TABASCO: A Fast, Simplified Model for Molecular Generation with Improved Physical Quality

Carlos Vonessen [*1 †] **Charles Harris** [*2] **Miruna Cretu** [*2] **Pietro Liò** [2]

Abstract

State-of-the-art models for 3D molecular generation are based on significant inductive biases—*SE*(3), permutation equivariance to respect symmetry and graph message-passing networks to capture local chemistry—yet the generated molecules still struggle with physical plausibility. We introduce TABASCO which relaxes these assumptions: The model has a standard non-equivariant transformer architecture, treats atoms in a molecule as sequences and reconstructs bonds deterministically after generation. The absence of equivariant layers and message passing allows us to significantly simplify the model architecture and scale data throughput. On the GEOM-Drugs benchmark TABASCO achieves state-of-the-art PoseBusters validity and delivers inference roughly $10\times$ faster than the strongest baseline, while exhibiting emergent rotational equivariance despite symmetry not being hard-coded. Our work offers a blueprint for training minimalist, high-throughput generative models suited to specialised tasks such as structure- and pharmacophore-based drug design. We provide a link to our implementation at `github.com/carlosinator/tabasco`.

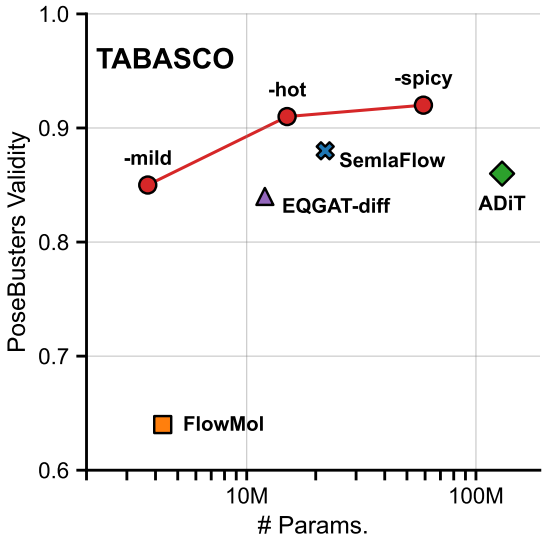

Figure 1: Comparison of POSEBUSTERS on GEOM-Drugs.

1. Introduction

In recent years, there has been growing interest in using diffusion models as generative methods for molecular design (Du et al., 2024; Schneuing et al., 2022; Hoogeboom et al., 2022; Vignac et al., 2023; Dunn & Koes, 2024; Irwin et al., 2025). Much of the literature converges on design principles believed to be essential for high-quality molecular generation. First, models are typically SE(3)-equivariant, ensuring that rotations or translations of input conformers yield correspondingly transformed outputs—a symmetry prior that serves as a strong inductive bias (Hoogeboom et al., 2022). Second, message-passing graph neural networks (GNNs) are widely used to capture many-hop, context-dependent interactions between atoms (Hoogeboom et al., 2022; Schneuing et al., 2022; Irwin et al., 2025; Dunn & Koes, 2024; Schneuing et al.). Third, recent work emphasises flow-matching objectives that rely on coupled optimal transport (OT) (Tong et al., 2023) or incorporate heavily structured, domain-informed priors (Dunn & Koes, 2024; Irwin et al., 2025). However, despite incorporating these inductive biases, current models continue to struggle with physical plausibility—often failing to produce chemically coherent structures or accurately recover fundamental features of protein–ligand binding (Buttenschoen et al., 2024; Harris et al., 2023).

In parallel, a growing body of work explores scaling up simpler model architectures—most notably Transformers (Vaswani et al., 2017)—across adjacent domains. A prominent example is AlphaFold3 (Abramson et al., 2024), which

[*]Core contributor, [†] Work done as a visiting student in Cambridge [1]ETH Zurich, Switzerland [2]University of Cambridge, United Kingdom. Correspondence to: Carlos Vonessen <cvonessen[at]ethz.ch>, Charles Harris <cch57[at]cam.ac.uk>.

Proceedings of the Workshop on Generative AI for Biology at the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

achieves strong performance on physical plausibility benchmarks (Buttenschoen et al., 2024) despite omitting many of the conventional inductive biases, including equivariance. Similarly, recent generative models for protein backbone design have demonstrated competitive results with minimal architectural complexity, provided they are scaled appropriately (Geffner et al., 2025). Simplification of model architecture and removal of inductive biases for conformer generation has also proven successful (Wang et al., 2024). In parallel to this work, (Joshi et al., 2025) explore using non-equivariant latent diffusion for generating small molecules.

In this work, we aim to distill the core components of diffusion-based molecular generation and ask: how much architectural complexity is necessary to build high-performing models? We introduce TABASCO (Transformer-based Atomistic Bondless Scalable Conformer Output), a stripped-down and scalable model that achieves state-of-the-art performance on unconditional molecular generation benchmarks. Despite its simplicity, TABASCO exceeds the physical plausibility of more complex models, as measured by POSEBUSTERS (Buttenschoen et al., 2024) validity, while being up to 10× faster at inference. Our contributions are as follows:

(i) **State-of-the-art physical quality on GEOM-Drugs.** TABASCO surpasses prior models such as FlowMol and SemlaFlow in POSEBUSTERS validity, achieving a 10× speed-up at sampling time.

(ii) **Lean, bond-free Transformer backbone.** Our model omits both bond inputs and equivariant layers, relying instead on a standard Transformer to generate high-quality coordinates. Chemoinformatics tools recover bonds post hoc, which maintains physical plausibility and focuses computational resources on exact coordinate generation.

(iii) **Physically-constrained last-mile correction.** We introduce a simple distance-bounds guidance step that improves POSEBUSTERS validity without requiring force-field-based relaxation or additional parameters.

(iv) **Emergent structure without explicit symmetry.** We analyse the model's equivariant behaviour despite the absence of SE(3) symmetry constraints, and investigate the role of positional encodings in improving model performance.

2. Background and Related Work

2.1. Flow-Matching Models

Flow-matching (FM) is a generative modelling framework that learns to transport samples from a source distribution (e.g., noise) to a target distribution (e.g., data) by directly estimating the time-dependent velocity field of a probability flow (Lipman et al., 2023; Albergo & Vanden-Eijnden, 2022).

Given a pair of samples $(\mathbf{x}_0, \mathbf{x}_1)$ from source and target distributions, one defines a continuous interpolation $\mathbf{x}_t = (1-t)\mathbf{x}_0 + t\mathbf{x}_1$, and a target velocity $u_t = \frac{\mathbf{x}_1 - \mathbf{x}0}{t(1-t)}$. A neural field $v_\theta(\mathbf{x}_t, t)$ is then trained to match this velocity using the squared error:

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{t,(\mathbf{x}_0,\mathbf{x}1)} \left[||v_\theta(\mathbf{x}_t, t) - u_t||_2^2 \right]. \tag{1}$$

Flow-matching enables efficient generation via deterministic integration (e.g., using an ODE solver), and has been shown to improve sampling speed and stability over score-based diffusion models (Dunn & Koes, 2024; Irwin et al., 2025).

2.2. Generative Models for 3D Molecule Design

Early works used standard continuous diffusion processes on coordinate and atomic features, where bond connectivity was determined by chemoinformatics software (Hoogeboom et al., 2022; Schneuing et al., 2022). This process often resulted in low-quality conformers that were not fully-connected or violated atomic valences. MiDi (Vignac et al., 2023) improved on this by applying discrete diffusion to both the atom types as well as generated a full bond matrix end-to-end, which significantly increased stabilty and bond connectivity. EQGAT-diff (Le et al., 2023) explored the design space of equivariant diffusion models, creating a custom attention-based equivariant architecture to allow for interaction between continuous and discrete features. Further work introduced more advanced model architectures (Morehead & Cheng, 2024; Hua et al., 2024), additional losses (Xu et al., 2024), alternative transport strategies (Song et al., 2023), and geometric latent diffusion (Xu et al., 2023; Joshi et al., 2025). FlowMol (Dunn & Koes, 2024) and SemlaFlow (Irwin et al., 2025) use flow-matching for generation of coordinates, atom types and bonds. Both methods proposed new architectures and showed great improvements in speed versus diffusion based approaches.

3. TABASCO: Fast, Simple, and High-Quality Molecule Generation

Overview and Motivation Our goal in this work is to identify the simplest possible model architecture that can generate physically realistic small molecules at scale. Our motivation stems from the observation that recent progress in protein structure generation has demonstrated the surprising power of non-equivariant Transformer architectures when scaled appropriately (Abramson et al., 2024; Geffner et al., 2025; Wang et al., 2024). Based on these results, we began our experiments with a deliberately stripped-down,

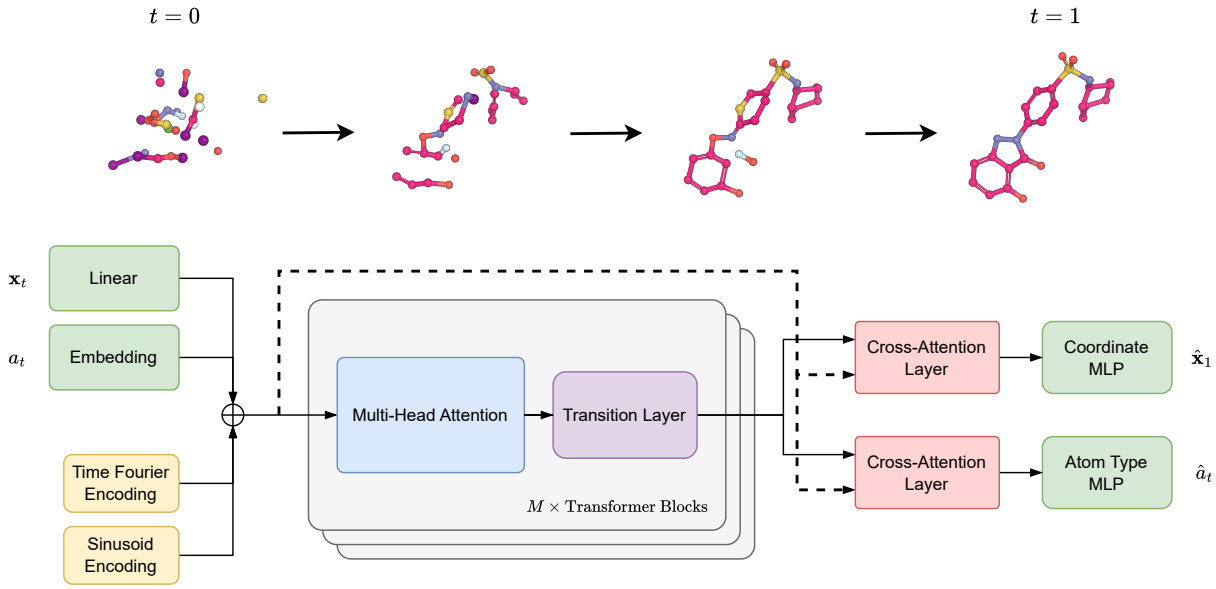

Figure 2: Top: Interpolation between noise and data. Bottom: TABASCO model architecture.

non-equivariant Transformer backbone for molecular generation.

We choose to exclude explicit bond information from the model. While most existing models treat bonds as a distinct modality, often processed with triangle attention or edge representations, we rely on standard chemoinformatics tools instead, which infer bonds reliably so long as the generated coordinates are physically sensible. We therefore hypothesised that if coordinate generation is sufficiently accurate, bond information becomes redundant. This perspective allowed us to further simplify the architecture while focusing on improving conformer quality.
Physical realism, as measured by POSEBUSTERS validity, is the primary metric guiding design decisions. Modules and heuristics in the approach that did not contribute to this metric were pruned, resulting in a lean, fast, and extensible model that maintains strong performance without relying on specialised architectural components.

3.1. Model Architecture

In contrast to most prior work in unconditional molecular generation, we adopt a simplified non-equivariant Transformer architecture (see Figure 2) without self-conditioning. Atom coordinates and types are jointly embedded along with time and sequence encodings. These are passed through a stack of standard Transformer blocks (Vaswani et al., 2017). We add a single cross-attention layer for each domain and process these outputs in MLP heads for atom types and coordinates. The resulting model is straightforward to implement, highly extensible, and dramatically faster at sampling time than previous equivariant or bond-aware approaches.

3.2. Training Objective

We optimize coordinates with Euclidean conditional flow-matching (CFM) (Tong et al., 2023; Albergo & Vanden-Eijnden, 2022) and atom types with discrete CFM which is parametrized based on the Discrete Flow Models (DFM) framework (Campbell et al., 2024). Concretely, consider a molecule with N atoms, ground-truth coordinates $\mathbf{x_1}$ and atom types a_1. Coordinates are partially noised with $\mathbf{x_t} = t \cdot \mathbf{x_1} + (1-t) \cdot \epsilon$, where the noise is distributed with $\epsilon \sim \mathcal{N}(0, I)$. Noisy atom types a_t are obtained by interpolating between atom type probabilities and sampling from a categorical distribution $a_t \sim \text{Cat}\left(t \cdot \delta(a_1) + (1-t) \cdot \delta(\frac{1}{N})\right)$, where $\delta(\cdot)$ creates a one-hot encoding (Campbell et al., 2024). During training the model takes $\mathbf{x_t}$ and a_t and learns to predict the endpoint of the trajectory. The continuous coordinate objective becomes

$$L_{\text{metric}}(\mathbf{x}) = \mathbb{E}_{\epsilon,t}\left[\frac{1}{N}||\hat{\mathbf{x}}_1^\theta(\mathbf{x}_t, t) - \mathbf{x}_1||_2^2\right]. \quad (2)$$

The discrete atom type objective is the cross-entropy loss

$$L_{\text{discrete}}(a) = \mathbb{E}_t\left[-\sum_i a_i \log(\hat{a}_1(a_t, t))\right]. \quad (3)$$

We combine these into a multi-objective formulation with

weighing factor $\lambda_{\text{discrete}} \in (0, 1]$, as

$$L_{\text{total}}(\mathbf{x}, a) = L_{\text{metric}}(\mathbf{x}) + \lambda_{\text{discrete}} \cdot L_{\text{discrete}}(a)\,. \quad (4)$$

During training we sample from $t \sim \text{Beta}(\alpha, 1)$, where α is a hyperparameter we ablate in Appendix B. As $t \to 1$ the model's behaviour approaches the identity function, due to the chosen endpoint formulation. To ensure the model can still learn precise atom placement even as losses approach zero as $t \to 1$, we weigh the loss with $\beta(t) \cdot L_{\text{total}}(\mathbf{x}_t, a_t)$ based on the sampled time t, with

$$\beta(t) = \min\left\{100, \frac{1}{(1-t)^2}\right\}. \quad (5)$$

3.3. Sampling

We generate molecules with TABASCO by simulating a system of coupled stochastic differential equations:

$$\begin{aligned} \mathrm{d}\mathbf{x}_t &= \mathbf{v}_t^\theta(\mathbf{x}_t, a_t)\mathrm{d}t + g(t)\,\mathbf{s}_t^\theta(\mathbf{x}_t, a_t)\mathrm{d}t \\ &\quad + \sqrt{2g(t)\gamma}\,\mathrm{d}W_t\,, \quad (6) \\ \partial p_t &= R_t(\mathbf{x}_t, a_t)^\top p_t \quad (7) \end{aligned}$$

where p_t describes the probability of each atom type at time t. We estimate the velocity with $\mathbf{v}_t = \frac{\mathbf{x}_1 - \mathbf{x}_t}{1-t}$ from the models endpoint prediction $\hat{\mathbf{x}}_1$ at time t, and the score with $\mathbf{s}_t = \frac{t\mathbf{v}_t - \mathbf{x}_t}{1-t}$. We refer the reader to (Geffner et al., 2025; Campbell et al., 2024) on which we base our coordinate and atom type sampling strategies, for more in-detail discussions. To improve sample quality, we apply a logarithmic discretization scheme on $t \in [0, 1]$ with more fine-grained steps near the end of denoising. We also scale the score $\mathbf{s_t}$ and the Gaussian noise component $\mathrm{d}W_t$ by $g(t)$, setting it to zero as $t \to 1$ (see Appendix B).

3.4. Ordering Atoms as Sequences

Transformers operate in a bag-of-tokens fashion unless provided with additional information about the absolute or relative positions of those tokens. Unlike text or protein sequences, small molecules lack a natural linear ordering that reflects their 3D structure. While formats such as SMILES and InChI offer consistent ways to linearise molecular graphs, the ordering in these representations does not strictly correspond to spatial proximity. However, the SMILES ordering is deterministically derived—typically via a depth-first traversal starting from a canonical root atom (Weininger et al., 1989)—which does impart some semantic structure. In practice, many neighbouring atoms in the SMILES string are also spatially or chemically proximate in the molecule. We hypothesise that this implicit locality helps the model establish a coarse structural scaffold early in the generation process (see lower trajectory in Figure 7). Accordingly, we include sinusoidal positional encodings based on the atom indices in the SMILES sequence, and ablate their effect in Section 4.3.

3.5. Physically Constrained Last-Mile Pose Guidance

Existing 3D molecule generators yield globally sound conformations but struggle with local stereochemical checks such as POSEBUSTERS. We find most violations stem from subtle coordinate drifts that accumulate near the end of the sampling trajectory ($t \to 1$). We therefore frame pose refinement as a *last-mile* problem and introduce a lightweight, differentiable guidance step that enforces simple physical distance bounds without force-field evaluation or relaxation.

Distance–bounds matrix. For every element pair we pre-compute lower and upper bounds $[L_{ij}, U_{ij}]$ over 1–5 bond separations, analogously to how POSEBUSTERS computes bounds on valid bond lengths and angles:

- **Lower bound L_{ij}:** sum of van-der-Waals radii minus 0.1 Å;
- **Upper bound U_{ij}:** cumulative covalent bond lengths along the shortest path.

These numbers match the limits used in Universal Force Field (UFF) relaxation but are looked up from a static table; *no* UFF energy, gradients, or optimisation is performed.

Two-phase sampling with distance-bounds guidance.

1. **Free denoising.** Run the standard sampler until $t = 0.99$, obtaining noised conformation $(\mathbf{x}_{0.99}, a_{0.99})$.
2. **Guided refinement.** In each remaining denoising step, convert the endpoint predicted coordinates to an RDKIT conformer and look up the physical bounds on atom pair distances $[L_{ij},\ U_{ij}]$ for each distance pair $d_{ij} = ||\mathbf{x}_{t,i} - \mathbf{x}_{t,j}||$. The loss on physical constraints is computed with

$$\mathcal{L}_{\text{phys}}(\mathbf{x}_t) = \sum_{i<j} \begin{cases} \big(d_{ij} - U_{ij}\big)^2, & d_{ij} > U_{ij}, \\ \big(L_{ij} - d_{ij}\big)^2, & d_{ij} < L_{ij}, \\ 0, & \text{otherwise}\,. \end{cases}$$

We back-propagate through the network and apply one gradient step to the inputs:

$$\mathbf{x}_t \;\leftarrow\; \mathbf{x}_t - \alpha_{\text{phys}} \frac{\partial \mathcal{L}_{\text{phys}}}{\partial \mathbf{x}_t}$$

If the molecule decoded at $t = 0.99$ is not RDKIT-valid, no guidance is applied to the sample.

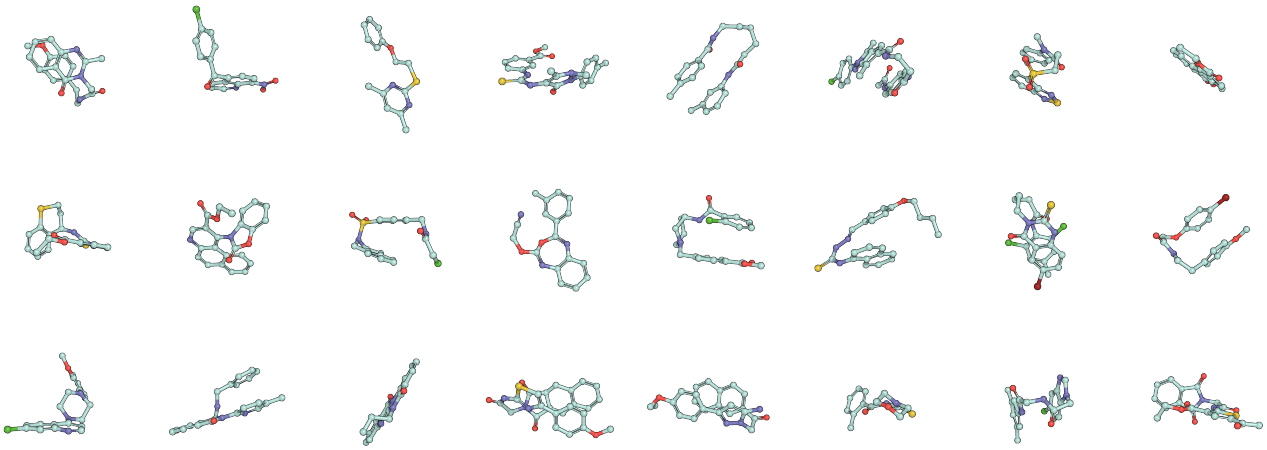

Figure 3: **Sampled molecules from TABASCO**.

4. Experiments

4.1. Experimental Setup

Training dataset We train TABASCO on GEOM-Drugs (Axelrod & Gomez-Bombarelli, 2022), a dataset of 1M high-quality conformers of drug-like molecules. We use the splits from Vignac et al. (2023) and, following Irwin et al. (2025), we discard molecules with more than 72 heavy atoms from the training dataset, accounting for 1% of the data. During testing, we sample from the distribution of molecule sizes in the test set, which was left unchanged.

Evaluation Metrics We evaluate generated molecules on several metrics: (i) **Validity**: Whether a molecule can be sanitized with RDKIT, (ii) **Novelty**: Whether the canonical SMILES of the molecule is not present in the training set, (iii) **Diversity**: Tanimoto similarity of molecule fingerprints, (iv) **Strain Energy** (Harris et al., 2023): Energy of the molecule compared to low energy conformers, (v) **Root Mean Square Deviation** (RMSD): When comparing molecules, averaged distance between the atoms of two molecules, (vi) **POSEBUSTERS** (Buttenschoen et al., 2024): Evaluates steric clashes, valid bond lengths and bond angles, double bond and aromatic ring flatness, and sufficiently low strain energy with respect to simulated conformers. We employ POSEBUSTERS as our main metric for measuring conformer quality, because its array of tests are designed to test for physical plausibility. A molecule is only considered POSEBUSTERS-valid if it passes all tests. In existing generative models for 3D molecule generation, most other metrics have been saturated (Irwin et al., 2025).

Training We train three TABASCO models at three sizes: Both TABASCO-mild (3.7M params.) and TABASCO-hot (15M params.) were trained on two 80GB A100 GPUs for 36 hours at a learning rate of 0.001. TABASCO-spicy (59M params.) was trained on the same resources for 72 hours with a learning rate of 0.0005 (see Appendix C). During training, we augment each batch with 8 random rotations of the same molecules to improve equivariance. We apply Exponential Moving Averaging (EMA) with decay strength 0.999 to the model weights, which we ablate in Section 4.4. We compare our models against EQGAT-diff (Le et al., 2023), FlowMol (Dunn & Koes, 2024), SemlaFlow (Irwin et al., 2025), and ADiT (Joshi et al., 2025) (see Appendix A).

4.2. TABASCO Achieves High Physical Quality

Our main results are shown in Table 1, example molecules are shown in Figure 3. TABASCO-spicy (59M), surpasses all prior methods in physical plausibility, raising the POSEBUSTERS validity from the previous state-of-the-art of 0.88 to 0.92 (see Figure 1). Interestingly, most of the gain is achieved by the 15M parameter TABASCO-hot variant, with only modest improvements from further scaling to 59M, suggesting diminishing returns beyond this point. All variants maintain strong molecular diversity ($\sim$0.89), indicating that architectural simplifications do not compromise sampling breadth and generalisation. Earlier models such as FlowMol, which performed well on traditional metrics, show significantly lower physical validity (0.64), further highlighting the need for domain-aware evaluation such as POSEBUSTERS. Despite their simplicity, TABASCO models generate molecules up to 100$\times$ faster than some prior baselines, offering a practical advantage for large-scale or iterative workflows. Finally, we find that guidance modestly improves POSEBUSTERS validity to 0.94, matching the training dataset, though at a 7$\times$ increase in compute

Table 1: Results on GEOM-Drugs. We generate 1,000 molecules for each method. [α]Due to computational constraints, we evaluate statistics on GEOM-Drugs on a random subset of 20K training molecules.

Method	# Params.	Validity ↑	Novelty ↑	Diversity ↑	POSEBUSTERS ↑	Strain Energy ↓	Time ↓ (s)
GEOM-Drugs[α]	-	1.0	0.0	0.90	0.94	-	-
EQGAT-diff	12M	0.94	0.94	0.90	0.84	360.19	4310.94
FlowMol	4.3M	0.81	0.81	0.91	0.64	34.20	362.22
SemlaFlow	22M	0.93	0.93	0.91	0.88	18.20	201.22
ADiT	150M	0.98	0.97	0.91	0.86	46.36	521.21
TABASCO-mild	3.7M	0.95	0.93	0.89	0.85	21.32	5.9
TABASCO-hot	15M	0.98	0.93	0.88	0.91	14.16	10.67
TABASCO-spicy	59M	0.97	0.90	0.89	0.92	15.07	19.77
TABASCO-spicy w/ guidance	59M	0.97	0.92	0.89	0.94	19.23	131.80

cost, suggesting its use may be best reserved for high-value targets or post hoc filtering.

4.3. The Effect of Sequential Ordering

Empirically, we observe that introducing sequence positional encodings yields higher quality molecules compared to treating atoms in a bag-of-words fashion. We compare the effect of positional encodings across several model scales (see Figure 4). We also show examples of failure modes we observed repeatedly in models without positional encodings in Figure 5. We hypothesize that this difference in generative quality may stem from early steps in molecule denoising, when the atomic coordinates are very noisy and positional encodings can provide a signal about relative positions. To test this, we sample from two 15M parameter TABASCO-models, one trained with sinusoid encodings and one without any encodings. We partially noise molecules to different $t \in [\tau, 1]$ and finish the denoising process with the models from that point. We do this to better isolate the sampling dynamics of the models at different noise levels.

Figure 4 shows how as τ increases, the performance difference of the models decreases and switches near the end of denoising. This suggests that the sampling trajectories in the model with positional encodings differ from those the model is trained on (see Appendix B). Furthermore, the higher POSEBUSTERS validity of the positional-encoding-free model towards the end of denoising suggests that in later stages of denoising its sampling trajectory is well aligned with training trajectories. This indicates that it is able to create high quality molecules when the final atom positions are apparent from its noisy coordinates, implying that as the final relative positions of atoms become more evident, positional encodings become less relevant.

4.4. Ablations

Table 2 summaries how three "optional" components affect the 15M-parameter TABASCO-hot model on GEOM-Drugs. Removing weight-EMA has almost no effect: all headline metrics change by ≤ 0.01 and diversity rises slightly. This shows that the model does not rely on EMA for chemical or geometric correctness. Performance is more sensitive to the coordination between input and output coordinate features. Eliminating the single cross-attention block lowers validity and novelty by $\sim$0.04 and, critically, drops POSEBUSTERS validity to 0.80. This indicates that coupling atom-type and coordinate information is necessary to resolve steric clashes and strain at this parameter scale. Positional encodings also prove critical. Without them, raw validity remains high (0.93) but POSEBUSTERS validity collapses to 0.70, revealing widespread geometric artifacts. The model can still generate chemically plausible graphs, yet struggles to arrange them in physically realistic 3D space (see Figure 5). In short, TABASCO's competitive accuracy does not depend on heavy symmetry priors, but it does require sequence position cues and cross-attention with the latent inputs; the EMA has negligible impact on performance. In Appendix B we further investigate the effect of removing random rotations alltogether, reducing the number of sampling steps and modifying the sampling strategy.

4.5. Evaluating Equivariance

We evaluate the quality of TABASCO's equivariance, as this is not encoded into the architecture. Similarly to previous work, we measure the deviation of the models prediction under random rotations (Karras et al., 2021; Bouchacourt et al., 2021). To control for numerical inaccuracies during sampling, rather than measuring the equivariance of fully denoised molecules, we measure the equivariance of the endpoint predictions of the model by partially noising molecules to different $t \in [0, 1]$. Given noisy coordinates $\mathbf{x}_t$ at time t, a random rotation R, and a function that at any timestep predicts the endpoint $\hat{\mathbf{x}}_1 = f(\mathbf{x}_t, t)$, we randomly rotate the input and apply the inverse rotation to the output, i.e. $Z(\mathbf{x}_t, t, R) = R^\top f(R\mathbf{x}_t, t)$. We estimate the relative equivariance error with

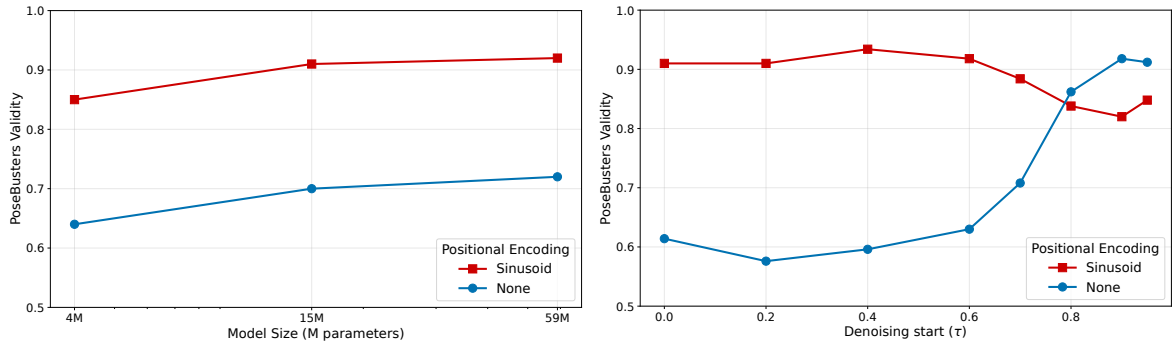

Figure 4: Left: Model performance across parameter scales with/without positional encodings. Right: 15M parameter model with/without positional encodings, POSEBUSTERS when starting denoising from different noise levels on test molecules.

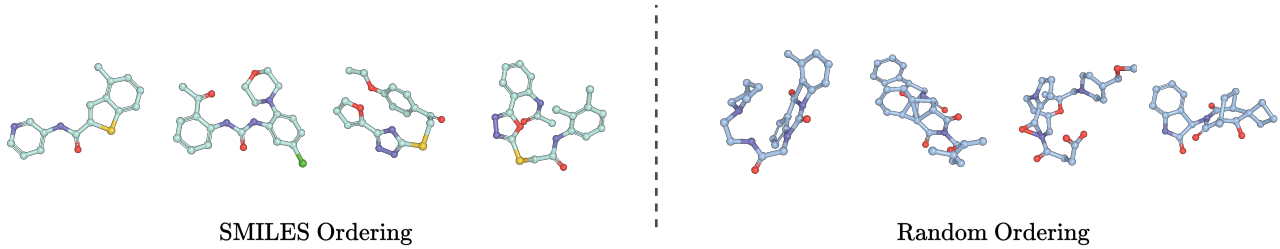

Figure 5: **Importance of atom ordering in TABASCO**. Molecules generated with sinusoidal positional encodings from SMILES order (left) are coherent and valid, while random atom ordering (right) yields more fragmented, implausible structures, highlighting SMILES' inductive bias for local structure during early denoising.

$$\epsilon_{\text{equiv}} = \text{Var}_R \left[\frac{Z(\mathbf{x}_t, t, R)}{||Z(\mathbf{x}_t, t, R)||} \right] . \tag{8}$$

We normalize the endpoint prediction per atom within random rotations of the same molecule to account for changes in scale during the sampling process and differing vector magnitudes. In an equivariant model one would have $Z(\mathbf{x}_t, t, R) = f(\mathbf{x}_t, t)$, which would trivially yield $\epsilon_{\text{equiv}} = 0$. In Figure 6 we show that the relative equivariance error is small across all t, and decreases further as denoising progresses. We observe significant differences in the relative equivariance error when comparing models with and without positional encodings of up to an order of magnitude at different points in sampling.

4.6. Effect of Physically-Constrained Guidance

In Table 3 we compare physically-constrained guidance to UFF relaxation of unguided molecules. Molecules are jointly denoised up to $t = 0.99$. In one experiment we allow for unconstrained relaxation and in another introduce a movement constraint of 0.1Å on each atoms original location. We choose $\alpha_{phys} = 0.01$ in all experiments.
Table 3 shows how distance-bounds guidance improves POSEBUSTERS validity, while preserving diversity and slightly increasing strain energy. Although distance-bounds guidance increases sampling time due to sequential bound computation and backpropagation, overall sampling remains faster than in prior approaches. We also observe that the incremental nature of the method largely preserves atom positions compared to the unguided baselines. Finally, we argue that this refinement preserves diversity, because the molecular hypothesis is essentially fixed by $t = 0.99$ and the model is only guided to create a lower-energy conformer of the same molecule.
The approach is model-agnostic and applies to any diffusion- or flow-based 3D generator that exposes gradients with respect to atom coordinates. Unlike force-conditioned samplers such as DiffForce (Kulytė et al., 2024), which back-propagate full molecular-mechanics gradients onto every atom at every reverse step, our method enforces pre-tabulated element-specific distance bounds and needs *no* energy evaluation or additional learnable parameters.

4.7. Discussion

Our findings align with recent trends toward simpler architectures, as seen in AlphaFold3's success without explicit symmetry constraints: Although SE(3) equivariance is often considered essential, our non-equivariant model learns equivariant representations up to small errors and

Table 2: Ablation study of model performance when removing components. Layer counts are adjusted to match model size where needed. Higher values are better.

Method	Validity	Novelty	Diversity	POSEBUSTERS
TABASCO-hot	0.98	0.93	0.88	0.91
w/o EMA	0.98	0.93	0.89	0.91
w/o cross-attention	0.94	0.89	0.89	0.80
w/o positional encoding	0.93	0.93	0.91	0.70

Table 3: Effect of distance-bounds guidance on POSEBUSTERS validity and runtime (single A100) on TABASCO-hot (15M) for 1000 molecules. RMSD is evaluated as a per-molecule mean with respect to the unguided baseline. $^{\gamma}$UFF calculations were performed on an M3 MacBook Pro.

Method	POSEBUSTERS↑	Strain Energy ↓	Diversity ↑	RMSD	Runtime ↓
Baseline	0.91	14.16	0.88	-	10.67
w/ UFF	0.94	4.74	0.88	0.226	14.21$^{\gamma}$
w/ Constr. UFF	0.93	11.15	0.88	0.084	23.42$^{\gamma}$
w/ guidance	0.94	19.23	0.89	0.132	75.65

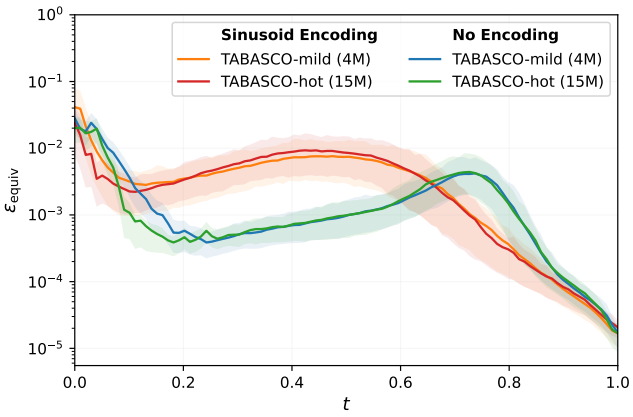

Figure 6: Relative equivariance error for TABASCO-hot (15M) and -mild (3.7M) with and without positional encodings. The error is normalized to the average atom coordinate magnitude.

achieves state-of-the-art performance on physical plausibility benchmarks, suggesting that enforced symmetries may be restrictive for some generation tasks. Our stripped-down architecture is easily extensible and only models coordinates and atom types explicitly. Conversely, omitting explicitly modelled bonds can limit conditioning when aiming to enforce valences or bond types (Peng et al., 2024). Physically-constrained guidance is shown to be effective for improving physical plausibility with minimal modifications, however compared to traditional methods like UFF Relaxation, it remains more expensive and converges to higher strain energies. Nevertheless, unguided generation yields a ten-fold speed improvement compared to previous methods, potentially making practical applications like large-scale virtual screening more feasible in the future.

5. Conclusion

In this work we present TABASCO, a non-equivariant generative model for 3D small molecule design that exhibits enhanced scalability and performance on physical plausibility compared to baselines. We study the importance of positional embeddings and atom ordering for small molecules, investigate the emergent equivariant properties of our model and the effects of scaling the model to large sizes. We note that minor effects emerged at scale unlike for previous work on other molecular modalities (Abramson et al., 2024; Geffner et al., 2025), and highlight the important design elements that are conducive to model performance. We hope that our model serves as a compelling example of how minimalist architectures can be effectively applied to molecular design and that our code base acts as an extensible tool for integration in drug-discovery workflows, through conditioned generation on relevant modalities or RL-based property optimization.

6. Societal Impact

Our model enables faster and more accessible generation of physically plausible molecular structures, supporting applications in drug discovery and computational chemistry.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630 (8016):493–500, 2024.

Albergo, M. S. and Vanden-Eijnden, E. Building normalizing flows with stochastic interpolants. *arXiv preprint arXiv:2209.15571*, 2022.

Axelrod, S. and Gomez-Bombarelli, R. Geom, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data*, 9(1):185, 2022.

Bouchacourt, D., Ibrahim, M., and Morcos, A. S. Grounding inductive biases in natural images:invariance stems from variations in data, 2021. URL https://arxiv.org/abs/2106.05121.

Buttenschoen, M., Morris, G. M., and Deane, C. M. Posebusters: Ai-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15(9):3130–3139, 2024.

Campbell, A., Yim, J., Barzilay, R., Rainforth, T., and Jaakkola, T. Generative flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design, 2024. URL https://arxiv.org/abs/2402.04997.

Du, Y., Jamasb, A. R., Guo, J., Fu, T., Harris, C., Wang, Y., Duan, C., Liò, P., Schwaller, P., and Blundell, T. L. Machine learning-aided generative molecular design. *Nature Machine Intelligence*, 6(6):589–604, 2024.

Dunn, I. and Koes, D. R. Mixed continuous and categorical flow matching for 3d de novo molecule generation, 2024. URL https://arxiv.org/abs/2404.19739.

Geffner, T., Didi, K., Zhang, Z., Reidenbach, D., Cao, Z., Yim, J., Geiger, M., Dallago, C., Kucukbenli, E., Vahdat, A., et al. Proteina: Scaling flow-based protein structure generative models. *arXiv preprint arXiv:2503.00710*, 2025.

Harris, C., Didi, K., Jamasb, A. R., Joshi, C. K., Mathis, S. V., Lio, P., and Blundell, T. L. Posecheck: Generative models for 3d structure-based drug design produce unrealistic poses. 2023.

Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. Equivariant diffusion for molecule generation in 3D. In Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., and Sabato, S. (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 8867–8887. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/hoogeboom22a.html.

Hua, C., Luan, S., Xu, M., Ying, R., Fu, J., Ermon, S., and Precup, D. Mudiff: Unified diffusion for complete molecule generation, 2024. URL https://arxiv.org/abs/2304.14621.

Irwin, R., Tibo, A., Janet, J. P., and Olsson, S. Semlaflow – efficient 3d molecular generation with latent attention and equivariant flow matching, 2025. URL https://arxiv.org/abs/2406.07266.

Joshi, C. K., Fu, X., Liao, Y.-L., Gharakhanyan, V., Miller, B. K., Sriram, A., and Ulissi, Z. W. All-atom diffusion transformers: Unified generative modelling of molecules and materials, 2025. URL https://arxiv.org/abs/2503.03965.

Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., and Aila, T. Alias-free generative adversarial networks, 2021. URL https://arxiv.org/abs/2106.12423.

Kulytė, P., Vargas, F., Mathis, S. V., Wang, Y. G., Hernández-Lobato, J. M., and Liò, P. Improving antibody design with force-guided sampling in diffusion models. *arXiv preprint arXiv:2406.05832*, 2024.

Le, T., Cremer, J., Noe, F., Clevert, D.-A., and Schütt, K. Navigating the design space of equivariant diffusion-based generative models for de novo 3d molecule generation. *arXiv preprint arXiv:2309.17296*, 2023.

Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling, 2023. URL https://arxiv.org/abs/2210.02747.

Morehead, A. and Cheng, J. Geometry-complete diffusion for 3d molecule generation and optimization, 2024. URL https://arxiv.org/abs/2302.04313.

Peng, X., Guo, R., Xu, Y., Guan, J., Jia, Y., Huang, Y., Zhang, M., Peng, J., Sun, J., Han, C., et al. Decipher fundamental atomic interactions to unify generative molecular docking and design. *bioRxiv*, pp. 2024–10, 2024.

Ramakrishnan, R., Dral, P. O., Rupp, M., and Von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.

Schneuing, A., Igashov, I., Dobbelstein, A. W., Castiglione, T., Bronstein, M. M., and Correia, B. Multi-domain distribution learning for de novo drug design. In *The Thirteenth International Conference on Learning Representations*.

Schneuing, A., Du, Y., Harris, C., Jamasb, A., Igashov, I., Du, W., Blundell, T., Lió, P., Gomes, C., Welling, M., et al. Structure-based drug design with equivariant diffusion models. *arXiv preprint arXiv:2210.13695*, 2022.

Song, Y., Gong, J., Xu, M., Cao, Z., Lan, Y., Ermon, S., Zhou, H., and Ma, W.-Y. Equivariant flow matching with hybrid probability transport, 2023. URL https://arxiv.org/abs/2312.07168.

Tong, A., Fatras, K., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Wolf, G., and Bengio, Y. Improving and generalizing flow-based generative models with minibatch optimal transport. *arXiv preprint arXiv:2302.00482*, 2023.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Vignac, C., Osman, N., Toni, L., and Frossard, P. Midi: Mixed graph and 3d denoising diffusion for molecule generation. In *Machine Learning and Knowledge Discovery in Databases: Research Track: European Conference, ECML PKDD 2023, Turin, Italy, September 18–22, 2023, Proceedings, Part II*, pp. 560–576, Berlin, Heidelberg, 2023. Springer-Verlag. ISBN 978-3-031-43414-3. doi: 10.1007/978-3-031-43415-0_33. URL https://doi.org/10.1007/978-3-031-43415-0_33.

Wang, Y., Elhag, A. A., Jaitly, N., Susskind, J. M., and Bautista, M. A. Swallowing the bitter pill: Simplified scalable conformer generation, 2024. URL https://arxiv.org/abs/2311.17932.

Weininger, D., Weininger, A., and Weininger, J. L. Smiles. 2. algorithm for generation of unique smiles notation. *Journal of chemical information and computer sciences*, 29(2):97–101, 1989.

Xu, C., Wang, H., Wang, W., Zheng, P., and Chen, H. Geometric-facilitated denoising diffusion model for 3d molecule generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(1):338–346, March 2024. ISSN 2159-5399. doi: 10.1609/aaai.v38i1.27787. URL http://dx.doi.org/10.1609/aaai.v38i1.27787.

Xu, M., Powers, A., Dror, R., Ermon, S., and Leskovec, J. Geometric latent diffusion models for 3d molecule generation, 2023. URL https://arxiv.org/abs/2305.01140.

A. Comparison to Previous Work

For all compared baselines we sample 1000 molecules with three random seeds on an A100 GPU. We report averages over the three runs.

EQGAT-diff We evaluated EQGAT-diff using the official codebase on GitHub[1] and the checkpoints linked there. We used the example evaluation script, which we edited to save molecules as outputted from reverse sampling, without any post-processing.

FlowMol We used the official implementation and the linked checkpoints on GitHub[2] and sampled molecules using the default script. For both GEOM[3] and QM9 (Ramakrishnan et al., 2014) we benchmarked againts the CTMC-based models.

Semla Flow We evaluated SemlaFlow using the sampling script and model checkpoints from GitHub[4]. We modified the sampling script to save all outputs from the model, as opposed to only valid molecules.

ADiT We benchmark ADiT by evaluating the molecules provided in the paper's GitHub repository, accessed on June 28th 2025[5].

Performance Comparison on QM9 We further train and evaluate our model on the benchmark dataset QM9 (Ramakrishnan et al., 2014) and compare the performance to previous methods in Table 4. We observe that TABASCO achieves very low novelty scores on QM9 along with POSEBUSTERS close to one. Given that QM9 represents a subset of physically plausible molecules from an enumeration containing up to nine heavy atoms, the low novelty score is not surprising, and is a testament to the fact that our model's outputs are confined within the constraints of physical plausibility. We further note that on QM9 adding positional encodings still helps with performance, but the performance gap is much smaller compared to TABASCO on GEOM-Drugs. A possible explanation for this is that the much smaller number of atoms per molecule compared to GEOM-Drugs makes it easier to distinguish atoms and place them with respect to each other even without positional encodings.

Table 4: Results on the QM9 Dataset

Method	# Params.	Validity ↑	Novelty ↑	Diversity ↑	POSEBUSTERS ↑	Strain Energy ↓
EQGAT-diff	12M	0.99	0.99	0.89	0.94	9.10
FlowMol	4.3M	0.97	0.97	0.92	0.92	17.81
SemlaFlow	22M	0.99	0.99	0.89	0.95	4.69
TABASCO-mild	3.7M	0.98	0.31	0.91	0.98	2.31
TABASCO-hot	15M	0.99	0.32	0.92	0.99	3.32
w/o pos. encodings	15M	1.00	0.34	0.93	0.93	17.10

B. Further Ablations

Time Distribution During Training Based on success in previous works, we choose the Beta-distribution for sampling the time t during training (Irwin et al., 2025; Geffner et al., 2025). We investigate the effect of different α values for the training time distribution $\text{Beta}(\alpha, 1)$ and ablate three values in Table 5. We observe significant changes in performance at sampling time when shifting the probability weight assigned to different times $t \in [0, 1]$ during training, and empirically find that $\text{Beta}(1.8, 1)$ yields the best results for our purpose.

[1] `https://github.com/jule-c/eqgat_diff/`, available under the MIT License

[2] `https://github.com/Dunni3/FlowMol/`, available under the MIT License

[3] Axelrod, Simon, et al. "GEOM, energy-annotated molecular conformations for property prediction and molecular generation." *Sci Data* 9, 185 (2022). Available under CC0 1.0.

[4] `https://github.com/rssrwn/semla-flow/`, available under the MIT License

[5] https://github.com/facebookresearch/all-atom-diffusion-transformer

Table 5: Ablation of α in the training time t distribution Beta$(\alpha, 1)$ on TABASCO-hot (15M) trained on GEOM-Drugs.

α	**Validity** ↑	**Novelty** ↑	**Diversity** ↑	**POSEBUSTERS** ↑
1.5	0.96	0.92	0.89	0.84
1.8	0.98	0.93	0.88	0.91
2.0	0.97	0.93	0.88	0.89

Table 6: Effect of four possible $g(t)$ parameterizations on TABASCO trained on GEOM-Drugs. For all $t > 0.9$ we set $g(t) = 0$ and use $\epsilon = 0.01$.

$g(t)$	**Validity** ↑	**Novelty** ↑	**Diversity** ↑	**POSEBUSTERS** ↑
0	0.96	0.95	0.90	0.83
$\frac{1}{t+\epsilon}$	0.98	0.93	0.88	0.91
$\frac{1}{t^2+\epsilon}$	0.97	0.93	0.89	0.91
$\frac{1-t}{t+\epsilon}$	0.98	0.94	0.89	0.89
w/o positional encodings				
0	0.94	0.93	0.91	0.69
$\frac{1}{t+\epsilon}$	0.89	0.87	0.91	0.26

Stochasticity Ablation Based on the approach in (Geffner et al., 2025), we investigate several choices for $g(t)$ in Eq. 7. Throughout this work, except when explicitly stated otherwise, we set $g(t)$ to zero for $t > 0.9$ to allow for precise placement of atoms towards the end of sampling. Table 6 compares the effect of four possible stochasticity functions. We observe that except for $g(t) = 0$ performance is very similar across all metrics. We trace the contrast between $g(t) = 0$ and the other parameterizations to a difference in sampling trajectories and show a comparison in Figure 7. In contrast to the trajectory of $g(t) = 0$ which is consistent with the training trajectory, the rest of the $g(t)$ functions have very large magnitudes close to $t = 0$, which empirically leads first to an explosion and then to a collapse of the atom vector magnitudes. In the collapsed state atoms are roughly arranged in a sequence and slowly grow into the finished molecule as $t \to 1$ (see Figure 7). This sudden rearrangement and growing into the finished molecule appears to yield better final molecules compared to when following the training trajectories more closely. This may also help explain the dip in POSEBUSTERS validity during partial molecule noising of TABASCO with positional encodings in Figure 4. In contrast to this, we observe in Table 6 that this explosion and collapse behaviour leads to much worse molecules when positional encodings are not added to the model, possibly because in the collapsed state atom coordinates are almost identical and they become very hard to distinguish without the positional encodings.

Number of Steps at Sampling We investigate the effect of reducing the number of sampling steps on molecule quality and ablate over several choices in Table 7. We observe that as little as 40 steps are necessary for TABASCO-hot to outperform previous methods on POSEBUSTERS. We further observe that additional steps have no effect on molecular quality.

Noise Scaling We investigate several values for γ to ablate the effect of noise scaling. In Figure 8 we compare POSEBUSTERS-validity for different noise scales and different $g(t)$ parameterizations. We observe that molecular quality remains high over several noise scales, and then collapses for $g(t) = \frac{1}{t+\epsilon}$.

The Effect of Random Rotations We study the effect of augmenting the data with random rotations during training on the sampled molecules. The batches in normal model training are augmented with seven copies of the same molecules (i.e. the effective batch size increases to eight times the original). All molecules in the augmented batch are then subjected individually to a random rotation. We study the effect of this operation by training two 15M TABASCO-hot models, one where no augmentations and only random rotations are applied, and one where neither augmentations nor rotations are applied. To approximately match the original training dynamics, we increase the batch size of these models to match the effective batch size of the original training procedure. We train the models with the same compute budget as previously

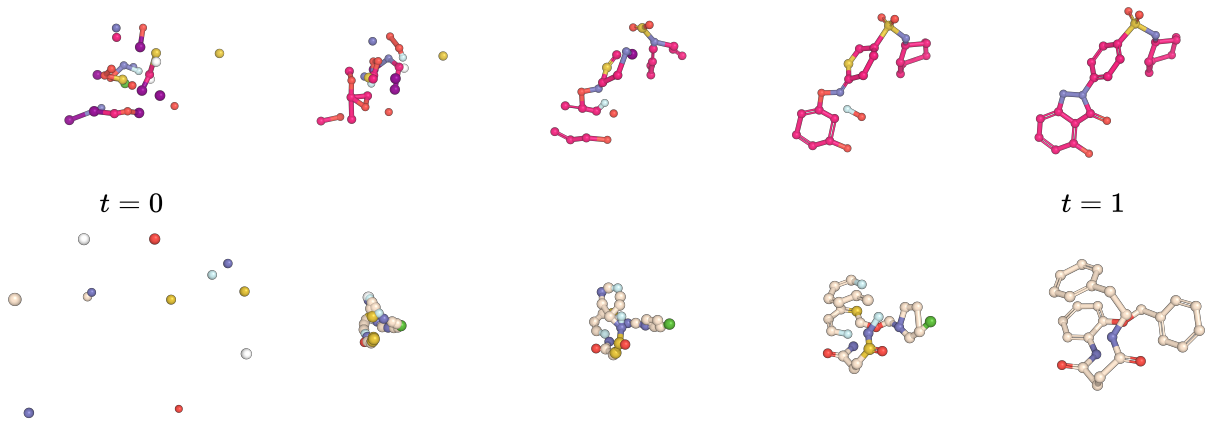

Figure 7: Snapshots of the sampling trajectories for two different molecules sampled from TABASCO-hot (15M) trained on GEOM-Drugs. The upper trajectory is sampled with $g(t) = 0$ and the lower one with $g(t) = \frac{1}{t+0.01}$

Table 7: Number of steps at sampling for TABASCO-hot (15M) trained on GEOM-Drugs. We additionally evaluate connectivity, which denotes the fraction of fully connected molecules.

# Steps	Validity ↑	Novelty ↑	Connectivity ↑	POSEBUSTERS ↑
10	0.99	0.98	0.00	0.00
20	1.00	0.99	0.00	0.00
30	0.98	0.97	0.99	0.81
40	0.98	0.94	0.99	0.91
50	0.99	0.95	1.00	0.91
100	0.98	0.94	1.00	0.91
200	0.98	0.93	1.00	0.89
500	0.98	0.94	1.00	0.91

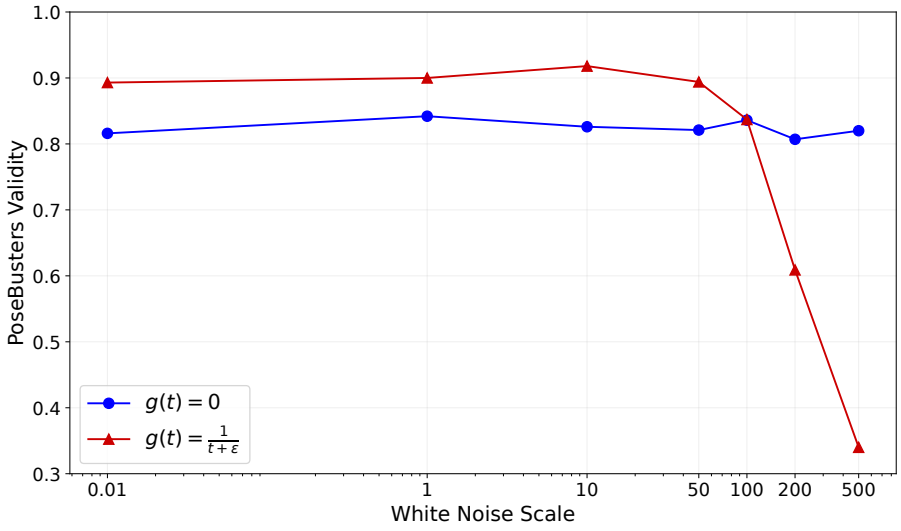

Figure 8: Comparison of POSEBUSTERS-validity across noise scales γ with different $g(t)$. In contrast to all other comparisons we set $g(t) = 0$ only beyond $t > 0.95$ to further augment the effect of adding noise.

Table 8: Performance comparison of additional runs trained on GEOM-Drugs without per-batch random rotations and without any random rotation.

Configuration	Validity ↑	Novelty ↑	Diversity ↑	POSEBUSTERS ↑
TABASCO-hot (15M)	0.98	0.93	0.89	0.91
w/o positional encoding	0.93	0.93	0.91	0.70
w/o batch augmentations	0.98	0.94	0.88	0.89
w/o random rotations	0.98	0.94	0.89	0.90

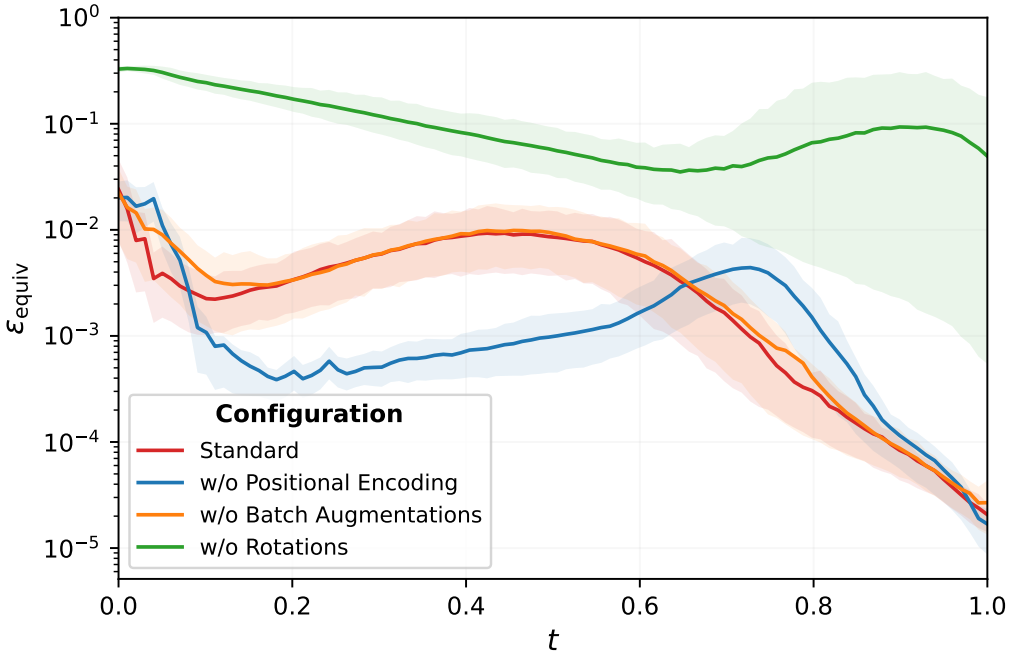

Figure 9: Comparison of the equivariance error to runs trained without per-batch random augmentations and without any random augmentation for TABASCO-hot (15M) trained on GEOM-Drugs.

allotted: two A100 GPUs over 36 hours. We compare the observed performance in Table 8 and visualize the equivariance error over time in Figure 9.

The results show how omitting any random rotations of the data leads to a high-performing model that has a significantly higher equivariance error than all other models. Simultaneously, randomly rotating data, but omitting intra-batch augmentations with further random rotations, does not worsen the equivariance error, but slightly hurts POSEBUSTERS performance. This suggests that random intra-batch augmentations do not improve model equivariance, but improve training, possibly because of higher-quality gradient steps induced by the random rotations. Finally, we verify that random rotations are also not strictly necessary to create models that generate high-quality molecules, though at the expense of a significantly worse equivariance error.

C. Extended Details on Models

We give an overview of key design features of TABASCO in Figure 2 and describe the unconditional sampling algorithm in detail in Algorithm 1. In this section we further elaborate on model architecture and give concrete values for relevant hyperparameters in Table 9.

Atom coordinates are encoded with a bias-free linear layer that scales to the model's hidden size. Discrete atom types are encoded through an embedding layer, where we model Carbon, Nitrogen, Oxygen, Fluorine, Sulfur, Chlorine, Bromine,

Table 9: Model hyperparameters across different model sizes

Hyperparam.	**TABASCO-mild**	**TABASCO-hot**	**TABASCO-spicy**
# Params.	3.711.369	14.795.529	59.082.249
Hidden size	128	256	512
# Transformer blocks	16	16	16
# Attn. heads	8	8	8
Train t distribution	Beta(1.8,1)	Beta(1.8,1)	Beta(1.8,1)
$\lambda_{\text{discrete}}$	0.1	0.1	0.1
Learning rate	0.001	0.001	0.0005
Optimizer	Adam	Adam	Adam
EMA-weight	0.999	0.999	0.999
Batch size	256	256	128
# Rotation Augs.	8	8	8
Effective batch size	2048	2048	1024
# GPUs	2	2	2
Training Duration	36h	36h	72h
# Sampling Steps	100	100	100
$g(t)$	$\frac{1}{t+0.01}$	$\frac{1}{t+0.01}$	$\frac{1}{t+0.01}$
γ	0.01	0.01	0.01

Iodine and a miscellaneous "$*$" atom, for all elements in the training set not contained within the previous list. We encode the time $t \in [0, 1]$ through a Fourier encoding, and each atoms location within the molecules's SMILES sequence through a standard sinusoid encoding. We tried concatenating these four vectors and creating a combined hidden representation with an MLP mapping from $\mathbb{R}^{4 \times \text{hidden dim}} \to \mathbb{R}^{\text{hidden dim}}$ but observed no difference in practice to simply adding the vector representations, and thus opted for this approach. Each transformer block applies layer-norm to the activations, then PyTorch multi-head attention, another layer-norm and a transition layer, where we include residual connections between the first two and second two components. This output is processed by two parallel PyTorch cross-attention layers one for atom types and for coordinates. Each consists of a self-attention block, a multi-head attention block, where the original combined hidden representation is used as key and value, and a feed-forward block. Both outputs are subsequently processed through domain-specific MLPs where the output atom coordinate MLP is also bias-free.

D. Further Analysis on Physical Guidance

Table 10: Comparison POSEBUSTERS validity when adding physically-constrained guidance. Evaluated on 1000 molecules on a single A100 GPU.

Method	**# Params.**	**Validity** ↑	**Novelty** ↑	**Diversity** ↑	**POSEBUSTERS** ↑	**Strain Energy** ↓	**Time** ↓ (s)
TABASCO-mild	3.7M	0.95	0.93	0.89	0.85	21.32	5.9
w/ guidance		0.96	0.95	0.89	0.91	26.53	60.86
TABASCO-hot	15M	0.98	0.93	0.88	0.91	14.16	10.67
w/ guidance		0.97	0.94	0.89	0.94	19.23	75.66
TABASCO-spicy	59M	0.97	0.90	0.89	0.92	15.07	19.77
w/ guidance		0.97	0.93	0.89	0.94	17.01	131.80

We provide a detailed description of our physically constrained guidance procedure in Algorithm 2 and provide full results on all model sizes in Table 10. We observe the largest improvement in POSEBUSTERS for TABASCO-small, and minor improvements in novelty for all model sizes. Simultaneously we consistently observe an increase in strain energy when applying physical guidance.

E. Limitations

The approach described in this work introduces several limitations. SMILES-derived positional encodings improve performance but can theoretically introduce systemic biases, that may limit the model when faced with unusual bond patterns or non-standard chemical structures. Furthermore, omitting explicit bond modeling creates a leaner model and simpler training objective, but limits control over valences and bond orders when sampling the model. While TABASCO exhibits emergent equivariance, in some areas such as molecular dynamics, where even small equivariance errors can prove problematic, this approximate equivariance may still be insufficient.

The physically-constrained guidance algorithm serves as a proof-of-concept for boosting the physical plausibility of molecules during sampling without requiring any modifications to training data, model architecture or parameter scale. Still, this approach is based on optimizing chemoinformatics heuristics for high-quality molecules and it dramatically increases sampling times.

Furthermore, while useful to quantify physical plausibility of 3D molecules, POSEBUSTERS cannot capture all aspects of molecular quality, and does not quantify additional very relevant metrics of interest: TABASCO does not address improvements in drug-likeness of molecules or synthetic accessibility.

Algorithm 1 Unconditional Sampling Algorithm

procedure EUCLIDEANSTEP($\mathbf{x}_t, \hat{\mathbf{x}}_1, t, \Delta t, g(\cdot), \gamma$)
  $\mathbf{v}_t \leftarrow \frac{1}{1-t}(\hat{\mathbf{x}}_1 - \mathbf{x}_t)$
  $\mathbf{s}_t \leftarrow g(t)\frac{t\mathbf{v}_t - \mathbf{x}_t}{1-t}$
  $\mathrm{d}W_t \leftarrow \sqrt{2\,\gamma\,g(t)}\,\mathcal{N}(0, I)$
  $\mathbf{x}_t \leftarrow (\mathbf{v}_t + \mathbf{s}_t + \mathrm{d}W_t)\Delta t$
  return $\mathbf{x}_t$
end procedure

procedure DISCRETEFLOWSTEP($a_t, \hat{p}_1, t, \Delta t$)
  $\mathbf{r}_t(i, \cdot) = \frac{\Delta t}{1-t}\hat{p}_1(i)$ ▷ $\hat{p}_1$ consists of softmax-normalized model logits
  $\mathbf{r}_t(i, a_t(i)) \leftarrow -\sum_{j \neq a_t(i)} \mathbf{r}_t(i, j)$ ▷ Make $\mathbf{r}_t$ zero mean
  $\mathbf{p}_{t+\Delta t}(i, j) \leftarrow 1_{a_t(i)=j} + \mathbf{r}_t(i, j)$
  $a_t(i) \leftarrow \text{Categorical}(\mathbf{p}_{t+\Delta t}(i, \cdot))$
  return a_t
end procedure

procedure SAMPLEMOLECULE($f, \{t_i\}_{i=0}^N$)
  $\mathbf{x} \leftarrow \mathcal{N}(0, I)$
  $a \leftarrow \text{Categorical}\left(\delta(\frac{1}{\#\text{ atom types}})\right)$
  for $i = 1$ **to** N **do**
    $\Delta t \leftarrow t_i - t_{i-1}$
    $(\hat{\mathbf{x}}_1, \hat{p}_1) \leftarrow \text{EndpointPrediction}(f, (\mathbf{x}, a), t_i)$
    $\mathbf{x} \leftarrow \text{EUCLIDEANSTEP}(\mathbf{x}_t, \hat{\mathbf{x}}_1, t_i, \Delta t)$
    $a \leftarrow \text{DISCRETEFLOWSTEP}(a_t, \hat{p}_1, t_i, \Delta t)$
  } SAMPLINGSTEP
  end for
  return $(\mathbf{x}, a)$
end procedure

Algorithm 2 Flow Matching with Physical Guidance

procedure PHYSICALGUIDANCE($f, (\mathbf{x}_t, a_t), t, \alpha$)
  $\hat{\mathbf{x}}_1, \hat{p}_1 \leftarrow$ EndpointPrediction($f, (\mathbf{x}_t, a_t), t$)
  $\hat{a}_1(i) = \text{argmax}_j\ \hat{p}_1(i, j)$
  bounds $\leftarrow$ GetPhysicalConstraints($\hat{\mathbf{x}}_1, \hat{a}_1$) ▷ Calls RDKIT `GetBoundsMatrix()`
  for each atom pair $(\mathbf{x}_{t,i}, \mathbf{x}_{t,j})$ in $\mathbf{x}_t$ **do** ▷ This nested loop is vectorized in practice
    $d_{ij} \leftarrow ||\mathbf{x}_{t,i} - \mathbf{x}_{t,j}||_2^2$
    if $d_{ij} < \text{bounds}_{ij}^{\min}$ **then** ▷ Can also regress towards the interval centre
      $\mathcal{L} \leftarrow \mathcal{L} + (d_{ij} - \text{bounds}_{ij}^{\min})^2$
    else if $d_{ij} > \text{bounds}_{ij}^{\max}$ **then**
      $\mathcal{L} \leftarrow \mathcal{L} + (d_{ij} - \text{bounds}_{ij}^{\max})^2$
    end if
  end for
  $\mathbf{x}_t \leftarrow \mathbf{x}_t - \alpha \cdot \text{sign}(\nabla_{\mathbf{x}_t}\mathcal{L})$ ▷ The sign-op slightly stabilizes updates in practice
  return $\mathbf{x}_t$
end procedure

procedure GUIDEDSAMPLING($f, (\mathbf{x}_0, a_0), \{t_i\}_{i=0}^N, t_{\text{guidance}}$)
  $(\mathbf{x}, a) \leftarrow (\mathbf{x}_0, a_0)$
  for $i = 1$ **to** N **do**
    $\Delta t \leftarrow t_i - t_{i-1}$
    if $t_i \geq t_{\text{guidance}}$ **then**
      $\mathbf{x} \leftarrow$ PHYSICALGUIDANCE($f, (\mathbf{x}, a), t_i, \alpha$)
    end if
    $(\mathbf{x}, a) \leftarrow$ SAMPLINGSTEP($f, (\mathbf{x}, a), t_i, \Delta t$)
  end for
  return $(\mathbf{x}, a)$
end procedure