Minimum-Excess-Work Guidance: Score-Based Sampling with Experimental Data or Sparse Restraints

Citation for the original published paper (version of record):
Kara, C., Höppe, T., Angelis, E. et al (2026). Minimum-Excess-Work Guidance: Score-Based Sampling with Experimental Data or Sparse Restraints. Journal of Chemical Theory and Computation, 22(11): 5838-5848. http://dx.doi.org/10.1021/acs.jctc.6c00080

JCTC Journal of Chemical Theory and Computation

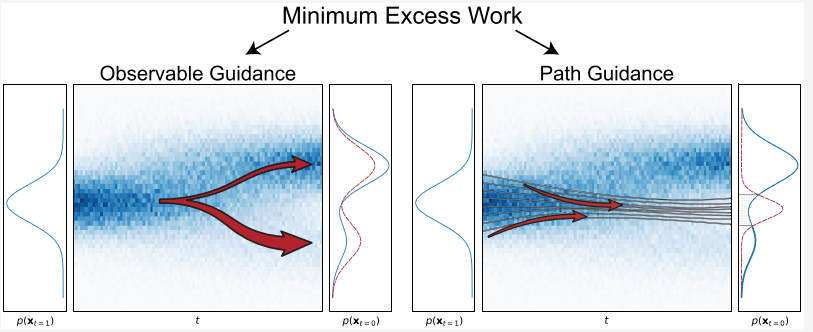
pubs.acs.org/JCTC Article

Minimum-Excess-Work Guidance: Score-Based Sampling with Experimental Data or Sparse Restraints

Christopher Kolloff,[#] Tobias Höppe,[#] Emmanouil Angelis,[#] Mathias Jacob Schreiner, Stefan Bauer, Andrea Dittadi, and Simon Olsson*

Cite This: https://doi.org/10.1021/acs.jctc.6c00080 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Surrogate models, such as Boltzmann generators (BGs) and emulators (BEs), based on deep generative models are becoming an important tool in molecular simulation. Often, we may want to use additional external information such as sparse experimental data to refine these models. However, there is no unique way to achieve this goal. Here, we propose a method inspired by thermodynamic work from statistical mechanics to regularize the guidance of pretrained probability flow generative models (e.g., continuous normalizing flows or diffusion models) to match additional sparse information. The regularization ensures that the *excess work* of the guidance procedure is minimized. We developed two guiding strategies based on this method: Path Guidance, which facilitates sampling of rare transition states by concentrating probability mass on user-defined subsets, and Observable Guidance, which aligns generated distributions with experimental observables while preserving entropy. We demonstrate the framework's versatility on two coarse-grained Boltzmann emulators, showcasing its ability to sample transition configurations and to correct systematic biases using experimental data on a variety of model protein systems. Finally, we provide bounds on the distributional differences between the guided and unguided distributions. The method bridges thermodynamic principles with modern generative architectures, offering a principled, efficient, and physics-inspired alternative to standard fine-tuning in data-scarce domains. Our results highlight improved sample efficiency and bias reduction, underscoring their applicability to molecular simulations and beyond.

INTRODUCTION

Probability flow generative models, such as normalizing flows[1−4] and diffusion models,[5,6] enable the modeling of complex, high-dimensional statistical distributions. These methods, are increasingly being used to build surrogate models for molecular dynamics such as Boltzmann Generators (BG),[7] Boltzmann Emulators (BE),[8−10] and implicit transfer operators (ITO),[11−13] collectively referred to as Generative MD.[14] These models generate samples through the integration of differential equations evolving a tractable distribution, e.g., a high-dimensional Gaussian, in a latent space, into an approximation of the data distribution. Although these models excel at general distribution learning, many scientific applications require precise control over generated samples to meet sparse observational constraints (e.g., limited transition-state configurations or partial density constraints from experiments). Current guidance methods struggle in data-scarce regimes as they typically rely on either specialized training or abundant reward signals. Existing approaches often involve fine-tuning,[15−17] incorporating conditional information during training,[18,19] or training an additional noise-aware discriminative model.[5]

While effective, these methods may be impractical in sparse-data regimes common in molecular and scientific applications. When the bias is large, the effective sample size of the reweighted distribution collapses, so any model retrained on

Received: January 16, 2026
Revised: April 12, 2026
Accepted: May 4, 2026

https://doi.org/10.1021/acs.jctc.6c00080
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

the reweighted data overfits to a small number of high-weight configurations rather than the full shifted distribution. For transferable foundation models such as BioEmu, principled fine-tuning would require evaluating a KL regularizer between the pretrained and fine-tuned models at each update step—computationally prohibitive at scale and potentially causing catastrophic forgetting of the base model's generalization. This motivates a new approach in which we apply minimal perturbations to the trained model at inference time, enabling controlled generation under very sparse constraints without modifying any model parameters. Inspired by statistical mechanics, we introduce an approach for regularizing guidance of probability flow generative models based on the principle of minimum excess work (MEW). In this context, "work" is a measure of the physical effort, e.g., energy, needed to transform a system from one macrostate to another, where a macrostate is characterized by a probability density function. MEW thereby acts as a natural, physics-inspired regularization scheme for guiding generative models. We develop the theoretical framework for MEW-based regularization of generative models, explicitly connecting it to optimal transport theory, and validate its effectiveness through extensive benchmarks across multiple scales and systems. In addition to introducing the MEW framework, we propose a simple yet effective form of path guidance tailored to sparse sampling problems. We specialize MEW guidance to two common challenges in molecular simulation. First, **Observable Guidance**: a bias-correction method that matches experimental observables while preserving the entropy of the reference ensemble via a minimum-excess-work regularizer. We validate this approach on a toy system and two coarse-grained protein Boltzmann emulators. With this approach, we thus correct the systematic bias in the base model and are therefore able to improve the prediction of unmeasured observables since they report on the same thermodynamics. Second, **Transition-State Sampling**: a path guidance-based sampling strategy that concentrates samples on user-specified regions, e.g., the low-probability transition region between states, which we evaluate on the coarse-grained Boltzmann emulator.[1]

BACKGROUND AND THEORY

Diffusion models learn a stochastic process that maps a simple prior distribution p_1 to an approximation p_0 of the data distribution q_0. This is typically done by reversing a known noising process governed by an Ornstein−Uhlenbeck SDE, $\mathrm{d}\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, t)\,\mathrm{d}t + g(t)\,\mathrm{d}\mathbf{w}_t$ with $\mathbf{f}(\mathbf{x}_t, t)$ linear in $\mathbf{x}_t$. This process induces a family of marginals q_t with simple forward transitions $q_t(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \alpha_t\mathbf{x}_0, \sigma_t^2\mathbf{I})$, where α_t and σ_t are determined by the SDE coefficients and $\mathbf{x}_0 \sim q_0(\mathbf{x})$ is drawn from the data distribution. Given q_1 and the score $\nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t)$, one can sample from q_0 via the time reversal:[20]

$$\mathrm{d}\mathbf{x}_t = [\mathbf{f}(\mathbf{x}_t, t) - g(t)^2\nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t)]\,\mathrm{d}t + g(t)\,\mathrm{d}\tilde{\mathbf{w}}_t, \quad \mathbf{x}_1 \sim q_1 \tag{1}$$

where $\tilde{\mathbf{w}}_t$ is a reverse-time Wiener process or via the *probability flow ODE*:[5,21]

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}_t, t) - \frac{1}{2}g(t)^2\nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t), \quad \mathbf{x}_1 \sim q_1 \tag{2}$$

both having the same time-marginals q_t as the forward process. In practice, the score is approximated by a score model $\mathrm{s}_\theta(\mathbf{x}_t, t)$, and a simple distribution $p_1 \approx q_1$ is used as initial distribution at $t = 1$:

$$\mathrm{d}\mathbf{x}_t = [\mathbf{f}(\mathbf{x}_t, t) - g(t)^2\mathrm{s}_\theta(\mathbf{x}_t, t)]\,\mathrm{d}t + g(t)\,\mathrm{d}\tilde{\mathbf{w}}_t, \quad \mathbf{x}_1 \sim p_1 \tag{3}$$

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}_t, t) - \frac{1}{2}g(t)^2\mathrm{s}_\theta(\mathbf{x}_t, t), \quad \mathbf{x}_1 \sim p_1 \tag{4}$$

We denote by $\{p_t\}_{t\in[0,1]}$ the probability path induced by eqs 3 or 4.

Equilibrium Sampling of the Boltzmann Distribution

A key challenge in statistical mechanics is to generate independent samples from the Boltzmann distribution

$$\mathbf{x} \sim p(\mathbf{x}) \propto \exp[-\beta U(\mathbf{x})] \tag{5}$$

where $\beta = (k_\mathrm{B}T)^{-1}$ is the inverse temperature and $U(\mathbf{x})$ is the potential energy of a configuration $\mathbf{x} \in \Omega \subseteq \mathbb{R}^d$. This distribution underlies the estimation of macroscopic observables, such as $\mathbb{E}_{p(\mathbf{x})}[O_i(\mathbf{x})]$, which allow for a direct comparison to experimental data. However, sampling from $p(\mathbf{x})$ is notoriously difficult due to the rugged energy landscape $U(\mathbf{x})$. Traditional methods such as Molecular Dynamics (MD) or Markov Chain Monte Carlo (MCMC) suffer from slow mixing and generate highly correlated samples that often fail to cross energy barriers between metastable states. This leads to biased estimates and poor coverage of transition configurations, i.e., regions in state space that are severely undersampled but mechanistically crucial. Recent work on Boltzmann Generators[7,22−26] addressed these challenges by learning direct mappings from simple priors to Boltzmann-like distributions. However, because these models are trained on data generated under a fixed potential energy function $U(\mathbf{x})$, any inaccuracy in U propagates into the learned distribution.[27,28] When experimental measurements are available that reveal such discrepancies, the question arises how to correct the generative model *post hoc* using sparse observational constraints. Separately, for applications such as transition-path sampling or mechanistic studies of folding, one often requires enhanced sampling of transition regions that are rare under the Boltzmann distribution. This is not due to the model being wrong but because the target application demands a different sampling distribution. In this work, we address both challenges by guiding a generative model using sparse experimental or structural information, leveraging a coarse-grained Boltzmann emulator inspired by Arts et al.[29] and show how our method can be integrated into a state-of-the-art Boltzmann emulator[10] to sample protein ensembles consistent with experimental data.

Maximum Entropy Reweighting is a broadly adopted technique to overcome force-field inaccuracies in potential energy models.[30−38] The result of this optimization is a tilted distribution which depends on a set of Lagrange multipliers, $\{\lambda_i\}$, each corresponding to an experimental observable of interest. The solution $p'(\mathbf{x}) \propto p(\mathbf{x})\,\exp(-\sum_{i=1}^{M}\lambda_i O_i(\mathbf{x}))$ minimizes the KL divergence from the reference distribution $p(\mathbf{x})$, subject to the constraints $\mathbb{E}_{p'(\mathbf{x})}[O_i(\mathbf{x})] = o_i$. A detailed derivation is provided in Supporting Information A.1 for the reader's convenience. However, this approach has two limitations in the generative setting. First, it operates on a fixed set of samples $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^M$. When the bias in $p(\mathbf{x})$ is substantial, the effective sample size collapses and the tilted distribution $p'(\mathbf{x})$ is represented by only a small fraction of

https://doi.org/10.1021/acs.jctc.6c00080
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

high-weight configurations. Second, one cannot simply retrain a new generative model on this tilted distribution to recover $p'(\mathbf{x})$. The retrained model would overfit to those same high-weight samples rather than learn the full shifted distribution, thus inheriting the degeneracy problem. These considerations motivate applying the maximum entropy principle directly within the generative process rather than as a *post hoc* or retraining step.

Loss Guidance is the process of adjusting the diffusion process to satisfy a target condition $\mathbf{y}$ without fine-tuning and has been explored in several prior works.[39−41] To sample from the conditional distribution $p(\mathbf{x}_0|\mathbf{y})$ *post hoc*, we can use the following identity: $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{y}) = \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)$. Obtaining $\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)$ typically requires training a separate model on the noisy states $\mathbf{x}_t$, as done in classifier guidance.[5] Alternatively, the posterior mean $\hat{\mathbf{x}}_t(\mathbf{x}_t) := \mathbb{E}_{p(\mathbf{x}_0|\mathbf{x}_t)}[\mathbf{x}_0]$ can be used as an estimate of the clean data $\mathbf{x}_0$. Using Tweedie's formula, the posterior mean can be expressed as $\mathbb{E}_{p(\mathbf{x}_0|\mathbf{x}_t)}[\mathbf{x}_0] = \frac{1}{\alpha_t}[\mathbf{x}_t + \sigma_t^2 \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)]$. This allows us to approximate the likelihood in data space via $\log p(\mathbf{y}|\hat{\mathbf{x}}_t(\mathbf{x}_t)) \approx l(\hat{\mathbf{x}}_t(\mathbf{x}_t), \mathbf{y})$, where l denotes a suitable differentiable loss function (e.g., cross-entropy or log-likelihood under a differentiable model). The gradient $\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\hat{\mathbf{x}}_t(\mathbf{x}_t))$ can then be computed by backpropagation. In practice, the mean is approximated using the score model $s_\theta(\mathbf{x}_t, t)$, allowing the score estimate to be updated as $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{y}) \approx s_\theta(\mathbf{x}_t, t) + \eta_t \nabla_{\mathbf{x}_t} l(\hat{\mathbf{x}}_t(\mathbf{x}_t), \mathbf{y})$ with η_t being a guiding strength function. We note that this procedure is a heurisitc approximation to exact posterior sampling. Replacing the posterior by a delta function incurs an approximation error that is severe at large t where the posterior is broad.[40,41] This limitation motivates our path guidance formulation, which avoids backpropagation through the posterior mean.

Work and Optimal Transport

In statistical mechanics, thermodynamic work W is the energy required to transform a system from a probabilistic state p to another p'. For a continuum system:

$$W = \iint \mathbf{J}(\mathbf{x}, t)\cdot\mathbf{F}(\mathbf{x}, t)\, \mathrm{d}\mathbf{x}\, \mathrm{d}t \tag{6}$$

where $\mathbf{J}(\mathbf{x}, t) = \mathbf{v}(\mathbf{x}, t)p_t(\mathbf{x})$ is the probability flux and $\mathbf{F}(\mathbf{x}, t)$ is the force applied to the system. This generalizes the classical work expression $W = \int F(\mathbf{x})\mathrm{d}\mathbf{x}$.[42] When the force and velocity field coincide (i.e., the Jacobian of the push-forward map associated with the velocity field is a diffeomorphism), they can be expressed as spatial gradients of a potential $u(\mathbf{x}, t)$.[43] Under these conditions, W becomes equivalent to the kinetic energy in the Benamou−Brenier formulation of optimal transport[44] and provides an upper bound on the squared 2-Wasserstein distance between the distributions:

$$W_2^2(p, p') \le \iint \|\mathbf{v}(\mathbf{x}, t)\|^2 p_t(\mathbf{x})\, \mathrm{d}\mathbf{x}\, \mathrm{d}t = W \tag{7}$$

where $\mathbf{v}$ and p satisfy $\frac{\partial}{\partial t} p_t(\mathbf{x}) = -\nabla_{\mathbf{x}}\cdot[p_t(\mathbf{x})\mathbf{v}(\mathbf{x}, t)]$. Minimizing W (the kinetic energy) yields the optimal transport map that transforms p into p' along the path requiring minimal energy; by eq 7, this simultaneously tightens the upper bound on $W_2^2(p, p')$. The idea of identifying probability paths minimizing the kinetic energy, or more generally a Lagrangian, has recently been applied to improve the efficiency of probability flow generative models.[45−52]

MINIMUM-EXCESS-WORK GUIDANCE

During the generative process, we transform a simple distribution $p_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ into a complex data distribution p_0 with support $\Omega \subseteq \mathbb{R}^d$ by solving the reverse-time SDE (3) or the ODE (4). To incorporate additional constraints and align the generative process with new information, we modify the drift of eqs 3 and 4 by introducing an additive perturbation to the score model:

$$\mathrm{d}\mathbf{x}_t = (\mathbf{f}(\mathbf{x}_t, t) - g(t)^2[s_\theta(\mathbf{x}_t, t) + \mathbf{h}_\nu(\mathbf{x}_t, t)])\, \mathrm{d}t + g(t)\, \mathrm{d}\tilde{\mathbf{w}}_t, \quad \mathbf{x}_1 \sim p_1 \tag{8}$$

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}_t, t) - \frac{1}{2} g(t)^2[s_\theta(\mathbf{x}_t, t) + \mathbf{h}_\nu(\mathbf{x}_t, t)], \quad \mathbf{x}_1 \sim p_1 \tag{9}$$

where $\mathbf{h}_\nu$: $\mathbb{R}^d \times [0, 1] \to \mathbb{R}^d$ is a time-dependent vector field.

The aim of MEW guidance is to satisfy a guidance objective for the guided distribution $p_0' \neq p_0$, while minimizing the *excess work* associated with $\mathbf{h}_\nu(\mathbf{x}_t, t)$, required to modify the probability density, p_0. We define the excess work in the context of an unperturbed and perturbed system described by the following ODEs over $t \in [0, 1]$ with $p_1 = p_1'$:

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = \mathbf{v}(\mathbf{x}_t, t), \qquad \frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = \mathbf{v}(\mathbf{x}_t, t) + \mathbf{u}(\mathbf{x}_t, t) \tag{10}$$

with the respective time-marginal densities p_t, p_t'. Loosely following eq 7, we define the excess work as $\Delta W := \iint \|\mathbf{u}(\mathbf{x}, t)\|^2 p_t'(\mathbf{x})\, \mathrm{d}\mathbf{x}\, \mathrm{d}t$. For the ODEs (4) and (9), the perturbation velocity is $\mathbf{u}(\mathbf{x}, t) = -\frac{1}{2} g(t)^2 \mathbf{h}_\nu(\mathbf{x}, t)$, so that $\|\mathbf{u}\|^2 = \frac{g(t)^4}{4} \|\mathbf{h}_\nu\|^2$, and the excess work becomes

$$\Delta W(\vartheta) = \iint \frac{g(t)^4}{4} \|\mathbf{h}_\nu(\mathbf{x}, t)\|^2 p_t'(\mathbf{x})\, \mathrm{d}\mathbf{x}\, \mathrm{d}t \tag{11}$$

To justify our choice of excess work as a regularizer, it is helpful to understand how perturbations affect the generated distribution. In particular, we would like p_0' to remain close to the reference base distribution p_0. While stability bounds of this type have appeared in the literature on ODEs and SDEs, we restate tailored versions here for completeness, with proofs in Supporting Information A.2 and A.3.

Proposition 1: *Let pt and p_t' be the distributions at time t obtained by solving the ODEs (4) and (9) backward in time from the same initial distribution p_1 at $t = 1$. Assume that the vector fields are measurable in time and L_t-Lipschitz in space with L_t integrable. Then,*

$$W_2^2(p_0, p_0') \le \int_0^1 w_W(t) \frac{g t^4}{4} \mathbb{E}_{\mathbf{x}\sim p_t'}[\|\mathbf{h}_\vartheta(\mathrm{x}, t)\|^2]\, \mathrm{d}t, \qquad w_W(t) := \mathrm{e}^{t + 2\int_0^t L_s\, \mathrm{d}s} \tag{12}$$

The weight $w_W(t)$ arises from a Grönwall argument on $\|\mathbf{x}_t - \mathbf{x}_t'\|$. The factor e^t comes from decoupling perturbation and displacement via Young's inequality, and $\mathrm{e}^{2\int_0^t L_s\, \mathrm{d}s}$ from the Lipschitz continuity of the base vector field.

Proposition 2: *Let p_t and p_t' be the distributions at time t induced by the reverse-time SDEs (3) and (8) starting from the*

https://doi.org/10.1021/acs.jctc.6c00080
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

same distribution p_1 at $t = 1$. Assume that both SDEs admit strong solutions and that $\mathbb{P}' \ll \mathbb{P}$, where $\mathbb{P}$, $\mathbb{P}'$ are the path measures induced by the SDEs on $C([0, 1], \mathbb{R}^d)$. Then,

$$D_{\mathrm{KL}}(p_0'||p_0) \le \int_0^1 w_{\mathrm{KL}}(t)\frac{gt^4}{4}\mathbb{E}_{x\sim p_t'}[||\mathbf{h}_\vartheta(x, t)||^2]\,\mathrm{d}t,$$
$$w_{\mathrm{KL}}(t) := \frac{2}{g(t)^2} \tag{13}$$

The weight $w_{\mathrm{KL}}(t)$ follows from Girsanov's theorem. The KL between path measures introduces a factor $\frac{1}{g(t)^2}$, which, after substituting the drift difference, yields $w_{\mathrm{KL}}(t) = \frac{2}{g(t)^2}$.

Since both bounds—for the KL divergence and the Wasserstein distance—are time-reweighted versions of the excess work ΔW (11), it serves as a natural choice of regularizer for guidance objectives.

We then optimize the parameters ϑ of the perturbation $\mathbf{h}_\nu$ by minimizing the following:

$$\mathcal{L}(\vartheta) = \mathcal{L}_1(\vartheta) + \gamma\Delta W(\vartheta) \tag{14}$$

where $\mathcal{L}_1(\vartheta)$ is a guidance objective and γ controls the regularization strength.

We now explore how this minimum-excess-work principle is applied in the two settings: (1) guidance based on expectations of observables and (2) targeted guidance toward a user-defined subspace.

Observable Guidance

In this section, we guide a diffusion model to align with data that reflects an expectation, using the MEW approach. Using a set of Lagrange multipliers $\Lambda = \{\lambda_1, ..., \lambda_M\}$ pre-estimated using, e.g., the algorithm outlined in Bottaro et al.,[53] we dynamically adjust the score by estimating an augmentation factor $\mathbf{h}_\vartheta$ that ensures $|\mathbb{E}_{p'(\mathbf{x})}[O_i(\mathbf{x})] - o_i| \le \varepsilon$. It is noted that traditional reweighting techniques typically apply bias *post hoc*, whereas our method adapts the *generative* process. We express the guidance factor as,

$$\mathbf{h}_\nu(\mathbf{x}_t, t) = -\eta_t(\vartheta)\sum_{i=1}^{M}\lambda_i\nabla_{\mathbf{x}_t}O_i(\hat{\mathbf{x}}_t(\mathbf{x}_t)) \tag{15}$$

In the same way that a score model $s_\theta(\mathbf{x}_t, t)$ approximates the gradient of the log probability, $\mathbf{h}_\nu(\mathbf{x}_t, t)$ represents the gradient of the observable function with respect to the latent variable $\mathbf{x}_t$. The coefficients λ_i steers the flow toward (or away from) directions favored by the experimental observable expectation, thus "adjusting" the score of the original model. The expression in eq 15 thus reflects the maximum entropy principle applied in a generative setting. Its amplitude is modulated by $\eta_t(\vartheta) = \eta_{\mathrm{init}}\exp(-\kappa(1-t))$, and our optimization strategy consists of learning the parameters $\vartheta = \{\eta_{\mathrm{init}}, \kappa\}$ of this scalar function. It is noted that we use the mean posterior estimation $\hat{\mathbf{x}}_t(\mathbf{x}_t)$ discussed in the background section, instead of using $\mathbf{x}_t$ directly.[39,40] Our optimization objective is 2-fold: we aim to reduce the discrepancies between the model predictions and experimental data while minimizing the excess work exerted by the augmentation. The former is a supervised loss defined as

$$\mathcal{L}_1(\vartheta) = \frac{1}{M}\sum_{i=1}^{M}(o_i^{\mathrm{exp}} - \mathbb{E}_{\mathbf{x}\sim p_0'}[O_i(\mathbf{x})])^2 \tag{16}$$

where o_i^{exp} denotes the experimental values and $\mathbb{E}_{\mathbf{x}\sim p_0'}[O_i(\mathbf{x})]$ denotes the expected values under the adjusted distribution p_0'. To balance accuracy with the principle of maximum entropy, we introduce a regularization term based on minimizing the excess work ΔW. Substituting the specific form of $\mathbf{h}_\nu(\mathbf{x}_t, t)$ from eq 15 into eq 11, we obtain

$$\Delta W(\vartheta) = \int_0^1 \frac{g(t)^4}{4}\left|\eta_t(\vartheta)|^2\mathbb{E}_{\mathbf{x}\sim p_t'}\left[\left|\left|\sum_{i=1}^{M}\lambda_i\nabla_{\mathbf{x}}O_i(\hat{\mathbf{x}}_t(\mathbf{x}))\right|\right|^2\right]\right.\mathrm{d}t \tag{17}$$

Path Guidance

In this setting, we assume access to a set of guiding samples $\mathcal{X}^g = \{\mathbf{x}^i\}_{i=1}^M$, each belonging to a target subset $A \subset \Omega$ of the sampling space. Assuming that A forms a coherent region rather than being scattered across distinct modes, we modify the score of the diffusion model to sample from a perturbed distribution p_0' that maximizes the probability mass placed in A. The objective does not prescribe a specific target density over A, but rather, the internal structure of p_0' within A is shaped implicitly by the MEW regularizer, which penalizes unnecessary deviations from the base model, effectively encouraging the guided distribution to approximate the conditional Boltzmann $\frac{\mathbb{I}_{\{\mathbf{x}\in A\}}p_0(\mathbf{x})}{Z_A}$. Training an emulator directly on A would be impractical in our setting, since we have access to only very few samples. Since $\mathcal{L}_1$ does not need to be differentiable, the objective can be formulated generally as

$$\mathcal{L}_1(\vartheta, \varphi) = 1 - \mathbb{E}_{\mathbf{x}\sim p_0'}[\mathbb{I}_{\{\mathbf{x}\in A\}}] \tag{18}$$

That is, $\mathcal{L}_1$ is the complement of the probability of drawing a sample within the target subset A under the guided distribution p_0'. Guiding the diffusion process toward the subset A can be done by taking advantage of the probability flow ODE (4), which holds the desirable property of providing unique *latent representations* of each data point, for any time step t. Starting from the guiding samples, we compute their trajectories by integrating eq 4 forward in time, obtaining the latent representations $\mathcal{X}_t^g = \{\mathbf{x}_t^i\}_{i=1}^M$ for time t. The set $\{\mathcal{X}_t^g\}_{t=0}^1$ defines a trajectory of latent representations that the model must follow to ensure its samples satisfy $\mathbf{x}' \in A$. Based on this trajectory, we can define the augmentation factor as

$$\mathbf{h}_{\nu,\varphi}(\mathbf{x}_t, t) := \eta_t(\vartheta)\nabla_{\mathbf{x}_t}\log\mathcal{K}_{h_t(\varphi)}(\mathbf{x}_t, \mathcal{X}_t^g) \tag{19}$$

with $\mathcal{K}_{h_t(\varphi)}(\mathbf{x}_t, \mathcal{X}_t^g) := \sum_{\mathbf{x}_t^i\in\mathcal{X}_t^g}\mathrm{K}_{h_t(\varphi)}(\mathbf{x}_t, \mathbf{x}_t^i)$, where K can be any differentiable kernel with time-dependent bandwidth $h_t(\varphi)$. By updating the score function using eq 19, we align the sampling trajectory with that of the guiding points, while regularizing the guidance strength via the same excess work penalty as in eq 17, now evaluated using the time-dependent KDE score $\mathbb{E}_{\mathbf{x}\sim p_t'}[\left|\left|\nabla_{\mathbf{x}}\log\mathcal{K}_{h_t(\varphi)}(\mathbf{x}, \mathcal{X}_t^g)\right|\right|^2]$. In practice, both $\eta_t(\vartheta)$ and $h_t(\varphi)$ are implemented as sigmoid functions with learnable parameters $\vartheta = (\vartheta_{\mathrm{init}}, \vartheta_g, \vartheta_s)$ and $\varphi = (\varphi_{\mathrm{init}}, \varphi_g, \varphi_s)$ (see Supporting Information B.4) and optimized for eq 14 using Bayesian optimization with Gaussian processes. The use of sigmoids allows the guidance to be stronger early in the trajectory, when $\mathbf{x}_t$ is close to the Gaussian prior, and the kernel signal is more stable, and weaker near $t = 0$, where the data

https://doi.org/10.1021/acs.jctc.6c00080
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

distribution is more complex and direct guidance is less reliable.

RESULTS

We now demonstrate the application of MEW guidance across several experimental setups. We first evaluate path and observable guidance on two toy setups and then proceed to showcase our approach on a coarse-grained protein Boltzmann Emulator.

Observable Guidance

Synthetic Data. We chose a fully controlled synthetic system to test MEW guidance. We set up a biased 1D quadruple-well diffusion model with an accessible ground-truth Boltzmann distribution[54] using only the expectation of a known observable (a four-component GMM) and injecting the corresponding Lagrange multiplier via eq 8 following Bottaro et al.[53] This simple test system displays two closely related properties in molecular dynamics: multimodality and metastability (the former being the statistical signature of the latter), while keeping the corresponding Boltzmann distribution is numerically accessible, allowing us to directly gauge our methods' ability to recover the unbiased distribution and to unambiguously test whether guidance alone corrects distributional bias: we observe a 10-fold reduction in $\mathrm{KL}(p_{\mathrm{GT}}\|p_{\mathrm{M}})$, where p_{GT} is the unbiased ground-truth density and p_{M} is either the biased reference or the guided model, (from 0.13 to 0.019 ± 0.002; see Supporting Information Table 2 for details) while matching the observable. We note that in this controlled toy setting, where densities are available analytically and ESS collapse is not severe, retraining a model directly on $p'(\mathbf{x})$ would likely yield comparable or better results. The decisive advantages of guidance over retraining arise in realistic scenarios (large systems, transferable foundation models, ESS collapse, and data scarcity) as discussed above. We also find MEW regularization is critical to prevent mode collapse and preserve distributional fidelity. See Supporting Information Figure 2 and Table 2 for overlays and metrics; ablation experiments are reported in Supporting Information Figure 10 and Table 3.

Coarse-Grained Protein Boltzmann Emulator (cgBE): Chignolin. To evaluate our method on a real-world task, we apply observable guidance to guide a pretrained cgBE to sample conformations of chignolin, a ten-residue mini-protein that serves as a standard benchmark in protein folding studies.[55−57] The base model is a coarse-grained Boltzmann Emulator trained on MD simulation data of chignolin (architecture and training details in Supporting Information B.1 and B.3). All guidance is applied at inference time, and no model parameters are modified. Our task is to correct systematic biases in the equilibrium sampling using only experimental measurements while preserving physical validity. This is a challenging task given the high-dimensional structured space and unknown ground-truth distribution.

Experimental Setup. We use folding free energy $\Delta G = -k_{\mathrm{B}}T \log\left(\frac{p_{\mathrm{folded}}}{p_{\mathrm{unfolded}}}\right)$ as our observable, which captures the relative stability of different protein conformations. The reference model p_{MD} shows significant bias in this metric (−1.27 kcal/mol vs experimental value of −1.87 kcal/mol[55]), making it a suitable test case. Model architecture and training details are provided in Supporting Information B.3.

Evaluation. Our guided model achieves substantial improvements across several metrics (see Table 1) while maintaining physical validity, which we verified through the analysis of bond lengths and torsion angles (Figures S3 to S5, in the Supporting Information). The guided model's folding free energy (-1.82 ± 0.01 kcal/mol) closely matches the target experimental value (−1.87 kcal/mol), reducing mean squared error by an order of magnitude from 0.6 to 0.05 kcal/mol. Additionally, the KL divergence from the reference MD trajectory improves from 0.329 to 0.005 ± 0.002, demonstrating better conservation of the properties of the reference distribution, including multimodality and entropy. Figure 1 visualizes these improvements.

Table 1. Quantitative Metrics Evaluating the Guidance Process: Expected Observables and KL Divergence

model $\mathcal{M}$	$\mathbb{E}_{p_{\mathcal{M}}(\mathbf{x})}[O(\mathbf{x})]$ (kcal/mol)	KL $(p'_{\mathrm{MD}}\|p_{\mathcal{M}})$
experimental	−1.87	
reference	−1.27	0.329
guided	-1.82 ± 0.01	0.005 ± 0.002

Panels A and B demonstrate that guidance successfully increases the population of folded states ($d(C_1^{\alpha}, C_9^{\alpha}) < 7.5$ Å), consistent with experimental observations. Panel C shows 50 superimposed generated structures, highlighting both the diversity and physical validity of our samples.

BioEmu: Homeodomain. Finally, we showcase our approach on the fast-folding homeodomain EnHD HTH fragment (44 residues) using BioEmu,[10] a transferable coarse-grained Boltzmann Emulator pretrained on MD data across a large set of protein systems, and a canonical model extensively studied experimentally[58,59] and computationally.[10,60] As with the chignolin experiments, BioEmu serves as the unconditional base model and is held fixed throughout; guidance adjusts only the drift of the reverse-time SDE at the inference time.

Experimental Setup. We generate ensembles with BioEmu and compare expected ${}^3J_{\mathrm{HN\text{-}HA}}$ couplings to experiment. These couplings report on backbone dihedrals and, thus, thermodynamic populations. In our experiments, we use the 10 most informative observables (details in Supporting Information B.3). We quantified the agreement between experiments and computational predictions by the Q-factor[61] and noticed that the unguided model shows a clear discrepancy ($Q = 0.147$), motivating the use of our approach. Training details are in Supporting Information B.3.

Evaluation. MEW markedly improves the agreement while preserving physical plausibility (Figure 2). Using only 10 experimental expectations, Q drops from 0.147 (unguided) to 0.037 (MEW); post hoc reweighting achieves 0.031 but with moderate weight degeneracy (relative ESS = 0.255), meaning fewer than 26% of samples contribute effectively. MEW avoids this importance-weight collapse by updating the generative process directly. Retraining BioEmu on the reweighted samples would face the same degeneracy. The model would overfit to high-weight configurations rather than learn the full shifted distribution, and updating BioEmu's full parameter set would be computationally prohibitive without dedicated infrastructure. After guidance, points cluster near the identity within experimental uncertainty; the inset ensemble shows broad conformational coverage without mode collapse, and a representative observable (residue 38) shifts toward the

https://doi.org/10.1021/acs.jctc.6c00080
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

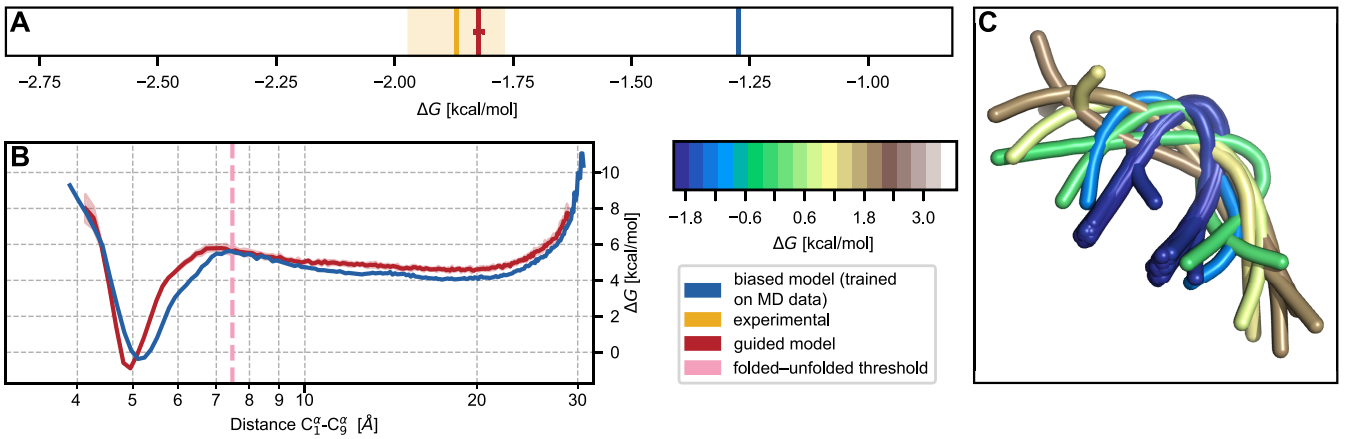


Figure 1. Observable guidance of chignolin. (A) Folding free energy comparison between reference model (blue), experimental data (yellow), and guided model (red). (B) Free energy profiles as a function of N- to C-terminal C^α distance. (C) Ensemble of 50 generated protein structures colored by their energy.

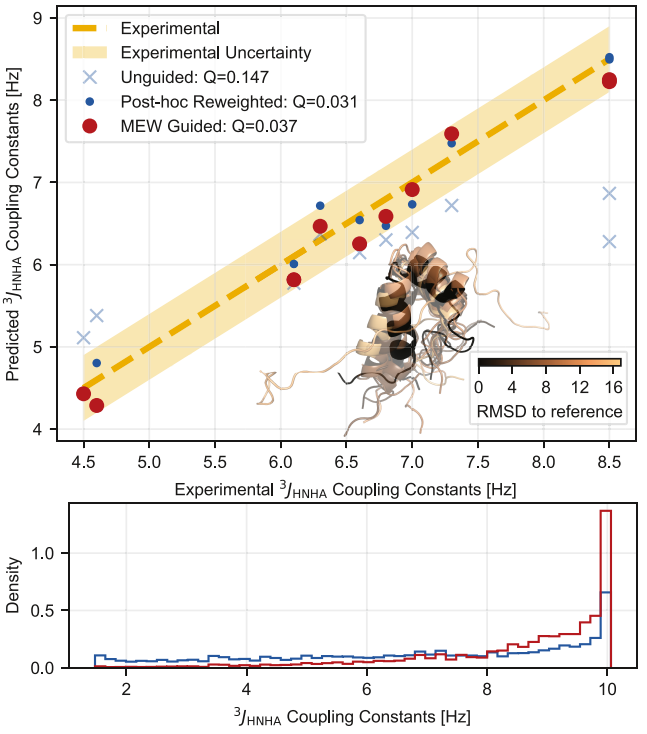


Figure 2. Observable guidance on EnHD. Top: 10 experimental ^3J vs unguided (light blue), MEW-guided (red), and post hoc reweighted predictions; inset: 10 sampled structures colored by RMSD. Bottom: representative histogram of residue 38 illustrating the guided population shift; blue (dashed) shows the unguided BioEmu predictions, red (solid) shows the MEW-guided predictions.

experimental region without variance loss. Full histograms and structural diagnostics are in Supporting Information Figures S8 and S9. Overall, MEW uses sparse 1D NMR readouts to make targeted, physically consistent adjustments to the sampling of EnHD using BioEmu.

These results demonstrate that guidance in the sparse-data regime with MEW regularization allows us to effectively align high-dimensional and highly structured generative models with experimental constraints without compromising local physical validity and maintaining global distributional properties such as multimodality.

Path Guidance

We now use the cgBE to evaluate path guidance with MEW regularization for upsampling high-energy transition configurations (states), which are critical for understanding the folding process of proteins. Due to their high energy (eq 5), these states account for only 1% of both the data and model distribution, making their successful upsampling a strong demonstration of our method's effectiveness. Consistent with our Observable Guidance section, we also use the chignolin mini-protein, to investigate the effectiveness of path guidance and will later show that we can scale our method to a 65-residue protein. To contextualize path guidance, we first introduced an alternative baseline.

Baseline. As a natural alternative to path guidance, we adapt loss guidance to our setting by using the log-likelihood of a KDE fitted on guiding points $\mathcal{X}_0^g = \{\mathbf{x}_0^i\}_{i=1}^M$. Specifically, we will change the perturbation kernel from eq 19 to $\mathcal{K}_{h_t(\varphi)}(\hat{\mathbf{x}}_t(\mathbf{x}_t), \mathcal{X}_0^g)$. While it appears similar to path guidance, the key difference lies in the space in which the KDE is computed. In path guidance, the kernel is applied along the trajectory $\{\mathcal{X}_t^g\}_{t=0}^{t=1}$, resulting in a distinct KDE for each time step t. In contrast, loss guidance computes the KDE in data space and estimates the likelihood with respect to the posterior mean $\hat{\mathbf{x}}_t(\mathbf{x}_t)$, which requires backpropagating through the model at every sampling step. Implementation details and ablation studies for two alternative baselines that do not augment the vector field are provided in Supporting Information B.4 and D.3. To explore the dynamics of both methods, we design a synthetic example to study the effect of different parameters (see Supporting Information D.1 for details).

Evaluation Criteria. We assessed the methods using three key metrics. First, we measure guiding success as the percentage of sampled transition configurations (see Supporting Information B.2 for details). Second, we evaluate the diversity among transition states using the Vendi score (VS)[62] to verify that our method generates novel samples rather than merely resampling the guiding data. Lastly, since we cannot evaluate the energy under the coarse-grained model, we instead ensure the physical validity of the generated samples under guidance by computing the Wasserstein distance (WD) between the bond-length distributions of generated and

https://doi.org/10.1021/acs.jctc.6c00080
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

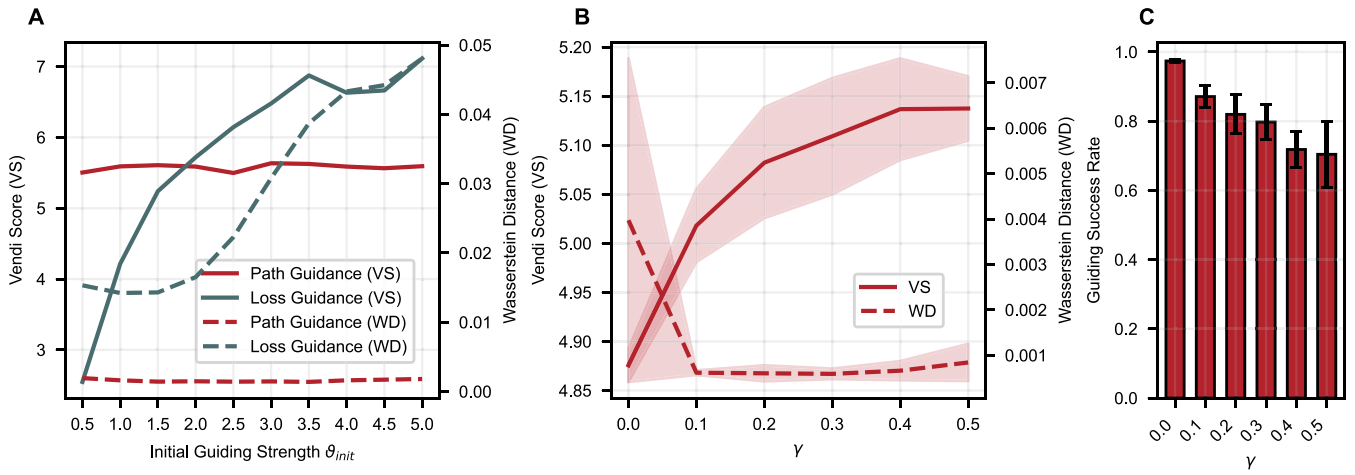


Figure 3. Path guidance vs loss guidance for sampling transition states. (A) Sample quality and diversity, measured by the Wasserstein Distance (WD) and Vendi score (VS), show that path guidance preserves diversity and quality even at high guiding strengths, whereas loss guidance deteriorates. (B) Without MEW regularization ($\gamma = 0$), sampled transition states tend to collapse and have no diversity (VS). Regularization also improves sample quality (WD). (C) Guiding success rate, measured as the percentage of transition states sampled, for different regularization strengths.

ground-truth samples, which quantifies how well our method preserves the local molecular structure.

Transition-State Sampling. For the transition configuration sampling task, we adapt the kernel to handle rigid-body transformations using the Kabsch algorithm[63] akin to that adopted in ref 64. Since we found loss guidance to be difficult to optimize in this application, we first performed a large grid search to identify the optimal parameters for a fair comparison.This analysis revealed that increasing the guidance strength deteriorates sample quality in loss guidance, preventing it from achieving meaningful guiding success (Figure S13B in the Supporting Information). The performance gap stems from two key disadvantages of the loss guidance.

First, it requires using the posterior mean to compute the augmentation factor, which, especially at large t, suffers from very high variance. Second, at small t, while the predictions become more accurate, the KDE fails to capture the distribution of the guiding points, as it is not well-suited for high data complexity. As a result, the loss signal can degrade the sampled data, as evident from the increasing Wasserstein distance as the guiding strength ϑ_{init} increases (Figure 3A). In contrast, path guidance circumvents this issue by applying stronger guidance for larger t, where the latent is primarily noise, and decreasing it while sampling. Notably, in Figure 3A, we observe that both quality and diversity remain largely unaffected by the initial guiding strength ϑ_{init}. We further investigate the difference between path and loss guidance in Supporting Information D.2.

After observing that loss guidance could not be reliably optimized, we conducted a separate set of experiments to evaluate path guidance within the MEW framework by optimizing the objective in eq 14. Disabling regularization ($\gamma = 0$) results in the highest guidance success rates (Figure 3C) but produces highly degenerate samples and reduced structural diversity, as indicated by the large variance in Wasserstein distance. In contrast, applying MEW regularization improves both sample quality and diversity (Figure 3B), while incurring only a modest reduction in guidance success. Overall, our results demonstrate that path guidance offers a strong alternative to loss guidance and that MEW regularization is essential for robust and physically meaningful sampling in data-sparse regimes. Hence, with only minor modifications, we can scale path guidance.

To assess whether Path Guidance scales to higher-dimensional systems, we evaluate it on the 65-residue Protein G[65] using the BioEmu cgBE.[10] Because BioEmu represents both backbone geometry and local frames, we adapt the KDE-based guidance term to operate on the rotation tangent space (see Supporting Information B.4.3 for details). We found that standard sampling barely represents the transition states (~0.1%), hence, already minor upsampling would be remarkably helpful. Our experiments show that path guidance with MEW regularization increases the proportion of sampled transition states to ~28% while preserving physical validity (see Figure 4). With MEW regularization, the Wasserstein Distance (WD) between guided and reference bond-length distribution is WD = 0. 08 ± 0.002, and the Vendi score is VS = 10. 88 ± 0.33. Without regularization, sample quality degrades (WD = 0. 09 ± 0.002) while the Vendi score remains within variation (VS = 10. 94 ± 0.31), consistent with our findings on chignolin and further supporting the necessity of MEW regularization for maintaining sample quality under guidance.

RELATED WORKS

Stochastic Optimal Control

MEW also naturally connects to recent advances in stochastic optimal control (SOC) applied to diffusion and flow-based generative models. In particular, to approaches which consider steering generative trajectories by balancing a task-specific objective, such as aligning with experimental observables or reward models with a regularization that penalizes deviation from a pretrained base model. In those works,[17,66−68] fine-tuning the diffusion model is framed as an SOC problem that minimizes the control effort while achieving alignment with downstream goals. Conceptually, the MEW principle plays an analogous role to the control cost in SOCs, regularizing path perturbations to preserve the prior's structure while achieving target objectives. This connection puts MEW within the broader trend of leveraging control-theoretic principles,

https://doi.org/10.1021/acs.jctc.6c00080
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

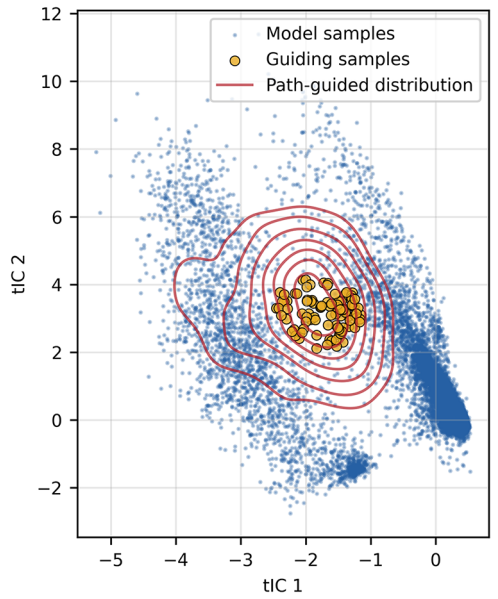


Figure 4. Path guidance on Protein G. Applying path guidance to Protein G increases sampling in the transition region while preserving biologically meaningful conformations. Shown is a 2D tICA projection of sampled configurations, highlighting samples under standard sampling (blue), transition-state regions used for guidance (yellow) and samples under guidance (red). The mode of the guided distribution is placed on the guidance samples and remains quantitatively consistent with the original transition-state ensemble.

including KL and f-divergence regularization, to derive principled, sample-efficient, and robust fine-tuning strategies for probabilistic generative models.

Transition Ensemble Sampling

Traditional methods such as transition-path sampling[69,70] use Monte Carlo in trajectory space, while recent machine learning approaches[71] employ neural networks but require extensive training data or predefined collective variables. Instead of explicit path sampling, we guide the generative process by using latent representations of known transition states. While related to recent work using Boltzmann generators,[72] our approach directly modifies the score function during sampling rather than performing MCMC moves between paths, enabling more efficient exploration of transition regions.

Reweighting with Experimental Data

Reweighting molecular dynamics simulations using experimental data has a long history in computational chemistry and biophysics. Theoretical work[30,31,34,73] adopted Jaynes[74] Maximum Entropy approach to the problem, following several early experimental studies[75−77] based on replica-averaged simulations, giving a theoretical foundation for these approaches. This work was later complemented by probabilistic and Bayesian perspectives,[32,38,53,78] some of which specifically focused on reweighing.[33,37,79,80]

LIMITATIONS

Despite the strong empirical performance of MEW guidance across a range of scientific settings, several limitations merit a consideration. These primarily stem from the assumptions underpinning the method's application, e.g., that physical observables or representative samples can be leveraged to correct expectation values or guide sampling in low-density regions. While this does not require perfect model accuracy, it does require the model to be sufficiently expressive and responsive to the guidance. If key modes are absent, then convergence to meaningful distributions may fail. Additionally, the current framework assumes differentiable observables, restricting its applicability in discrete or nondifferentiable domains.

Computational Cost

Parameter estimation (Bayesian optimization over η_{init}, κ) incurs a one-time cost of minutes for the toy system and up to a few hours for larger protein systems on a single GPU; full details are in Supporting Information B.3−B.4. Once estimated, guided sampling costs almost the same as unguided sampling, since no gradients through the score network are required. This is a higher upfront cost than post hoc reweighting, but the resulting guided generator can be reused indefinitely without maintaining a large weighted sample archive. Compared to enhanced MD sampling, MEW Path Guidance is substantially cheaper. Sampling the transition region of Protein G via unbiased MD would require simulation times on the order of the exchange time scale (~65 μs[57]), far exceeding the few GPU-hours needed for guidance.

CONCLUSION

In this work, we introduced minimum-excess-work (MEW) guidance, a physics-inspired framework for guidance of pretrained probability flow generative models with sparse external information by regularizing *excess work*. Our analysis shows that this thermodynamically motivated regularization is closely connected to upper bounds on the Wasserstein distance and the KL divergence between the reference and the guided distributions. We demonstrated the effectiveness of MEW regularization in two settings: *observable guidance* and *path guidance*. These approaches enable the alignment with sparse experimental constraints and targeted sampling in low-density regions while maintaining model flexibility. By penalizing excess work, our method reduces bias and enhances the sampling of rare, physically meaningful configurations, without degrading sample quality. Our results position MEW guidance as a principled and effective tool for bias correction and informed exploration in data-scarce scientific applications, such as the refinement of coarse-grained force fields against experimental data or the generation of starting conditions for unbiased MD or transition-path sampling.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.6c00080.

Proofs of propositions, experimental details, and additional results on toy systems (PDF)

AUTHOR INFORMATION

Corresponding Author

▽**Simon Olsson** − *Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, SE-41296 Gothenburg, Sweden;* orcid.org/0000-0002-3927-7897; Email: simonols@chalmers.se

https://doi.org/10.1021/acs.jctc.6c00080
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Authors

Christopher Kolloff − *Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, SE-41296 Gothenburg, Sweden; Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States*

Tobias Höppe − *Technical University of Munich, 80333 Munich, Germany; Helmholtz AI, Helmholtz Munich, 85764 Neuherberg, Germany*

Emmanouil Angelis − *Technical University of Munich, 80333 Munich, Germany; Helmholtz AI, Helmholtz Munich, 85764 Neuherberg, Germany*

Mathias Jacob Schreiner − *Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, SE-41296 Gothenburg, Sweden*

Stefan Bauer − *Technical University of Munich, 80333 Munich, Germany; Helmholtz AI, Helmholtz Munich, 85764 Neuherberg, Germany*

[▽]**Andrea Dittadi** − *Technical University of Munich, 80333 Munich, Germany; Helmholtz AI, Helmholtz Munich, 85764 Neuherberg, Germany; Max Planck Institute for Intelligent Systems, 72076 Tubingen, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.6c00080

Author Contributions

[#]C.K., T.H., and E.A. contributed equally to this work.

Notes

The authors declare no competing financial interest.
[▽]A.D. and S.O. are joint last authors.

■ ACKNOWLEDGMENTS

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation, the Helmholtz Foundation Model Initiative, the Helmholtz Association, the Fulbright Foreign Student Program sponsored by the U.S. Department of State and the Swiss-American Fulbright Commission, and the Chalmers Academic Excellence Program. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUPITER|JUWELS[81] at Jülich Supercomputing Centre (JSC). Preliminary results were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at Alvis (Project No.: NAISS 2023/22-1289), partially funded by the Swedish Research Council through Grant Agreement No. 2022-06725. The authors also acknowledge the MIT Office of Research Computing and Data for providing high-performance computing resources that have contributed to the research results reported within this paper. T.H. and A.D. acknowledge support from G-Research.

■ ADDITIONAL NOTE

[1]Code to reproduce the result are available on the dedicated repository: https://github.com/olsson-group/minimumexcesswork/.

■ REFERENCES

(1) Papamakarios, G.; Nalisnick, E.; Rezende, D. J.; Mohamed, S.; Lakshminarayanan, B. Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res.* **2021**, *22*, 1−64.

(2) Liu, Q. Rectified flow: A marginal preserving approach to optimal transport, arXiv:2209.14577. arXiv.org e-Print archive. https://arxiv.org/abs/2209.14577 2022.

(3) Chen, R. T.; Rubanova, Y.; Bettencourt, J.; Duvenaud, D. K. Neural ordinary differential equations. *Adv. Neural Inf. Process. Syst.* **2018**, *31*.

(4) Lipman, Y.; Chen, R. T. Q.; Ben-Hamu, H.; Nickel, M.; Le, M. *Flow Matching for Generative Modeling*, Eleventh International Conference on Learning Representations; 2023.

(5) Song, Y.; Sohl-Dickstein, J.; Kingma, D. P.; Kumar, A.; Ermon, S.; Poole, B. Score-based generative modeling through stochastic differential equations, arXiv:2011.13456. arXiv.org e-Print archive. https://arxiv.org/abs/2011.13456 2020.

(6) Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840−6851.

(7) Noé, F.; Olsson, S.; Köhler, J.; Wu, H. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* **2019**, *365*, No. eaaw1147.

(8) Jing, B.; Corso, G.; Chang, J.; Barzilay, R.; Jaakkola, T. Torsional diffusion for molecular conformer generation. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 24240−24253.

(9) Diez, J. V.; Atance, S. R.; Engkvist, O.; Olsson, S. Generation of conformational ensembles of small molecules via surrogate model-assisted molecular dynamics. *Mach. Learn.: Sci. Technol.* **2024**, *5*, No. 025010.

(10) Lewis, S.; Hempel, T.; Jiménez-Luna, J.; Gastegger, M.; Xie, Y.; Foong, A. Y.; Satorras, V. G.; Abdin, O.; Veeling, B. S.; Zaporozhets, I.; et al. Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science* **2025**, *389*, No. eadv9817.

(11) Schreiner, M.; Winther, O.; Olsson, S. *Implicit Transfer Operator Learning: Multiple Time-Resolution Models for Molecular Dynamics*, Thirty-seventh Conference on Neural Information Processing Systems; 2023.

(12) Diez, J. V.; Schreiner, M.; Engkvist, O.; Olsson, S. Boltzmann priors for Implicit Transfer Operators, arXiv:2410.10605. arXiv.org e-Print archive. https://arxiv.org/abs/2410.10605 2024.

(13) Diez, J. V.; Schreiner, M.; Olsson, S. Transferable Generative Models Bridge Femtosecond to Nanosecond Time-Step Molecular Dynamics. 2025.

(14) Olsson, S. Generative molecular dynamics. *Curr. Opin. Struct. Biol.* **2026**, *96*, No. 103213.

(15) Wallace, B.; Dang, M.; Rafailov, R.; Zhou, L.; Lou, A.; Purushwalkam, S.; Ermon, S.; Xiong, C.; Joty, S.; Naik, N.Diffusion model alignment using direct preference optimization *Pattern Recognition* 2024, pp 8228−8238.

(16) Black, K.; Janner, M.; Du, Y.; Kostrikov, I.; Levine, S. Training diffusion models with reinforcement learning, arXiv:2305.13301. arXiv.org e-Print archive. https://arxiv.org/abs/2305.13301 2023.

(17) Domingo-Enrich, C.; Drozdzal, M.; Karrer, B.; Chen, R. T. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control, arXiv:2409.08861. arXiv.org e-Print archive. https://arxiv.org/abs/2409.08861 2024.

(18) Ho, J.; Salimans, T. Classifier-free diffusion guidance, arXiv:2207.12598. arXiv.org e-Print archive. https://arxiv.org/abs/2207.12598 2022.

(19) Nichol, A.; Dhariwal, P.; Ramesh, A.; Shyam, P.; Mishkin, P.; McGrew, B.; Sutskever, I.; Chen, M. Glide: Towards photorealistic image generation and editing with text-guided diffusion models, arXiv:2112.10741. arXiv.org e-Print archive. https://arxiv.org/abs/2112.10741 2021,.

(20) Anderson, B. D. Reverse-time diffusion equation models. *Stochastic Processes and their Applications* **1982**, *12*, 313−326.

(21) Maoutsa, D.; Reich, S.; Opper, M. Interacting Particle Solutions of Fokker−Planck Equations Through Gradient−Log−Density Estimation. *Entropy* **2020**, *22*, 802.

(22) Klein, L.; Krämer, A.; Noé, F. Equivariant flow matching *Adv. Neural Inf. Process. Syst.* 2024; Vol. *36*.
(23) Midgley, L. I.; Stimper, V.; Antoran, J.; Mathieu, E.; Schölkopf, B.; Hernández-Lobato, J. M. *SE(3) Equivariant Augmented Coupling Flows*, Thirty-seventh Conference on Neural Information Processing Systems' 2023.
(24) Köhler, J.; Klein, L.; Noe, F.*Equivariant Flows: Exact Likelihood Generative Learning for Symmetric Densities*, Proceedings of the 37th International Conference on Machine Learning; 2020; pp 5361−5370.
(25) Moqvist, S.; Chen, W.; Schreiner, M.; Nüske, F.; Olsson, S. Thermodynamic Interpolation: A Generative Approach to Molecular Thermodynamics and Kinetics. *J. Chem. Theory Comput.* **2025**, *21*, 2535−2545, DOI: 10.1021/acs.jctc.4c01557.
(26) Tan, C. B.; Bose, J.; Lin, C.; Klein, L.; Bronstein, M. M.; Tong, A. *Scalable Equilibrium Sampling with Sequential Boltzmann Generators*, Forty-second International Conference on Machine Learning.
(27) Kolloff, C.; Mazur, A.; Marzinek, J. K.; Bond, P. J.; Olsson, S.; Hiller, S. Motional clustering in supra-τc conformational exchange influences NOE cross-relaxation rate. *J. Magn. Reson.* **2022**, *338*, No. 107196.
(28) Klein, L.; Foong, A.; Fjelde, T.; Mlodozeniec, B.; Brockschmidt, M.; Nowozin, S.; Noé, F.; Tomioka, R. Timewarp: Transferable acceleration of molecular dynamics by learning time-coarsened dynamics *Adv. Neural Inf. Process. Syst.* 2024; Vol. *36*.
(29) Arts, M.; Garcia Satorras, V.; Huang, C.-W.; Zugner, D.; Federici, M.; Clementi, C.; Noé, F.; Pinsler, R.; van den Berg, R. Two for one: Diffusion models and force fields for coarse-grained molecular dynamics. *J. Chem. Theory Comput.* **2023**, *19*, 6151−6159.
(30) Pitera, J. W.; Chodera, J. D. On the use of experimental observations to bias simulated ensembles. *J. Chem. Theory Comput.* **2012**, *8*, 3445−3451.
(31) Cavalli, A.; Camilloni, C.; Vendruscolo, M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J. Chem. Phys.* **2013**, *138*, No. 094112, DOI: 10.1063/1.4793625.
(32) Olsson, S.; Frellsen, J.; Boomsma, W.; Mardia, K. V.; Hamelryck, T. Inference of structure ensembles of flexible biomolecules from sparse, averaged data. *PLoS One* **2013**, *8*, No. e79439.
(33) Olsson, S.; Strotz, D.; Vögeli, B.; Riek, R.; Cavalli, A. The dynamic basis for signal propagation in human Pin1-WW. *Structure* **2016**, *24*, 1464−1475.
(34) Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Combining experiments and simulations using the maximum entropy principle. *PLoS Comput. Biol.* **2014**, *10*, No. e1003406.
(35) White, A. D.; Voth, G. A. Efficient and minimal method to bias molecular simulations with experimental data. *J. Chem. Theory Comput.* **2014**, *10*, 3023−3030.
(36) Beauchamp, K. A.; Pande, V. S.; Das, R. Bayesian energy landscape tilting: towards concordant models of molecular ensembles. *Biophys. J.* **2014**, *106*, 1381−1390.
(37) Hummer, G.; Köfinger, J. Bayesian ensemble refinement by replica simulations and reweighting. *J. Chem. Phys.* **2015**, *143*, No. 243150, DOI: 10.1063/1.4937786.
(38) Bonomi, M.; Camilloni, C.; Cavalli, A.; Vendruscolo, M. Metainference: ABayesian inference method for heterogeneous systems. *Sci. Adv.* **2016**, *2*, No. e1501177.
(39) Bansal, A.; Chu, H.-M.; Schwarzschild, A.; Sengupta, S.; Goldblum, M.; Geiping, J.; Goldstein, T.*Universal guidance for diffusion models*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2023; pp 843−852.
(40) Chung, H.; Kim, J.; Mccann, M. T.; Klasky, M. L.; Ye, J. C. *Diffusion Posterior Sampling for General Noisy Inverse Problems*, Eleventh International Conference on Learning Representations; 2023.
(41) Song, J.; Zhang, Q.; Yin, H.; Mardani, M.; Liu, M.-Y.; Kautz, J.; Chen, Y.; Vahdat, A.Loss-guided diffusion models for plug-and-play controllable generation *Mach. Learn.* 2023, pp 32483−32498.
(42) Sekimoto, K. *Stochastic Energetics*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; pp 150−151.
(43) Brenier, Y. Polar factorization and monotone rearrangement of vector-valued functions. *Commun. Pure Appl. Math.* **1991**, *44*, 375−417.
(44) Benamou, J.-D.; Brenier, Y. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik* **2000**, *84*, 375−393.
(45) Tong, A.; Huang, J.; Wolf, G.; van Dijk, D.; Krishnaswamy, S. TrajectoryNet: A Dynamic Optimal Transport Network for Modeling Cellular Dynamics. 2020.
(46) Tong, A.; Malkin, N.; Huguet, G.; Zhang, Y.; Rector-Brooks, J.; Fatras, K.; Wolf, G.; Bengio, Y.Improving and generalizing flow-based generative models with minibatch optimal transport, arXiv:2302.00482. arXiv.org e-Print archive. https://arxiv.org/abs/2302.00482 2023.
(47) Klein, L.; Krämer, A.; Noé, F. Equivariant flow matching. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 59886−59910.
(48) Irwin, R.; Tibo, A.; Janet, J. P.; Olsson, S. *SemlaFlow − Efficient 3D Molecular Generation with Latent Attention and Equivariant Flow Matching*, 28th International Conference on Artificial Intelligence and Statistics; 2025.
(49) Shaul, N.; Chen, R. T. Q.; Nickel, M.; Le, M.; Lipman, Y.*On Kinetic Optimal Probability Paths for Generative Models*, Proceedings of the 40th International Conference on Machine Learning; 2023; pp 30883−30907.
(50) Albergo, M. S.; Boffi, N. M.; Vanden-Eijnden, E. Stochastic interpolants: A unifying framework for flows and diffusions, arXiv:2303.08797. arXiv.org e-Print archive. https://arxiv.org/abs/2303.08797 2023,.
(51) Neklyudov, K.; Brekelmans, R.; Severo, D.; Makhzani, A.*Action Matching: Learning Stochastic Dynamics from Samples*, Proceedings of the 40th International Conference on Machine Learning; 2023; pp 25858−25889.
(52) Neklyudov, K.; Brekelmans, R.; Tong, A.; Atanackovic, L.; Liu, Q.; Makhzani, A. A Computational Framework for Solving Wasserstein Lagrangian Flows. 2023.
(53) Bottaro, S.; Bengtsen, T.; Lindorff-Larsen, K. Integrating molecular simulation and experimental data: a Bayesian/maximum entropy reweighting approach. *Struct. Bioinf.: Methods Protoc.* **2020**, *2112*, 219−240.
(54) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **2011**, *134*, No. 174105, DOI: 10.1063/1.3565032.
(55) Honda, S.; Yamasaki, K.; Sawada, Y.; Morii, H. 10 residue folded peptide designed by segment statistics. *Structure* **2004**, *12*, 1507−1518.
(56) Satoh, D.; Shimizu, K.; Nakamura, S.; Terada, T. Folding free-energy landscape of a 10-residue mini-protein, chignolin. *FEBS Lett.* **2006**, *580*, 3422−3426.
(57) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517−520.
(58) Religa, T. L.; Johnson, C. M.; Vu, D. M.; Brewer, S. H.; Dyer, R. B.; Fersht, A. R. The helix-turn-helix motif as an ultrafast independently folding domain: The pathway of folding of Engrailed homeodomain. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 9272−9277.
(59) Religa, T. L. Comparison of multiple crystal structures with NMR data for engrailed homeodomain. *J. Biomol. NMR* **2008**, *40*, 189−202.
(60) Mayor, U.; Johnson, C. M.; Daggett, V.; Fersht, A. R.; Kazmirski, S. L.; Wong, K. B.; Freund, S. M. V.; Tan, Y.-J. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 13518−13522.
(61) Bax, A. Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci.* **2003**, *12*, 1−16.
(62) Friedman, D.; Dieng, A. B. The vendi score: A diversity evaluation metric for machine learning, arXiv:2210.02410. arXiv.org e-Print archive. https://arxiv.org/abs/2210.02410 2022.

(63) Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr., Sect. A* **1976**, *32*, 922−923.
(64) Pasarkar, A. P.; Bencomo, G. M.; Olsson, S.; Dieng, A. B. Vendi sampling for molecular simulations: Diversity as a force for faster convergence and better exploration. *J. Chem. Phys.* **2023**, *159*, No. 144108, DOI: 10.1063/5.0166172.
(65) Nauli, S.; Kuhlman, B.; Le Trong, I.; Stenkamp, R. E.; Teller, D.; Baker, D. Crystal structures and increased stabilization of the protein G variants with switched folding pathways NuG1 and NuG2. *Protein Sci.* **2002**, *11*, 2924−2931.
(66) Uehara, M.; Zhao, Y.; Black, K.; Hajiramezanali, E.; Scalia, G.; Diamant, N. L.; Tseng, A. M.; Biancalani, T.; Levine, S. Fine-tuning of continuous-time diffusion models as entropy-regularized control, arXiv:2402.15194. arXiv.org e-Print archive. https://arxiv.org/abs/2402.15194 2024.
(67) Han, Y.; Razaviyayn, M.; Xu, R. Stochastic Control for Fine-tuning Diffusion Models: Optimality, Regularity, and Convergence, arXiv:2412.18164. arXiv.org e-Print archive. https://arxiv.org/abs/2412.18164 2024.
(68) Tang, W. Fine-tuning of diffusion models via stochastic control: entropy regularization and beyond, arXiv:2403.06279. arXiv.org e-Print archive. https://arxiv.org/abs/2403.06279 2024.
(69) Bolhuis, P. G.; Dellago, C.; Geissler, P. L.; Chandler, D. Transition path sampling: throwing ropes over mountains in the dark. *J. Phys.:Condens. Matter* **2000**, *12*, No. A147, DOI: 10.1088/0953-8984/12/8A/316.
(70) Cabriolu, R.; Skjelbred Refsnes, K. M.; Bolhuis, P. G.; van Erp, T. S. Foundations and latest advances in replica exchange transition interface sampling. *J. Chem. Phys.* **2017**, *147*, No. 152722, DOI: 10.1063/1.4989844.
(71) Liu, B.; Boysen, J. G.; Unarta, I. C.; Du, X.; Li, Y.; Huang, X. Exploring transition states of protein conformational changes via out-of-distribution detection in the hyperspherical latent space. *Nat. Commun.* **2025**, *16*, No. 349.
(72) Plainer, M.; Stärk, H.; Bunne, C.; Günnemann, S. Transition path sampling with boltzmann generator-based mcmc moves, arXiv:2312.05340. arXiv.org e-Print archive. https://arxiv.org/abs/2312.05340 2023.
(73) Roux, B.; Weare, J. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *J. Chem. Phys.* **2013**, *138*, No. 084107, DOI: 10.1063/1.4792208.
(74) Jaynes, E. T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, No. 620.
(75) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **2005**, *433*, 128−132.
(76) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Mapping Long-Range Interactions in α-Synuclein using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2005**, *127*, 476−477.
(77) Cesari, A.; Reißer, S.; Bussi, G. Using the Maximum Entropy Principle to Combine Simulations and Solution Experiments. *Computation* **2018**, *6*, No. 15.
(78) Bonomi, M.; Pellarin, R.; Vendruscolo, M. Simultaneous determination of protein structure and dynamics using cryo-electron microscopy. *Biophys. J.* **2018**, *114*, 1604−1613.
(79) Olsson, S.; Wu, H.; Paul, F.; Clementi, C.; Noé, F. Combining experimental and simulation data of molecular processes via augmented Markov models. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 8265−8270.
(80) Kolloff, C.; Olsson, S. Rescuing off-equilibrium simulation data through dynamic experimental data with dynAMMo. *Mach. Learn.: Sci. Technol.* **2023**, *4*, No. 045050, DOI: 10.1088/2632-2153/ad10ce.
(81) Alvarez, D. JUWELS Cluster and Booster: Exascale Pathfinder with Modular Supercomputing Architecture at Juelich Supercomputing Centre. *J. Large-Scale Res. Facil.* **2021**, *7*, No. A183, DOI: 10.17815/jlsrf-7-183.

https://doi.org/10.1021/acs.jctc.6c00080
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX