VESPER: ONLINE LEARNING OF FREE ENERGIES WITH A MULTI-TEMPERED VARIATIONAL OBJECTIVE

Anonymous authors
Paper under double-blind review

ABSTRACT

Computing a molecular free-energy landscape over collective variables (CVs) is a learning problem without a dataset: the microscopic potential is known, but samples of the CV distribution must be generated by molecular dynamics (MD), which rarely crosses the barriers of interest. We introduce VESPER, which learns a neural free energy while using it to bias the simulations that produce its own training data. VESPER minimizes a single variational objective that combines several tempering levels, each flattening the landscape by a different amount. Every level is minimized by the same physical free energy, so all of them train one model together, without reweighting. At each level, biased MD configurations are compared with auxiliary walkers that explore the flattened model directly in CV space, and replica exchange between levels further accelerates exploration. We prove that the coupled sampling and learning dynamics converge under log-Sobolev and Polyak–Łojasiewicz conditions. On alanine dipeptide VESPER matches or improves on established enhanced-sampling methods; it is self-consistent from three to six CVs on alanine tetrapeptide, reproduces the reference barrier for methane dehydrogenation on a nickel catalyst with a machine-learned potential, and unfolds and refolds the protein chignolin with seventeen CVs.

1 INTRODUCTION

Molecular dynamics (MD) is a central tool for studying the structure and dynamics of molecular systems in physics, chemistry, and biology. A recurring goal is to compute the free-energy landscape over a set of collective variables (CVs), defined as the negative logarithm of their marginal density under the equilibrium distribution. This landscape quantifies the relative stability of metastable states and the barriers between them. Given samples from this marginal, learning it would be a standard estimation problem. In MD, however, no such data are available: the microscopic potential is known but configurations must be generated by simulation, and brute-force MD cannot generate them in the regimes of interest. Barriers are crossed on microsecond-to-millisecond timescales, while integration requires femtosecond timesteps (Hénin et al., 2022). Free-energy learning is therefore coupled to a sampling problem, and the model must accelerate the simulation while it is being learned.

We introduce VESPER (Variational Extended Space Parallel Ensemble Regression), which learns a neural free energy G_θ while using it to bias the simulations that produce its own training data. VESPER minimizes a single variational objective: a positively weighted sum of components, one per tempering level, each flattening the landscape by a different amount. Each component is convex in the free energy and is minimized by the physical free energy at every tempering level. The sum therefore has the same minimizer, and every level contributes its gradient to one model without reweighting to the physical ensemble. Taken individually, each component belongs to the pseudo-spherical and Rényi families of variational objectives (Yu et al., 2021; Fujisawa & Eguchi, 2008; Birrell et al., 2021), and its gradient coincides with the instantaneous update of well-tempered variationally enhanced sampling (Valsson & Parrinello, 2014; 2015; Bonati et al., 2019). VESPER contributes an online estimator of this gradient that needs no equilibrium data, together with the observation that all tempering levels can be pooled into one objective for one model.

The gradient of each component is the difference between two averages of $\nabla_\theta G_\theta$: one over MD configurations evolving under the microscopic potential plus a bias built from G_θ, the other over

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

persistent auxiliary walkers evolving directly in CV space under a tempered copy of G_θ. At the optimum both populations sample the same flattened distribution, and the walkers interact with the molecular system only through the shared parameters, so, unlike temperature-accelerated MD (Maragliano & Vanden-Eijnden, 2006; Abrams & Tuckerman, 2008), no adiabatic limit is needed to identify the free energy. Replica exchange can additionally carry configurations discovered under strongly flattened biases down the ladder.

Our main contributions are:

- **A single objective with a common minimizer.** Every tempering level defines a component that is convex in the free energy and minimized by the physical free energy, so their weighted sum is one objective for one model, to which every level contributes directly rather than serving only as a sampler for the physical level (Proposition 1).
- **Online estimation from coupled populations.** The gradient is estimated from biased MD and auxiliary CV-space walkers that evolve alongside the parameters, with no pre-existing dataset or adiabatic separation. We prove convergence of the coupled sampling–learning dynamics under log-Sobolev and Polyak–Łojasiewicz conditions (Appendix A.6).
- **Replica exchange on the learned free energy.** The acceptance ratio for exchange between adjacent levels involves only the learned free energy in CV space, not the microscopic energy, so it is independent of system size, and exchange at fixed model preserves each level's sampling law.
- **Evaluation in the online-sampling regime.** On alanine dipeptide VESPER matches or improves on established enhanced-sampling methods, and on analytic landscapes the pooled ladder removes the need to tune the tempering strength; it scales from three to six CVs on alanine tetrapeptide, reproduces the reference barrier for methane dehydrogenation on Ni(511) with a machine-learned potential, and samples folded and unfolded chignolin with seventeen dihedral CVs (Section 3).

2 METHODOLOGY

2.1 COLLECTIVE VARIABLES AND FREE ENERGY

Let $x \in \mathcal{X}$ denote a molecular configuration, $U(x)$ its potential energy, and $\varrho(x) = Z^{-1}e^{-\beta U(x)}$ its equilibrium density at inverse temperature $\beta = (k_B T)^{-1}$. A collective-variable map $s : \mathcal{X} \to \mathcal{S}$ reduces the d-dimensional configuration space to a D-dimensional CV domain, typically with $D \ll d$. The equilibrium CV density and its associated free energy are

$$p(s) = \int_{\mathcal{X}} \delta(s - s(x))\varrho(x)\,\mathrm{d}x, \qquad G_\star(s) = -\beta^{-1}\log p(s). \tag{1}$$

The microscopic potential supplies the forces for simulation, whereas the marginal $p(s)$ must be inferred from configurations generated by the simulation. VESPER represents $G_\star$ by a differentiable model G_θ and uses it to bias the molecular dynamics that generate its training configurations.

2.2 THE VARIATIONAL OBJECTIVE

Fix an auxiliary inverse temperature $0 < \hat\beta < \beta$ and set $\alpha = \hat\beta/\beta \in (0, 1)$. We define the component

$$\mathcal{L}_{\hat\beta}[G] = \frac{1}{\hat\beta}\log \underbrace{\int_{\mathcal{S}} e^{-\hat\beta G(s)}\,\mathrm{d}s}_{=:A_G} + \frac{1}{\beta - \hat\beta}\log \underbrace{\int_{\mathcal{X}} e^{(\beta-\hat\beta)G(s(x))}\varrho(x)\,\mathrm{d}x}_{=:B_G}. \tag{2}$$

The two terms are the log-partition functions of a tempered version of the model on $\mathcal{S}$ and of the equilibrium density tilted by the model. They define the auxiliary CV law

$$P_{G,\hat\beta}(s) = \frac{e^{-\hat\beta G(s)}}{A_G} = \frac{e^{-\alpha\beta G(s)}}{A_G}, \tag{3}$$

and the biased microscopic law together with its CV marginal,

$$\nu_{G,\hat\beta}(x) = \frac{e^{(\beta-\hat\beta)G(s(x))}\varrho(x)}{B_G}, \qquad Q_{G,\hat\beta}(s) = \frac{p(s)e^{(\beta-\hat\beta)G(s)}}{B_G}. \tag{4}$$

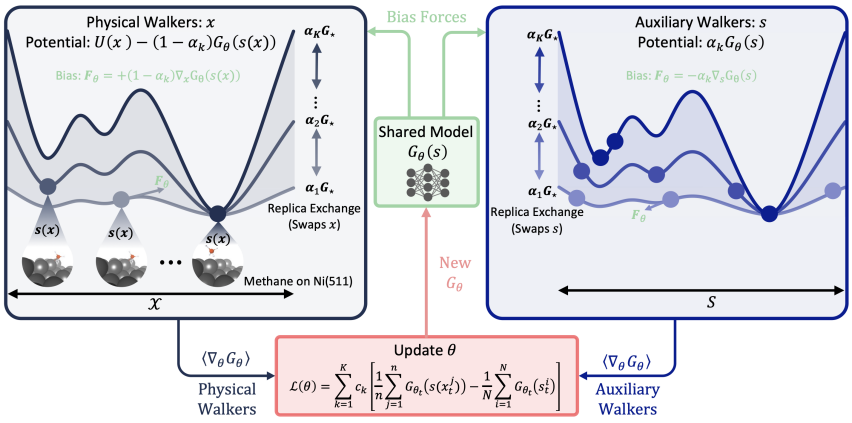


Figure 1: Coupled sampling and learning in VESPER. Each tempering level k carries two walker populations, and all levels share one model G_θ. Left: physical walkers evolve in configuration space under $U(x) - (1-\alpha_k)G_\theta(s(x))$ and are mapped to CV space by $s(x)$; replica exchange swaps configurations between adjacent levels. Right: auxiliary walkers evolve directly in CV space under $\alpha_k G_\theta(s)$. The two populations supply the averages in equation 10, whose c_k-weighted difference updates θ; the updated model then sets the forces on both populations. At convergence, both populations at level k sample the same flattened landscape $\alpha_k G_\star$; during learning, the CVs of the physical walkers instead follow $Q_{G_\theta,k} \propto p\, e^{(\beta-\hat\beta_k)G_\theta}$.

The first variation of equation 2 in a direction h is $\mathbb{E}_{Q_{G,\hat\beta}} h - \mathbb{E}_{P_{G,\hat\beta}} h$, so G is stationary exactly when the two CV laws coincide:

$$\frac{e^{-\hat\beta G}}{A_G} = \frac{p\, e^{(\beta-\hat\beta)G}}{B_G} \iff e^{-\beta G} \propto p \iff G = G_\star + C. \tag{5}$$

The tempering exponent cancels: for every α, the stationary point is the physical free energy. The second variation, $\hat\beta \operatorname{Var}_{P_{G,\hat\beta}}(h) + (\beta - \hat\beta)\operatorname{Var}_{Q_{G,\hat\beta}}(h)$, is nonnegative, so $\mathcal{L}_{\hat\beta}$ is convex and its minimizer set is $\{G_\star + C : C \in \mathbb{R}\}$ (Proposition S1). At the minimizer,

$$P_{G_\star,\hat\beta} = Q_{G_\star,\hat\beta} = \pi_\alpha, \qquad \pi_\alpha \propto e^{-\alpha\beta G_\star} \propto p^\alpha, \tag{6}$$

so both populations explore the same flattened landscape, with barriers reduced by the factor α. Near the minimizer, $\mathcal{L}_{\hat\beta}[G] - \mathcal{L}_{\hat\beta}[G_\star] \approx \frac{\beta}{2}\operatorname{Var}_{\pi_\alpha}(G - G_\star)$: α does not change the target, only where errors are weighted, with smaller α giving more weight to high free-energy regions.

For a parametric model G_θ, the chain rule gives

$$\nabla_\theta \mathcal{L}_{\hat\beta}[G_\theta] = \mathbb{E}_{\nu_{G_\theta,\hat\beta}}[\nabla_\theta G_\theta(s(x))] - \mathbb{E}_{P_{G_\theta,\hat\beta}}[\nabla_\theta G_\theta(s)]\,. \tag{7}$$

Neither expectation requires knowing p. The first is over configurations sampled by MD biased with G_θ, and the second over walkers sampling the tempered model in CV space (Section 2.4).

2.3 The combined objective

Choose $0 \le \hat\beta_1 < \cdots < \hat\beta_K \le \beta$, set $\alpha_k = \hat\beta_k/\beta$, and write $P_{G,k}$, $\nu_{G,k}$, $Q_{G,k}$ for the laws of level k. The component $\mathcal{L}_{\hat\beta_k}$ is given by equation 2 for $0 < \hat\beta_k < \beta$ and extended by continuity to $\hat\beta_k = \beta$ and, when $\mathcal{S}$ is compact, to $\hat\beta_k = 0$:

$$\mathcal{L}_\beta[G] = \frac{1}{\beta}\log\int_{\mathcal{S}} e^{-\beta G(s)}\,\mathrm{d}s + \mathbb{E}_p[G], \qquad \mathcal{L}_0[G] = -\mathbb{E}_u[G] + \frac{1}{\beta}\log\mathbb{E}_p\left[e^{\beta G}\right], \tag{8}$$

where u is the uniform law on $\mathcal{S}$, and $\mathcal{L}_0$ is taken after subtracting the constant $\hat\beta^{-1}\log|\mathcal{S}|$ from equation 2. The first is the maximum-likelihood objective, $\beta^{-1}\mathrm{KL}(p \,\|\, q_G)$ up to a constant with

$q_G \propto e^{-\beta G}$: its physical walkers are unbiased. The second is the fully flattened level, whose auxiliary walkers are exact uniform samples.

Proposition 1 (VESPER objective and minimizer). *Let $c_k > 0$ with $\sum_{k=1}^{K} c_k = 1$, and define*

$$\mathcal{L}_{\mathrm{VESPER}}[G] = \sum_{k=1}^{K} c_k \, \mathcal{L}_{\hat{\beta}_k}[G], \tag{9}$$

with components given by equation 2, or by equation 8 if $\hat{\beta}_1 = 0$ or $\hat{\beta}_K = \beta$. Under the integrability assumptions of Appendix A.4, $\mathcal{L}_{\mathrm{VESPER}}$ is convex in G, invariant under addition of a constant, and has minimizer set $\{G_\star + C : C \in \mathbb{R}\}$. In a realizable model family, every minimizer therefore represents $G_\star$ up to a constant. For a differentiable model G_θ,

$$\nabla_\theta \mathcal{L}_{\mathrm{VESPER}}[G_\theta] = \sum_{k=1}^{K} c_k \left\{ \mathbb{E}_{\nu_{G_\theta,k}}[\nabla_\theta G_\theta(s(x))] - \mathbb{E}_{P_{G_\theta,k}}[\nabla_\theta G_\theta(s)] \right\}. \tag{10}$$

Every level, including strongly flattened ones, thus trains the same model directly, without reweighting between levels. The weights c_k and exponents α_k leave the target unchanged but set how errors are weighted: near the minimizer, $\mathcal{L}_{\mathrm{VESPER}}[G] - \mathcal{L}_{\mathrm{VESPER}}[G_\star] \approx \frac{\beta}{2} \sum_k c_k \, \mathrm{Var}_{\pi_{\alpha_k}}(G - G_\star)$, which also governs the fit, to leading order, when the model family cannot represent $G_\star$ exactly.

2.4 Coupled sampling and optimization

We estimate the expectations in equation 10 with two populations at each level, all sharing G_θ (Fig. 1). At level k, n_k physical walkers evolve under the biased potential $U(x) - (1-\alpha_k) G_\theta(s(x))$, and N_k auxiliary walkers evolve directly in CV space under the tempered potential $\alpha_k G_\theta(s)$:

$$\begin{aligned} m_s \ddot{s}_t^{i,k} + \gamma_s \dot{s}_t^{i,k} &= -\alpha_k \nabla_s G_{\theta_t}(s_t^{i,k}) + \sqrt{2\gamma_s \beta^{-1}}\, \eta_t^{i,k}, \\ m_x \ddot{x}_t^{j,k} + \gamma_x \dot{x}_t^{j,k} &= -\nabla_x U(x_t^{j,k}) + (1-\alpha_k) \nabla_x G_{\theta_t}(s(x_t^{j,k})) + \sqrt{2\gamma_x \beta^{-1}}\, \tilde{\eta}_t^{j,k}, \end{aligned} \tag{11}$$

where m_s, m_x are masses, γ_s, γ_x friction coefficients, and $\eta_t^{i,k}, \tilde{\eta}_t^{j,k}$ independent white noises. The bias force on the molecular system is $\nabla_x G_\theta(s(x)) = J_s(x)^\top \nabla_s G_\theta(s(x))$, with $J_s = \partial s / \partial x \in \mathbb{R}^{D \times d}$. For fixed θ, these dynamics have configurational invariant laws $P_{G_\theta,k}$ and $\nu_{G_\theta,k}$ (Appendix A.3), so their walker averages estimate the two expectations in equation 10. The parameters follow the corresponding gradient flow with γ_θ^{-1} the learning rate,

$$\gamma_\theta \dot{\theta}_t = \sum_{k=1}^{K} c_k \left[\frac{1}{N_k} \sum_{i=1}^{N_k} \nabla_\theta G_{\theta_t}(s_t^{i,k}) - \frac{1}{n_k} \sum_{j=1}^{n_k} \nabla_\theta G_{\theta_t}(s(x_t^{j,k})) \right]. \tag{12}$$

Three features of this construction matter in practice.

- **Timescales.** Since θ moves, the walkers track moving targets, and the averages in equation 12 approximate the gradient when θ evolves more slowly than the walkers relax. Appendix A.6 makes this quantitative for the overdamped dynamics: objective error and sampling lag decay together below a learning-rate threshold set by the log-Sobolev constants of the sampling laws, a bound on the learning rate rather than an adiabatic limit.
- **Cost.** Auxiliary walkers live in the D-dimensional CV space and need one network evaluation per step, so N_k can be large (10^4–10^5 in the molecular runs) with cost dominated by the biased MD.
- **Persistence.** Auxiliary walkers persist between parameter updates, as in persistent contrastive divergence (Tieleman, 2008), staying close to equilibrium in the slowly changing tempered model.

In the implementation, equation 12 is discretized: sampling epochs alternate with optimizer steps, and batches may be reused across several steps with within-batch importance weights (Algorithm 1, Appendix B.7). When equilibrium configurations from ϱ are available, the physical average can instead be computed from them with self-normalized weights (Appendix A.9).

2.5 REPLICA EXCHANGE ACROSS TEMPERING LEVELS

The common-minimizer property does not require the levels to communicate. Sampling, however, benefits from it: strongly flattened levels cross barriers readily, while weakly flattened ones, including the physical level, may remain trapped. We therefore couple adjacent levels by replica exchange (Sugita & Okamoto, 1999; Affentranger et al., 2006; Earl & Deem, 2005), performed between parameter updates with G_θ held fixed. For a symmetric swap proposal between levels i and j, the Metropolis acceptance probability for physical walkers is

$$A_x(x^i, x^j) = \min\Big\{1, \exp\Big[(\hat{\beta}_i - \hat{\beta}_j)\big(G_\theta(s(x^i)) - G_\theta(s(x^j))\big)\Big]\Big\}, \tag{13}$$

and $A_s(s^i, s^j)$ is identical with $s(x^i)$, $s(x^j)$ replaced by s^i, s^j. Each swap preserves the product law $\prod_k \nu_{G_\theta,k}$ or $\prod_k P_{G_\theta,k}$, so exchange leaves every level's sampling distribution, and hence the gradient estimator, unchanged (Appendix A.5). All physical replicas share the potential U and the thermostat temperature, so the microscopic energy and the Maxwell factors cancel from the physical ratio and only the bias difference remains: both acceptance ratios depend only on the learned free energy in the D-dimensional CV space. Whereas in temperature replica exchange energy fluctuations grow with system size and the ladder must be refined accordingly, here the spacing is set by the variation of G_θ between adjacent levels and is independent of system size; physical and auxiliary populations are exchanged independently.

2.6 RELATED WORK

Adaptive-bias enhanced sampling. Umbrella sampling, adaptive biasing force, metadynamics and its well-tempered variant, and on-the-fly probability enhanced sampling accelerate MD with restraining or adaptively built CV-dependent biases during the simulation (Torrie & Valleau, 1977; Darve & Pohorille, 2001; Darve et al., 2008; Laio & Parrinello, 2002; Barducci et al., 2008; Invernizzi & Parrinello, 2020). Neural representations of the free energy or bias have been trained from mean forces, histograms, or both (Schneider et al., 2017; Sidky & Whitmer, 2018; Guo et al., 2018; Sevgen et al., 2020; Zubieta Rico et al., 2025); reinforced dynamics couples such training to an uncertainty-gated bias (Zhang et al., 2018). Most similar to VESPER is variationally enhanced sampling (VES) (Valsson & Parrinello, 2014), in its well-tempered (Valsson & Parrinello, 2015) and neural (Bonati et al., 2019) forms. With bias $V_\theta = -(1-\alpha)G_\theta$ and target $t_\theta \propto e^{-\alpha\beta G_\theta}$ held fixed during differentiation, the VES gradient equals $(1-\alpha)\nabla_\theta \mathcal{L}_{\hat{\beta}}[G_\theta]$, so a single VESPER level reproduces the instantaneous well-tempered VES update as the gradient of an explicit convex objective, with the target set by the current model and sampled by auxiliary walkers rather than by a separate recursive estimate (Appendix D.1).

Replica ladders with CV biases. Replica exchange improves exploration by swapping configurations between replicas with different temperatures or Hamiltonians (Sugita & Okamoto, 1999; Affentranger et al., 2006; Earl & Deem, 2005). Bias-exchange metadynamics assigns different CVs to different replicas (Piana & Laio, 2007). Replica exchange with collective-variable tempering uses a ladder of well-tempered biases of increasing strength, which is the closest analogue of VESPER's physical population (Gil-Ley & Bussi, 2015). There the flattened replicas serve only as samplers, with the free energy estimated from the unbiased replica or by reweighting to it; in VESPER every level contributes a gradient to the same objective (Proposition 1), and swaps depend only on the learned CV free energy.

Extended-variable methods. Temperature-accelerated MD, driven adiabatic free-energy dynamics, and related schemes couple the CVs to hot auxiliary variables through a stiff spring. They recover the free energy in the limit of adiabatic separation between auxiliary and molecular motion (Maragliano & Vanden-Eijnden, 2006; Abrams & Tuckerman, 2008; Rosso et al., 2002; Chen et al., 2012; Awasthi & Nair, 2017), possibly followed by mean-force reconstruction (Maragliano & Vanden-Eijnden, 2008). VESPER's auxiliary walkers are not coupled to the molecular coordinates: they sample the tempered model and interact with the MD only through the shared parameters, so the free energy is identified by the variational principle rather than an adiabatic limit, and the remaining requirement is a bound on the learning rate (Appendix A.6).

Energy-based models and tempered objectives. With $g = \beta G$ and $\gamma = \alpha - 1$, the single-level objective is the pseudo-spherical contrastive divergence (PS-CD) objective (Yu et al., 2021), a member of the γ-cross-entropy family (Fujisawa & Eguchi, 2008), and admits a Rényi variational representation (Birrell et al., 2021); at $\alpha = 1$ it reduces to maximum likelihood. These objectives are used with a fixed dataset, and PS-CD reweights model samples to estimate the tempered expectation. VESPER instead samples both tilted laws directly: the tempered model by persistent auxiliary walkers, as in contrastive divergence (Hinton, 2002) with persistent chains (Tieleman, 2008), and the tilted data law by biased MD, which removes the need for equilibrium data (Appendix D). Parallel tempering has also been used to train restricted Boltzmann machines (Desjardins et al., 2010; Cho et al., 2010), with only the untempered chain entering the gradient; in VESPER every level contributes.

Learned samplers. Boltzmann generators and non-equilibrium transport samplers learn maps or flows that sample a known energy (Noé et al., 2019; Dibak et al., 2022; Invernizzi et al., 2022; Albergo & Vanden-Eijnden, 2025), and annealed importance sampling and Jarzynski-type estimators weight paths between distributions (Neal, 2001; Jarzynski, 1997; Carbone et al., 2023). These methods target the full configurational distribution; VESPER targets a CV free energy through directly simulated tilted populations, and the two are complementary.

3 Numerical results

We test VESPER on six systems. For the molecular systems, times are per replica, so a K-level ladder uses K times the MD of a single level; on the analytic landscapes every method has the same number of walkers. Table 1 lists the levels and walkers of each system and Table 2 the MD per seed of every method. Errors are reported with the additive constant removed (Appendix B.8).

3.1 Analytic landscapes: balancing basins and barriers, and scaling with dimension

We first use analytic landscapes, where the free energy is known exactly, to ask whether a ladder can be accurate at once in the low-energy basins and in the high-energy regions between them, and whether that accuracy survives as the number of biased coordinates grows. On the Müller–Brown surface (Müller & Brown, 1979), shown in Fig. 2A–C, the two error measures pull in opposite directions. The RMSE over the full box is dominated by the high-energy regions that only strongly tempered walkers visit and is lowest for $\alpha = 0.05$ ($237.7 \pm 2.5\, k_B T$ at $t = 80$, three seeds) and highest for $\alpha = 0.5$ (243.9 ± 0.5); the Boltzmann-weighted $D_{\mathrm{KL}}(\rho_\theta \,\|\, \rho)$ is dominated by the basins and inverts that order (0.010 for $\alpha = 0.5$ and 0.25, 0.10 for $\alpha = 0.05$). No single level is good on both. The exchange ladder, with no α to choose, is within 1% of the best RMSE (240.0 ± 1.2) and within a factor of two of the best divergence (0.017 ± 0.013): its cold levels fit the basins, its hot levels fit the barriers, and all train one model. Unbiased sampling ($\alpha = 1$) never leaves the starting basin (Appendix C.2).

To test scaling, we use a rotated four-well landscape with $13.5\, k_B T$ barriers in $d = 2$ to 100 biased coordinates, with one network and one budget at every d, so that neither capacity nor budget changes with dimension (Appendix C.3). We report the RMSE normalized by the standard deviation σ of the reference free energy (1 is a flat surface), the divergence, and the time at which the error first stays below $0.25\,\sigma$ (Fig. 2D–F; three seeds). The four-level exchange ladder ($\alpha = 0.5$, 0.25, 0.1, 0.05) works at every dimension: its normalized error stays between 0.19 and 0.42 from $d = 2$ to 100, within a factor 1.6 of the best single level everywhere, and at $d = 100$ it is the most accurate arm, with a normalized error of 0.20 against 0.42 for the best single level, and the only one to reach $0.25\,\sigma$ within the budget. Every single level is best at some dimension and fails at another: $\alpha = 0.05$ is best at $d = 2$ and worse than a flat surface from $d = 5$; $\alpha = 0.25$ ties for best at $d = 10$ and loses the landscape at $d = 100$; $\alpha = 0.5$ is the best single level from $d = 5$ up, tied with $\alpha = 0.25$ at $d = 10$, but the worst tempered level at $d = 2$; $\alpha = 1$ fails at every d (normalized error from 1.2 to above 10^4). The divergence ranks them similarly. Exchange, not population, carries the result in high dimension: the ladder without exchange is as accurate or more accurate from $d = 5$ to 50 but reaches only 0.46 at $d = 100$, and ladders with four times the walkers per level reach 0.44 with exchange and 0.39 without.

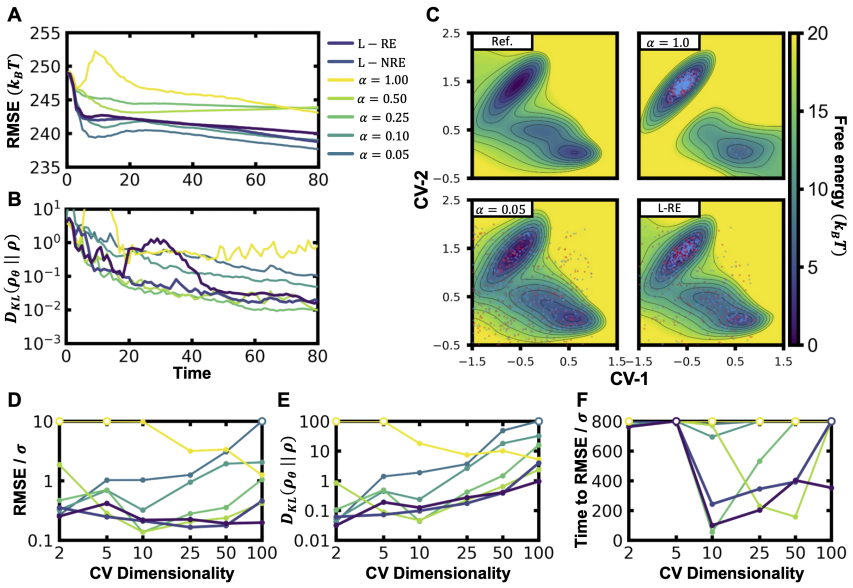


Figure 2: Analytic landscapes. (A) RMSE over the full evaluation box and (B) $D_{\mathrm{KL}}(\rho_\theta \,\|\, \rho)$ against reduced time on the Müller–Brown surface (three-seed means, k_BT) for single levels at $\alpha = 1$, 0.5, 0.25, 0.1, 0.05 and a five-level ladder with (L-RE) and without (L-NRE) exchange, 1,000 walkers per population. (C) Reference, and VESPER surfaces at $t = 80$ for $\alpha = 1$, $\alpha = 0.05$ and the exchange ladder on a common 0–20 k_BT scale; red and blue dots are physical and auxiliary walkers. (D–F) Four-well landscape with 13.5 k_BT barriers in $d = 2$–100 biased coordinates, one width-1024 network and budget at every d, three-seed medians, ladders with four levels ($\alpha = 0.5$, 0.25, 0.1, 0.05): (D) RMSE over the standard deviation σ of the reference free energy (1 is a flat surface), (E) $D_{\mathrm{KL}}(\rho_\theta \,\|\, \rho)$, (F) reduced time at which the error first stays below $0.25\,\sigma$; open markers: off-scale or not reached.

3.2 Alanine dipeptide: comparison with established methods

Alanine dipeptide (ADP) in explicit water, with the two backbone dihedrals (ϕ, ψ) as CVs, is a standard benchmark with a reference surface from 100 ns of adaptive-biasing-force sampling (Zubieta Rico et al., 2025). We score all methods by the RMSE over the full 100×100 periodic grid after removing the mean offset. Unbiased VESPER ($\alpha = 1$) never leaves its initial basin and remains inaccurate at 10 ns (13.27 ± 0.26 kJ mol^{-1}). With tempering, VESPER explores the full (ϕ, ψ) plane and recovers the reference FES (Fig. 3). At 10 ns per replica, the RMSE is 2.19 ± 0.02 kJ mol^{-1} for a single level at $\alpha = 0.01$, and 2.02 ± 0.24 and 2.05 ± 0.09 kJ mol^{-1} for four-level ladders with and without exchange. Metadynamics reaches 2.61 ± 0.22, (neural) VES 4.72 ± 0.25, and SIREN-based combined force-frequency sampling (S-CFF) 2.04 ± 0.51 kJ mol^{-1}. Metadynamics and S-CFF converge fastest initially, but VESPER is more accurate at late times and on par with the best baseline. The ladders, at four times the aggregate MD, give no significant gain here: in two dimensions a single level at $\alpha = 0.01$ already flattens the landscape almost completely.

3.3 Alanine tetrapeptide: increasing the number of CVs

To test how VESPER scales with the number of CVs, we bias three (ϕ_1, ϕ_2, ϕ_3) and then all six $(\phi_1, \phi_2, \phi_3, \psi_1, \psi_2, \psi_3)$ backbone dihedrals of alanine tetrapeptide (ATP) in vacuum. The two models share one architecture and differ only in the input layer (163 and 247 parameters). No converged reference exists in six dimensions, so we test self-consistency: the two-dimensional Boltzmann marginals of the three-CV and six-CV models should agree, and they do (Fig. 6). On the (ϕ_1, ϕ_2), (ϕ_1, ϕ_3) and (ϕ_2, ϕ_3) planes, the squared Pearson correlations between the two models are 0.951, 0.895 and 0.890, with RMSEs of 5.5, 8.1 and 7.5 kJ mol^{-1}. On this system the ladder becomes relevant. After 5 ns per replica, the six-dimensional surfaces of the four-level ladder correlate with their own 10 ns estimates at $r = 0.913 \pm 0.012$ with exchange and 0.941 ± 0.028 without, against 0.804 ± 0.179 for a single level. The ladder thus converges faster and more consistently across seeds,

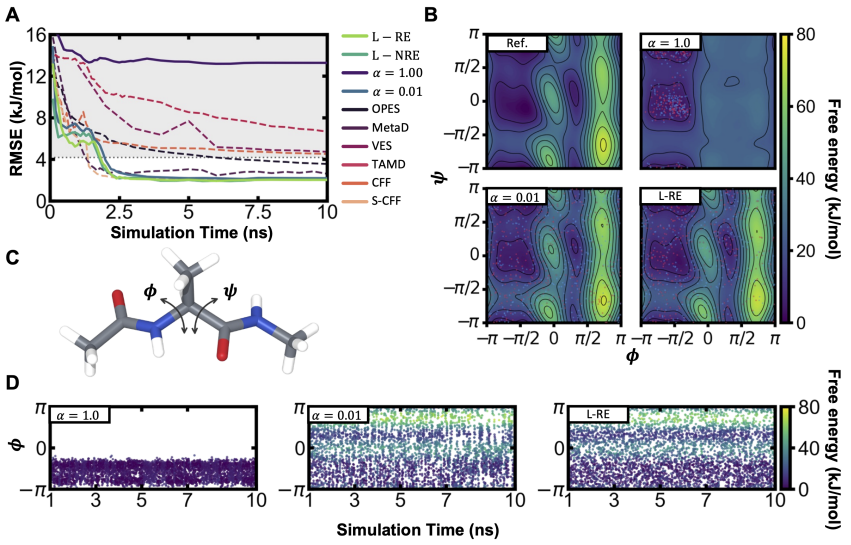


Figure 3: Alanine dipeptide. (A) RMSE against the reference surface (Zubieta Rico et al., 2025) over 0–10 ns per replica, averaged over three seeds; VESPER solid, baselines dashed; grey shading above 1 kcal mol^{-1} (4.184 kJ mol^{-1}). L-RE, L-NRE: four-level ladders with and without exchange; MetaD: metadynamics; CFF, S-CFF: combined force-frequency sampling and its SIREN variant. (B) Reference (top left), and VESPER surfaces for $\alpha = 1$ (top right), a single level at $\alpha = 0.01$ (bottom left) and the exchange ladder (bottom right), on a common scale; red and blue dots are physical and auxiliary walkers. (C) ADP with ϕ, ψ labelled (Humphrey et al., 1996). (D) ϕ of the physical walkers over 1–10 ns, coloured by the learned free energy.

at four times the aggregate MD. On the (ϕ_1, ϕ_2) plane, the exchange ladder also agrees best with the three-CV model ($r^2 = 0.951$, versus 0.932 without exchange and 0.927 for a single level).

3.4 Methane dehydrogenation on Ni(511)

We next test if VESPER at a single level at $\alpha = 0.05$ can resolve a surface catalytic reaction driven by a machine-learned interatomic potential with three CVs. We study C–H bond breaking of methane on a stepped Ni(511) surface at 700 K, using an Allegro potential (Musaelian et al., 2023). Two of the CVs are the C–Ni and C–H coordination numbers of Jin et al. (2025); the third, S_1, the fractional position of the C atom along the first surface lattice vector, distinguishes the step site (edge Ni atoms at $S_1 = 0.70$) from the terraces. In all three seeds the C–H bond breaks within 0.4 ns and, in two of the seeds, re-forms about 0.1 ns later. The learned surface reproduces the physisorbed and dissociated basins and the transition region reported by Jin et al. (2025) (Fig. 4). At 7.5 ns, the free-energy difference between the physisorbed state and the transition region across three seeds is 1.12 ± 0.25 eV, which agrees with the reference value of 1.2 eV (Jin et al., 2025). Prior work could bias only the two coordination numbers and located the reaction site by reweighting observables. With S_1 as a third CV, the learned surface shows directly that the reaction occurs at the step, with the transition state at $S_1 \approx 0.65$, next to the edge atoms, consistent with their result.

3.5 Chignolin: seventeen CVs

To test a high-dimensional molecular system, we learn a free energy over the seventeen backbone dihedrals of chignolin (CLN025 in TIP3P water at 340 K; Appendix B.3). One network and a twelve-level geometric ladder from $\alpha = 1$ to $1/18$, with exchange, are run over four seeds. High-dimensional VES has been applied to chignolin (Shaffer et al., 2016); here one model is trained with data from every level, and the ladder unfolds and refolds chignolin (Fig. 5A). Every level starts folded; averaged over seeds, unfolded levels outnumber folded ones after 3.9 ns per replica. Following individual configurations, which exchange moves between levels without altering them, the four seeds unfold chignolin 48 times and refold it 9 times within 25 ns per replica (each state held for 20 ps). The unbiased level stays folded (99.3% of frames within 0.20 nm), the folded fraction

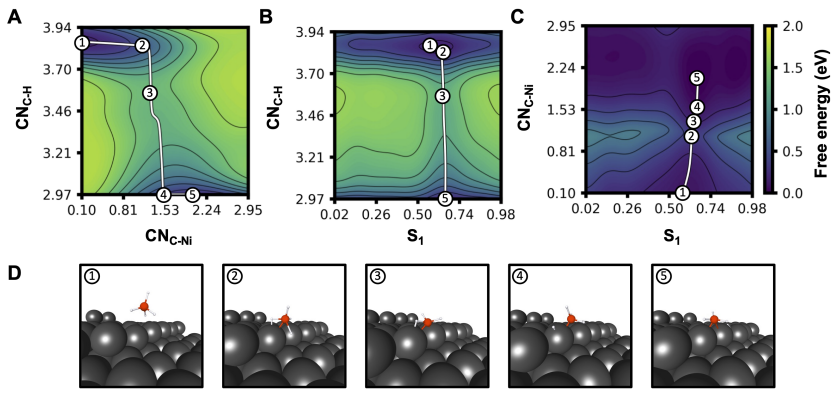


Figure 4: Methane on Ni(511): three-CV free energy. (A–C) Two-dimensional Boltzmann marginals in the planes (C–Ni, C–H coordination), (S_1, C–H coordination) and (S_1, C–Ni coordination). White line: path on the mean surface computed with the string method (E et al., 2007), from reactant (1) through the saddle (3) to product (5). (D) Trajectory frames corresponding to the labelled points.

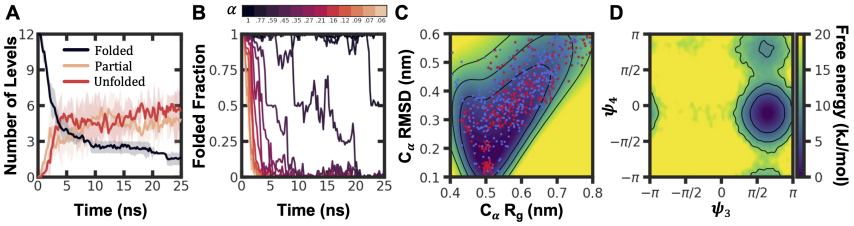


Figure 5: Chignolin with seventeen dihedral CVs (four seeds; twelve levels from $\alpha = 1$ to $1/18$). (A) Number of levels holding a folded (C_α RMSD below 0.20 nm), partially folded or unfolded (above 0.40 nm) configuration, mean and range over seeds, and (B) folded fraction of each level, over the first 25 ns per replica. (C, D) Learned free energy of the unbiased level at the end of the run in C_α radius of gyration and RMSD (dots: MD frames, red, and auxiliary walkers, blue, from all levels) and in (ψ_3, ψ_4); contours every 5 kJ mol^{-1}.

falls along the ladder (Fig. 5B), and exchange acceptance between adjacent levels is 30–93% (50% next to the unbiased level). The learned free energy of the unbiased level has its minimum in the folded region and a single valley, below 5 kJ mol^{-1} out to an RMSD of about 0.5 nm, along which most MD frames and auxiliary walkers of all levels lie (Fig. 5C). In (ψ_3, ψ_4), its minimum lies 0.07 rad from that of the unbiased MD (Fig. 5D).

4 Discussion and Conclusion

VESPER learns a CV free energy while generating its own training data: a neural model G_θ biases the MD, auxiliary walkers sample tempered copies of it, and the two populations give the gradient of one variational objective. Since every tempering level has the physical free energy as its minimizer, all levels train one model, and the tempering strength sets how errors are weighted rather than what is learned. In practice, the pooled ladder is near the best single level at every dimension without tuning α, and beats it in high dimension. VESPER is on par with the best established method on alanine dipeptide, and converges faster and more consistently with a ladder on six-CV alanine tetrapeptide. A small grid-free network also resolves a three-CV surface reaction and seventeen-CV chignolin.

Like any CV-based method, VESPER accelerates only barriers along the CVs; orthogonal slow motions still limit the physical walkers. Our convergence analysis assumes overdamped dynamics and a Polyak–Łojasiewicz condition and excludes exchange, batch reuse and practical optimizers; a K-level ladder costs K times the MD of one level, which pays off in high dimensions but not in low ones. In high dimensions, validation rests on self-consistency rather than converged references, and the stability controls for batch reuse (Appendix B.7) add per-system hyperparameters. Natural extensions are joint CV learning, adaptive level weights, and combining online sampling with existing trajectories (Appendix A.9).

REPRODUCIBILITY STATEMENT

The main text establishes the shared-target ladder result; Appendix A.6 states and proves a conditional convergence theorem for coupled population dynamics. Full assumptions, stationary laws and weighted estimators are derived in Appendix A. Physical protocols, reference surfaces and scoring conventions are specified in Appendices B and B.9; controlled-landscape protocols are given in Appendix C, and the algorithm is given in Appendix E. Code, run configurations and analysis scripts will be released publicly upon publication.

AI USE STATEMENT

In this work, we used generative AI tools under the authors' direction to help implement methods in code, run experiments, and produce figures for results. We also used them to check the proofs in the appendix. We also used them to draft and edit parts of the paper and to search and summarize literature. We have reviewed all AI-assisted work: code was verified against analytical references and unit tests, reported numbers were checked against the runs that produced them, and cited works were checked against their sources. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work develops and evaluates a computational method for free-energy estimation in molecular simulation. It involves no human subjects, personal data, or field data; all molecular systems are standard published benchmarks (alanine di- and tetrapeptide, chignolin, methane on a nickel surface) simulated with public force fields and a machine-learned interatomic potential from the cited literature, and all other results are on analytic landscapes. The method is general-purpose sampling and estimation software; we see no direct dual-use risk beyond that of molecular simulation in general. The compute used is reported in the appendix.

REFERENCES

Jerry B. Abrams and Mark E. Tuckerman. Efficient and direct generation of multidimensional free energy surfaces via adiabatic dynamics without coordinate transformations. *The Journal of Physical Chemistry B*, 112(49):15742–15757, December 2008. ISSN 1520-6106. doi: 10.1021/jp805039u.

Roman Affentranger, Ivano Tavernelli, and Ernesto E. Di Iorio. A novel Hamiltonian replica exchange MD protocol to enhance protein conformational space sampling. *Journal of Chemical Theory and Computation*, 2(2):217–228, 2006. doi: 10.1021/ct050250b.

Michael S. Albergo and Eric Vanden-Eijnden. NETS: A non-equilibrium transport sampler. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 1026–1055. PMLR, 2025.

Shalini Awasthi and Nisanth N. Nair. Exploring high dimensional free energy landscapes: Temperature accelerated sliced sampling. *The Journal of Chemical Physics*, 146(9):094108, 2017. doi: 10.1063/1.4977704.

Shitanshu Bajpai, Brian K. Petkov, Muchen Tong, Charlles R. A. Abreu, Nisanth N. Nair, and Mark E. Tuckerman. An interoperable implementation of collective-variable based enhanced sampling methods in extended phase space within the OpenMM package. *Journal of Computational Chemistry*, 44(28):2166–2183, 2023. ISSN 1096-987X. doi: 10.1002/jcc.27182.

Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Physical Review Letters*, 100(2):020603, January 2008. doi: 10.1103/PhysRevLett.100.020603.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, 2022. doi: 10.1038/s41467-022-29939-5.

Jeremiah Birrell, Paul Dupuis, Markos A. Katsoulakis, Luc Rey-Bellet, and Jie Wang. Variational representations and neural network estimation of Rényi divergences. *SIAM Journal on Mathematics of Data Science*, 3(4):1093–1116, 2021. doi: 10.1137/20M1368926.

Luigi Bonati, Yue-Yu Zhang, and Michele Parrinello. Neural networks-based variationally enhanced sampling. *Proceedings of the National Academy of Sciences*, 116(36):17641–17647, September 2019. doi: 10.1073/pnas.1907975116.

Davide Carbone, Mengjian Hua, Simon Coste, and Eric Vanden-Eijnden. Efficient training of energy-based models using Jarzynski equality. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Ming Chen, Michel A. Cuendet, and Mark E. Tuckerman. Heating and flooding: a unified approach for rapid generation of free energy surfaces. *The Journal of Chemical Physics*, 137(2):024102, 2012. doi: 10.1063/1.4733389.

KyungHyun Cho, Tapani Raiko, and Alexander Ilin. Parallel tempering is efficient for learning restricted Boltzmann machines. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 2010. doi: 10.1109/IJCNN.2010.5596837.

Eric Darve and Andrew Pohorille. Calculating free energies using average force. *The Journal of Chemical Physics*, 115(20):9169–9183, November 2001. ISSN 0021-9606. doi: 10.1063/1.1410978.

Eric Darve, David Rodríguez-Gómez, and Andrew Pohorille. Adaptive biasing force method for scalar and vector free energy calculations. *The Journal of Chemical Physics*, 128(14):144120, April 2008. ISSN 0021-9606. doi: 10.1063/1.2829861.

Guillaume Desjardins, Aaron Courville, Yoshua Bengio, Pascal Vincent, and Olivier Delalleau. Tempered Markov chain Monte Carlo for training of restricted Boltzmann machines. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 9 of *Proceedings of Machine Learning Research*, pp. 145–152, 2010. URL https://proceedings.mlr.press/v9/desjardins10a.html.

Manuel Dibak, Leon Klein, Andreas Krämer, and Frank Noé. Temperature steerable flows and Boltzmann generators. *Physical Review Research*, 4(4):L042005, 2022. doi: 10.1103/PhysRevResearch.4.L042005.

Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *The Journal of Chemical Physics*, 126(16):164103, April 2007. ISSN 0021-9606. doi: 10.1063/1.2720838.

David J. Earl and Michael W. Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–3916, 2005. doi: 10.1039/B509983H.

Peter Eastman, Jason Swails, John D. Chodera, Robert T. McGibbon, Yutong Zhao, Kyle A. Beauchamp, Lee-Ping Wang, Andrew C. Simmonett, Matthew P. Harrigan, Chaya D. Stern, Rafal P. Wiewiora, Bernard R. Brooks, and Vijay S. Pande. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology*, 13(7):e1005659, 2017. doi: 10.1371/journal.pcbi.1005659.

Hironori Fujisawa and Shinto Eguchi. Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis*, 99(9):2053–2081, 2008. doi: 10.1016/j.jmva.2008.02.004.

Alejandro Gil-Ley and Giovanni Bussi. Enhanced conformational sampling using replica exchange with collective-variable tempering. *Journal of Chemical Theory and Computation*, 11(3):1077–1085, 2015. doi: 10.1021/ct5009087. Correction: J. Chem. Theory Comput. 11, 5554 (2015), doi:10.1021/acs.jctc.5b00981.

Ashley Z. Guo, Emre Sevgen, Hythem Sidky, Jonathan K. Whitmer, Jeffrey A. Hubbell, and Juan J. de Pablo. Adaptive enhanced sampling by force-biasing using neural networks. *The Journal of Chemical Physics*, 148(13):134108, 2018. doi: 10.1063/1.5020733.

Jérôme Hénin, Tony Lelièvre, Michael R. Shirts, Omar Valsson, and Lucie Delemotte. Enhanced sampling methods for molecular dynamics simulations. *Living Journal of Computational Molecular Science*, 4(1):1583, 2022. doi: 10.33011/livecoms.4.1.1583.

Geoffrey E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14(8):1771–1800, 2002. doi: 10.1162/089976602760128018.

Richard Holley and Daniel Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. *Journal of Statistical Physics*, 46(5-6):1159–1194, 1987. doi: 10.1007/BF01011161.

Shinya Honda, Toshio Akiba, Yusuke S. Kato, Yoshito Sawada, Masakazu Sekijima, Miyuki Ishimura, Ayako Ooishi, Hideki Watanabe, Takayuki Odahara, and Kazuaki Harata. Crystal structure of a ten-amino acid protein. *Journal of the American Chemical Society*, 130(46):15327–15331, 2008. doi: 10.1021/ja8030533.

Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006. doi: 10.1002/prot.21123.

William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996. doi: 10.1016/0263-7855(96)00018-5.

Michele Invernizzi and Michele Parrinello. Rethinking metadynamics: From bias potentials to probability distributions. *The Journal of Physical Chemistry Letters*, 11(7):2731–2736, 2020. doi: 10.1021/acs.jpclett.0c00497.

Michele Invernizzi, Andreas Krämer, Cecilia Clementi, and Frank Noé. Skipping the replica exchange ladder with normalizing flows. *The Journal of Physical Chemistry Letters*, 13(50):11643–11649, 2022. doi: 10.1021/acs.jpclett.2c03327.

Pavel Izmailov, Dmitrii Podoprikhin, Timur Garipov, Dmitry Vetrov, and Andrew Gordon Wilson. Averaging weights leads to wider optima and better generalization. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 876–885, 2018.

Christopher Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690–2693, 1997. doi: 10.1103/PhysRevLett.78.2690.

Yezhi Jin, Yinan Xu, Jireh S. García Sánchez, Gustavo R. Pérez-Lemus, Pablo F. Zubieta Rico, Massimiliano Delferro, and Juan J. de Pablo. Free-energy landscapes and surface dynamics in methane activation on Ni(511) via machine learning and enhanced sampling. *ACS Catalysis*, 15 (11):8931–8942, June 2025. doi: 10.1021/acscatal.5c00724.

William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983. doi: 10.1063/1.445869.

Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923, 1976. doi: 10.1107/S0567739476001873.

Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, October 2002. doi: 10.1073/pnas.202427399.

Kaizhao Liang, Lizhang Chen, Bo Liu, and Qiang Liu. Cautious optimizers: Improving training with one line of code. In *International Conference on Learning Representations*, 2026.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, 2015. doi: 10.1021/acs.jctc.5b00255.

Luca Maragliano and Eric Vanden-Eijnden. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chemical Physics Letters*, 426(1-3):168–175, July 2006. ISSN 0009-2614. doi: 10.1016/j.cplett.2006.05.062.

Luca Maragliano and Eric Vanden-Eijnden. Single-sweep methods for free energy calculations. *The Journal of Chemical Physics*, 128(18):184110, 2008. doi: 10.1063/1.2907241.

Klaus Müller and Leo D. Brown. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theoretica Chimica Acta*, 53(1):75–93, 1979. doi: 10.1007/BF00547608.

Albert Musaelian, Simon Batzner, Anders Johansson, Lixin Sun, Cameron J. Owen, Mordechai Kornbluth, and Boris Kozinsky. Learning local equivariant representations for large-scale atomistic dynamics. *Nature Communications*, 14(1):579, 2023. doi: 10.1038/s41467-023-36329-y.

Radford M. Neal. Annealed importance sampling. *Statistics and Computing*, 11(2):125–139, 2001. doi: 10.1023/A:1008923215028.

Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019. doi: 10.1126/science.aaw1147.

Stefano Piana and Alessandro Laio. A bias-exchange approach to protein folding. *The Journal of Physical Chemistry B*, 111(17):4553–4559, 2007. doi: 10.1021/jp067873l.

Boris T. Polyak and Anatoli B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, 30(4):838–855, 1992. doi: 10.1137/0330046.

Lula Rosso, Peter Mináry, Zhongwei Zhu, and Mark E. Tuckerman. On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles. *The Journal of Chemical Physics*, 116(11):4389–4402, 2002. doi: 10.1063/1.1448491.

David Ruppert. Efficient estimations from a slowly convergent Robbins-Monro process. Technical Report 781, School of Operations Research and Industrial Engineering, Cornell University, 1988.

Elia Schneider, Luke Dai, Robert Q. Topper, Christof Drechsel-Grau, and Mark E. Tuckerman. Stochastic neural network approach for learning high-dimensional free energy surfaces. *Physical Review Letters*, 119(15):150601, 2017. doi: 10.1103/PhysRevLett.119.150601.

Emre Sevgen, Ashley Z. Guo, Hythem Sidky, Jonathan K. Whitmer, and Juan J. de Pablo. Combined force-frequency sampling for simulation of systems having rugged free energy landscapes. *Journal of Chemical Theory and Computation*, 16(3):1448–1455, 2020. doi: 10.1021/acs.jctc.9b00883.

Patrick Shaffer, Omar Valsson, and Michele Parrinello. Enhanced, targeted sampling of high-dimensional free-energy landscapes using variationally enhanced sampling, with an application to chignolin. *Proceedings of the National Academy of Sciences*, 113(5):1150–1155, 2016. doi: 10.1073/pnas.1519712113.

Hythem Sidky and Jonathan K. Whitmer. Learning free energy landscapes using artificial neural networks. *The Journal of Chemical Physics*, 148(10):104111, 2018. doi: 10.1063/1.5018708.

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems*, volume 33, pp. 7462–7473, 2020.

Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2):141–151, 1999. ISSN 0009-2614. doi: 10.1016/S0009-2614(99)01123-9.

Tijmen Tieleman. Training restricted Boltzmann machines using approximations to the likelihood gradient. In *Proceedings of the 25th International Conference on Machine Learning (ICML)*, pp. 1064–1071, 2008. doi: 10.1145/1390156.1390290.

Glenn M. Torrie and John P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, February 1977. ISSN 0021-9991. doi: 10.1016/0021-9991(77)90121-8.

Omar Valsson and Michele Parrinello. Variational approach to enhanced sampling and free energy calculations. *Physical Review Letters*, 113(9):090601, August 2014. ISSN 1079-7114. doi: 10.1103/PhysRevLett.113.090601.

Omar Valsson and Michele Parrinello. Well-tempered variational approach to enhanced sampling. *Journal of Chemical Theory and Computation*, 11(5):1996–2002, 2015. doi: 10.1021/acs.jctc.5b00076.

Lantao Yu, Jiaming Song, Yang Song, and Stefano Ermon. Pseudo-spherical contrastive divergence. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL `https://proceedings.neurips.cc/paper_files/paper/2021/file/bc5fcb0018cecacba559dc512740091b-Paper.pdf`.

Linfeng Zhang, Han Wang, and Weinan E. Reinforced dynamics for enhanced sampling in large atomic and molecular systems. *The Journal of Chemical Physics*, 148(12):124113, 2018. doi: 10.1063/1.5019675.

Pablo F. Zubieta Rico, Ludwig Schneider, Gustavo R. Pérez-Lemus, Riccardo Alessandri, Siva Dasetty, Trung D. Nguyen, Cintia A. Menéndez, Yiheng Wu, Yezhi Jin, Yinan Xu, Samuel Varner, John A. Parker, Andrew L. Ferguson, Jonathan K. Whitmer, and Juan J. de Pablo. PySAGES: flexible, advanced sampling methods accelerated with GPUs. *npj Computational Materials*, 10(1):35, February 2024. ISSN 2057-3960. doi: 10.1038/s41524-023-01189-z.

Pablo F. Zubieta Rico, Gustavo R. Pérez-Lemus, and Juan J. de Pablo. Efficient sampling of free energy landscapes with functions in Sobolev spaces. *The Journal of Chemical Physics*, 162(8):084109, 2025. doi: 10.1063/5.0221263.

A PROOFS

We retain the notation of the main text, with G in energy units and $\alpha = \hat{\beta}/\beta$. When considering a single tempering level, we write P_G, ν_G, and Q_G for $P_{G,\hat{\beta}}$, $\nu_{G,\hat{\beta}}$, and $Q_{G,\hat{\beta}}$, respectively.

A.1 CONVEXITY AND THE PHYSICAL FREE-ENERGY MINIMIZER

Let p be the physical CV density in equation 1. We work on its support $\mathcal{S}$, where $p > 0$ almost everywhere, and for the chosen $0 < \alpha < 1$ assume that

$$C_\alpha = \int_{\mathcal{S}} p(s)^\alpha \, \mathrm{d}s < \infty. \tag{14}$$

We consider finite-valued measurable free energies G for which the two normalizers

$$A_G = \int_{\mathcal{S}} e^{-\hat{\beta}G(s)} \, \mathrm{d}s, \qquad B_G = \int_{\mathcal{S}} p(s) e^{(\beta-\hat{\beta})G(s)} \, \mathrm{d}s \tag{15}$$

are finite and positive.

The auxiliary density, biased microscopic density and its CV marginal are

$$P_G(s) = \frac{e^{-\hat{\beta}G(s)}}{A_G}, \qquad \nu_G(x) = \frac{e^{(\beta-\hat{\beta})G(s(x))}\varrho(x)}{B_G}, \tag{16}$$

$$Q_G(s) = \frac{p(s)e^{(\beta-\hat{\beta})G(s)}}{B_G}. \tag{17}$$

Proposition S1 (Convexity and the free-energy minimizer)**.** *The functional in equation 2 is convex in G, invariant under addition of a constant, and has minimizer set $\{G_\star + C\}$. Its minimum is*

$$\mathcal{L}_{\hat{\beta}}[G_\star] = \frac{\log C_\alpha}{\beta\alpha(1-\alpha)}. \tag{18}$$

At this minimum, $P_{G_\star} = Q_{G_\star} = p^\alpha / C_\alpha$.

Proof. Each term of equation 2 is convex by Hölder's inequality. Adding a constant c to G changes the two terms by $-c$ and c, so their sum is unchanged. To identify the minimizer, differentiate in a bounded direction h:

$$D\mathcal{L}_{\hat{\beta}}[G](h) = \mathbb{E}_{Q_G} h - \mathbb{E}_{P_G} h, \tag{19}$$

$$D^2\mathcal{L}_{\hat{\beta}}[G](h,h) = \hat{\beta}\,\mathrm{Var}_{P_G}(h) + (\beta - \hat{\beta})\,\mathrm{Var}_{Q_G}(h) \geq 0. \tag{20}$$

The first variation vanishes for every such h exactly when $P_G = Q_G$. Using equation 16 and equation 17, this condition becomes

$$\frac{e^{-\hat{\beta}G(s)}}{A_G} = \frac{p(s)e^{(\beta-\hat{\beta})G(s)}}{B_G} \quad \Longleftrightarrow \quad G(s) = -\beta^{-1}\log p(s) + C.$$

To see that this candidate attains the minimum, apply Hölder's inequality once more:

$$C_\alpha = \int_{\mathcal{S}} (e^{-\hat{\beta}G})^{1-\alpha}(pe^{(\beta-\hat{\beta})G})^{\alpha}\,\mathrm{d}s \leq A_G^{1-\alpha}B_G^{\alpha}.$$

Taking logarithms gives the lower bound in equation 18. Equality holds precisely when the two normalized factors coincide, hence precisely at $G = G_\star + C$. This candidate is admissible by equation 14, and substitution gives $P_{G_\star} = Q_{G_\star} = p^\alpha/C_\alpha$. Finally, applying the chain rule to equation 19 gives the parameter gradient in equation 7, provided differentiation under the integrals is justified. □

A.2 KL CONTROL AND THE PHYSICAL-TEMPERATURE LIMIT

Proposition S2 (Loss and KL control)**.** *Let $\pi_\alpha = p^\alpha/C_\alpha$. Under the assumptions of Proposition S1 and integrability of G and $\log p$ under π_α,*

$$\mathcal{L}_{\hat{\beta}}[G] - \mathcal{L}_{\hat{\beta}}[G_\star] = \frac{\mathrm{KL}(\pi_\alpha\|P_G)}{\hat{\beta}} + \frac{\mathrm{KL}(\pi_\alpha\|Q_G)}{\beta - \hat{\beta}}. \tag{21}$$

In particular,

$$\mathrm{KL}(\pi_\alpha\|P_G) \leq \hat{\beta}\big(\mathcal{L}_{\hat{\beta}}[G] - \mathcal{L}_{\hat{\beta}}[G_\star]\big), \qquad \mathrm{KL}(\pi_\alpha\|Q_G) \leq (\beta - \hat{\beta})\big(\mathcal{L}_{\hat{\beta}}[G] - \mathcal{L}_{\hat{\beta}}[G_\star]\big). \tag{22}$$

At $\hat{\beta} = \beta$, assume $Z_G = \int e^{-\beta G}\,\mathrm{d}s < \infty$, $\mathbb{E}_p|G| < \infty$, and $\mathbb{E}_p|\log p| < \infty$. With $q_G = e^{-\beta G}/Z_G$, the extension in equation 8 satisfies

$$\mathcal{L}_\beta[G] - \mathcal{L}_\beta[G_\star] = \beta^{-1}\,\mathrm{KL}(p\|q_G). \tag{23}$$

Proof. Substituting the expressions for P_G and Q_G into the two divergences gives

$$\mathrm{KL}(\pi_\alpha\|P_G) = \mathbb{E}_{\pi_\alpha}\log\pi_\alpha + \hat{\beta}\mathbb{E}_{\pi_\alpha}G + \log A_G,$$
$$\mathrm{KL}(\pi_\alpha\|Q_G) = \mathbb{E}_{\pi_\alpha}\log\pi_\alpha - \mathbb{E}_{\pi_\alpha}\log p - (\beta - \hat{\beta})\mathbb{E}_{\pi_\alpha}G + \log B_G.$$

Dividing by $\hat{\beta}$ and $\beta - \hat{\beta}$, respectively, and adding cancels the expectations of G. Since $\log\pi_\alpha = \alpha\log p - \log C_\alpha$, the remaining terms are $\mathcal{L}_{\hat{\beta}}[G] - \log C_\alpha/[\beta\alpha(1-\alpha)]$. This proves equation 21; nonnegativity of either divergence gives equation 22.

At the physical endpoint, the definition in equation 8 gives directly

$$\mathrm{KL}(p\|q_G) = \mathbb{E}_p\log p + \beta\mathbb{E}_p G + \log Z_G = \beta\mathcal{L}_\beta[G] + \mathbb{E}_p\log p.$$

Since $q_{G_\star} = p$, subtracting the same identity at $G_\star$ proves equation 23.

To recover this endpoint as a limit of the interior objectives, assume in addition that $\int e^{-\hat{\beta}_0 G}\,\mathrm{d}s < \infty$ for some $0 < \hat{\beta}_0 < \beta$ and $\mathbb{E}_p e^{\varepsilon_0 G} < \infty$ for some $\varepsilon_0 > 0$. The bound $e^{-\hat{\beta}G} \leq e^{-\hat{\beta}_0 G} + e^{-\beta G}$ then gives $\hat{\beta}^{-1}\log A_G \to \beta^{-1}\log Z_G$ by dominated convergence. Writing $\varepsilon = \beta - \hat{\beta} \downarrow 0$, the difference quotient $(e^{\varepsilon G} - 1)/\varepsilon$ decreases to G and lies between the integrable functions G and $(e^{\varepsilon_0 G} - 1)/\varepsilon_0$. Thus $\mathbb{E}_p e^{\varepsilon G} = 1 + \varepsilon\mathbb{E}_p G + o(\varepsilon)$, and the second term tends to $\mathbb{E}_p G$ as required. □

A.3 DUAL FORMULATION AND COUPLED DYNAMICS

For a probability density ρ on $\mathcal{S}$, define its entropy relative to the CV reference measure by $\mathrm{Ent}_{\mathrm{ds}}(\rho) = \int_{\mathcal{S}} \rho \log \rho \, \mathrm{d}s$. Let ν be a probability law on configuration space. Define

$$\mathcal{J}(G, \rho, \nu) = \mathbb{E}_\nu[G \circ s] - \mathbb{E}_\rho G - \frac{\mathrm{Ent}_{\mathrm{ds}}(\rho)}{\hat{\beta}} - \frac{\mathrm{KL}(\nu \| \varrho)}{\beta - \hat{\beta}}. \tag{24}$$

Proposition S3 (Dual formulation and stationary sampling laws)**.** *For $0 < \hat{\beta} < \beta$ and admissible G,*

$$\mathcal{L}_{\hat{\beta}}[G] = \sup_{\rho,\nu} \mathcal{J}(G, \rho, \nu), \qquad \min_G \mathcal{L}_{\hat{\beta}}[G] = \min_G \sup_{\rho,\nu} \mathcal{J}(G, \rho, \nu). \tag{25}$$

The maximizing laws are P_G and ν_G, and the pushforward of ν_G is Q_G. Suppose the potentials and CV map are smooth, the Langevin equations are well posed, and the boundaries are periodic, reflecting, or confining with sufficient decay. The fixed-model dynamics in equation 11 then preserve the configurational laws P_G and ν_G.

Proof. The dual identity follows by expanding the relative entropies:

$$\mathcal{J}(G, \rho, \nu) = \mathcal{L}_{\hat{\beta}}[G] - \frac{\mathrm{KL}(\rho \| P_G)}{\hat{\beta}} - \frac{\mathrm{KL}(\nu \| \nu_G)}{\beta - \hat{\beta}}. \tag{26}$$

The two KL terms are nonnegative and vanish at $\rho = P_G$ and $\nu = \nu_G$, proving equation 25. The expansion is immediate when all terms are finite; the Gibbs variational formula gives the same supremum otherwise.

For fixed ρ and ν, differentiating $\mathcal{J}$ with respect to θ, under the integrals, gives $\mathbb{E}_\nu \nabla_\theta G_\theta \circ s - \mathbb{E}_\rho \nabla_\theta G_\theta$. At the maximizing laws this is $\nabla_\theta \mathcal{L}_{\hat{\beta}}[G_\theta]$. Taking the negative derivative at each level, weighting by c_k, and replacing the expectations by walker averages gives the parameter update in equation 12. Meanwhile, Wasserstein ascent in ρ and ν, with mobilities α/γ_s and $(1-\alpha)/\gamma_x$, gives

$$\partial_t \rho = \gamma_s^{-1} \nabla_s \cdot (\alpha \rho \nabla_s G + \beta^{-1} \nabla_s \rho), \tag{27}$$

$$\partial_t \nu = \gamma_x^{-1} \nabla_x \cdot \big(\nu \nabla_x [U - (1-\alpha) G \circ s] + \beta^{-1} \nabla_x \nu\big). \tag{28}$$

Their particle realizations are the overdamped sampling equations. Adding momenta and Langevin thermostats yields the inertial realization in equation 11 with the same configurational targets.

The CV marginal of ν_G follows from the definition of p: for any bounded test function h,

$$\int_{\mathcal{X}} h(s(x)) \nu_G(x) \, \mathrm{d}x = \frac{1}{B_G} \int_{\mathcal{S}} h(s) p(s) e^{(\beta - \hat{\beta}) G(s)} \, \mathrm{d}s.$$

Thus $s_\# \nu_G = Q_G$. To check invariance, note that the potentials αG and $U - (1-\alpha) G \circ s$ at inverse temperature β have Gibbs factors $e^{-\hat{\beta} G}$ and $e^{-\beta U + (\beta - \hat{\beta}) G \circ s}$. These make the probability currents in equation 27 and equation 28 vanish. For inertial dynamics, the corresponding phase-space laws include Maxwell momentum factors: Hamiltonian transport preserves energy and phase volume, while friction and noise preserve the Maxwell factors. Under ergodicity, time averages therefore recover the required expectations.

The sampling laws also hold at $\hat{\beta} = \beta$, although the dual functional in equation 24 is singular there. The auxiliary potential is then G, with invariant density $P_{G,\beta} \propto e^{-\beta G}$, and the unbiased physical dynamics preserve ϱ. □

A.4 THE COMMON-MINIMIZER LADDER

Proposition S4 (Weighted excess loss and level gradients)**.** *Assume Proposition S1 at every interior level and the moment and normalizer conditions of Proposition S2 at an endpoint level. For $c_k > 0$ with $\sum_k c_k = 1$, the ladder functional is convex and its minimizer set is $\{G_\star + C\}$. Its parameter gradient is given by equation 10. These statements imply Proposition 1.*

Proof. For each level, Propositions S1 and S2 give

$$\Delta_k[G] := \mathcal{L}_{\hat{\beta}_k}[G] - \mathcal{L}_{\hat{\beta}_k}[G_\star] \geq 0, \qquad \Delta_k[G] = 0 \iff G = G_\star + C.$$

This includes the physical endpoint by equation 23. The excess of the combined objective is $\sum_k c_k \Delta_k[G]$. Since every c_k is positive, this sum vanishes exactly when every component vanishes, proving the common-minimizer claim. Each component is convex and invariant under constant shifts, so the weighted sum has both properties as well. At the endpoint, these properties follow directly from equation 8: the log-partition term is convex, the expectation is linear, and their shifts cancel. If the model family contains $G_\star + C$, it attains the same minimum.

Finally, differentiating the finite sum gives

$$\nabla_\theta \mathcal{L}_{\mathrm{VESPER}}[G_\theta] = \sum_k c_k \nabla_\theta \mathcal{L}_{\hat{\beta}_k}[G_\theta].$$

Substitution of equation 7 proves equation 10. At the endpoint, differentiating equation 8 gives $\mathbb{E}_p \nabla_\theta G_\theta - \mathbb{E}_{P_{G_\theta,\beta}} \nabla_\theta G_\theta$, assuming differentiation under the integral is justified. This has the same form, since $\nu_{G,K} = \varrho$ and $s_\# \varrho = p$. Thus each level contributes expectations under its own sampling laws, with coefficient c_k. □

A.5 Replica-exchange acceptance

Proposition S5 (Exchange preserves the level product laws)**.** *Fix G for which all level distributions are normalizable, and consider replicas of the same physical system at inverse temperature β, with auxiliary inverse temperatures $\hat{\beta}_k$. A symmetric proposal to swap two physical configurations, accepted with A_x in equation 13, is reversible with respect to $\prod_k \nu_{G,k}$. The analogous auxiliary swap with acceptance A_s is reversible with respect to $\prod_k P_{G,k}$.*

Proof. For an auxiliary swap between levels i and j, the ratio of the swapped and original product densities is

$$\frac{P_{G,i}(s^j) P_{G,j}(s^i)}{P_{G,i}(s^i) P_{G,j}(s^j)} = \exp\Big[(\hat{\beta}_i - \hat{\beta}_j)(G(s^i) - G(s^j))\Big].$$

For a physical swap, write $G_i = G(s(x^i))$ and $G_j = G(s(x^j))$. The common physical factors $\varrho(x^i)\varrho(x^j)$ and the normalizers cancel, leaving

$$\begin{aligned}\log \frac{\nu_{G,i}(x^j)\nu_{G,j}(x^i)}{\nu_{G,i}(x^i)\nu_{G,j}(x^j)} &= (\beta - \hat{\beta}_i)G_j + (\beta - \hat{\beta}_j)G_i \\ &\quad - (\beta - \hat{\beta}_i)G_i - (\beta - \hat{\beta}_j)G_j \\ &= (\hat{\beta}_i - \hat{\beta}_j)(G_i - G_j).\end{aligned}$$

Thus the two log ratios are

$$\ell_s = (\hat{\beta}_i - \hat{\beta}_j)[G(s^i) - G(s^j)], \qquad \ell_x = (\hat{\beta}_i - \hat{\beta}_j)[G_i - G_j]. \tag{29}$$

In either case, let T denote the swap and Π the product density. Since $T^2 z = z$, the Metropolis acceptance $A(z) = \min\{1, \Pi(Tz)/\Pi(z)\}$ satisfies $\Pi(z)A(z) = \min\{\Pi(z), \Pi(Tz)\} = \Pi(Tz)A(Tz)$. Together with the symmetric proposal, this proves detailed balance. With equal masses and a common thermostat, the Maxwell factors also cancel if momenta are exchanged; leaving momenta in place preserves the same independent Maxwell law. Composing pair-exchange kernels therefore preserves Π. A joint symmetric swap of both populations instead uses the sum $\ell_x + \ell_s$ as its log acceptance ratio. □

A.6 Convergence of the coupled population dynamics

We now consider the overdamped population dynamics with continuous parameter descent and no replica exchange. The aim is to control the objective error together with the lag of the sampling distributions behind their changing targets.

First suppose that all levels satisfy $0 < \alpha_k < 1$, with $c_k > 0$ and $\sum_k c_k = 1$; we treat the physical endpoint below. It is convenient to use dimensionless quantities,

$$g_\theta = \beta G_\theta, \qquad F(\theta) = \beta \mathcal{L}_{\mathrm{VESPER}}[G_\theta], \qquad f_\theta(s) = \nabla_\theta g_\theta(s).$$

so the instantaneous targets are

$$P_k^\theta(s) = \frac{e^{-\alpha_k g_\theta(s)}}{A_k(\theta)}, \qquad \nu_k^\theta(x) = \frac{\varrho(x) e^{(1-\alpha_k) g_\theta(s(x))}}{B_k(\theta)}. \tag{30}$$

Let $\rho_{k,t}$ and $\nu_{k,t}$ denote the evolving auxiliary and physical laws. Suppressing the time subscript inside expectations, their equations are

$$\begin{aligned} \partial_t \rho_k &= d_{s,k} \nabla_s \cdot \left(\rho_k \nabla_s \log \frac{\rho_k}{P_k^{\theta_t}} \right), & d_{s,k} &= (\beta \gamma_{s,k})^{-1}, \\ \partial_t \nu_k &= d_{x,k} \nabla_x \cdot \left(\nu_k \nabla_x \log \frac{\nu_k}{\nu_k^{\theta_t}} \right), & d_{x,k} &= (\beta \gamma_{x,k})^{-1}, \\ \dot\theta_t &= -\eta \sum_k c_k \left(\mathbb{E}_{\nu_k}[f_{\theta_t} \circ s] - \mathbb{E}_{\rho_k} f_{\theta_t} \right). \end{aligned} \tag{31}$$

Here $\eta = (\beta \gamma_\theta)^{-1}$ matches the normalization in equation 12. The physical drift is available directly from $\nabla_x \log \nu_k^\theta = -\beta \nabla_x U + (1 - \alpha_k) \nabla_x (g_\theta \circ s)$.

Measure the sampling lag by

$$H_t = \sum_k c_k \left\{ \frac{\mathrm{KL}(\rho_{k,t} \| P_k^{\theta_t})}{\alpha_k} + \frac{\mathrm{KL}(\nu_{k,t} \| \nu_k^{\theta_t})}{1 - \alpha_k} \right\}. \tag{32}$$

These weights will cancel the tempering factors when we differentiate the relative entropies.

Let $F_* = \min_\theta F(\theta)$ be finite and attained. We make the following assumptions on an open neighborhood $\mathcal{O}$ of the initial sublevel $\{\theta : F(\theta) \le F(\theta_0) + H_0\}$:

- (C1) The normalizers are finite, g_θ is differentiable in θ and smooth enough in space for equation 31, and $\sup_{\theta \in \mathcal{O}, s} \|f_\theta(s)\| \le B < \infty$. Differentiation under the integrals is justified.
- (C2) The evolution is well posed from initial laws with finite relative entropy. Entropy differentiation and integration by parts are valid, with periodic or zero-flux boundaries, or sufficient decay on unbounded domains.
- (C3) The instantaneous laws satisfy uniform log-Sobolev inequalities on their respective spaces:
$$I(\zeta \| P_k^\theta) \ge 2\lambda_{s,k} \mathrm{KL}(\zeta \| P_k^\theta), \qquad I(\zeta \| \nu_k^\theta) \ge 2\lambda_{x,k} \mathrm{KL}(\zeta \| \nu_k^\theta),$$
where $I(\zeta \| \pi) = \int \|\nabla \log(\zeta/\pi)\|^2 \, \mathrm{d}\zeta$. With constants positive and uniform on $\mathcal{O}$, set $a = \min_k \{d_{s,k} \lambda_{s,k}, d_{x,k} \lambda_{x,k}\} > 0$.
- (C4) The objective satisfies the Polyak–Łojasiewicz inequality $\|\nabla F(\theta)\|^2 \ge 2\mu (F(\theta) - F_*)$ on $\mathcal{O}$, for some $\mu > 0$.

In (C3), the physical inequality is required for the full microscopic law. Uniform constants follow, for example, from the Holley–Stroock perturbation lemma when the model gives a uniformly bounded tilt of a reference law satisfying a log-Sobolev inequality (Holley & Stroock, 1987).

Theorem S6 (Convergence of the two-population ladder)**.** *Under (C1)–(C4), for $0 < \eta < a/B^2$, the evolution in equation 31 obeys*

$$F(\theta_t) - F_* + H_t \le e^{-\kappa t} \big[F(\theta_0) - F_* + H_0 \big], \qquad \kappa = 2 \min\{\eta\mu, a - \eta B^2\}. \tag{33}$$

Thus the objective error and the weighted sampling lag decay together.

Proof. Let e denote the error in the population gradient, and write

$$\begin{aligned} H_{s,k} &= \mathrm{KL}(\rho_k \| P_k^{\theta_t}), \qquad H_{x,k} = \mathrm{KL}(\nu_k \| \nu_k^{\theta_t}), \\ e_k &= (\mathbb{E}_{\nu_k} - \mathbb{E}_{\nu_k^{\theta_t}})[f_{\theta_t} \circ s] - (\mathbb{E}_{\rho_k} - \mathbb{E}_{P_k^{\theta_t}}) f_{\theta_t}, \qquad e = \sum_k c_k e_k. \end{aligned}$$

By equation 10, the parameter equation is $\dot{\theta} = -\eta(\nabla F + e)$.

To differentiate H, first note that normalization of the target densities gives

$$\nabla_\theta \log P_k^\theta = -\alpha_k(f_\theta - \mathbb{E}_{P_k^\theta} f_\theta), \qquad \nabla_\theta \log \nu_k^\theta = (1-\alpha_k)(f_\theta \circ s - \mathbb{E}_{\nu_k^\theta}[f_\theta \circ s]).$$

Integration by parts in equation 31 gives

$$\dot{H}_{s,k} = -d_{s,k} I(\rho_k \| P_k^{\theta_t}) + \alpha_k \dot{\theta}^\top (\mathbb{E}_{\rho_k} - \mathbb{E}_{P_k^{\theta_t}}) f_{\theta_t}, \tag{34}$$

$$\dot{H}_{x,k} = -d_{x,k} I(\nu_k \| \nu_k^{\theta_t}) - (1-\alpha_k)\dot{\theta}^\top (\mathbb{E}_{\nu_k} - \mathbb{E}_{\nu_k^{\theta_t}})[f_{\theta_t} \circ s]. \tag{35}$$

Multiplying by c_k/α_k and $c_k/(1-\alpha_k)$, respectively, and summing over levels combines the moving-target terms into $-\dot{\theta}^\top e$. The log-Sobolev inequalities in (C3) therefore give

$$\dot{H} \le -2aH - \dot{\theta}^\top e. \tag{36}$$

The gradient error is controlled by the same entropies. For a vector-valued function f with $\|f\| \le B$, applying Pinsker's inequality to unit scalar projections gives $\|(\mathbb{E}_\zeta - \mathbb{E}_\pi) f\| \le \sqrt{2} B \sqrt{\mathrm{KL}(\zeta \| \pi)}$. Applying this to both populations and then using weighted Cauchy–Schwarz yields

$$\begin{aligned} \|e\|^2 &\le 2B^2 \left[\sum_k c_k (\sqrt{H_{s,k}} + \sqrt{H_{x,k}}) \right]^2 \\ &\le 2B^2 \left[\sum_k c_k \{\alpha_k + (1-\alpha_k)\} \right] \left[\sum_k c_k \left\{ \frac{H_{s,k}}{\alpha_k} + \frac{H_{x,k}}{1-\alpha_k} \right\} \right] = 2B^2 H. \end{aligned} \tag{37}$$

Now let $V = F - F_* + H$. Combining equation 36 with the parameter equation cancels the mixed gradient terms:

$$\begin{aligned} \dot{V} &\le (\nabla F - e)^\top \dot{\theta} - 2aH \\ &= -\eta(\nabla F - e)^\top (\nabla F + e) - 2aH \\ &= -\eta \|\nabla F\|^2 + \eta \|e\|^2 - 2aH \\ &\le -\eta \|\nabla F\|^2 - 2(a - \eta B^2) H \le -\kappa V. \end{aligned}$$

The last inequality uses (C4), and Gronwall's inequality gives equation 33. This estimate also gives $F(\theta_t) \le F_* + V_t \le F(\theta_0) + H_0$. Applying the argument up to a possible first exit from $\mathcal{O}$ therefore rules out such an exit and proves the estimate for all time. □

The physical endpoint. The same argument covers a ladder with $\alpha_K = 1$, where the last physical replica has the fixed target $\nu_K^\theta = \varrho$. The gradient formula in equation 10 still holds under the endpoint moment conditions of Proposition S2, as shown in Proposition S4. To adapt the entropy H, keep the terms for $k < K$ and replace the last pair by

$$c_K \{\mathrm{KL}(\rho_K \| P_K^\theta) + w\, \mathrm{KL}(\nu_K \| \varrho)\}, \qquad w > 0.$$

Write H' for this modified entropy and use (C1)–(C4) on an open neighborhood of $\{\theta : F(\theta) \le F(\theta_0) + H_0'\}$, with the physical log-Sobolev inequality at level K imposed on ϱ.

Since ϱ does not depend on θ, its relative entropy has no moving-target term. Thus, with $e_{x,K} = (\mathbb{E}_{\nu_K} - \mathbb{E}_\varrho)[f_\theta \circ s]$, the entropy estimate becomes

$$\dot{H}' \le -2aH' - \dot{\theta}^\top (e - c_K e_{x,K}).$$

Pinsker's inequality and the same weighted Cauchy–Schwarz estimate give

$$\|e\|^2 \le 2B^2(1 + c_K/w) H', \qquad c_K^2 \|e_{x,K}\|^2 \le 2B^2 (c_K/w) H'.$$

For $V' = F - F_* + H'$, the mixed terms therefore cancel as before, except for the contribution from the last physical replica:

$$\begin{aligned} \dot{V}' &\le -\eta \|\nabla F\|^2 + \eta \|e\|^2 - \eta c_K (\nabla F + e)^\top e_{x,K} - 2aH' \\ &\le -\tfrac{\eta}{2} \|\nabla F\|^2 - \big[2a - \eta B^2 (3 + 5c_K/w)\big] H'. \end{aligned}$$

Here we used Young's inequality in the form $|c_K(\nabla F+e)^\top e_{x,K}| \le \frac{1}{2}\|\nabla F\|^2+\frac{1}{2}\|e\|^2+c_K^2\|e_{x,K}\|^2$. For $0 < \eta < 2a/[B^2(3+5c_K/w)]$, (C4) then gives

$$V_t' \le e^{-\kappa' t}V_0', \qquad \kappa' = \min\{\eta\mu,\, 2a - \eta B^2(3+5c_K/w)\}.$$

As in the proof above, $F(\theta_t) \le F(\theta_0)+H_0'$ keeps the trajectory in the region where the assumptions hold. Without (C4), integrating the bound on $\dot V'$ gives the time-averaged estimate in equation 38, with $2V_0'/(\eta T)$ on the right.

Corollary S7 (Time-averaged stationarity)**.** *Under (C1)–(C3) and $0 < \eta < a/B^2$, for every $T > 0$,*

$$\frac{1}{T}\int_0^T \|\nabla F(\theta_t)\|^2 \,\mathrm{d}t \le \frac{F(\theta_0) - F_* + H_0}{\eta T}. \tag{38}$$

Proof. The proof above gives $\dot V \le -\eta\|\nabla F\|^2 - 2(a-\eta B^2)H$ before invoking (C4). Integrating over $[0,T]$ and using $V_T \ge 0$ proves the bound. □

A.7 General reference measures

The same variational argument applies to a general reference measure. Let r be sigma-finite and let the probability measure p have density $f = \mathrm{d}p/\mathrm{d}r > 0$ r-almost everywhere. For $r = \mathrm{d}s$, this f is the physical CV density $p(s)$. Writing $g = \beta G$, define

$$\ell_\alpha[g] = \frac{1}{\alpha}\log\int e^{-\alpha g}\,\mathrm{d}r + \frac{1}{1-\alpha}\log\int e^{(1-\alpha)g}\,\mathrm{d}p. \tag{39}$$

For the physical Lebesgue reference, $\ell_\alpha[\beta G] = \beta\mathcal{L}_{\hat\beta}[G]$. When $Z_g = \int e^{-g}\,\mathrm{d}r < \infty$, the normalized residual model is

$$q_g(\mathrm{d}y) = Z_g^{-1}e^{-g(y)}r(\mathrm{d}y). \tag{40}$$

Proposition S8 (Reference-measure minimizer)**.** *For $0 < \alpha < 1$ with $M(\alpha) = \int f^\alpha\,\mathrm{d}r < \infty$, the functional in equation 39, restricted to g for which both integrals are finite, is minimized exactly at $g_\star = -\log f + C$. At the minimum, the two tilted laws coincide with $f^\alpha\mathrm{d}r/M(\alpha)$.*

Proof. As in the proof of Proposition S1, Hölder's inequality gives

$$M(\alpha) = \int (e^{-\alpha g})^{1-\alpha}(fe^{(1-\alpha)g})^\alpha\,\mathrm{d}r \le \left(\int e^{-\alpha g}\,\mathrm{d}r\right)^{1-\alpha}\left(\int e^{(1-\alpha)g}\,\mathrm{d}p\right)^{\alpha}.$$

Equality requires the two factors to be proportional, which is equivalent to $g = -\log f + C$. At this choice, both normalized densities relative to r equal $f^\alpha/M(\alpha)$. □

A.8 Positive-weight effective sample size

Proposition S9 (Population normalized ESS)**.** *At the population optimum, the fixed-data weights are proportional to $f^{\alpha-1}$. If $M(2\alpha-1) < \infty$, their population normalized effective sample size is*

$$\frac{(\mathbb{E}_p w)^2}{\mathbb{E}_p w^2} = \frac{M(\alpha)^2}{M(2\alpha-1)}. \tag{41}$$

For independent $y_j \sim p$ with $w_j = w(y_j)$, the empirical quantity $(\sum_j w_j)^2/[n\sum_j w_j^2]$ converges to this value almost surely. At $\alpha = 1/2$, the denominator is $M(0) = r(\operatorname{supp} p)$.

Proof. The scale of the weights cancels in the ratio, so take $w = f^{\alpha-1}$. Since $\mathrm{d}p = f\mathrm{d}r$, we have $\mathbb{E}_p w = \int f^\alpha\,\mathrm{d}r = M(\alpha)$ and $\mathbb{E}_p w^2 = \int f^{2\alpha-1}\,\mathrm{d}r = M(2\alpha-1)$, proving equation 41. The assumed second moment allows the strong law of large numbers to be applied to both empirical moments, giving the almost-sure limit. □

A.9 LEARNING FROM EXISTING DATA

When equilibrium configurations $x_j^* \sim \varrho$ are already available, their CV values $y_j = s(x_j^*)$ can replace the physical walkers. The physical expectation in equation 7 is then estimated using normalized importance weights

$$\bar{w}_j(\theta) = \frac{e^{(\beta-\hat{\beta})G_\theta(y_j)}}{\sum_\ell e^{(\beta-\hat{\beta})G_\theta(y_\ell)}}. \tag{42}$$

This leads to the following weighted-data consistency result.

Proposition S10 (Weighted-data gradient)**.** *Fix θ and let x_j^* be independent samples from ϱ, with $y_j = s(x_j^*)$. Suppose $B_{G_\theta} < \infty$ and $\mathbb{E}_p[e^{(\beta-\hat{\beta})G_\theta}\|\nabla_\theta G_\theta\|] < \infty$. Then the weights in equation 42 satisfy*

$$\sum_{j=1}^{n} \bar{w}_j(\theta)\nabla_\theta G_\theta(y_j) \longrightarrow \mathbb{E}_{\nu_{G_\theta}} \nabla_\theta G_\theta \circ s \quad \textit{almost surely as } n \to \infty. \tag{43}$$

Moreover, the corresponding weighted population objective has minimizer $G_\star + C$.

Proof. Using the CV marginal in equation 17, the required expectation can be written as

$$\mathbb{E}_{\nu_{G_\theta}} \nabla_\theta G_\theta \circ s = \frac{\mathbb{E}_p[e^{(\beta-\hat{\beta})G_\theta}\nabla_\theta G_\theta]}{\mathbb{E}_p[e^{(\beta-\hat{\beta})G_\theta}]}.$$

The assumed moments allow the strong law of large numbers to be applied to both sample averages. Their ratio converges almost surely because the limiting denominator is positive, proving equation 43.

Together with the auxiliary expectation, this yields the gradient in equation 7 of the original population objective, whose minimizer is $G_\star + C$ by Proposition S1. The exponential weighting is essential: using unweighted samples from p would instead give the functional stationarity condition $P_G = p$, equivalently $G = G_\star/\alpha + C$. □

The auxiliary walkers still sample P_{G_θ}, so the resulting update is a weighted-data version of the same learning rule. With $g = \beta G$ and $\gamma = \alpha - 1$, the objective coincides with the PS-CD/γ-cross-entropy objective (Yu et al., 2021; Fujisawa & Eguchi, 2008), with the Rényi variational representation described in Appendix D (Birrell et al., 2021).

The same ratio argument also yields within-population sample-reuse weights. If a batch was generated at G_0, then the current-to-generating density ratios are proportional to

$$w^{(s)} = e^{-\hat{\beta}(G-G_0)}, \qquad w^{(x)} = e^{(\beta-\hat{\beta})(G-G_0)\circ s}. \tag{44}$$

Normalization within each pool removes the unknown partition functions:

$$\mathbb{E}_{P_{G_0}}[w^{(s)} f] = \frac{A_G}{A_{G_0}}\,\mathbb{E}_{P_G} f, \qquad \mathbb{E}_{\nu_{G_0}}[w^{(x)} f \circ s] = \frac{B_G}{B_{G_0}}\,\mathbb{E}_{\nu_G} f \circ s.$$

Hence, if G and G_0 are admissible, f is integrable under the current law, and the batch satisfies a law of large numbers under its generating law, the corresponding self-normalized estimators are consistent. Their weight concentration is monitored by the ESS criterion described in Appendix E.

The single-level update also admits a scalar-objective interpretation in self-consistent well-tempered VES. For $V_\theta = -(1-\alpha)G_\theta$ and the instantaneous target $t_\theta = P_{G_\theta}$, differentiating the VES functional while holding t fixed gives

$$\nabla_\theta \Omega[V_\theta; t]|_{t=t_\theta} = (1-\alpha)\nabla_\theta \mathcal{L}_{\hat{\beta}}[G_\theta]. \tag{45}$$

The target convention and the recursive neural VES update are detailed in Appendix D.1 (Valsson & Parrinello, 2014; 2015; Bonati et al., 2019).

B PHYSICAL SYSTEMS: MODELS AND PROTOCOLS

This appendix gives the model, force field, collective-variable and network settings behind Section 3, together with the trust region, readout and gauge conventions used for every number reported there.

B.1 ALANINE DIPEPTIDE

Explicit-solvent alanine dipeptide, AMBER99SB (Hornak et al., 2006) with TIP3P water (Jorgensen et al., 1983) and particle-mesh Ewald electrostatics, 298.15 K, 2 fs timestep, bonds to hydrogen constrained. The CVs are the backbone dihedrals ϕ, ψ, both periodic. G_θ is a SIREN (Sitzmann et al., 2020) with one hidden layer of width 32, $\omega_0 = 16$ and a periodic input embedding. Each replica carries 10^4 auxiliary walkers with $\gamma_s = 10$ and $\gamma_x = 100$; training uses AdamW (Loshchilov & Hutter, 2019) at learning rate 3×10^{-4}, up to 80 inner iterations per epoch, and epochs of 1000 MD steps, 50 per checkpoint (100 ps). The single replica uses $\alpha = 0.01$. Both four-level ladders use $\alpha = (0.1, 0.05, 0.02, 0.01)$ with equal loss weights. L-RE attempts exchange of physical configurations between adjacent levels every epoch; L-NRE uses the same ladder without exchange. The single replica starts from the same configurations as the OPES runs. The reference surface is the free energy from 100 ns of adaptive-biasing-force sampling by Zubieta Rico et al. (2025), on a 100×100 periodic grid. Figure 3 reports mean-aligned RMSE over this full grid; OPES and TAMD are scored only on cells where their native free-energy estimators are finite. VESPER predictions use parameter averages over the trailing eight available checkpoints, or all available checkpoints at earlier times. RMSE curves average the three seed-specific errors; the displayed surfaces average the mean-aligned predictions and shift the resulting minimum to zero. Physical points overlaid on the maps are drawn from the last nanosecond and auxiliary points from the final saved populations, balanced across seeds and ladder levels.

B.2 ALANINE TETRAPEPTIDE

Gas-phase Ace-$(Ala)_3$-NMe, AMBER99SB (Hornak et al., 2006) without cutoff, 298.15 K, 2 fs timestep, bonds to hydrogen constrained. The three-CV model biases (ϕ_1, ϕ_2, ϕ_3) and the six-CV model adds ψ_1, ψ_2 and ψ_3. G_θ is a SIREN with hidden widths $(14, 4)$, $\omega_0 = 16$ and a periodic input embedding, so the two models differ only in the input layer (163 and 247 parameters). Each replica carries 2×10^4 auxiliary walkers with $\gamma_s = \gamma_x = 100$; training uses AdamW at 3×10^{-4}, up to 80 inner iterations, and epochs of 1000 steps, 100 per checkpoint (200 ps). The single replica uses $\alpha = 0.01$, and the ladder arms are those of ADP. Cooling is off for both models, with constant-step training throughout. Each model is read as the parameter-SWA10 average over checkpoints from 8.2–10.0 ns.

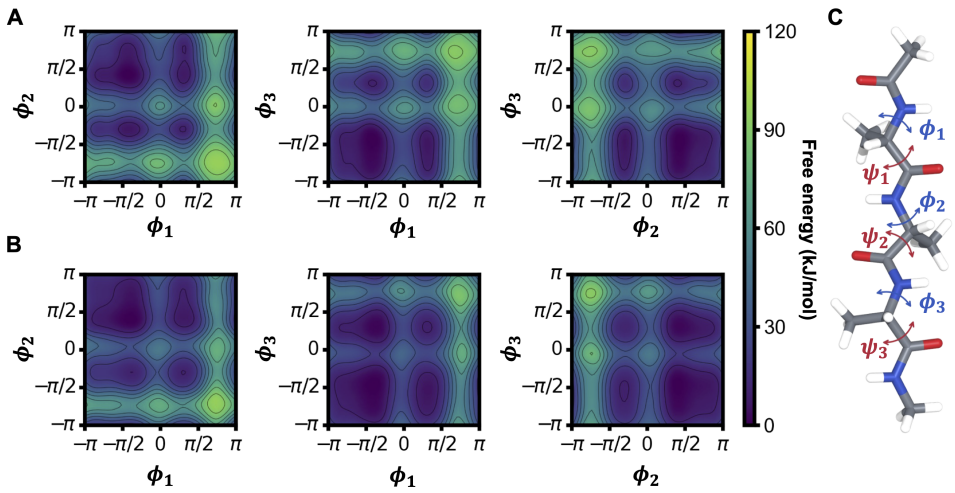


Figure 6: Alanine tetrapeptide: consistency across CV dimension. Two-dimensional Boltzmann marginals of (A) the three-CV and (B) the six-CV model, each from the three-seed mean surface of the four-level ladder with exchange. Columns: (ϕ_1, ϕ_2), (ϕ_1, ϕ_3), (ϕ_2, ϕ_3). (C) ATP with ϕ bonds in blue and ψ bonds in red.

The ATP figure uses the physical Boltzmann marginal at 298.15 K. The three full physical FES predictions are averaged before marginalization. The retained (ϕ_1, ϕ_2), (ϕ_1, ϕ_3) and (ϕ_2, ϕ_3) planes are complete 64×64 grids; six-CV maps integrate omitted ψ coordinates on 32-point axes and

the remaining omitted ϕ on its 64-point grid, while three-CV maps integrate the omitted ϕ on a 256-point grid. Independent IID estimates use 65,536 samples per retained point for three-CV maps and 262,144 for six-CV maps; scrambled-Sobol estimates use the same sample counts. Maximum pooled ensemble RMS discrepancies from the grid are 0.03864 kJ mol^{-1} for IID and 0.00562 kJ mol^{-1} for Sobol. Display surfaces are shifted to zero minimum and use a common 0–120 kJ mol^{-1} scale with 10 kJ mol^{-1} contours. The three-to-six-CV squared-Pearson correlations over all plane cells are 0.951, 0.895 and 0.890.

B.3 CHIGNOLIN

CLN025 (Honda et al., 2008) in explicit solvent, ff14SB (Maier et al., 2015) with TIP3P (Jorgensen et al., 1983), 340 K, 2 fs timestep, with its seventeen backbone dihedrals as CVs; structural deviation is measured by Cα RMSD after Kabsch superposition (Kabsch, 1976). G_θ is a SIREN with hidden widths $(48, 24)$, $\omega_0 = 1$ and a periodic input embedding (2,881 parameters). Twelve replicas span a geometric ladder from $\alpha = 1$ to $\alpha = 1/18$, every level exchanging physical configurations with its neighbours every epoch; each carries 10^5 auxiliary walkers. Training uses AdamW with a gradient clip of $10\,k_BT$, up to 100 inner iterations and epochs of 5000 steps, 20 per checkpoint (200 ps), at constant step throughout. Four seeds run to 26.2 ns per replica, and the readout is the final checkpoint. Frames are saved every 5 ps, $5{,}240$ per level and seed; state fractions, exchange acceptance and event counts pool the four seeds over the whole run, and every unfolding and refolding event falls within the first 25 ns. Folded and unfolded are defined as mean Cα RMSD below 0.20 nm and above 0.40 nm; an unfolding or refolding event requires the new state to persist for 20 ps and is counted along each configuration's identity. The learned surfaces are estimated from $25{,}000$ auxiliary walkers of the unbiased level drawn at random from each seed at the readout ($100{,}000$ in all), and the MD frames shown in Fig. 5C come from the last nanosecond. For the radius-of-gyration and RMSD map, each walker's dihedrals are placed on a backbone with the bond lengths, bond angles, peptide ω and Pro4 ϕ of the native structure, and $F = -k_BT \ln p$ ($k_BT = 2.83$ kJ mol^{-1}), with p an adaptive Gaussian kernel density estimate. The dihedral maps are product von Mises kernel density estimates, compared after mean alignment with the same estimate on the unbiased level's MD frames where the MD free energy is below 10 kJ mol^{-1}. Over (ϕ_2, ψ_2), (ψ_3, ψ_4) and (ϕ_5, ψ_5), the learned minimum lies 0.42, 0.07 and 0.19 rad from the MD minimum, at points where the MD free energy is 1.3, 0.2 and 0.5 kJ mol^{-1} above its minimum, with RMS differences of 1.9, 2.2 and 2.0 kJ mol^{-1}. Accepted exchanges permute configurations between levels, so every configuration keeps its identity (Appendix A.5).

B.4 CATALYTIC METHANE DEHYDROGENATION

Methane on stepped Ni(511) at 700 K uses an Allegro machine-learned interatomic potential (Musaelian et al., 2023), deployed through the NequIP toolchain (Batzner et al., 2022) and driven by OpenMM (Eastman et al., 2017). The CVs are the two coordination numbers of Jin et al. (2025) and the periodic fractional coordinate S_1 of the carbon atom along the first surface lattice vector. The system has 85 atoms, 25 held fixed, in a triclinic cell. A soft restraint keeps the system in the CV ranges: $S_1 \in [0, 1]$, CNi $\in [0.05, 3.02]$, and CH $\in [2.95, 3.98]$. G_θ is a SIREN with hidden widths $(14, 4)$ and $\omega_0 = 16$; one replica at $\alpha = 0.05$ carries 6.5×10^4 auxiliary walkers, with up to 100 inner iterations and epochs of 5000 steps, 20 per checkpoint, at a 0.25 fs timestep, so one checkpoint is 25 ps; gradient clip $50\,k_BT$, $\gamma_s = 100$, $\gamma_x = 10$, Langevin friction 9.82 ps^{-1}, constant step. The reweighted inner step uses the full trust region of Appendix B.7. Three seeds run to checkpoint 300. Each checkpoint-300 physical free-energy prediction uses deployed parameter EMA decay 0.9.

B.5 CATALYSIS FREE-ENERGY PROJECTIONS

With S_1 the periodic fractional coordinate and y, z the C–Ni and C–H coordination numbers, the two-dimensional free energies shown in Fig. 4 are Boltzmann marginals of the equal-weight mean of the three gauge-aligned checkpoint-300 EMA predictions,

$$F_R(q_R) = -k_BT \log \int_{D_O(q_R)} \exp[-G_{\mathrm{mean}}(q_R, q_O)/(k_BT)]\, \mathrm{d}q_O,$$

at $T = 700$ K, with ordinary measure in the original CV coordinates and integration box $D = S_1[0, 1] \times y[0.02, 3] \times z[2.90, 3.94]$, while the visible route uses $S_1[0.02, 0.98] \times y[0.10, 2.95] \times$

Table 1: Levels per molecular system: tempering exponents α, number of levels K, and walkers per level.

System	α	K	Physical	Auxiliary
ADP	0.1, 0.05, 0.02, 0.01	4	1	10^4
ATP, $D = 3, 6$	0.1, 0.05, 0.02, 0.01	4	1	2×10^4
Methane/Ni(511)	0.05	1	1	6.5×10^4
Chignolin	$18^{-k/11}$, $k = 0, \ldots, 11$	12	1	10^5

$z[2.97, 3.94]$. For these surfaces, 128- versus 256-point Gauss–Legendre marginalization differs by at most 2.8×10^{-14} eV at 70 selected coordinates per displayed panel; an independent Monte Carlo check over the same 70 selected coordinates, with 131,072 draws per selected coordinate, has maximum absolute error 0.001124 eV. Each map is shifted to its own minimum for display. The fitted route's $G(3) - G(1) = 1.12$ eV is a bounded surface difference rather than a global or kinetic barrier.

B.6 Ladder configuration by system

Table 1 lists the levels of every molecular system. Each level is one physical trajectory; the auxiliary walkers enter only the training loss. The single-level runs of ADP and ATP use $\alpha = 0.01$, and ADP also has a single-level run at $\alpha = 1$.

B.7 Reweighted inner steps and the step-size controller

Within an epoch the walker samples generated under G_{θ_0} are frozen while several parameter steps are taken and reweighted by self-normalized importance sampling at every step (Section 2.4). A trust region bounds the log-weights: a soft threshold $\delta = 5$ on their 95th percentile, a hard threshold $\delta_{\mathrm{hard}} = 8$ on their maximum, and a floor of 0.2 on the effective sample size of the pool. Exceeding the hard threshold or falling below the floor ends the epoch and forces fresh trajectory data; exceeding the soft threshold multiplies the parameter step by $1/q^2$ with $q = \|\log w\|_{95}/\delta$. The trust region is a stability response to a measured degradation of the importance-sampling estimator, inactive whenever the weights stay inside it, and leaves the variational objective unchanged. It is used on ADP, ATP and catalysis; chignolin trains on fresh samples only.

On ADP, the inner-iteration budget and learning rate are governed by a sample-based controller rather than a predefined schedule. Nothing is keyed to the checkpoint or epoch index: the onset of a decay phase latches when the cumulative parameter travel $\sum_k \|\theta_k - \theta_{k-1}\|$ over checkpoints exceeds a budget of 160, after which the learning rate is multiplied by 0.95 and the inner-iteration count by 0.7 per checkpoint. ATP, chignolin and catalysis use constant-step training; ATP cooling is off for every tested dimension.

B.8 Readout, optimizer settings and RMSE gauge

Parameters trained at constant step do not settle onto a point, so each network is read out as a uniform weight-space average $\bar{\theta}$ over its trailing checkpoints (equation 49), the Polyak–Ruppert or stochastic-weight-averaging readout (Polyak & Juditsky, 1992; Ruppert, 1988; Izmailov et al., 2018): eight checkpoints on ADP and ten on ATP. Catalysis uses the deployed parameter EMA readout with decay 0.9 at checkpoint 300, and chignolin reports its final checkpoint (26.2 ns per replica). AdamW (Loshchilov & Hutter, 2019) uses decoupled weight decay $w = 0.01$ in every run, and the peptide runs apply the cautious mask of Liang et al. (2026), which zeroes coordinates whose Adam step is not descent-aligned with the current gradient and rescales the rest; both leave the VESPER objective of equation 2 unchanged.

RMSE is reported after removing the additive constant. *Mean alignment*, used for ADP, removes the mean difference between the learned and reference surfaces on the scored region; *minimum alignment*, used for catalysis, sets each surface's minimum on that region to zero. Flat-surface baselines, where reported, use the same region and gauge. ATP dimensional-consistency comparisons use complete retained planes, squared-Pearson correlations and mean-aligned RMSEs after physical

Table 2: Replicas K, simulated time per replica at the readout, and aggregate MD per seed (K times the time per replica).

System	Method	K	ns/replica	ns/seed	Seeds
ADP	VESPER, $\alpha = 0.01$ or 1	1	10	10	3
ADP	VESPER ladders (L-RE, L-NRE)	4	10	40	3
ADP	Each baseline	1	10	10	3
ATP, $D = 3,\ 6$	VESPER, $\alpha = 0.01$	1	10	10	3
ATP, $D = 3,\ 6$	VESPER ladders (L-RE, L-NRE)	4	10	40	3
Methane/Ni(511)	VESPER, $\alpha = 0.05$	1	7.5	7.5	3
Chignolin	VESPER ladder, $\alpha = 1$ to $1/18$	12	26.2	314.4	4

Boltzmann marginalization; their numerical integration checks are compared with the corresponding grid baseline.

B.9 Baseline settings

The ADP and ATP comparisons in Sections 3.2 and 3.3 use the Hamiltonians and CVs specified above and a 2 fs timestep. The ADP figure uses the first 10 ns per replica; ATP comparisons use the horizons stated in the main text. Times are per physical replica, so a four-level ladder uses four times the aggregate MD of a single replica. The same-α and no-exchange controls retain the ladder's architecture and population sizes. ADP VESPER single-replica, OPES and TAMD runs share initial configurations. Means and sample standard deviations use the trajectories that reach the reported horizon.

OPES uses bias factor 100 and barrier parameter 40 kJ mol^{-1} (Invernizzi & Parrinello, 2020); metadynamics deposits 1.2 kJ mol^{-1} hills every 500 steps (Laio & Parrinello, 2002); TAMD uses UFEDMM (Bajpai et al., 2023). Readouts are the OPES kernel density, the negative metadynamics hill sum, and the TAMD auxiliary-variable histogram at its own temperature. CFF (Sevgen et al., 2020) and its SIREN variant (S-CFF) (Zubieta Rico et al., 2025) use their saved native free-energy functions. ADP errors use the full-grid and finite-estimator-cell conventions in Appendix B.1; ATP dimensional-consistency metrics use complete retained planes after physical Boltzmann marginalization, with no external ABF surface used as a target.

Neural VES. We use the released implementation of Bonati et al. (2019), commit `953f6923d1ac`, with PLUMED 2.5.0 and LibTorch 1.0.0. Two compatibility repairs bound the periodic-input constructor by the periodicity-vector length and initialize inverse temperature from the supplied temperature. The bias factor is 100, with 50 grid bins per periodic CV and ReLU widths 48–24–12 for ADP. Adam starts at 10^{-3}; parameter and target updates occur every 500 MD steps with the released system-specific learning-rate controllers. Errors use the native free-energy exports.

B.10 Compute

Table 2 lists the simulated time per replica at the readout and the resulting aggregate MD per seed; several runs continued beyond the readout, and the table counts MD up to it. The ADP baselines are OPES, metadynamics, TAMD, neural VES, CFF and S-CFF. VESPER, CFF and S-CFF ran on NVIDIA H100, H200, L40S and RTX PRO 6000 GPUs, and the other baselines on CPUs, so wall-clock times are not comparable across methods. For scale, one chignolin seed takes about 100 h on one H100 GPU, which runs all twelve replicas, and one methane seed takes about 140–180 GPU-hours, mostly on L40S GPUs.

B.11 Reported molecular results

Table 3 collects the values quoted in Sections 3.2 to 3.5, with the readout behind each.

Table 3: Values reported in the main text for the molecular systems. ADP: 10 ns per replica, SWA8 readout, three seeds. ATP: 10 ns, SWA10, three seeds, three-seed mean surfaces for the plane comparisons. Methane: 7.5 ns, EMA readout, three seeds. Chignolin: first 25 ns per replica, four seeds. RMSEs are mean ± sample SD over seeds.

System	Quantity	Value
ADP	RMSE, VESPER $\alpha = 1$ (kJ mol^{-1})	13.27 ± 0.26
ADP	RMSE, VESPER $\alpha = 0.01$	2.19 ± 0.02
ADP	RMSE, L-RE / L-NRE	2.02 ± 0.24 / 2.05 ± 0.09
ADP	RMSE, MetaD/VES/S-CFF	2.61 ± 0.22/4.72 ± 0.25/2.04 ± 0.51
ATP	r^2, 3-CV vs 6-CV marginals, three planes	0.951 / 0.895 / 0.890
ATP	RMSE of the same planes (kJ mol^{-1})	5.5 / 8.1 / 7.5
ATP	r, 6-CV, 5 vs 10 ns, L-RE/L-NRE/single	0.913 ± 0.012/0.941 ± 0.028/0.804 ± 0.179
ATP	r^2 vs 3-CV on (ϕ_1, ϕ_2), L-RE/L-NRE/single	0.951/0.932/0.927
Methane/Ni(511)	ΔG, physisorbed to transition region (eV)	1.12 ± 0.25 (ref. 1.2)
Chignolin	Unfolding / refolding events	48 / 9
Chignolin	Folded fraction of the unbiased level	99.3%
Chignolin	Exchange acceptance between adjacent levels	30–93%
Chignolin	Unfolded levels outnumber folded from	3.9 ns
Chignolin	(ψ_3, ψ_4) minimum offset from unbiased MD	0.07 rad

C Controlled landscapes

C.1 Controlled evidence for learning from every level

Pooling native gradients on a five-dimensional landscape. The clearest test of the shared-minimizer construction (Proposition 1) holds the sampling ladder fixed and changes only which levels learn. On a five-dimensional multi-basin landscape with a five-level ladder ($\alpha = 1$, 0.5, 0.25, 0.1, 0.05) and a fixed total of 1,000 walkers per population (200 per level), a physical population started in one basin, a uniform auxiliary population and three seeds, the all-level exchange ladder reaches a final parameter-averaged full-5D RMSE of $0.233\ k_BT$, against 0.404 when the same exchange sampler trains only from its $\alpha = 0.5$ level, the most accurate of the three levels tested as sole learner (1.38 for $\alpha = 1$ and 0.55 for $\alpha = 0.25$): a 42.3% reduction, with a raw error–time integral 44.0% lower (1.78 against 3.17), both in all three paired seeds. Without exchange the ladder reaches 0.230, so the gain comes from pooling the native gradients of all levels. The pooled ladder comes within 7% of a separate single-level run at its best tuned α (0.218 at $\alpha = 0.5$, with all 1,000 walkers per population and no ladder; the best of $\alpha = 1$, 0.5, 0.25, 0.1, 0.05) without that value having to be found in advance. Means ± sample standard deviations over the three paired seeds are 0.233 ± 0.075 with all levels, 0.404 ± 0.069 with the $\alpha = 0.5$ level only, 0.230 ± 0.035 without exchange and 0.218 ± 0.041 for the single level.

Exchange on trapped landscapes. In an earlier campaign with a width-32 network and a five-level ladder that included $\alpha = 1$, exchange accelerated convergence where the cold level is trapped. On four-well landscapes in two, five and ten dimensions whose cold level cannot leave its starting basin, and on a scaled Müller–Brown surface, we compared the five-level exchange ladder with the best single level of each landscape over 812 runs (four to eight seeds per arm). The exchange ladder reaches a fixed error threshold earlier than every single level in $7/8$ seeds at five dimensions and in $8/8$ at ten, and at five dimensions its final error is lower by a paired ratio of 0.83 ($6/8$ seeds, signed-rank $p = 0.039$), or 0.78 weighted by coverage ($8/8$, $p = 0.004$). Without exchange the same ladder never reaches the threshold from five dimensions up, and a ten-fold slower swap cadence performs like the default. The advantage is largest with a trapped cold level and limited network capacity (compare Appendix C.3, where at width 1024 the ladder without exchange reaches the threshold at $d = 10$–50): in two dimensions and on the Müller–Brown surface, a well-placed single level is at least as accurate.

Table 4: Müller–Brown errors at $t = 320$: Boltzmann-weighted FES RMSE in k_BT of the parameter average, mean $\pm$ sample SD over four seeds, against the tempering exponent at 1000 walkers per population and against the walker count at $\alpha = 0.25$. A flat surface scores 1.281.

α	RMSE	Walkers	RMSE
0.05	0.109 ± 0.018	1	1.949 ± 0.749
0.1	0.098 ± 0.026	10	0.377 ± 0.160
0.25	0.104 ± 0.024	100	0.169 ± 0.082
0.5	0.093 ± 0.029	1000	0.104 ± 0.024
1	0.602 ± 0.495	10000	0.070 ± 0.017

C.2 MÜLLER–BROWN PROTOCOL

Figure 2A–C. The ground-truth energy is 0.1 times the standard Müller–Brown potential at $\beta = 1$ on the physical box $[-1.5, 1] \times [-0.5, 2]$, with the identity CV map. Seven arms share every setting except the tempering and, for the ladders, exchange: single levels at $\alpha = 1$, 0.5, 0.25, 0.1, 0.05 with 1,000 physical and 1,000 auxiliary walkers, and a five-level ladder ($\alpha = 1$, 0.5, 0.25, 0.1, 0.05; 200 walkers per level and population; normalized level weights 0.2) with and without exchange of physical and auxiliary walkers (every 100 steps, 200 attempted swaps). All physical walkers start within ± 0.025 of the deepest minimum, near $(-0.558, 1.442)$, and the auxiliary walkers start uniformly over the box; every arm of a seed receives the same initial arrays and initial model. The network has width 32, Adam at 3×10^{-4}, reduced timesteps $\Delta t_x = \Delta t_s = 0.002$, unit frictions, force clips of 100 and 500, and reflecting boundaries; 40,000 steps give $t = 80$. Three seeds (987011, 987029, 987053). Curves show the causal trailing-half parameter average: at each plotted time T the raw parameters saved in $[T/2, T]$ are averaged, so no point uses a future state.

Errors are evaluated on a 241×241 grid over $[-1.5, 1.5] \times [-0.5, 2.5]$, which extends beyond the simulation box so that the high-energy rim is scored. The RMSE removes the additive constant and weights every cell equally; because the rim energies are large it is dominated by the barrier and rim regions, which is why its scale is 237–252 k_BT and why the most strongly tempered level is best on it. $D_{\mathrm{KL}}(\rho_\theta \,\|\, \rho)$ normalizes $\rho \propto e^{-F}$ and $\rho_\theta \propto e^{-G_\theta}$ on the same grid and is dominated by the basins. At $t = 80$ (mean $\pm$ sample SD over three seeds) the RMSE is 237.7 ± 2.5 ($\alpha = 0.05$), 238.8 ± 1.4 (L-NRE), 239.0 ± 1.1 ($\alpha = 0.1$), 240.0 ± 1.2 (L-RE), 243.1 ± 2.0 ($\alpha = 1$), 243.7 ± 0.4 ($\alpha = 0.25$) and 243.9 ± 0.5 ($\alpha = 0.5$); the divergence is 0.010 ± 0.008 ($\alpha = 0.5$), 0.010 ± 0.007 ($\alpha = 0.25$), 0.017 ± 0.013 (L-RE), 0.019 ± 0.013 (L-NRE), 0.049 ± 0.034 ($\alpha = 0.1$), 0.102 ± 0.138 ($\alpha = 0.05$) and 0.82 ± 1.14 ($\alpha = 1$). The physical walkers of the $\alpha = 1$ arm stay in the starting basin (355 of the 360 overlaid positions from $t = 40$–80 lie nearest the deepest minimum), so its surface is unconstrained outside it. The maps in Fig. 2C are three-seed means of the final trailing-half averages after aligning the additive constants, each shifted to its own minimum; the reference map carries no walkers. The walker overlays pool 40 physical and 40 auxiliary positions per seed from each of $t = 40$, 60 and 80.

Tempering and population sweeps. Separate sweeps at $t = 320$ (160,000 steps) use four seeds, 0–3, uniform starts and a network of width 32. The tempering sweep is $\alpha \in \{0.05, 0.1, 0.25, 0.5, 1\}$ at 1,000 walkers per population; the population sweep changes the physical and auxiliary walker counts together through $1, 10, 100, 1000, 10000$ at $\alpha = 0.25$ and fixed integration and update counts, so a larger population also means proportionally more sampling work. Errors here are Boltzmann-weighted, mean-gauge FES RMSEs of trailing parameter averages, evaluated on fixed reference samples drawn from a 1200×1200 inverse-CDF discretization of the Müller–Brown density (Table 4). Any $\alpha \in [0.05, 0.5]$ reaches 0.09–0.11 k_BT against 1.28 k_BT for a flat surface, while $\alpha = 1$ remains at 0.60 k_BT; the error falls from 1.95 k_BT with one walker to 0.070 k_BT with 10^4.

C.3 FOUR-WELL DIMENSIONAL SCALING

Landscape. The teacher is a four-well Gaussian mixture in a two-dimensional plane, embedded in d coordinates by a random rotation and completed by harmonic directions of stiffness 10 (well width $\sigma_w = 0.3$, well separation 1.6, reflecting box $[-3, 3]^d$). The adjacent-well barriers are 13.5 k_BT, so a cold population started in one well does not cross within the budget, and the identity CV map is

Table 5: Four-well landscape: NRMSE and D_{KL} of the exchange ladder against the best single level at each dimension (three-seed medians), and the same ladder without exchange and with four times the walkers per level. Open entries in Fig. 2D–F correspond to values above the axis or not reached within the budget.

	NRMSE				D_{KL}	
d	L-RE	best single (α)	L-NRE	L-RE, $4\times$ walkers	L-RE	best single (α)
2	0.25	0.26 (0.05)	0.35	0.13	0.03	0.04 (0.05)
5	0.42	0.28 (0.5)	0.25	0.27	0.19	0.09 (0.5)
10	0.22	0.14 (0.25)	0.21	0.26	0.12	0.04 (0.25)
25	0.22	0.21 (0.5)	0.17	0.23	0.27	0.22 (0.5)
50	0.19	0.24 (0.5)	0.18	0.20	0.40	0.64 (0.5)
100	0.20	0.42 (0.5)	0.46	0.44	0.95	2.36 (0.5)

used. The two-dimensional core is identical at every d; the reference is 40,000 i.i.d. draws from the teacher's Boltzmann distribution, and all errors are evaluated on all d coordinates.

Protocol. $d \in \{2, 5, 10, 25, 50, 100\}$. The same three-hidden-layer network of width 1024 is used at every d; it is the smallest width at which a perfect-information regression of the exact free energy into the same network (i.i.d. Boltzmann samples, exact energies, 160,000 Adam updates of batch 1,000) learns the four-well plane at $d = 100$, so capacity does not limit any arm (widths 128, 256 and 512 do not learn the $d = 100$ plane; width 1024 reaches a plane error of 0.03–0.13 k_BT against 1.29 for a flat surface at every d). Every arm uses 1,000 physical and 1,000 auxiliary walkers in total, both started in the deepest well, 160,000 overdamped steps at $\Delta t = 0.005$ (reduced time 800), Adam at 3×10^{-5}, a bias-corrected parameter EMA with time constant 2,000 steps for the convergence readout and a trailing-half parameter average for the endpoint. The learning rate was fixed by a stability sweep before the campaign: at width 1024 a rate of 3×10^{-4} makes every arm oscillate by tens of k_BT, whereas 3×10^{-5} is stable for singles and ladders at $d = 25$ and 100. Arms: single levels at $\alpha = 1$, 0.5, 0.25, 0.1, 0.05, and four-level ladders ($\alpha = 0.5$, 0.25, 0.1, 0.05; 250 walkers per level and population) with and without symmetric exchange of physical and auxiliary walkers. The $\alpha = 1$ level is omitted from the ladders because at this width its unbiased population, confined to one well, drives the model to non-finite energies. As a control on population rather than temperature, the same ladders were also run with 1,000 walkers per level and population. Three seeds per arm, 162 cells; one cell (the $\alpha = 1$ level at $d = 2$, seed 2) diverged to non-finite energies and is absent. The campaign, its selection rules and readouts were fixed before any cell was read.

Readouts. $\mathrm{NRMSE} = \mathrm{RMSE}/\sigma$, where the RMSE removes the additive constant and σ is the Boltzmann-weighted standard deviation of the reference free energy, so that 1 is a flat surface and $R^2 = 1 - \mathrm{NRMSE}^2$; $D_{\mathrm{KL}}(\rho_\theta \,\|\, \rho)$ on the same reference samples; and the reduced time at which the EMA readout's NRMSE first stays below 0.25 to the end of the run ($R^2 > 0.94$). All are three-seed medians (Table 5).

Observations. Against the single levels, the exchange ladder is within a factor 1.6 of the best on NRMSE at every d and is best at $d = 50$ and 100; on the divergence it is best at $d = 2$, 50 and 100, second at 5 and 25 and third at 10. At $d = 100$ it is the most accurate of all arms on both measures. It is the only arm whose median reaches $\mathrm{NRMSE} < 0.25$ at $d = 100$ (reduced time 352); the best single level, $\alpha = 0.5$, reaches it at $d = 5$–50, only at the end of the budget at $d = 5$ and 10, but not at $d = 100$, and the exchange ladder itself does not reach it at $d = 5$. The identity of the best single level changes with d and every single level fails somewhere: $\alpha = 0.05$ and 0.1 exceed a flat surface from $d = 5$ and $d = 50$; $\alpha = 0.25$ reaches 1.05 at $d = 100$; $\alpha = 0.5$ is the worst tempered level at $d = 2$ (1.86) and falls to 0.42 at $d = 100$; $\alpha = 1$ fails at every d (NRMSE 1.2 at $d = 100$, 3 at $d = 25$ and 50, and 10–3 $\times 10^4$ at $d \leq 10$). At $d = 100$ the ladder without exchange (0.46) and the ladders with four times the walkers per level (0.44 with exchange, 0.39 without) are both far from the exchange ladder (0.20), so exchange, not population, carries the high-dimensional result. At $d = 5$–25 a well-chosen $\alpha = 0.5$ single level is more accurate than the ladder by up to $1.6\times$ on NRMSE, so the ladder's advantage is robustness to d without tuning α and accuracy where sampling

is hardest, not superiority at every d. With three seeds, one exchange-ladder seed at $d = 100$ ended at 0.57 after reaching 0.30, and one no-exchange seed reached 0.19; the medians above should be read with that spread.

D CONNECTIONS TO VES, PS-CD, AND RÉNYI OBJECTIVES

The following identities connect the VESPER objective to well-tempered VES, pseudo-spherical contrastive divergence, and a Rényi variational principle. They also clarify how VESPER uses these connections to learn a physical free energy from two evolving populations and to combine tempering levels in one model.

D.1 WELL-TEMPERED VES AND AN EXPLICIT SHARED-ENERGY OBJECTIVE

Let p be the physical CV density and t the VES target. The fixed-target functional of Valsson & Parrinello (2014) is

$$\Omega[V;t] = \beta^{-1} \log \mathbb{E}_p e^{-\beta V} + \int t(s)V(s)\,\mathrm{d}s.$$

Differentiating with t held fixed gives

$$\nabla_\theta \Omega[V_\theta;t] = -\mathbb{E}_{\mathrm{biased}} \nabla_\theta V_\theta + \mathbb{E}_t \nabla_\theta V_\theta.$$

Set $V_\theta = -(1-\alpha)G_\theta$ and $t_\theta = P_{G_\theta}$, with $0 < \alpha < 1$. The biased physical law is then ν_{G_θ}, and substitution yields

$$\nabla_\theta \Omega[V_\theta;t]|_{t=t_\theta} = (1-\alpha)\left\{\mathbb{E}_{\nu_{G_\theta}} \nabla_\theta G_\theta \circ s - \mathbb{E}_{P_{G_\theta}} \nabla_\theta G_\theta\right\} = (1-\alpha)\nabla_\theta \mathcal{L}_{\hat{\beta}}[G_\theta].$$

This proves equation 45 for every admissible model for which differentiation under the integrals is valid. The factor $1-\alpha$ can be absorbed into the step size for ordinary gradient descent with the same parameterization and sample estimate.

VESPER thus gives the instantaneous self-consistent update an explicit scalar objective, convex in the free energy. This point requires holding the VES target fixed during differentiation. Taking the total derivative of $\Omega[V_\theta; t_\theta]$ also differentiates the target and adds

$$\int V_\theta(s) \nabla_\theta t_\theta(s)\,\mathrm{d}s = \alpha\beta(1-\alpha)\,\mathrm{Cov}_{t_\theta}(G_\theta, \nabla_\theta G_\theta),$$

where we used $\nabla_\theta t_\theta = -\alpha\beta t_\theta(\nabla_\theta G_\theta - \mathbb{E}_{t_\theta}\nabla_\theta G_\theta)$. In contrast, differentiating $\mathcal{L}_{\hat{\beta}}$ directly gives the desired two-population update, with its minimum and KL control established in Propositions S1 and S2.

The distinction also appears in how the target evolves. Well-tempered VES sets $t \propto p^\alpha$ (Valsson & Parrinello, 2015). In neural VES, the free-energy relation $\widehat{F} = -V - \beta^{-1}\log t$ and the previous target $t_{n-1} \propto e^{-\alpha\beta\widehat{F}_{n-1}}$ give the recursive update (Bonati et al., 2019)

$$\widehat{F}_n = -V_n + \alpha\widehat{F}_{n-1} + C_n, \qquad t_n \propto e^{-\alpha\beta\widehat{F}_n}. \tag{46}$$

At a fixed point, $(1-\alpha)\widehat{F} = -V + C$, so $V = -(1-\alpha)G$ gives $\widehat{F} = G + \mathrm{const}$. During learning, VESPER instead defines the target directly from the current G_θ and represents it with persistent auxiliary walkers. One model therefore determines both the molecular bias and the auxiliary potential, without a separate recursion for the free-energy estimate. The walker laws track these changing targets through the coupled dynamics analyzed in Appendix A.6.

This formulation extends directly to the combined objective: every $\mathcal{L}_{\hat{\beta}_k}$ has the same physical minimizer, so all levels train one G_θ through their native gradients (Proposition 1). In VES notation, the interior-level update is reproduced by weighting the corresponding fixed-target gradients by $c_k/(1-\alpha_k)$. The rescaled gradient has the finite physical-endpoint limit

$$\lim_{\alpha\uparrow 1} \frac{1}{1-\alpha}\,\nabla_\theta \Omega[-(1-\alpha)G_\theta;t]|_{t=P_{G_\theta}} = \mathbb{E}_p \nabla_\theta G_\theta - \mathbb{E}_{P_{G_\theta,\beta}} \nabla_\theta G_\theta,$$

whenever the two expectations converge. VESPER includes this endpoint through equation 8. Its energy parameterization therefore lets strongly biased and unbiased replicas contribute to one variational objective, with no cross-level importance weights or separate reconstruction of a free energy from each bias.

D.2 PS-CD AND DIRECT SAMPLING OF THE TWO TILTED POPULATIONS

With $g = \beta G$, the dimensionless objective ℓ_α is given by equation 39. Let $q_g = e^{-g}/Z_g$ be the model density relative to a reference measure r, with $Z_g < \infty$. Substituting $\log q_g = -g - \log Z_g$ and $\gamma = \alpha - 1$ gives

$$\ell_\alpha[g] = \frac{1}{1+\gamma} \log \int q_g^{1+\gamma} \, \mathrm{d}r - \frac{1}{\gamma} \log \int q_g^{\gamma} \, \mathrm{d}p = \mathcal{L}_\gamma^{\mathrm{PS}}[g]. \tag{47}$$

The first term contributes $-\log Z_g$ and the second $+\log Z_g$, so the normalizer cancels. This is the PS-CD objective of Yu et al. (2021, Eq. 17), in the γ-cross-entropy family (Fujisawa & Eguchi, 2008), whenever both integrals are finite. VESPER uses $-1 < \gamma < 0$, which weights observations more strongly where the model density is small.

For the Lebesgue reference, write $q_\theta = q_{G_\theta} = e^{-\beta G_\theta}/Z_{G_\theta}$. The two populations in the gradient are exactly the laws used in the main text:

$$\begin{aligned} P_{G_\theta}(s) &= \frac{q_\theta(s)^\alpha}{\int q_\theta(u)^\alpha \, \mathrm{d}u}, \\ Q_{G_\theta}(s) &= \frac{p(s) q_\theta(s)^{\alpha-1}}{\int p(u) q_\theta(u)^{\alpha-1} \, \mathrm{d}u}. \end{aligned} \tag{48}$$

Differentiating the two log integrals gives $\nabla_\theta \ell_\alpha[g_\theta] = \mathbb{E}_{Q_{G_\theta}} \nabla_\theta g_\theta - \mathbb{E}_{P_{G_\theta}} \nabla_\theta g_\theta$, recovering equation 7 after division by β. At the physical endpoint, $P_{G_\theta} = q_\theta$ and $Q_{G_\theta} = p$, recovering the maximum-likelihood gradient.

The computational difference is how these expectations are obtained. PS-CD uses a fixed dataset for the positive term and reweights samples from q_θ to estimate the tempered model expectation (Yu et al., 2021). VESPER directly evolves auxiliary walkers in αG_θ, removing that temperature-change importance weighting from the fresh-sample gradient. In molecular applications, it also generates Q_{G_θ} online through the biased microscopic law ν_{G_θ}. This makes the same variational principle usable when equilibrium configurations are not available in advance: sampling supplies new configurations while their gradients update the free-energy model (Appendix A.3). Combining several such levels retains the physical free-energy target and lets every population contribute to learning (Appendix A.4).

D.3 RÉNYI VARIATIONAL IDENTITY AND ITS SAMPLING REALIZATION

For a probability reference r and $0 < \alpha < 1$, set $h = -g$ in equation 39. Then

$$-\ell_\alpha[g] = \frac{1}{\alpha - 1} \log \int e^{(\alpha-1)h} \, \mathrm{d}p - \frac{1}{\alpha} \log \int e^{\alpha h} \, \mathrm{d}r,$$

which is the variational expression of Birrell et al. (2021, Theorem 3.1). With $f = \mathrm{d}p/\mathrm{d}r > 0$ r-almost everywhere and $M(\alpha) = \int f^\alpha \, \mathrm{d}r$, Proposition S8 gives $g_\star = -\log f + C$. Substituting this minimizer into the two integrals yields

$$\inf_g \ell_\alpha[g] = \frac{\log M(\alpha)}{\alpha(1-\alpha)} = -\frac{1}{\alpha} D_\alpha(p\|r), \qquad D_\alpha(p\|r) = \frac{\log M(\alpha)}{\alpha - 1}.$$

The factor $1/\alpha$ accounts for the divergence convention in that reference. If f vanishes on a set of positive r-measure, the infimum need not be attained by a finite free energy.

The value of this identity for VESPER is that its optimizer is also the quantity needed to drive sampling. For $r = \mathrm{d}s$, $g_\star = \beta G_\star$ up to a constant. VESPER uses G_θ to bias microscopic dynamics and to evolve the auxiliary walkers, obtaining both gradient expectations without evaluating the unknown CV density p or first collecting equilibrium data (Appendix A.3). The KL decomposition in Proposition S2 connects the objective excess to the discrepancy of P_G and Q_G from their common tempered target, and Proposition 1 lets several orders train the same physical free energy. The variational identity thereby supports a coupled sampling and reconstruction procedure.

For a reference of finite mass R, normalization to r/R subtracts $(\log R)/\alpha$ from ℓ_α and leaves its gradients unchanged. On infinite-volume CV domains we instead use the exponential integrals directly, with the integrability conditions of Appendices A.1 and A.7. This preserves the same free-energy minimizer without introducing an artificial uniform probability reference.

E ALGORITHM AND IMPLEMENTATION DETAILS

VESPER is implemented in PySAGES (Zubieta Rico et al., 2024). We alternate molecular and auxiliary sampling with parameter updates, retaining the auxiliary population between epochs. Neural models provide both parameter derivatives for learning and spatial derivatives for biasing. When several parameter updates reuse one batch, within-population importance weights account for the change from its generating model. The pseudocode and weight formulas are given below; model architectures, integration parameters, optimizer settings, exchange intervals, and readouts are specified for each experiment in Appendix B.

Algorithm 1 describes physical sampling with an energy-unit model G_θ and level ratios $\alpha_k = \hat{\beta}_k/\beta$. Each epoch generates physical and auxiliary samples, optionally exchanges configurations (in the molecular implementation, physical configurations only; Appendix B), and updates the shared model. The per-system protocols in Appendix B specify sampler lengths, optimizer budgets, sample reuse, exchange frequency, and readout.

Algorithm 1 Physical VESPER ladder

Input: potential U, CV map s, inverse temperature β; levels $0 < \hat{\beta}_1 < \cdots < \hat{\beta}_K \le \beta$; weights $c_k > 0$, $\sum_k c_k = 1$; model G_θ; sampler, optimizer, exchange, and readout settings.
Initialize θ and the physical and auxiliary populations on every level.
repeat
  Set $\theta_0 \leftarrow \theta$. Advance physical walkers in $U - (1-\alpha_k)G_{\theta_0} \circ s$ and auxiliary walkers in $\alpha_k G_{\theta_0}$, at inverse temperature β.
  If exchange is enabled, attempt adjacent-level swaps using equation 13.
  Save batches $\{x_k^j\}_{j=1}^{n_k}$ and $\{s_k^i\}_{i=1}^{N_k}$, together with their generating model G_{θ_0}.
  for each parameter update allowed by the protocol **do**
    for $k = 1, \ldots, K$ **do**
      if sample reweighting is enabled **then**
        $\Delta G \leftarrow G_\theta - G_{\theta_0}$.
        $w_j^{(x,k)} \leftarrow e^{(\beta-\hat{\beta}_k)\Delta G(s(x_k^j))}$, $w_i^{(s,k)} \leftarrow e^{-\hat{\beta}_k \Delta G(s_k^i)}$.
        Normalize the weights separately in each pool.
      else
        $\bar{w}_j^{(x,k)} \leftarrow 1/n_k$, $\bar{w}_i^{(s,k)} \leftarrow 1/N_k$.
      end if
      $h_k \leftarrow \sum_j \bar{w}_j^{(x,k)} \nabla_\theta G_\theta(s(x_k^j)) - \sum_i \bar{w}_i^{(s,k)} \nabla_\theta G_\theta(s_k^i)$.
    end for
    If an enabled sample-reuse threshold is exceeded, end the inner loop and generate fresh samples.
    Update θ with the configured optimizer using $\sum_k c_k h_k$.
  end for
  Save model checkpoints and sampling diagnostics.
until the stopping condition or simulation budget is reached
Output: the checkpoint or averaged readout selected by the protocol.

Ladder controls. To isolate gradient pooling, use the same level values, initial states, replica counts, exchange schedule, and simulation budgets in the all-level and sampling-only arms. The first uses $\sum_k c_k h_k$ with normalized weights; the second uses only h_j from a single level j, whose objective is $\mathcal{L}_{\hat{\beta}_j}$ (the likelihood endpoint when $\alpha_j = 1$); Appendix C.1 reports $\alpha_j = 0.5$. A ladder without exchange retains all level gradients while disabling all swap moves. Comparisons record aggregate microscopic integration steps, auxiliary force evaluations, optimizer work, and elapsed time.

Sample reuse. The weights of equation 44 supply the two density ratios for a batch generated at G_{θ_0}. Reweighted estimators also appear in Carbone et al. (2023) and Invernizzi & Parrinello (2020). The molecular implementation can monitor the 95th percentile and maximum of $|\log w|$, with soft and hard thresholds 5 and 8, and the normalized ESS $(\sum w)^2/[n \sum w^2]$, with floor 0.2. These

statistics measure weight concentration during batch reuse. Activation, step reduction, and refresh rules are specified per system in Appendix B.7.

Existing-data updates. For equilibrium observations $x_j^* \sim \varrho$, replace physical sampling and its reuse weights by equation 42.

Checkpoint and averaged readouts. A trailing weight-space average produces a network with parameters

$$\bar{\theta} = \frac{1}{|W|} \sum_{k \in W} \theta_k. \tag{49}$$

The checkpoint set W, stride, and window are specified per experiment (Polyak & Juditsky, 1992; Izmailov et al., 2018). Fixed-checkpoint and last-iterate readouts use their corresponding θ_k.