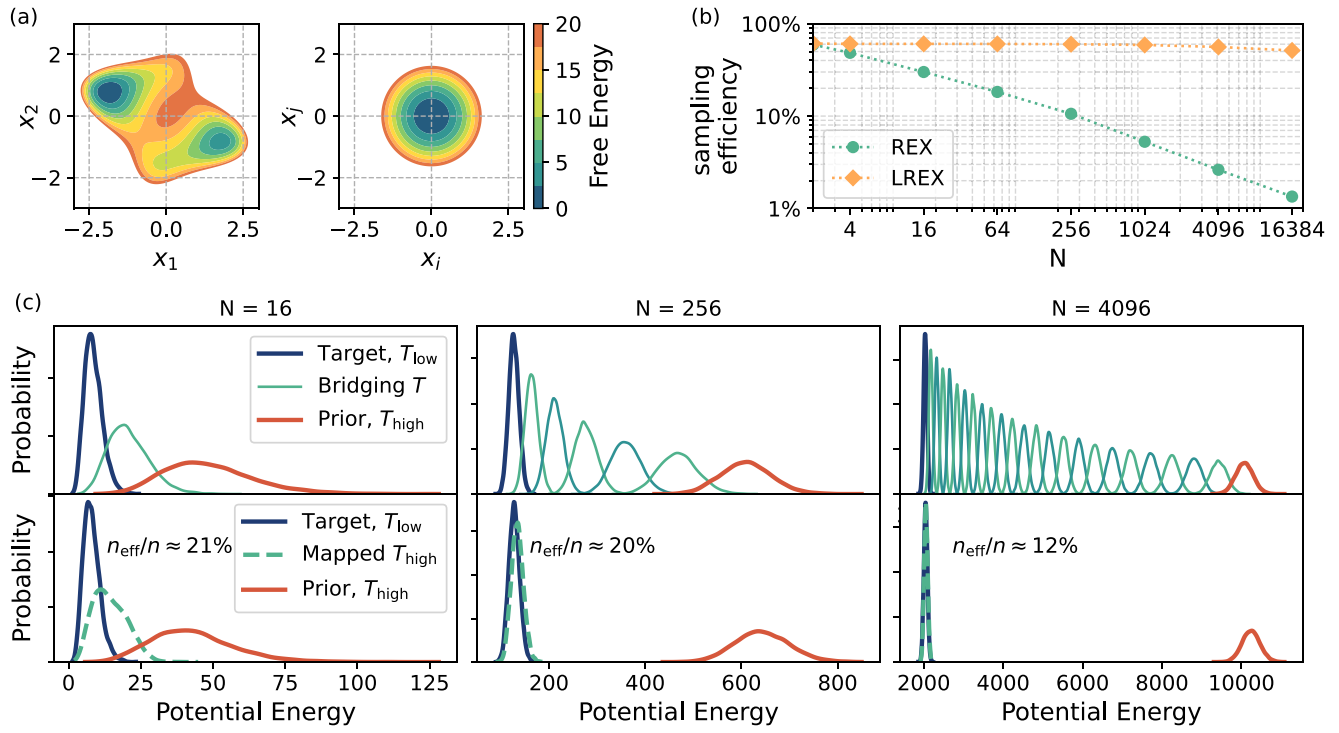


Figure 1. N-dimensional double-well system. (a) Apart from the first two dimensions, x_1 and x_2, all the other degrees of freedom, x_i, $i \in \{3, ..., N\}$, are uncoupled harmonic oscillators. (b) Sampling efficiency of REX and LREX, as a function of the system size N, in a log−log plot. See description in the text for details on how it is estimated. (c) Potential energy distribution for different dimensions N. In standard REX (top row) the number of replicas needed grows with the number of degrees of freedom, while in LREX (bottom row) the normalizing flow makes the distribution overlap, allowing for direct exchanges between prior and target at any N. The sampling efficiency, eq 4, of the NF is also reported.

$$\alpha_{\mathrm{LREX}} = \min\left\{1, \frac{p(\mathbf{x}'_q)}{q'(\mathbf{x}'_q)}\frac{q'(\mathbf{x}_p)}{p(\mathbf{x}_p)}\right\} = \min\{1, w_f(\mathbf{x}_q)w_{f^{-1}}(\mathbf{x}_p)\} \quad (6)$$

where $\mathbf{x}_p$ and $\mathbf{x}_q$ are the current configurations of the prior and target replica, respectively, and $w_{f^{-1}}$ are the weights of the inverse mapping, defined as in eq 3. If the exchange is accepted, the MD simulation of the prior continues from the configuration $\mathbf{x}'_q = f(\mathbf{x}_q)$, while the target continues from $f^{-1}(\mathbf{x}_p)$. New velocities are randomly assigned from the respective equilibrium distributions.

To demonstrate the advantages of the LREX approach over standard REX and Boltzmann generators, we present three examples. The first system considered is a particle moving with Langevin dynamics in a N-dimensional double-well potential, shown in Figure 1a. The first two dimensions, with coordinates x_1 and x_2, feel the 2D potential introduced in ref 28, while all the other $N - 2$ are subject to a harmonic potential (details in the SI). The target and prior distribution are obtained from the same system, but at different temperatures. The target distribution is at reduced temperature $T_{\mathrm{low}} = 1$, where transitions between the two basins are extremely rare, while the prior has $T_{\mathrm{high}} = 5$, which instead can be sampled efficiently also with a short MD run.

Figure 1c shows a visual comparison of standard REX (top row) and the proposed LREX method (bottom row). In replica exchange, the number of intermediate temperatures grows with the system size, independently of the fact that the added degrees of freedom are trivial Gaussian noise. Moreover, as the number of replicas increases, the mixing time required to equilibrate the simulation also increases, further limiting the efficiency of the method.[29] In LREX, instead, only the highest and lowest temperatures need to be sampled, because the NF easily learns a transformation that makes the high-temperature configuration space overlap with the low-temperature one. The NF architecture and training procedure are kept identical for all sizes and are described in detail in the SI. Although the size of the neural network increases with N, the slowdown in training time and efficiency is minor; thus, the overall scaling of LREX with system size is more favorable than the one of standard REX, for all the systems we studied.[30] Figure 1b presents the sampling efficiency of REX and LREX as a function of system size N. Other expanded ensemble methods would have similar scaling as REX (see Figure S2 in SI). The sampling efficiency is estimated by considering the importance weights of all the replicas and is an upper bound estimate strictly valid only in the infinite simulation limit (see SI). In a finite REX simulation, the efficiency can be further reduced by long time correlations due to a slow mixing time or small acceptance rate (see Figure S1 in SI). We also do not include the training cost for LREX, which would shift the curve down by a fixed small amount, depending on the total simulation length.

Next, we consider alanine dipeptide in vacuum. This molecule has two long-lived metastable basins that can be identified by its ϕ torsion angle. The target is to sample the system at $T_{\mathrm{low}} = 300$ K, and as prior we consider a very high temperature, $T_{\mathrm{high}} = 1000$ K, at which there are no sampling issues. With standard REX, one would need about four parallel replicas of the system to ensure proper overlap and good acceptance rate for the exchanges (see SI). Alanine dipeptide is

https://doi.org/10.1021/acs.jpclett.2c03327
J. Phys. Chem. Lett. 2022, 13, 11643−11649

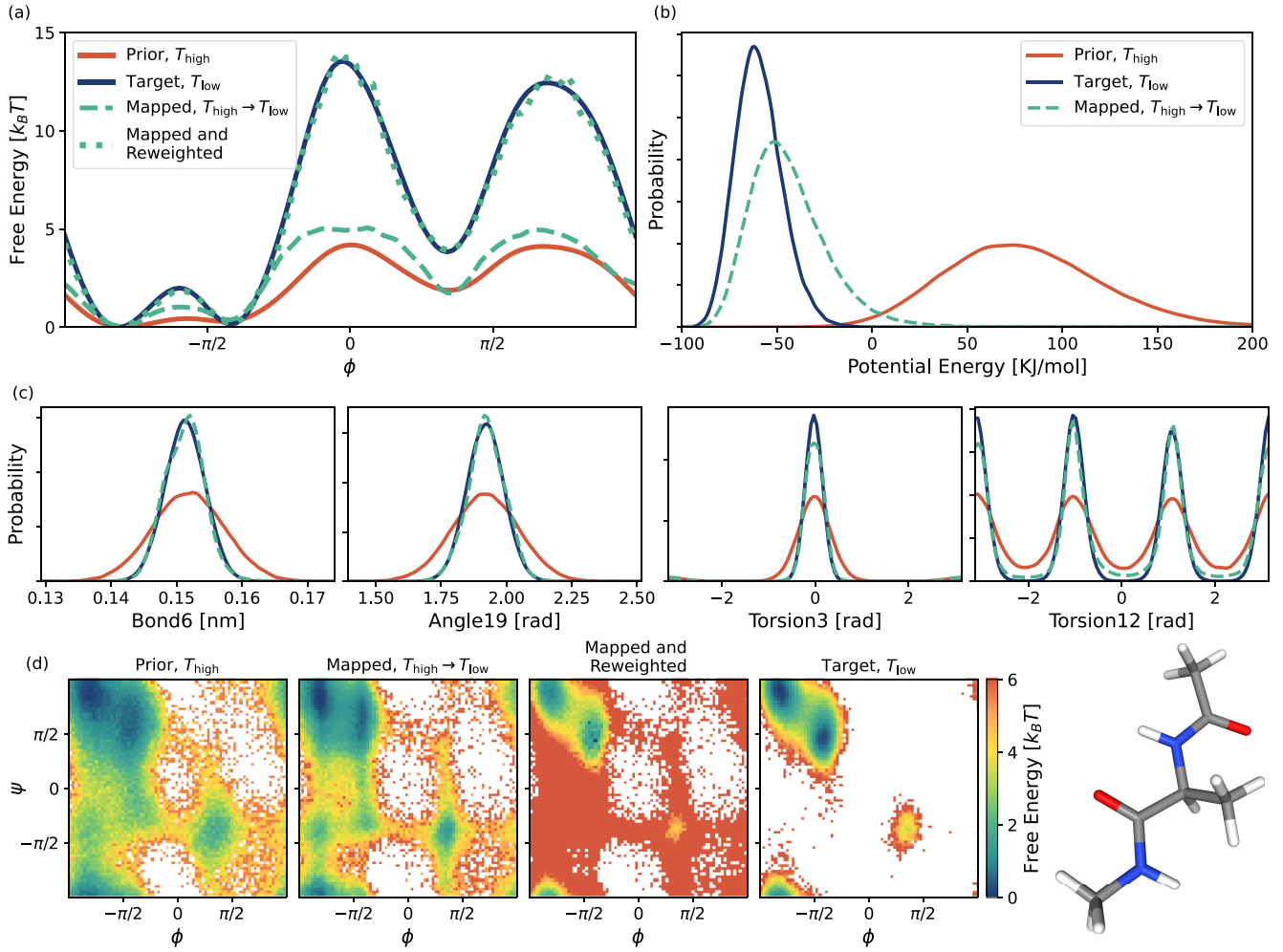


Figure 2. Alanine dipeptide in vacuum. (a) Free energy surface along the ϕ torsion angle, for the prior (T_{high} = 1000 K), the target distribution (T_{low} = 300 K), and the effect of the trained NF with and without reweighting. The NF does not move mass from one metastable state to the other, but the reweighting procedure corrects for this bias. (b) Energy distribution of the prior, target, and mapped distributions. A lack of overlap in the energy distribution implies a lack of overlap in configuration space, and without such overlap reweighting is not possible. (c) 1D marginal distributions over some of the internal coordinates of the molecule (48 in total). Notice that the NF has to learn not only the 1D marginals but also higher dimensional correlations in the probability distribution. (d) Histogram of the ϕ and ψ torsion angles sampled from the prior $q(\mathbf{x})$, the mapped $q'(\mathbf{x})$, the mapped and reweighted $w_f(\mathbf{x})q'(\mathbf{x})$, and the target distribution $p(\mathbf{x})$. The effect of the NF mapping is to bring configurations closer to the minima, as expected in a lower-temperature distribution.

composed of 22 atoms; thus, its configuration space has 66 dimensions, $\mathbf{x} \in \mathbb{R}^{66}$. Following refs 15 and 31, we remove global rotations and translations by modeling the NF in internal coordinates (bonds, angles, and torsion angles). This also makes it easy to constrain hydrogen atoms at fixed bond length and leaves us with a total of 48 degrees of freedom. To train the NF, we perform a short 20 ns MD simulation, from which we take 20,000 configurations at regular intervals of 500 MD steps, which are used as prior samples for the training. Thanks to our choice of prior distributions, we can use a simple flow architecture with only a few coupling layers.[32] Three epochs are enough to reach a good sampling efficiency $n_{eff}/n > 1\%$, and the entire training takes only a few minutes on a single consumer-grade CPU. Overall, the training is orders of magnitude faster than for previous Boltzmann generators of alanine dipeptide.[27,31,33] The reason for this speedup is 2-fold. First, our prior is closer to the target, so the flow needs to learn a simpler transformation. Second, the fact that LREX combines NFs with local sampling of the target distribution allows one to settle for a smaller sampling efficiency, without compromising on the accuracy at which observables can be computed.[25]

To get a sense of the effect of the trained NF map, we report in Figure 2c the 1D marginal distribution of some of the internal coordinates and in Figure 2d histograms over the 2D ϕ and ψ plane. Intuitively, the NF map pushes high-temperature configurations toward more low-temperature-like configurations, with smaller fluctuations from the local minima. The NF map is far from perfect, but, as indicated by Figure 2b, it makes the two distributions overlap, so that reweighting can be used to reconstruct the target. This is shown in Figure 2a, where the free energy surface (FES) along ϕ is reported. Interestingly, the NF does not map configurations from one metastable state to the other; it mostly makes local adjustments. This ensures that no mode collapse occurs during the NF energy-based training, which is one of the main current limitations of energy-based training for NFs.[12] In the case of alanine dipeptide, the NF we obtain is good enough that there is no need to perform the full LREX procedure; it is possible to

https://doi.org/10.1021/acs.jpclett.2c03327
J. Phys. Chem. Lett. 2022, 13, 11643−11649

accurately compute the torsion angle FES by direct reweighting, similarly as in Rizzi et al.[17] However, by running the prior and target in parallel and swapping configurations according to the LREX procedure, the sampling efficiency further increases, matching the one of standard REX for this system. In the SI we report the case of a target distribution at a much lower temperature of T_{low} = 100 K, where LREX is significantly more efficient than REX.

As a last example we consider the alanine tetrapeptide molecule. It has three pairs of $\phi-\psi$ torsion angles, for a total of 8 metastable basins that can be identified by the signs of the three ϕ angles. The system has 42 atoms, roughly double that of alanine dipeptide, for a total of 98 internal coordinates. This time we choose a less trivial prior and target, to show that LREX can be used not only across temperatures but also when the potential energy between prior and target differs. As the target we take the system in implicit solvent at T_{low} = 300 K, while for the prior we use the system in vacuum at T_{high} = 1000 K and also add a restraint potential $U_r(\boldsymbol{\psi})$ acting on the three ψ angles,

$$U_r(\boldsymbol{\psi}) = \sum_{i=1}^{3} k \sin^2(\psi_i/2) \tag{7}$$

with k = 100 kJ/mol. The chosen prior distribution not only has no phase space overlap with the target, as in the previous two examples, but it also has a smaller number of local minima in some metastable basins, as can be seen in Figure 3a. We use the same NF architecture as for alanine dipeptide (see SI), and the same amount of training data, obtained from just 20 ns of MD from the prior distribution. Also in this case the training requires only a few epochs and is computationally extremely cheap. Compared to the alanine dipeptide case, we obtain a lower sampling efficiency, $n_{eff}/n \approx 0.1\%$, which indicates a poorer phase space overlap. As shown in Figure 3c, the resulting NF fails to sample the correct target distribution when directly used via reweighting, as in Boltzmann generators and LFEP. Using it in a LREX setup is a better strategy and allows for an accurate reconstruction of the FES, Figure 3d. To perform LREX, we concurrently run 100 ns of MD for both the prior and the target replica, attempting an exchange every 1 ps, according to eq 6. The acceptance rate obtained is around 2%, and about 15% of the accepted exchanges are useful exchanges, meaning that they make the target MD replica jump from one metastable basin into a different one. Although it is possible to train a larger NF and use more MD data to better learn the target distribution, this would be computationally much more expensive to the extent of being less efficient than a simpler REX with many replicas. Thanks to the LREX setup, we can make good use of a less than perfect NF, significantly improving efficiency and accuracy over the standard reweighting strategy of Boltzmann generators.

In conclusion, this letter presents the learned replica exchange method that combines replica exchange and normalizing flows, resulting in improved performance over the two methods used separately. The LREX method has three phases: (i) run a short simulation to gather samples from the prior distribution, (ii) use these samples for energy-based training of the NF, (iii) run parallel simulations of the prior and target, attempting an exchange at regular intervals with acceptance given by eq 6. An important aspect of the method is that it is possible to estimate the efficiency of LREX before performing the final production run, simply by computing n_{eff} over the training data. If the system is particularly challenging and n_{eff} remains very low despite increasing the training or tuning the hyperparameters of the NF, it is possible to expand LREX by dividing the problem into smaller parts and training the NF to map to an intermediate system. This provides a straightforward way to improve the accuracy of the sampling by gradually increasing the number of replicas, in analogy with REX. It should be noted that in the worst case scenario, when the NF is completely wrong, LREX is equivalent to plain MD.

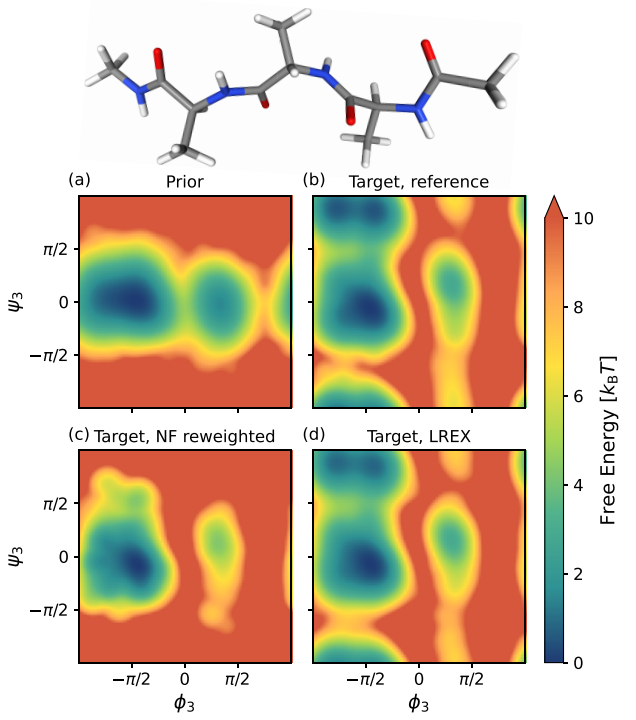


Figure 3. Free energy surfaces of alanine tetrapeptide along the ϕ_3-ψ_3 torsion angles. (a) The prior distribution is in vacuum at T_{high} = 1000 K and feels an extra restraint potential $U_r(\boldsymbol{\psi})$, eq 7. It can be efficiently sampled with plain MD. (b) The target distribution is in implicit solvent at T_{low} = 300 K. The reference FES is obtained with adaptive-bias enhanced sampling on the three ϕ angles.[34] (c) FES obtained from the trained NF, by mapping and reweighting the prior samples. The NF struggles to correctly reconstruct some of the FES minima. However, (c) using the same NF to exchange configurations between the prior and the target in an LREX scheme allows for an accurate estimate of the target FES. See Figure S6 in the SI for the 1D projection of these FES.

Compared to standard REX, using LREX allows for a drastic reduction of the number of replicas that have to be simulated in parallel, which can lead to a significant reduction in computation cost, especially when machine-learning potentials are employed. This could allow the use of prior distributions that would otherwise be impractical for REX, such as coarse-grained versions of the target system.[35] Unfortunately, a simple scaling law with system size is not available for NF, so new experiments are needed to assess feasibility on larger systems.[30] Compared to other setups that use NFs as Boltzmann generators, LREX can be significantly more efficient and handle larger systems, due to the similarity between prior and target. It can also be more accurate due to the combination with local MD sampling. The main limitations of LREX are due to the current limitations of normalizing flows. For example, NFs have not yet been successfully applied to systems

https://doi.org/10.1021/acs.jpclett.2c03327
J. Phys. Chem. Lett. 2022, 13, 11643−11649

with explicit solvent, except in cases where only a subset of the system has been considered.[36,37] However, NFs have been significantly improved in recent years, and any future improvements to the architecture or training of normalizing flows, making them more efficient or expressive, can also be readily exploited in LREX. For this reason, we expect LREX to become a useful tool for molecular simulations, especially as normalizing flows' capabilities are developing rapidly, allowing larger and more diverse systems to be handled.

■ ASSOCIATED CONTENT

Data Availability Statement

All the codes and data needed to reproduce the simulations are available at https://github.com/invemichele/learned-replica-exchange. Molecular dynamics simulations were performed with the OpenMM software.[38] For the reference simulations the on-the-fly probability enhanced sampling method[11,34,39] has been used, as implemented in PLUMED.[40,41] The normalizing flows have been implemented using the `bgflow` library (https://github.com/noegroup/bgflow).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.2c03327.

Double-well: computational details, figure of low temperature trajectory, figure and further comments on sampling efficiency. Alanine dipeptide: computational details, figure of training loss, figure for lower temperature. Alanine tetrapeptide: computational details, figure of FES for all other angles in 2D and 1D (PDF)

Transparent Peer Review report available (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Michele Invernizzi − *Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany;* orcid.org/0000-0002-3328-6557; Email: michele.invernizzi@fu-berlin.de

Frank Noé − *Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany; Department of Physics, Freie Universität Berlin, 14195 Berlin, Germany; Department of Chemistry, Rice University, 77005 Houston, United States; Microsoft Research AI4Science, 10178 Berlin, Germany*; Email: franknoe@microsoft.com

Authors

Andreas Krämer − *Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany;* orcid.org/0000-0002-7699-3083

Cecilia Clementi − *Department of Physics, Freie Universität Berlin, 14195 Berlin, Germany; Department of Chemistry and Center for Theoretical Biological Physics, Rice University, 77005 Houston, United States;* orcid.org/0000-0001-9221-2358

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.2c03327

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Jonas Köhler for useful discussions. M.I. acknowledges support from the Swiss National Science Foundation through an Early Postdoc.Mobility fellowship and the Humboldt Foundation for a Postdoctoral Research Fellowship. F.N. and C.C. acknowledge funding from Deutsche Forschungsgemeinschaft (DFG) via CRC 1114 Project B08. F.N. and A.K. acknowledge funding from European Research Council (ERC) via CoG 772230 "ScaleCell".

■ REFERENCES

(1) Swendsen, R. H.; Wang, J.-S. Replica Monte Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607−2609.

(2) Earl, D. J.; Deem, M. W. Parallel Tempering: Theory, Applications, and New Perspectives. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910.

(3) Tabak, E. G.; Vanden-Eijnden, E. Density Estimation by Dual Ascent of the Log-Likelihood. *Commun. Math. Sci.* **2010**, *8*, 217−233.

(4) Tabak, E. G.; Turner, C. V. A Family of Nonparametric Density Estimation Algorithms. *Comm. Pure Appl. Math.* **2013**, *66*, 145−164.

(5) Rezende, D. J.; Mohamed, S. Variational Inference with Normalizing Flows. *Proc. 32th Int. Conf. Mach. Learn.* **2015**, *37*, 1530−1538.

(6) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431−9438.

(7) Mey, A. S.; Allen, B. K.; Bruce Macdonald, H. E.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; et al. Best Practices for Alchemical Free Energy Calculations [Article v1.0]. Living. *LiveCoMS* **2020**, *2*, 18378.

(8) Hukushima, K.; Nemoto, K. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604−1608.

(9) Marinari, E.; Parisi, G. Simulated Tempering: A New Monte Carlo Scheme. *Europhys. Lett. EPL* **1992**, *19*, 451−458.

(10) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. New Approach to Monte Carlo Calculation of the Free Energy: Method of Expanded Ensembles. *J. Chem. Phys.* **1992**, *96*, 1776−1783.

(11) Invernizzi, M.; Piaggi, P. M.; Parrinello, M. Unified Approach to Enhanced Sampling. *Phys. Rev. X* **2020**, *10*, 41034.

(12) Ballard, A. J.; Jarzynski, C. Replica Exchange with Nonequilibrium Switches. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12224−12229.

(13) Nilmeier, J. P.; Crooks, G. E.; Minh, D. D. L.; Chodera, J. D. Nonequilibrium Candidate Monte Carlo Is an Efficient Tool for Equilibrium Simulation. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, E1009−E1018.

(14) Deighan, M.; Bonomi, M.; Pfaendtner, J. Efficient Simulation of Explicitly Solvated Proteins in the Well-Tempered Ensemble. *J. Chem. Theory Comput.* **2012**, *8*, 2189−2192.

(15) Noé, F.; Olsson, S.; Köhler, J.; Wu, H. Boltzmann Generators: Sampling Equilibrium States of Many-Body Systems with Deep Learning. *Science* **2019**, *365*, eaaw1147.

(16) Kish, L. Sampling Organizations and Groups of Unequal Sizes. *Am. Sociol. Rev.* **1965**, *30*, 564.

(17) Rizzi, A.; Carloni, P.; Parrinello, M. Targeted Free Energy Perturbation Revisited: Accurate Free Energies from Mapped Reference Potentials. *J. Phys. Chem. Lett.* **2021**, *12*, 9449−9454.

(18) Matthews, A. G. D. G.; Arbel, M.; Rezende, D. J.; Doucet, A. Continual Repeated Annealed Flow Transport Monte Carlo. *Proc. 39th Int. Conf. Mach. Learn.* **2022**, *162*, 15196−15219.

(19) Jarzynski, C. Targeted Free Energy Perturbation. *Phys. Rev. E* **2002**, *65*, 5.

(20) 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.
(21) Coretti, A.; Falkner, S.; Geissler, P.; Dellago, C.Learning Mappings between Equilibrium States of Liquid Systems Using Normalizing Flows. *arXiv***2022**, 2208.10420.
(22) Falkner, S.; Coretti, A.; Romano, S.; Geissler, P.; Dellago, C.Conditioning Normalizing Flows for Rare Event Sampling. *arXiv***2022**, 2207.14530.
(23) Wu, H.; Köhler, J.; Noé, F. Stochastic Normalizing Flows. *NIPS'20: Proc. 34th Int. Conf. Neural Inf. Process. Syst.* **2020**, *33*, 5933−5944.
(24) Sbailò, L.; Dibak, M.; Noé, F. Neural Mode Jump Monte Carlo. *J. Chem. Phys.* **2021**, *154*, 074101.
(25) Gabrié, M.; Rotskoff, G. M.; Vanden-Eijnden, E. Adaptive Monte Carlo Augmented with Normalizing Flows. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, e2109420119.
(26) Arbel, M.; Matthews, A. G. D. G.; Doucet, A. Annealed Flow Transport Monte Carlo. *Proc. 38th Int. Conf. Mach. Learn.* **2021**, *139*, 318−330.
(27) Midgley, L. I.; Stimper, V.; Simm, G. N. C.; Schölkopf, B.; Hernández-Lobato, J. M.Flow Annealed Importance Sampling Bootstrap. *arXiv***2022**, 2208.01893.
(28) Invernizzi, M.; Parrinello, M. Making the Best of a Bad Situation: A Multiscale Approach to Free Energy Calculation. *J. Chem. Theory Comput.* **2019**, *15*, 2187−2194.
(29) Abraham, M. J.; Gready, J. E. Ensuring Mixing Efficiency of Replica-Exchange Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2008**, *4*, 1119−1128.
(30) Abbott, R.; Albergo, M. S.; Botev, A.; Boyda, D.; Cranmer, K.; Hackett, D. C.; Matthews, A. G. D. G.; Racanière, S.; Razavi, A.; Rezende, D. J.; et al.Aspects of Scaling and Scalability for Flow-Based Sampling of Lattice QCD. *arXiv***2022**, 2211.07541.
(31) Köhler, J.; Krämer, A.; Noé, F.Smooth Normalizing Flows. *Adv. Neural Inf. Process. Syst.***2021**, *34*, 2796−2809.
(32) Durkan, C.; Bekasov, A.; Murray, I.; Papamakarios, G. Neural Spline Flows. *NIPS'19: Proc. 33rd Int. Conf. Neural Inf. Process. Syst.* **2019**, 7511−7522.
(33) Dibak, M.; Klein, L.; Krämer, A.; Noé, F. Temperature Steerable Flows and Boltzmann Generators. *Phys. Rev. Res.* **2022**, *4*, L042005.
(34) Invernizzi, M.; Parrinello, M. Rethinking Metadynamics: From Bias Potentials to Probability Distributions. *J. Phys. Chem. Lett.* **2020**, *11*, 2731−2736.
(35) Lyman, E.; Zuckerman, D. M. Resolution Exchange Simulation with Incremental Coarsening. *J. Chem. Theory Comput.* **2006**, *2*, 656−666.
(36) Köhler, J.; Chen, Y.; Krämer, A.; Clementi, C.; Noé, F.Flow-Matching − Efficient Coarse-Graining of Molecular Dynamics without Forces. *arXiv***2022**, 2203.11167.
(37) Wang, Y.; Herron, L.; Tiwary, P. From Data to Noise to Data for Mixing Physics across Temperatures with Generative Artificial Intelligence. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, e2203656119.
(38) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; et al. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLOS Comput. Biol.* **2017**, *13*, e1005659.
(39) Invernizzi, M.; Parrinello, M. Exploration vs Convergence Speed in Adaptive-Bias Enhanced Sampling. *J. Chem. Theory Comput.* **2022**, *18*, 3988−3996.
(40) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(41) The PLUMED consortium.. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nat. Methods* **2019**, *16*, 670−673.

https://doi.org/10.1021/acs.jpclett.2c03327
J. Phys. Chem. Lett. 2022, 13, 11643−11649