

---

# Physics-informed generative model for drug-like molecule conformers

---

Anonymous Author(s)

Affiliation

Address

email

## Abstract

1 Diffusion-based methods have been successfully applied to molecule conformer  
2 generation using implicit physical modeling. In contrast, conventional, rules-  
3 based approaches employ an explicit physical model such as a classical force  
4 field parameterization. In order to combine the advantages of both approaches, we  
5 present a diffusion-based, physics-informed denoising model (PIDM) for conformer  
6 generation that is constructed from molecule subgraph patterns borrowed from  
7 classical force fields. The result is a model that is resistant to overfitting and  
8 explainable. Using recent advances in denoising score matching, we naturally  
9 separate the task of training and generation while providing a smooth transition  
10 between deterministic and stochastic generative schemes that adapt to any number  
11 of denoising steps. We demonstrate conformer generation quality that outperforms  
12 the current state-of-the-art while employing a fraction of parameters.<sup>1</sup>

## 13 1 Introduction

14 Conformer generation is the process of identifying a valid and useful set of atomic coordinates for  
15 a given molecule. Because it plays a crucial role in structure-based drug-discovery [1], over four  
16 decades of effort has been invested in conventional rules-based approaches [2, 3, 4, 5, 6, 7, 8, 9].  
17 Recent advances in generative techniques in deep learning, particularly in diffusion-based models  
18 of image generation [10, 11, 12, 13] and point-cloud generation [14, 15], suggest that generative  
19 techniques could lead to a renaissance in this field.

20 We describe in this work is a method of conformer generation using a physics-informed, denoising  
21 model (PIDM). By taking advantage of established methods employed in classical force fields, we  
22 have constructed a diffusion-based model that is explainable, transferable, and robust. Building  
23 upon recent theoretical advancements [16], we employ a flexible method of generation that smoothly  
24 adapts to either deterministic or stochastic modes and naturally scales to any number of denoising  
25 steps. This is the first time that a physics-informed approach has been applied in a denoising model  
26 of this type, the first that is explainable, and the first to feature deterministic generation. The result  
27 outperforms the current state-of-the-art while using ten times fewer steps. In addition, we demonstrate  
28 a proof-of-concept guided technique that permits targeted generation.

29 To build a useful tool, we need to generate molecule conformers that are physically meaningful. The  
30 naive approach is to directly train a generic, implicit model of physical viability. A drawback is that  
31 the number of possible drug-like compounds is too large (by some estimates, as large as  $10^{33}$  [17])  
32 to cover in a dataset. Another issue is that the number of viable conformations for each drug-like  
33 molecule is intractable in many cases [7]. Therefore, we are forced to limit ourselves to a suitable  
34 approach that can readily transfer to molecules and conformations outside a training set.

---

<sup>1</sup>Code is available at [TBA]

35 The first step is to define what constitutes a viable conformer. We use physics as a guide. To appreciate  
 36 what constitutes a local minima in molecular energy, consider classic force field parameterizations,  
 37 an established methodology [18]. Force fields are commonly formulated as the sum of “bonded”  
 38 and “nonbonded” contributions [19, 20, 21]. The former include contributions associated with  
 39 intramolecular bonds and the latter captures longer-distance interactions. The bonded contributions  
 40 in a classical force field are conventionally divided into four terms: bonds, bends, proper torsions and  
 41 improper torsions [19, 20, 21]. Each term is identified with a subgraph of a specific topology and one  
 42 characteristic scalar parameter (Fig 1). These terms will become important in the construction of our  
 43 model.

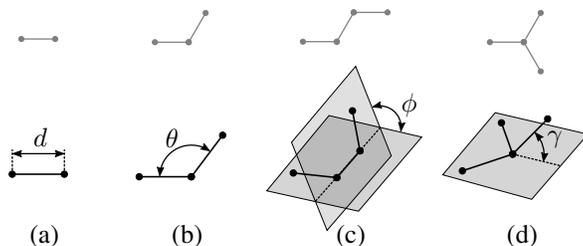


Figure 1: Force fields typically include bonded terms associated with (a) bond lengths, (b) bend angles, (c) proper torsions, and (d) improper torsions.

44 In the context of classical force fields, the set of viable molecule conformers are commonly defined  
 45 in terms of fixed bond lengths, bond angles, and improper torsions. The constraint on proper torsions  
 46 is cyclic in the angle  $\phi$  such that energy is commonly parameterized as a function of  $n\phi - \phi_0$ , where  
 47  $n$  is an integer and  $\phi_0$  is a reference angle. The freedom to select one of  $n$  torsion angles is one way  
 48 a molecule can exhibit multiple conformations. We will refer to such freedom as *torsional space*.

49 Nonbonded contributions to energies depend on external conditions, such as whether the molecule is  
 50 solvated or bound to a protein [22]. As such, in many applications, nonbonded energies are ignored or  
 51 attenuated during the process of conformer generation, to avoid biases [23]. We employ this strategy  
 52 in this work. It can be considered adequate for tools that perform their own search in torsional space,  
 53 such as flexible ligand docking [24, 25, 26, 27, 28].

54 In addition to locating energy minima, a useful conformer generator must respect stable stereochem-  
 55 istry, specifically chirality [29] and cis-trans isomerism [30].

## 56 1.1 Related Works

57 Conventional approaches use a variety of schemes [4, 5, 6, 7, 8, 9, 24]. Balloon [5] is a conformer  
 58 generator based on a multiobjective genetic algorithm. ETKDGV3 [9] is a knowledge-based generator  
 59 provided by RDKit [31] based on distance geometry. The OMEGA toolkit [7, 32] employs a fragment  
 60 library combined with rules-based sampling. Although widely adopted, conventional methods rely  
 61 on hand-tuned algorithms and are typically limited in accuracy due to an overdependence on classical  
 62 force fields parameterization.

63 Several strategies have been employed for learned models, such as energy gradients [33, 34, 35], Gibbs  
 64 sampling [36], and conditional variational encoders [37, 38]. The drawback of these approaches is  
 65 that the energy of disordered molecule systems is difficult to characterize directly due to singularities  
 66 and large energy barriers. GeoMol [39] learns local structure and applies incremental construction.  
 67 Since incremental construction is poorly suited to cycles, it fails to reproduce all but the simplest ring  
 68 systems. GeoDiff [40] is a state-of-the-art stochastic diffusion model. It follows conventions most  
 69 closely related to “denoising diffusion probabilistic models” (DDPM) [10], employs 793,858 weights  
 70 and uses 5,000 steps for generation. GeoDiff attempts to model nonbonded distances which requires  
 71 it to sample torsional space during training. This is not only undesirable (since torsional space is  
 72 physically ambiguous) but is also likely the reason for a high level of computational complexity.

73 Others have proposed methods that generate novel molecules in 3D space [41, 42, 43, 44, 45, 46],  
 74 which is a related but different task than reported here.

75 **2 Model Design**

76 Our goal is to construct a model that can generate acceptable conformers for any drug-like molecule  
 77 when provided with just the atom composition, connectivity, and stereochemistry. Recent work [16]  
 78 has demonstrated that the design of diffusion-based models for images can be generalized around  
 79 the concept of denoising score matching [47]. We apply this approach to conformer generation by  
 80 representing our model as a denoising function  $D$  that provides an estimate of the true coordinates  $\mathbf{x}$   
 81 of a molecule when provided with coordinates that have been smeared by a centered, uncorrelated  
 82 Gaussian of width  $\sigma$

$$\mathbf{x} \approx D(\mathcal{N}(\mathbf{x}; 0, \sigma^2 \mathbf{I}), \sigma; \mathbf{a}), \quad (1)$$

83 where  $\mathbf{a}$  is a suitable embedding that represents the composition of the molecule (*i.e.* its atom types  
 84 and connectivity).

85 The overall structure of the model is shown in Fig 2 and consists of two major components: a graph  
 86 transformer network to build a useful atom embedding and a series of bonded subcomponents whose  
 87 outputs are summed together for coordinate prediction.

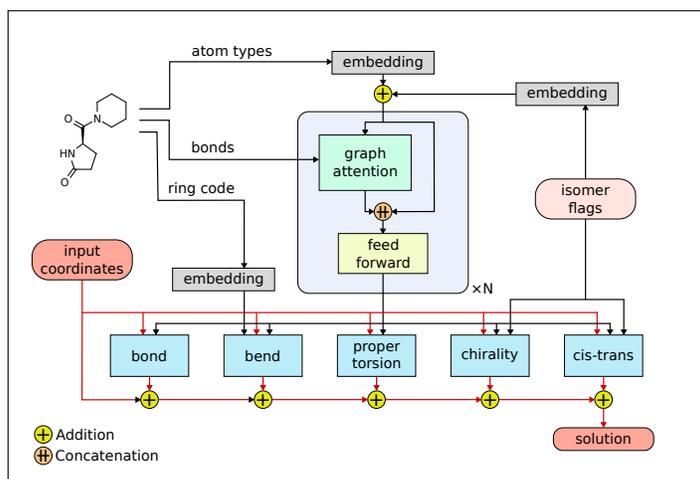


Figure 2: A schematic of the denoising model.

88 The purpose of the graph transformer network is to place the atoms of each molecule into a suitably  
 89 descriptive embedding space that can be employed by the bonded components. It starts with a  
 90 embedding in which atoms are distinguished by their element, formal charge, and hybridization. The  
 91 latter is taken from the algorithm built into the RDKit cheminformatics library [31]. All hydrogens  
 92 are treated as explicit. Atoms that are associated with explicit chirality or cis-trans isomerisms are  
 93 flagged by the addition of a global vector reserved for this purpose.

94 The initial atom embedding is refined by multiple layers of a graph transformer network, based on  
 95 GATv2 [48], and configured to use bonds as graph edges. No edge labeling (such as bond order) is  
 96 employed, since connectivity along with atom identity is sufficient to describe relevant chemistry [49].  
 97 To preserve a form of atom self-identity, rather than use self edges (a poor physical analog), the input  
 98 to the graph attention network is concatenated to the output.

99 Each of the bonded components have a similar structure in which a multilayer perceptron (MLP)  
 100 is used to calculate a correction to atom positions as a displacement along a vector. The details,  
 101 outlined in the supplemental materials, are summarized here. For the bond component, the vector  
 102 is the difference  $\delta_{ij}$  of the two atom positions. The input to the MLP is the concatenation after a  
 103 suitable normalization of  $\{|\delta_{ij}|, \sigma, \mathbf{a}_i, \mathbf{a}_j\}$ , where  $\mathbf{a}$  are the atom embeddings.

104 For the bend component, the vector  $\delta_{ik}$  is between the two outer atoms. The MLP is fed  
 105  $\{|\delta_{ik}|, \sigma, \mathbf{a}_i, \mathbf{a}_j, \mathbf{a}_k, \mathbf{c}_{ijk}\}$ , where  $\mathbf{c}_{ijk}$  is an embedding capturing the size of the rings (if any) to  
 106 which the bend belongs. The latter is important because the GATv2 graph network, like all message  
 107 passing networks, is incapable of detecting cycles [50].

108 For the proper torsion component, the vector  $\delta_{il}$  is between the two outer atoms. The MLP is fed  
 109  $\{|\delta_{il}|, \sigma, \mathbf{a}_i, \mathbf{a}_j, \mathbf{a}_k, \mathbf{a}_l, \sin \phi, \cos \phi\}$ . The torsional angle  $\phi$  is needed to account for torsional space.

110 The chirality component is constructed out of the improper torsions associated with each chiral  
 111 atom. The vector is the normal of the plane defined by three of the four atoms. The MLP is fed  
 112  $\{d_{\perp}, d_{\parallel}, \sigma, \mathbf{a}_i, \mathbf{a}_j, \mathbf{a}_k, \mathbf{a}_l\}$ , where  $d_{\perp}$  ( $d_{\parallel}$ ) is the distance of the fourth atom out of (along) the plane.  
 113 The cis-trans component is constructed out of the proper torsions associated with each constrained  
 114 bond. The vector is the difference  $\delta_{ijkl}$  between the average positions of the outer and inner atoms.  
 115 The MLP is fed  $\{|\delta_{ijkl}|, \chi_{ijkl}, \sigma, \mathbf{a}_i, \mathbf{a}_j, \mathbf{a}_k, \mathbf{a}_l\}$ , where  $\chi_{ijkl}$  is the requested cis-trans state.

116 For the models reported here, an atom embedding of dimension 50 is used throughout. Four graph  
 117 transformer layers are employed. The MLP for each component uses two hidden layers. The result is  
 118 a model with a total of 135,080 weights, with 63,480 reserved for the molecule graph and 71,240  
 119 in the geometry components. Experiments in increasing the atom embedding dimension or adding  
 120 additional transformer layers produced only marginal improvement in loss.

### 121 3 Datasets

122 We are interested in high-quality conformers of drug-like molecules. Two synthetic, publicly available  
 123 datasets fit this role: QMugs [51] and GEOM-drugs [52]. Each contain several hundred thousand  
 124 drug-like molecules with conformers optimized (in vacuum) using the GFN2-xTB semiempirical  
 125 quantum mechanical method [53, 54]. We discard about 2% of the GEOM-drugs molecules due to  
 126 conformer inconsistency at the graph level. Otherwise, we accept all conformers in both data sets.  
 127 Both are randomly divided into their own training (80%), validation (10%), and test (10%) subsets.

128 To measure molecule similarity, we use the Tanimoto measure applied to an ECFP6 fingerprint [55]  
 129 folded to 1024 bits. At a threshold of 0.9 (0.8), we find 5.6% (6.4%) of the compounds in QMugs  
 130 overlap with those in GEOM-drugs. Although the two datasets contain mostly different molecules,  
 131 the underlying physics will be the same. To test dataset independence, we have chosen to train two  
 132 versions of our model on the two corresponding training subsets, which we will label PIDM[QMugs]  
 133 and PIDM[GEOM-drugs]. Both models and others will be compared to a single benchmark dataset,  
 134 whose preparation is described below.

135 For the benchmark dataset, we started from the QMugs test subset. To ensure independence and  
 136 guard against data leakage, we filter each molecule against the contents of the QMugs training subset,  
 137 the entire GEOM-drugs dataset, and internally using a Tanimoto threshold of 0.7. The annotated  
 138 chirality and cis-trans isomerism are then validated against the public PubChem database [56]. We  
 139 also queried PubChem for a copy of the first ten of their generated conformers [23] for comparison.  
 140 Molecules that could not be validated or did not have a PubChem conformer were discarded. The  
 141 final result is 15,763 fully annotated, independent molecules reserved for benchmarks.

### 142 4 Training

143 In this work, we follow the modular scheme for diffusion-based models as proposed by Karras et  
 144 al. [16] in which training and generation are separate tasks that need not share the same noising  
 145 schedule. For the purposes of training, we chose to evenly sample from the set  $\{\sigma_1 \dots \sigma_N\}$ :

$$\sigma_i = \begin{cases} \left( \sigma_{\max}^{1/\rho} + \frac{i-1}{N-1} \left( \sigma_{\min}^{1/\rho} - \sigma_{\max}^{1/\rho} \right) \right)^{\rho} & 1 \leq i < N \\ 0 & i = N \end{cases}. \quad (2)$$

146 The total loss  $\mathcal{L}$  is calculated as the weighted sum of the contribution from each sample  $\sigma_i$

$$\mathcal{L} = \sum_{i=1}^N \frac{1}{\sqrt{\sigma_i^2 + \epsilon^2}} \mathcal{L}(\sigma_i) \quad (3)$$

147 with

$$\mathcal{L}(\sigma) = \mathbb{E}_{\mathbf{x} \sim \text{data}, \mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})} \| \mathbf{D}(\mathbf{x} + \mathbf{n}, \sigma) - \mathbf{x} \|_2^2. \quad (4)$$

148 Standard parameter values are  $N = 100$ ,  $\sigma_{\max} = 8 \text{ \AA}$ ,  $\sigma_{\min} = 10^{-5} \text{ \AA}$ ,  $\epsilon = 10^{-5} \text{ \AA}$ , and  $\rho = 6$ .

149 PIDM[QMugs] is trained using a fixed schedule of 100 epochs, taking approximately 50 hours on  
 150 a single RTX 3090 (Fig. 3a). PIDM[GEOM-drugs] is trained using a fixed schedule of 25 epochs,  
 151 taking approximately 170 hours (Fig. 3b).

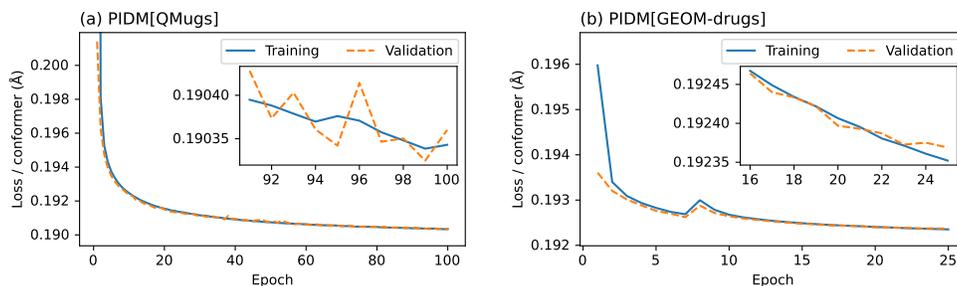


Figure 3: Loss per conformer as calculated during model training for (a) QMugs and (b) GEOM-drugs. Plotted are losses calculated for the training set and for an independent validation set of 1/8 the size.

152 Building the model with explicit components has the advantage of explainability. Once provided with  
 153 an example molecule, each of the bonded compounds can be probed for specific atoms, a process that  
 154 was invaluable during development. An example is shown in Fig 4. More examples are included in  
 155 the supplemental materials.

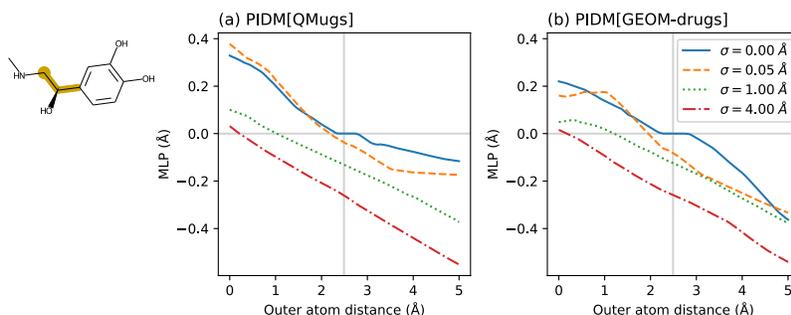


Figure 4: Example bend correction from (a) PIDM[QMugs] and (b) PIDM[GEOM-drugs] for an atom involved in the ethanol group of adrenaline. Corrections for various values of  $\sigma$  are plotted. The vertical gray line is the expected atom distance, obtained from a separate GFN2-xTB optimization. As the distance deviates from expectations, the model applies larger corrections. As  $\sigma$  approaches zero, the correction vanishes once the correct geometry is achieved. Despite being trained on independent data sets, the two models learn a similar behavior.

## 156 5 Generation

157 We adopt a score-based, probability-flow framework [13] in order to generate conformers from our  
 158 denoising model. As is typical in this approach, we consider a multidimensional Wiener process  
 159 applied to molecule coordinates  $\mathbf{x}$  over a time interval  $t \in [0, 1]$ :

$$p_t(\mathbf{y}(t)|\mathbf{x}; \sigma(t)) = \mathcal{N}(\mathbf{y}(t); \mathbf{x}, \sigma(t)^2 \mathbf{I}), \quad (5)$$

160 where  $\mathbf{y}(t)$  are the resulting random coordinates and  $\sigma(t)$  is a width schedule we are free to choose to  
 161 suit our task, with the only requirement that  $\lim_{t \rightarrow 0} \sigma(t) = 0$ . For generation, we start by sampling  
 162 from a random Gaussian distribution  $\mathcal{N}(0, \sigma(1)^2 \mathbf{I})$  as an approximation for  $\mathbf{y}(1)$  and solve for the  
 163 corresponding reverse process (denoising) to obtain  $\mathbf{y}(0)$  as a candidate solution for  $\mathbf{x}$ .

164 To construct a solution for the reverse process, we identify the marginal distribution  $p(\mathbf{y}; \sigma)$  as

$$p(\mathbf{y}; \sigma) = \int p_t(\mathbf{y}|\mathbf{x}; \sigma) p(\mathbf{x}) d\mathbf{x}, \quad (6)$$

165 where the  $t$  dependence is implicit and  $p(\mathbf{x})$  represents the marginal distribution of the training data.  
 166 We can use  $p(\mathbf{y}; \sigma)$  to express the time dependence of  $\mathbf{y}$  as a *probability-flow* ODE [16]:

$$d\mathbf{y} = -\dot{\sigma} \nabla_{\mathbf{y}} \log p(\mathbf{y}; \sigma) dt, \quad (7)$$

167 where  $\nabla_{\mathbf{y}} \log p(\mathbf{y}; \sigma)$  is the *score function* and  $\dot{\sigma}$  is the time derivative. In a score-based framework,  
 168 there is a direct relationship between the score function and our denoising model  $D$  [16, 47]:

$$\nabla_{\mathbf{y}} \log p_t(\mathbf{y}; \sigma) \approx \frac{1}{\sigma^2} (D(\mathbf{y}, \sigma) - \mathbf{y}) . \quad (8)$$

169 This important result connects our denoising model to the conformer generation process.

170 In our implementation, we have selected a linear function  $\sigma(t) = \alpha t$ , where  $\alpha$  is a scale parameter in  
 171 units of Å. Applying this selection to Eq. 7 and 8 results in a simple form for the probability-flow  
 172 ODE:

$$\frac{d\mathbf{y}}{dt} = (\mathbf{y} - D(\mathbf{y}; \alpha t)) / t. \quad (9)$$

173 Our conformer generation process is the numerical solution to this equation, calculated in steps of  $t$   
 174 in reverse, and using as initial conditions  $\mathbf{y}(1) \sim \mathcal{N}(0, \alpha^2 \mathbf{I})$ .

175 Inspired by work elsewhere [16], we solve Eq. 9 using Heun’s 2<sup>nd</sup>-order method, augmented by a  
 176 form of backtracking (Algorithm 1). The backtracking provides an option to add additional noise to  
 177 the generation process. We begin by dividing the interval  $[0, 1]$  into a fixed partition  $\{t_i\}$  over which  
 178 we iterate in order to calculate a set of intermediate solutions  $\{\mathbf{y}_i\}$ . Instead of relying on solving on  
 179 the partition  $\{t_i\}$ , we substitute modified values for the upper bound  $t_i$  of each subinterval:

$$\tilde{t}_i = \beta t_i \quad \tilde{\mathbf{y}}_i \sim \mathcal{N}(\mathbf{y}_i; 0, \lambda^2 \alpha^2 t_i^2 (\beta^2 - 1)), \quad (10)$$

180 where  $\beta \geq 1$  and  $\lambda \geq 0$  are fixed parameters. This has the effect of introducing Gaussian noise  
 181 at each step of the solution. For  $\lambda = 1$ , the amount of added noise compensates for the change in  
 182 subinterval size.

---

**Algorithm 1** Conformer generation.

---

```

1: procedure GENERATE( $D(\mathbf{y}, \sigma), \{t_i\}, \alpha, \beta, \lambda$ )
2:    $\mathbf{y} \leftarrow \mathcal{N}(0, \alpha^2 \mathbf{I})$  ▷ Prepare random initial state
3:   for  $i \leftarrow 1$  to  $|t|$  do
4:      $\tilde{t} \leftarrow \beta t_i$  ▷ Widen effective subinterval
5:      $\tilde{\mathbf{y}} \leftarrow \mathcal{N}(\mathbf{y}; 0, \lambda^2 \alpha^2 \tilde{t}_i^2 (\beta^2 - 1) \mathbf{I})$  ▷ Add noise
6:      $\mathbf{d}_1 \leftarrow (\tilde{\mathbf{y}} - D(\tilde{\mathbf{y}}, \alpha \tilde{t})) / \tilde{t}$  ▷ Evaluate  $d\mathbf{y}/dt$ 
7:      $\mathbf{y} \leftarrow \tilde{\mathbf{y}} + (t_{i+1} - \tilde{t}) \mathbf{d}_1$  ▷ Solve
8:     if  $t_{i+1} > 0$  then
9:        $\mathbf{d}_2 \leftarrow (\mathbf{y} - D(\mathbf{y}, \alpha t_{i+1})) / t_{i+1}$  ▷ Apply 2nd-order correction
10:       $\mathbf{y} \leftarrow \tilde{\mathbf{y}} + \frac{1}{2} (t_{i+1} - \tilde{t}) (\mathbf{d}_1 + \mathbf{d}_2)$ 
11:       $\mathbf{y} \leftarrow \mathbf{y} - \langle \mathbf{y} \rangle$  ▷ Remove center of mass
12:   return  $\mathbf{y}$ 

```

---

183 For reasons of convenience, we remove an overall center-of-mass during each generation step. The  
 184 correction is small and quality of output is not affected.

185 If we generate using  $\lambda = 0$ , no noise is added during the intermediate steps, resulting in deterministic  
 186 generation. Combined with  $\beta > 0$ , the algorithm is equivalent to pretending that each intermediate  
 187 value  $\mathbf{y}_i$  belongs to a solution sampled from a larger value of  $\sigma$ . This has the effect of overcorrecting,  
 188 which improves accuracy in our case.

189 Alternatively, if we generate using  $\lambda > 0$  and  $\beta > 0$ , we inject noise during each step for stochastic  
 190 generation. Both stochastic and deterministic approaches have been used for image generation, with  
 191 impressive results [10, 11, 12, 13, 16, 57, 58, 59].

192 To apply our algorithm, we use the partition  $\{t_1 \dots t_N\}$  of a given size  $N$  and final step size  $t_\epsilon$ :

$$t_i = \begin{cases} t_\epsilon^{i/(N-1)} & 1 \leq i < N \\ 0 & i = N \end{cases} . \quad (11)$$

193 The quality of generated output is reasonably stable for a large range of parameter values. The results  
 194 reported here use  $t_\epsilon = 0.0006$ ,  $\alpha = 2.5$  Å, and  $\beta = 5$ . Quality improves marginally if the solution is

195 calculated using more steps at a proportional cost in processing time. To quantify this trade off, we  
196 report on results for  $N = 100, 200,$  and  $500$ . We also report results for both deterministic ( $\lambda = 0$ )  
197 and stochastic ( $\lambda = 1$ ) generation.

198 Shown in Fig 5 are random examples of generated conformers, using PIDM[QMugs], deterministic  
199 generation, and 500 steps. More examples are available in the supplemental materials.

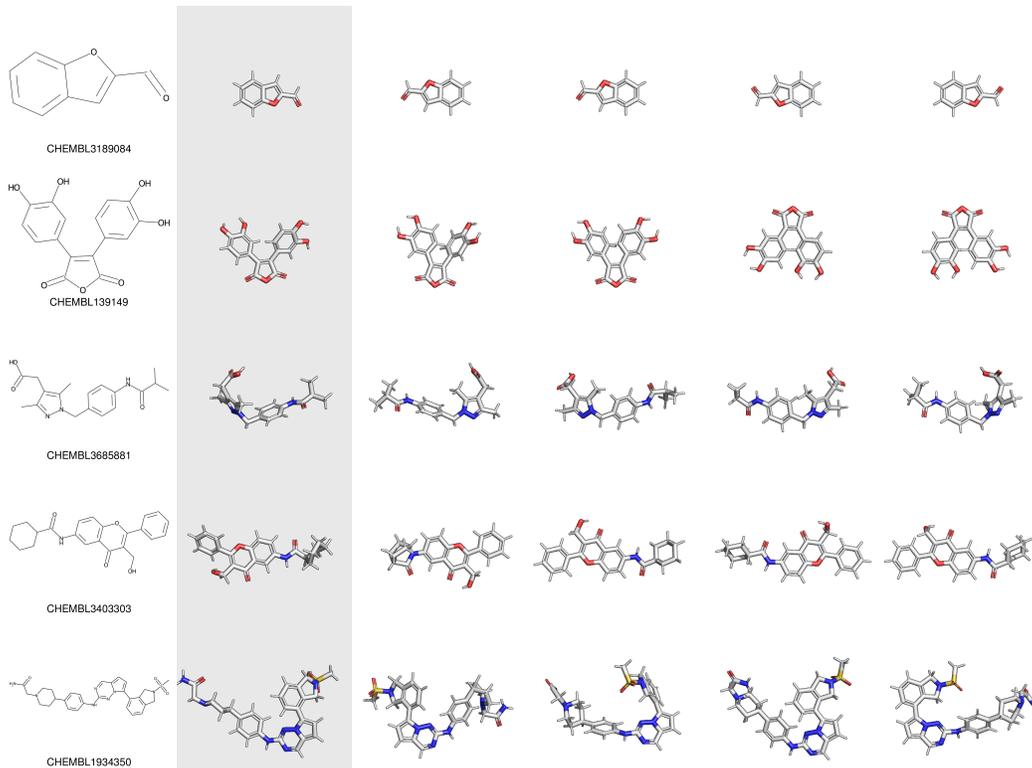


Figure 5: Example conformer output for molecules randomly selected from the benchmark dataset. Shown in the second column from the left (grey background) is the first conformer from QMugs. Shown on the right are unfiltered output from PIDM[QMugs] using deterministic generation and 500 steps. All molecule renderings are oriented by principal component.

## 200 6 Experiments

201 Our objective is to generate molecules with valid bonded geometry. To demonstrate accuracy, ten  
202 random conformers were generated by PIDM[QMugs] and PIDM[GEOM-drugs] for each of the  
203 molecules in the benchmark set. Each result was compared against the ground truth represented by  
204 the corresponding conformers in QMugs. Overall errors in bond length  $d$  and bend angle  $\theta$  were  
205 measured using the mean absolute deviation (MAD). Because proper torsions have multiple favored  
206 values, the MAD for  $\phi$  was limited to values generated with  $\pm 30^\circ$  of the angle found in the reference  
207 conformer. Failures to reproduce the desired chirality or cis-trans isomerism were recorded as a  
208 fraction of total occurrence of associated improper or proper torsion.

209 Results are shown in Table 1. Also shown are results obtained from other conformer generation  
210 solutions. In all cases, ten conformers were requested, although some conformer solutions provide  
211 less than the requested number under certain circumstances. The conformers provided by PubChem  
212 are calculated under the PubChem3D scheme [23] based on the OMEGA toolkit [32]. ETKDGv3  
213 has known deficiencies and is often followed by gradient optimization based on the MMFF94 force  
214 field [6, 21]. We used the GeoMol checkpoint (GEOM-drugs) provided by the authors.

215 We tested the GeoDiff checkpoint trained on GEOM-drugs provided by the authors. This model  
216 performs well on  $d$  and  $\theta$ , but struggles with  $\phi$ . GeoDiff has no mechanism for enforcing chirality

Table 1: Performance data for PIDM[QMugs] and PIDM[GEOM-drugs] for 100, 200 and 500 steps, using both deterministic and stochastic schemes, and with optional guided generation ( $\eta > 0$ , see section 7). Also shown are other methods. All statistics are measured against a single, independent benchmark dataset. Best values in each category are highlighted.

		Mean absolute deviation			Inconsistency rate	
		$d$ (Å)	$\theta$ (rad)	$\phi$ (rad)	chirality	cis-trans
PIDM[QMugs]						
Deterministic	100	0.0042	0.015	0.036	0.031	0.033
Deterministic	200	0.0038	0.013	0.027	0.022	0.009
Deterministic	500	0.0036	0.012	0.023	0.013	0.002
Deterministic	500, $\eta = 0.5$	0.0035	0.012	0.025	0.028	0.001
Deterministic	500, $\eta = 1$	0.0036	0.012	0.026	0.057	0.001
Stochastic	100	0.0051	0.021	0.079	0.112	0.027
Stochastic	200	0.0047	0.019	0.069	0.081	0.013
Stochastic	500	0.0045	0.018	0.062	0.057	0.004
PIDM[GEOM-drugs]						
Deterministic	100	0.0044	0.015	0.034	0.031	0.034
Deterministic	200	0.0040	0.013	0.027	0.023	0.015
Deterministic	500	0.0037	0.012	0.023	0.015	0.005
Deterministic	500, $\eta = 0.5$	0.0036	0.012	0.025	0.029	0.002
Deterministic	500, $\eta = 1$	0.0036	0.012	0.025	0.048	0.003
Pubchem3D (OMEGA)		0.0075	0.020	0.020	0.020	0.014
ETKDGv3		0.0183	0.039	0.019	0.049	0.017
ETKDGv3+MMFF94		0.0081	0.017	0.021	0.049	0.017
Balloon		0.0082	0.018	0.032	0.002	0.011
GeoMol		0.0125	0.030	0.042	0.032	0.087
GeoDiff		0.0051	0.017	0.170	0.500	0.263

217 nor cis-trans isomerism, an omission that appears to be an oversight rather than a limitation of the  
 218 approach.

## 219 7 Guided Generation

220 Our model makes no attempt to predict distances between nonbonded atoms. This suggests that there  
 221 are degrees of freedom available during generation that are being ignored. Addressing this deficiency  
 222 could provide a valuable measure of control over the torsional space.

223 Before experimenting, we require a control. The torsional space sampled in the synthetic QMugs and  
 224 GEOM-drugs data sets were established by their authors based on explicit yet arbitrary criteria, which  
 225 makes them uninteresting. Instead, we can refer to experimental data. The authors of the OMEGA  
 226 toolkit selected two small experimental sets for this purpose [32]: 480 molecules from the Cambridge  
 227 Structural Database (CSD) and 197 ligands from the PDB. As shown in Table 2, RMSD performance  
 228 of our model lags that obtained from OMEGA and RDKit+MMFF94.

229 Consider a modified, probability-flow ODE:

$$\frac{d\mathbf{y}}{dt} = (\mathbf{y} - D(\mathbf{y}; \alpha t) - F(\mathbf{y})) / t, \quad (12)$$

230 where  $F(\mathbf{y})$  is introduced to guide generation in a desired fashion. Solving for Eq. 12 in place  
 231 of Eq. 9 provides a mechanism for guided generation where  $F(\mathbf{y})$  serves as a form of conditional  
 232 score [60].

233 As a simple proof-of-concept, consider  $F(\mathbf{y}) = \eta \sum (b^2 + \delta^2)^{-5}$ , summed over all nonbonded atom  
 234 pairs, where  $\delta$  is the distance between atoms,  $b = 0.7\text{\AA}$ , and  $\eta$  is an overall magnitude. This term is  
 235 analogous to a repulsive force similar to what is found in the van der Waals interaction. The result is  
 236 a modest improvement in RMSD statistics on the CSD and PDB experimental data sets, as shown in  
 237 Table 2. Statistics on bond parameters (Table 1) are largely unaffected, with the notable exception of  
 238 an increase in chirality failure rates.

Table 2: RMSD statistics ( $\text{\AA}$ , heavy atoms only) on conformers generated by PIDM (deterministic with 500 steps) and various other methods compared to experimental data from the CSD and PDB. Mean and median are calculated on the closest conformer out of  $N$  generated.

Model		N	CSD		PDB	
			Mean	Median	Mean	Median
PIDM[QMugs]	Undirected	1000	0.74	0.84	0.90	0.98
PIDM[QMugs]	$\eta = 0.5$	1000	0.53	0.58	0.70	0.75
PIDM[QMugs]	$\eta = 1$	1000	0.51	0.55	0.67	0.73
PIDM[GEOM-drugs]	Undirected	1000	0.78	0.91	0.96	1.10
PIDM[GEOM-drugs]	$\eta = 0.5$	1000	0.54	0.59	0.81	0.81
PIDM[GEOM-drugs]	$\eta = 1$	1000	0.51	0.55	0.69	0.76
OMEGA <sup>a</sup>		—	0.51	0.44	0.67	0.51
RDKit+MMFF94		1000	0.43	0.48	0.53	0.64

<sup>a</sup> Published statistics [32]

## 239 8 Discussion

240 The PIDM[QMugs] and PIDM[GEOM-drugs] models, despite being trained on datasets with little  
 241 overlap, perform similarly on all benchmarks. Probes indicate that both models have learned similar  
 242 solutions. We contend that this level of robustness could only have been achieved by capturing the  
 243 important, underlying physics. Unlike conventional approaches to conformer generation, which either  
 244 rely on curated template libraries or manually-tabulated force field parameterizations, the physics in  
 245 our model is entirely learned.

246 PIDM is explainable. The advantage this provides cannot be overstated. The ability to probe the  
 247 internal structure during development permitted a level of experimentation that would not have been  
 248 possible in a black-box implementation.

249 For generation, we numerically solve for the probability-flow ODE, using a form of oversampling.  
 250 By separating the training and generation tasks, we were free to explore multiple approaches without  
 251 retraining. Using deterministic generation, we were able to produce conformers of reasonable quality  
 252 in as little as 100 steps. This is in stark contrast with GeoDiff [40], the current state-of-the-art  
 253 diffusion model, which employs stochastic generation in 5,000 steps.

254 For reproducing bonded parameters, PIDM with 500 steps has comparable performance on average  
 255 to conventional methods (such as OMEGA and ETKDGV3). It outperforms GeoDiff with 1/6 the  
 256 number of weights.

### 257 8.1 Limitations

258 Conformer generation for molecules with chemical groups or atom types outside the training set  
 259 may perform poorly or fail. Molecules with certain challenging topologies, such as a central ring  
 260 with several large branches, may perform poorly (an example is given in the supplemental materials).  
 261 Conformer quality is expected to degrade as molecules grow in size beyond  $\sim 200$  heavy atoms.

262 The atom embedding used in our model was generated using GATv2 [48]. We suspect that using  
 263 a type of graph network that can capture the same quality of atom-type information while also  
 264 recognizing cycles would improve performance.

## 265 9 Conclusion

266 Presented is PIDM, a diffusion-based model inspired by the bonded components of classical force  
 267 fields. Parameters were trained on high-quality conformers from the QMugs and GEOM-drugs  
 268 data sets. Learning appears robust, transferable, and explainable. Both deterministic and stochastic  
 269 generation schemes are demonstrated. Average performance on the reproduction of bonded parameters  
 270 is comparable to conventional conformer generation tools. A simple example of guided generation is  
 271 successful at improving torsional sampling when compared to experimental data.

## 272 References

- 273 [1] N. Schaduangrat, S. Lampa, S. Simeon, M. P. Gleeson, O. Spjuth, and C. Nantasenamat, "Towards  
274 reproducible computational drug discovery," *Journal of Cheminformatics*, vol. 12, p. 9, Jan. 2020.
- 275 [2] R. E. Bruccoleri and M. Karplus, "Chain closure with bond angle variations," *Macromolecules*, vol. 18,  
276 no. 12, pp. 2767–2773, 1985.
- 277 [3] G. M. Crippen, T. F. Havel, *et al.*, *Distance geometry and molecular conformation*, vol. 74. Research  
278 Studies Press Taunton, 1988.
- 279 [4] J. Gasteiger, C. Rudolph, and J. Sadowski, "Automatic generation of 3D-atomic coordinates for organic  
280 molecules," *Tetrahedron Computer Methodology*, vol. 3, pp. 537–547, Jan. 1990.
- 281 [5] M. J. Vainio and M. S. Johnson, "Generating Conformer Ensembles Using a Multiobjective Genetic  
282 Algorithm," *Journal of Chemical Information and Modeling*, vol. 47, pp. 2462–2474, Nov. 2007.
- 283 [6] S. Riniker and G. A. Landrum, "Better Informed Distance Geometry: Using What We Know To Improve  
284 Conformation Generation," *Journal of Chemical Information and Modeling*, vol. 55, pp. 2562–2574, Dec.  
285 2015. Publisher: American Chemical Society.
- 286 [7] P. C. D. Hawkins, "Conformation Generation: The State of the Art," *Journal of Chemical Information and  
287 Modeling*, vol. 57, pp. 1747–1756, Aug. 2017. Publisher: American Chemical Society.
- 288 [8] N.-O. Friedrich, C. de Bruyn Kops, F. Flachsenberg, K. Sommer, M. Rarey, and J. Kirchmair, "Bench-  
289 marking Commercial Conformer Ensemble Generators," *Journal of Chemical Information and Modeling*,  
290 vol. 57, pp. 2719–2728, Nov. 2017. Publisher: American Chemical Society.
- 291 [9] S. Wang, J. Witek, G. A. Landrum, and S. Riniker, "Improving Conformer Generation for Small Rings and  
292 Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences," *Journal of  
293 Chemical Information and Modeling*, vol. 60, pp. 2044–2058, Apr. 2020. Publisher: American Chemical  
294 Society.
- 295 [10] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," June 2020.
- 296 [11] A. Q. Nichol and P. Dhariwal, "Improved Denoising Diffusion Probabilistic Models," in *Proceedings of the  
297 38th International Conference on Machine Learning*, pp. 8162–8171, PMLR, July 2021. ISSN: 2640-3498.
- 298 [12] D. Kingma, T. Salimans, B. Poole, and J. Ho, "Variational Diffusion Models," in *Advances in Neural  
299 Information Processing Systems*, vol. 34, pp. 21696–21707, Curran Associates, Inc., 2021.
- 300 [13] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-Based Generative  
301 Modeling through Stochastic Differential Equations," Feb. 2022.
- 302 [14] S. Luo and W. Hu, "Diffusion Probabilistic Models for 3D Point Cloud Generation," pp. 2837–2845, 2021.
- 303 [15] G. Yang, X. Huang, Z. Hao, M.-Y. Liu, S. Belongie, and B. Hariharan, "PointFlow: 3D Point Cloud  
304 Generation with Continuous Normalizing Flows," Sept. 2019. arXiv:1906.12320 [cs].
- 305 [16] T. Karras, M. Aittala, T. Aila, and S. Laine, "Elucidating the design space of diffusion-based generative  
306 models," in *Advances in Neural Information Processing Systems* (A. H. Oh, A. Agarwal, D. Belgrave, and  
307 K. Cho, eds.), 2022.
- 308 [17] P. G. Polishchuk, T. I. Madzhidov, and A. Varnek, "Estimation of the size of drug-like chemical space  
309 based on GDB-17 data," *Journal of Computer-Aided Molecular Design*, vol. 27, pp. 675–679, Aug. 2013.
- 310 [18] D. L. Mobley, É. Dumont, J. D. Chodera, and K. A. Dill, "Comparison of charge models for fixed-charge  
311 force fields: small-molecule hydration free energies in explicit solvent," *The Journal of Physical Chemistry  
312 B*, vol. 111, no. 9, pp. 2242–2254, 2007.
- 313 [19] P. Dauber-Osguthorpe and A. T. Hagler, "Biomolecular force fields: where have we been, where are we  
314 now, where do we need to go and how do we get there?," *Journal of Computer-Aided Molecular Design*,  
315 vol. 33, pp. 133–203, Feb. 2019.
- 316 [20] D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev,  
317 C. Simmerling, B. Wang, and R. J. Woods, "The Amber biomolecular simulation programs,"  
318 *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1668–1688, 2005. \_eprint:  
319 <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.20290>.

- 320 [21] T. A. Halgren, "Merck molecular force field. I. Basis, form, scope, parameterization, and performance of  
321 MMFF94," *Journal of Computational Chemistry*, vol. 17, pp. 490–519, Apr. 1996.
- 322 [22] S. Ehlert, M. Stahn, S. Spicher, and S. Grimme, "Robust and Efficient Implicit Solvation Model for Fast  
323 Semiempirical Methods," *Journal of Chemical Theory and Computation*, vol. 17, pp. 4250–4261, July  
324 2021. Publisher: American Chemical Society.
- 325 [23] E. E. Bolton, S. Kim, and S. H. Bryant, "PubChem3D: Conformer generation," *Journal of Cheminformatics*,  
326 vol. 3, p. 4, Jan. 2011.
- 327 [24] C. R. Corbeil and N. Moitessier, "Docking Ligands into Flexible and Solvated Macromolecules. 3. Impact  
328 of Input Ligand Conformation, Protein Flexibility, and Water Molecules on the Accuracy of Docking  
329 Programs," *Journal of Chemical Information and Modeling*, vol. 49, pp. 997–1009, Apr. 2009. Publisher:  
330 American Chemical Society.
- 331 [25] L. Guo, Z. Yan, X. Zheng, L. Hu, Y. Yang, and J. Wang, "A comparison of various optimization algorithms  
332 of protein-ligand docking programs by fitness accuracy," *Journal of Molecular Modeling*, vol. 20, July  
333 2014.
- 334 [26] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, "Docking and scoring in virtual screening for drug  
335 discovery: methods and applications," *Nature Reviews Drug Discovery*, vol. 3, pp. 935–949, Nov. 2004.
- 336 [27] N. S. Pagadala, K. Syed, and J. Tuszynski, "Software for molecular docking: a review," *Biophysical  
337 Reviews*, vol. 9, pp. 91–102, Apr. 2017.
- 338 [28] G. Corso, H. StÅrkk, B. Jing, R. Barzilay, and T. Jaakkola, "DiffDock: Diffusion Steps, Twists, and Turns  
339 for Molecular Docking," Oct. 2022. arXiv:2210.01776 [physics, q-bio].
- 340 [29] W. Brooks, W. Guida, and K. Daniel, "The Significance of Chirality in Drug Design and Development,"  
341 *Current topics in medicinal chemistry*, vol. 11, no. 7, pp. 760–770, 2011.
- 342 [30] C. Dugave and L. Demange, "Cis- trans isomerization of organic molecules and biomolecules: implications  
343 and applications," *Chemical reviews*, vol. 103, no. 7, pp. 2475–2532, 2003.
- 344 [31] G. Landrum, P. Tosco, B. Kelley, s. sriniker, g. gedeck, N. NadineSchneider, R. Vianello, R. Ric, A. Dalke,  
345 B. Cole, A. AlexanderSavelyev, M. Swain, S. Turk, D. N, A. Vaucher, E. Kawashima, M. Wójcikowski,  
346 D. Probst, g. godin, D. Cosgrove, A. Pahl, J. JP, F. Berenger, s. strets123, J. JLVarjo, N. O'Boyle,  
347 P. Fuller, J. Jensen, G. Sforna, and D. Gavid, "RDKit: Open-source cheminformatics," 2021. 10.5281/zen-  
348 odo.3732262.
- 349 [32] P. C. D. Hawkins, A. G. Skillman, G. L. Warren, B. A. Ellingson, and M. T. Stahl, "Conformer Generation  
350 with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and  
351 Cambridge Structural Database," *Journal of Chemical Information and Modeling*, vol. 50, pp. 572–584,  
352 Apr. 2010. Publisher: American Chemical Society.
- 353 [33] S. Luo, C. Shi, M. Xu, and J. Tang, "Predicting Molecular Conformation via Dynamic Graph Score  
354 Matching," in *Advances in Neural Information Processing Systems*, vol. 34, pp. 19784–19795, Curran  
355 Associates, Inc., 2021.
- 356 [34] C. Shi, S. Luo, M. Xu, and J. Tang, "Learning Gradient Fields for Molecular Conformation Generation,"  
357 in *Proceedings of the 38th International Conference on Machine Learning*, pp. 9558–9568, PMLR, July  
358 2021. ISSN: 2640-3498.
- 359 [35] M. Xu, S. Luo, Y. Bengio, J. Peng, and J. Tang, "Learning Neural Generative Dynamics for Molecular  
360 Conformation Generation," Mar. 2021. arXiv:2102.10240 [physics].
- 361 [36] E. Mansimov, O. Mahmood, S. Kang, and K. Cho, "Molecular Geometry Prediction using a Deep  
362 Generative Graph Neural Network," *Scientific Reports*, vol. 9, p. 20381, Dec. 2019. Number: 1 Publisher:  
363 Nature Publishing Group.
- 364 [37] G. N. C. Simm and J. M. HernÃ;ndez-Lobato, "A Generative Model for Molecular Distance Geometry,"  
365 Aug. 2020. arXiv:1909.11459 [cs, stat].
- 366 [38] M. Xu, W. Wang, S. Luo, C. Shi, Y. Bengio, R. Gomez-Bombarelli, and J. Tang, "An End-to-End  
367 Framework for Molecular Conformation Generation via Bilevel Programming," in *Proceedings of the 38th  
368 International Conference on Machine Learning*, pp. 11537–11547, PMLR, July 2021. ISSN: 2640-3498.
- 369 [39] O. Ganea, L. Pattanaik, C. Coley, R. Barzilay, K. Jensen, W. Green, and T. Jaakkola, "GeoMol: Torsional  
370 Geometric Generation of Molecular 3D Conformer Ensembles," in *Advances in Neural Information  
371 Processing Systems*, vol. 34, pp. 13757–13769, Curran Associates, Inc., 2021.

- 372 [40] M. Xu, L. Yu, Y. Song, C. Shi, S. Ermon, and J. Tang, “GeoDiff: a Geometric Diffusion Model for  
373 Molecular Conformation Generation,” Mar. 2022. Number: arXiv:2203.02923 arXiv:2203.02923 [cs,  
374 q-bio].
- 375 [41] E. Hoogetboom, V. G. Satorras, C. Vignac, and M. Welling, “Equivariant Diffusion for Molecule Generation  
376 in 3D,” in *Proceedings of the 39th International Conference on Machine Learning*, pp. 8867–8887, PMLR,  
377 June 2022. ISSN: 2640-3498.
- 378 [42] G. N. Simm, R. Pinsler, G. Csányi, and J. M. Hernández-Lobato, “Symmetry-aware actor-critic for 3d  
379 molecular design,” *arXiv preprint arXiv:2011.12747*, 2020.
- 380 [43] N. W. Gebauer, M. Gastegger, S. S. Hessmann, K.-R. Müller, and K. T. Schütt, “Inverse design of 3d  
381 molecular structures with conditional generative neural networks,” *Nature communications*, vol. 13, no. 1,  
382 p. 973, 2022.
- 383 [44] S. Luo, J. Guan, J. Ma, and J. Peng, “A 3d generative model for structure-based drug design,” *Advances in  
384 Neural Information Processing Systems*, vol. 34, pp. 6229–6239, 2021.
- 385 [45] D. Flam-Shepherd, A. Zhigalin, and A. Aspuru-Guzik, “Scalable fragment-based 3d molecular design with  
386 reinforcement learning,” *arXiv preprint arXiv:2202.00658*, 2022.
- 387 [46] J. Guan, W. W. Qian, X. Peng, Y. Su, J. Peng, and J. Ma, “3d equivariant diffusion for target-aware  
388 molecule generation and affinity prediction,” *arXiv preprint arXiv:2303.03543*, 2023.
- 389 [47] P. Vincent, “A Connection Between Score Matching and Denoising Autoencoders,” *Neural Computation*,  
390 vol. 23, pp. 1661–1674, July 2011. Conference Name: Neural Computation.
- 391 [48] S. Brody, U. Alon, and E. Yahav, “How Attentive are Graph Attention Networks?,” *arXiv:2105.14491 [cs]*,  
392 Jan. 2022. arXiv: 2105.14491.
- 393 [49] E. D. Glendening and F. Weinhold, “Natural resonance theory: II. Natural bond order  
394 and valency,” *Journal of Computational Chemistry*, vol. 19, no. 6, pp. 610–  
395 627, 1998. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.10021>  
396 987X%2819980430%2919%3A6%3C610%3A%3AAID-JCC4%3E3.0.CO%3B2-U.
- 397 [50] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, “How Powerful are Graph Neural Networks?,” Tech. Rep.  
398 arXiv:1810.00826, arXiv, Feb. 2019. arXiv:1810.00826 [cs, stat] type: article.
- 399 [51] C. Isert, K. Atz, J. Jiménez-Luna, and G. Schneider, “QMugs, quantum mechanical properties of drug-like  
400 molecules,” *Scientific Data*, vol. 9, p. 273, June 2022. Number: 1 Publisher: Nature Publishing Group.
- 401 [52] S. Axelrod and R. Gómez-Bombarelli, “GEOM, energy-annotated molecular conformations for property  
402 prediction and molecular generation,” *Scientific Data*, vol. 9, p. 185, Apr. 2022. Number: 1 Publisher:  
403 Nature Publishing Group.
- 404 [53] S. Grimme, C. Bannwarth, and P. Shushkov, “A Robust and Accurate Tight-Binding Quantum Chemical  
405 Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems  
406 Parametrized for All spd-Block Elements (Z = 1–86),” *Journal of Chemical Theory and Computation*,  
407 vol. 13, pp. 1989–2009, May 2017. Publisher: American Chemical Society.
- 408 [54] C. Bannwarth, S. Ehlert, and S. Grimme, “GFN2-xTB—An Accurate and Broadly Parametrized Self-  
409 Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent  
410 Dispersion Contributions,” *Journal of Chemical Theory and Computation*, vol. 15, pp. 1652–1671, Mar.  
411 2019. Publisher: American Chemical Society.
- 412 [55] D. Rogers and M. Hahn, “Extended-Connectivity Fingerprints,” *Journal of Chemical Information and  
413 Modeling*, vol. 50, pp. 742–754, May 2010.
- 414 [56] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu,  
415 *et al.*, “Pubchem 2023 update,” *Nucleic Acids Research*, vol. 51, no. D1, pp. D1373–D1380, 2023.
- 416 [57] J. Song, C. Meng, and S. Ermon, “Denoising Diffusion Implicit Models,” Oct. 2022. arXiv:2010.02502  
417 [cs].
- 418 [58] A. Bansal, E. Borgnia, H.-M. Chu, J. S. Li, H. Kazemi, F. Huang, M. Goldblum, J. Geiping, and T. Goldstein,  
419 “Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise,” Aug. 2022. arXiv:2208.09392 [cs].
- 420 [59] Y. Song, P. Dhariwal, M. Chen, and I. Sutskever, “Consistency Models,” Mar. 2023. arXiv:2303.01469 [cs,  
421 stat].
- 422 [60] C. Luo, “Understanding Diffusion Models: A Unified Perspective,” Aug. 2022. arXiv:2208.11970 [cs].