ANCHORSET: STRUCTURE-ANCHORED SET-VALUED CANONICALIZATION FOR NON-EQUIVARIANT 3D MOLECULAR GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Equivariant diffusion models have achieved strong performance in 3D molecular generation by encoding Euclidean symmetries directly into the denoising architecture. Their specialized geometric operations, however, can increase computational cost and constrain the scalability and flexibility of the generative backbone. Existing canonicalization can support non-equivariant generation, but it can become unstable when molecular geometry does not define a unique orientation, replacing rotational redundancy with an arbitrary or perturbation-sensitive frame. We introduce **AnchorSet**, a structure-anchored, set-valued canonicalization framework that combines chemically informed pose estimation with explicit preservation of unresolved frame ambiguity. AnchorSet estimates molecular orientation from structure-conditioned anchors that combine a chemical-scaffold prior with learnable atom-type weights, biasing pose estimation toward stable molecular cores rather than flexible peripheral atoms. When the geometry does not identify a unique orientation, it preserves a compact set of symmetry-equivalent frames related by proper signed permutations, enabling a standard non-equivariant generator to operate in canonical coordinates. Across QM9 and GEOM-Drugs, AnchorSet outperforms prior non-equivariant approaches and matches or surpasses equivariant diffusion models in sample quality. Its stability under global rotations and non-anchor perturbations demonstrates that structure-aware set-valued canonicalization is a practical, scalable alternative to built-in equivariance.

1 INTRODUCTION

3D molecular generation aims to produce chemically valid molecules with plausible geometries for downstream modeling and molecular design. Because molecular coordinates are equivalent under global rotations, equivariant generative models explicitly preserve $SO(3)$ symmetry and provide strong geometric modeling capabilities. EGNN, EDM, and GeoLDM, for example, employ equivariant message passing or diffusion to model molecular conformations and structures (Satorras et al., 2021; Hoogeboom et al., 2022; Xu et al., 2023). However, equivariant computation propagates and constrains geometric quantities throughout the network, with computational and memory requirements that typically increase with molecular size, edge count, and network depth. Equivariant architectures may also be coupled to particular message-passing mechanisms, coordinate-update rules, and noise parameterizations, complicating scaling and component replacement. Reducing the geometric burden on the generator while retaining the ability to handle rotational equivalence is therefore a practical challenge in 3D molecular generation.

Explicit canonicalization moves the handling of rotational symmetry from the generator to its input. Each molecule is first mapped to a coordinate frame in which a non-equivariant backbone can operate with a reduced geometric burden. Analytic constructions use PCA, inertia tensors, or Procrustes/Kabsch alignment (Kabsch, 1976); they are computationally efficient and interpretable, but eigenvectors become sensitive to perturbations when eigengaps are small (Yu et al., 2015), and principal-axis frames retain sign and axis-permutation ambiguities (Puny et al., 2022). Learned canonicalization provides a general architecture-agnostic alternative; RADM instantiates this idea for molecular generation by coupling learned rotational alignment to an autoencoder and latent diffusion model (Ding & Hofmann, 2025). Theory and unified analyses further show that symmetric

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.

objects need not admit a unique continuous canonical section (Dym et al., 2024; Ma et al., 2024). A single deterministic frame is consequently not always appropriate: canonicalization may need to retain equivalent transformations or use a set-valued representation.

A learned single-frame pipeline places downstream reconstruction and generation under one estimated rotation, as in RADM. Whether hydrogens, flexible terminal groups, or conformation-dependent side chains make that estimate unstable is not established by RADM; here it is a design hypothesis tested directly through subset-specific perturbations. Analytic PCA removes the learned rotation but does not remove identifiability failures: linear, approximately symmetric, or very small structures can have small eigengaps, making axis ordering sensitive to perturbations. More generally, forcing a unique frame can turn intrinsic equivalence into an arbitrary target when a continuous canonical section does not exist. Recent canonical and quotient-space generative models formalize alternative treatments of this redundancy and demonstrate them in molecular generation. The central challenge is therefore to use chemically meaningful structural evidence when it is informative while explicitly retaining equivalent frames when the geometry is not identifiable.

We propose **AnchorSet**, a structure-anchored, set-valued canonicalization method for canonical-coordinate latent diffusion. AnchorSet extracts a stable pose when geometry is informative and retains equivalent candidates in symmetric or degenerate cases. The representation thus captures both the principal directions of the molecular structure and unavoidable frame ambiguity. Three complementary design choices support this approach. First, a chemical scaffold prior and atom-type weights define structure-conditioned anchors, allowing the chemical core to drive pose estimation. Second, a degeneracy-aware candidate mapping retains equivalent frame choices where a unique orientation cannot be identified, from which an operational frame is selected. Third, a common evaluation protocol measures robustness to global rotations and conformational perturbations, together with the reliability of downstream latent generation. AnchorSet thereby combines structural anchoring, ambiguity handling, and consistency analysis within one canonicalization-to-diffusion pipeline, while distinguishing candidate-representation evidence from generation performance. On QM9, these evaluations show improved rotational consistency and perturbation stability alongside stronger non-equivariant latent generation under matched configurations. Together, these properties give researchers and drug-discovery practitioners a practical, auditable way to build non-equivariant 3D molecular generators that preserve chemically meaningful structure while reducing sensitivity to arbitrary orientation.

2 Related Work

2.1 Single-frame canonicalization and learned alignment

Explicit canonicalization selects a coordinate reference before generation, treating rotational equivalence as a representation-preprocessing problem. Analytic alignment based on PCA, inertia tensors, or Procrustes/Kabsch transformations is simple and auditable, but depends on stable principal directions, identifiable eigenvector signs, and consistent reference conformations. Kaba et al. develop learned canonicalization as a general architecture-agnostic framework rather than a molecular generator (Kaba et al., 2023). RADM provides a molecular instance, combining learned rotational alignment with an autoencoder and latent diffusion for non-equivariant generation.

Within the broader landscape of 3D generation, EDM uses equivariant networks for molecular diffusion, GeoDiff extends diffusion to molecular conformation generation (Xu et al., 2022), and GeoLDM explores compression and acceleration through geometric latent diffusion. Wang et al. demonstrate that non-equivariant models can remain competitive for conformation generation conditioned on a given 2D molecular structure (Wang et al., 2024), while Joshi et al. scale non-equivariant Transformer backbones through rotation augmentation (Joshi et al., 2025). Yao et al. incorporate sample-dependent frames and deterministic canonicalization into variable-size generation (Yao et al., 2026). Inference-time DIST provides a complementary corrective sampling procedure for diffusion backbones, including RADM (Wan et al., 2026). Design studies of equivariant molecular diffusion provide complementary comparisons of architectural choices (Le et al., 2024). At the molecular-graph level, motif generation, graph SDEs, and discrete graph diffusion address permutation-aware topology generation rather than 3D frame ambiguity (Jin et al., 2020; Jo et al., 2022; Vignac et al., 2023). Analytic and learned alignment emphasize geometric transparency and data adaptation, respectively, but both commonly assign one deterministic frame per molecule. An-

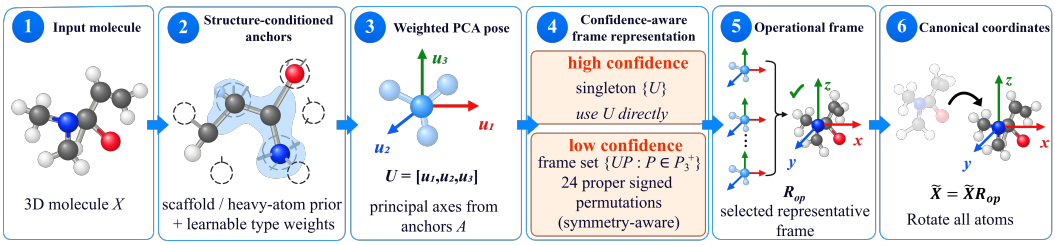


Figure 1: Structure-anchored set-valued canonicalization in AnchorSet.

alytic frames can become unstable near degeneracies, whereas the sensitivity of learned estimates to hydrogens, flexible side chains, or conformational noise remains an empirical question. AnchorSet retains explicit alignment's interpretability and auditability while treating structural evidence and frame identifiability jointly.

2.2 Symmetry handling and set-valued representations

Frame averaging replaces a unique canonical representative with a finite equivariant set and averages predictions over that set (Puny et al., 2022). Subsequent theory proves that continuous canonical frames do not exist in general and motivates weighted frames that preserve continuity (Dym et al., 2024); a unified canonicalization perspective clarifies the corresponding expressivity and stability trade-offs (Ma et al., 2024). Probabilistic symmetrization and probabilistic symmetry breaking instead learn or sample distributions over transformations (Kim et al., 2023; Lawrence et al., 2025). Symmetry-aware learned canonicalization brings related ideas into generative modeling by learning an orbit representative for a non-equivariant generator (Sareen et al., 2025).

For molecular generation, Canonical Diffusion maps molecular graphs to representatives under permutation and rigid-motion symmetries before fitting an unconstrained generator (Zhou et al., 2026); Quotient-Space Diffusion instead defines the stochastic process directly on the symmetry quotient (Xu et al., 2026). InertialAR uses an inertial canonical frame and atom reordering for autoregressive 3D generation (Li et al., 2025), while frame-based diffusion compares global, local, and pre-canonicalized invariant frames (Guo et al., 2025). These approaches distinguish single-frame, set-valued, stochastic, and quotient-space treatments. AnchorSet retains the 24 proper signed-permutation candidates of a PCA frame and selects one operational representative for the downstream generator. The finite set handles discrete sign and axis-permutation ambiguity but not continuous rotations within repeated-eigenvalue subspaces.

3 Method

3.1 Method Overview

AnchorSet targets non-equivariant 3D molecular generation in canonical coordinates. Figure 1 illustrates AnchorSet's structure-anchored, set-valued canonicalization for non-equivariant 3D molecular generation. Its design addresses two coupled questions: which atoms should determine a molecule's global orientation, and how should orientation ambiguity be represented when geometric evidence is insufficient? The model estimates principal axes from learnable anchors defined by a structural prior, retains discrete candidates for low-confidence axes, and encodes the selected coordinates as per-atom latents for a non-equivariant diffusion model. Structural anchoring determines the evidence used for pose estimation, candidate representations handle discrete ambiguity, and the autoencoder (AE) and DiT respectively perform representation reconstruction and distribution modeling.

Our central hypothesis is that, when the canonical frame is determined primarily by structurally stable anchors and the discrete equivalence of the axes is retained at low confidence, both global input rotations and conformational perturbations of non-anchor atoms should have a smaller effect on the operational frame. Under the same generator and training budget, this representational improvement should further appear as a smaller rotation-induced likelihood drift. Accordingly, support for the

proposed mechanism requires consistent improvements in frame consistency, perturbation stability, and rotation-conditioned likelihood rather than an isolated gain in a generation metric.

Given atomic attributes H, molecular graph G, and row-vector coordinates $X \in \mathbb{R}^{n\times 3}$, we remove the valid-atom centroid to obtain $\bar{X}$. For any $Q \in SO(3)$, $\bar{X}$ and $\bar{X}Q$ represent the same conformation. The canonicalizer outputs a candidate rotation set $\mathcal{R}_\alpha(\bar{X}, G, H)$, selects the rotation used for encoding, R_{op}, and forms $\tilde{X} = \bar{X}R_{\mathrm{op}}$. Its four modules are structural anchoring, candidate coordinate frames, canonical autoencoding, and non-equivariant latent diffusion.

$$(\bar{X}, G, H) \xrightarrow{C_\alpha} (\mathcal{R}_\alpha, R_{\mathrm{op}}, \tilde{X}) \xrightarrow{E_{\eta_0}} (\mu_x, \mu_h) \xrightarrow{+\sigma_{\mathrm{AE}}\xi} z_0.$$

This expression denotes the mapping from training data to the latent representation, where C_α comprises the first two modules. In the first stage, $D_\psi(z_0)$ reconstructs $(\tilde{X}, H)$; in the second stage, the representation mapping is fixed and a denoiser ϵ_θ is trained on z_0. Unconditional generation starts from Gaussian noise, runs the reverse diffusion process, and finally invokes D_ψ for decoding; it does not require a given molecule as input.

The first two modules define AnchorSet's structure-anchored candidate representation; the latter two follow RADM's AE/LDM formulation. The set represents discrete orientation ambiguity, but each encoding receives one selected $(\tilde{X}, H)$ and candidate latents are not aggregated or marginalized.

3.2 Structure-Anchored Pose Estimation

Local displacements of hydrogen atoms and peripheral substituents can alter whole-molecule geometric statistics without changing the principal chemical core. We therefore restrict the support of pose estimation to the rings and linker fragments in the Bemis–Murcko scaffold (Bemis & Murcko, 1996), allowing core geometry to determine the coordinate directions. Let $m_i \in \{0, 1\}$ denote the final valid-anchor mask. If fewer than three scaffold atoms are available, we fall back to the heavy-atom mask; if fewer than three heavy atoms remain, we fall back to the all-atom mask. This fallback guarantees an available input for direction estimation but does not guarantee an identifiable frame; stability still depends on the actual coordinates and the subsequent candidate rule.

Within the support set, the relative contributions of different atom types can still affect the principal axes. For each type t_i, we introduce a shared scalar α_{t_i} and use a strictly positive weight to modulate its contribution. The anchor centroid and covariance are

$$w_i = m_i\big(\delta_w + \mathrm{softplus}(\alpha_{t_i})\big), \qquad \mu_a = \frac{\sum_i w_i \bar{x}_i}{\sum_i w_i}, \qquad S = \frac{\sum_i w_i (\bar{x}_i - \mu_a)^\top (\bar{x}_i - \mu_a)}{\sum_i w_i}.$$

Here $\delta_w > 0$ is the lower bound on the weight. The mask determines which atoms participate in estimation, whereas the type parameters determine their relative contributions within the support set; the weights are optimized through the AE reconstruction objective and shared across molecules. Because scaffold annotations and atom types are invariant to global rotation, this parameterization preserves rotational invariance of the weights. The structural mask defines the support of pose estimation, while type weights adjust the relative strength of geometric evidence within that support; the candidate module then handles frame ambiguity induced by the weighted geometry.

We eigendecompose S and order its eigenvectors by decreasing eigenvalue to obtain the principal-axis matrix U. The axis signs are determined by the type-weighted third moments described in Section 3.3, and the third axis is corrected to enforce $\det(U) = 1$. The anchor centroid μ_a is used only to estimate the directions. Applying the rotation to globally centered all-atom coordinates $\bar{X}$ retains hydrogens and non-anchor atoms in the reconstruction and generation targets.

This construction converts the structural prior into an explicit coordinate transformation. Under an input rotation, the weights remain unchanged and $S' = Q^\top S Q$; if the eigenvalues are distinct and the signs are identifiable, then $U' = Q^\top U$, and therefore $\bar{X}QU' = \bar{X}U$. The next module handles the remaining cases in which the discrete principal-axis choice cannot be determined stably.

3.3 Candidate Coordinate Frames and Representative Selection

Structural anchors constrain the evidence used for pose estimation but do not guarantee a unique PCA orientation. In particular, the signs of eigenvectors must still be determined. Third moments

can determine these signs, but the decision becomes sensitive to numerical perturbations when positive and negative contributions nearly cancel. Let $a_{ij} = (\bar{x}_i - \mu_a)u_j$ be the projection of atom i onto PCA axis j, and let $\kappa_i \in \{1, \dots, T_a\}$ be the integer code associated with a fixed atom-category order. The sign confidence of the first two axes is

$$c_j = \frac{\left|\sum_i w_i \kappa_i a_{ij}^3\right|}{\max\left(\sum_i w_i \kappa_i |a_{ij}^3|, \varepsilon_c\right)}, \qquad c = \min(c_1, c_2).$$

Here $\varepsilon_c > 0$ prevents a zero denominator. The first two axis signs and right-handedness determine the complete frame; confidence below τ_c activates the candidate representation:

$$\mathcal{R}_\alpha = \begin{cases} \{U\}, & c \geq \tau_c, \\ \{UP : P \in \mathcal{P}_3^+\}, & c < \tau_c, \end{cases}$$

$$\mathcal{P}_3^+ = \{P : P \text{ is a signed-permutation matrix, } \det(P) = 1\}.$$

$\mathcal{P}_3^+$ contains 24 elements, covering orientation-preserving sign and permutation choices of the three axes, consistent with principal-axis frame-set constructions used in frame averaging (Puny et al., 2022). If the PCA bases of the original and rotated inputs differ only by these transformations after compensating for the input rotation, the complete candidate sets coincide; one-to-one correspondence between candidate indices is not required. Every candidate lies in $SO(3)$ and preserves input chirality. This representation handles discrete principal-axis ambiguity but excludes continuous rotations within repeated-eigenvalue subspaces.

Candidate sampling during training and representative selection during evaluation. For high-confidence inputs, we set $R_{\text{op}} = U$ directly. For low-confidence inputs, the main model samples $R_{\text{op}} \sim \text{Uniform}(\mathcal{R}_\alpha)$ during both AE and LDM training, exposing the two stages to different discrete coordinate choices. Unlike the probabilistic symmetrization approaches discussed above, AnchorSet uses one uniformly sampled candidate per forward pass and does not marginalize candidate latents. At validation and test time, a fixed fingerprint rule selects a reproducible representative; the AE parameters remain frozen in the second stage, but candidate sampling during training is retained. Specifically, let

$$s = \max\left\{\sqrt{\frac{\sum_i w_i \|\bar{x}_i - \mu_a\|_2^2}{\sum_i w_i}}, \varepsilon_s\right\}, \qquad \|v\|_2 = 1, \qquad d_{ik} = \frac{(\bar{x}_i - \mu_a)R_k v}{s},$$

where $\varepsilon_s > 0$ is a scale floor and v is a fixed projection direction. Using projection coefficient β and clipping amplitude b, we compute the fingerprint of each candidate as

$$f_k = \sum_i w_i \kappa_i \exp(\text{clip}(\beta d_{ik}, -b, b)), \qquad R_{\text{op}} = R_{\arg\max_k f_k}.$$

The normalization compares candidates using the relative geometric distribution of the anchors; the nonlinear projection moment distinguishes different sign and axis-permutation choices, and summation over atoms avoids dependence on node ordering. Category codes are fixed during both training and evaluation. When the candidate sets coincide and the maximum score is unique, scoring selects the same geometric representative. Ties or continuous degeneracies can still make the representative unstable. We therefore evaluate candidate-set and single-frame consistency separately.

3.4 Canonical Autoencoder

After canonicalization, the AE converts atomic attributes and coordinates into a continuous per-atom representation, with reconstruction providing the learning signal for anchor weights. Three coordinate channels per atom and F_z attribute dimensions preserve geometry and node correspondence.

Encoder. The EGNN encoder E_{η_0} is kept fixed after initialization. It outputs latent means $\mu_x \in \mathbb{R}^{n\times 3}$ and $\mu_h \in \mathbb{R}^{n\times F_z}$, to which Gaussian noise with amplitude σ_{AE} is added. Noise is applied only to valid nodes, and the coordinate part is centered:

$$(\mu_x, \mu_h) = E_{\eta_0}(\tilde{X}, H), \qquad z_0 = [\mu_x, \mu_h] + \sigma_{\text{AE}}\xi \in \mathbb{R}^{n\times(3+F_z)}.$$

Decoder. The representation preserves the number of atoms and node correspondence rather than pooling the molecule into a single vector. The non-equivariant GNN decoder D_ψ treats the coordinate and attribute channels of z_0 as ordinary node features and propagates information through edge-message MLPs, aggregation, and residual node updates. It predicts 3D coordinates, atom-category logits, and enabled charge channels, then re-centers output coordinates over valid nodes:

$$(\widehat{\widehat{X}}, \widehat{H}) = D_\psi(z_0).$$

The purpose of the non-equivariant decoder is to make the reconstruction error sensitive to the canonical coordinate direction, thereby linking orientation selection to the reconstruction task. Fixing the encoder concentrates first-stage learning on the anchor type weights and decoder. Although the encoder parameters are not updated, input gradients can still propagate to the canonicalizer. Both the reconstruction target and the output use the selected canonical coordinate frame.

3.5 Non-equivariant Latent Diffusion

Once the AE has established a continuous per-atom representation, a Diffusion Transformer (DiT) learns its distribution (Peebles & Xie, 2023). At the end of the first stage, all canonicalizer and AE parameters are frozen, while training candidates continue to be sampled according to Section 3.3. An input linear layer maps each atom's latent to a token; the time step is encoded by a sinusoidal embedding and an MLP, which then modulates self-attention and MLP sublayers in each block through adaLN-Zero. The output layer returns a latent-shaped noise prediction:

$$\epsilon_\theta(z_t, t) \in \mathbb{R}^{n\times(3+F_z)}.$$

The model uses a valid-edge mask for padding and inter-atomic attention and recenters the predicted coordinate noise. Because atomic coordinates are already contained in the latent, no atom-index positional encoding is added. Rotation handling is confined to the canonicalization layer, allowing the denoiser to use ordinary tokens and attention. For unconditional sampling, we draw the atom count from $p(n)$, run T reverse-diffusion steps from matching-dimensional Gaussian noise z_T, and invoke the frozen decoder:

$$n \sim p(n), \qquad z_T \sim \mathcal{N}_0(0, I) \xrightarrow[\text{reverse diffusion}]{\epsilon_\theta} \hat{z}_0 \xrightarrow{D_\psi} (\widehat{\widehat{X}}, \widehat{H}).$$

Here $\mathcal{N}_0$ combines zero-centroid coordinate noise with ordinary Gaussian feature noise. Sampling does not require a source-molecule scaffold and does not recompute the frame at each denoising step. The output is interpreted in canonical coordinates.

3.6 Two-stage Optimization

In the first stage, we jointly optimize the anchor type parameters α and decoder parameters ψ, while keeping the encoder parameters η_0 fixed. The objective is

$$\mathcal{L}_{\mathrm{AE}}(\alpha, \psi) = \mathbb{E}_{(X,G,H),R_{\mathrm{op}},\xi}\left[\ell_{\mathrm{coord}}(\widehat{\widehat{X}}, \tilde{X}) + \ell_{\mathrm{type}}(\widehat{H}, H) + \ell_{\mathrm{charge}}(\widehat{H}, H)\right].$$

The coordinate term is a masked mean-squared error over valid nodes; the type term is cross-entropy; when charge channels are enabled, a masked mean-squared error is added. Appendix A gives the exact reduction conventions. Low-confidence samples use uniform one-candidate sampling per forward pass. Anchor weights are updated through the reconstruction path of the selected PCA frame; signs, trigger conditions, and candidate indices are discrete decisions. This step needs only a locally differentiable eigendecomposition; candidate selection need not be differentiable.

In the second stage, the first-stage model is fixed. Following the ϵ-parameterized denoising diffusion objective and RADM's latent formulation (Ho et al., 2020; Ding & Hofmann, 2025), we construct z_0 with the same σ_{AE} and apply the forward diffusion process:

$$z_t = \alpha_t z_0 + \sigma_t \epsilon, \qquad t \sim \mathrm{Uniform}\{0, \ldots, T\}, \qquad \mathcal{L}_{\mathrm{LDM}}(\theta) = \frac{1}{2}\mathbb{E}\left[\frac{\|\epsilon - \epsilon_\theta(z_t, t)\|_2^2}{(3+F_z)N_{\max}}\right].$$

Here $N_{\max}$ is the padded node dimension used for batching; the error is masked by valid nodes, and coordinate noise is restricted to the zero-centroid subspace. This is an ϵ-parameterization; we display only the MSE term that depends on θ, while the terminal-prior KL retained in the implementation is fixed. The two stages respectively determine the representation mapping and fit the latent distribution. When comparing generation performance within a dataset, we keep the first-stage parameters fixed and use the same diffusion architecture and noise schedule.

Table 1: QM9 AE canonicalization diagnostics under random rotations. Flips indicate frame errors above 90°; perturbation columns report drift at $\sigma = 0.1$ (lower is better).

AE variant	rot. p95 (°)	flips	H drift (°)	side-chain drift (°)
All-Proc	178.337	969.7	3.034	0.677
Heavy-Proc	177.061	918.0	7.14×10^{-6}	1.467
Murcko-Single	0.000370	3.0	0	0
Murcko-Set	**0.000363**	**1.3**	0	0
Learnable-Set	0.000406	1.7	0	0

3.7 Model Implementation

Model framework. Both datasets use the same two-stage pipeline: a structure-anchored canonicalizer maps molecules to canonical coordinates, a one-layer EGNN encoder and non-equivariant GNN decoder learn per-atom latents, and a DiT-B denoiser models the frozen latent distribution. The encoder is fixed during AE training, after which the canonicalizer and AE are frozen for LDM training. DiT-B uses 12 blocks, hidden width 768, 12 attention heads, and MLP ratio 4; both datasets use $\sigma_{\mathrm{AE}} = 0.00316$ and the polynomial_2 schedule with noise precision 10^{-5}.

QM9 configuration. QM9 uses the Murcko prior and fallback procedure in Section 3.2, with $F_z = 1$ (four channels per atom) and $T = 1000$ diffusion steps. Decoder depth, anchor-weight initialization, candidate threshold, and training schedules are given in Appendix A.

GEOM-Drugs configuration. GEOM-Drugs uses a heavy-atom prior with learnable atom-type weights, $F_z = 2$ (five channels per atom), and $T = 1000$ diffusion steps. Canonicalizer variants retain the same AE and LDM capacities; decoder and optimization details are reported in Appendix A.

4 Experiments

Evaluation protocol. We evaluate AnchorSet on QM9 (Ramakrishnan et al., 2014) with five AE canonicalizers: all-atom Procrustes (All-Proc), heavy-atom Procrustes (Heavy-Proc), fixed Murcko PCA (Murcko-Single), Murcko PCA with candidate selection (Murcko-Set), and the learnable-weight candidate model (Learnable-Set). Frame, perturbation, and paired LDM tests use 1,280 molecules; the LDM test applies three rotations per molecule with matched diffusion noise, and generation uses 10,000 samples per configuration. GEOM-Drugs (Axelrod & Gómez-Bombarelli, 2022) uses a separate heavy-atom protocol. Comparisons include ENF (Garcia Satorras et al., 2021), G-SchNet (Gebauer et al., 2019), EDM (Hoogeboom et al., 2022), GeoLDM (Xu et al., 2023), RADM variants (Ding & Hofmann, 2025), EquiFM (Song et al., 2023), UniGEM (Feng et al., 2025), GeoBFN (Song et al., 2024), GOAT (Hong et al., 2025), and MOLTD (Zhang et al., 2025).

4.1 Canonical-frame consistency

Table 1 separates the alignment choices. All-Proc and Heavy-Proc retain 177–178° p95 errors and hundreds of flips; removing hydrogens suppresses hydrogen drift but leaves Heavy-Proc sensitive to flexible heavy atoms. Scaffold PCA reduces the p95 below 4.1×10^{-4} degrees and leaves at most three flips. Murcko-Set further lowers the single-frame mean and flip count by 54.5% and 55.6%, respectively, with a candidate rate of about 7.3%; Learnable-Set reduces complete candidate-set mean and maximum errors by 86.3% and 88.9% relative to Murcko-Set (Figure S2).

The perturbation columns provide the same support-isolation result: Murcko-prior frames remain unchanged under the tested displacements, while Heavy-Proc transfers sensitivity from hydrogen to flexible heavy atoms. The perturbation subsets and fallback boundary are detailed in Appendix C.

Learnable-Set also gives the lowest test Kabsch RMSD (0.00168 versus 0.00169–0.00175), with RDKit validity and molecule stability remaining within 97.58–97.89% and 96.88–97.19%, respectively. Appendix Tables 5–8 provide the complete reconstruction and gate results.

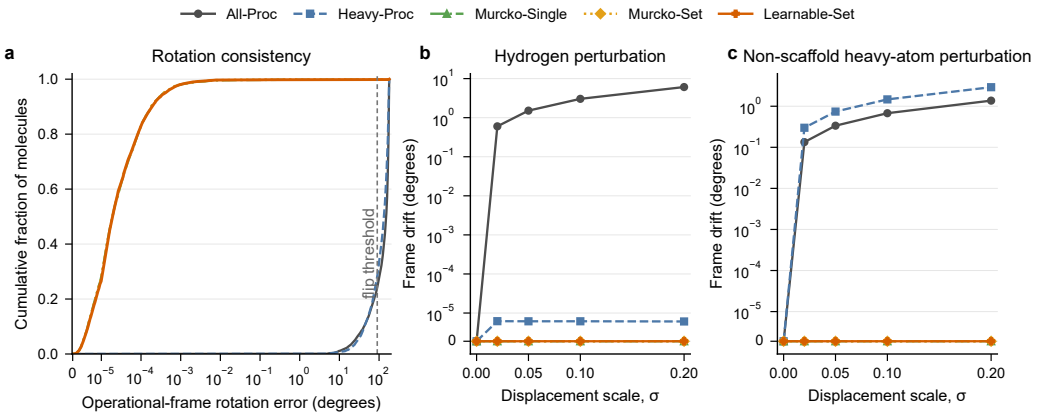


Figure 2: Rotation consistency and perturbation robustness on QM9. **a,** Cumulative distribution of operational-frame rotation error over 1,280 molecules; the dashed line marks 90°. **b,c,** Mean frame drift under hydrogen and non-scaffold heavy-atom perturbations; bands show one SD.

Table 2: QM9 and GEOM-Drugs generation comparison. Data is the dataset reference. All metrics are percentages and higher is better.

			QM9			GEOM-Drugs	
Method	Atom Sta.	Mol. Sta.	Valid	$V \times U$	$V \times U \times N$	Atom Sta.	Valid
Data	99.0	95.2	97.7	97.7		86.5	99.9
ENF	85.0	4.9	40.2	39.4			
G-SchNet	95.7	68.1	85.5	80.3			
EDM	98.7	82.0	91.9	90.7	59.6	81.3	92.6
GeoLDM	98.9	89.4	93.8	92.7	53.9	84.4	99.3
GDM-aug	97.6	71.6	90.4	89.5	66.8	77.1	91.8
GraphLDM-aug	97.9	78.7	90.5	89.5		79.6	98.0
RADM DiT-B	98.5	87.3	94.1	91.7		85.0	99.3
UniGEM	99.0	89.8	95.0	93.2			
EquiFM	98.9	88.3	94.7		53.7	84.1	98.9
GeoBFN	98.6	87.2	93.0		64.4	78.9	93.1
GOAT	99.2		92.9		72.3	84.8	96.2
MOLTD	99.4	92.5	96.0		67.0	86.9	95.3
AnchorSet	99.2	91.3	95.5	93.6	69.5	86.4	99.1

Figure 2 gives a visual summary of the same advantage. In panel a, the AnchorSet variants concentrate almost all molecules near zero rotation error, while All-Proc and Heavy-Proc remain dominated by large errors beyond the 90° flip threshold; Murcko-Set further reduces the residual tail to 1.3 flips on average, and Learnable-Set preserves sub-millidegree accuracy while sharply improving complete candidate-set matching. Panels b and c show zero frame drift for the AnchorSet variants across the full displacement range, compared with 6.08° hydrogen drift for All-Proc and 2.91° non-scaffold heavy-atom drift for Heavy-Proc at $\sigma = 0.20$. Together, the three panels make the AnchorSet gain visible as a stable frame under both global rotations and local structural perturbations.

4.2 Performance Analysis

QM9. Against published QM9 generators (Table 2), AnchorSet improves molecule stability and validity over RADM DiT-B, GeoLDM, and UniGEM. This comparison is particularly relevant to RADM DiT-B because both use a non-equivariant latent generator; the gains show that the complete AnchorSet pipeline achieves better molecular quality in this setting. MOLTD remains ahead of AnchorSet on both measures, so the result is a strong but mixed ranking rather than a sweep of the benchmark. Atom stability is already high for several methods, yet their molecule stability varies more: reliable atoms do not automatically yield a stable complete molecule. Among generators

Table 3: QM9 LDM end-to-end ablation of the frozen AE canonicalizer. All configurations use the same latent diffusion backbone. All metrics are percentages and higher is better.

LDM (AE canonicalizer)	Atom Sta.	Mol. Sta.	Valid	$V \times U$	$V \times U \times N$
All-Atom	97.78	78.91	90.90	88.91	45.60
Heavy-Atom	97.68	77.88	90.30	88.51	45.48
Murcko-Set	98.67	83.39	91.53	90.40	64.52
Learnable-Set	**99.20**	**91.30**	**95.50**	**93.60**	**69.50**

reporting $V \times U$, AnchorSet is highest, narrowly above UniGEM. Its $V \times U \times N$ ranks second among reported values, behind GOAT, and exceeds MOLTD's reported joint rate.

GEOM-Drugs. AnchorSet reaches 86.4% Atom Stability and 99.1% Validity on GEOM-Drugs (Table 2). Its Atom Stability is 1.4 points above RADM DiT-B and 2.0 points above GeoLDM, nearly matching the 86.5% dataset reference and trailing MOLTD by 0.5 points. Its Validity is 0.2 points below GeoLDM and RADM DiT-B, while exceeding EquiFM, GraphLDM-aug, GOAT, and MOLTD. These two metrics show that the structure-anchored non-equivariant pipeline preserves high chemical validity as molecular size and flexibility increase.

4.3 Generation ablation

Table 3 holds the DiT backbone fixed while changing the canonical-coordinate pipeline. Heavy-Atom slightly trails All-Atom throughout, so excluding hydrogens from alignment does not improve generation despite suppressing hydrogen-induced AE frame drift (Table 1). Murcko-Set instead improves molecule stability and $V \times U \times N$ much more than atom stability. The larger whole-molecule gain suggests that scaffold-based frames help preserve chemically coherent structures; the parallel rise in the joint rate shows that this improvement carries through to valid, distinct, novel samples. Learnable-Set advances every metric again, with the largest gain in molecule stability. Its stronger AE candidate-set matching offers a consistent explanation for the improved output. Together, these results favor the complete structure-anchored configuration.

4.4 LDM rotational likelihood consistency

The paired LDM test asks whether AE frame consistency carries through to the non-equivariant generator by measuring rotation-induced changes in generator NLL. All-Proc and Heavy-Proc have median $|\Delta\mathrm{NLL}|$ near 11 and p95 above 100; Murcko-Set and Learnable-Set reduce the medians to 4.6×10^{-5} and 6.1×10^{-5}, with p95 below 7.1×10^{-4}.

For the two set models, candidate selection triggers in 7.42% and 7.66% of molecules, respectively. Non-triggered mean drift is 3.31×10^{-4} and 3.08×10^{-4}; triggered p95 is 0.00425 and 0.00432, but rare cases raise the triggered means to 0.457 and 1.189 and the maxima to 42.3 and 162.4. Heavy-atom fallback occurs in 10.08% of molecules, yet its mean drift remains below 2.2×10^{-4} for both models. Errors do not vary monotonically across eigengap quartiles. Thus the set representation removes the dominant discrete failure mode without eliminating the extreme tail.

5 Conclusion

We introduced AnchorSet, a structure-anchored, set-valued canonicalization framework for non-equivariant 3D molecular generation. Across QM9, structural anchors and candidate frames improve rotational and perturbation stability, while the matched-budget generation results show the strongest chemical metrics among the four tested configurations for the complete model. On GEOM-Drugs, atom stability and validity remain competitive under the dataset-specific heavy-atom protocol. The interface remains limited to $SO(3)$ rotations and discrete PCA sign and axis-permutation ambiguity; reflections, continuous repeated-eigenvalue subspaces, explicit chirality objectives, and unreliable scaffold annotations remain outside the present evidence.

AI USE STATEMENT

Generative AI tools were used to improve the clarity and readability of the manuscript and to assist with revisions to the schematic in Figure 1. The authors reviewed all AI-assisted work and take full responsibility for the content of the paper.

REPRODUCIBILITY STATEMENT

To facilitate reproduction, Section 3 specifies the canonicalization procedure, learning objectives, model architecture, and optimization settings, while Section 4 describes the datasets, comparisons, and evaluation protocol. Appendices A–F provide implementation details, the evaluation protocol, perturbation and reconstruction results, gate checks, and alignment diagnostics. An anonymized package containing the source code, configuration files, and instructions for training and evaluation will accompany the submission.

REFERENCES

Simon Axelrod and Rafael Gómez-Bombarelli. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data*, 9(1):185, 2022. doi: 10.1038/s41597-022-01288-4.

Guy W. Bemis and Mark A. Murcko. The properties of known drugs. 1. molecular frameworks. *Journal of Medicinal Chemistry*, 39(15):2887–2893, 1996. doi: 10.1021/jm9602928.

Yuhui Ding and Thomas Hofmann. Scalable non-equivariant 3D molecule generation via rotational alignment. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 13814–13824. PMLR, 13–19 Jul 2025.

Nadav Dym, Hannah Lawrence, and Jonathan W. Siegel. Equivariant frames and the impossibility of continuous canonicalization. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 12228–12267. PMLR, 21–27 Jul 2024.

Shikun Feng, Yuyan Ni, Lu Yan, Zhi-Ming Ma, Wei-Ying Ma, and Yanyan Lan. UniGEM: A unified approach to generation and property prediction for molecules. In *International Conference on Learning Representations*, 2025.

Victor Garcia Satorras, Emiel Hoogeboom, Fabian B. Fuchs, Ingmar Posner, and Max Welling. E(n) equivariant normalizing flows. In *Advances in Neural Information Processing Systems*, volume 34, pp. 4181–4192. Curran Associates, Inc., 2021.

Niklas W. A. Gebauer, Michael Gastegger, and Kristof T. Schütt. Symmetry-adapted generation of 3D point sets for the targeted discovery of molecules. In *Advances in Neural Information Processing Systems*, volume 32, pp. 7566–7578. Curran Associates, Inc., 2019.

Mohan Guo, Cong Liu, and Patrick Forré. Frame-based equivariant diffusion models for 3D molecular generation, 2025.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020.

Haokai Hong, Wanyu Lin, and Kay Chen Tan. GOAT: Accelerating 3D molecule generation via jointly geometric optimal transport. In *International Conference on Learning Representations*, 2025.

Emiel Hoogeboom, Víctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3D. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 8867–8887. PMLR, 2022.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Hierarchical generation of molecular graphs using structural motifs. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 4839–4848. PMLR, 13–18 Jul 2020.

Jaehyeong Jo, Seul Lee, and Sung Ju Hwang. Score-based generative modeling of graphs via the system of stochastic differential equations. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 10362–10383. PMLR, 17–23 Jul 2022.

Chaitanya K. Joshi, Xiang Fu, Yi-Lun Liao, Vahe Gharakhanyan, Benjamin Kurt Miller, Anuroop Sriram, and Zachary Ward Ulissi. All-atom diffusion transformers: Unified generative modelling of molecules and materials. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 28393–28417. PMLR, 2025.

Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 15546–15566. PMLR, 23–29 Jul 2023.

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

Jinwoo Kim, Dat Nguyen, Ayhan Suleymanzade, Hyeokjun An, and Seunghoon Hong. Learning probabilistic symmetrization for architecture agnostic equivariance. In *Advances in Neural Information Processing Systems*, volume 36, pp. 18582–18612. Curran Associates, Inc., 2023. doi: 10.52202/075280-0816.

Greg Landrum. RDKit: Open-source cheminformatics. Software, 2006.

Hannah Lawrence, Vasco Portilheiro, Yan Zhang, and Sékou-Oumar Kaba. Improving equivariant networks with probabilistic symmetry breaking, 2025.

Tuan Le, Julian Cremer, Frank Noé, Djork-Arné Clevert, and Kristof T. Schütt. Navigating the design space of equivariant diffusion-based generative models for de novo 3D molecule generation. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 26244–26266, 2024.

Haorui Li, Weitao Du, Yuqiang Li, Hongyu Guo, and Shengchao Liu. InertialAR: Autoregressive 3D molecule generation with inertial frames, 2025. Accepted at ICML 2026; formal proceedings metadata not yet available.

George Ma, Yifei Wang, Derek Lim, Stefanie Jegelka, and Yisen Wang. A canonicalization perspective on invariant and equivariant learning. In *Advances in Neural Information Processing Systems*, volume 37, pp. 60936–60979. Curran Associates, Inc., 2024. doi: 10.52202/079017-1948.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4172–4182. IEEE, 2023. doi: 10.1109/ICCV51070.2023.00387.

Omri Puny, Matan Atzmon, Heli Ben-Hamu, Ishan Misra, Aditya Grover, Edward J. Smith, and Yaron Lipman. Frame averaging for invariant and equivariant network design. In *International Conference on Learning Representations*, 2022.

Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(1):140022, 2014. doi: 10.1038/sdata.2014.22.

Kusha Sareen, Daniel Levy, Arnab Kumar Mondal, Sékou-Oumar Kaba, Tara Akhound-Sadegh, and Siamak Ravanbakhsh. Symmetry-aware generative modeling through learned canonicalization, 2025.

Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9323–9332. PMLR, 2021.

Yuxuan Song, Jingjing Gong, Minkai Xu, Ziyao Cao, Yanyan Lan, Stefano Ermon, Hao Zhou, and Wei-Ying Ma. Equivariant flow matching with hybrid probability transport for 3D molecule generation. In *Advances in Neural Information Processing Systems*, volume 36, pp. 549–568, 2023. URL https://openreview.net/forum?id=hHUZ5V9XFu.

Yuxuan Song, Jingjing Gong, Yanru Qu, Hao Zhou, Mingyue Zheng, Jingjing Liu, and Wei-Ying Ma. Unified generative modeling of 3D molecules via bayesian flow networks. In *International Conference on Learning Representations*, 2024.

Clément Vignac, Igor Krawczuk, Antoine Siraudin, Bohan Wang, Volkan Cevher, and Pascal Frossard. DiGress: Discrete denoising diffusion for graph generation. In *International Conference on Learning Representations*, 2023.

Fang Wan, Wenhan Gao, Jingxiang Qu, and Yi Liu. Diffuse and steer: Corrective sampling for stable 3D molecular diffusion. In *ReALM-GEN Workshop at ICLR 2026*, 2026. URL https://openreview.net/pdf?id=Se6N9P6apP. Accepted workshop paper.

Yuyang Wang, Ahmed A. A. Elhag, Navdeep Jaitly, Joshua M. Susskind, and Miguel Ángel Bautista. Swallowing the bitter pill: Simplified scalable conformer generation. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 50400–50418. PMLR, 21–27 Jul 2024.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. GeoDiff: A geometric diffusion model for molecular conformation generation. In *International Conference on Learning Representations*, 2022.

Minkai Xu, Alexander S. Powers, Ron O. Dror, Stefano Ermon, and Jure Leskovec. Geometric latent diffusion models for 3D molecule generation. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 38592–38610. PMLR, 2023.

Yixian Xu, Yusong Wang, Shengjie Luo, Kaiyuan Gao, Tianyu He, Di He, and Chang Liu. Quotient-space diffusion models. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 137094–137136, 2026.

Weichi Yao, Cameron Gruich, Bryan R. Goldsmith, and Yixin Wang. Fixed-dimensional latent flow for generating variable-size 3D molecules, 2026.

Yi Yu, Tengyao Wang, and Richard J. Samworth. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015. doi: 10.1093/biomet/asv008.

Zhilong Zhang, Yuxuan Song, Yichun Wang, Jingjing Gong, Hanlin Wu, Dongzhan Zhou, Hao Zhou, and Wei-Ying Ma. Accelerating 3D molecule generative models with trajectory diagnosis. In *Advances in Neural Information Processing Systems*, volume 38, pp. 39099–39127, 2025.

Cai Zhou, Zijie Chen, Zian Li, Jike Wang, Kaiyi Jiang, Pan Li, Rose Yu, Muhan Zhang, Stephen Bates, and Tommi Jaakkola. Rethinking diffusion models with symmetries through canonicalization with applications to molecular graph generation, 2026.

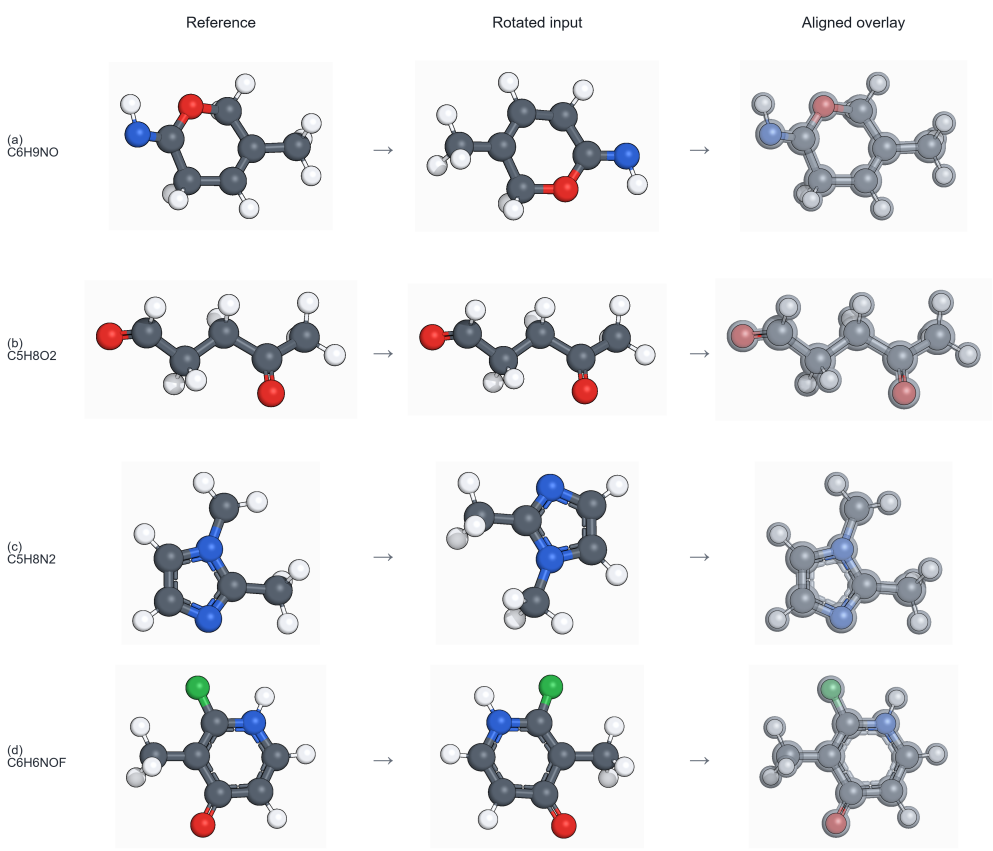


Figure S1: Qualitative rigid-alignment checks on representative held-out QM9 molecules. Each row shows the same conformation in its dataset orientation (left), after a prescribed global rotation (middle), and after all-atom Kabsch recovery (right). The recovered structure is rendered in element colors and the reference in transparent gray. These panels document the controlled rotation and visualization procedure.

A Implementation Details

The sign-confidence denominator floor is $\varepsilon_c = 10^{-8}$. The candidate fingerprint uses scale floor $\varepsilon_s = 10^{-6}$, projection direction $v = (1, \sqrt{2}, \sqrt{3})^\top / \|(1, \sqrt{2}, \sqrt{3})\|_2$, projection coefficient $\beta = 0.5$, and clipping amplitude $b = 10$.

QM9 uses the Murcko prior and fallback procedure from Section 3.2; $\delta_w = 0.05$, type weights start near 1, and $\tau_c = 10^{-3}$. Its AE has a nine-layer decoder of width 256 and $F_z = 1$ (four channels per atom), with 200 AE epochs and 6,000 LDM epochs. GEOM-Drugs uses a heavy-atom prior with learnable atom-type weights, a four-layer decoder of width 256, $F_z = 2$ (five channels per atom), 4 AE epochs, and 55 LDM epochs; both datasets use $T = 1000$.

Both datasets use batch size 64 for QM9 AE, 32 for GEOM-Drugs AE, and 256 for LDM training. All stages use AdamW with AMSGrad, initial learning rate 10^{-4}, and weight decay 10^{-12}. AE training uses cosine annealing with minimum learning rates 10^{-6} (QM9) and 10^{-5} (GEOM-Drugs); neither AE uses gradient clipping or EMA. The LDMs use gradient clipping with EMA decays 0.999 and 0.9999, respectively.

The AE coordinate squared error is first summed over the three coordinate channels, then weighted by the valid-node mask and averaged over the padded node dimension; the charge squared error uses the same mask. The current atom-category cross-entropy is averaged over the padded node dimension without an additional node mask, and this convention is kept across all comparisons.

The main AE uses no KL regularizer, rotation-consistency loss, or hydrogen/side-chain perturbation loss; input rotations and hydrogen/side-chain perturbations are used only for evaluation. The PCA path does not use a learned rotation encoder or head; the valid-edge mask excludes padding nodes and self-edges. Eigengap is diagnostic only and does not trigger candidates. The fingerprint always uses category indices and does not assume invariance to a reindexing of category channels.

B Evaluation Protocol

QM9 retains atom types, graphs, hydrogens, centered coordinates, project partitions, and verified Murcko mappings. The AE alignment and gate audits use validation molecules, paired reconstruction uses separate validation and test partitions, and the LDM likelihood audit uses test molecules. GEOM-Drugs is split by parent SMILES so conformers cannot cross partitions. For rotations,

$$d_{SO(3)}(A, B) = \operatorname{atan2}\left(\frac{\| \operatorname{vee}(A^\top B - B^\top A)\|_2}{2}, \frac{\operatorname{tr}(A^\top B) - 1}{2}\right).$$

Candidate sets use directed nearest matching after rotation compensation. Perturbation tests displace hydrogens or non-scaffold heavy atoms, remove subset displacement mean, and recompute the frame. Generation reports Atom and Molecule Stability, Validity, Uniqueness, Valid×Unique, and Novelty when supported.

The visualization provides a qualitative check of the rotation protocol used by the quantitative audits. Across molecules with oxygen, nitrogen, and fluorine substituents, the prescribed global rotation changes the displayed pose while preserving atom identities and connectivity; after rigid recovery, the colored structure overlaps the transparent reference throughout the scaffold and peripheral groups. This consistent overlap supports the interpretation of the numerical rotation and reconstruction metrics as geometric alignment measures.

C Perturbation Results

Table 4: AE frame drift in degrees under non-anchor perturbations at $\sigma = 0.1$; mean $\pm$ SD across repeated evaluations. Hydrogen perturbations cover 1,280 validation molecules per evaluation; side-chain perturbations use only molecules with eligible non-scaffold heavy atoms.

AE	H mean	H p95	side-chain mean	side-chain p95
All-Proc	3.034 ± 0.134	8.464 ± 0.252	0.677 ± 0.022	2.297 ± 0.182
Heavy-Proc	$7.14 \times 10^{-6} \pm 1.15 \times 10^{-7}$	$1.48 \times 10^{-5} \pm 8.52 \times 10^{-7}$	1.467 ± 0.068	4.804 ± 0.524
Murcko-Single	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Murcko-Set	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Learnable-Set	0 ± 0	0 ± 0	0 ± 0	0 ± 0

Table 4 reveals a clear hierarchy in how atom selection controls frame stability. All-Proc is sensitive to hydrogen displacement, with a mean drift of 3.034° and a p95 of 8.464°. Removing hydrogens from pose estimation suppresses this response in Heavy-Proc, but its mean and p95 side-chain drift rise to 1.467° and 4.804°, more than twice the corresponding All-Proc values. In contrast, all three scaffold-anchored variants maintain zero measured drift under both perturbations. Structural anchoring therefore isolates the global frame from flexible peripheral geometry instead of merely transferring sensitivity from hydrogens to side chains. For AnchorSet, this stable support complements the candidate representation: the scaffold protects the frame from local conformational noise, while the candidate set handles the remaining orientation ambiguity of the anchored core.

D Reconstruction Results

Table 5: Paired AE reconstruction results on 1,280 molecules per partition. The valid and test Kabsch columns use their respective partitions; all other columns use test molecules. Coordinate quantities are RMSD or pairwise distance MAE; RDKit validity uses the RDKit toolkit (Landrum, 2006); validity and stability are percentages.

AE	valid Kabsch	test Kabsch	coord. RMSD	pairwise MAE	RDKit valid.	atom stab.	mol. stab.
All-Proc	0.001814	0.001695	0.001813	0.002064	97.89	99.56	97.19
Heavy-Proc	0.001855	0.001746	0.001867	0.002162	97.89	99.56	97.19
Murcko-Single	0.001895	0.001750	0.001909	0.002164	97.81	99.51	96.88
Murcko-Set	0.001850	0.001740	0.001884	0.002150	97.58	99.52	96.88
Learnable-Set	0.001797	0.001681	0.001831	0.002121	97.81	99.54	97.11

Table 6: Additional AE reconstruction diagnostics. Atom-type accuracy, charge accuracy, and graph connectivity are percentages; throughput is measured in molecules per second.

AE	atom-type acc.	charge acc.	charge MAE	connectivity	throughput (mol/s)
All-Proc	100	100	0.002378	100	18.38
Heavy-Proc	100	100	0.002168	100	20.08
Murcko-Single	100	100	0.002175	100	16.47
Murcko-Set	100	100	0.002002	100	19.72
Learnable-Set	100	100	0.002055	100	19.47

Table 5 shows that Learnable-Set achieves the lowest Kabsch-aligned reconstruction error on both partitions, reaching 0.001797 on validation molecules and 0.001681 on test molecules. Its unaligned coordinate RMSD of 0.001831 remains close to the best value of 0.001813 from All-Proc, indicating that the gain comes from more consistent frame recovery rather than a trade-off in coordinate fidelity. The chemical reconstruction measures are likewise preserved: Learnable-Set reaches 100% atom-type and charge accuracy, 100% graph connectivity, 97.81% RDKit validity, and 97.11% molecule stability. Relative to Murcko-Set, learning the anchor weights lowers test Kabsch RMSD from 0.001740 to 0.001681 and raises molecule stability from 96.88% to 97.11%. AnchorSet therefore improves aligned reconstruction while retaining the atom-level and molecular integrity of the decoded structures.

E GATE CHECKS

Table 7: Effect of candidate gating on AE frame consistency. Each slash-separated entry lists the three repeated evaluations in the same order. Gating changes the selected operational representative while leaving the complete candidate set unchanged.

AE variant	gate	rot. mean (degrees)	flips	low-conf. mean (degrees)
Murcko-Set	off	1.562 / 1.285 / 1.733	11 / 9 / 12	13.886 / 11.423 / 15.398
Murcko-Set	on	0.985 / 0.877 / 1.135	7 / 6 / 8	8.751 / 7.798 / 10.090
Learnable-Set	off	1.942 / 1.577 / 1.694	14 / 11 / 12	17.630 / 14.310 / 15.375
Learnable-Set	on	0.848 / 1.155 / 0.991	6 / 8 / 7	7.696 / 10.480 / 8.992

Table 7 shows that candidate gating improves every reported operational measure for both set-valued models. For Murcko-Set, the mean rotation error falls by 31.8–36.9% across the three evaluations, while the flip count decreases from 9–12 to 6–8. Learnable-Set shows an even larger reduction in two evaluations and lowers its flips from 11–14 to 6–8. The low-confidence mean follows the same pattern, confirming that the improvement is concentrated in the geometrically ambiguous cases for which a single PCA frame is least reliable. Because the complete candidate set is unchanged, these gains arise from selecting a more consistent representative rather than altering the underlying molecular geometry. Candidate gating thus converts the redundancy retained by AnchorSet into a direct reduction of discrete frame failures.

Table 8: Sensitivity of candidate use and flip counts to the gate threshold. Candidate fractions are percentages; flip entries list the three repeated evaluations.

Threshold	Murcko cand.	Learnable cand.	Murcko flips	Learnable flips
10^{-4}	7.34	7.34	9 / 7 / 9	6 / 9 / 9
10^{-3}	11.25	11.02	7 / 6 / 8	6 / 8 / 7
10^{-2}	24.53	24.53	7 / 6 / 8	6 / 8 / 7

Table 8 identifies a clear saturation point for candidate activation. Increasing the threshold from 10^{-4} to 10^{-3} raises candidate use from 7.34% to about 11% and reduces flips in five of the six model–evaluation combinations. Increasing it further to 10^{-2} more than doubles candidate use to 24.53%, but produces no additional reduction in flips. The 10^{-3} setting therefore captures the ambiguity associated with the observed discrete failures while avoiding candidate expansion for many already stable molecules. The same saturation pattern for Murcko-Set and Learnable-Set also indicates that the gate responds consistently to the shared geometric confidence criterion.

F Alignment Diagnostics

Table 9: Complete QM9 validation frame-consistency statistics for the five AEs. Values are means ± SD across three evaluations of 1,280 molecules. Rotation, low-confidence (sign confidence below 10^{-3}), and candidate-set quantities are in degrees; candidate is the fraction of molecules using the candidate path. Candidate-set statistics are reported for all PCA variants, and a flip is an operational-frame error above 90°.

AE	rot. mean	rot. p95	low-conf. mean	low-conf. p95	low-conf. max	cand. mean	cand. max	cand. (%)	flips
All-Proc	128.493±0.770	178.337±0.161	126.449±8.718	177.854±1.405	178.920±0.933	–	–	0	969.7±8.1
Heavy-Proc	119.938±0.772	177.061±0.234	112.425±8.026	174.243±3.057	178.532±1.284	–	–	0	918.0±12.1
Murcko-Single	0.422±0.244	0.000370±0.000028	5.546±2.426	0.003021±0.001908	180.000±0	0.004703±0.007605	5.751±9.646	0	3.0±1.7
Murcko-Set	0.192±0.078	0.000363±0.000024	2.645±1.115	0.001435±0.000501	180.000±0	0.004703±0.007605	5.751±9.646	7.344±1.025	1.3±0.6
Learnable-Set	0.235±0.081	0.000406±0.000019	3.459±1.517	0.001663±0.001026	180.000±0	0.000646±0.000540	0.639±0.639	7.083±1.032	1.7±0.6

Table 10: AE rotation error stratified by normalized covariance eigengap quartile. Q1 contains the smallest eigengaps and Q4 the largest; entries are mean degrees ± SD.

AE	Q1	Q2	Q3	Q4
All-Proc	132.152±1.711	126.803±2.409	128.903±3.887	126.114±5.857
Heavy-Proc	121.210±1.034	123.276±1.444	116.534±0.520	118.732±3.860
Murcko-Single	0.564±0.562	0.563±0	0.375±0.650	0.188±0.325
Murcko-Set	0.394±0.635	0±0	0±0	0.375±0.325
Learnable-Set	0.752±0.323	0.188±0.325	0±0	0±0

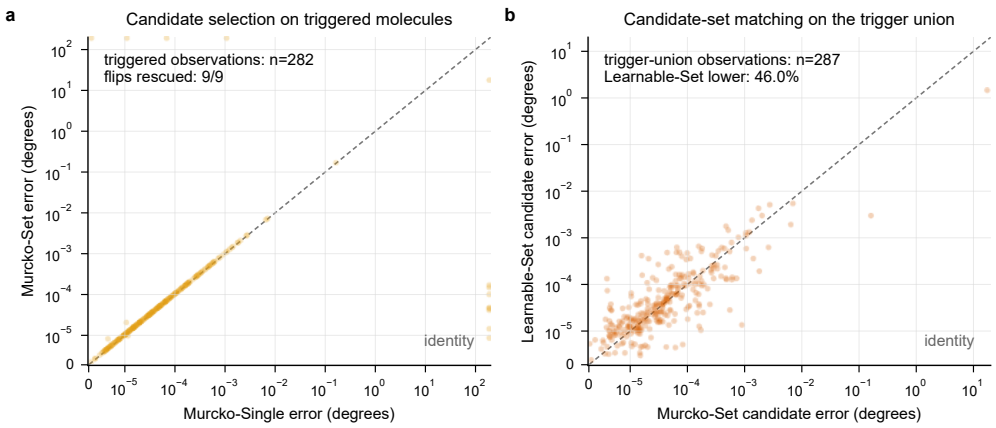


Figure S2: Candidate selection resolves discrete frame failures on triggered molecules. **a,** Paired operational-frame errors for Murcko-Single and Murcko-Set on candidate-triggered cases. Points below the identity line favor candidate selection; all nine Murcko-Single flips fall below the $90°$ criterion with Murcko-Set. **b,** Candidate-set matching error for Murcko-Set and Learnable-Set on the union of cases that trigger either candidate path. Points below the identity line favor Learnable-Set.

Table 9 shows the strongest separation between conventional alignment and structure-anchored canonicalization. All-Proc and Heavy-Proc retain mean rotation errors above $119°$, p95 errors above $177°$, and more than 900 flips. Introducing scaffold-based PCA reduces the p95 error to approximately 4×10^{-4} degrees and the flip count to at most three, demonstrating that a chemically stable core provides a far more reliable orientation signal than all-atom or heavy-atom alignment.

The set-valued variants further improve the difficult tail. Relative to Murcko-Single, Murcko-Set more than halves the operational mean error, low-confidence mean error, and flip count. Learnable-Set provides the most accurate complete-set correspondence: its candidate-set mean falls from 0.004703 to 0.000646 degrees and its candidate-set maximum from 5.751 to 0.639 degrees, reductions of 86.3% and 88.9%, respectively. These complementary gains show how AnchorSet combines two distinct strengths: candidate selection suppresses discrete representative flips, while learnable structural weights sharpen the agreement between the full sets of equivalent frames.

Figure S2 makes these improvements visible at the case level. Every one of the nine Murcko-Single flips among the triggered cases falls below $90°$ with Murcko-Set, showing that the candidate mechanism corrects the dominant discrete failures rather than merely lowering an average. The Learnable-Set comparison places a large share of cases below the identity line, consistent with its substantially lower candidate-set mean and maximum. Together, the aggregate statistics and paired cases show that AnchorSet improves both the typical alignment and the rare ambiguous configurations that dominate the error tail.

The eigengap analysis in Table 10 confirms that the advantage is not confined to one covariance regime. Both Procrustes baselines remain between $116.5°$ and $132.2°$ across all quartiles, whereas every scaffold-based variant stays below $0.752°$ and reaches zero mean error in several strata. Even with a $10.10\% \pm 0.51\%$ heavy-atom fallback rate, the structure-anchored variants maintain consistently small quartile means. The remaining $180°$ events are rare tail cases, while the overall results show that AnchorSet is robust across both small- and large-eigengap molecules.