

# Reinforcement Learning for Multi-Scale Molecular Modeling

Jun Zhang<sup>1</sup>, Yao-Kun Lei<sup>2</sup>, Yi Isaac Yang<sup>3,†</sup> and Yi Qin Gao<sup>2,3,‡</sup>

<sup>1</sup>*Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany*

<sup>2</sup>*Institute of Theoretical and Computational Chemistry, College of Chemistry and Molecular Engineering, Peking University, 100871 Beijing, China.*

<sup>3</sup>*Institute of Systems Biology, Shenzhen Bay Laboratory, 518055 Shenzhen, China*

Corresponding authors: (Y.I.Y) <sup>†</sup> yangyi@szbl.ac.cn and (Y.Q.G) <sup>‡</sup> gaoyq@pku.edu.cn

## Abstract

Molecular simulations are widely applied in the study of chemical and bio-physical systems of interest. However, the accessible timescales of atomistic simulations are limited, and extracting equilibrium properties of systems containing rare events remains challenging. Two distinct strategies are usually adopted in this regard: either sticking to the atomistic level and performing enhanced sampling, or trading details for speed by leveraging coarse-grained models. Although both strategies are promising, either of them, if adopted individually, exhibits severe limitations. In this paper we propose a machine-learning approach to take advantage of both strategies. In this approach, simulations on different scales are executed simultaneously and benefit mutually from their cross-talks: Accurate coarse-grained (CG) models can be inferred from the fine-grained (FG) simulations; In turn, FG simulations can be boosted by the guidance of CG models. Our method grounds on unsupervised and reinforcement learning, defined by a variational and adaptive training objective, and allows end-to-end training of parametric models. Through multiple experiments, we show that our method is efficient and flexible, and performs well on challenging chemical and bio-molecular systems.

**Keywords:** Coarse graining, Machine learning, Deep learning, Enhanced sampling, Artificial neural networks

## Significance Statement

Enhanced sampling and coarse graining are two widely adopted approaches that expand the scope of molecular simulations into experiment-relevant scales. We propose here a machine-learning approach to combine coarse graining with importance sampling, so that simulations on different scales can benefit mutually: Accurate coarse-grained models can be inferred from fine-grained simulations, whereas fine-grained simulations can be boosted by the guidance of coarse-grained models. This new multi-scale modeling framework incorporates state-of-the-art deep unsupervised and reinforcement learning techniques, allowing novel deep neural-network-based molecular models to be adaptively and variationally optimized.

## I. Introduction

Molecular simulations, particularly all-atom and *ab initio* molecular dynamics (MD), have furthered our understanding of many chemical and bio-physical processes (1, 2). In molecular simulations, interactions between particles (e.g., atoms, residues or molecules) are described by an (potential) energy function  $U(\mathbf{R})$  of the configuration  $\mathbf{R}$ . To investigate such systems, one is often not interested in the exact energy, but the free-energy, or the equilibrium distribution, of some reduced descriptors, e.g., collective variables (CV) (3) or CG variables (4),  $\mathbf{s}(\mathbf{R})$ , as a function of  $\mathbf{R}$ :

$$p(\mathbf{s}) = \frac{\int e^{-\beta U(\mathbf{R})} \delta(\mathbf{s} - \mathbf{s}(\mathbf{R})) d\mathbf{R}}{\int e^{-\beta U(\mathbf{R})} d\mathbf{R}} = \frac{e^{-\beta F(\mathbf{s})}}{Z} \quad (1)$$

$$F(\mathbf{s}) = -\frac{1}{\beta} [\log p(\mathbf{s}) + \log Z] \quad (2)$$

where  $Z = \int e^{-\beta U(\mathbf{R})} d\mathbf{R}$  is the partition function,  $\beta$  is the inverse temperature and  $\delta$  denotes the Dirac delta function. Equation (2) holds up to an arbitrary additive constant. Usually  $\mathbf{s}$  is selected to be slowly changing variables governing the process of interest, and the rest of degrees of freedom (DOF) can be treated in the mean-field fashion (5, 6). Under this setting,  $F(\mathbf{s})$  becomes a CG description of the original thermodynamic system, and simulations performed under  $F(\mathbf{s})$  are generally much faster than those run on the FG potential (i.e.,  $U(\mathbf{R})$ ) because  $\text{Dim}(\mathbf{s}) \ll \text{Dim}(\mathbf{R})$  (where  $\text{Dim}(\cdot)$  denotes the dimensionality) despite the loss of finer details. Therefore, Eqs. (1-2) are also known as the principle of thermodynamic consistency for coarse graining (5). However, practical implementation of Eqs. (1-2) is hindered by two major issues: (1) How can one approximate a reliable analytical form for the CG potential  $F(\mathbf{s})$  given access to samples drawn from  $U(\mathbf{R})$ ? (2) How can one draw equilibrium samples from  $U(\mathbf{R})$ , which are further used to infer  $F(\mathbf{s})$ ? The former is known as the coarse graining problem, while the latter as the importance sampling problem, and both are of particular importance in physics, chemistry and biology.

Conventionally, if  $\mathbf{s}$  is low-dimensional (say,  $\text{Dim}(\mathbf{s}) \leq 3$ ), non-parametric methods like kernel density estimation (KDE) (7) can be adopted to infer  $F(\mathbf{s})$ , but they become quickly infeasible as  $\text{Dim}(\mathbf{s})$  increases. Artificial neural networks (ANNs) and deep learning may offer extra flexibility and expressivity to this end (8, 9). For instance, several recent studies proposed supervised learning approaches to fit  $F(\mathbf{s})$  by ANNs (10-12). However, fitting  $F(\mathbf{s})$  as a regression problem has several drawbacks: (1) it necessitates gridding the space of  $\mathbf{s}$ , which would be computationally prohibitive for large  $\text{Dim}(\mathbf{s})$ ; (2) it is rather data-inefficient because calculating  $F(\mathbf{s})$  at one point needs a large amount of samples from  $U(\mathbf{R})$  at the neighborhood of  $\mathbf{s}$  (i.e.,  $\delta(\mathbf{s} - \mathbf{s}(\mathbf{R}))$ ); (3) Since regression learning lays equal importance over any point in the probability measure space regardless of the mass density distribution, it is not suitable for fitting an imbalanced distribution where the accuracy of areas with higher mass density should be more emphasized.

An alternative view of inferring  $p(\mathbf{s})$  is provided by statistical modeling and machine learning community, where density estimation of high-dimensional data has been a long-standing goal (9, 13). Particularly, a recent burst of work sparks new ideas to exploit deep learning to this end, giving rise to an active research field of generative learning and unsupervised models including variational auto-encoders (14), auto-regressive (15) and normalizing-flow models (16). Nevertheless, these methods are all footed on certain simple prior distributions, hence the complexity of the distributions resulted from these methods is bound by the manifold structure of the prior distribution. Therefore, they are known to suffer from issues like mode-dropping (i.e., violating the ergodicity assumption in statistical physics) and often assigning probability mass to areas unwarranted by

the data (17). On the other hand, generative learning with energy-based models (EBMs) can be dated back to even longer before (18-20), which in principle can fit arbitrarily complex distributions due to the flexibility and plasticity of the energy landscapes (21, 22). Based on EBMs we propose a variational approach to the CG problem, i.e., to infer an analytic form for the complex free-energy surface (FES) without supervision. Furthermore, our new approach allows simulations on different scales, which are launched simultaneously, cross talk and benefit from each other. We will demonstrate that the inferred CG potential can in turn help enhance the sampling of the FG model in a reinforced and adaptive manner.

## II. Methods

### 1. Coarse graining with thermodynamic consistency

We propose a variational approach, called variational inference of free-energy (VIFE), to approximate the (possibly high-dimensional) FES,  $F(\mathbf{s})$ , by parametric models. Specifically, we denote the approximate free-energy function and the associated probability distribution as  $F_\theta(\mathbf{s})$  and  $p_\theta(\mathbf{s})$  respectively (where  $\theta$  are optimizable parameters), and define a strict divergence  $D(p||p_\theta)$  between  $p$  and  $p_\theta$ . A strict divergence, including but not limited to Kullback-Leibler divergence  $D_{\text{KL}}(p||p_\theta)$  (23) and Wasserstein divergence  $D_W(p||p_\theta)$  (24), satisfies the condition that  $D(p||p_\theta) \geq 0$  where the equality holds if and only if  $p = p_\theta$ , hence can be used as a variational objective (25-27). Particularly, the gradient of  $D_{\text{KL}}(p||p_\theta)$  w.r.t.  $\theta$  takes the following form in Eq. (3) (see SI or Ref. (26) for the derivation),

$$\nabla_\theta D_{\text{KL}}(p||p_\theta) = \langle \nabla_\theta F_\theta(\mathbf{s}) \rangle_{p(\mathbf{s})} - \langle \nabla_\theta F_\theta(\mathbf{s}) \rangle_{p_\theta(\mathbf{s})} \quad (3)$$

where  $\langle f(x) \rangle_{p(x)}$  denotes the expectation value of  $f$ , a function of  $x$ , over a distribution  $p(x)$ . The gradient for  $D_W(p||p_\theta)$  has been separately derived in Targeted Adversarial Learning Optimized Sampling (TALOS) (28) and it shares dramatic similarity with Eq. (3). In this paper we perform experiments exclusively according to Eq. (3) and leave  $D_W$  for future research.

The rest of the problem is to evaluate the two expectation values in Eq. (3). If we have access to, say, equilibrium MD samples drawn from  $U(\mathbf{R})$ , the distribution of which is denoted as  $p_{\text{FG}}(\mathbf{s})$ , we can use  $p_{\text{FG}}$  to approximate  $p$  and optimize  $F_\theta$  w.r.t. a surrogate objective  $D_{\text{KL}}(p_{\text{FG}}||p_\theta)$ , minimizing which is equivalent to maximum likelihood estimation. The remaining task is to calculate  $\langle \nabla_\theta F_\theta(\mathbf{s}) \rangle_{p_\theta(\mathbf{s})}$ . Since the analytical forms of  $F_\theta(\mathbf{s})$  and  $\nabla_\theta F_\theta(\mathbf{s})$  are both available, we can perform CG simulations on  $F_\theta(\mathbf{s})$  to estimate  $\langle \nabla_\theta F_\theta(\mathbf{s}) \rangle_{p_\theta(\mathbf{s})}$ . Due to the simplicity of  $F_\theta(\mathbf{s})$  (i.e., the number of DOFs in  $\mathbf{s}$  is usually limited), the CG simulations are usually computationally economic and converge relatively fast, hence brute-force simulations via Monte Carlo (MC) sampling or Langevin dynamics generally suffice. We note here that VIFE can be viewed as a more flexible and general extension of the inversed Monte Carlo method (29). VIFE also inherits the merit of the compositionality from EBMs: In VIFE, it is straightforward to introduce some restraints or physics-based prior knowledge in the form of additive energies.

### 2. Reinforced VIFE (RE-VIFE)

In a more common setting where  $p_{\text{FG}}$  is not available beforehand, we have to perform sampling over  $U(\mathbf{R})$  from scratch. Since  $\text{Dim}(\mathbf{R})$  is usually very large, i.e.,  $\text{Dim}(\mathbf{R}) \gg \text{Dim}(\mathbf{s})$ , estimating ensemble averages over  $p(\mathbf{s})$  is often infeasible for brute-force FG (e.g., all-atom or *ab initio*) simulations. A plethora of enhanced sampling methods have been developed trying to solve this issue, and there exist several excellent reviews on this topic (3, 30-32). Here we will show that VIFE, combined

with reinforced imitation learning, provides a new solution to this problem. We name our new approach RE-VIFE, which exhibits several compelling merits: (1) RE-VIFE is able to handle high-dimensional CG variables while most of enhanced sampling methods become ineffective as  $\text{Dim}(\mathbf{s})$  grows; (2) RE-VIFE casts a well-defined optimization problem which allows cross-fertilization with deep learning; (3) RE-VIFE formulates an adaptive training objective which can be optimized variationally, hence ensuring the efficiency and convergence of the method.

Mathematically, if we have an optimal  $F_\theta(\mathbf{s})$  which is equal to the ground-true  $F(\mathbf{s})$ , and perform sampling under  $U - F_\theta$ , we will arrive at a uniform distribution over  $\mathbf{s}$ . Therefore, one is motivated to employ  $F_\theta(\mathbf{s})$  as a bias potential in order to achieve a flattened distribution over  $\mathbf{s}$  which may originally involve high free-energy barriers (33). However, the situation is complicated by the errors in  $F_\theta$ : Since  $F_\theta$  is optimized w.r.t. available FG samples (usually corresponding to metastable states in MD simulations), its value can be very inaccurate in those under-sampled regions (e.g., transition regions). Therefore, directly inserting  $F_\theta$  as the bias potential could be non- or even counter-productive. We note here that similar problems where one has to deal with moving distributions and partial sampling are commonly encountered and addressed in Reinforcement Learning (RL) (34). Inspired by RL, we introduce a two-timescale learning scheme, where a bias potential  $V_\phi(\mathbf{s})$  with  $\phi$  denoting optimizable parameters (equivalent to a policy function in RL) is separately trained in addition to  $F_\theta(\mathbf{s})$  which can now be viewed as a value function in the spirit of actor-critic RL (35). As in variationally enhanced sampling (VES) (27) or TALOS (28), we can define a target distribution  $p_T(\mathbf{s})$  where the free-energy barrier is lowered (in other words, less-visited regions are more encouraged) according to  $F_\theta(\mathbf{s})$  (see SI for more information about  $p_T$ ), then optimize  $V_\phi(\mathbf{s})$  by minimizing a strict divergence, for instance,  $D_{\text{KL}}(p_T||p_\phi)$ ,

$$\nabla_\phi D_{\text{KL}}(p_T||p_\phi) = \langle \nabla_\phi V_\phi(\mathbf{s}) \rangle_{p_T} - \langle \nabla_\phi V_\phi(\mathbf{s}) \rangle_{p_\phi} \quad (4)$$

where  $p_\phi$  denotes the distribution induced by  $U + V_\phi$ . The separate parametrization allows us to use an imbalanced learning schedule for the free-energy function  $F_\theta$  and the bias potential  $V_\phi$ . Particularly, we can train  $F_\theta(\mathbf{s})$  based on the latest  $p_{\text{FG}}(\mathbf{s})$  (reweighted from  $p_\phi$ ) with a higher rate, and update  $V_\phi(\mathbf{s})$  in a more conservative manner. In terms of imitation learning,  $F_\theta(\mathbf{s})$  plays the role of a leader that coins a moving target based on the current density estimation, while  $V_\phi(\mathbf{s})$  learns to tune the policy in order to catch up (Fig. 1). A dramatic advantage of such leader-chaser scheme lies in the fact that  $F_\theta$  along with  $p_T$  is constructed based on the samples drawn from simulations under  $V_\phi$ , so  $p_T$  and  $p_\phi$  always share substantial overlap; otherwise  $D_{\text{KL}}$  would fall victim to the notorious vanishing gradient issue (24). Remarkably, such sort of separate parametrization and two-timescale updated rule were also adopted in generative adversarial networks (GANs) (36, 37), double Q-learning (38) as well as other reinforcement and imitation learning settings (39, 40).

In RE-VIFE, albeit  $F_\theta(\mathbf{s})$  may be imperfect globally (especially at the beginning of training), it still gives relatively accurate density estimation of the recently visited regions, according to which  $V_\phi(\mathbf{s})$  can be improved locally via Eq. (4) and gradually enhance the sampling efficiency. On the other hand, as the FG sampling is enhanced by  $V_\phi(\mathbf{s})$ , we can get a better approximation of  $p(\mathbf{s})$ , thus gradually push  $F_\theta(\mathbf{s})$  to the optimum according to Eq. (3). Specifically, it can be shown that if we choose

$$p_T = p_\theta^{\frac{1}{\gamma}} \propto \exp\left(-\frac{\beta}{\gamma} F_\theta\right) \quad (5)$$

where  $\gamma > 1$  is the well-tempering factor (41), then

$$\operatorname{argmin}_{V_\phi} D_{\text{KL}}(p_T || p_\phi) = \left(\frac{1}{\gamma} - 1\right) F(\mathbf{s}) \quad (6)$$

Equation (6) concludes that the optimal  $V_\phi$  is equivalent to the well-tempered free energy. From this perspective, metadynamics (33) can be viewed as a special case of RE-VIFE if we adopt KDE in replacement of  $F_\theta(\mathbf{s})$  and use accumulative Gaussians as  $V_\phi(\mathbf{s})$ . In this sense, RE-VIFE is a generalization of metadynamics to large  $\text{Dim}(\mathbf{s})$  and parametric bias potential energy functions. The assembled training algorithm of RE-VIFE is summarized in Algorithm S1 in Supplemental Information.

### 3. Parametrization in (RE-)VIFE

In (RE-)VIFE, we train parametric models to approximate  $F_\theta$  and  $V_\phi$ , with  $\theta$  and  $\phi$  denoting optimizable parameters. Since (RE-)VIFE is a variational approach, many functional forms are optional as long as they have sufficient variational flexibility. As in VES and TALOS, we can use certain orthonormal basis functions to expand  $F_\theta$  and/or  $V_\phi$  in low-dimensional spaces (27, 28). With some caveats, we can also exploit ANNs to approximate  $F_\theta$  and  $V_\phi$ . Compared to other functional approximation methods, ANNs are more flexible and expressive: On the one hand, ANNs allow us to cheaply approximate high-dimensional functions which would be intractable otherwise; On the other hand, the gradients or forces produced by ANNs are not necessarily smooth enough, which may cause instabilities for numerical integration in simulations. To attack the issue of irregular gradients while maximally harnessing the expressivity, the architecture of ANNs should be carefully designed, and special regularization techniques may be needed to smooth the gradients (see SI for more details).

## III. Results

### 1. Illustration on a toy model

We benchmarked VIFE on a 2-dimensional 3-well toy model (42). The contour map of the potential energy surface (PES),  $U(x, y)$ , is shown in Fig. 2A. Langevin dynamics simulation was first performed on  $U(x, y)$  and yielded 50,000 samples, which were used as  $p_{\text{FG}}$ . We then choose the coarse-graining variable  $\mathbf{s} = (x, y)$  (we desire to reproduce  $U(x, y)$  instead of coarse-graining), and train a parametric model to infer  $F_\theta(x, y)$  based on the available samples. An ANN with specially designed architecture was adopted as  $F_\theta$  and the model was optimized via Eq. (3) (see SI for more details about model setup and training details). We trained VIFE for 100 epochs, and plotted the estimated  $D_{\text{KL}}$  between  $p_\theta$  and  $p_{\text{FG}}$  along the training progress (Fig. 2B). It can be seen that  $D_{\text{KL}}$  between the two distributions quickly diminished, and the optimization of  $F_\theta$  converged within 50 epochs, demonstrating the efficiency and efficacy of VIFE. Note that Eq. (3) can be viewed as the gradient of the following surrogate loss function,

$$L(\theta) = \langle F_\theta(\mathbf{s}) \rangle_{p(\mathbf{s})} - \langle F_\theta(\mathbf{s}) \rangle_{p_\theta(\mathbf{s})} \quad (7)$$

hence we also showed the evolution of  $L(\theta)$  during training. Intriguingly, we observed in Fig. 2B that there is an ups-and-downs pattern in the curve of  $L(\theta)$ , with the oscillating amplitude vanishes throughout training. This is an expected phenomenon because Eq. (7) is reminiscent of the training objective of Wasserstein-GAN (24), which solves a saddle-point (or mini-max) problem thus entails an oscillating loss curve. Besides  $D_{\text{KL}}$ , for EBMs, another important indicator of the quality of

models is whether or not  $F_\theta$  of FG and CG samples are identically (or similarly) distributed. Therefore, we examined the two distributions accordingly (Fig. S1), and found that they almost overlap perfectly, proving the resulting  $F_\theta$  is of high quality.

We then drew the contour plot of  $F_\theta$  in Fig. 2C. Compared to the original PES (Fig. 2A),  $F_\theta$  preserves all the three local minima correctly, showing no signs of mode-dropping. This is a particular advantage of EBMs over other generative methods, since ergodicity is vital to investigation of thermodynamic systems. Besides, the overall contour and landscape of  $F_\theta$  also resemble  $U(x, y)$ , especially on regions with higher densities. In contrast, deviations seem to be more significant in the transition regions. This arises from the fact that we only collected 50,000 FG samples (most of which locate around the local minima), and the transition regions are not adequately covered by available samples (as shown in Fig. 2D). If more FG samples over the transition regions can be obtained,  $F_\theta$  would be further improved, and such consideration is a strong motivation behind RE-VIFE.

Now that we obtained the analytical form of  $F_\theta$  based merely on samples from  $U(x, y)$  (rather than knowing the mathematical form of  $U(x, y)$ ), we can locate free-energy minima and cluster samples. We exploited Nelder-Mead algorithm (43) to minimize the simulation samples over  $F_\theta$ , and observed that all the samples finally fall into three distinct local minima, shown by different colored symbols in Fig. 2C. We also colored the simulation samples according to their final minimizers, as shown in Fig. 2D, and found that the noisily distributed samples were indeed assigned to different metastable states quite reasonably. This result implies important potential application of VIFE in identifying free-energy minima (or metastable states) and clustering noisy high-dimensional samples, which is demanded by many mechanism analysis and kinetic modeling methods.

## 2. Coarse grain Chignolin under physics restraints

Particle-based CG models are widely adopted in bio-molecular modeling (44). Conventionally, such models are premised on some empirical forms of force fields and fitted w.r.t. experimental and/or high-level calculation data. Since the force fields are generally designed to reflect certain physics restraints, the number of parameters are relatively limited and bottlenecks the expressivity and flexibility of the resulting models. In contrast, it is also possible to construct a CG potential entirely through an ANN in order to improve the expressivity, however, the model would require excessive amount of samples for training and might not generalize well due to the large parameter space. VIFE can help combine the best of the two worlds if the CG potential takes a hybrid form,

$$F_\theta(\mathbf{s}) = F_\Theta(\mathbf{s}) + F_{\text{FF}}(\mathbf{s}) \quad (8)$$

where  $F_\Theta(\mathbf{s})$  is a trainable parametric model; while  $F_{\text{FF}}(\mathbf{s})$  is a force-field-like term accounting for some *a priori* knowledge or physics restraints. Equation (8) can be interpreted from two perspectives: On the one hand,  $F_\Theta(\mathbf{s})$  can be regarded as a correction term for the traditional force fields. On the other hand,  $F_{\text{FF}}(\mathbf{s})$  serves as a reasonable prior distribution which effectively restricts the hypothesis space of  $F_\theta$  thus expedites the training.

We tested this idea on a mini-protein Chignolin using long all-atom MD simulation trajectories contributed by Lindorff-Larsen *et al.* (45). The positions of backbone carbon and nitrogen atoms are selected as the CG variables  $\mathbf{s}$ .  $F_{\text{FF}}(\mathbf{s})$  contains physical restraints over bonds and angles, whereas  $F_\Theta(\mathbf{s})$  is an ANN accounting for many-body interactions, and the hybrid  $F_\theta$  is trained according to Eq. (3) (see more details about model setups and training details in SI). As shown in Fig. 3A, after training  $F_\theta$  over c.a. 500,000 FG samples for 5 epochs (one training epoch takes about one hour in wall-clock time on a single

NVIDIA GeForce GTX 1650 GPU card), the distributions of  $F_\theta$  for the FG and CG samples overlap fairly well, indicating that the resulting CG model is likely to reproduce the conformational distributions of FG samples. In order to examine whether  $F_\theta$  captures the important protein conformations, we characterized the generated CG samples from  $F_\theta$  with two widely-used metrics: the root-mean-squared deviation (rmsd) w.r.t. the folded structure and the radius of gyration (Rg). We compared the distribution of the generated samples against the real ones in terms of these two metrics (Fig. 3B), and found that the CG potential faithfully reproduced the overall conformational features. Specifically, a sharp local minimum corresponds to the compact native conformation (small rmsd and small Rg) is well defined by  $F_\theta$ , meanwhile the distribution of the extended unfolded structures is preserved as well. Furthermore, we showcase several typical CG structures, superimposed with their best-aligned FG counterparts. It can be shown that sampling from  $F_\theta$  can yield diverse and realistic CG backbone conformations comparable to FG ones, including the folded structure (C1 in Fig. 3C), intermediates (C2 and C3 in Fig. 3C) and unfolded structures (C4 and C5 in Fig. 3C), demonstrating that the CG model obtained by VIFE is free of mode-dropping and is able to capture the conformational features of protein backbones.

### 3. Reinforced sampling of alanine dipeptide in explicit water

Next, we proposed to construct a CG model for a prototypical bio-molecular system, alanine dipeptide (Ala2) in explicit water without available FG samples. Specifically, the backbone torsional angles were chosen to be the CG variables (Fig. 4A), that is,  $\mathbf{s} = (\phi, \varphi)$ , and our goal is to infer a reasonable FES,  $F_\theta(\mathbf{s})$ . However, different from the previous toy model, isomerization of  $(\phi, \varphi)$  involves relatively high barrier, thus brute-force FG simulations of Ala2 converge too slowly to obtain an accurate estimate of  $\langle \nabla_\theta F_\theta(\mathbf{s}) \rangle_{p_{\text{FG}}}$ . Therefore, we adopted RE-VIFE to enhance the sampling over  $\mathbf{s}$ . Technically, we simultaneously launched two simulations: one FG (all-atom) MD simulation under a bias potential  $V_\phi(\mathbf{s})$ , and a CG MC simulation over  $F_\theta(\mathbf{s})$ . Both  $F_\theta$  and  $V_\phi$  are initialized to be zero everywhere. After a period of FG sampling (40 ps in length) biased by  $V_\phi$ , we reweight the yielded FG samples to represent  $p_{\text{FG}}$ , and optimize  $F_\theta$  according to Eq. (3). Based on the newly-trained  $F_\theta$ , a well-tempered target distribution  $p_{\text{T}}$  (see SI for more details) is established, w.r.t. which the bias potential  $V_\phi$  is optimized.  $V_\phi$  is then fed into the FG simulations of Ala2 and yields an updated collection of samples representing  $p_{\text{FG}}$ . This procedure constitutes one iteration of RE-VIFE, and the entire process continues till the convergence criteria are met.

We tracked how  $F_\theta$  of FG simulation samples and CG ones distribute (Fig. 4B), provided that the similarity (or overlap) between these two distributions is a good indicator of convergence. We trained the FG and CG models by RE-VIFE for 8 ns (or equivalently, 200 iterations), and found that the two distributions overlap well and that both spread for a relatively wide range (implying no mode-dropping).

Noteworthy, in such a short simulation length, it is impossible for brute-force MD to produce equilibrium samples covering all important metastable states. To illustrate how RE-VIFE helps enhance the sampling of FG models, we presented the 1-ns trajectories for torsions  $\phi$  and  $\varphi$  produced by vanilla MD in contrast to those produced by simulations biased by  $V_\phi$  in Fig. 4C. It can be seen that isomerization of torsion  $\phi$  is fairly frequent in biased MD but hardly found in vanilla MD. Similarly, rotation of torsion  $\varphi$  is also boosted significantly by  $V_\phi$ .

One may wonder how  $V_\phi$  looks like and why it is able to boost the FG sampling so efficiently. By drawing the contour map of the bias potential (Fig. 4D), we observed that the optimized  $V_\phi$  appears complementary to the ground-true FES of  $\mathbf{s} = (\phi, \varphi)$

(a reference FES of  $\mathbf{s}$  was presented in Fig. S2). We also superimposed some randomly selected FG samples produced under  $V_\phi$  over the contour map (Fig. 4D), demonstrating an excellent coverage over both the free-energy minima and the transition regions. Therefore, samples from  $V_\phi$  are better representatives of  $p_{\text{FG}}$  and can be reliably used to optimize the CG models. The final CG model ( $F_\theta$ ) optimized via RE-VIFE, which can be regarded as a variationally approximated FES for Ala2 in explicit water, is shown in Fig. 4E. We found that  $F_\theta$  not only captures all known metastable states of Ala2 w.r.t.  $(\phi, \varphi)$  (i.e., no mode-dropping), but also quantitatively agrees well with the reference FES (Fig S2). This example demonstrates that simulations on multiple scales can be bridged by RE-VIFE and that CG models can be reliably inferred even without access to FG samples *a priori*.

#### 4. Reinforced sampling of chemical reactions in explicit solvent

Chemical reactions in condensed phase are notoriously known to be difficult to simulate, and enhanced sampling approaches are often needed to this end. We thus employed RE-VIFE to boost the simulation of a Claisen rearrangement reaction (Fig. 5A), which involves relatively high energy barrier, in the media of ionic liquid (see more details about simulation setup in SI). According to previous studies (46), a linear combination of the breaking/forming bonds (i.e.,  $d1$  and  $d2$  in Fig. 5A) is selected as the CV  $s$ , and the target distribution over  $s$  takes a Lorentzian form (47) which is developed to help enhance the sampling of the transition state regions (see more details about the CV, target distribution and training details in SI). The reactant is treated quantum mechanically (QM) while the solvent treated molecular mechanically (MM). One iteration of RE-VIFE consists of 60-ps QM/MM simulation. As shown in Fig. 5B, after 100 iterations of RE-VIFE, the chemical transitions can take place back-and-forth within several nano-seconds by virtue of RE-VIFE (i.e., the enhanced reaction rate is about  $0.1 \text{ ns}^{-1}$ ). Noteworthy, as can be inferred from  $F_\theta$  (Fig. 5C) which approximates the equilibrium potential of mean force over  $s$  (the putative reaction coordinate), the reaction rate is indeed very slow (less than  $0.1 \text{ s}^{-1}$  according to transition state theory). Therefore, within less than 10 ns simulation and training time, RE-VIFE achieves an acceleration of the reaction rate by nearly 9 orders of magnitude.

To understand how RE-VIFE achieves such astounding performance, we examined the target distribution  $p_T$  and the optimized bias potential  $V_\phi(s)$ . It can be seen from Fig. 5C that, based on the well-trained  $F_\theta$ , the target distribution automatically emphasizes on the TS region while understates the reactant and product regions. In order to arrive at the target distribution, the bias potential  $V_\phi$  is optimized in the way that the energy barrier is lowered and the metastable wells are lifted (Fig. 5C). By doing so, RE-VIFE enables one to sample various molecular configurations efficiently (Fig. 5E), based on which we can investigate interesting thermodynamic properties of the reaction. For instance, the products of this reaction consists of a pair of enantiomers (Figs. 5A and 5E), the sampling of which is challenging for many sampling methods like umbrella sampling (6). We constructed the 2-dimensional FES (Fig. 5D) spanned by the chirality order parameter  $q_C$  (the definition of  $q_C$  can be found in SI) (48) and  $s$  via KDE over samples generated by RE-VIFE. We found that both enantiomers are adequately sampled. More importantly, the final FES appears fairly symmetric w.r.t.  $q_C$ , indicating that the distributions of the two enantiomers are almost identical, agreeing with the fact that the enantiomeric excess of this reaction should be zero. This result also demonstrates that the sampling of such a complex chemical reaction has converged.

## IV. Concluding Remarks



Despite that all-atom and *ab initio* MD simulations have assisted scientists gain insights over many important physical, chemical and biological processes, their applications to complex systems containing rare events are limited, because experimentally related timescales of such systems (like protein folding and chemical reactions) are well beyond the reachable scope for even the most powerful supercomputers. Two distinct strategies are separately developed to combat this issue: either to perform enhanced sampling over the atomistic model, or to leverage CG models at the cost of losing atomic details. However, traditional enhanced sampling methods can neither scale well to large system sizes, nor transfer well to different system types due to the requirement of system-specific expert knowledge. On the other end, although different attempts exist to build CG models incorporating atom-level knowledge, transitioning from atomistic models to CG models still remains challenging.

In this paper we developed a machine-learning approach, (RE-)VIFE, to connect FG and CG models. In (RE-)VIFE, simulations on different scales can benefit from each other: CG models are optimized w.r.t. the FG simulations hence incorporating information on finer scales; In turn, FG simulations are enhanced under the guidance of CG models. Mathematically, (RE-)VIFE belongs to the realm of unsupervised and reinforcement learning. The variational and self-adaptive training objective allows end-to-end and online training of parametric models like ANNs. Through several experiments we show that (RE-)VIFE is able to yield flexible CG models more rapidly than traditional CG methods, moreover, it can also boost the sampling efficiency of chemical reactions and conformational transitions of prototypical biomolecules by several orders of magnitude.

In (RE-)VIFE, CG models can be variationally inferred based merely on equilibrium FG samples, thus involving less artifacts and computational cost than existing methods. This feature allows researchers to fully exploit the available atomistic simulations in order to construct transferrable CG models. More importantly, (RE-)VIFE also allows one to construct CG models even without access to FG samples *a priori*. We remark here that although the CG models obtained this way are able to reproduce the thermodynamic properties of finer-scale simulations, the dynamics is not necessarily correct. In terms of Langevin dynamics, our CG models only provide a reliable description of the drifting field, yet the diffusion field remains unknown. Inferring the diffusion field based on the available drifting field can be an interesting direction for further studies.

In the presented examples, (RE-)VIFE is composed of one CG-FG cycle over merely two scales, because we assume that sampling over CG models can be readily achieved with simple simulation techniques and converges much faster than the FG simulations. While this is true for most cases because CG models are designed for computational tractability, some CG models for very large or complex systems may entail heavy computation and may also suffer from the sampling issue like FG models. If coarser-grained models were built on top of the CG model, they could possibly help enhance the CG sampling in a way similar to how the CG model boosts the FG sampling. Following this line, it is appealing to develop models over a cascade of scales via (RE-)VIFE. Besides, if one only needs to enhance the sampling over certain trans-rotationally invariant CVs via RE-VIFE, parametric energy models with a normalized partition function may be preferred as  $F_\theta$  due to simpler sampling procedure. We leave these interesting ideas for future research

## Acknowledgements

The authors thank Xing Che, Frank Noé and Lijiang Yang for useful discussion. This research was supported by National Natural Science Foundation of China [21573006, 21233002 and 91427304 to Y.Q.G] and the National Key Research and

Development Program of China [2017YFA0204702 to Y.Q.G.]. The author J.Z. thanks the Alexander von Humboldt Foundation for supporting this research.

## Associated Content

Supplementary texts including introduction to derivation of VIFE training objectives, simulation setups and supplemental figures are provided in Supplemental Information.

**Author contributions:** J.Z., Y.K.L., Y.I.Y and Y.Q.G. designed the research; J.Z., Y.K.L. and Y.I.Y performed the research; J.Z., Y.K.L., Y.I.Y and Y.Q.G. analyzed the data; J.Z., Y.K.L., Y.I.Y and Y.Q.G. wrote the paper.

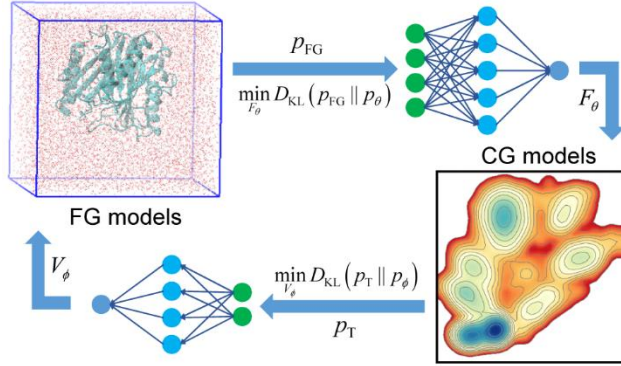
**Notes:** The authors declare no conflict of interest.

## REFERENCES

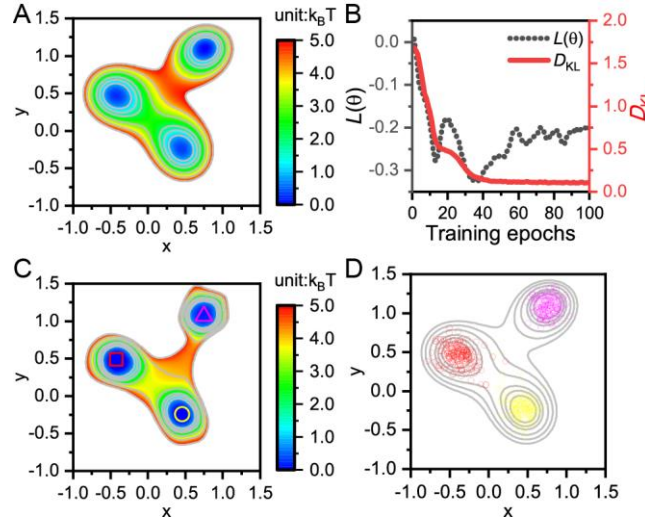
1. Frenkel D & Smit B (2001) *Understanding molecular simulation: from algorithms to applications* (Elsevier).
2. Tuckerman ME (2002) Ab initio molecular dynamics: basic concepts, current trends and novel applications. *Journal of Physics: Condensed Matter* 14(50):R1297.
3. Fiorin G, Klein ML, & Hémin J (2013) Using collective variables to drive molecular dynamics simulations. *Molecular Physics* 111(22-23):3345-3362.
4. Voth GA (2008) *Coarse-graining of condensed phase and biomolecular systems* (CRC press).
5. Noid W, *et al.* (2008) The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *The Journal of chemical physics* 128(24):244114.
6. Torrie GM & Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* 23(2):187-199.
7. Silverman BW (2018) *Density estimation for statistics and data analysis* (Routledge).
8. LeCun Y, Bengio Y, & Hinton G (2015) Deep learning. *nature* 521(7553):436.
9. Goodfellow I, Bengio Y, & Courville A (2016) *Deep learning* (MIT press).
10. Schneider E, Dai L, Topper RQ, Drechsel-Grau C, & Tuckerman ME (2017) Stochastic neural network approach for learning high-dimensional free energy surfaces. *Physical review letters* 119(15):150601.
11. Zhang L, Wang H, & E W (2018) Reinforced dynamics for enhanced sampling in large atomic and molecular systems. *The Journal of chemical physics* 148(12):124113.
12. Wang J, *et al.* (2019) Machine learning of coarse-grained molecular dynamics force fields. *ACS central science*.
13. Vapnik VN (1999) An overview of statistical learning theory. *IEEE transactions on neural networks* 10(5):988-999.
14. Kingma DP & Welling M (2013) Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.
15. Van den Oord A, Kalchbrenner N, Espeholt L, Vinyals O, & Graves A (2016) Conditional image generation with pixelcnn decoders. *Advances in neural information processing systems*, pp 4790-4798.
16. Dinh L, Krueger D, & Bengio Y (2014) Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*.
17. Abbasnejad ME, Shi Q, Hengel Avd, & Liu L (A Generative Adversarial Density Estimator. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
18. Salakhutdinov R, Mnih A, & Hinton G (2007) Restricted Boltzmann machines for collaborative filtering. *Proceedings of the 24th international conference on Machine learning*, (ACM), pp 791-798.
19. Salakhutdinov R & Hinton G (2009) Deep Boltzmann Machines. in *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, eds David van D & Max W (PMLR, Proceedings of Machine Learning Research), pp 448--455.
20. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences* 79(8):2554-2558.
21. Wales D (2003) *Energy landscapes: Applications to clusters, biomolecules and glasses* (Cambridge University Press).
22. Du Y & Mordatch I (2019) Implicit generation and generalization in energy-based models. *arXiv preprint arXiv:1903.08689*.

23. Kullback S & Leibler RA (1951) On information and sufficiency. *The annals of mathematical statistics* 22(1):79-86.
24. Arjovsky M, Chintala S, & Bottou L (2017) Wasserstein generative adversarial networks. *International conference on machine learning*, pp 214-223.
25. Shell MS (2008) The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *The Journal of chemical physics* 129(14):144108.
26. Chaimovich A & Shell MS (2011) Coarse-graining errors and numerical optimization using a relative entropy framework. *The Journal of chemical physics* 134(9):094112.
27. Valsson O & Parrinello M (2014) Variational approach to enhanced sampling and free energy calculations. *Physical review letters* 113(9):090601.
28. Zhang J, Yang YI, & Noé F (2019) Targeted Adversarial Learning Optimized Sampling.
29. Lyubartsev AP (2005) Multiscale modeling of lipids and lipid bilayers. *European Biophysics Journal* 35(1):53.
30. Abrams C & Bussi G (2014) Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy* 16(1):163-199.
31. Okamoto Y (2004) Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *Journal of Molecular Graphics and Modelling* 22(5):425-439.
32. Yang YI, Shao Q, Zhang J, Yang L, & Gao YQ (2019) Enhanced sampling in molecular dynamics. *The Journal of Chemical Physics* 151(7):070902.
33. Laio A & Parrinello M (2002) Escaping free-energy minima. *Proceedings of the National Academy of Sciences* 99(20):12562-12566.
34. Sutton RS & Barto AG (2018) *Reinforcement learning: An introduction* (MIT press).
35. Grondman I, Busoniu L, Lopes GA, & Babuska R (2012) A survey of actor-critic reinforcement learning: Standard and natural policy gradients. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42(6):1291-1307.
36. Goodfellow I, et al. (2014) Generative adversarial nets. *Advances in neural information processing systems*, pp 2672-2680.
37. Heusel M, Ramsauer H, Unterthiner T, Nessler B, & Hochreiter S (2017) Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems*, pp 6626-6637.
38. Van Hasselt H, Guez A, & Silver D (2016) Deep reinforcement learning with double q-learning. *Thirtieth AAAI conference on artificial intelligence*.
39. Botvinick M, et al. (2019) Reinforcement learning, fast and slow. *Trends in cognitive sciences*.
40. Fujimoto S, van Hoof H, & Meger D (2018) Addressing function approximation error in actor-critic methods. *arXiv preprint arXiv:1802.09477*.
41. Barducci A, Bussi G, & Parrinello M (2008) Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical review letters* 100(2):020603.
42. Tiwary P & Berne B (2017) Predicting reaction coordinates in energy landscapes with diffusion anisotropy. *The Journal of chemical physics* 147(15):152701.
43. Singer S & Nelder J (2009) Nelder-mead algorithm. *Scholarpedia* 4(7):2928.
44. Kmiecik S, et al. (2016) Coarse-Grained Protein Models and Their Applications. *Chemical Reviews* 116(14):7898-7936.
45. Lindorff-Larsen K, Piana S, Dror RO, & Shaw DE (2011) How fast-folding proteins fold. *Science* 334(6055):517-520.
46. Zhang J, et al. (2017) Rich Dynamics Underlying Solution Reactions Revealed by Sampling and Data Mining of Reactive Trajectories. *ACS Central Science* 3(5):407-414.
47. Debnath J, Invernizzi M, & Parrinello M (2019) Enhanced Sampling of Transition States. *Journal of Chemical Theory and Computation* 15(4):2454-2459.
48. Zhang J, Yang YI, Yang L, & Gao YQ (2015) Dynamics and Kinetics Study of "In-Water" Chemical Reactions by Enhanced Sampling of Reactive Trajectories. *The Journal of Physical Chemistry B* 119(45):14505-14514.

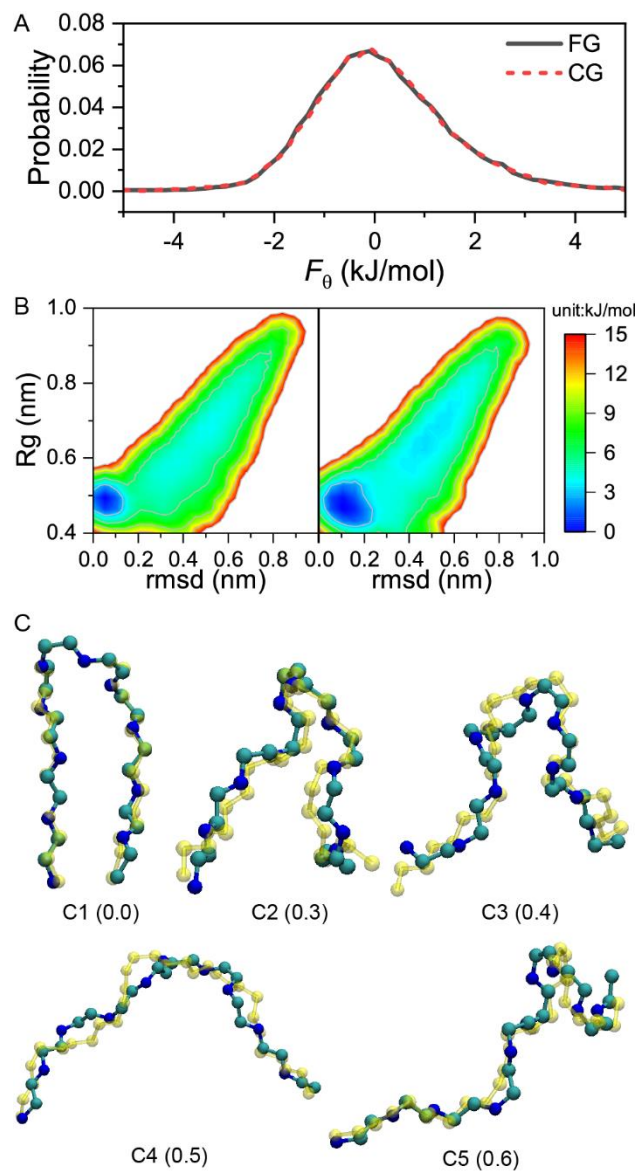
## Figures



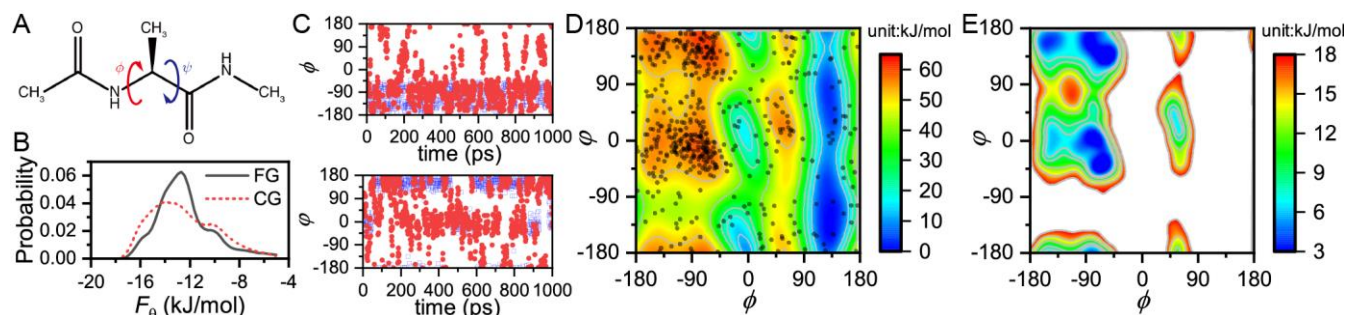
**Figure 1.** Illustration of RE-VIFE. Given simulation samples from fine-grained (FG) models ( $p_{FG}$ ), a coarse-grained (CG) potential (or free-energy)  $F_\theta$  can be variationally approximated. In turn, a target distribution  $p_T$  can be defined based on CG simulations running over  $F_\theta$ , according to which a bias potential  $V_\phi$  can be variationally optimized to boost the FG simulation. This illustration consists of only one CG-FG cycle, yet additional cycles corresponding to more scales can be added on top in the same manner.



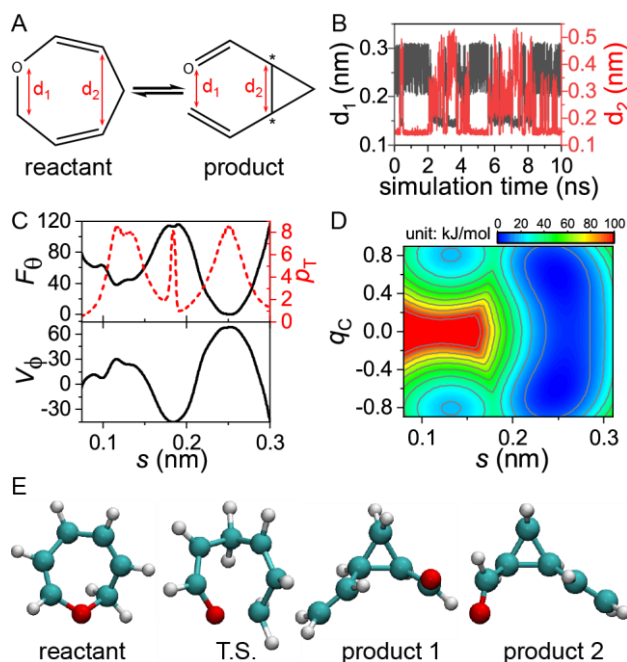
**Figure 2.** Application of VIFE on a 2D toy model. (A) 2D potential energy surface (PES),  $U(x, y)$  of the toy model. (B) Evolution of effective loss function  $L(\theta)$  defined in Eq. (5) (black dotted line) and KL-divergence  $D_{KL}(p_{FG} || p_\theta)$  (red solid line) against training epochs by VIFE. (C) The contour map corresponds to the optimized  $F_\theta$  by VIFE, the three symbols (magenta triangle, red square and yellow circle) represent three free-energy minima found in  $F_\theta$ . (D) Hollow circles are simulation samples produced by Langevin dynamics on  $U(x, y)$  colored by which local minimum they are minimized to over  $F_\theta$ . Colors are in line with the symbols in panel C. The contour map of  $U(x, y)$  is shown in grey as background.



**Figure 3.** Coarse grain Chignolin via VIFE. (A) Distributions of  $F_0$  for all-atom MD simulation samples (black solid line) and for CG simulation samples (red dashed line). (B) 2D FES spanned by rmsd and  $R_g$  of all-atom MD simulation samples (left panel) and CG simulation samples (right panel). (C) Representative CG structures colored by cyan (carbon atoms) and blue (nitrogen atoms), superimposed with their best-aligned FG counterparts (colored by transparent yellow). Numbers in parenthesis are the rmsd values (in unit of nm) of CG structures w.r.t. the folded structure.



**Figure 4.** RE-VIFE sampling of Ala2 in explicit water. (A) Chemical structure of Ala2 and two coarse-graining variables: the torsions  $\phi$  and  $\psi$ . (B) Distributions of  $F_\theta$  for all-atom MD simulation samples (black solid line) and for CG simulation samples (red dotted line). (C) 1-ns simulation trajectories projected on torsion  $\phi$  (upper panel) and torsion  $\psi$  (lower panel). Blue squares correspond to vanilla MD, red dots to MD biased by  $V_\phi$ . (D) The contour map of  $V_\phi$  optimized via RE-VIFE. Superimposed black dots are representative samples produced by the enhanced MD simulation under  $V_\phi$ . (E) Contour map of  $F_\theta$  optimized via RE-VIFE.



**Figure 5.** RE-VIFE sampling of chemical reactions in ionic liquid. (A) The retro-Claisen rearrangement under study. The reactant is a 7-member-ring ether while the product is an aldehyde containing a 3-member ring. Two potential chiral carbons in the product are indicated by star symbols. The breaking and forming bonds are denoted as  $d_1$  and  $d_2$ , respectively. (B) Trajectories of  $d_1$  (black) and  $d_2$  (red) during RE-VIFE simulations. (C) Upper panel: the optimized CG potential  $F_\theta(s)$  (black solid line) in unit of kJ/mol and the associated (unnormalized) target distribution  $p_T(s)$  (red dashed line). Lower panel: the optimized bias potential  $V_\theta$  in unit of kJ/mol. (D) 2D FES over the CG variable  $s$  and chirality order parameter  $q_c$ . (E)

Snapshots of various molecular configurations during RE-VIFE. From left to right: reactant, transition state (T.S) and two enantiomer products.

## Supplemental Information

### Reinforcement Learning for Multi-Scale Molecular Modeling

Jun Zhang<sup>1</sup>, Yao-Kun Lei<sup>2</sup>, Yi Isaac Yang<sup>3,†</sup> and Yi Qin Gao<sup>2,3,‡</sup>

<sup>1</sup>*Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany*

<sup>2</sup>*Institute of Theoretical and Computational Chemistry, College of Chemistry and Molecular Engineering, Peking University, 100871 Beijing, China.*

<sup>3</sup>*Institute of Systems Biology, Shenzhen Bay Laboratory, 518055 Shenzhen, China*

Corresponding authors: (Y.I.Y) <sup>†</sup> yangyi@szbl.ac.cn and (Y.Q.G) <sup>‡</sup> gaoyq@pku.edu.cn



## TABLE OF CONTENTS

<b>Part I. Supplemental Texts</b>		
I	Training objective of (RE-)VIFE	1
II	Parametric models in (RE-)VIFE	1
III	Target distribution for RE-VIFE	2
IV	Connection between RE-VIFE and other sampling methods	3
V	Algorithm for RE-VIFE	4
<b>Part II. Simulation Details</b>		
I	VIFE for numerical toy model	5
II	VIFE for Chignolin	5
III	RE-VIFE for Alanine dipeptide (Ala2)	5
IV	RE-VIFE for Claisen rearrangement	6
<b>Part III. Supplemental Figures</b>		7

## PART I. SUPPLEMENTAL TEXTS

### I. Training objective of (RE-)VIFE

Given the target distribution  $p(\mathbf{s})$  and an approximate distribution  $p_\theta(\mathbf{s})$ , there are some practical ways to define the divergence between the two. The most commonly used one is the Kullback-Leibler divergence or relative entropy.

#### A. Kullback-Leibler divergence

$$D_{\text{KL}}(p \parallel p_\theta) = \int p(\mathbf{s}) \log \left( \frac{p(\mathbf{s})}{p_\theta(\mathbf{s})} \right) d\mathbf{s} \quad (\text{S1})$$

where both distributions are assumed to admit densities (i.e. absolutely continuous) with respect to a same measure. KL-divergence is known to be asymmetric and possibly infinite when there is non-overlapping area between the two distributions. Therefore, effective utilization of KL divergence as a optimization objective requires substantial overlap between  $p(\mathbf{s})$  and  $p_\theta(\mathbf{s})$  (1).

#### B. Training objective of VIFE

Now we consider the derivatives of Eq. (S1) w.r.t. the parameters  $\theta$ , namely,  $\nabla_\theta D_{\text{KL}}(p \parallel p_\theta)$ . First, we can rewrite  $p$  and  $p_\theta$  in terms of the free energy,

$$p(\mathbf{s}) = \frac{\exp(-\beta F(\mathbf{s}))}{Z} \quad (\text{S2})$$

$$p_\theta(\mathbf{s}) = \frac{\exp(-\beta F_\theta(\mathbf{s}))}{Z_\theta} \quad (\text{S3})$$

following which Eq. (S1) becomes

$$D_{\text{KL}}(p \parallel p_\theta) = \int [\beta F_\theta(\mathbf{s}) - \beta F(\mathbf{s})] p(\mathbf{s}) d\mathbf{s} + \int (\log Z_\theta - \log Z) d\mathbf{s} \quad (\text{S4})$$

It can be found that  $D_{\text{KL}}(p \parallel p_\theta)$  is now a functional of  $F_\theta$ , hence we can obtain the functional derivatives,

$$\frac{\delta Z_\theta}{\delta F_\theta} = -\beta \exp(-\beta F_\theta) \quad (\text{S5})$$

$$\frac{\delta D_{\text{KL}}(p \parallel p_\theta)}{\delta F_\theta} = \frac{\beta \exp(-\beta F)}{Z} - \frac{\beta \exp(-\beta F_\theta)}{Z_\theta} \quad (\text{S6})$$

Notice that Eq. (S6) is equivalent to

$$\frac{\delta D_{\text{KL}}}{\delta F_\theta} = \beta p(\mathbf{s}) - \beta p_\theta(\mathbf{s}) \quad (\text{S7})$$

so we finally arrive at the derivatives of  $D_{\text{KL}}$  w.r.t.  $\theta$ ,

$$\begin{aligned} \frac{\partial D_{\text{KL}}}{\partial \theta} &= \int \frac{\delta D_{\text{KL}}}{\delta F_\theta} \frac{\partial F_\theta(\mathbf{s})}{\partial \theta} d\mathbf{s} \\ &= \int \beta \frac{\partial F_\theta(\mathbf{s})}{\partial \theta} (p(\mathbf{s}) - p_\theta(\mathbf{s})) d\mathbf{s} \\ &= \left\langle \beta \frac{\partial F_\theta(\mathbf{s})}{\partial \theta} \right\rangle_{p(\mathbf{s})} - \left\langle \beta \frac{\partial F_\theta(\mathbf{s})}{\partial \theta} \right\rangle_{p_\theta(\mathbf{s})} \end{aligned} \quad (\text{S8})$$

We remark here that Eq. (S8) holds for arbitrary fixed  $p(\mathbf{s})$  when  $p_\theta(\mathbf{s})$  is a learnable distribution used to approximate  $p(\mathbf{s})$ . So it can be easily generalized to the training objective of RE-VIFE (Eq. (4) in the main text). Besides, the derivation of Eq. (S8) shares dramatic similarity with the training objective of TALOS (2).

### II. Parametric models in (RE-)VIFE

There is relatively large margin for the specific functional forms for  $F_\theta$  (and/or  $V_\phi$ ). For example, it can be a linear expansion of certain basis functions as in VES where  $\theta$  are the expansion coefficients (3); or can be a non-linear neural network, where  $\theta$  are the built-in parameters of neural network (4).

#### A. Orthonormal polynomials or functions

If  $\text{Dim}(\mathbf{s}) \leq 3$ , we recommend orthonormal basis functions as  $F_\theta$  and  $V_\phi$ , and the expansion coefficients are the learnable parameters. For periodic CG variables  $\mathbf{s}$ , Fourier expansions can be adopted. For non-periodic  $\mathbf{s}$ , Legendre or Chebyshev polynomials can be used (3). Orthonormal polynomials usually yield smooth energy function, so generally no additional regularization is needed.

#### B. Building blocks for neural network models

ANNs are expressive parametric models which can be used as  $F_\theta$  and  $V_\phi$ . Modern deep learning models have evolved based on specially designed architecture. Here we briefly introduce some building blocks which may be useful for constructing the deep models in (RE-)VIFE.

##### 1. Multi-layer perceptron (MLP)

MLPs are most commonly seen ANNs consisting of fully connected hidden layer. The input of a MLP,  $\mathbf{s}$ , should translational invariant features of the molecular system. More importantly, each dimension of the input vector  $\mathbf{s}$  should be indexable. MLP will transform the input vector to hidden features, and finally yields an output vector.

##### 2. Transformer & Attention

If  $\mathbf{s}$  consists of a sequence (or set) of feature vectors,  $(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n)$ , (e.g.,  $\mathbf{s}$  can be a series of features describing each amino residue along a protein), then the Transformer-like model (5) based on attention mechanism can be

constructed. Specifically, the attention mechanism learns a pair of Query vector and Key vector, and regard  $\mathbf{s}$  as the Value vector, then perform multi-head self-attention (5) to transform  $(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n)$  into a new series of vectors  $(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n)$ . Based on the new feature vectors, one can employ MLP to point-wisely transform  $\mathbf{z}_i$  to an output vector. One merit of Transformer-based model is resulting potential can be transferable to systems of different sizes and containing various types of particles. We refer readers interested in this topic to relevant reference (5).

### 3. Message-passing neural networks (MPNN)

Molecular systems consisting of particles (e.g., atoms) can be viewed as a graph, where vertices (or nodes) represent the particles while the interactions between particles can be modeled by graphical edges. MPNNs (6) are generalization of graph neural network family, hence can be used to model molecular systems. MPNNs exhibit a nice property that preserves the trans-rotational invariance and permutation invariance of the many-particle system. There are several recently developed MPNNs that specifically aim at molecular systems, including SchNet (7) and PhysNet (8). Both models directly learn a function based on the Cartesian coordinates and the type of the particles. Such models are potentially useful in constructing particle-based CG models for biomolecules.

### C. Neural allocative potentials (NAP)

We note here that it is not the absolute value but the difference of energy makes physical sense. If we do not exploit this property, the absolute value of the energy produced by ANNs may grow rapidly but the energy difference does not. This is harmful to our purpose, because it was recently revealed that the smoothness and continuity of ANNs become vulnerable if the absolute values of built-in parameters are too large hence the weight matrices of ANNs become ill-conditioned (9). Given above considerations, we first choose a lower bound and an upper bound for the bias potential, and "quantized" this energy range into  $K$  fixed levels  $\{E_k\}_{k=1..K}$ , based on which we propose the following functional form,

$$F_\theta(\mathbf{s}) = \sum_{k=1}^K \alpha_k(\mathbf{s}; \theta) E_k \quad (\text{S9})$$

$$\sum_{k=1}^K \alpha_k(\mathbf{s}; \theta) = 1 \quad (\text{S10})$$

where  $\alpha_k(\mathbf{s}; \theta)$  corresponds to the output of an ANN with a soft-max output layer. The associated force takes the following form,

$$-\nabla_{\mathbf{s}} F_\theta(\mathbf{s}) = \sum_{k=1}^K E_k \nabla_{\mathbf{s}} \alpha_k(\mathbf{s}; \theta) \quad (\text{S11})$$

In other words, we transform the problem of learning a scalar into learning a simplex  $\{\alpha_k\}_{k=1..K}$ . Following this form, NAP is trained to allocate proper amount of energy to

configuration  $\mathbf{s}$ , rather than estimate the absolute value of the bias potential. Therefore, the training of  $F_\theta(\mathbf{s})$  is more robust and the gradient will be well-behaving.

### D. Gradient regularization

In order to ensure the to-be-optimized function  $F_\theta(\mathbf{s})$  to be smooth enough for integration, we introduce several techniques to effectively regularize the gradient,  $\nabla_{\mathbf{s}} F_\theta(\mathbf{s})$ .

#### 1. Weight clipping

One simple and intuitive approach is to clip the weights (1) of the neural network  $F_\theta$  (e.g.,  $\theta \in \mathcal{W}[-0.01, 0.01]$ ) after each gradient update. The reason behind is that, if the parameter space  $\mathcal{W}$  for  $\theta$  is compact, then all the functions  $F_\theta$  will be  $K$ -Lipschitz for some  $K$  that only depends on  $\mathcal{W}$  and not the individual weights. In practice, the clipping range is a very important hyper-parameter, which can neither be too large (reduced compactness) nor too small (reduced capacity of  $F_\theta$ ). Besides, usually this method requires batch-normalization (10) in order to achieve a robust and good performance.

#### 2. Spectral normalization

One more recent approach to stabilize the training of GAN can also be applied in (RE-)VIFE, which is to perform spectral normalization for the weight matrices  $\theta$  in the energy function  $F_\theta$ . The basic idea is to constrain the Lipschitz constant of the function by restricting the spectral norm of each layer. Compared to other normalization techniques, spectral normalization does not require extra hyper-parameter tuning (setting the spectral norm of all weight layers to 1 consistently performs well in practice). Moreover, the computational cost is also relatively small. We refer readers interested in this technique to the reference (9).

#### 3. Drift removal

In experiments we found that removing the net drift of the energy function will help stabilize training. To be specific, we added a regularization term to the overall loss function, to restrict the mean value of the energy distribution of FG samples to be close to zero. This trick works possibly because the output of the network is anchored to a nearly zero-mean value, so the absolute value of the output (and the built-in parameters) will not be too large.

## III. Target distribution for RE-VIFE

### A. Well-Tempered (WT) distribution

Given the density estimation, i.e.,  $F_\theta(\mathbf{s})$ , we can define the target distribution in WT form (11), which is equivalent to a partially flattened FES,

$$p_T(\mathbf{s}) = \frac{e^{-\frac{\beta}{\gamma} F_\theta(\mathbf{s})} d\mathbf{s}}{\int e^{-\frac{\beta}{\gamma} F_\theta(\mathbf{s})} d\mathbf{s}} \quad (\text{S12})$$

where  $\gamma > 1$  is the WT factor.

### B. Lorentzian-coupled distribution

As recently proposed by Debnath et al. (12), the target distribution may take the following form:

$$p_T(\mathbf{s}) = \frac{l(\mathbf{s}) p_{WT}(\mathbf{s}; \gamma)}{\int l(\mathbf{s}) p_{WT}(\mathbf{s}; \gamma) d\mathbf{s}} \quad (\text{S13})$$

$$l(\mathbf{s}) = \frac{\zeta}{\zeta^2 + \left[ \beta \frac{dF_\theta(\mathbf{s})}{d\mathbf{s}} \right]^2} \quad (\text{S14})$$

where  $\zeta$  is a scaling factor and  $p_{WT}(\mathbf{s}; \gamma)$  is a pre-defined WT distribution with the WT-factor  $\gamma$  (Eq. S12). More introduction to this kind of target distribution can be found in the reference (12). The advantage of such a Lorentzian-type target is that, the transition-state region (where  $dF_\theta/d\mathbf{s} \approx 0$ ) can be further enhanced along with metastable states. This property is very useful in some scenarios where the sampling of transition state is crucial as in transition path sampling.

## IV. Connection between RE-VIFE and other sampling methods

### A. Metadynamics

Metadynamics (13) is a powerful tool in enhanced sampling of rare events if low-dimensional CV  $\mathbf{s}$  can be defined *a priori*. However, metadynamics cannot be readily applied to cases where  $\text{Dim}(\mathbf{s}) \geq 3$ . As introduced in the main text, RE-VIFE can be viewed as a generalization of metadynamics into large  $\text{Dim}(\mathbf{s})$  and parametric bias potential functions.

(1) Metadynamics exploits KDE to coin the target distribution which will fail in large  $\text{Dim}(\mathbf{s})$ . In contrast, RE-VIFE uses parametric models to perform density estimation which is even applicable to large  $\text{Dim}(\mathbf{s})$ .

(2) Metadynamics accumulated non-parametric Gaussian as bias potential which will be intractable for large  $\text{Dim}(\mathbf{s})$ ; while RE-VIFE implements parametric learning to construct the bias potential, thus being more flexible and scalable for complex systems.

In summary, RE-VIFE will be more useful if one wants to boost the sampling efficiency but does not have enough expert knowledge to determine low-dimensional CV.

### B. VES

The mathematical form of VIFE training objective is almost identical to VES (3), although it is derived from a different starting point. Both VES and RE-VIFE aims to minimize the KL-divergence between the sampled

distribution and a target distribution. Nevertheless, they can be distinguished by several remarkable differences:

(1) The target distribution in VES has to be defined carefully because VES would suffer from gradient vanishing or exploding issues when the distributions scarcely overlap. Usually this is a very demanding task. In contrast, RE-VIFE follows actor-critic learning and coins a target distribution in a metadynamics-like fashion, thus always ensures a useful gradient for optimization.

(2) Optimization in VES relies on higher-order derivatives thus involving higher computational cost, while RE-VIFE employs the state-of-the-art first-order optimization techniques from deep learning thus being robust, fast and economic.

(3) The functional form of bias potentials supported by vanilla VES is limited to linear expansions, thus the dimension of the CV space cannot be too large. In contrast, RE-VIFE supports any differentiable functions (such as neural networks) as valid forms and can be easily extended to ultra-large dimensions.

### C. TALOS

Both TALOS (2) and RE-VIFE are closely connected to actor-critic reinforcement learning. They both parametrize a pair of value function and policy function in order to boost the rare events. But there are several key differences demarcating the two approaches:

(1) In TALOS, the value network is a critic (or discriminator)  $D_w$  as in GAN; whereas in RE-VIFE, the value network  $F_\theta$  is a density estimator.

(2) TALOS optimizes the policy network (i.e., the bias potential) to minimize the Wasserstein-1 distance between the sampled distribution and the target; while RE-VIFE minimizes the KL-divergence for the same purpose.

(3) The target distribution in TALOS is manually selected *a priori*, the choice of which is sometimes tricky. While in RE-VIFE, the target distribution is automatically determined and adaptively updated according to the current density estimation.

(4) The critic and policy networks in TALOS can be defined on different vector spaces; while in RE-VIFE, both must operate on the same space.

We mark here that the strengths of TALOS and RE-VIFE are highly complementary. It remains as an interesting research direction that one may combine both methods, that is, to minimize the Wasserstein-1 distance between the sampled distribution and an adaptively determined target distribution.

---

**Algorithm S1.** Reinforced Variational Inference of Free Energy (RE-VIFE)

---

- 1:   **Input:** Initialize value network  $F_\theta$  and policy network  $V_\phi$ ; Learning rate  $\alpha_\theta$   
and  $\alpha_\phi$  for  $\theta$  and  $\phi$ , respectively.
  - 2:   **While**  $\phi$  do not converge, **do**
  - 3:     Run MD under  $V_\phi$ , collect samples to calculate  $\langle V_\phi(\mathbf{s}) \rangle_{p_\phi}$  ▷ run FG simulations
  - 4:     Define  $p_{\text{FG}}(\mathbf{s}) \propto \exp(\beta V_\phi(\mathbf{s}))$  and calculate  $\langle F_\theta(\mathbf{s}) \rangle_{p_{\text{FG}}}$  ▷ reweight biased samples
  - 5:     **For**  $t = 0 \rightarrow m$  ▷ train  $F_\theta$  for  $m$  iterations
  - 6:       Draw samples from  $F_\theta$ , calculate  $\langle F_\theta(\mathbf{s}) \rangle_{p_\theta}$ . ▷ run CG simulations
  - 7:       Calculate  $\mathcal{L}(\theta) = \langle F_\theta(\mathbf{s}) \rangle_{p_{\text{FG}}} - \langle F_\theta(\mathbf{s}) \rangle_{p_\theta}$
  - 8:        $\theta \leftarrow \text{Adam}(\nabla_\theta \mathcal{L}(\theta), \theta, \alpha_\theta)$  ▷ update value network
  - 9:     **End For**
  - 10:    Define target distribution  $p_{\text{T}}$  according to  $F_\theta$
  - 11:    Calculate  $\mathcal{L}(\phi) = \langle V_\phi(\mathbf{s}) \rangle_{p_{\text{T}}} - \langle V_\phi(\mathbf{s}) \rangle_{p_\phi}$
  - 12:     $\phi \leftarrow \text{Adam}(\nabla_\phi \mathcal{L}(\phi), \phi, \alpha_\phi)$  ▷ update bias potential
  - 13: **End While**
-

## PART II. SIMULATION DETAILS

### I. VIFE for numerical model: 2-dimensional 3-well potential

#### A. Simulation setup

The potential energy function of the 2D 3-well potential (14) takes the following form given the inverse temperature  $\beta$ :

$$\begin{aligned} \beta U(x, y) = & -16 \exp\{-2(x+0.5)^2 - 2(y-0.5)^2\} \\ & -18 \exp\{-2(x-0.8)^2 - 2(y-1.2)^2\} \\ & -16 \exp\{-2(x-0.5)^2 - 2(y+0.3)^2\} + 0.5(x^6 + y^6) \end{aligned}$$

For the overdamped Langevin simulation, we chose  $\beta = 1$ , and the diffusion tensor was set to be:

$$\mathbf{D} = \begin{pmatrix} D_{xx} & 0 \\ 0 & D_{yy} \end{pmatrix} = \begin{pmatrix} \frac{1}{5\beta} & 0 \\ 0 & \frac{1}{5\beta} \end{pmatrix}$$

The resulting white-noised Langevin dynamics was simulated with a discrete time integration step of 0.01. 50,000 samples in total were collected every 100 integration steps.

#### B. Functional form

$F_\theta$  is constructed on the 2D  $(x, y)$  space via a MLP with two hidden layers and a soft-max output layer. Hidden layers consist of 64 units with hyperbolic tangent (tanh) as activation function; output layer consists of 11 logits which uniformly discretize the energy range from -5 to 5. Namely, the output layer takes the form of NAP. Spectral normalization was exploited to regularize the gradients.

#### C. Training details

We trained  $F_\theta$  with a mini-batch size of 500. Samples from  $F_\theta$  was drawn through importance sampling. The default Adam optimizer with a learning rate of  $10^{-4}$  was adopted. Optimization was performed on Tensorflow (v1.15.0).

### II. VIFE for Chignolin

#### A. Data source

The all-atom simulation data of chignolin is taken from reference (15), which contains over 500,000 samples in total. These samples were used to approximate  $p_{\text{FG}}$ .

#### B. Functional form

The hybrid  $F_\theta$  is composed of two terms:  $F_{\text{FF}}$  and  $F_\theta$ , respectively.  $F_{\text{FF}}$  contains restraints over bonds and angles. Specifically, we impose harmonic-form restraints as  $F_{\text{FF}}$ ,

$$F_{\text{FF}}(\mathbf{s}) = \frac{1}{2} k (\mathbf{s} - \langle \mathbf{s} \rangle_{\text{FG}})^2$$

where  $\mathbf{s}$  is a bond or angle,  $\langle \mathbf{s} \rangle_{\text{FG}}$  is the average value of the corresponding  $\mathbf{s}$  from the FG simulations. The stiffness factor  $k$  is set to be 100 kJ/mol. We also manually set the torsional angle  $\omega$  to be  $\pi$ .

$F_\theta$  is designed to account for many-body interactions based on all phi and psi torsions. We expand each torsional angle with trigonal functions and use them as the input for the torsion. For example, for torsional angle  $\varphi$ ,

$$\mathbf{s}(\varphi) = (\cos \varphi, \dots, \cos n\varphi, \sin \varphi, \dots, \sin n\varphi)$$

In order to construct a model potentially transferrable to different proteins, a standard Transformer model (5) was chosen as  $F_\theta$  based on self-attention mechanisms. Position embedding (5) was used to encode the information of the amino-acid sequence. Code is adapted from GitHub: <https://github.com/Kyubyong/transformer>.

#### C. Training details

We regularized the model by removing the drift of mean energy of the FG samples.  $F_\theta$  was optimized with a mini-batch size of 512. The default Adam optimizer with  $\beta_1 = 0$  and  $\beta_2 = 0.999$  was used as recommended by Ref. (16), and a warm-up learning rate schedule as recommended by Ref. (5) was adopted. Optimization was performed on Tensorflow (v1.15.0).

### III. RE-VIFE for Alanine dipeptide (Ala2)

#### A. Simulation setup

For the alanine dipeptide in aqueous solution, no ions were added since the terminal of the alanine was neutrally blocked (namely, ACE-ALA-NME) surrounded by 384 SPCE water molecules (17). All the simulations were executed on AMBER17 package (18) using FF99SB force field (19) parameters. The aqueous solution system was put in a rectangular simulation box with periodic boundaries on. SHAKE algorithm (20) was adopted to constrain all covalent bonds involving hydrogen atoms, and a 2 fs time step was permitted. The system underwent a standard relaxation procedure and equilibrated to an NTP ensemble (300 K, 1 atm). To equilibrate the system to the appropriate volume, the pressure of the system was adjusted to 1 atm by the Berendsen weak-coupling algorithm (21) with the relaxation time constants of 0.2 ps under another 1 ns long normal MD. For production run, samples were collected every 1 ps from 10 parallel walkers (22).

#### B. Functional forms

Both  $F_\theta$  and  $V_\phi$  are functions of the 2D  $(\phi, \varphi)$  space. Since  $\phi$  and  $\varphi$  are periodic variables, we construct both functions using the Fourier polynomial:

$$F_\theta = \sum_{i=0}^{i=N} \sum_{j=0}^{j=N} \theta_{ij} [\cos(i\phi) + \sin(i\phi)] [\cos(j\phi) + \sin(j\phi)]$$

where  $N = 8$  for both  $F_\theta$  and  $V_\phi$ .

### C. Target distribution

In order to train  $V_\phi$ , we chose a WT-form target distribution with  $1/\gamma = 0.4$  based on  $F_\theta$  according to Eq. (S12).

### D. Training details

In each RE-VIFE iteration, we ran MD simulations for 40 ps to collect samples as  $p_{FG}$ . Before every training step of  $V_\phi$ ,  $F_\theta$  was first trained for 5 steps (i.e.,  $m = 5$  in Algorithm S1). We adopted the Averaged SGD algorithm as recommended by Refs. (23) and (3) for each optimization step, and the update step-size for  $V_\phi$  and  $F_\theta$  is 0.1. Codes for optimization was inherited from Ref. (2) and (3). In total, RE-VIFE was performed for 200 iterations, which is equivalent to 8-ns MD simulation.

## IV. RE-VIFE for Claisen rearrangement

### A. Simulation setup

The simulation was performed at the QM/MM interface on AMBER14 MD platform. The self-consistent charge density functional tight-binding (SCC-DFTB) method (24) was adopted to approximate the quantum mechanical Hamiltonian of the reactant molecule. The solvent is a kind of ionic liquid, containing a pair of soluble ion pairs termed as  $[C_2mim]^+[NTf_2]^-$ . We adopted the classical force field developed by Sieffert and Wipff (25, 26) to describe the solvent molecules (or ions) and SHAKE was imposed on the solvent. No additional ions were added.

The system underwent a standard relaxation procedure and equilibrated to an NTP ensemble (300 K, 1 atm) lasting for 1-ns long normal MD. A cutoff of 10.0 Å was applied for calculating nonbonding interactions. All the simulations were performed with a 1-fs time integration step (no SHAKE on QM-treated molecule) and with periodic boundary condition. In production MD run, samples were collected every 0.5 ps from 8 parallel walkers.

### B. Functional forms

In reference of the work by Zhang *et al.* (27), we set  $s = 0.82d_2 - 0.18d_1$  as a one dimensional CV for enhanced sampling.

So both  $F_\theta$  and  $V_\phi$  are functions of  $s$ , and they are expanded by Legendre polynomials up to the same order. For instance,  $F_\theta$  takes the following form,

$$F_\theta(s) = \sum_{i=1}^{i=N} \theta_i f_i(\bar{s})$$

where  $f_i$  denotes the  $i$ -th order Legendre polynomial,  $N = 50$ , and  $\bar{s}$  linearly rescales  $s$  to the range of  $[0,1]$ , with  $s_{\min} = 0$  and  $s_{\max} = 0.35$  nm,

$$\bar{s} = \left( s - \frac{s_{\min} + s_{\max}}{2} \right) / \left( \frac{s_{\max} - s_{\min}}{2} \right)$$

### C. Target distribution

In order to enhance the chemical transition, we optimized  $V_\phi$  according to a Lorentzian-coupled target distribution (Eqs. S13 and S14), where  $p_{WT}(s; \gamma)$  in Eq. (S13) takes a WT factor  $1/\gamma = 0.2$  and  $\zeta = 500$ .

### D. Training details

In each RE-VIFE iteration, we ran MD simulations for 60 ps to collect samples as  $p_{FG}$ . Before every training step of  $V_\phi$ ,  $F_\theta$  was first trained for 5 steps (i.e.,  $m = 5$  in Algorithm S1). Similarly to Ala2, we also adopted the Averaged SGD algorithm as for each optimization step, and the update step-size for  $V_\phi$  and  $F_\theta$  is 0.5. In total, RE-VIFE was performed for 100 iterations, which is equivalent to 6-ns MD simulation.

### E. Characterizing the chirality

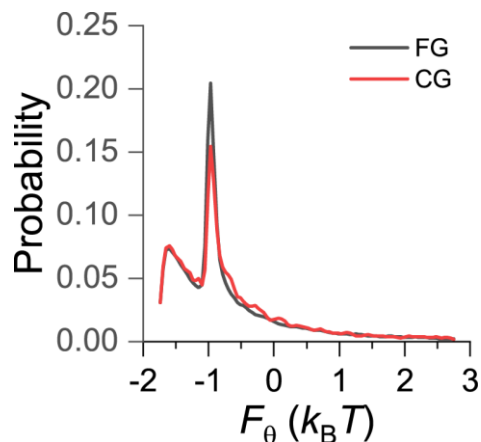
We define the chirality order parameter  $q_C$  as follows (28). First we pick the position of central carbon (corresponding to the chirality site) as  $\mathbf{R}_C$ , and order the attached four atoms from  $\mathbf{R}_0$  to  $\mathbf{R}_3$  according to conventions (Fig. S3). We then define several geometric vectors (Fig. S3):  $\mathbf{v}_i = \mathbf{R}_i - \mathbf{R}_C$ . In order to judge the chirality of the central carbon atom, we first calculate the cross product of two in-plane vectors,

$$\mathbf{u} = (\mathbf{v}_2 - \mathbf{v}_3) \times (\mathbf{v}_1 - \mathbf{v}_2)$$

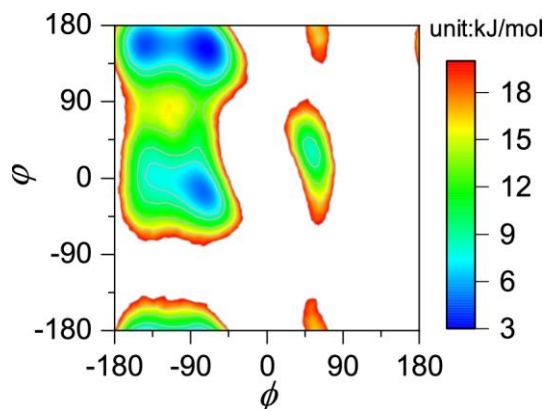
Finally, the chirality order parameter  $q_C$  is defined as the cosine of the angle between  $\mathbf{u}$  and the pseudo-normal vector  $\mathbf{v}_0$ , e vectors,

$$q_C = \frac{\mathbf{u} \cdot \mathbf{v}_0}{\|\mathbf{u}\| \|\mathbf{v}_0\|}$$

### PART III. SUPPLEMENTAL FIGURES

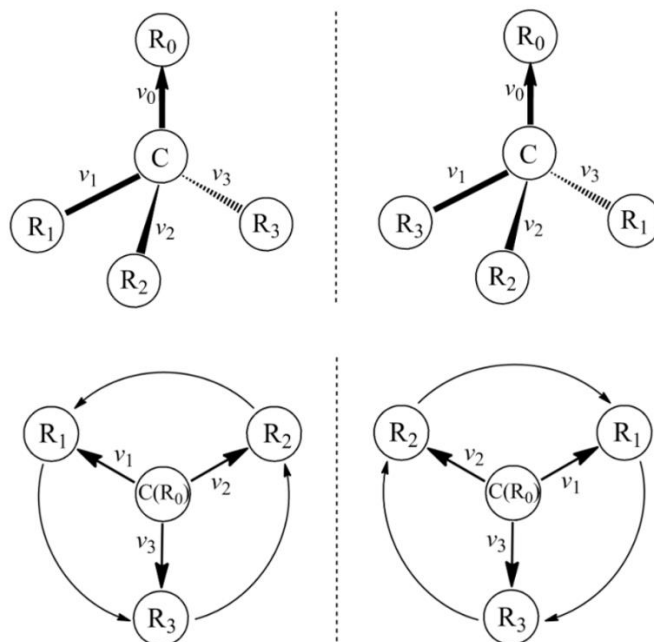


**Figure S1.** Distributions of  $F_\theta$  for real samples (termed as FG) drawn from Langevin dynamics simulation on the numerical model potential (black), and for generated samples (termed as CG) drawn from the inferred  $F_\theta$  (red).



**Figure S2.** Reference FES for torsional angles  $(\phi, \varphi)$  of Ala2. The reference FES is obtained via kernel density estimation over 750 ns MD simulation of Ala2 in explicit water. The data is accessible to the public (<https://github.com/markovmodel/mdshare>).





**Figure S3.** The chiral carbon atoms in the product. We order the attached atoms to the central carbon by the ordering  $R_3 > R_2 > R_1 > R_0$ . Upper panels: Illustration of the ordering of the central chiral carbon atom and the geometric vectors. Lower panels: Clockwise sequence of the attached groups corresponds to R-configuration, and a counter-clockwise sequence to S-configuration. This figure is adapted from Ref. (28) with permission.

## REFERENCES

1. Arjovsky M, Chintala S, & Bottou L (2017) Wasserstein generative adversarial networks. *International conference on machine learning*, pp 214-223.
2. Zhang J, Yang YI, & Noé F (2019) Targeted Adversarial Learning Optimized Sampling.
3. Valsson O & Parrinello M (2014) Variational approach to enhanced sampling and free energy calculations. *Physical review letters* 113(9):090601.
4. Schneider E, Dai L, Topper RQ, Drechsel-Grau C, & Tuckerman ME (2017) Stochastic neural network approach for learning high-dimensional free energy surfaces. *Physical review letters* 119(15):150601.
5. Vaswani A, *et al.* (2017) Attention is all you need. *Advances in Neural Information Processing Systems*, pp 5998-6008.
6. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, & Dahl GE (2017) Neural message passing for quantum chemistry. *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, (JMLR.org), pp 1263-1272.
7. Schütt KT, Sauceda HE, Kindermans P-J, Tkatchenko A, & Müller K-R (2018) SchNet—A deep learning architecture for molecules and materials. *The Journal of Chemical Physics* 148(24):241722.
8. Unke OT & Meuwly M (2019) PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *Journal of Chemical Theory and Computation* 15(6):3678-3693.
9. Miyato T, Kataoka T, Koyama M, & Yoshida Y (2018) Spectral Normalization for Generative Adversarial Networks. *International Conference on Learning Representations*.
10. Ioffe S & Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.
11. Barducci A, Bussi G, & Parrinello M (2008) Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical review letters* 100(2):020603.
12. Debnath J, Invernizzi M, & Parrinello M (2019) Enhanced sampling of transition states. *Journal of chemical theory and computation* 15(4):2454-2459.
13. Laio A & Parrinello M (2002) Escaping free-energy minima. *Proceedings of the National Academy of Sciences* 99(20):12562-12566.
14. Tiwary P & Berne B (2017) Predicting reaction coordinates in energy landscapes with diffusion anisotropy. *The Journal of chemical physics* 147(15):152701.
15. Lindorff-Larsen K, Piana S, Dror RO, & Shaw DE (2011) How fast-folding proteins fold. *Science* 334(6055):517-520.
16. Du Y & Mordatch I (2019) Implicit generation and generalization in energy-based models. *arXiv preprint arXiv:1903.08689*.
17. Berendsen H, Grigera J, & Straatsma T (1987) The missing term in effective pair potentials. *Journal of Physical Chemistry* 91(24):6269-6271.
18. Case D, *et al.* (AMBER 2017, 2017. *San Francisco: University of California*.
19. Hornak V, *et al.* (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* 65(3):712-725.
20. Ryckaert J-P, Ciccotti G, & Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* 23(3):327-341.
21. Berendsen HJC, Postma JPM, Gunsteren WFv, DiNola A, & Haak JR (1984) Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* 81(8):3684-3690.
22. Raiteri P, Laio A, Gervasio FL, Micheletti C, & Parrinello M (2006) Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *The journal of physical chemistry B* 110(8):3533-3539.
23. Bach F & Moulines E (2013) Non-strongly-convex smooth stochastic approximation with convergence rate  $O(1/n)$ . *neural information processing systems*, pp 773-781.
24. Elstner M, *et al.* (1998) Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* 58:7260-7268.
25. Sieffert N & Wipff G (2006) The [BMI][Tf2N] Ionic Liquid/Water Binary System: A Molecular Dynamics Study of Phase Separation and of the Liquid-Liquid Interface. *The Journal of Physical Chemistry B* 110(26):13076-13085.
26. Fu J, *et al.* (2016) Structural Characteristics of Homogeneous Hydrophobic Ionic Liquid-HNO<sub>3</sub>-H<sub>2</sub>O Ternary System: Experimental Studies and Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* 120(23):5194-5202.
27. Zhang J, *et al.* (2017) Rich Dynamics Underlying Solution Reactions Revealed by Sampling and Data Mining of Reactive Trajectories. *ACS Central Science* 3(5):407-414.
28. Zhang J, Yang YI, Yang L, & Gao YQ (2015) Dynamics and Kinetics Study of “In-Water” Chemical Reactions by Enhanced Sampling of Reactive Trajectories. *The Journal of Physical Chemistry B* 119(45):14505-14514.