SEAL-BG: SCALING BOLTZMANN GENERATION WITH EQUIVARIANT PERCEPTION AND LOCAL-COORDINATE DENSITIES

Anonymous authors
Paper under double-blind review

ABSTRACT

Estimating protein thermodynamics, including conformational populations and free-energy differences, requires sampling molecular equilibrium ensembles, but correlated trajectories and slow conformational transitions make this costly. Boltzmann generators (BGs) offer independent molecular proposals with evaluable densities, enabling energy-based statistical correction. Autoregressive Boltzmann Generators (ArBG) demonstrate this approach using causal Transformers over Cartesian coordinate tokens, with single-system benchmarks extending to Chignolin (10 amino acids). Scaling ArBG to larger proteins motivates presentations that exploit molecular symmetries while distinguishing local geometry from global conformational variation. We introduce SEAL-BG ($\underline{\mathrm{SE}}(3)$-$\underline{\mathrm{E}}$quivariant $\underline{\mathrm{A}}$utoregression in $\underline{\mathrm{L}}$ocal Coordinates for $\underline{\mathrm{B}}$oltzmann $\underline{\mathrm{G}}$eneration), an atom-autoregressive BG that decouples geometric perception from probabilistic generation. A causal $\mathrm{SE}(3)$-equivariant geometric encoder encodes the Cartesian coordinates of previously generated atoms, while normalized local-coordinate conditionals model bond lengths, bond angles, and periodic torsions. This separation combines equivariant Cartesian conditioning with structured local-coordinate densities, while analytic local Jacobians preserve an evaluable Cartesian proposal density. We evaluate on four released peptide benchmarks and scale to Trp-cage (20 amino acids, 284 atoms) and Homeodomain (54 amino acids, 947 atoms) using separate within-system ensembles. SEAL-BG achieves competitive energy and torsion distances on the peptide benchmarks and improves raw total-energy overlap relative to ArBG from 0.8651 to 0.8926 on Trp-cage and from 0.4176 to 0.7704 on Homeodomain. Across the evaluated training budgets, SEAL-BG uses geometric-mean factors of $15.2\times$ fewer training floating-point operations (FLOPs) and $18.1\times$ fewer A100 GPU-hours than ArBG. These results support separating Cartesian perception from local-coordinate generation as a route toward scalable Boltzmann proposals.

1 INTRODUCTION

Thermodynamic properties of a molecule depend on the distribution of probability across accessible conformations rather than on a single low-energy structure. For an N-atom system with Cartesian coordinates $\mathbf{r} \in \mathbb{R}^{3N}$ and reduced potential energy $u(\mathbf{r})$, let Ω denote configuration space and Z the partition function. The equilibrium probability density p^{eq} is the Boltzmann distribution

$$p^{\mathrm{eq}}(\mathbf{r}) = Z^{-1} \exp[-u(\mathbf{r})], \qquad Z = \int_{\Omega} \exp[-u(\mathbf{r})] \, \mathrm{d}\mathbf{r}.$$

These ensembles are often multimodal, with metastable conformations separated by high-energy regions (Wirnsberger et al., 2020; Tuckerman, 2023).

To sample equilibrium ensembles, molecular dynamics (MD) explores conformational space by integrating equations of motion with small time steps to resolve fast local vibrations (Frenkel & Smit, 1996). Transitions between metastable conformations can occur on much longer timescales, so trajectories may spend many steps within individual basins and produce strongly correlated samples (Shaw et al., 2010). Enhanced-sampling methods can improve exploration, but reliable mixing

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.

across relevant conformations remains difficult for complex molecular systems (Hénin et al., 2022; Syed et al., 2022).

Boltzmann generators (BGs) complement MD by drawing independent configurations from a learned proposal $q(\mathbf{r})$, whose evaluable density enables importance reweighting toward the target distribution (Noé et al., 2019). To be useful at manageable computational cost, BGs must capture both global conformational diversity and local geometric preferences encoded by the force field, which is the potential-energy function (MacKerell et al., 1998).

Many existing BGs use normalizing flows to construct evaluable proposal densities by transforming a simple base distribution through an invertible map (Noé et al., 2019; Köhler et al., 2020; Rezende & Mohamed, 2015). Discrete flows permit efficient density evaluation but require layers with tractable inverses and Jacobian determinants, restricting architectural choices (Dinh et al., 2017). Continuous flows allow more flexible architectures, but sampling requires numerical integration, while log-density evaluation adds the cost of integrating the vector field's divergence (Chen et al., 2018; Lipman et al., 2024).

Autoregressive Boltzmann Generators (ArBG) (Rehman et al., 2026c) avoid these constraints by using a causal Transformer to predict discretized Cartesian coordinate components, enabling direct generation and single-pass log-density evaluation without invertible neural networks or numerical integration . However, this parameterization provides no explicit inductive bias for molecular internal geometry or rigid-motion symmetry, forcing the model to learn bond-length, bond-angle, and torsional preferences, as well as rotational and translational equivalences, directly from Cartesian data. This learning burden becomes increasingly challenging as molecular size and conformational complexity grow.

To address these limitations, we introduce SEAL-BG ($\underline{\mathrm{SE}}(3)$-$\underline{\mathrm{E}}$quivariant $\underline{\mathrm{A}}$utoregression in $\underline{\mathrm{L}}$ocal Coordinates for $\underline{\mathrm{B}}$oltzmann $\underline{\mathrm{G}}$eneration), which separates geometric perception from probabilistic generation. A causal $\mathrm{SE}(3)$-equivariant Transformer encodes the Cartesian coordinates of previously generated atoms, while a normalized local-coordinate head generates the next atom through its distance, angle, and signed torsion. Radial and angular means are parameterized relative to force-field reference geometry, and full-prefix conditioning preserves their dependence on the global conformation.

Our contributions are:

- We introduce SEAL-BG, combining SE(3)-equivariant Cartesian conditioning with normalized local-coordinate autoregression and analytic Jacobians for single-pass log-density evaluation.
- We evaluate SEAL-BG across peptide and protein systems using energy and torsion metrics to assess agreement between generated and MD reference distributions. It achieves competitive energy and torsion metrics on the alanine-peptide and Chignolin benchmarks, and improves energy agreement over ArBG on Trp-cage and Homeodomain.
- Across all six systems, SEAL-BG requires substantially lower training floating-point operations (FLOPs) and A100 GPU-hours than other baselines.

2 Background and Problem Formulation

2.1 Equilibrium Proposals and Reweighting

For an N-atom system with Cartesian coordinates $\mathbf{r} \in \mathbb{R}^{3N}$, let $U(\mathbf{r})$ denote the potential energy and $u(\mathbf{r}) = U(\mathbf{r})/(k_{\mathrm{B}}T)$ the reduced potential. Here k_{B} is the Boltzmann constant and T is the absolute temperature. The equilibrium distribution is

$$p^{\mathrm{eq}}(\mathbf{r}) = Z^{-1}\exp[-u(\mathbf{r})], \qquad Z = \int \exp[-u(\mathbf{r})]\,\mathrm{d}\mathbf{r}. \tag{1}$$

A Boltzmann generator with trainable parameters θ learns a proposal density $q_\theta(\mathbf{r})$ that approximates $p^{\mathrm{eq}}(\mathbf{r})$, combining independent proposal sampling with importance reweighting to account for the remaining distributional mismatch (Noé et al., 2019). Given M samples $\mathbf{r}^{(m)} \sim q_\theta$, $m = 1, \ldots, M$,

self-normalized importance sampling (SNIS) (Owen, 2013) assigns unnormalized weights w_m and normalized weights $\overline{w}_m$ as

$$w_m = \frac{\exp[-u(\mathbf{r}^{(m)})]}{q_\theta(\mathbf{r}^{(m)})}, \qquad \overline{w}_m = \frac{w_m}{\sum_{j=1}^{M} w_j}. \tag{2}$$

The unknown partition function Z cancels when the weights are normalized, so reweighting does not require its evaluation. Reweighting increases the contribution of sampled configurations underrepresented relative to the target and reduces that of overrepresented configurations. Our sampling benchmarks use the clipped-SNIS protocol described in Section 4.

2.2 Free-Energy Estimation

For each thermodynamic state k, with potential energy $U_k(\mathbf{r})$ and temperature T_k, we define the dimensionless reduced potential as $u_k(\mathbf{r}) = U_k(\mathbf{r})/(k_\mathrm{B}T_k)$, the partition function as $Z_k = \int \exp[-u_k(\mathbf{r})]\,\mathrm{d}\mathbf{r}$, and the reduced free energy as $f_k = -\log Z_k$.

We are interested in the free-energy difference between two physical states $s \in \{0, 1\}$, given by $\Delta f = f_1 - f_0 = -\log(Z_1/Z_0)$. For states evaluated at a common temperature T, the corresponding free-energy difference in energy units is $\Delta F = k_\mathrm{B}T\,\Delta f$.

The Bennett acceptance ratio (BAR) estimator uses equilibrium samples from both physical states and evaluates both reduced potentials on both sample sets (Bennett, 1976). Its reliability depends on the overlap between the two state distributions. When the endpoint overlap is insufficient, intermediate states can be introduced along a path connecting the two physical states. The multistate Bennett acceptance ratio (MBAR) estimator then combines samples from the endpoint and intermediate states, at additional sampling cost (Shirts & Chodera, 2008). For K states along such a path, let M_k be the number of samples from state k, and let $\{\mathbf{r}^{(m)}\}_{m=1}^{M}$ denote the pooled samples, with $M = \sum_{k=0}^{K-1} M_k$. Let $\hat{f}_i$ denote the MBAR estimate of state i's reduced free energy. Evaluating all reduced potentials on the pooled samples gives

$$\hat{f}_i = -\log \sum_{m=1}^{M} \frac{\exp[-u_i(\mathbf{r}^{(m)})]}{\sum_{k=0}^{K-1} M_k \exp[\hat{f}_k - u_k(\mathbf{r}^{(m)})]}, \qquad i = 0, \ldots, K-1. \tag{3}$$

The estimates are defined up to a common additive constant that cancels in free-energy differences; for $K = 2$, MBAR reduces to BAR.

To avoid sampling additional intermediate physical states, Ding & Zhang (2021) fit a separate generative model to MD samples from each of the two physical endpoint states $s \in \{0, 1\}$, thereby defining a paired learned reference state s^* for each endpoint. The model density q_{s^*} is normalized by construction, so its reduced potential, partition function, and reduced free energy satisfy

$$u_{s^*}(\mathbf{r}) = -\log q_{s^*}(\mathbf{r}), \qquad Z_{s^*} = \int q_{s^*}(\mathbf{r})\,\mathrm{d}\mathbf{r} = 1, \qquad f_{s^*} = -\log Z_{s^*} = 0. \tag{4}$$

Applying Equation 3 with $K = 2$ to each endpoint–reference pair (s, s^*), using samples from the corresponding physical endpoint and learned reference, yields an estimate $\hat{f}_s$ because $f_{s^*} = 0$ by construction. The endpoint difference is then $\Delta\hat{f} = \hat{f}_1 - \hat{f}_0$. These physical–reference comparisons require sufficient overlap and densities defined with respect to a common coordinate measure.

3 Methodology

3.1 Autoregressive Cartesian Proposal

Autoregressive factorization. For an N-atom molecule with topology $\mathcal{G}$, let $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$ denote its Cartesian configuration. We represent the configuration relative to the first atom by defining $\tilde{\mathbf{r}}_i = \mathbf{r}_i - \mathbf{r}_1$. Under a fixed atom ordering determined by the molecular topology, the generated Cartesian prefix for atom $i \geq 2$ is $\tilde{\mathbf{r}}_{<i} = (\tilde{\mathbf{r}}_1, \ldots, \tilde{\mathbf{r}}_{i-1})$. The conformational density over relative Cartesian coordinates factorizes autoregressively as

$$q_\theta^{\mathrm{shape}}(\mathbf{r} \mid \mathcal{G}) = \prod_{i=2}^{N} q_{\theta,i}^{\mathrm{cart}}\left(\tilde{\mathbf{r}}_i \mid \tilde{\mathbf{r}}_{<i}, \mathcal{G}\right), \tag{5}$$

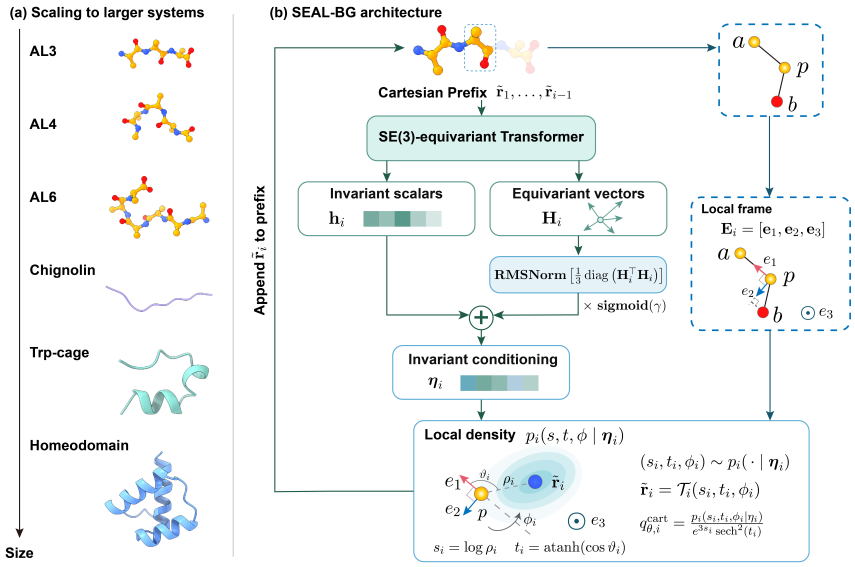


Figure 1: Overview of SEAL-BG. (a) Evaluated molecular systems, ranging from the alanine peptides AL3, AL4, and AL6 to Chignolin, Trp-cage, and Homeodomain. (b) At autoregressive step i, a causal SE(3)-equivariant Transformer encodes the generated Cartesian prefix $\tilde{\mathbf{r}}_{<i}$ into invariant scalar features h_i and equivariant vector features H_i, which are combined into the invariant conditioning representation η_i. A normalized density head models the next atom in local coordinates defined by a parent atom p and reference atoms a and b, all belonging to the generated prefix. The sampled local coordinates are mapped to the Cartesian position $\tilde{\mathbf{r}}_i$ and appended to the prefix, while an analytic Jacobian converts the local density p_i into the Cartesian conditional density $q^{\mathrm{cart}}_{\theta,i}$.

where $q^{\mathrm{cart}}_{\theta,i}$ is the next-atom conditional density with respect to Cartesian coordinates.

Equivariant prefix encoding. To parameterize each conditional, a causal SE(3)-equivariant Transformer encoder $\mathcal{E}_\theta$, adapted from Jiao et al. (2026), processes the generated Cartesian prefix. The encoder additionally receives fixed molecular metadata specifying the source and target atoms, element and residue identities, atom roles, sequence positions, and causal bond connectivity; no target coordinate is provided. For the next atom i, it returns scalar features h_i and vector features H_i, with $h_i \in \mathbb{R}^d$ and $H_i \in \mathbb{R}^{3\times d}$:

$$(h_i, H_i) = \mathcal{E}_\theta\left(\tilde{\mathbf{r}}_{<i}, \mathcal{G}\right). \tag{6}$$

For a proper rotation $R \in \mathrm{SO}(3)$, these features transform as

$$\mathcal{E}_\theta\left(R\tilde{\mathbf{r}}_{<i}, \mathcal{G}\right) = (h_i, RH_i). \tag{7}$$

Because the input uses relative coordinates, the encoder is invariant to a global translation. Let γ be a learned scalar gate. We combine the scalar and vector features into the conditioning representation

$$\eta_i = h_i + \mathrm{sigmoid}(\gamma)\ \mathrm{RMSNorm}\left[\tfrac{1}{3}\,\mathrm{diag}\left(H_i^\top H_i\right)\right], \tag{8}$$

where diag extracts the diagonal of a matrix and RMSNorm denotes root-mean-square normalization (Zhang & Sennrich, 2019). Because $(RH_i)^\top(RH_i) = H_i^\top H_i$, η_i is invariant to global rotation and can parameterize the local-coordinate density of the next atom. Feature alignment and causal evaluation are detailed in Appendix A.1.

3.2 Local-Coordinate Generation and Density Evaluation

Local frame and internal coordinates. For atom i, let p be its parent atom and let a and b be two additional reference atoms. All three atoms belong to the generated prefix and are selected deterministically from the molecular topology. Assuming that they define a nondegenerate frame,

we construct the right-handed orthonormal basis

$$e_1 = \frac{\tilde{\mathbf{r}}_a - \tilde{\mathbf{r}}_p}{\|\tilde{\mathbf{r}}_a - \tilde{\mathbf{r}}_p\|}, \quad v = \left(I - e_1 e_1^\top\right)(\tilde{\mathbf{r}}_b - \tilde{\mathbf{r}}_p), \quad e_2 = \frac{v}{\|v\|}, \quad e_3 = e_1 \times e_2. \tag{9}$$

The frame is centered at $\tilde{\mathbf{r}}_p$.

Let $d_i = \tilde{\mathbf{r}}_i - \tilde{\mathbf{r}}_p$ denote the displacement of atom i from its parent. Its local position is described by the radius $\rho_i = \|d_i\| > 0$, angle $\vartheta_i = \arccos(d_i^\top e_1/\rho_i) \in (0, \pi)$, and signed torsion $\varphi_i = \operatorname{atan2}(d_i^\top e_3, d_i^\top e_2) \in [-\pi, \pi)$. We transform the radial and angular variables as $s_i = \log \rho_i$ and $t_i = \operatorname{atanh}(\cos \vartheta_i)$, while retaining the periodic torsion φ_i.

The inverse map from local variables to relative Cartesian coordinates is

$$\mathcal{T}_i(s, t, \varphi) = \tilde{\mathbf{r}}_p + e^s \left[\tanh(t) e_1 + \operatorname{sech}(t)\left(\cos\varphi\, e_2 + \sin\varphi\, e_3\right)\right]. \tag{10}$$

The initial atoms, for which a complete local frame is unavailable, use the normalized orientation factors described in Appendix A.3.

Normalized local conditional density. Conditioned on the invariant representation η_i, the density head models the transformed local coordinates (s, t, φ). We partition the torsion circle into K equal bins B_k, $k = 1, \ldots, K$, each of width $\Delta = 2\pi/K$. The model predicts the bin probabilities w_{ik} and, within each bin, Gaussian parameters for s and t, while φ is uniform within the bin. The resulting normalized conditional density is

$$p_i(s, t, \varphi \mid \eta_i) = \sum_{k=1}^{K} w_{ik}\, \mathcal{N}\left(s; \mu_{ik}^{\rho}, (\sigma_{ik}^{\rho})^2\right) \mathcal{N}\left(t; \mu_{ik}^{\mathrm{ang}}, (\sigma_{ik}^{\mathrm{ang}})^2\right) \frac{\mathbf{1}\{\varphi \in B_k\}}{\Delta}. \tag{11}$$

The radial and angular means are parameterized as bounded residuals around force-field reference geometry, providing a local geometric prior while retaining dependence on the complete Cartesian prefix through η_i. The complete density-head parameterization and numerical bounds are given in Appendix A.2. Given the prefix and torsion bin, s and t are conditionally independent, while their marginal dependence is induced by the mixture over torsion bins.

Cartesian density. The local density p_i is defined with respect to (s, t, φ), whereas the autoregressive proposal in Equation 5 is defined with respect to the relative Cartesian position $\tilde{\mathbf{r}}_i$. We therefore apply an analytic change of variables. For fixed prefix coordinates, let $J_i = \partial \mathcal{T}_i / \partial(s, t, \varphi)$ denote the local-to-Cartesian Jacobian. Its absolute determinant is

$$|J_i| = e^{3s} \operatorname{sech}^2(t). \tag{12}$$

The Cartesian conditional density is therefore

$$q_{\theta,i}^{\mathrm{cart}}\left(\tilde{\mathbf{r}}_i \mid \tilde{\mathbf{r}}_{<i}, \mathcal{G}\right) = \frac{p_i(s_i, t_i, \varphi_i \mid \eta_i)}{e^{3s_i} \operatorname{sech}^2(t_i)}. \tag{13}$$

Equivalently, its log density is

$$\log q_{\theta,i}^{\mathrm{cart}}\left(\tilde{\mathbf{r}}_i \mid \tilde{\mathbf{r}}_{<i}, \mathcal{G}\right) = \log p_i\left(s_i, t_i, \varphi_i \mid \eta_i\right) - 3s_i - \log \operatorname{sech}^2(t_i). \tag{14}$$

For a complete frame, only the bin containing φ_i contributes to the local density. Writing each conditional log density as $\ell_i = \log q_{\theta,i}^{\mathrm{cart}}$, the conformational log density over relative Cartesian coordinates is

$$\log q_\theta^{\mathrm{shape}}(\mathbf{r} \mid \mathcal{G}) = \sum_{i=2}^{N} \ell_i. \tag{15}$$

Thus, SEAL-BG generates each atom in a local coordinate system while retaining an analytically evaluable density over relative Cartesian coordinates. On nondegenerate coordinate charts, the normalized local conditionals, invariant density parameters, and covariant local frames ensure normalization and consistent transformation under proper rotations, as formalized in Proposition 1 and Appendix A.3.

3.3 Training Loss and Sampling

Training objective. For an effective batch of B MD configurations $\mathbf{r}^{(b)}$, $b = 1, \ldots, B$, we combine maximum likelihood with log-dispersion regularization (LDR) (Schopmans et al., 2026). Define $z_b = \log q_\theta^{\text{shape}}(\mathbf{r}^{(b)}) + u(\mathbf{r}^{(b)})$ and $\bar{z} = B^{-1} \sum_{b=1}^{B} z_b$. At training step m, the objective is

$$\mathcal{L} = -\frac{1}{3NB} \sum_{b=1}^{B} \log q_\theta^{\text{shape}} \left(\mathbf{r}^{(b)}\right) + \frac{\lambda(m)}{B} \sum_{b=1}^{B} |z_b - \bar{z}|. \tag{16}$$

The first term maximizes the likelihood of the observed MD configurations, while the second reduces the dispersion of $\log q_\theta^{\text{shape}} + u$ within each batch. For the target Boltzmann distribution, $\log p^{\text{eq}}(\mathbf{r}) + u(\mathbf{r}) = -\log Z$, which is constant over configuration space. LDR therefore encourages the learned density to satisfy the same relation on the observed MD support. It uses cached MD energies and requires no on-policy energy or force evaluations.

Teacher-forced evaluation and autoregressive sampling. During training and density evaluation, teacher forcing evaluates all next-atom conditionals in parallel on an observed configuration. During generation, atoms are sampled sequentially. At each step, the encoder processes the current Cartesian prefix, the density head samples local coordinates (s_i, t_i, φ_i), and Equation 10 reconstructs the next relative Cartesian position $\tilde{\mathbf{r}}_i$. The sampled atom is then appended to the prefix, and the procedure continues until the complete configuration has been generated. Sampling and rescoring use the same normalized proposal.

4 Experiments

Setup. Following the released ArBG setup (Rehman et al., 2026c), we evaluate AL3 (33 atoms), AL4 (42 atoms), AL6 (63 atoms), and Chignolin (138 atoms) using the same force-field parameters, datasets, and evaluation pipeline. For Trp-cage (284 atoms) and Homeodomain (947 atoms), both models use shared 300 K MD ensembles generated with CHARMM36 (Best et al., 2012) and OBC2 implicit solvent (Onufriev et al., 2004), with polar solvation disabled. Within each system, the models share the topology, evaluation protocol, and held-out reference ensemble. Simulation and dataset details are provided in Appendix B.1.

Metrics. Let $g(\mathbf{r})$ denote the evaluated quantity, $\mathcal{I}$ the retained generated-sample indices, M the number of MD reference samples, and δ_a a point mass at a. For corrected metrics, generated samples receive importance weights $w_i = \exp[-u(\mathbf{r}^{(i),\text{BG}})]/q_\theta(\mathbf{r}^{(i),\text{BG}})$, where $u = U/(k_\text{B}T)$; raw metrics retain all generated samples with equal weights. With $\overline{w}_i = w_i / \sum_{k\in\mathcal{I}} w_k$, the compared distributions are

$$P_{\text{BG}}^g = \sum_{i\in\mathcal{I}} \overline{w}_i \, \delta_{g(\mathbf{r}^{(i),\text{BG}})}, \qquad P_{\text{MD}}^g = \frac{1}{M} \sum_{j=1}^{M} \delta_{g(\mathbf{r}^{(j),\text{MD}})}. \tag{17}$$

We compute 2-Wasserstein distances using $g(\mathbf{r}) = u(\mathbf{r})$ for E-W_2 and the torsion-angle vector for T-W_2. Energy distances are reported in $k_\text{B}T$, and periodic torsion distances in radians; lower values are better.

For corrected metrics, we follow ArBG's clipped-SNIS protocol, discarding samples with the largest 0.2% of weights before renormalization and multinomial resampling (Rehman et al., 2026c). Clipping introduces additional bias, and correction need not improve every metric. Sampling counts, reporting windows, and component definitions are given in Appendix B.3.

Baselines. For AL3, AL4, AL6, and Chignolin, we use the baselines from Rehman et al. (2026c), including the enhanced equivariant continuous normalizing flow ECNF++, SBG (Tan et al., 2025), regression training of normalizing flows (RegFlow) (Rehman et al., 2026b), Few-Step Accurate Likelihoods for Continuous Flows (FALCON and FALCON-A) (Rehman et al., 2026a), Generative Infinite-Vocabulary Transformers (GIVT) (Tschannen et al., 2024), and molecular variants of MoL- and GMM-PixelCNN++ (Salimans et al., 2017; Rehman et al., 2026c). Our primary baseline is ArBG (Rehman et al., 2026c), which autoregressively models discretized Cartesian coordinate components with a causal Transformer and provides the closest architectural comparison to SEAL-BG.

For Trp-cage and Homeodomain, we train and evaluate ArBG using the same data and protocol as SEAL-BG. Further evaluation details are provided in Appendix B.3.

4.1 Alanine Peptides and Chignolin

Table 1: Results on alanine peptides and Chignolin. The labels E and T denote energy and torsion, respectively. Lower E-W_2 ($k_\mathrm{B}T$) and T-W_2 (radians) are better. Boldface and underlining denote the best and second-best values in each column, respectively.

Method	AL3		AL4		AL6		Chignolin	
	E-W_2 ↓	T-W_2 ↓	E-W_2 ↓	T-W_2 ↓	E-W_2 ↓	T-W_2 ↓	E-W_2 ↓	T-W_2 ↓
ECNF++	2.206	0.962	5.638	1.002	10.668	1.902	–	–
RegFlow	0.853	1.577	3.277	2.342	–	–	–	–
SBG (SMC)	0.598	0.503	1.007	1.039	1.189	1.444	10.819	3.778
FALCON-A	1.385	0.343	2.929	1.094	1.211	1.163	–	–
FALCON	0.544	0.452	0.686	0.858	0.892	1.256	–	–
GIVT	1.354	0.343	1.033	1.113	1.206	1.527	45.646	3.031
MoL-PixelCNN++	0.506	1.024	1.643	1.415	1.429	1.264	140.717	3.391
GMM-PixelCNN++	0.249	0.364	1.434	0.806	1.164	1.285	23.339	3.007
ArBG	<u>0.202</u>	<u>0.312</u>	<u>0.449</u>	<u>0.592</u>	<u>0.328</u>	<u>1.094</u>	**1.723**	**2.632**
SEAL-BG	**0.177**	0.316	**0.382**	**0.583**	**0.299**	1.146	<u>1.738</u>	<u>2.778</u>

In Table 1, SEAL-BG achieves the lowest E-W_2 on all three alanine systems and the lowest T-W_2 on AL4. On AL6, its lower energy distance comes with a higher torsional distance than ArBG, while on Chignolin it remains competitive with ArBG.

4.2 Trp-cage and Homeodomain

Figure 2(a,d) compares raw total-energy distributions; panels (c,f) report histogram overlap, ranging from 0 (no overlap) to 1 (identical histograms). The improvement over ArBG is modest on Trp-cage and larger on Homeodomain. SEAL-BG also has less high-energy overflow than ArBG, with 0.76% versus 3.40% on Trp-cage and 1.60% versus 11.88% on Homeodomain above $500\,k_\mathrm{B}T$ relative to the reference median. Additional generated conformations are shown in Appendix E.

Panels (c,f) show higher angle and dihedral energy overlaps but lower bond and Lennard–Jones (LJ) overlaps than ArBG. After clipped SNIS, SEAL-BG also obtains a lower energy Wasserstein distance than ArBG on both systems while maintaining comparable torsional distances, as shown in panels (b,e).

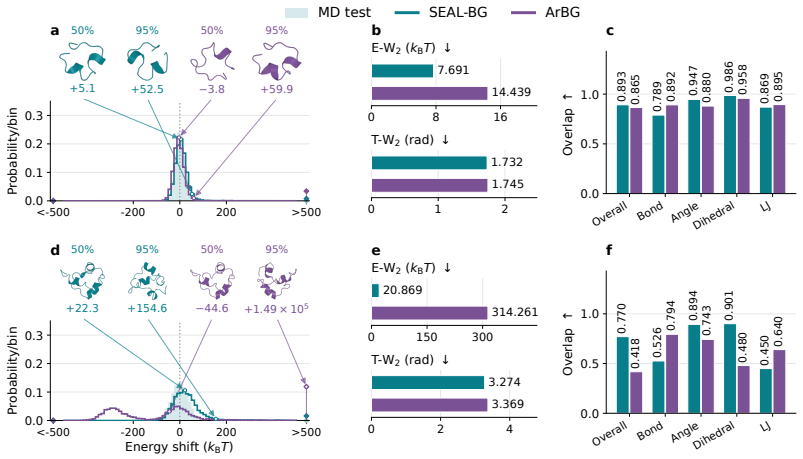


Figure 2: Proposal quality on Trp-cage (top) and Homeodomain (bottom). (a,d) Raw total-energy distributions, shifted by the MD reference median in $k_\mathrm{B}T$. Structures are generated samples near each model's 50th and 95th energy percentiles; signed labels below them give potential-energy offsets from the MD median in $k_\mathrm{B}T$. (b,e) Clipped-SNIS E-W$_2$ ($k_\mathrm{B}T$) and T-W$_2$ (radians); lower is better. (c,f) Raw overall and component energy overlaps; higher is better.

4.3 Computational Cost Analysis

Figure 3 compares sampling quality with training FLOPs and A100 training time. SEAL-BG has the lowest training cost among the plotted methods on every system, with the largest separation from ArBG on Chignolin and the smallest on Homeodomain. This efficiency likely stems from its symmetry- and geometry-aware inductive biases.

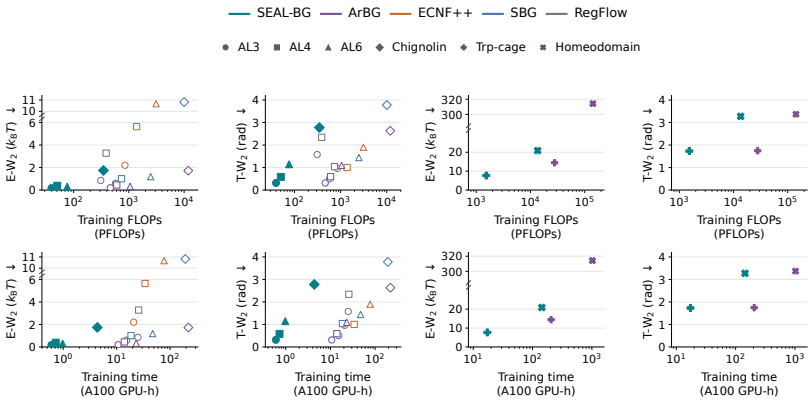


Figure 3: Sampling quality and computational cost across the evaluated molecular systems.

4.4 Application: Alchemical Free Energy

We estimate the polar-solvation coupling free energy of Trp-cage in OBC2 implicit solvent at 300 K by scaling only the charges entering the generalized Born (GB) term with $\lambda \in [0, 1]$:

$$u_\lambda(\mathbf{r}) = u_{\mathrm{MM}}(\mathbf{r}) + u_{\mathrm{np}}(\mathbf{r}) + \lambda^2 u_{\mathrm{GB}}(\mathbf{r}). \tag{18}$$

Here u_{MM} is the reduced CHARMM36 bonded and vacuum nonbonded energy, u_{np} the charge-independent nonpolar solvation term, and u_{GB} the GB polar term at full charges. The endpoints switch polar solvation off and on while retaining intramolecular electrostatics and nonpolar solvation, so the target is a polar-solvation coupling free energy rather than a full vacuum-to-solvent transfer free energy. Following Section 2.2, we train a normalized model at each endpoint and apply BAR between its generated samples and the corresponding MD samples to obtain $\Delta\hat{f} = \hat{f}_1 - \hat{f}_0$ (Bennett, 1976; Ding & Zhang, 2021). We compare against direct endpoint BAR and an 11-window MD/MBAR reference at $\lambda = 0, 0.1, \ldots, 1$ (Shirts & Chodera, 2008). Simulation and implementation details are provided in Appendix D.

Using 3,000 generated configurations per endpoint, SEAL-BG gives $\Delta\hat{f} = -213.376$, an absolute deviation of 8.762 from MBAR, compared with 14.315 for ArBG and 33.320 for direct endpoint BAR. Its across-seed SD is also smaller, 0.130 versus 0.495 for ArBG.

Figure 4 places the model estimates alongside the MD/MBAR window scan. The A100 timing projections give 146.698 GPU-hours for the shared two-endpoint MD baseline and approximately 7.5 additional GPU-hours for training both SEAL-BG endpoint models and generating one 3,000-sample set per endpoint. Its reported deviation from the reference is comparable to the five- to six-window MD/MBAR estimates, whose three or four additional MD windows add 220.046–293.395 projected GPU-hours.

5 Related Work

5.1 Boltzmann Generator

Boltzmann generators combine normalized learned proposals with importance weighting to target equilibrium ensembles (Noé et al., 2019). Their flow-based constructions build on invertible density transformations (Rezende & Mohamed, 2015; Dinh et al., 2017), incorporating molecular symmetries through equivariant flows (Köhler et al., 2020) and periodic coordinate topology

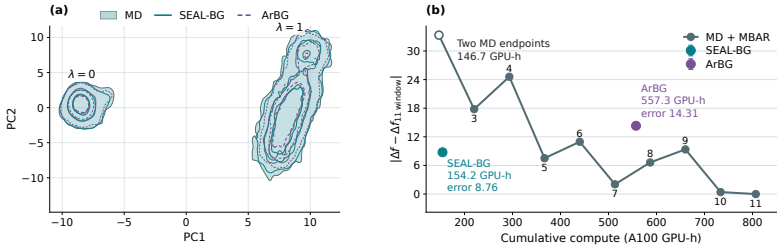


Figure 4: Trp-cage endpoint conformations and free-energy error versus estimated compute. (a) MD ensembles and model proposals for GB polar solvation switched off ($\lambda = 0$) and fully on ($\lambda = 1$). (b) The open point denotes the shared two-endpoint MD cost; model points add training and generation (3,000 samples per endpoint). Costs are A100 timing excluding energy evaluation, BAR, and search. The curve gives the 2–11-window MD/MBAR scan, labeled by window count.

through internal-coordinate flows (Köhler et al., 2021). Continuous flows (Chen et al., 2018) admit regression-based training such as flow matching (Lipman et al., 2023; Liu et al., 2023), although likelihood evaluation requires integrating the vector-field divergence (Lipman et al., 2024). Sequential Boltzmann generators combine Transformer flows with sequential Monte Carlo and provide the enhanced equivariant-flow baseline ECNF++ (Tan et al., 2025); RegFlow regresses transport targets (Rehman et al., 2026b), FALCON and FALCON-A target few-step flow evaluation (Rehman et al., 2026a), and BoltzNCE learns approximate likelihoods (Aggarwal et al., 2025).

5.2 SE(3)-Equivariant Geometric Networks

Tensor field networks encode scalar, vector, and higher-order tensor features (Thomas et al., 2018), SE(3)-Transformers introduce equivariant attention (Fuchs et al., 2020), and equivariant graph neural networks provide coordinate-based message passing (Satorras et al., 2021). Our causal encoder uses the scalar–vector representation of a memory-efficient equivariant Transformer (Jiao et al., 2026), while separate local-coordinate heads supply normalized conditionals.

5.3 Autoregressive Models for Molecular Generation

Autoregressive density models use tractable conditionals, including discretized logistic mixtures in PixelCNN++ (Salimans et al., 2017) and Gaussian mixtures in GIVT (Tschannen et al., 2024). For three-dimensional molecular generation, G-SchNet uses rotation-invariant geometric representations (Gebauer et al., 2019), Symphony parameterizes spatial distributions with spherical harmonics (Daigavane et al., 2024), and Quetzal couples a causal Transformer to a diffusion position head (Cheng et al., 2026). InertialAR uses canonical inertial frames and hierarchical atom decoding (Li et al., 2026), whereas ArBG models Cartesian coordinate tokens and adapts mixture-density baselines to Boltzmann generation (Rehman et al., 2026c). SEAL-BG conditions analytically normalized local-coordinate densities on an SE(3)-equivariant Cartesian prefix representation.

6 Conclusion

SEAL-BG combines SE(3)-equivariant geometric conditioning with local-coordinate autoregression, incorporating molecular symmetry and geometry into a model with analytically evaluable densities. Across peptide and protein benchmarks, it achieves competitive sampling metrics and improved raw total-energy matching on larger systems, with substantially lower training compute.

Tractable densities also connect molecular generation to thermodynamic estimation by allowing SEAL-BG to serve as a learned reference distribution. In the Trp-cage case study, SEAL-BG yields free-energy estimates approaching the reference, with substantially lower additional training and generation costs than the intermediate-state MD sampling when endpoint MD data are shared.

Together, these findings highlight the potential of geometry-aware autoregressive models to support both molecular sampling and quantitative thermodynamic calculations.

AI USE STATEMENT

Generative AI tools were used to assist with manuscript organization, translation, drafting, and language editing, including improving clarity and concision, refining transitions, checking the consistency of notation and equations, and editing tables and figure captions. They were also used to help organize experiment records for presentation, but were not used to generate experimental data or numerical results. All AI-assisted material was reviewed and revised by the authors, who take full responsibility for the final content of the paper.

REPRODUCIBILITY STATEMENT

The model architecture, density construction, training objective, and autoregressive sampling procedure are described in Section 3. Appendix A provides the encoder details, density-head parameterization, and derivations supporting normalization and geometric covariance. Appendix B documents the molecular systems, dataset splits, MD protocol, training hyperparameters, checkpoint-selection criterion, sampling procedure, and evaluation metrics. Additional sampling results and energy-component distributions are reported in Appendix C, the free-energy case study is specified in Appendix D.

REFERENCES

Rishal Aggarwal, Jacky Chen, Nicholas Matthew Boffi, and David Koes. BoltzNCE: Learning likelihoods for boltzmann generation with stochastic interpolants and noise contrastive estimation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=FXTg2P8OQz.

C. H. Bennett. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *Journal of Computational Physics*, 22(2):245–268, October 1976. doi: 10.1016/0021-9991(76)90078-4.

Robert B. Best, Xiao Zhu, Jihyun Shim, Pedro E. M. Lopes, Jeetain Mittal, Michael Feig, and Jr. MacKerell, Alexander D. Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles. *Journal of Chemical Theory and Computation*, 8(9):3257–3273, 09 2012. ISSN 1549-9618. doi: 10.1021/ct300400x. URL https://doi.org/10.1021/ct300400x.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf.

Austin Henry Cheng, Chong Sun, and Alan Aspuru-Guzik. Scalable autoregressive 3d molecule generation, 2026. URL https://openreview.net/forum?id=AxdOmqDdIo.

Ameya Daigavane, Song Eun Kim, Mario Geiger, and Tess Smidt. Symphony: Symmetry-equivariant point-centered spherical harmonics for 3d molecule generation. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=MIEnYtlGyv.

Xinqiang Ding and Bin Zhang. Deepbar: A fast and exact method for binding free energy computation. *The Journal of Physical Chemistry Letters*, 12(10):2509–2515, 03 2021. ISSN 1948-7185. doi: 10.1021/acs.jpclett.1c00189. URL https://doi.org/10.1021/acs.jpclett.1c00189.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real NVP. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=HkpbnH9lx.

Peter Eastman, Raimondas Galvelis, Raúl P. Peláez, Charlles R. A. Abreu, Stephen E. Farr, Emilio Gallicchio, Anton Gorenko, Michael M. Henry, Frank Hu, Jing Huang, Andreas Krämer, Julien Michel, Joshua A. Mitchell, Vijay S. Pande, João PGLM Rodrigues, Jaime Rodriguez-Guerra,

Andrew C. Simmonett, Sukrit Singh, Jason Swails, Philip Turner, Yuanqing Wang, Ivy Zhang, John D. Chodera, Gianni De Fabritiis, and Thomas E. Markland. Openmm 8: Molecular dynamics simulation with machine learning potentials. *The Journal of Physical Chemistry B*, 128(1):109–116, 01 2024. ISSN 1520-6106. doi: 10.1021/acs.jpcb.3c06662. URL https://doi.org/10.1021/acs.jpcb.3c06662.

Daan Frenkel and Berend Smit. Understanding molecular simulation: from algorithms to applications. 1996. URL https://api.semanticscholar.org/CorpusID:62612337.

Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. Se(3)-transformers: 3d roto-translation equivariant attention networks. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 1970–1981. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/15231a7ce4ba789d13b722cc5c955834-Paper.pdf.

Niklas Gebauer, Michael Gastegger, and Kristof Schütt. Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/a4d8e2a7e0d0c102339f97716d2fdfb6-Paper.pdf.

Jérôme Hénin, Tony Lelièvre, Michael R. Shirts, Omar Valsson, and Lucie Delemotte. Enhanced sampling methods for molecular dynamics simulations [article v1.0]. *Living Journal of Computational Molecular Science*, 4(1):1583, December 2022. ISSN 2575-6524. doi: 10.33011/livecoms.4.1.1583. URL http://dx.doi.org/10.33011/livecoms.4.1.1583.

Rui Jiao, Xiangzhe Kong, Yinjun Jia, Yijia Zhang, Ziyi Yang, Yang Liu, and Jianzhu Ma. Scalable peptide design via memory-efficient equivariant transformer, 2026. URL https://arxiv.org/abs/2606.25006.

Keller Jordan, Yuchen Jin, Vlado Boza, Jiacheng You, Franz Cesista, Laker Newhouse, and Jeremy Bernstein. Muon: An optimizer for hidden layers in neural networks, 2024. URL https://kellerjordan.github.io/posts/muon/.

Jonas Köhler, Leon Klein, and Frank Noe. Equivariant flows: Exact likelihood generative learning for symmetric densities. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 5361–5370. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/kohler20a.html.

Jonas Köhler, Andreas Krämer, and Frank Noe. Smooth normalizing flows. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 2796–2809. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/167434fa6219316417cd4160c0c5e7d2-Paper.pdf.

Haorui Li, weitao Du, Yuqiang Li, Hongyu Guo, and Shengchao Liu. InertialAR: Autoregressive 3d molecule generation with inertial frames, 2026. URL https://openreview.net/forum?id=de4bbZedv3.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling, 2023. URL https://arxiv.org/abs/2210.02747.

Yaron Lipman, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky T. Q. Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat. Flow matching guide and code, 2024. URL https://arxiv.org/abs/2412.06264.

Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=XVjTT1nw5z.

Jr. MacKerell, A. D., D. Bashford, M. Bellott, Jr. Dunbrack, R. L., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, 04 1998. ISSN 1520-6106. doi: 10.1021/jp973084f. URL https://doi.org/10.1021/jp973084f.

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. URL https://www.science.org/doi/abs/10.1126/science.aaw1147.

Alexey Onufriev, Donald Bashford, and David A. Case. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics*, 55(2):383–394, 2004. doi: https://doi.org/10.1002/prot.20033. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.20033.

Art B. Owen. *Monte Carlo theory, methods and examples*. https://artowen.su.domains/mc/, 2013.

Danyal Rehman, Tara Akhound-Sadegh, Artem Gazizov, Yoshua Bengio, and Alexander Tong. FALCON: Few-step accurate likelihoods for continuous flows. In *The Fourteenth International Conference on Learning Representations*, 2026a. URL https://openreview.net/forum?id=FbssShlI4N.

Danyal Rehman, Oscar Davis, Jiarui Lu, Jian Tang, Michael M. Bronstein, Yoshua Bengio, Alexander Tong, and Joey Bose. Efficient regression-based training of normalizing flows for boltzmann generators. In *The Fourteenth International Conference on Learning Representations*, 2026b. URL https://openreview.net/forum?id=ctdnzPxDI3.

Danyal Rehman, Charlie B. Tan, Yoshua Bengio, Joey Bose, and Alexander Tong. Autoregressive boltzmann generators. In *Forty-third International Conference on Machine Learning*, 2026c. URL https://openreview.net/forum?id=75AYDsndHP.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In Francis Bach and David Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 1530–1538, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/rezende15.html.

Tim Salimans, Andrej Karpathy, Xi Chen, and Diederik P. Kingma. PixelCNN++: Improving the pixelCNN with discretized logistic mixture likelihood and other modifications. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=BJrFC6ceg.

Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9323–9332. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/satorras21a.html.

Henrik Schopmans, Christopher von Klitzing, and Pascal Friederich. Efficient training of boltzmann generators using off-policy log-dispersion regularization. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=cKoqWcyVvd.

David E. Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, Michael P. Eastwood, Joseph A. Bank, John M. Jumper, John K. Salmon, Yibing Shan, and Willy Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, 2010. doi: 10.1126/science.1187409. URL https://www.science.org/doi/abs/10.1126/science.1187409.

Michael R. Shirts and John D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics*, 129(12), September 2008. ISSN 1089-7690. doi: 10.1063/1.2978177. URL http://dx.doi.org/10.1063/1.2978177.

Saifuddin Syed, Alexandre Bouchard-Côté, George Deligiannidis, and Arnaud Doucet. Non-reversible parallel tempering: A scalable highly parallel mcmc scheme. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(2):321–350, 04 2022. ISSN 1369-7412. doi: 10.1111/rssb.12464. URL https://doi.org/10.1111/rssb.12464.

Charlie B. Tan, Joey Bose, Chen Lin, Leon Klein, Michael M. Bronstein, and Alexander Tong. Scalable equilibrium sampling with sequential boltzmann generators. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Forty-second International Conference on Machine Learning, ICML 2025, Vancouver, BC, Canada, July 13-19, 2025*, volume 267 of *Proceedings of Machine Learning Research*. PMLR / OpenReview.net, 2025. URL https://proceedings.mlr.press/v267/tan25a.html.

Nathaniel Thomas, Tess E. Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. *CoRR*, abs/1802.08219, 2018. URL http://arxiv.org/abs/1802.08219.

Michael Tschannen, Cian Eastwood, and Fabian Mentzer. Givt: Generative infinite-vocabulary transformers, 2024. URL https://arxiv.org/abs/2312.02116.

Mark E. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press, 08 2023. ISBN 9780198825562. doi: 10.1093/oso/9780198825562.001.0001. URL https://doi.org/10.1093/oso/9780198825562.001.0001.

Peter Wirnsberger, Andrew J. Ballard, George Papamakarios, Stuart Abercrombie, Sébastien Racanière, Alexander Pritzel, Danilo Jimenez Rezende, and Charles Blundell. Targeted free energy estimation via learned mappings. *The Journal of Chemical Physics*, 153(14), October 2020. ISSN 1089-7690. doi: 10.1063/5.0018903. URL http://dx.doi.org/10.1063/5.0018903.

Biao Zhang and Rico Sennrich. Root mean square layer normalization. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/1e8a19426224ca89e83cef47f1e7f53b-Paper.pdf.

Zhijun Zhang, Xinzijian Liu, Kangyu Yan, Mark E. Tuckerman, and Jian Liu. Unified efficient thermostat scheme for the canonical ensemble with holonomic or isokinetic constraints via molecular dynamics. *The Journal of Physical Chemistry A*, 123(28):6056–6079, 07 2019. ISSN 1089-5639. doi: 10.1021/acs.jpca.9b02771. URL https://doi.org/10.1021/acs.jpca.9b02771.

A METHOD DETAILS

A.1 CARTESIAN PREFIX ENCODER

The encoder interface in Equation 6 consumes relative Cartesian coordinates together with fixed molecular metadata. At prefix position j, the coordinate input is $\tilde{\mathbf{r}}_j$, while the metadata identify both the source atom j and the next target atom $j+1$. Element, residue, atom-role, and sequence-position embeddings encode these identities, and causal bond connectivity is supplied by the molecular topology. No target coordinate is provided.

The output at prefix position j consists of scalar features $h_{j+1} \in \mathbb{R}^d$ and vector features $H_{j+1} \in \mathbb{R}^{3\times d}$, indexed by the next atom they predict. Their transformation law is given in Equation 7. To predict atom i, the density head reads h_i and H_i from prefix position $i-1$ and constructs η_i using Equation 8. The main-text notation $\mathcal{E}_\theta$ selects this final-prefix output rather than the complete sequence of encoder features.

With stochastic regularization disabled, causality requires the features at position $i-1$ to be identical whether they are computed within a longer observed sequence or from the prefix ending at that position. A teacher-forced pass can therefore evaluate all next-atom conditionals in parallel, whereas autoregressive generation exposes one prefix at a time. The same encoder and density head are used for teacher-forced evaluation and cached incremental generation.

The encoder contains eight layers, 256 channels, and eight attention heads. It operates on relative Cartesian geometry; local coordinates parameterize the output density but do not replace the Cartesian prefix representation.

A.2 LOCAL FRAMES AND DENSITY PARAMETERIZATION

Causal reference atoms. For each atom i, the molecular topology and fixed generation order specify a parent $p(i)$, an angle anchor $a(i)$, and, when available, a torsion anchor $b(i)$. These reference maps are fixed before training and reused during training, density evaluation, and generation. Every available reference precedes atom i in the generation order:

$$p(i) < i, \qquad a(i) < i, \qquad b(i) < i.$$

Consequently, the local frame never depends on a future coordinate. For atoms with all three references available, the references must define a nondegenerate frame. The frame, internal coordinates, and reconstruction map are defined in Equations 9–10. The initial atoms, for which one or more anchors are unavailable, are handled by the normalized conditionals in Appendix A.3.

Density-head parameterization. The reported model uses $K = 64$ piecewise-uniform torsion bins in Equation 11. Let $(\rho_i^0, \vartheta_i^0)$ denote the force-field reference radius and angle for atom i, expressed in model units. Learned weights W_w and bias b_w produce the torsion-bin probabilities. For each bin k, learned affine maps f_k^ν and g_k^ν produce the mean residuals and log standard deviations for $\nu \in \{\rho, \mathrm{ang}\}$:

$$\begin{aligned} w_i &= \mathrm{softmax}(W_w \eta_i + b_w), \\ \log \sigma_{ik}^\nu &= \underline{\ell}_\nu + \left(\overline{\ell}_\nu - \underline{\ell}_\nu\right) \mathrm{sigmoid}\left(g_k^\nu(\eta_i)\right), \\ \mu_{ik}^\rho &= \log \rho_i^0 + b_\rho \tanh\left(f_k^\rho(\eta_i)\right), \\ \mu_{ik}^{\mathrm{ang}} &= \mathrm{atanh}\left(\cos \vartheta_i^0\right) + b_{\mathrm{ang}} \tanh\left(f_k^{\mathrm{ang}}(\eta_i)\right). \end{aligned} \tag{19}$$

The residual bounds are

$$b_\rho = 0.15, \qquad b_{\mathrm{ang}} = 0.75,$$

and the log-standard-deviation ranges are

$$[\underline{\ell}_\rho, \overline{\ell}_\rho] = [-5.5, -2.0], \qquad [\underline{\ell}_{\mathrm{ang}}, \overline{\ell}_{\mathrm{ang}}] = [-4.5, -0.5].$$

Reference radii are divided by the model coordinate unit before their logarithms are taken. These bounded residuals keep the predicted radial and angular distributions near the force-field reference geometry while allowing their parameters to depend on the complete generated prefix through η_i.

A.3 Normalization and Geometric Properties

Proposition 1 (Normalization and geometric covariance) *On nondegenerate coordinate charts, the normalized local conditionals define a normalized conformational density over relative Cartesian coordinates. If the conditional parameters are rotation invariant and the causal local frames transform covariantly under proper rotations, the corresponding Cartesian conditionals are covariant under those rotations.*

The result follows from conditional change of variables, the block-triangular structure of the sequential coordinate map, and iterative integration.

Chart determinant and conditional normalization. For fixed causal anchors, spherical coordinates satisfy

$$\mathrm{d}\tilde{\mathbf{r}} = \rho^2 \sin\vartheta \,\mathrm{d}\rho \,\mathrm{d}\vartheta \,\mathrm{d}\varphi.$$

Because

$$\mathrm{d}\rho = \rho \,\mathrm{d}s, \qquad \left|\frac{\mathrm{d}\vartheta}{\mathrm{d}t}\right| = \sin\vartheta,$$

the absolute determinant of the reconstruction map in Equation 10 is

$$|J_i| = \rho^3 \sin^2\vartheta = e^{3s}\,\mathrm{sech}^2(t). \tag{20}$$

Each mixture component in Equation 11 integrates to its bin probability w_{ik}, and $\sum_k w_{ik} = 1$. The coordinate chart is one-to-one away from the polar axis and the identified torsion boundary, both of which are null sets under the continuous density. Applying the change-of-variables formula therefore gives

$$\int q^{\mathrm{cart}}_{\theta,i}\left(\tilde{\mathbf{r}}_i \mid \tilde{\mathbf{r}}_{<i}, \mathcal{G}\right) \mathrm{d}\tilde{\mathbf{r}}_i = 1. \tag{21}$$

For the complete sequential transformation, each $\tilde{\mathbf{r}}_i$ depends on its own local coordinates and on preceding coordinates, but not on future coordinates. The joint Jacobian is therefore block triangular, with diagonal blocks J_i. Derivatives through the prefix contribute only off-diagonal blocks and do not alter the product of diagonal determinants. Integrating Equation 5 from $\tilde{\mathbf{r}}_N$ backwards gives

$$\int q^{\mathrm{shape}}_{\theta}\left(\tilde{\mathbf{r}}_{2:N} \mid \mathcal{G}\right) \mathrm{d}\tilde{\mathbf{r}}_{2:N} = 1. \tag{22}$$

No invertibility of the neural conditioning network is required.

Incomplete local frames. For the initial atoms, a complete local frame may not yet be available. We abbreviate $q^{\mathrm{cart}}_{\theta,i}$ as q^{cart}_i and suppress its fixed prefix and topology arguments.

If an atom has a parent but no angle anchor, its direction is uniform over the sphere:

$$q^{\mathrm{cart}}_i(\tilde{\mathbf{r}}_i) = \frac{\sum_{k=1}^{K} w_{ik}\mathcal{N}\left(s; \mu^{\rho}_{ik}, (\sigma^{\rho}_{ik})^2\right)}{4\pi\rho^3}. \tag{23}$$

If an angle anchor is available but no torsion anchor is available, the azimuthal direction is uniform:

$$q^{\mathrm{cart}}_i(\tilde{\mathbf{r}}_i) = \frac{\sum_{k=1}^{K} w_{ik}\mathcal{N}\left(s; \mu^{\rho}_{ik}, (\sigma^{\rho}_{ik})^2\right)\mathcal{N}\left(t; \mu^{\mathrm{ang}}_{ik}, (\sigma^{\mathrm{ang}}_{ik})^2\right)}{2\pi\rho^3\,\mathrm{sech}^2(t)}. \tag{24}$$

The factors 4π and 2π normalize the uniform solid-angle and azimuthal distributions, respectively. These conditionals avoid assigning a physical torsion to an atom for which no torsion-defining frame exists.

Geometric covariance. Let $R \in \mathrm{SO}(3)$ be a proper rotation and let $\mathbf{d}$ be a global translation:

$$\mathbf{r}'_i = R\mathbf{r}_i + \mathbf{d}.$$

The relative coordinates satisfy

$$\tilde{\mathbf{r}}'_i = \mathbf{r}'_i - \mathbf{r}'_1 = R\tilde{\mathbf{r}}_i,$$

so the global translation cancels. Under the rotation, Equation 9 gives

$$e_j \mapsto Re_j, \qquad j = 1, 2, 3,$$

where the cross-product relation uses $\det R = 1$.

Let $\mathcal{T}'_i$ denote the reconstruction map constructed from the rotated prefix. Because the conditional parameters are rotation invariant,

$$\mathcal{T}'_i(s, t, \varphi) = R\mathcal{T}_i(s, t, \varphi).$$

Since $|\det R| = 1$, the corresponding Cartesian conditional satisfies

$$q^{\mathrm{cart}}_{\theta,i}\left(R\tilde{\mathbf{r}}_i \mid R\tilde{\mathbf{r}}_{<i}, \mathcal{G}\right) = q^{\mathrm{cart}}_{\theta,i}\left(\tilde{\mathbf{r}}_i \mid \tilde{\mathbf{r}}_{<i}, \mathcal{G}\right). \tag{25}$$

The uniform orientation factors for incomplete frames obey the same relation. This proves the geometric statement in Proposition 1.

Global translations leave the relative-coordinate representation unchanged. Reflections are not included in the symmetry assumption: the right-handed local frames retain signed torsions and therefore distinguish mirror configurations.

Coordinate units. Let ℓ denote the physical length represented by one model coordinate unit, so that $\tilde{\mathbf{r}}^{\mathrm{phys}} = \ell\tilde{\mathbf{r}}$. Because the conformational density contains $N - 1$ three-dimensional relative coordinates, its value in physical units is

$$\log q^{\mathrm{shape}}_{\theta,\mathrm{phys}} = \log q^{\mathrm{shape}}_{\theta,\mathrm{model}} - 3(N-1)\log \ell. \tag{26}$$

Proposal and target densities must be evaluated with respect to the same coordinate units when constructing importance weights.

Numerical support. The preceding statements apply to nondegenerate coordinate charts in exact arithmetic. The implementation guards radius and frame norms at 10^{-6} and clamps cosine values near the poles. These numerical safeguards prevent undefined operations but are not invertible coordinate transformations; they are therefore distinct from the analytic normalization result above.

B Experimental Protocol

B.1 Datasets and Molecular Systems

Table 2 summarizes the systems and dataset splits used in the experiments. The released benchmark regime follows the data configuration of ArBG and contains AL3, AL4, AL6, and Chignolin. The scale-up regime uses the TC10b Trp-cage structure from PDB ID 2JOF and the Engrailed Homeodomain structure from PDB ID 1ENH.

Table 2: Dataset splits and molecular sizes used in the experiments. Counts denote MD frames.

Source	System	Atoms	Train	Val.	Test
ArBG release	AL3	33	100,000	20,000	180,000
	AL4	42	100,000	20,000	580,000
	AL6	63	100,000	20,000	180,000
	Chignolin	138	100,000	20,000	20,000
This study	Trp-cage	284	200,000	25,000	25,000
	Homeodomain	947	200,000	25,000	25,000

The Trp-cage system contains 20 residues and is constructed from PDB ID 2JOF, chain A, residues 1–20. The Homeodomain system contains 54 residues and 947 atoms and is constructed from PDB ID 1ENH, chain A, residues 2–55. Both systems are prepared with charged termini.

For the scale-up systems, both methods use the same molecular topology, force-field parameters, MD ensemble, and energy-evaluation procedure. The dynamics are generated with OpenMM using CHARMM36 and OBC2 implicit solvent at 300 K (Eastman et al., 2024). The solute and solvent dielectric constants are 1 and 78.5, respectively; the Debye screening parameter is zero, and non-bonded interactions use no cutoff. Both systems follow Equation 18: only GB particle charges are multiplied by λ, giving a λ^2 dependence of the polar GB energy at fixed coordinates. The charge-independent nonpolar surface-area term, bonded terms, Lennard–Jones interactions, and direct Coulomb interactions remain unchanged. All scale-up sampling-quality results use the $\lambda = 0$ ensemble, which retains nonpolar solvation and intramolecular electrostatics. The Trp-cage free-energy application additionally uses $\lambda = 1$ and intermediate ensembles; these are thermodynamic states with different Hamiltonians, not predefined conformational basins. A fixed split is used for training, validation, and testing.

After minimization, Langevin-middle dynamics (Zhang et al., 2019) use 300 K, 1 ps^{-1} friction, a 4 fs step, and 4 amu hydrogen mass repartitioning. We discard 500,000 equilibration steps (2 ns) and run 2.5×10^9 production steps (10 μs) per λ window, saving one frame every 10,000 steps (40 ps) for 250,000 frames per window. Each window evolves at fixed λ without an umbrella bias.

B.2 Training and Autoregressive Sampling

Table 3: Model architecture, density parameterization, and training hyperparameters used for SEAL-BG. For each molecular system, the checkpoint with the lowest validation loss is selected.

Item	Value
Model architecture	
Geometric trunk depth	8 layers
Feature channels	256
Attention heads	8
Attention head dimension	32
Feed-forward expansion	3
Causal bond-adapter layers	1
Local density parameterization	
Torsion density	64 piecewise-uniform periodic bins
Radial residual bound	0.15
Radial log-scale range	$[-5.5, -2.0]$
Angular residual bound	0.75
Angular log-scale range	$[-4.5, -0.5]$
Optimization	
Effective batch size	512
MD-frame sampling	Independent sampling with replacement
Optimizer	Muon
Learning rate	0.003
Weight decay	0.01
Learning-rate schedule	One-cycle cosine
Schedule warmup fraction	0.05
Numerical precision	FP32 with TensorFloat-32 enabled
LDR coefficient	0.00247619
Training budget and model selection	
Training steps (AL3 / AL4 / AL6 / Chignolin / Trp-cage / Homeodomain)	10k / 10k / 10k / 20k / 30k / 60k
Checkpoint selection	Lowest validation loss

Training configuration. The training hyperparameters are summarized in Table 3. Each molecular system is trained with an effective batch size of 512, random rotation augmentation, and Muon

optimization (Jordan et al., 2024). We train for 10k updates on each of AL3, AL4, and AL6, 20k updates on Chignolin, 30k updates on Trp-cage, and 60k updates on Homeodomain. The LDR coefficient $\lambda(m)$ follows a cosine warmup and reaches 0.00247619 after 20k updates. Among the saved checkpoints for each system, we select the checkpoint with the lowest validation loss, without using test-set metrics for model selection.

Batch-centered regularization. The LDR center in Equation 16 is the effective-batch mean rather than a median or a detached constant. Define the centered residual

$$\delta_b = z_b - \bar{z}$$

and the unweighted regularizer

$$\mathcal{L}_{\mathrm{LDR}} = \frac{1}{B} \sum_b |\delta_b|.$$

A subgradient is

$$\nabla_\theta \mathcal{L}_{\mathrm{LDR}} = \frac{1}{B} \sum_b \left[\mathrm{sign}(\delta_b) - \frac{1}{B} \sum_j \mathrm{sign}(\delta_j) \right] \nabla_\theta \log q_\theta^{\mathrm{shape}} \left(\mathbf{r}^{(b)} \right). \tag{27}$$

The energy labels are fixed. This expression includes the derivative through the batch mean and is invariant to a constant batchwise shift of the log density. Consistency on observed MD configurations does not by itself guarantee coverage of unobserved conformational basins.

Autoregressive sampling. Generation begins with the first relative coordinate fixed at $\tilde{\mathbf{r}}_1 = 0$. At each subsequent step, the geometric encoder processes the current Cartesian prefix, and the local density head produces the parameters of the next-atom conditional distribution. For an atom with a complete local frame, let k_i denote its sampled torsion-bin index. The local variables are sampled as

$$\begin{aligned} k_i &\sim \mathrm{Cat}(w_i), \\ s_i &\sim \mathcal{N}\left(\mu^{\rho}_{ik_i}, (\sigma^{\rho}_{ik_i})^2\right), \\ t_i &\sim \mathcal{N}\left(\mu^{\mathrm{ang}}_{ik_i}, (\sigma^{\mathrm{ang}}_{ik_i})^2\right), \\ \varphi_i &\sim \mathrm{Unif}(B_{k_i}). \end{aligned} \tag{28}$$

Here, $\mathrm{Cat}(w_i)$ is the categorical distribution over torsion bins, and $\mathrm{Unif}(B_{k_i})$ is uniform within the selected bin. For the initial atoms without a complete local frame, we use the normalized orientation conditionals in Appendix A.3.

Equation 10 maps the sampled local variables (s_i, t_i, φ_i) to the next relative Cartesian position $\tilde{\mathbf{r}}_i$, which is then appended to the prefix. The procedure continues until all atoms have been generated. When centered Cartesian coordinates are required, they are recovered as

$$\mathbf{r}_i = \tilde{\mathbf{r}}_i - \frac{1}{N} \sum_{j=1}^{N} \tilde{\mathbf{r}}_j. \tag{29}$$

Summing the Cartesian conditional log densities in Equation 14 over the generated atoms gives the conformational log density $\log q_\theta^{\mathrm{shape}}$ in Equation 15. Sampling and rescoring use the same normalized conditional distributions.

B.3 Evaluation

Evaluation sample counts. For the released peptide benchmarks, each model generates 200,000 candidate conformations and the released ArBG evaluator computes energy and torsion distances on 10,000 configurations (Rehman et al., 2026c). Trp-cage uses 10,000 generated configurations and 10,000 shared MD reference frames per method, whereas Homeodomain uses 5,000 of each.

Benchmark distances. The upstream evaluator is pinned to AutoBG commit `21624a80504b3199b291514c37a49cccd19c8817`. E-W_2 is the square root of the one-dimensional optimal-transport cost; energy distances are reported in units of $k_{\mathrm{B}}T$, with numerical values computed from the reduced potential u. For torsion vectors ϕ, ψ with torsions indexed by j, let d^2_{tor} denote the squared periodic ground cost used by T-W_2:

$$d^2_{\mathrm{tor}}(\phi, \psi) = \sum_j \min(|\phi_j - \psi_j|, 2\pi - |\phi_j - \psi_j|)^2. \tag{30}$$

Clipped SNIS discards the largest 0.2% of weights before resampling.

Large-system distributions. The common MD reference is fixed within each system and shared across methods. Trp-cage uses 10,000 proposals/reference frames and Homeodomain uses 5,000. Energy plots subtract the common MD median and retain underflow and overflow mass. Regular-bin probabilities are divided by the full sample count. Energy decompositions and model densities are evaluated on the same proposal coordinates.

Mirror-related energy bimodality. The two raw Homeodomain ArBG energy modes in Figure 6 are associated with populations of opposite molecular handedness. At the reported operating point, 2,445 of the 5,000 proposals have all 53 non-glycine Cα chirality signs inverted relative to the MD reference, 2,417 match at every center, and 138 have mixed signs. Here chirality is diagnosed by the sign of the scalar triple product of the N–Cα, C–Cα, and Cβ–Cα vectors; this Cα diagnostic is distinct from the upstream evaluator's broader topology-based detector. The fully inverted and reference-handed groups have median total energies of -3510.6 and $-3234.0\,k_{\mathrm{B}}T$, respectively, and median dihedral energies of 726.8 and $1006.1\,k_{\mathrm{B}}T$. Their bond, angle, Lennard–Jones, and electrostatic component medians each differ by less than $2.2\,k_{\mathrm{B}}T$. These associations support a mirror-related, torsion-dominated interpretation of the splitting and show incomplete stereochemical selectivity of the trained ArBG proposal on this system.

C Additional Results

C.1 Peptide Metrics

Table 4 reports the raw and clipped-SNIS distances for the selected SEAL-BG peptide configurations. The two sets of metrics characterize direct proposal quality and post-correction quality, respectively, and are not interchangeable.

Table 4: Small-system SEAL-BG metrics before and after clipped SNIS. E and T denote E-W_2 ($k_{\mathrm{B}}T$) and T-W_2 (radians). Each row uses 200,000 proposals and 10,000 metric samples.

System	Raw E	Raw T	Clipped E	Clipped T
AL3	156.457	0.328	0.177	0.316
AL4	153.265	0.564	0.382	0.583
AL6	148.609	1.049	0.299	1.146
Chignolin	477.621	2.758	1.738	2.778

C.2 Large-System Metrics

Tables 5 and 6 report raw component overlaps and raw or clipped-SNIS distances.

Table 5: Raw large-system energy-component and total-energy overlaps. Both methods use a common fixed MD reference within each system.

System	Method	Bond	Angle	Dihedral	LJ	Coulomb	Total
Trp-cage	SEAL-BG	0.7894	0.9468	0.9861	0.8691	0.7649	0.8926
Trp-cage	ArBG	0.8921	0.8799	0.9578	0.8948	0.9574	0.8651
Homeodomain	SEAL-BG	0.5258	0.8936	0.9008	0.4496	0.8752	0.7704
Homeodomain	ArBG	0.7938	0.7430	0.4798	0.6402	0.8846	0.4176

Table 6: Large-system W_2 metrics. Clipped E and T use the released top-0.2% weight removal and multinomial-resampling convention.

System	Method	Raw E	Raw T	Clipped E	Clipped T
Trp-cage	SEAL-BG	98.5360	1.2299	7.6908	1.7323
Trp-cage	ArBG	210.5458	1.3897	14.4391	1.7447
Homeodomain	SEAL-BG	113.4645	2.7749	20.8695	3.2738
Homeodomain	ArBG	360.2891	3.4532	314.2612	3.3691

C.3 Energy-Component Distributions

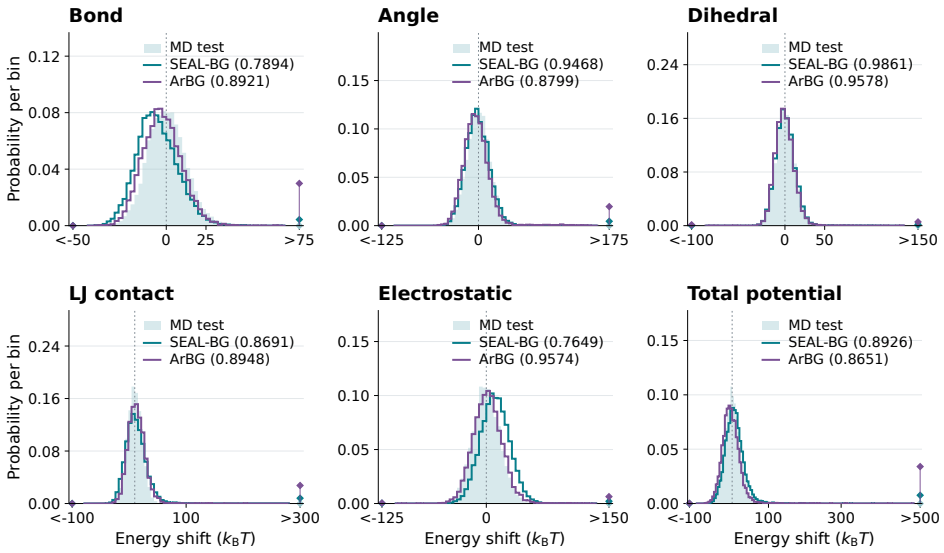


Figure 5: Trp-cage raw component distributions for SEAL-BG, ArBG, and 10,000 fixed MD test frames. Each model distribution contains 10,000 configurations. Legends give formal overlap; diamonds retain all probability outside the display window. Angle includes Urey–Bradley terms, dihedral includes improper and CMAP terms, and electrostatics include the recorded implicit-solvent contribution. No energy-based sample filtering is applied.

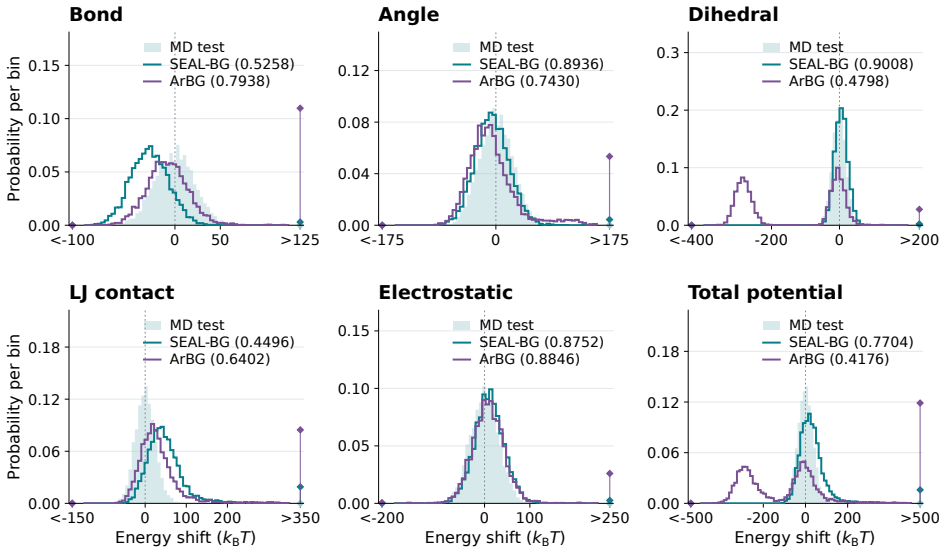


Figure 6: Homeodomain raw component distributions for SEAL-BG, ArBG, and 5,000 fixed MD test frames.

D ALCHEMICAL FREE-ENERGY CASE STUDY

Endpoint Hamiltonians and MD. The TC10b topology, charged termini, 300 K temperature, and CHARMM36/OBC2 parameters follow the setup described in Appendix B.1. For the windows $\lambda = 0, 0.1, \ldots, 1$, the system builder scales every particle charge in OpenMM's `CustomGBForce` by λ, while leaving the charges in the direct nonbonded force unchanged. The Born radii depend on the molecular geometry and fixed radius parameters, whereas the GB self and pair energies depend on squared charges and charge products, respectively. Equation 18 therefore holds exactly at fixed coordinates.

The charge-independent nonpolar surface-area term is retained. Consequently, $\lambda = 0$ is not a pure vacuum endpoint, while $\lambda = 1$ recovers the complete CHARMM36/OBC2 potential. All windows use the same atomic masses, molecular topology, and MD protocol described in Appendix B.1, and each contributes 250,000 saved frames from a 10 μs production trajectory.

Model estimates. Table 7 reports the mean reduced free-energy estimates obtained from eight sampling seeds for SEAL-BG and ArBG. Each seed uses 3,000 generated configurations per endpoint. We additionally report the direct endpoint BAR estimate and use the estimate obtained from the full set of windows as the reference.

Table 7: Trp-cage reduced free-energy estimates relative to the reference. Estimates, deviations, and SDs are dimensionless. Model estimates are reported as the mean and SD across eight sampling seeds.

Method	$\Delta\hat{f}$	Abs. deviation	SD
SEAL-BG	-213.376	8.762	0.130
ArBG	-236.453	14.315	0.495
MD endpoint BAR	-255.458	33.320	–

Window comparison. Table 8 compares the reduced free-energy estimates obtained from the evaluated window subsets.

Table 8: Trp-cage reduced free-energy estimates from selected window subsets. The second column lists 10λ to compactly specify the window locations. Estimates and absolute deviations from the reference are dimensionless.

Windows	10λ	$\Delta\hat{f}$	Abs. deviation
2	$0, 10$	-255.458	33.320
3	$0, 5, 10$	-239.934	17.796
4	$0, 3, 7, 10$	-246.735	24.597
5	$0, 3, 5, 8, 10$	-214.619	7.519
6	$0, 2, 4, 6, 8, 10$	-211.178	10.960
7	$0, 2, 3, 5, 7, 8, 10$	-224.182	2.044
8	$0, 1, 3, 4, 6, 7, 9, 10$	-228.763	6.625
9	$0, 1, 3, 4, 5, 6, 8, 9, 10$	-212.772	9.366
10	$\{0, \ldots, 10\} \setminus \{5\}$	-222.560	0.422

E GENERATED MOLECULAR CONFORMATIONS

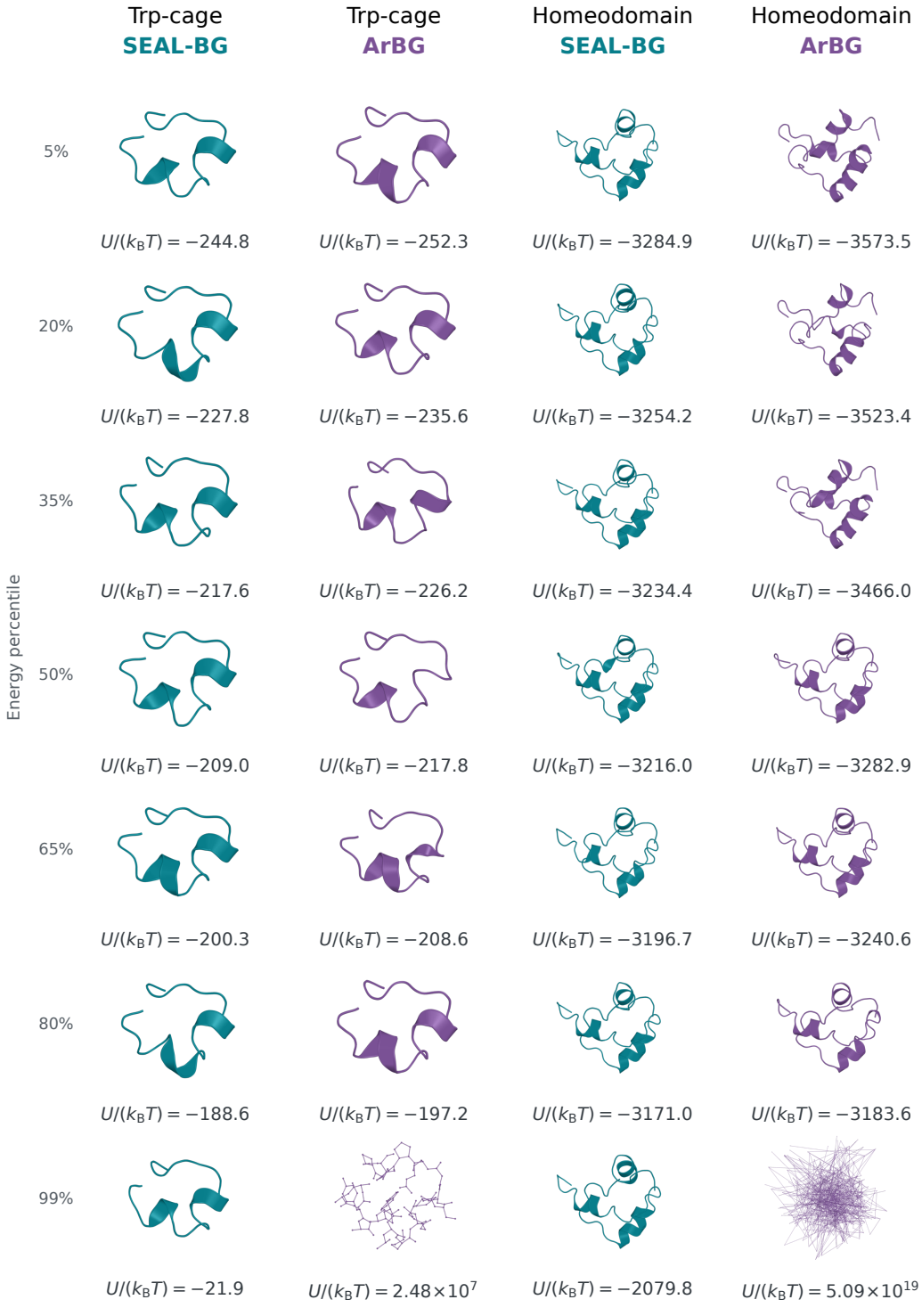


Figure 7: Generated samples of Trp-cage and Homeodomain from SEAL-BG and ArBG.

F Ablation Study

We compare SEAL-BG with a variant that removes log-dispersion regularization (LDR), keeping the model architecture, training hyperparameters, and evaluation settings otherwise unchanged.

Table 9: Ablation results on AL3, AL4, AL6, and Chignolin. E-W_2 is reported in $k_B T$ and T-W_2 in radians; lower is better for all metrics.

System	Model	Raw		Clip+SNIS	
		E-W_2	T-W_2	E-W_2	T-W_2
AL3	SEAL-BG	156.457408	0.327763	0.176652	0.315733
	SEAL-BG without LDR	160.182727	0.354600	0.253245	0.339317
AL4	SEAL-BG	153.265103	0.563638	0.381922	0.583175
	SEAL-BG without LDR	153.133782	0.577567	0.508478	0.566827
AL6	SEAL-BG	148.609026	1.049300	0.298793	1.145833
	SEAL-BG without LDR	164.151214	1.052725	0.816030	1.241256
Chignolin	SEAL-BG	477.620916	2.758284	1.738461	2.778315
	SEAL-BG without LDR	509.928591	2.742922	3.135599	2.884350

Removing LDR increases clipped-SNIS E-W_2 on all four systems, suggesting that LDR improves the consistency between the learned proposal and the Boltzmann target under importance reweighting. Its effect on T-W_2 is modest and generally favorable, with improvements in three of the four systems, indicating that LDR does not substantially compromise torsional agreement.