# MAKE DIFFUSION MODELS GREAT AGAIN IN STRUCTURE-BASED DRUG DESIGN

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Recent Bayesian and flow-based generators have challenged diffusion's leading position in structure-based drug design (SBDD). This progress motivates us to re-examine the modeling choices within molecular diffusion, particularly its forward process and intermediate atom-type representations. We introduce **MEGA**, a unified molecular diffusion framework that redesigns both for joint generation of coordinates and atom types. MEGA directly specifies Gaussian coordinate marginals and categorical parameter trajectories on the simplex, enabling explicit control over both forward noising paths. It separates atom-type parameter evolution from network observations, using stochastic Concrete samples to provide graded classwise evidence beyond hard one-hot observations. To further support learning under high noise and the sampling cold start, we augment MEGA with **GeoRank**, a geometric prior that provides persistent geometric slot identities. On CrossDocked2020, MEGA alone leads the compared methods in 12 of 15 reported benchmark statistics, including all Vina evaluations, PoseBusters validity, and strain energy. It achieves 89.5% PoseBusters validity and a mean Vina score of $-6.96$ kcal/mol. GeoRank's geometric prior improves MEGA's ability to model target distributions of bond lengths, bond angles, functional groups, and ring structures, reducing their Jensen–Shannon distances (JSDs) by 20.2%, 15.1%, 40.4%, and 24.7%, respectively, relative to MEGA alone. These results show that diffusion models remain among the strongest approaches to SBDD.

## 1 INTRODUCTION

Structure-based drug design (SBDD) requires generating ligands with favorable binding interactions and physically plausible structures (Anderson, 2003). A central challenge is to jointly model continuous atomic coordinates and discrete atom types, whose coupling determines molecular geometry and chemical organization.

Generative approaches to SBDD have progressed from autoregressive construction, exemplified by Pocket2Mol (Peng et al., 2022), to equivariant diffusion models that jointly refine ligand coordinates and atom types (Guan et al., 2023a; Schneuing et al., 2024). Subsequent methods introduced decomposed priors and interaction-aware conditioning (Guan et al., 2023b; Huang et al., 2024). More recently, BFN-based MolCRAFT (Qu et al., 2024) and flow-based MolPIF (Jin et al., 2026) have demonstrated strong performance. These advances motivate us to revisit two design choices within molecular diffusion: its forward process and intermediate atom-type representations.

The first choice concerns how coordinate and atom-type information evolves during corruption. Gaussian coordinate perturbations and categorical atom-type transitions determine the evidence available for joint reconstruction at each noise level. Building on marginal-based diffusion formulations (Kingma et al., 2021; Lu et al., 2022), we directly specify Gaussian coordinate marginals together with a categorical parameter trajectory governed by a shared signal schedule. This formulation makes the evolution of coordinate marginals and atom-type parameters explicit, providing a common framework for forward-process design.

The second choice concerns the atom-type observations supplied to the denoiser. A sampled one-hot observation represents a single atom type without expressing graded support for alternative identities. We therefore separate the categorical parameters governing corruption from the observations presented

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.

to the network. Stochastic Concrete observations (Maddison et al., 2017; Jang et al., 2017) retain graded classwise evidence, providing richer chemical information for joint coordinate–type denoising.

These choices form **MEGA**, a molecular diffusion framework combining directly prescribed coordinate marginals and continuous atom-type observations within a shared, pocket-conditioned denoiser. Our ablations show that both the corruption schedules and observation representations contribute to physical plausibility and geometric fidelity.

High-noise denoising presents a complementary challenge. As illustrated in Figure 1(a), heavily corrupted atoms provide ambiguous evidence about their neighbors and spatial roles. We therefore propose **GeoRank**, an optional module that supplies persistent geometric organization. GeoRank maps clean training-ligand geometry to ordered atom slots through virtual-anchor or spectral ordering (Figure 1(b–d)). These geometry-informed slots retain their identities throughout corruption, helping the denoiser learn structural roles when local geometric cues become unreliable. During generation, ranks are randomly assigned to the initial noise slots and remain fixed throughout denoising.

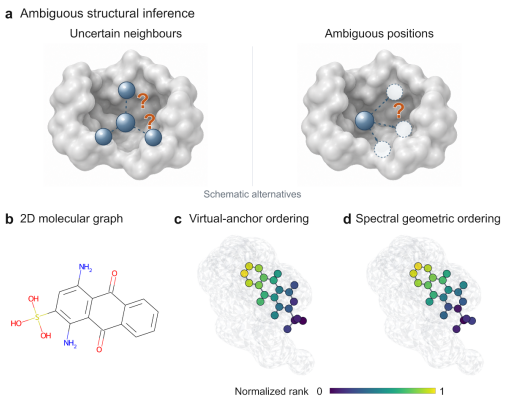



Figure 1: **Motivation for GeoRank. (a)** At high noise, atoms have uncertain ligand neighbors and pocket positions. **(b)** Reference molecular graph. **(c,d)** Virtual-anchor and spectral ordering assign normalized geometric ranks, indicated by color.

On CrossDocked2020, MEGA ranks first in 12 of 15 reported benchmark statistics among the compared methods, including all Vina statistics, PB-valid rate, strain-energy quartiles, and synthetic accessibility score. These results address key priorities in drug design: estimated binding affinity, physical plausibility, and synthetic accessibility. Adding GeoRank improves distributional fidelity, reducing bond-length, bond-angle, functional-group, and ring JSDs by $20.2\%$, $15.1\%$, $40.4\%$, and $24.7\%$, respectively, relative to MEGA alone. These findings show that diffusion models remain among the strongest approaches to SBDD.

Our contributions are summarized as follows:

- **A unified molecular diffusion framework.** We propose MEGA, explicitly defining the forward evolution of coordinates and atom-type parameters while separating observation representations from distribution parameters for joint geometric and chemical modeling.
- **Geometric priors for high-noise modeling.** We introduce GeoRank to address weakened structural cues under high noise through persistent geometric conditioning. Variational analysis and experiments support its benefits for geometric distribution modeling.
- **Strong performance on CrossDocked2020.** MEGA leads the compared methods in 12 of 15 reported benchmark statistics, while GeoRank further improves bond-length, bond-angle, functional-group, and ring distributions.

## 2 Related Work

**Density-based and autoregressive generation.** LiGAN generated atomic densities followed by atom fitting and bond inference (Ragoza et al., 2022), while spatial atom-occurrence densities supported autoregressive sampling (Luo et al., 2021). Pocket2Mol modeled equivariant spatial and bonding relations, and GraphBP generated atoms sequentially in local spherical coordinates (Peng et al., 2022; Liu et al., 2022). These approaches construct ligands through density decoding or sequential growth, whereas MEGA jointly refines whole-ligand coordinates and atom types.

**Equivariant diffusion for SBDD.** TargetDiff jointly denoised coordinates and categorical atom types (Guan et al., 2023a). DecompDiff added scaffold–arm priors, bond diffusion, and validity guidance (Guan et al., 2023b), while DiffSBDD extended diffusion to inpainting and property

optimization (Schneuing et al., 2024). Building on marginal and SNR-based formulations (Song et al., 2020a; Kingma et al., 2021), MEGA focuses on the design of joint corruption trajectories and intermediate atom-type observations. GeoRank further addresses the scarcity of structural cues under high noise.

**Bayesian and flow-based generation.** MolCRAFT modeled continuous distribution parameters, and MolPilot optimized modality-specific schedules (Qu et al., 2024; Qiu et al., 2025). MolPIF interpolated Gaussian and Dirichlet parameters; PAFlow introduced interaction-guided flow matching; FLOWR combined continuous and categorical flows (Jin et al., 2026; Zhou et al., 2025; Cremer et al., 2026). These methods motivate attention to probability paths and multimodal representations. Within a diffusion framework, MEGA separates prescribed categorical parameter trajectories from stochastic Concrete observations (Maddison et al., 2017; Jang et al., 2017), coupling them with a posterior-consistent objective.

# 3 METHOD

We first present MEGA, a unified molecular diffusion framework that specifies coordinate marginals, categorical parameter trajectories, continuous atom-type observations, and their shared denoiser. We then introduce GeoRank, a geometric prior that provides persistent slot roles at high noise and can be used with the unified framework. Section 3.1 defines the pocket-conditioned generation task, Section 3.2 presents MEGA's coordinate and atom-type formulation, and Section 3.3 describes GeoRank during training and generation.

## 3.1 PROBLEM FORMULATION

We formulate SBDD as conditional generative modeling: given a protein pocket, the goal is to generate ligand molecules with compatible three-dimensional binding conformations and favorable interactions. The target (protein pocket) and ligand molecule can be represented as $\mathcal{P} = \{(\boldsymbol{x}_i^{\mathcal{P}}, \boldsymbol{v}_i^{\mathcal{P}})\}_{i=1}^{N_{\mathcal{P}}}$ and $\mathcal{M} = \{(\boldsymbol{x}_i^{\mathcal{M}}, \boldsymbol{v}_i^{\mathcal{M}})\}_{i=1}^{N_{\mathcal{M}}}$, respectively. Here, $N_{\mathcal{P}}$ (resp. $N_{\mathcal{M}}$) denotes the number of atoms in the protein pocket $\mathcal{P}$ (resp. the ligand molecule $\mathcal{M}$). $\boldsymbol{x}_i^{\mathcal{P}} \in \mathbb{R}^3$ represents the 3D coordinates of the protein atom, and $\boldsymbol{v}_i^{\mathcal{P}} \in \mathbb{R}^{N_f}$ represents protein atom features such as element types and amino acid types, with $N_f$ representing the number of such features. Here, $\boldsymbol{x}_i^{\mathcal{M}} \in \mathbb{R}^3$ and $\boldsymbol{v}_i^{\mathcal{M}} \in \mathbb{R}^K$ represent the atom coordinates and atom types of a ligand molecule, respectively. The variable $K$ indicates the number of features used to characterize these atom types. At inference, $N_{\mathcal{M}}$ can be sampled from an empirical pocket-conditioned distribution or predicted by a neural network. Chemical bonds are reconstructed in post-processing.

## 3.2 FORWARD CORRUPTION OF COORDINATES AND ATOM TYPES

We specify the noisy inputs to the denoiser separately for geometry and atom identity. Coordinates follow prescribed Gaussian marginals with a compatible linear SDE (Song et al., 2020b), while atom types follow a categorical parameter trajectory observed through Concrete sampling. This separation makes both the corruption parameters and the representations presented to the network explicit.

### 3.2.1 COORDINATE MARGINALS AND COMPATIBLE SDE

For simplicity, we describe the process for one ligand atom and omit the superscript $\mathcal{M}$ and atom index $i$. Following DPM-Solver (Lu et al., 2022), we define

$$q_{0t}(\boldsymbol{x}_t \mid \boldsymbol{x}_0) = \mathcal{N}\big(\boldsymbol{x}_t; \alpha(t)\boldsymbol{x}_0, \sigma^2(t)\mathbf{I}_3\big)\,, \tag{1}$$

where $\boldsymbol{x}_0$ denotes the clean coordinates, and $\alpha(t)$ and $\sigma(t)$ specify the cumulative signal amplitude and noise standard deviation. Equivalently, a noisy state can be sampled directly as $\boldsymbol{x}_t = \alpha(t)\boldsymbol{x}_0 + \sigma(t)\boldsymbol{\epsilon}$, with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3)$.

Conventional diffusion formulations typically start from local forward transitions, represented in continuous time by the following linear SDE:

$$\mathrm{d}\boldsymbol{x}_t = f(t)\boldsymbol{x}_t\,\mathrm{d}t + g(t)\,\mathrm{d}\mathbf{w}_t. \tag{2}$$

The marginal corruption is then determined by these local dynamics. We instead specify the marginal schedules $\alpha(t)$ and $\sigma(t)$ directly and obtain compatible SDE coefficients through

$$f(t) = \frac{\mathrm{d}\log\alpha(t)}{\mathrm{d}t}, \qquad g^2(t) = \frac{\mathrm{d}\sigma^2(t)}{\mathrm{d}t} - 2f(t)\sigma^2(t). \tag{3}$$

This marginal-first parameterization makes the signal and noise levels encountered during training explicit, while retaining a corresponding continuous-time forward process.

For $s < t$, its exact single-step transition is

$$q_{st}(\boldsymbol{x}_t \mid \boldsymbol{x}_s) = \mathcal{N}\Big(\boldsymbol{x}_t; \alpha_{t|s}\boldsymbol{x}_s, \sigma_{t|s}^2\mathbf{I}_3\Big), \tag{4}$$

where the single-step coefficients are determined by the cumulative schedules:

$$\alpha_{t|s} = \frac{\alpha(t)}{\alpha(s)}, \qquad \sigma_{t|s}^2 = \sigma^2(t) - \frac{\alpha^2(t)}{\alpha^2(s)}\sigma^2(s). \tag{5}$$

The nonincreasing SNR ensures $\sigma_{t|s}^2 \geq 0$. For an infinitesimal step $\Delta t$, these coefficients satisfy $\alpha_{t+\Delta t|t} = 1 + f(t)\Delta t + o(\Delta t)$ and $\sigma_{t+\Delta t|t}^2 = g^2(t)\Delta t + o(\Delta t)$, establishing the correspondence between the Gaussian single-step kernel and the forward SDE.

### 3.2.2 Atom-Type Parameters and Continuous Observations

To preserve graded atom-type information, MEGA separates categorical parameter evolution from sampled observations, drawing on the distribution-parameter perspective of Bayesian Flow Networks (Graves et al., 2023; Qu et al., 2024). For a clean one-hot atom type $\boldsymbol{v}_0 \in \{\mathbf{e}_1, \ldots, \mathbf{e}_K\}$, we define its clean categorical parameter as $\boldsymbol{\theta}_0 = \boldsymbol{v}_0$. The parameter trajectory is given by

$$\boldsymbol{\theta}_t = \alpha(t)\boldsymbol{\theta}_0 + \big(1 - \alpha(t)\big)\mathbf{u}, \qquad \mathbf{u} = \frac{\mathbf{1}_K}{K}, \tag{6}$$

where $\alpha(t) \in [0, 1]$ follows the same nonincreasing signal schedule as in the coordinate corruption process (Section 3.2.1). Thus, $\boldsymbol{\theta}_t$ is a convex interpolation between the clean one-hot label and the uniform prior: decreasing $\alpha(t)$ progressively removes clean-type information, and $\alpha(1) = 0$ yields $\boldsymbol{\theta}_1 = \mathbf{u}$.

To obtain the representation used by the denoising network, we sample a continuous observation $\mathbf{y}_t \in \Delta^{K-1}$:

$$\mathbf{y}_t \mid \boldsymbol{\theta}_t \sim \mathrm{Concrete}(\boldsymbol{\theta}_t, \tau(t)), \qquad \mathbf{y}_t = \mathrm{softmax}\left(\frac{\log\boldsymbol{\theta}_t + \mathbf{g}}{\tau(t)}\right), \tag{7}$$

where $g_k \overset{\text{iid}}{\sim} \mathrm{Gumbel}(0, 1)$, $\tau(t) > 0$ controls the softness of the observation, and the logarithm and softmax operate over the $K$ atom types.

Here, $\boldsymbol{\theta}_t$ specifies the categorical corruption level, whereas $\mathbf{y}_t$ provides a stochastic representation of that parameter. The denoising network receives $\mathbf{y}_t$ together with the noisy coordinates, protein pocket, and time, and predicts the clean categorical parameter $\widehat{\boldsymbol{\theta}}_0 \in \Delta^{K-1}$.

**Training objective.** The shared denoiser predicts clean coordinates $\widehat{\mathbf{X}}_0$ and atom-type probabilities $\widehat{\boldsymbol{\theta}}_0$ from $(\mathbf{X}_t, \mathbf{Y}_t, t, \mathcal{P})$, where $\mathbf{Y}_t$ collects the continuous observations. With $N = N_{\mathcal{M}}$ and $t^- = \max(t - 1/T, 0)$ for $T = 1000$, we optimize

$$\mathcal{L}_x = \mathbb{E}\left[\frac{w_x(t)}{N}\sum_{i=1}^{N}\|\widehat{\boldsymbol{x}}_{0,i} - \boldsymbol{x}_{0,i}\|_2^2\right], \quad \mathcal{L}_v = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} D_{\mathrm{KL}}\Big(\boldsymbol{\theta}_{t^-,i} \,\Big\|\, \widehat{\boldsymbol{\theta}}_{t^-,i}\Big)\right], \tag{8}$$

$$\mathcal{L} = \mathcal{L}_x + \lambda_v\mathcal{L}_v, \qquad \lambda_v = 2. \tag{9}$$

The expectation is over ligand–pocket pairs, $t \sim \mathcal{U}(0, 1)$, and corruption noise. Here $\boldsymbol{\theta}_{t^-,i}$ and $\widehat{\boldsymbol{\theta}}_{t^-,i}$ are obtained by applying Equation (6) to the clean type $\boldsymbol{v}_{0,i}$ and its prediction $\widehat{\boldsymbol{\theta}}_{0,i}$, respectively; $w_x(t)$ is an SNR-based coordinate weight. MEGA+GeoRank uses the same objective without an additional rank-supervision loss. Weighting details, the minibatch formulation, and derivations are given in Appendix A.

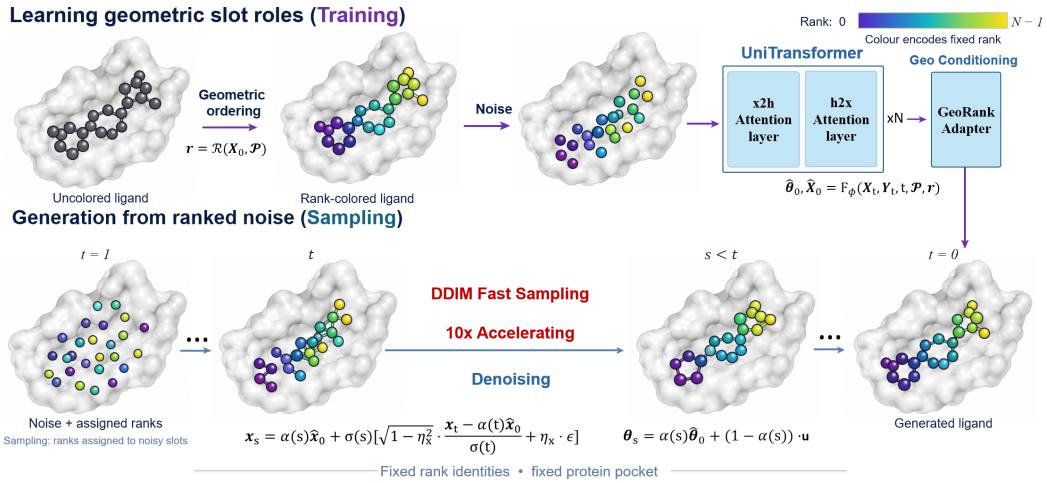



Figure 2: **Overview of MEGA with the optional GeoRank module.** During training (top), geometric ordering assigns fixed ranks to ligand atoms. GeoRank adds feature and coordinate residuals after complete UniTransformer layers. During sampling (bottom), randomly ranked noisy slots are refined through DDIM-style (Song et al., 2020a) coordinate updates and atom-type parameter interpolation, with fixed rank identities and protein pocket. Sampling performance is reported in Appendix G.1.

### 3.3 Geometric Slot Prior

At high noise levels, the corrupted coordinates $\mathbf{X}_t$ and atom-type observations $\mathbf{Y}_t$ provide limited information about the clean ligand. The denoiser must therefore infer both the geometric roles of individual atoms and their collective spatial organization from weakly informative inputs. This ambiguity makes high-noise denoising difficult and creates a corresponding cold-start problem during generation, when ligand slots are initialized from exchangeable noise.

Our key insight is that *a consistent geometric ordering can provide structural roles for noisy ligand slots without requiring reference coordinates at inference*. We introduce GeoRank, a geometric slot prior that assigns each training atom a rank derived from the clean conformation and retains this identity throughout corruption. Figure 2 illustrates how GeoRank connects training on geometrically ordered ligand atoms with generation from ranked noise. We first define the geometric slot identities and two ordering rules in Section 3.3.1, then explain how the resulting ranks are used during training and generation in Section 3.3.2.

#### 3.3.1 Constructing Geometry-Derived Slot Identities

Let $N = N_{\mathcal{M}}$ and let $\mathbf{X}_0 \in \mathbb{R}^{N\times 3}$ denote the clean ligand coordinates centered at the protein-pocket centroid. We assign each atom a unique rank through a geometric ordering rule $\mathcal{R}$:

$$\mathbf{r} = \mathcal{R}(\mathbf{X}_0, \mathcal{P}), \qquad \{r_i\}_{i=1}^{N} = \{0, \dots, N-1\}, \qquad q_i = \frac{r_i}{\max(N-1, 1)}. \tag{10}$$

The rank provides a discrete slot identity, while $q_i$ describes its relative position in the ordering. For nondegenerate configurations with unambiguous ordering, the geometric rules below are invariant to joint rigid transformations of the complex and equivariant to ligand reindexing:

$$\mathcal{R}(\Pi\mathbf{X}_0, \mathcal{P}) = \Pi\,\mathcal{R}(\mathbf{X}_0, \mathcal{P}), \tag{11}$$

where $\Pi$ is a permutation matrix. Thus, slot identities follow the atoms under reindexing rather than depending on their original indices.

We consider two complementary ways to extract this ordering from three-dimensional geometry.

**Virtual-anchor ordering.** To capture global organization relative to the pocket, we select the ligand atom farthest from the pocket centroid and place a virtual anchor beyond it:

$$a = \arg\max_i \|\boldsymbol{x}_{0,i}\|_2, \qquad \mathbf{z}_{\text{anc}} = \kappa \boldsymbol{x}_{0,a}, \qquad r_i = \text{rank}_{\uparrow}\left(\|\boldsymbol{x}_{0,i} - \mathbf{z}_{\text{anc}}\|_2\right). \tag{12}$$

Here, $\kappa > 1$, $\boldsymbol{x}_{0,i}$ is the coordinate of atom $i$, and $\mathrm{rank}_{\uparrow}$ assigns zero-based ranks in ascending order. For distinct sorting distances, the selected atom $a$ receives rank zero. The remaining ranks describe an ordering by distance from the virtual anchor, providing a pocket-relative geometric organization without retaining the anchor coordinates as network inputs. This ordering captures a global geometric relation; it does not require adjacent ranks to be adjacent in the chemical bond graph.

**Spectral geometric ordering.** As an alternative emphasizing local geometric connectivity, we construct a symmetric $k$-nearest-neighbor graph $\mathcal{E}_{\mathrm{geo}}$ from the clean ligand coordinates:

$$w_{ij} = \mathbf{1}[(i,j) \in \mathcal{E}_{\mathrm{geo}}] \exp\left(-\frac{\|\boldsymbol{x}_{0,i} - \boldsymbol{x}_{0,j}\|_2^2}{2\ell^2}\right), \qquad \mathbf{D}_{ii} = \sum_j w_{ij},$$
$$\mathbf{L} = \mathbf{I}_N - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}, \qquad \mathbf{L}\mathbf{f} = \lambda_2 \mathbf{f}, \qquad r_i = \mathrm{rank}_{\uparrow}(f_i), \tag{13}$$

where $\mathbf{W} = (w_{ij})$, $\mathbf{D}$ is the degree matrix, $\ell > 0$ is the distance bandwidth, and $\mathbf{f}$ is an eigenvector associated with the second-smallest eigenvalue. For a connected graph with a simple second eigenvalue, we resolve the sign ambiguity by choosing the orientation that assigns the farthest atom $a$ the smaller rank among the two reversed orderings.

This construction extracts a one-dimensional ordering from local three-dimensional connectivity. Together, the two variants provide complementary geometric biases: virtual-anchor ordering emphasizes pocket-relative global organization, whereas spectral ordering emphasizes neighborhood structure. Both orderings encode geometric slot roles; chemical bonds are reconstructed after generation.

#### 3.3.2 Rank Conditioning in Training and Generation

During training, we compute $\mathbf{r} = \mathcal{R}(\mathbf{X}_0, \mathcal{P})$ once from the clean ligand and keep it fixed throughout corruption. The denoiser therefore learns a rank-conditioned prediction:

$$\widehat{\mathbf{X}}_0, \widehat{\boldsymbol{\theta}}_0 = F_\phi\left(\mathbf{X}_t, \mathbf{Y}_t, t, \mathcal{P}; \mathbf{r}\right). \tag{14}$$

Here $\phi$ denotes the network parameters. Independent GeoRank adapters follow complete UniTransformer layers 3 and 6, adding ligand-only feature and coordinate residuals (Appendix G.2). GeoRank can improve the optimal variational fit of the reverse process by preserving informative geometric conditions; Appendix E gives the theory and supporting validation-loss evidence (Figure 5).

At inference, we sample a uniformly random permutation of $\{0, \ldots, N-1\}$ independently of the initial ligand noise, assign it to the ligand slots, and retain it throughout denoising. This initialization matches the permutation-symmetrized training distribution at the zero-signal endpoint, which factorizes into independent ligand noise and a uniform rank permutation.

No reference ligand conformation is required at inference. Instead, the geometric roles learned during training provide a structural prior for organizing initially noisy slots into a coherent conformation. GeoRank thereby connects geometry-informed training with structured early-stage generation, supporting high-noise learning and addressing the sampling cold start when coordinate and atom-type evidence is weak.

## 4 Experiments

### 4.1 Experimental Setup

**Dataset and generation protocol.** Following previous works (Qu et al., 2024; Guan et al., 2023a), we adopted filtered CrossDocked2020 (Francoeur et al., 2020): pose RMSD below 1 Å, a 30% sequence-identity split, approximately 100,000 training pairs, and 100 test pockets. For evaluation, we sampled 100 molecules for each test protein.

**Baselines.** We compared MEGA with autoregressive methods AR (Luo et al., 2021) and Pocket2Mol (Peng et al., 2022), diffusion-based methods TargetDiff (Guan et al., 2023a) and DecompDiff (Guan et al., 2023b), the Bayesian flow network MolCRAFT (Qu et al., 2024), and the flow-based model MolPIF (Jin et al., 2026). These baselines cover different approaches to jointly generating ligand coordinates and atom types.

Table 1: CrossDock benchmark results. MEGA+GeoRank uses virtual-anchor ordering. Arrows indicate the preferred direction. Best and second-best distinct generated results are bolded and underlined. Clash mean counts protein–ligand clashes; clash ratio is the fraction of ligand atoms involved in a clash.

| Methods | PB-Valid | Vina Score (↓) | | Vina Min (↓) | | Vina Dock (↓) | | Strain Energy (↓) | | | Connected | QED | SA | Clashes (↓) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Avg. (↑) | Avg. | Med. | Avg. | Med. | Avg. | Med. | 25% | 50% | 75% | Avg. (↑) | Avg. (↑) | Avg. (↑) | Mean | Clash ratio |
| CrossDock | 95.0% | -6.36 | -6.46 | -6.71 | -6.49 | -7.45 | -7.26 | 34 | 107 | 196 | – | 0.48 | 0.73 | 7.44 | 0.12 |
| AR | 59.0% | -5.75 | -5.64 | -6.18 | -5.88 | -6.75 | -6.62 | 259 | 595 | 2286 | 93.5% | 0.51 | 0.63 | **5.88** | 0.11 |
| Pocket2Mol | 72.3% | -5.14 | -4.70 | -6.42 | -5.82 | -7.15 | -6.79 | 102 | 189 | 374 | 96.3% | 0.57 | 0.76 | 7.72 | 0.16 |
| TargetDiff | 50.5% | -5.47 | -6.30 | -6.64 | -6.83 | -7.80 | -7.91 | 369 | 1243 | 13871 | 90.4% | 0.48 | 0.58 | 13.21 | 0.16 |
| DecompDiff | 71.7% | -5.19 | -5.27 | -6.03 | -6.00 | -7.03 | -7.16 | 115 | 421 | 1424 | 82.9% | 0.51 | 0.66 | 10.05 | 0.15 |
| MolCRAFT | 84.6% | -6.55 | -6.95 | -7.21 | -7.14 | -7.67 | -7.82 | 83 | 195 | 510 | 96.7% | 0.50 | 0.67 | 9.02 | 0.13 |
| MolPIF | 83.0% | -6.64 | -7.02 | -7.41 | -7.28 | -8.09 | -8.13 | 65 | 150 | 375 | **97.5%** | **0.59** | 0.72 | 8.75 | 0.12 |
| MEGA | **89.5%** | **-6.96** | **-7.28** | **-7.54** | **-7.54** | **-8.21** | **-8.35** | **28** | **85** | **203** | 96.7% | 0.55 | **0.82** | 7.31 | **0.10** |
| MEGA + GeoRank | 88.5% | -6.57 | -6.96 | -7.18 | -7.15 | -7.78 | -7.86 | 63 | 155 | 363 | 97.1% | 0.57 | 0.72 | 8.06 | 0.11 |

**Evaluation metrics.** Binding affinity and conformational quality were evaluated. AutoDock Vina (Eberhardt et al., 2021) provides **Vina Score** for direct evaluation of generated poses, **Vina Min** after local minimization, and **Vina Dock** after redocking. We report means and medians in kcal/mol, with lower scores indicating more favorable estimated binding. Conformational quality was measured using PoseBusters validity (**PB-valid**), **strain-energy quartiles**, and protein–ligand clashes, reported as the **mean clash count** and the **clash ratio** (Gu et al., 2021; Buttenschoen et al., 2024; Harris et al., 2023), capturing physical plausibility, intramolecular stability, and steric compatibility with the pocket. Jensen–Shannon distances (**JSD**) between generated and reference bond-length, bond-angle, and dihedral distributions further assess local geometric fidelity. Conformational checks used raw poses before relaxation. **Connectedness**, **QED** (Bickerton et al., 2012), and normalized synthetic accessibility (**SA**; higher is better) provide complementary molecular-quality measures.

### 4.2 MEGA Leads the Main Generation Benchmark

Table 1 shows that the proposed MEGA achieves strong binding scores together with chemical and physical plausibility. **MEGA ranks first in 12 of the 15 reported metric columns among the compared generative methods.** These include all Vina statistics and strain-energy quartiles, PB-valid, SA, and clash ratio, demonstrating MEGA's ability to model molecular structures consistent with physical and chemical constraints.

**Leading binding performance across all Vina evaluations.** MEGA achieves mean Vina Score, Min, and Dock values of $-6.96$, $-7.54$, and $-8.21$ kcal/mol, respectively, surpassing MolPIF, the strongest reported baseline on these metrics, by $0.32$, $0.13$, and $0.12$ kcal/mol. The advantage also holds across all three median statistics. The direct-score improvement shows that favorable estimated pocket interactions are present before external structural optimization, with the advantage retained after minimization and redocking.

**Strong binding scores with superior physical plausibility.** MEGA simultaneously achieves the highest PB-valid rate of $89.5\%$, exceeding MolCRAFT and MolPIF by $4.9$ and $6.5$ percentage points, respectively. Its strain-energy quartiles of $28/85/203$ are all the lowest among the compared generative methods, with a median $43.3\%$ below MolPIF's. MEGA further achieves the lowest clash ratio of $0.10$ and the second-lowest clash mean of $7.31$, behind AR's $5.88$, while retaining the highest normalized SA score of $0.82$. Relative to the representative diffusion baselines, MEGA raises PB-valid from TargetDiff's $50.5\%$ and DecompDiff's $71.7\%$ to $89.5\%$, while reducing their strain-energy medians of $1243$ and $421$ to $85$. Together, these results support redesigning molecular diffusion to improve generation quality; Appendix H examines the forward-process and atom-type representation choices.

Table 2: Geometric distribution fidelity and ring frequencies. MEGA+GeoRank uses virtual-anchor ordering. Bond/angle/torsion use weighted JSD; FG/Ring denote functional-group/ring JSD (lower is better). Bold/underline mark best/second-best values. Frequencies are percentages. Fused rings share at least two atoms within a molecule's smallest set of smallest rings (SSSR).

| Method | Generated | Distribution JSD ↓ | | | | | Fused | Ring size (%) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | molecules | Bond | Angle | Torsion | FG | Ring | (%) | 3 | 4 | 5 | 6 |
| Reference | – | – | – | – | – | – | 48.00 | 4.0 | 0.0 | 49.0 | 84.0 |
| AR | 9,295 | 0.5427 | 0.4868 | 0.5390 | 0.2763 | 0.3252 | 55.91 | 50.8 | 0.8 | 35.8 | 71.9 |
| Pocket2Mol | 9,831 | 0.4673 | 0.4585 | 0.4498 | 0.2191 | 0.1286 | 64.56 | 0.3 | 0.1 | 38.0 | 88.6 |
| TargetDiff | 9,036 | 0.3514 | 0.3874 | 0.3836 | 0.3030 | 0.2348 | 58.76 | 0.0 | 7.3 | 57.0 | 76.1 |
| DecompDiff | 8,292 | <u>0.3151</u> | 0.3642 | 0.3310 | **0.1985** | 0.2415 | 40.17 | 5.4 | 8.3 | 51.5 | 65.6 |
| MolPIF | 9,959 | 0.3238 | 0.3685 | 0.3410 | 0.2849 | 0.1567 | 50.57 | 0.0 | 1.4 | 40.1 | 89.9 |
| MolCRAFT | 9,667 | 0.3195 | <u>0.3253</u> | **0.2806** | 0.2537 | <u>0.1268</u> | 52.74 | 0.0 | 0.6 | 47.0 | 85.1 |
| MEGA | 9,928 | 0.3384 | 0.3709 | 0.3307 | 0.3397 | 0.1600 | 37.11 | 1.5 | 1.0 | 37.4 | 95.0 |
| MEGA + GeoRank | 9,889 | **0.2701** | **0.3150** | <u>0.3280</u> | <u>0.2025</u> | **0.1205** | 49.21 | 1.8 | 0.7 | 46.3 | 88.3 |

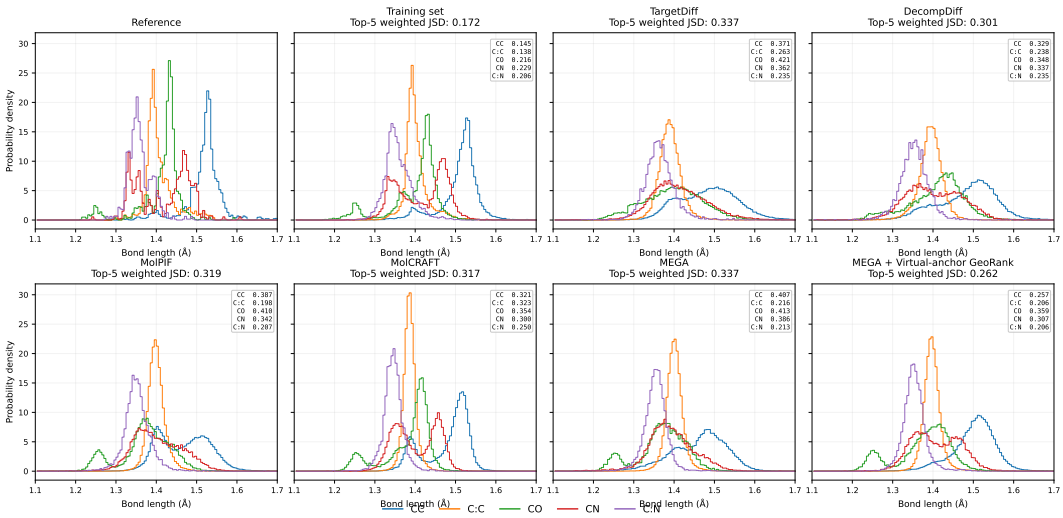



Figure 3: Bond-length densities for the five most frequent categories, comparing reference and training ligands with six generators. Panel titles report top-five weighted JSD against the reference; insets give category-wise JSD (lower is better). MEGA+GeoRank uses virtual-anchor ordering. Colons denote aromatic bonds.

### 4.3 GeoRank Advances Geometric Distribution Modeling

GeoRank supplies persistent geometric slot identities to improve MEGA's modeling of local geometry and structural motifs. These distribution-level gains complement molecule-level validity and docking scores.

**Local geometric fidelity.** Table 2 shows that MEGA+GeoRank reduces bond-length and bond-angle JSD from $0.3384/0.3709$ to $0.2701/0.3150$, improvements of $20.2\%$ and $15.1\%$ over MEGA. **Both scores outperform every compared baseline.** Figure 3 further shows the lowest top-five weighted bond-length JSD among generators ($0.262$, versus $0.337$ for MEGA and $0.301$ for DecompDiff). GeoRank suppresses excess CC density near $1.4$ Å and shifts the dominant peak toward the reference near $1.53$ Å. Across the full category evaluation, it improves all eight bond-length and seven of eight bond-angle categories over MEGA, with the best unweighted category means of $0.295/0.376$ (Tables 7 and 8, Appendix I).

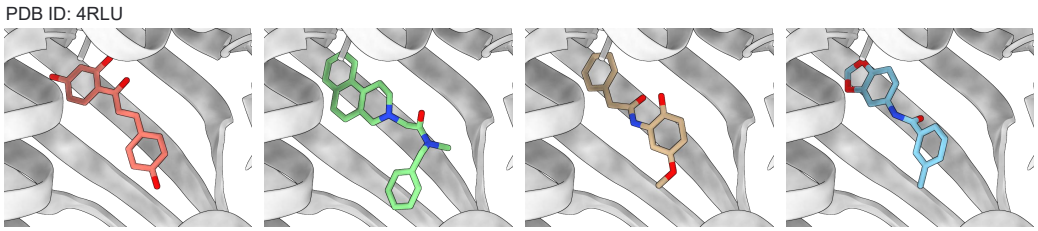



Figure 4: Example ligands in the 4RLU pocket with QED, normalized SA (higher is better), and strain energy (lower is better). Panels show the reference ligand and samples from MolCRAFT, MEGA, and MEGA+GeoRank with virtual-anchor ordering.

**Functional-group and ring distributions.** GeoRank reduces functional-group JSD from $0.3397$ to $0.2025$ ($40.4\%$), approaching DecompDiff's best value of $0.1985$, and achieves the lowest ring JSD of $0.1205$, down from $0.1600$ ($24.7\%$). Five-membered ring frequency rises from $37.4\%$ to $46.3\%$ toward the reference $49.0\%$, while six-membered ring frequency falls from $95.0\%$ to $88.3\%$ toward $84.0\%$. Together with the local geometry results, these changes show improved recovery of the structural preferences of reference ligands.

**Denoising and generation quality.** MEGA+GeoRank lowers validation denoising loss through most of training, mainly in the coordinate term (Appendix E.4). Under virtual-anchor conditioning, the coordinate contribution to the path evidence lower bound (ELBO) improves on all 10 evaluated test molecules, by $0.5767$ per dimension on average (Appendix E.5). Generation retains an $88.5\%$ PB-valid rate, $1.0$ percentage point below MEGA; mean Vina Score, Min, and Dock increase by $0.39/0.36/0.43$ kcal/mol, remaining close to MolCRAFT. Appendix J compares the two GeoRank orderings.

### 4.4 Case Study on the 4RLU Pocket

Figure 4 illustrates molecular quality in the 4RLU pocket. Compared with the displayed MolCRAFT sample, the MEGA sample has higher QED ($0.8276$ versus $0.7175$) and SA ($0.9400$ versus $0.8800$), with lower strain energy ($29.00$ versus $68.45$). The MEGA+GeoRank sample combines the highest QED among the displayed ligands ($0.8970$) with an SA of $0.9300$ and a strain energy of $20.83$, closer to the reference value of $14.22$. These examples illustrate the combination of favorable molecular properties and low conformational strain in pocket-conditioned generation.

## 5 Conclusion

MEGA presents a unified view of molecular diffusion by explicitly defining the forward evolution of continuous coordinates and categorical atom-type parameters while separating these parameters from the observations supplied to the denoiser. This formulation enables direct control over the forward process and preserves graded chemical evidence for joint geometric and chemical modeling. We further identify a common challenge in molecular diffusion: as corruption increases, local structural cues become unreliable, making molecular organization difficult to learn and recover. GeoRank addresses this challenge through persistent geometric priors, with variational analysis and empirical results supporting its effectiveness in modeling molecular geometry. On CrossDocked2020, MEGA leads the compared methods in 12 of 15 reported benchmark statistics, while GeoRank further improves bond-length, bond-angle, functional-group, and ring distributions. Together, these findings demonstrate the potential of unified diffusion design and geometric conditioning to advance molecular generation, reaffirming diffusion models as one of the strongest approaches to structure-based drug design.

## 6 AI USE STATEMENT

During code development, AI tools were used to debugging and refinement of the experimental code. For scientific figures, generative AI tools were used to create or refine selected visual elements. AI tools were also used for language editing to improve clarity and readability. All AI-assisted outputs were reviewed and verified by the authors before being incorporated into this work.

## REFERENCES


Amy C Anderson. The process of structure-based drug design. *Chemistry & biology*, 10(9):787–797, 2003.

Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. In *Advances in Neural Information Processing Systems*, 2021. URL https://arxiv.org/abs/2107.03006.

G Richard Bickerton, Gaia V Paolini, Jérémy Besnard, Sorel Muresan, and Andrew L Hopkins. Quantifying the chemical beauty of drugs. *Nature chemistry*, 4(2):90–98, 2012.

Martin Buttenschoen, Garrett M. Morris, and Charlotte M. Deane. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15:3130–3139, 2024. doi: 10.1039/D3SC04185A. URL https://doi.org/10.1039/D3SC04185A.

Julian Cremer, Ross Irwin, Alessandro Tibo, Jon Paul Janet, Simon Olsson, and Djork-Arné Clevert. FLOWR: flow matching for structure-aware de novo, interaction- and fragment-based ligand generation. *Nature Computational Science*, 6:565–574, 2026. doi: 10.1038/s43588-026-00998-8. URL https://www.nature.com/articles/s43588-026-00998-8.

Jerome Eberhardt, Diogo Santos-Martins, Andreas F. Tillack, and Stefano Forli. AutoDock Vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling*, 61(8):3891–3898, 2021. doi: 10.1021/acs.jcim.1c00203. URL https://doi.org/10.1021/acs.jcim.1c00203.

Paul G. Francoeur, Tomohide Masuda, Jocelyn Sunseri, Andrew Jia, Richard B. Iovanisci, Ian Snyder, and David R. Koes. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *Journal of Chemical Information and Modeling*, 60(9):4200–4215, 2020. doi: 10.1021/acs.jcim.0c00411. URL https://doi.org/10.1021/acs.jcim.0c00411.

Alex Graves, Rupesh Kumar Srivastava, Timothy Atkinson, and Faustino Gomez. Bayesian flow networks. *arXiv preprint arXiv:2308.07037*, 2023. URL https://arxiv.org/abs/2308.07037.

Shuo Gu, Matthew S Smith, Ying Yang, John J Irwin, and Brian K Shoichet. Ligand strain energy in large library docking. *Journal of chemical information and modeling*, 61(9):4331, 2021.

Jiaqi Guan, Wesley Wei Qian, Xingang Peng, Yufeng Su, Jian Peng, and Jianzhu Ma. 3D equivariant diffusion for target-aware molecule generation and affinity prediction. In *International Conference on Learning Representations*, 2023a. URL https://arxiv.org/abs/2303.03543.

Jiaqi Guan, Xiangxin Zhou, Yuwei Yang, Yu Bao, Jian Peng, Jianzhu Ma, Qiang Liu, Liang Wang, and Quanquan Gu. DecompDiff: Diffusion models with decomposed priors for structure-based drug design. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 11827–11846, 2023b. URL https://proceedings.mlr.press/v202/guan23a.html.

Charles Harris, Kieran Didi, Arian R. Jamasb, Chaitanya K. Joshi, Simon V. Mathis, Pietro Lió, and Tom Blundell. Benchmarking generated poses: How rational is structure-based drug design with generative models? *arXiv preprint arXiv:2308.07413*, 2023. URL https://arxiv.org/abs/2308.07413.

Emiel Hoogeboom, Didrik Nielsen, Priyank Jaini, Patrick Forré, and Max Welling. Argmax flows and multinomial diffusion: Learning categorical distributions. In *Advances in Neural Information Processing Systems*, 2021. URL https://arxiv.org/abs/2102.05379.

Zhilin Huang, Ling Yang, Xiangxin Zhou, Zhilong Zhang, Wentao Zhang, Xiawu Zheng, Jie Chen, Yu Wang, Bin Cui, and Wenming Yang. Protein-ligand interaction prior for binding-aware 3d molecule diffusion models. In *International Conference on Learning Representations*, volume 2024, pp. 18453–18474, 2024.

Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with Gumbel-Softmax. In *International Conference on Learning Representations*, 2017. URL https://arxiv.org/abs/1611.01144.

Yaowei Jin, Junjie Wang, Yufan Tang, Wenkai Xiang, Duanhua Cao, Dan Teng, Zhehuan Fan, Jiacheng Xiong, Xia Sheng, Chuanlong Zeng, Duo An, Mingyue Zheng, Shuangjia Zheng, and Qian Shi. MolPIF: a parameter interpolation flow model for molecule generation. *Bioinformatics*, 42(6):btag323, 2026. doi: 10.1093/bioinformatics/btag323. URL https://doi.org/10.1093/bioinformatics/btag323.

Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. In *Advances in Neural Information Processing Systems*, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/b578f2a52a0229873fefc2a4b06377fa-Abstract.html.

Meng Liu, Youzhi Luo, Kanji Uchino, Koji Maruhashi, and Shuiwang Ji. Generating 3D molecules for target protein binding. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pp. 13912–13924, 2022. URL https://proceedings.mlr.press/v162/liu22m.html.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *Advances in Neural Information Processing Systems*, volume 35, pp. 5775–5787, 2022.

Shitong Luo, Jiaqi Guan, Jianzhu Ma, and Jian Peng. A 3D generative model for structure-based drug design. In *Advances in Neural Information Processing Systems*, volume 34, pp. 6229–6239, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/314450613369e0ee72d0da7f6fee773c-Abstract.html.

Chris J. Maddison, Andriy Mnih, and Yee Whye Teh. The Concrete distribution: A continuous relaxation of discrete random variables. In *International Conference on Learning Representations*, 2017. URL https://arxiv.org/abs/1611.00712.

Xingang Peng, Shitong Luo, Jiaqi Guan, Qi Xie, Jian Peng, and Jianzhu Ma. Pocket2Mol: Efficient molecular sampling based on 3D protein pockets. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pp. 17644–17655, 2022. URL https://proceedings.mlr.press/v162/peng22b.html.

Keyue Qiu, Yuxuan Song, Zhehuan Fan, Peidong Liu, Zhe Zhang, Mingyue Zheng, Hao Zhou, and Wei-Ying Ma. Piloting structure-based drug design via modality-specific optimal schedule. In *International Conference on Machine Learning*, 2025. URL https://arxiv.org/abs/2505.07286.

Yanru Qu, Keyue Qiu, Yuxuan Song, Jingjing Gong, Jiawei Han, Mingyue Zheng, Hao Zhou, and Wei-Ying Ma. MolCRAFT: Structure-based drug design in continuous parameter space. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 41749–41768, 2024. URL https://proceedings.mlr.press/v235/qu24a.html.

Matthew Ragoza, Tomohide Masuda, and David Ryan Koes. Generating 3D molecules conditional on receptor binding sites with deep generative models. *Chemical Science*, 2022. doi: 10.1039/D1SC05976A. URL https://pubs.rsc.org/en/content/articlelanding/2022/sc/d1sc05976a.

Arne Schneuing, Charles Harris, Yuanqi Du, Kieran Didi, Arian Jamasb, Ilia Igashov, Weitao Du, Carla Gomes, Tom Blundell, Pietro Lió, Max Welling, Michael Bronstein, and Bruno Correia. Structure-based drug design with equivariant diffusion models. *Nature Computational Science*, 2024. doi: 10.1038/s43588-024-00737-x. URL https://www.nature.com/articles/s43588-024-00737-x.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020a.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020b.

Jingyuan Zhou, Hao Qian, Shikui Tu, and Lei Xu. Prior-guided flow matching for target-aware molecule design with learnable atom number. In *Advances in Neural Information Processing Systems*, 2025. URL https://arxiv.org/abs/2509.01486.

# A TRAINING LOSSES

## A.1 NOTATION AND TRAINING SETUP

We retain the main-text notation: $\boldsymbol{x}_{0,i}$ is the clean coordinate of ligand atom $i$ in the protein-centered model coordinate system, $\boldsymbol{v}_{0,i}$ is its one-hot type, and $\widehat{\boldsymbol{\theta}}_{0,i} = \mathrm{softmax}(\mathbf{z}_{\phi,i})$ is the predicted clean-type distribution. The network receives noisy coordinates and Concrete observations and jointly predicts $(\widehat{\mathbf{X}}_0, \widehat{\boldsymbol{\theta}}_0)$. For each molecule, sample $t \sim \mathcal{U}(0,1)$ and use

$$\alpha(t) = 1 - b^{1-t}, \qquad \sigma^2(t) = 1 - \alpha(t), \qquad b = 0.009. \tag{15}$$

The type process uses the same $\alpha(t)$, with $K = 23$ and Concrete temperature $\tau(t) = 0.4$. Here $\alpha(t)$ is the coordinate mean coefficient, with variance specified by $\sigma^2(t) = 1 - \alpha(t)$. The endpoint $t = 1$ gives the standard Gaussian coordinate prior and uniform categorical parameters. At $t = 0$, the corruption schedule has $\alpha(0) = 1 - b$ and $\sigma^2(0) = b$; the subscript 0 on clean targets denotes the uncorrupted data. Corrupted quantities evaluated at zero time are written explicitly as $\boldsymbol{x}_{t=0}$, $\boldsymbol{\theta}_{t=0}$, and $\mathbf{y}_{t=0}$ when needed.

Let $B$ denote the batch size and $N_m = N_{\mathcal{M}_m}$ the number of ligand atoms in molecule $m$. We use $T = 1000$ for the number of training intervals, distinguishing it from the atom counts.

## A.2 COORDINATE RECONSTRUCTION LOSS

The position loss is

$$\mathcal{L}_x = \frac{1}{B} \sum_{m=1}^{B} \frac{w_x(t_m)}{N_m} \sum_{i=1}^{N_m} \left\| \widehat{\boldsymbol{x}}_{0,i}^{(m)} - \boldsymbol{x}_{0,i}^{(m)} \right\|_2^2. \tag{16}$$

Its weight uses the coefficient SNR and the inherited discrete-time mapping:

$$\mathrm{SNR}_x(t) = \frac{\alpha^2(t)}{\sigma^2(t)} = \frac{\alpha^2(t)}{1 - \alpha(t)}, \tag{17}$$

$$\widetilde{t} = 1 - \frac{\lfloor (1-t)T \rfloor + 1}{T}, \qquad w_x(t) = \max\left\{1, \frac{\sqrt{T}}{2} \mathrm{SNR}_x(\widetilde{t})\right\}. \tag{18}$$

For $0 < t < 1$, $\widetilde{t}$ lies on the grid from 0 to $1 - 1/T$. This mapping is used only for coordinate weighting; the type loss below uses $t^- = \max(t - 1/T, 0)$.

**Relation to noise prediction.** For one atom, the Gaussian marginal gives $\boldsymbol{x}_t = \alpha(t)\boldsymbol{x}_0 + \sigma(t)\boldsymbol{\epsilon}$. The corresponding noise estimate and MSE identity are

$$\boldsymbol{\epsilon}_\phi = \frac{\boldsymbol{x}_t - \alpha(t)\widehat{\boldsymbol{x}}_0}{\sigma(t)}, \qquad \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\phi = \frac{\alpha(t)}{\sigma(t)}(\widehat{\boldsymbol{x}}_0 - \boldsymbol{x}_0),$$
$$\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\phi\|_2^2 = \mathrm{SNR}_x(t)\|\widehat{\boldsymbol{x}}_0 - \boldsymbol{x}_0\|_2^2. \tag{19}$$

This identity motivates SNR weighting of clean-coordinate prediction. Equation (18) additionally retains the MolPIF discrete-time scaling $\sqrt{T}/2$ and applies a floor of 1 so that coordinate supervision remains active at high noise. These scaling and flooring choices specify the training objective in addition to the noise-prediction identity.

## A.3 ADJACENT-TIME PARAMETER KL FOR ATOM TYPES

At $t^- = \max(t - 1/T, 0)$, apply Equation (6) to the clean type and the predicted clean probabilities. For atom $i$, this gives

$$\begin{aligned} \boldsymbol{\theta}_{t^-,i} &= \alpha(t^-)\boldsymbol{v}_{0,i} + (1 - \alpha(t^-))\mathbf{u}, \\ \widehat{\boldsymbol{\theta}}_{t^-,i} &= \alpha(t^-)\widehat{\boldsymbol{\theta}}_{0,i} + (1 - \alpha(t^-))\mathbf{u}. \end{aligned} \tag{20}$$

The atom-type loss is

$$\mathcal{L}_v = \frac{1}{B} \sum_{m=1}^{B} \frac{1}{N_m} \sum_{i=1}^{N_m} D_{\mathrm{KL}}\left(\boldsymbol{\theta}_{t_m^-,i}^{(m)} \,\middle\|\, \widehat{\boldsymbol{\theta}}_{t_m^-,i}^{(m)}\right), \tag{21}$$

where $D_{\mathrm{KL}}(\boldsymbol{\theta}\|\widehat{\boldsymbol{\theta}}) = \sum_{k=1}^{K} \theta_k \log(\theta_k/\widehat{\theta}_k)$ for a single atom. This follows by applying the same forward parameter map to the clean one-hot target and the predicted clean probabilities, then comparing the resulting categorical distributions. Concrete sampling supplies the network input; the loss compares deterministic parameter vectors conditional on that input, without an additional target-sampling step. It supervises the full predicted distribution and uses the same parameter map as the type-sampling update. Thus, $\mathcal{L}_v$ is the *adjacent-time parameter-space KL* between categorical probability vectors; Concrete samples serve as network observations.

As the target-time signal $\alpha(t^-)$ decreases, both vectors approach the uniform prior and the loss weakens. With fixed $T = 1000$, $t \to 1$ gives $t^- \to 1 - 1/T$, so the limiting signal is small but nonzero. Near the data endpoint, supervision is close to clean-label supervision, while retaining the smoothing from $\alpha(0) = 1 - b$. Appendix D gives the clean-posterior consistency and gradient analysis using the same parameter notation.

**Combined objective and GeoRank.** MEGA and MEGA+GeoRank use Equation (9) with $\lambda_v = 2$. GeoRank supplies conditioning through the network adapter; both configurations are trained with the same coordinate and atom-type losses, without auxiliary rank supervision.

## B Coordinate process derivations

As in Section 3.2.1, we write $\boldsymbol{x}_0$ for a clean atom coordinate and $\boldsymbol{x}_t$ for its corrupted state. To distinguish the low-noise endpoint from the clean target, we explicitly write $\boldsymbol{x}_{t=0}$ for a draw at zero corruption time. Let $\beta(t) = b^{1-t}$, with $b \in (0, 1)$; the training setting is $b = 0.009$.

$$q_{0t}(\boldsymbol{x}_t \mid \boldsymbol{x}_0) = \mathcal{N}(\alpha(t)\boldsymbol{x}_0, \sigma^2(t)\mathbf{I}_3), \qquad \boldsymbol{x}_t = \alpha(t)\boldsymbol{x}_0 + \sigma(t)\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3). \tag{22}$$

$$\alpha(t) = 1 - \beta(t), \qquad \sigma^2(t) = \beta(t), \qquad \mathrm{SNR}_x(t) = \frac{(1-\beta(t))^2}{\beta(t)}. \tag{23}$$

### B.1 Endpoints and monotonicity

Let $c_b = -\log b > 0$ and $\beta(t) = e^{-c_b(1-t)}$. Then $\beta'(t) = c_b\beta(t)$, $\alpha'(t) = -c_b\beta(t) < 0$, and $(\sigma^2(t))' = c_b\beta(t) > 0$. Moreover,

$$\mathrm{SNR}_x(t) = \beta(t)^{-1} - 2 + \beta(t), \qquad \frac{d}{dt}\mathrm{SNR}_x(t) = c_b(\beta(t) - \beta(t)^{-1}) < 0 \quad (0 \le t < 1). \tag{24}$$

Thus the coefficient SNR decreases continuously from $(1-b)^2/b$ to zero. At $b = 0.009$, this maximum is approximately 109.12. After undoing coordinate scaling, the additive noise has standard deviation $2\sqrt{b} \approx 0.1897$ Å per Cartesian component.

### B.2 Compatible coordinate transitions

**Proposition 1** (Compatible coordinate transitions)**.** *For $0 \le s < t \le 1$, with $s < 1$, set $\alpha_{t|s} = \alpha(t)/\alpha(s)$ and $\sigma^2_{t|s} = \sigma^2(t) - \alpha^2_{t|s}\sigma^2(s)$. The schedule in Equation (23) has $\sigma^2_{t|s} \ge 0$. Consequently,*

$$q_{st}(\boldsymbol{x}_t \mid \boldsymbol{x}_s) = \mathcal{N}(\alpha_{t|s}\boldsymbol{x}_s, \sigma^2_{t|s}\mathbf{I}_3) \tag{25}$$

*defines compatible transitions on any finite ordered grid, starting with the specified low-noise distribution at $t = 0$.*

**Proof.** For $s < t < 1$, monotone SNR implies $\sigma^2(t)/\alpha^2(t) \ge \sigma^2(s)/\alpha^2(s)$, hence $\sigma^2_{t|s} \ge 0$. Multiplying $\boldsymbol{x}_s$ by $\alpha_{t|s}$ yields conditional mean $\alpha(t)\boldsymbol{x}_0$. Adding independent noise of covariance $\sigma^2_{t|s}\mathbf{I}_3$ yields covariance $\sigma^2(t)\mathbf{I}_3$. The transitions compose because, for $s < t < u$,

$$\alpha_{u|t}\alpha_{t|s} = \alpha_{u|s}, \qquad \sigma^2_{u|t} + \alpha^2_{u|t}\sigma^2_{t|s} = \sigma^2_{u|s}. \tag{26}$$

At the terminal time, $\alpha_{1|s} = 0$ and $\sigma^2_{1|s} = 1$, giving an independent standard Gaussian on a finite grid. This proves consistency of the stated transition family.

### B.3 Continuous-time representation

On every compact subinterval of $[0, 1)$, choose

$$\mathrm{d}\boldsymbol{x}_t = f(t)\boldsymbol{x}_t\,\mathrm{d}t + g(t)\,\mathrm{d}\mathbf{w}_t, \qquad f(t) = \frac{\alpha'(t)}{\alpha(t)}, \qquad g^2(t) = (\sigma^2(t))' - 2f(t)\sigma^2(t). \tag{27}$$

Initialize $\boldsymbol{x}_{t=0} \mid \boldsymbol{x}_0 \sim \mathcal{N}((1-b)\boldsymbol{x}_0, b\mathbf{I}_3)$ independently of Brownian increments. Its mean $\boldsymbol{\mu}$ and scalar variance $v$ solve $\boldsymbol{\mu}' = f\boldsymbol{\mu}$ and $v' = 2fv + g^2$, giving Equation (22). For our schedule,

$$f(t) = -\frac{c_b\beta(t)}{1-\beta(t)}, \qquad g^2(t) = \frac{c_b\beta(t)(1+\beta(t))}{1-\beta(t)} > 0. \tag{28}$$

The SDE represents the prescribed marginals on $[0, 1)$, with coefficients singular at the terminal limit and marginals converging to $\mathcal{N}(\mathbf{0}, \mathbf{I}_3)$ as $t \uparrow 1$. Our finite-grid sampler initializes directly from this terminal Gaussian and applies the updates in Appendix F.

For the coordinate-only marginal conditioned on a fixed pocket, standard score-based theory gives reverse drift $f(t)\boldsymbol{x}_t - g^2(t)\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}_t \mid \mathcal{P})$ with decreasing time, and probability-flow drift $f(t)\boldsymbol{x}_t - \frac{1}{2}g^2(t)\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}_t \mid \mathcal{P})$ (Song et al., 2020b). These coordinate drifts characterize the continuous-time representation. The implemented joint coordinate–type sampler is specified by Equations (55) and (56).

### B.4 Conditional reverse kernel on a finite grid

Gaussian conditioning gives

$$q(\boldsymbol{x}_s \mid \boldsymbol{x}_t, \boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{\mu}_{s|t}, \widetilde{\sigma}^2_{s|t}\mathbf{I}_3), \tag{29}$$

$$\boldsymbol{\mu}_{s|t} = \alpha(s)\boldsymbol{x}_0 + \frac{\alpha_{t|s}\sigma^2(s)}{\sigma^2(t)}(\boldsymbol{x}_t - \alpha(t)\boldsymbol{x}_0), \tag{30}$$

$$\widetilde{\sigma}^2_{s|t} = \sigma^2(s) - \frac{\alpha^2_{t|s}\sigma^4(s)}{\sigma^2(t)} = \frac{\sigma^2(s)\sigma^2_{t|s}}{\sigma^2(t)}. \tag{31}$$

Averaging this kernel over the clean-ligand posterior gives the reverse distribution conditioned on $\boldsymbol{x}_t$ and the pocket. The sampler uses the predicted clean state as a plug-in approximation. A generalized update written as $\boldsymbol{x}_s = \alpha(s)\widehat{\boldsymbol{x}}_0 + \sqrt{\sigma^2(s) - \nu^2_{s|t}}\boldsymbol{\epsilon}_\phi + \nu_{s|t}\boldsymbol{\xi}$ recovers the oracle conditional kernel when $\nu^2_{s|t} = \widetilde{\sigma}^2_{s|t}$. In the notation of Equation (55), this requires

$$\eta^2_{s|t} = 1 - \frac{\alpha^2(t)\sigma^2(s)}{\alpha^2(s)\sigma^2(t)}, \tag{32}$$

This time-dependent setting recovers the conditional kernel; the main sampler uses the coupling with fixed $\eta_x = 1$ described below.

### B.5 Marginal preservation under clean-state prediction

If $\boldsymbol{x}_t = \alpha(t)\boldsymbol{x}_0 + \sigma(t)\boldsymbol{\epsilon}$ and $\widehat{\boldsymbol{x}}_0 = \boldsymbol{x}_0$, then $\boldsymbol{\epsilon}_\phi = \boldsymbol{\epsilon}$. Independent $\boldsymbol{\epsilon}, \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3)$ imply $\sqrt{1-\eta_x^2}\boldsymbol{\epsilon} + \eta_x\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3)$ for $0 \le \eta_x \le 1$. Consequently, the coordinate update preserves the prescribed marginal when the clean state is predicted exactly. For a learned denoiser, reconstruction accuracy depends on its clean-state predictions. The type update in Equation (56) remains in the simplex for every predicted probability vector. For a clean-class posterior $\boldsymbol{\rho}$, its induced categorical parameter is $\alpha(s)\boldsymbol{\rho} + (1-\alpha(s))\mathbf{u}$. By the Concrete argmax law, this update preserves posterior-averaged winning-class probabilities. Its continuous observation density is the Concrete distribution at that parameter, which generally differs from a mixture over clean-class Concrete densities.

## C Categorical semantics and information

The atom-type probability trajectory builds on categorical diffusion (Hoogeboom et al., 2021; Austin et al., 2021). We use the main-text notation throughout: $\boldsymbol{\theta}_t$ is a categorical parameter vector, $\mathbf{y}_t$ is its

Concrete observation, and $\mathbf{u} = \mathbf{1}_K/K$. For a clean class $c$, $\boldsymbol{v}_0 = \mathbf{e}_c$ and

$$\boldsymbol{\theta}_t = \alpha(t)\mathbf{e}_c + (1-\alpha(t))\mathbf{u}, \qquad y_{t,k} = \frac{\exp((\log\theta_{t,k} + g_k)/\tau(t))}{\sum_j \exp((\log\theta_{t,j} + g_j)/\tau(t))}. \tag{33}$$

Here the $g_k$ are independent standard Gumbel variables and $\tau(t) > 0$. Per-atom observations are collected in $\mathbf{Y}_t$; $\theta_{t,k}$ and $y_{t,k}$ denote scalar components of their respective vectors.

### C.1 Exact winning-class calibration

**Proposition 2** (Exact winning-class calibration)**.** *For a single Concrete draw with positive parameters and $\tau(t) > 0$, $\mathbf{y}_t$ lies in the simplex interior and*

$$\theta_{t,c} - \theta_{t,k} = \alpha(t) \quad (k \neq c), \qquad \Pr(\arg\max_k y_{t,k} = c \mid c) = \alpha(t) + \frac{1-\alpha(t)}{K}. \tag{34}$$

*More generally,* $\Pr(\arg\max_k y_{t,k} = j \mid c) = \theta_{t,j}$.

**Proof.** Convexity gives $\boldsymbol{\theta}_t \in \Delta^{K-1}$; if $\alpha(t) < 1$, all parameters are strictly positive. Softmax returns an interior simplex point. The parameter gap follows by subtraction. To prove the argmax law, write $E_k = -\log U_k \sim \mathrm{Exp}(1)$ and $g_k = -\log E_k$. Maximizing $\log\theta_{t,k} + g_k$ is equivalent to minimizing $E_k/\theta_{t,k}$. These are independent exponential variables with rates $\theta_{t,k}$, so the minimum has index $k$ with probability $\theta_{t,k}/\sum_j \theta_{t,j} = \theta_{t,k}$. Positive temperature and softmax preserve order. At signal coefficient 1, the one-hot case is defined by its limit.

For a single Concrete observation, $\boldsymbol{\theta}_t$ specifies categorical winning probabilities. Its componentwise expectation $\mathbb{E}[\mathbf{y}_t \mid c]$ is generally a different quantity at finite temperature. MEGA uses a single draw as specified in Algorithm 1.

### C.2 Information in categorical observations

Let $C$ be the clean class and $A_t = \arg\max_k y_{t,k}$. With a uniform class prior, this is a symmetric categorical channel with correct-class probability $p_{\mathrm{hit}}(t) = \alpha(t) + (1-\alpha(t))/K$ and each off-diagonal probability $p_{\mathrm{off}}(t) = (1-\alpha(t))/K$. Direct evaluation gives

$$I(C; A_t) = \log K + p_{\mathrm{hit}}(t)\log p_{\mathrm{hit}}(t) + (K-1)p_{\mathrm{off}}(t)\log p_{\mathrm{off}}(t), \tag{35}$$

using natural logarithms and $0\log 0 = 0$. This expression characterizes the winning-class observation under a uniform prior. For an arbitrary clean-class prior $\boldsymbol{\rho}$, including imbalanced atom types,

$$\begin{aligned} I(C; A_t) = {} & H(\alpha(t)\boldsymbol{\rho} + (1-\alpha(t))\mathbf{u}) \\ & + p_{\mathrm{hit}}(t)\log p_{\mathrm{hit}}(t) + (K-1)p_{\mathrm{off}}(t)\log p_{\mathrm{off}}(t). \end{aligned} \tag{36}$$

Because $A_t$ is a deterministic function of $\mathbf{y}_t$, the data-processing inequality gives $I(C; A_t) \leq I(C; \mathbf{y}_t)$. The continuous observation therefore retains at least the class information available in its hard assignment. Joint geometric and chemical prediction is learned through the shared denoiser.

### C.3 Temperature changes coordinates of the representation

For two positive temperatures $\tau, \tau'$, coupled through the same Gumbel variables at a fixed time $t$,

$$y_{t,k}^{(\tau')} = \frac{(y_{t,k}^{(\tau)})^{\tau/\tau'}}{\sum_j (y_{t,j}^{(\tau)})^{\tau/\tau'}}. \tag{37}$$

For known deterministic temperatures $0 < \tau, \tau' < \infty$, this map is bijective on the simplex interior. In exact arithmetic, $I(C; \mathbf{y}_t^{(\tau)}) = I(C; \mathbf{y}_t^{(\tau')})$. Temperature thus controls the geometry of the continuous representation while preserving its class information under these conditions. It also affects network optimization, numerical precision, and gradients; MEGA uses $\tau = 0.4$.

# D PARAMETER KL: RISK, GRADIENTS, AND IDENTIFIABILITY

## D.1 CLEAN-POSTERIOR CONSISTENCY

Fix the complete denoiser input $\mathcal{W}$, including time $t$, and let $\boldsymbol{\rho} = \Pr(C = \cdot \mid \mathcal{W})$ denote the true clean-class posterior. At target time $s = t^-$, write

$$\begin{aligned} \boldsymbol{\theta}_s^{(c)} &= \alpha(s)\mathbf{e}_c + (1-\alpha(s))\mathbf{u}, \qquad \boldsymbol{\rho}_s = \alpha(s)\boldsymbol{\rho} + (1-\alpha(s))\mathbf{u}, \\ \widehat{\boldsymbol{\theta}}_s &= \alpha(s)\widehat{\boldsymbol{\theta}}_0 + (1-\alpha(s))\mathbf{u}. \end{aligned} \tag{38}$$

The superscript $(c)$ indicates conditioning on clean class $c$; $\widehat{\boldsymbol{\theta}}_0$ is a candidate clean-type prediction.

**Proposition 3** (Clean-posterior consistency)**.** *For $0 < \alpha(s) < 1$, the conditional risk $R_{\mathcal{W}}(\widehat{\boldsymbol{\theta}}_0) = \mathbb{E}_{C|\mathcal{W}}[D_{\mathrm{KL}}(\boldsymbol{\theta}_s^{(C)} \| \widehat{\boldsymbol{\theta}}_s)]$ satisfies*

$$R_{\mathcal{W}}(\widehat{\boldsymbol{\theta}}_0) - R_{\mathcal{W}}(\boldsymbol{\rho}) = D_{\mathrm{KL}}(\boldsymbol{\rho}_s \| \widehat{\boldsymbol{\theta}}_s) \geq 0. \tag{39}$$

*Its unique minimizer over the closed simplex is $\widehat{\boldsymbol{\theta}}_0 = \boldsymbol{\rho}$. For $\alpha(s) = 1$, the same conclusion holds with the usual extended-value KL convention; for $\alpha(s) = 0$, all predictions minimize the risk.*

**Proof.** Fix $\mathcal{W}$ and let $\rho_c = \Pr(C = c \mid \mathcal{W})$. Since $\sum_c \rho_c \boldsymbol{\theta}_s^{(c)} = \boldsymbol{\rho}_s$,

$$R_{\mathcal{W}}(\widehat{\boldsymbol{\theta}}_0) = \sum_c \rho_c \sum_k \theta_{s,k}^{(c)} \log \theta_{s,k}^{(c)} - \sum_k \left( \sum_c \rho_c \theta_{s,k}^{(c)} \right) \log \widehat{\theta}_{s,k} \tag{40}$$

$$= h(\mathcal{W}) - \sum_k \rho_{s,k} \log \widehat{\theta}_{s,k}, \tag{41}$$

where $h(\mathcal{W})$ does not depend on the prediction. Subtracting the risk at $\widehat{\boldsymbol{\theta}}_0 = \boldsymbol{\rho}$ yields Equation (39). For $0 < \alpha(s) < 1$, $\boldsymbol{\rho}_s$ and $\widehat{\boldsymbol{\theta}}_s$ have positive entries. Equality of their KL requires equality of the vectors, and the affine parameter map is injective when $\alpha(s) > 0$, hence $\widehat{\boldsymbol{\theta}}_0 = \boldsymbol{\rho}$. For $\alpha(s) = 1$ this reduces to cross-entropy; for $\alpha(s) = 0$ the loss is identically zero. If the optimum is on a simplex boundary, finite logits attain only an infimum. The argument conditions on the full input, including the pocket, all noisy atom states, atom count, and time; it allows dependencies among clean atom labels.

This population-risk result establishes clean-posterior consistency of the parameter-space objective at every target time with positive signal, including the adjacent-time choice $s = t^-$ used in MEGA. Appendix E separately analyzes attainable variational fit and the effect of geometric conditioning.

## D.2 EXACT LOGIT GRADIENT

For one atom with clean type $\boldsymbol{v}_0 = \mathbf{e}_c$, let $\widehat{\boldsymbol{\theta}}_0 = \mathrm{softmax}(\mathbf{z}_\phi)$, $\boldsymbol{\theta}_s = \alpha(s)\boldsymbol{v}_0 + (1-\alpha(s))\mathbf{u}$, and $\ell_v = D_{\mathrm{KL}}(\boldsymbol{\theta}_s \| \widehat{\boldsymbol{\theta}}_s)$. Differentiation gives

$$\frac{\partial \ell_v}{\partial z_{\phi,j}} = \alpha(s)\widehat{\theta}_{0,j} \left( \sum_k \frac{\widehat{\theta}_{0,k}\theta_{s,k}}{\widehat{\theta}_{s,k}} - \frac{\theta_{s,j}}{\widehat{\theta}_{s,j}} \right). \tag{42}$$

When $\alpha(s) = 1$ and the predicted probabilities are positive, this becomes $\widehat{\theta}_{0,j} - \mathbf{1}[j = c]$, the CE gradient. Around $\alpha(s) = 0$, write $\boldsymbol{\theta}_s = \mathbf{u} + \alpha(s)(\boldsymbol{v}_0 - \mathbf{u})$ and $\widehat{\boldsymbol{\theta}}_s = \mathbf{u} + \alpha(s)(\widehat{\boldsymbol{\theta}}_0 - \mathbf{u})$. A second-order expansion of KL around $\mathbf{u}$ gives

$$\ell_v = \frac{K\alpha^2(s)}{2} \|\boldsymbol{v}_0 - \widehat{\boldsymbol{\theta}}_0\|_2^2 + O(\alpha^3(s)). \tag{43}$$

For fixed $K$,

$$\nabla_{\mathbf{z}_\phi} \ell_v = K\alpha^2(s) \left( \mathrm{diag}(\widehat{\boldsymbol{\theta}}_0) - \widehat{\boldsymbol{\theta}}_0 \widehat{\boldsymbol{\theta}}_0^\top \right) (\widehat{\boldsymbol{\theta}}_0 - \boldsymbol{v}_0) + O(\alpha^3(s)). \tag{44}$$

Both the loss and its logit gradient scale quadratically with $\alpha(s)$ near the uniform prior for fixed logits. In MEGA, $s = t^-$ retains a small positive signal as $t \to 1$ for fixed $T$. At the data endpoint, $\alpha(0) = 1 - b$ yields smoothed clean-label supervision. These expressions characterize how the prescribed trajectory modulates atom-type supervision across noise levels.

# E GeoRank: variational fit and likelihood

We follow the comparison of optimal variational bounds in IPDiff (Huang et al., 2024). Here the forward corruption is fixed, and GeoRank enriches the reverse model through persistent geometric conditioning.

## E.1 Improved attainable variational fit

Fix the pocket $\mathcal{P}$ and ligand atom count $N$, with conditioning on $N$ implicit below. Let $\mathcal{M} = (\mathbf{X}_0, \mathbf{V}_0)$ be the clean ligand and $\mathbf{r} = \mathcal{R}(\mathbf{X}_0, \mathcal{P})$ its rank assignment. For a finite time grid $0 < t_1 < \cdots < t_S = 1$, write $\mathbf{Z}_j = (\mathbf{X}_{t_j}, \mathbf{Y}_{t_j})$ and fix a normalized forward path density $q(\mathbf{Z}_{1:S} \mid \mathcal{M}, \mathcal{P})$ with the prescribed corruption marginals. Consider normalized reverse kernels and a shared, rank-independent terminal prior:

$$p^r_\phi(\mathcal{M}, \mathbf{Z}_{1:S} \mid \mathcal{P}, \mathbf{r}) = p_{\text{prior}}(\mathbf{Z}_S \mid \mathcal{P})\, p^r_\phi(\mathcal{M} \mid \mathbf{Z}_1, \mathcal{P}, \mathbf{r}) \prod_{j=2}^{S} p^r_\phi(\mathbf{Z}_{j-1} \mid \mathbf{Z}_j, \mathcal{P}, \mathbf{r}). \tag{45}$$

The corresponding conditional evidence lower bound (ELBO) is

$$\mathcal{B}^r_\phi(\mathcal{M}, \mathcal{P}) = \mathbb{E}_q \left[ \log \frac{p^r_\phi(\mathcal{M}, \mathbf{Z}_{1:S} \mid \mathcal{P}, \mathbf{r})}{q(\mathbf{Z}_{1:S} \mid \mathcal{M}, \mathcal{P})} \right] \le \log p^r_\phi(\mathcal{M} \mid \mathcal{P}, \mathbf{r}). \tag{46}$$

Superscript 0 denotes the corresponding model and bound without rank conditioning; expectations over clean ligands are taken under the data distribution.

**Proposition 4** (Attainable variational bound)**.** *If disabling the rank adapter recovers the base model family, then, for the same forward path and terminal prior,*

$$\sup_\phi \mathbb{E}_{\text{data}} \mathcal{B}^r_\phi(\mathcal{M}, \mathcal{P}) \ge \sup_\phi \mathbb{E}_{\text{data}} \mathcal{B}^0_\phi(\mathcal{M}, \mathcal{P}). \tag{47}$$

*Thus GeoRank admits a no-worse optimal conditional ELBO, equivalently a no-larger optimal variational bound on conditional negative log likelihood.*

**Proof.** Disabling the adapter reproduces every base path density and its ELBO, independently of $\mathbf{r}$. The inequality follows by optimizing over the larger family, with all time steps sharing the same network parameters.

To characterize when conditioning helps, let $\mathbf{U}_j = \mathbf{Z}_{j-1}$ for $j \ge 2$ and $\mathbf{U}_1 = \mathcal{M}$, and set $\mathcal{W}_j = (\mathbf{Z}_j, \mathcal{P}, N, t_j)$. For unrestricted reverse densities and finite conditional entropies,

$$\begin{aligned} &\inf_p \mathbb{E}[-\log p(\mathbf{U}_j \mid \mathcal{W}_j)] - \inf_p \mathbb{E}[-\log p(\mathbf{U}_j \mid \mathcal{W}_j, \mathbf{r})] \\ &\qquad = I(\mathbf{U}_j; \mathbf{r} \mid \mathcal{W}_j) \ge 0. \end{aligned} \tag{48}$$

The minimizing densities are the true conditional distributions; their expected log-density ratio is the conditional mutual information. Ranks with positive residual information therefore strictly improve optimal reverse fitting, including when noise obscures geometric information.

## E.2 Connection to the implemented training objective

The same conditioning argument applies directly to our losses. Let $\mathcal{W} = (\mathbf{X}_t, \mathbf{Y}_t, t, \mathcal{P}, N)$, $\mathbf{m} = \mathbb{E}[\mathbf{X}_0 \mid \mathcal{W}]$, and $\mathbf{m}^r = \mathbb{E}[\mathbf{X}_0 \mid \mathcal{W}, \mathbf{r}]$. For atom types, define the clean-class posterior vectors $\boldsymbol{\rho}_i = \mathbb{E}[\boldsymbol{v}_{0,i} \mid \mathcal{W}]$ and $\boldsymbol{\rho}^r_i = \mathbb{E}[\boldsymbol{v}_{0,i} \mid \mathcal{W}, \mathbf{r}]$. At $s = t^-$, their smoothed counterparts are

$$\boldsymbol{\rho}_{s,i} = \alpha(s)\boldsymbol{\rho}_i + (1 - \alpha(s))\mathbf{u}, \qquad \boldsymbol{\rho}^r_{s,i} = \alpha(s)\boldsymbol{\rho}^r_i + (1 - \alpha(s))\mathbf{u}. \tag{49}$$

Let $\mathcal{L}^{\star,0}_x, \mathcal{L}^{\star,0}_v$ and $\mathcal{L}^{\star,r}_x, \mathcal{L}^{\star,r}_v$ denote the infima of the population losses over all predictors using $\mathcal{W}$ and $(\mathcal{W}, \mathbf{r})$, respectively. Then

$$\mathcal{L}^{\star,0}_x - \mathcal{L}^{\star,r}_x = \mathbb{E}\left[ \frac{w_x(t)}{N} \|\mathbf{m}^r - \mathbf{m}\|_F^2 \right] \ge 0, \tag{50}$$

$$\mathcal{L}^{\star,0}_v - \mathcal{L}^{\star,r}_v = \mathbb{E}\left[ \frac{1}{N} \sum_{i=1}^{N} D_{\text{KL}}(\boldsymbol{\rho}^r_{s,i} \| \boldsymbol{\rho}_{s,i}) \right] \ge 0. \tag{51}$$

**Proof.** The coordinate identity follows from the orthogonal decomposition of conditional squared error; the implemented positive weight $w_x(t)/N$ is measurable with respect to $\mathcal{W}$. For types, apply the excess-risk identity in Proposition 3 conditional on $(\mathcal{W}, \mathbf{r})$ to the prediction $\boldsymbol{\rho}_i$, then average to obtain Equation (51). Informative ranks can therefore reduce both Bayes risks under our implemented losses. For a finite shared network, disabling the adapter likewise ensures a no-worse best attainable combined objective. Section E.4 evaluates the improvement achieved during training.

### E.3 Likelihood after marginalizing ranks

Generation samples ranks rather than observing a reference ligand. Let $\mathfrak{S}_N$ be the set of rank permutations and use the prior $p_{\mathrm{rank}}(\mathbf{r} \mid \mathcal{P}, N) = 1/N!$. Under permutation symmetrization and a permutation-equivariant rank rule, this matches the training rank marginal. The ligand likelihood is

$$p_\phi^{\mathrm{Geo}}(\mathcal{M} \mid \mathcal{P}) = \frac{1}{N!} \sum_{\mathbf{r} \in \mathfrak{S}_N} p_\phi^r(\mathcal{M} \mid \mathcal{P}, \mathbf{r}), \qquad \log p_\phi^{\mathrm{Geo}}(\mathcal{M} \mid \mathcal{P}) \geq \mathcal{B}_\phi^r(\mathcal{M}, \mathcal{P}) - \log N!. \tag{52}$$

The bound uses the deterministic training assignment $q_{\mathrm{rank}}(\mathbf{r} \mid \mathcal{M}, \mathcal{P}) = \delta_{\mathcal{R}(\mathbf{X}_0, \mathcal{P})}$; its KL to the uniform rank prior is $\log N!$, independent of model parameters. Moreover, the disabled-adapter setting makes every mixture component equal to the base model, so model-family inclusion gives

$$\sup_\phi \mathbb{E}_{\mathrm{data}} \log p_\phi^{\mathrm{Geo}}(\mathcal{M} \mid \mathcal{P}) \geq \sup_\phi \mathbb{E}_{\mathrm{data}} \log p_\phi^0(\mathcal{M} \mid \mathcal{P}). \tag{53}$$

For the normalized path model above, this is a model-family result: GeoRank supports equally good or better optimal likelihood. The implemented sampler uses parameter-space supervision and final deterministic decoding; its training objective is characterized by the Bayes-risk identities in Equations (50)–(51). Empirical denoising performance is evaluated below.

### E.4 Validation evidence for variational fitting

Figure 5 compares the validation denoising objectives of MEGA and MEGA with GeoRank over training. It includes all recorded validation events and separates the coordinate and atom-type terms of the objective in Equation (9).

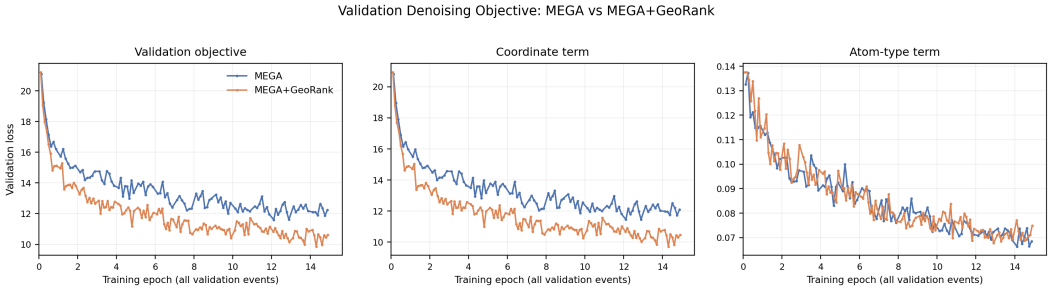



Figure 5: Validation denoising losses for MEGA and MEGA with GeoRank across all recorded validation events. The panels show the total objective (left), coordinate term (middle), and atom-type term (right) against training epoch. GeoRank lowers the total objective primarily through the coordinate term, while atom-type losses remain comparable.

After the initial transient, MEGA+GeoRank maintains a lower total validation loss across the remaining training trajectory. The coordinate curves closely track this separation, whereas the atom-type curves largely overlap and approach similar values. The observed gain is concentrated in coordinate fitting. The adapter updates both hidden features and coordinates (Appendix G.2), with comparable atom-type loss maintained during training.

These curves provide empirical support for the variational-fitting analysis: the additional geometric conditions improve the measured denoising objective, with the dominant coordinate reduction consistent with Equation (50). Figure 5 measures the implemented denoising objective, aggregated over sampled diffusion times at each validation event. Together with the population-risk analysis, it supports improved geometric fitting under GeoRank conditioning.

### E.5 PAIRED PATHWISE VARIATIONAL-BOUND EVALUATION

We evaluated MEGA (epoch 10) and MEGA + Virtual-anchor GeoRank (fine-tuning step 2500) on the first 10 molecules in the test-set order. Each molecule used $M = 100$ sampled paths with $S = 100$ discretization steps, processed in batches of 10 paths. Both models used the same molecules, ordering, and paired random seed (1234). Evaluation conditions on the protein pocket and reference atom count $N$; the GeoRank configuration additionally conditions on ranks $\mathbf{r} = \mathcal{R}(\mathbf{X}_0, \mathcal{P})$ computed from the clean ligand by virtual-anchor ordering. The resulting estimates measure atom-count-conditional fit for MEGA and rank-conditional fit for MEGA+GeoRank.

**Path densities and estimators.** Let $\mathcal{C} = (\mathcal{P}, N)$ for MEGA and $\mathcal{C} = (\mathcal{P}, N, \mathbf{r})$ for MEGA+GeoRank. For proposal paths $\mathbf{Z}_{1:S}^{(m)}$, we compute

$$\ell_m = \log \frac{p_\phi(\mathcal{M}, \mathbf{Z}_{1:S}^{(m)} \mid \mathcal{C})}{q(\mathbf{Z}_{1:S}^{(m)} \mid \mathcal{M}, \mathcal{C})}, \qquad \widehat{\mathcal{B}}_{\mathrm{path}} = \frac{1}{M} \sum_{m=1}^{M} \ell_m, \qquad \widehat{\mathcal{B}}_{\mathrm{IW},M} = \log \left( \frac{1}{M} \sum_{m=1}^{M} e^{\ell_m} \right). \quad (54)$$

The mean path estimate decomposes into coordinate and atom-type terms, $\widehat{\mathcal{B}}_{\mathrm{path}} = \widehat{\mathcal{B}}_x + \widehat{\mathcal{B}}_v$. We report IW-NLL per atom as $-\widehat{\mathcal{B}}_{\mathrm{IW},M}/N$. Coordinates were pocket-centered. The coordinate proposal used the Gaussian Markov transitions induced by the prescribed marginals, with analytic reverse conditionals from Equation (31). Model densities used the DDIM-style marginal reverse kernel with $\eta_x = 1$ and a Gaussian clean-coordinate decoder of standard deviation $0.094868$ model units ($0.189737$ Å). For atom types, proposal parameters were obtained from the clean labels, and model reverse parameters from predicted clean-type probabilities. Each state used one Concrete draw and its analytic density. The density implementation agreed with PyTorch's `RelaxedOneHotCategorical` to a maximum absolute error of $1.36 \times 10^{-7}$.

Table 3: Paired variational-bound evaluation on 10 test molecules. Here MEGA+GeoRank denotes the Virtual-anchor GeoRank variant. Values are molecule-wise means in natural-log units; normalized metrics are divided by $3N$ (coordinate dimensions) or $N$ (atoms) before averaging. Positive improvement favors MEGA+GeoRank, including the reduction in IW-NLL. Intervals are paired Student-$t$ 95% confidence intervals across molecules; wins count favorable paired differences. Bold marks the better mean.

| Metric | MEGA | MEGA+GeoRank | Improvement | 95% CI | Wins |
|---|---|---|---|---|---|
| Coordinate Path ELBO ↑ | $-9394.908$ | $\mathbf{-9349.383}$ | $+45.526$ | $[25.651, 65.401]$ | $10/10$ |
| Coordinate Path ELBO / dim ↑ | $-118.3653$ | $\mathbf{-117.7886}$ | $+0.5767$ | $[0.3813, 0.7721]$ | $10/10$ |
| Atom-type Path ELBO / atom ↑ | $-179.1475$ | $\mathbf{-177.3536}$ | $+1.7938$ | $[-5.0536, 8.6413]$ | $7/10$ |
| Joint Path ELBO ↑ | $-14110.197$ | $\mathbf{-14020.482}$ | $+89.716$ | $[-73.543, 252.975]$ | $8/10$ |
| IW-NLL / atom ↓ | $489.5806$ | $\mathbf{486.3666}$ | $+3.2140$ | $[-5.5570, 11.9850]$ | $8/10$ |

**Coordinate-path improvement.** MEGA+GeoRank improves coordinate Path ELBO on all 10 molecules, with a mean gain of $0.5767$ per dimension and a 95% confidence interval of $[0.3813, 0.7721]$; a two-sided sign test gives $p = 0.00195$. This consistent improvement across molecules directly supports better coordinate reverse-path fitting under geometric rank conditioning. The atom-type and joint path terms, together with IW-NLL, also improve on average, with their paired intervals reported in Table 3.

**Importance-weight statistics.** For the 100-path IW estimate, the mean effective sample size is $1.0005$ for MEGA and $1.0825$ for MEGA+GeoRank; the largest normalized path weight averages $99.98\%$ and $96.55\%$, respectively. These statistics show that the IW estimates are concentrated on a few paths. The paired coordinate Path ELBO, which averages log weights across all sampled paths, provides the primary evidence for improved variational fitting and complements the validation denoising results in Section E.4.

# F TRAINING AND GENERATION ALGORITHMS

Algorithms 1 and 2 summarize training and generation. We use the schedules and losses in Appendix A; $\mathbf{Y}_t$ collects the per-atom Concrete observations. All Gaussian noise matrices have independent standard normal entries. The rank argument $\mathbf{r}$ in $F_\phi$ is included only for MEGA+GeoRank, whose residual adapter is specified in Appendix G.2. Algorithm 1 shows the standard corruption procedure.

**Algorithm 1** Training procedure for MEGA and MEGA+GeoRank

**Input:** Ligand–pocket dataset $\mathcal{D}$, denoiser $F_\phi$, batch size $B$, $T = 1000$, $\lambda_v = 2$.
**Output:** Trained denoiser $F_\phi$.

**while** training is not complete **do**
  Sample a minibatch of $B$ ligand–pocket pairs from $\mathcal{D}$.
  **for** each pair $(\mathcal{M}, \mathcal{P})$ in the minibatch **do**
    Center the complex on the protein pocket and scale coordinates; obtain $(\mathbf{X}_0, \mathbf{V}_0)$.
    **if** GeoRank is enabled, compute $\mathbf{r} \leftarrow \mathcal{R}(\mathbf{X}_0, \mathcal{P})$ and keep it fixed.
    Sample $t \sim \mathcal{U}(0, 1)$ and $\mathbf{E} \sim \mathcal{N}(0, I)$.
    Perturb coordinates: $\mathbf{X}_t \leftarrow \alpha(t)\mathbf{X}_0 + \sigma(t)\mathbf{E}$.
    Set $\boldsymbol{\theta}_{t,i} \leftarrow \alpha(t)\boldsymbol{v}_{0,i} + (1 - \alpha(t))\mathbf{u}$ for each ligand atom $i$.
    Draw $\mathbf{y}_{t,i} \sim \mathrm{Concrete}(\boldsymbol{\theta}_{t,i}, \tau(t))$ independently and collect $\mathbf{Y}_t$.
    Predict $(\widehat{\mathbf{X}}_0, \widehat{\boldsymbol{\theta}}_0) \leftarrow F_\phi(\mathbf{X}_t, \mathbf{Y}_t, t, \mathcal{P}; \mathbf{r})$.
    Set $t^- \leftarrow \max(t - 1/T, 0)$ and compute the target and predicted type parameters by Equation (20).
  **end for**
  Compute $\mathcal{L}_x$ and $\mathcal{L}_v$ by Equations (16) and (21), using $w_x(t)$ in Equation (18); average within molecules, then across the batch.
  Update $\phi$ with Adam to minimize $\mathcal{L} = \mathcal{L}_x + \lambda_v \mathcal{L}_v$.
**end while**

**Sampling updates.** For $s < t$, predict the clean state from $(\mathbf{X}_t, \mathbf{Y}_t, t, \mathcal{P})$, additionally using $\mathbf{r}$ for GeoRank. Define $\mathbf{E}_\phi = (\mathbf{X}_t - \alpha(t)\widehat{\mathbf{X}}_0)/\sigma(t)$. With $0 \le \eta_x \le 1$ and independent standard Gaussian noise $\boldsymbol{\Xi}$, the coordinate update is

$$\mathbf{X}_s = \alpha(s)\widehat{\mathbf{X}}_0 + \sigma(s)\left(\sqrt{1 - \eta_x^2}\mathbf{E}_\phi + \eta_x \boldsymbol{\Xi}\right). \tag{55}$$

For each atom, update the type parameter and sample its observation:

$$\widehat{\boldsymbol{\theta}}_{s,i} = \alpha(s)\widehat{\boldsymbol{\theta}}_{0,i} + (1 - \alpha(s))\mathbf{u}, \qquad \mathbf{y}_{s,i} \sim \mathrm{Concrete}(\widehat{\boldsymbol{\theta}}_{s,i}, \tau(s)). \tag{56}$$

As in the coordinate derivation, $\mathbf{X}_{t=0}$ denotes the final low-noise state, while $\mathbf{X}_0$ denotes clean coordinates. Similarly, $\mathbf{Y}_{t=0}$ is the final sampled type input, not the clean one-hot target.

**Algorithm 2** Sampling procedure for MEGA and MEGA+GeoRank

**Input:** Protein pocket $\mathcal{P}$, trained denoiser $F_\phi$, pocket-conditioned atom-count prior, $S = 100$, coordinate noise level $\eta_x$.
**Output:** Generated ligand molecule $\mathcal{M}$.

Sample $N = N_{\mathcal{M}}$ from the pocket-conditioned atom-count prior.
Center and scale the pocket coordinates; retain the inverse transform.
Sample $\mathbf{X}_1 \sim \mathcal{N}(0, I)$ and, independently for each atom, $\mathbf{y}_{1,i} \sim \mathrm{Concrete}(\mathbf{u}, \tau(1))$.
**if** GeoRank is enabled, draw a uniform permutation $\mathbf{r}$ of $\{0, \dots, N-1\}$ independently of the noise and keep it fixed.
**for** $j = S, S-1, \dots, 1$ **do**
    Set $t \leftarrow j/S$ and $s \leftarrow (j-1)/S$.
    Predict $(\widehat{\mathbf{X}}_0, \widehat{\boldsymbol{\theta}}_0) \leftarrow F_\phi(\mathbf{X}_t, \mathbf{Y}_t, t, \mathcal{P}; \mathbf{r})$.
    Update $\widehat{\boldsymbol{\theta}}_{s,i}$ and sample $\mathbf{y}_{s,i}$ for each atom using Equation (56); collect $\mathbf{Y}_s$.
    Compute $\mathbf{E}_\phi \leftarrow (\mathbf{X}_t - \alpha(t)\widehat{\mathbf{X}}_0)/\sigma(t)$.
    Draw $\boldsymbol{\Xi} \sim \mathcal{N}(0, I)$ and update $\mathbf{X}_s$ using Equation (55).
**end for**
Make the final clean prediction: $(\widehat{\mathbf{X}}_0, \widehat{\boldsymbol{\theta}}_0) \leftarrow F_\phi(\mathbf{X}_{t=0}, \mathbf{Y}_{t=0}, 0, \mathcal{P}; \mathbf{r})$.
Undo coordinate scaling and centering on $\widehat{\mathbf{X}}_0$.
Decode each atom class from $\arg\max_k \widehat{\theta}_{0,i,k}$ and reconstruct bonds to obtain $\mathcal{M}$.
**return** $\mathcal{M}$.

The sampler uses $S+1$ denoiser evaluations, including the final clean prediction. Within each reverse step, Concrete observations are drawn before the coordinate noise, preserving the random-number order. Intermediate atom types remain stochastic even when $\eta_x = 0$.

## G Implementation specification and reproducibility

**Training protocol.** We trained MEGA on the CrossDocked2020 pocket10 split for 15 epochs using a single NVIDIA GeForce RTX 4090 D GPU with 48 GB of memory. The recorded training time was 84,184 s (23.38 h), comprising 187,484 optimizer updates. We used a batch size of 8 and Adam with an initial learning rate of $5 \times 10^{-4}$, $(\beta_1, \beta_2) = (0.95, 0.999)$, and zero weight decay. Model parameters were tracked using an exponential moving average with decay 0.999. The learning rate was controlled by ReduceLROnPlateau with reduction factor 0.6, patience 10, and a minimum learning rate of $10^{-6}$.

**Atom-type vocabulary.** Rather than representing an atom solely by its element, we use a 23-class ligand vocabulary defined by the tuple $(Z, \text{hybridization}, \text{aromaticity})$, where $Z$ is the atomic number and hybridization is assigned by RDKit. This vocabulary distinguishes chemically different local states, including aliphatic $sp^3$ carbon, non-aromatic $sp^2$ carbon, and aromatic $sp^2$ carbon. The atom-type variable is consequently represented on a 23-dimensional categorical simplex and modeled by the Concrete parameter-space diffusion process described in the main text.

Table 4: MEGA configuration.

| Component | Setting |
|---|---|
| Coordinate signal/variance | $\alpha(t) = 1 - 0.009^{1-t}$; $\sigma^2(t) = 0.009^{1-t}$ |
| Type path and temperature | Shared signal $\alpha(t)$; one Concrete draw; $\tau(t) = 0.4$ |
| Vocabulary and coordinate scale | $K = 23$ joint $(Z, \text{hybridization}, \text{aromaticity})$ classes; divide centered coordinates by 2.0 |
| Prediction and weights | Clean coordinates and clean type probabilities; $\lambda_v = 2$ |
| Time and coordinate weighting | $t \sim \mathcal{U}(0,1)$; $T = 1000$; $w_x(t) = \max\{1, (\sqrt{T}/2)\,\mathrm{SNR}_x(\widetilde{t})\}$, with $\widetilde{t}$ in Equation (18) |
| Backbone | UniO2Net (`UniTransformerO2TwoUpdateGeneral`); 9 layers |
| Features and attention | Width 128; 16 heads; 20 radial Gaussian features |
| Graph and pocket features | 32-nearest-neighbor graph; 27-dimensional protein features |
| Time input | Simple one-dimensional time embedding |
| Training | CrossDocked2020 pocket10; 15 epochs; batch size 8; 187,484 optimizer updates; 84,184 s (23.38 h) on one RTX 4090 D (48 GB) |
| Optimizer | Adam; learning rate $5 \times 10^{-4}$; $(\beta_1, \beta_2) = (0.95, 0.999)$; weight decay 0 |
| Learning-rate schedule | ReduceLROnPlateau; factor 0.6; patience 10; minimum learning rate $10^{-6}$ |
| Sampling | $S = 100$ intervals; $\eta_x = 1$; 101 denoiser evaluations |
| EMA | Decay 0.999 |

### G.1 Effect of the number of sampling steps

We evaluated the same trained MEGA model with $S \in \{10, 20, 50, 100, 200\}$ reverse sampling steps. Since the sampler performs a final clean-state prediction, each setting requires $S + 1$ denoiser evaluations. Figure 6 shows that more steps generally improve generation success and synthetic accessibility while substantially reducing median strain energy. The affinity trend is not monotonic: the Vina score improves up to 50 steps and then degrades slightly, while QED decreases as the number of steps grows. We therefore use $S = 100$ as the default, which provides high generation success, low strain, and competitive affinity without the doubled denoising cost of 200 steps.

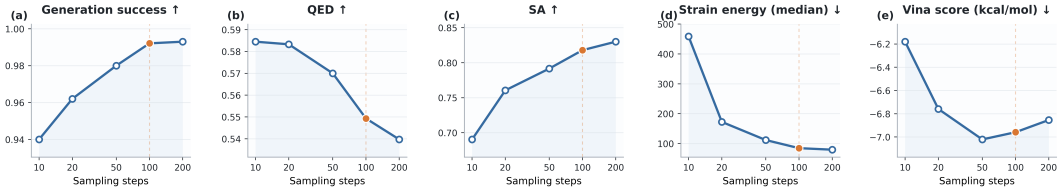



Figure 6: Effect of the number of reverse sampling steps on MEGA. Higher generation success, QED, and SA, together with lower median strain energy and Vina score, indicate better performance. The model is trained with 1000 diffusion steps.

### G.2 GeoRank residual adapter

**Placement and information flow.** An independent adapter follows backbone layers $\ell \in \{3, 6\}$. Each layer first completes its $x \to h$ and $h \to x$ updates, after which the adapter adds two residuals:

$$\begin{aligned} (\widetilde{\mathbf{H}}, \widetilde{\mathbf{X}}) &= F^{(\ell)}_{x2h+h2x}(\mathbf{H}^{(\ell)}, \mathbf{X}^{(\ell)}), \\ \mathbf{H}^{(\ell+1)} &= \widetilde{\mathbf{H}} + \Delta\mathbf{H}_{\text{rank}}, \qquad \mathbf{X}^{(\ell+1)} = \widetilde{\mathbf{X}} + \Delta\mathbf{X}_{\text{rank}}. \end{aligned} \tag{57}$$

The adapter operates only on ligand atoms and leaves protein states unchanged. Its feature residual enters subsequent backbone layers and can therefore affect both coordinate and atom-type predictions. Its coordinate residual is an additional equivariant update; the original $x2h$ and $h2x$ formulas are unchanged.

**Rank graph and messages.** Within each ligand, connect directed edges $(i, j) \in \mathcal{E}_r$ when $|r_i - r_j| \in \{1, 2, 4, 8\}$. The graph stays fixed during denoising, while messages read current features and

geometry. Let $\mathcal{N}_r(j) = \{i : (i,j) \in \mathcal{E}_r\}$, $d_{ij} = \|\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j\|_2$, $\mathbf{u}_{ij} = (\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)/(d_{ij} + \varepsilon)$, and $\delta r_{ij} = (r_i - r_j)/\max(N-1, 1)$, with $N = N_{\mathcal{M}}$ and numerical stabilizer $\varepsilon > 0$. For a shared ligand diffusion time $t$, the edge message is

$$\mathbf{z}_{ij} = \mathrm{MLP}\Big([\mathrm{LN}(\widetilde{\mathbf{h}}_j), \mathrm{LN}(\widetilde{\mathbf{h}}_i), \mathrm{RBF}(d_{ij}), \mathrm{RBF}(|\delta r_{ij}|), \delta r_{ij}, t]\Big), \tag{58}$$

where LN denotes layer normalization and RBF denotes radial basis features. A feature projection $W_h$ and scalar head $\psi_x$ produce raw residuals

$$\mathbf{a}_j^h = W_h\big(\mathrm{Mean}_{i\in\mathcal{N}_r(j)}\, \mathbf{z}_{ij}\big), \qquad \mathbf{a}_j^x = \mathrm{Mean}_{i\in\mathcal{N}_r(j)}\, \mathbf{u}_{ij}\, \psi_x(\mathbf{z}_{ij}). \tag{59}$$

We center the coordinate residual within each ligand, $\mathbf{a}_j^x \leftarrow \mathbf{a}_j^x - N^{-1}\sum_k \mathbf{a}_k^x$, so the rank branch introduces no net translation.

**Magnitude control and time gate.** All scales are ligand-wise RMS magnitudes: $s_h$ for backbone features, $s_{\Delta x}^{\mathrm{backbone}}$ for backbone coordinate updates, and $s_{a^h}, s_{a^x}$ for raw adapter residuals. With $\rho_h = \rho_x = 0.10$, we apply

$$\Delta\mathbf{h}_j^{\mathrm{rank}} = \rho_h \gamma_{\mathrm{geo}}(t) s_h \frac{\mathbf{a}_j^h}{1 + s_{a^h}}, \qquad \Delta\mathbf{x}_j^{\mathrm{rank}} = \rho_x \gamma_{\mathrm{geo}}(t) s_{\Delta x}^{\mathrm{backbone}} \frac{\mathbf{a}_j^x}{1 + s_{a^x}}. \tag{60}$$

Thus each residual's RMS is bounded by $10\%$ of its corresponding backbone scale. The gate enables the adapter at high noise:

$$\gamma_{\mathrm{geo}}(t) = \begin{cases} 0, & t \le 0.65, \\ 3\zeta(t)^2 - 2\zeta(t)^3, & 0.65 < t < 0.90, \\ 1, & t \ge 0.90, \end{cases} \qquad \zeta(t) = \frac{t - 0.65}{0.25}. \tag{61}$$

Both output heads, $W_h$ and $\psi_x$, are zero-initialized, so the inserted adapter initially reproduces the base checkpoint with $\Delta\mathbf{H}_{\mathrm{rank}} = \Delta\mathbf{X}_{\mathrm{rank}} = 0$. Invariant edge inputs yield invariant scalar weights, while the relative directions rotate with the coordinates. Ligand-wise centering and RMS scaling preserve this equivariance.

## H Ablations of forward-process design and atom-type representation

**Variants and evaluation.** We examine the two modeling choices underlying MEGA. **Base** denotes MEGA, using the proposed forward marginals and continuous Concrete atom-type observations. **Linear** and **Cosine** use forward marginals derived from conventional DDPM processes with linear and cosine schedules, respectively. **One-hot** ablates the continuous atom-type representation by using one-hot observations. Tables 5 and 6 report molecular quality and strain energy, and distribution fidelity, respectively.

Table 5: Molecular-quality and strain-energy ablations. Rates are percentages; Vina scores are in kcal/mol. The best and second-best distinct values are bolded and underlined, respectively.

| Methods | PB-Valid | Vina Score (↓) | Vina Min (↓) | Strain Energy (↓) | | | Mol. stable | Validity | Complete | QED | SA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Avg. (↑) | Avg. | Avg. | 25% | 50% | 75% | Avg. (↑) | Avg. (↑) | Avg. (↑) | Avg. (↑) | Avg. (↑) |
| Base | **89.5%** | **-6.959** | **-7.542** | **28.30** | **84.67** | **202.77** | <u>97.0%</u> | **99.3%** | **96.7%** | <u>0.549</u> | **0.818** |
| Linear | 48.6% | -5.622 | -6.606 | 412.34 | 1302.21 | 9566.71 | 76.0% | 85.8% | 72.1% | 0.532 | 0.535 |
| Cosine | 52.1% | -6.180 | -6.808 | 278.06 | 770.24 | 3184.81 | 79.9% | 89.4% | 69.9% | 0.514 | 0.585 |
| One-hot | <u>72.4%</u> | <u>-6.378</u> | <u>-7.240</u> | <u>110.94</u> | <u>278.84</u> | <u>780.62</u> | **97.3%** | <u>94.8%</u> | <u>91.8%</u> | **0.565** | <u>0.694</u> |

**Forward-process design.** Both DDPM variants reduce validity, completeness, and normalized SA, weaken Vina scores, and increase all four distribution JSDs relative to Base. Validity drops from $0.993$ to $0.858$ and $0.894$ for Linear and Cosine, while mean Vina Score changes from $-6.959$ to $-5.622$ and $-6.180$ kcal/mol. Their ring-type JSDs also rise from $0.160$ to $0.462$ and $0.472$. PB-valid falls from $0.895$ to $0.486$ and $0.521$, respectively, showing a substantial loss of physical plausibility. The joint decline in molecular quality and geometric fidelity supports the importance of the forward marginals in molecular diffusion.

Table 6: Distribution-fidelity ablations. Bond, angle, and torsion report weighted JSD; ring-type reports ring-distribution JSD. Lower is better; bold marks the best values.

| Method | Bond JSD ↓ | Angle JSD ↓ | Torsion JSD ↓ | Ring-type JSD ↓ |
|---|---|---|---|---|
| Base | **0.338** | **0.371** | **0.331** | 0.160 |
| Linear | 0.456 | 0.486 | 0.487 | 0.462 |
| Cosine | 0.421 | 0.479 | 0.437 | 0.472 |
| One-hot | 0.434 | 0.436 | 0.433 | **0.157** |

**Conformational strain.** Table 5 further links both modeling choices to intramolecular plausibility. Base has the lowest strain energy at all three quartiles. Median strain increases from $84.67$ to $1302.21$ with Linear and $770.24$ with Cosine; their upper quartiles reach $9566.71$ and $3184.81$, compared with $202.77$ for Base. One-hot also increases median strain to $278.84$ and the upper quartile to $780.62$. Together with PB-valid, these results support the contribution of the proposed forward marginals and continuous type observations to generating physically plausible conformations.

**Continuous atom-type observations.** One-hot achieves lower validity and completeness than Base, with weaker SA and Vina scores. Its bond, angle, and torsion JSDs increase from $0.338/0.371/0.331$ to $0.434/0.436/0.433$. Its PB-valid rate is $0.724$, compared with $0.895$ for Base. These results support retaining graded classwise information in the denoiser's atom-type input when modeling molecular geometry. One-hot retains the highest QED ($0.565$ versus $0.549$) and a slightly lower ring-type JSD ($0.157$ versus $0.160$). Together, the ablations support both forward-process design and continuous type observations as contributors to MEGA's molecular modeling performance.

## I Category-wise bond and angle distributions

Tables 7 and 8 report eight bond-length and eight bond-angle categories for the eight generative methods, extending the five bond categories in Figure 3. Colons denote aromatic bonds and equals signs denote double bonds. The Avg. rows average the eight listed categories equally, whereas Table 2 reports weighted aggregate JSD.

Table 7: Category-wise bond-length JSD (↓). MEGA+GeoRank uses virtual-anchor ordering. Avg. is the unweighted mean over the eight listed categories. The best and second-best distinct values are bolded and underlined, respectively.

| Bond | AR | Pocket2Mol | TargetDiff | DecompDiff | MolPIF | MolCRAFT | MEGA | MEGA + GeoRank |
|---|---|---|---|---|---|---|---|---|
| CC | 0.612 | 0.498 | 0.371 | 0.329 | 0.387 | <u>0.321</u> | 0.407 | **0.257** |
| C:C | 0.464 | 0.415 | 0.263 | 0.238 | **0.198** | 0.323 | 0.216 | <u>0.206</u> |
| CO | 0.490 | 0.453 | 0.421 | **0.348** | 0.410 | <u>0.354</u> | 0.413 | 0.359 |
| CN | 0.475 | 0.424 | 0.362 | 0.337 | 0.342 | **0.300** | 0.386 | <u>0.307</u> |
| C:N | 0.550 | 0.485 | 0.235 | 0.235 | <u>0.207</u> | 0.250 | 0.213 | **0.206** |
| OP | 0.677 | 0.526 | 0.451 | 0.457 | 0.367 | **0.342** | 0.348 | <u>0.347</u> |
| C=O | 0.555 | 0.510 | 0.461 | 0.390 | 0.335 | 0.334 | <u>0.327</u> | **0.293** |
| O=P | 0.626 | 0.581 | 0.507 | 0.452 | 0.397 | **0.358** | 0.389 | <u>0.382</u> |
| Avg. | 0.556 | 0.486 | 0.384 | 0.348 | 0.330 | <u>0.323</u> | 0.337 | **0.295** |

Table 8: Category-wise bond-angle JSD (↓). MEGA+GeoRank uses virtual-anchor ordering. Avg. is the unweighted mean over the eight listed categories. The best and second-best distinct values are bolded and underlined, respectively.

| Angle | AR | Pocket2Mol | TargetDiff | DecompDiff | MolPIF | MolCRAFT | MEGA | MEGA + GeoRank |
|---|---|---|---|---|---|---|---|---|
| CCC | 0.372 | 0.380 | 0.345 | 0.338 | 0.372 | <u>0.301</u> | 0.395 | **0.261** |
| C:C:C | 0.569 | 0.480 | 0.283 | 0.255 | 0.208 | **0.174** | <u>0.188</u> | 0.189 |
| CCO | 0.477 | 0.475 | 0.440 | 0.389 | 0.435 | **0.347** | 0.394 | <u>0.354</u> |
| C:C:N | 0.536 | 0.506 | 0.454 | <u>0.430</u> | 0.432 | 0.451 | 0.451 | **0.421** |
| CCN | 0.474 | 0.443 | 0.437 | 0.417 | 0.418 | **0.380** | 0.432 | <u>0.383</u> |
| CNC | 0.549 | <u>0.498</u> | 0.521 | 0.505 | 0.506 | 0.504 | 0.516 | **0.486** |
| COC | 0.498 | 0.495 | 0.502 | 0.489 | 0.503 | **0.466** | 0.553 | <u>0.487</u> |
| C:N:C | 0.619 | 0.580 | 0.495 | 0.463 | <u>0.440</u> | 0.453 | 0.447 | **0.424** |
| Avg. | 0.512 | 0.482 | 0.435 | 0.411 | 0.414 | <u>0.384</u> | 0.422 | **0.376** |

MEGA+GeoRank lowers JSD relative to MEGA in all eight bond-length categories and seven of the eight bond-angle categories; the aromatic C:C:C angle changes only slightly (0.188 to 0.189). MEGA+GeoRank achieves the best JSD in three bond-length categories and four bond-angle categories, and the lowest unweighted averages of 0.295 and 0.376, respectively. These results complement the weighted aggregate comparison by showing that the gains extend across multiple local chemical environments.

## J Comparison of GeoRank orderings

We compare MEGA with MEGA augmented by virtual-anchor or spectral GeoRank (Section 3.3.1). The virtual-anchor variant is denoted MEGA+GeoRank in the main benchmark tables. Table 9 reports molecular performance, strain energy, and weighted geometric JSD; Table 10 provides the category-wise results.

Table 9: Molecular quality, strain energy, and geometric fidelity of GeoRank variants. Rates are percentages; Vina scores and strain energy are in kcal/mol. Distribution metrics are weighted JSD. <u>Bold and underlined entries indicate the best and second-best distinct available values.</u>

*Molecular quality*

| Methods | Generated molecules | PB-Valid Avg. (↑) | Vina Score (↓) Avg. | Vina Min (↓) Avg. | Vina Dock (↓) Avg. | Complete Avg. (↑) | Validity Avg. (↑) | QED Avg. (↑) | SA Avg. (↑) |
|---|---|---|---|---|---|---|---|---|---|
| MEGA | 9,928 | **89.46%** | **-6.9587** | **-7.5422** | **-8.2083** | 96.71% | **99.28%** | 0.5493 | **0.8179** |
| MEGA + Virtual-anchor GeoRank | 9,889 | 88.50% | -6.5681 | -7.1755 | -7.7802 | <u>97.06%</u> | <u>98.89%</u> | <u>0.5654</u> | 0.7229 |
| MEGA + Spectral GeoRank | 9,881 | <u>89.40%</u> | <u>-6.5938</u> | <u>-7.2531</u> | <u>-7.9631</u> | **98.52%** | 98.81% | **0.5721** | <u>0.7412</u> |

*Molecular stability, strain energy, and geometric fidelity*

| Methods | Mol. stable (↑) | Strain Energy (↓) | | | Distribution JSD (↓) | | |
|---|---|---|---|---|---|---|---|
| | | 25% | 50% | 75% | Bond | Angle | Torsion |
| MEGA | <u>97.0%</u> | **28.30** | **84.67** | **202.77** | 0.3384 | <u>0.3709</u> | 0.3307 |
| MEGA + Virtual-anchor GeoRank | 96.97% | 63.19 | 154.75 | 362.54 | **0.2701** | **0.3150** | <u>0.3280</u> |
| MEGA + Spectral GeoRank | **98.72%** | <u>45.51</u> | <u>128.18</u> | <u>311.43</u> | <u>0.2762</u> | **0.3150** | **0.3253** |

**Geometric fitting and molecular quality.** Both orderings improve bond-length, bond-angle, and torsion JSD over MEGA. MEGA + Virtual-anchor GeoRank gives the lowest bond-length JSD (0.2701), while MEGA + Spectral GeoRank gives the lowest torsion JSD (0.3253). Both attain

an angle JSD of $0.3150$ at the reported precision. MEGA + Spectral GeoRank achieves molecular stability of $98.72\%$, completeness of $98.52\%$, and QED of $0.5721$. Compared with MEGA + Virtual-anchor GeoRank, it reduces median strain from $154.75$ to $128.18$ and improves Vina Score and Vina Min from $-6.5681/-7.1755$ to $-6.5938/-7.2531$. MEGA itself retains the best reported Vina scores, SA, and strain energy.

Table 10: Category-wise bond-length and bond-angle JSD with weighted means for GeoRank variants. Lower is better; bold and underlined entries mark the best and second-best distinct values. Colons denote aromatic bonds and equals signs denote double bonds.

| Category | MEGA | MEGA + Virtual-anchor GeoRank | MEGA + Spectral GeoRank |
|---|---|---|---|
| *Bond-length JSD* | | | |
| CC | 0.4068 | **0.2573** | <u>0.2664</u> |
| C:C | 0.2165 | **0.2063** | <u>0.2145</u> |
| CO | 0.4126 | <u>0.3594</u> | **0.3579** |
| CN | 0.3862 | **0.3072** | <u>0.3158</u> |
| C:N | 0.2133 | **0.2065** | <u>0.2092</u> |
| OP | 0.3479 | <u>0.3471</u> | **0.3454** |
| C=O | 0.3272 | **0.2931** | <u>0.3121</u> |
| O=P | 0.3889 | <u>0.3822</u> | **0.3677** |
| Weighted mean | 0.3384 | **0.2701** | <u>0.2762</u> |
| *Bond-angle JSD* | | | |
| CCC | 0.3953 | **0.2609** | <u>0.2689</u> |
| C:C:C | <u>0.1884</u> | 0.1894 | **0.1810** |
| CCO | 0.3943 | <u>0.3541</u> | **0.3435** |
| C:C:N | 0.4507 | **0.4210** | <u>0.4257</u> |
| CCN | 0.4320 | **0.3826** | <u>0.3863</u> |
| CNC | 0.5163 | **0.4863** | <u>0.4973</u> |
| COC | 0.5532 | <u>0.4867</u> | **0.4746** |
| C:N:C | 0.4470 | **0.4239** | <u>0.4373</u> |
| Weighted mean | <u>0.3709</u> | **0.3150** | **0.3150** |

**Category-level differences.** MEGA + Virtual-anchor GeoRank outperforms MEGA + Spectral GeoRank in five of the eight bond-length categories and five of the eight bond-angle categories. Spectral ordering is better for CO, OP, and O=P bond lengths and for C:C:C, CCO, and COC bond angles, indicating complementary effects across local chemical environments.

**Relation to the ordering design.** These differences align with the motivations in Section 3.3.1. Virtual-anchor ordering encodes the global organization of the ligand relative to the protein pocket and achieves better bond-length distribution fitting. Spectral ordering captures geometric connectivity within the ligand, consistent with its lower strain across all three quantiles and lower torsion JSD. It also achieves more favorable Vina Score and Vina Min values. The two variants thus show complementary strengths in local geometry and conformational quality.