

Research Article

Unified Guidance for Geometry-Conditioned Molecular Generation

Sirine Ayadi^{1,2}, Leon Hetzel^{1,2,3}, Johanna Sommer^{1,2}, Fabian Theis^{1,2,3}, Stephan Günnemann^{1,2}

1. School of Computation, Information and Technology, Technical University of Munich, Germany; 2. Munich Data Science Institute, Technical University of Munich, Germany; 3. Center for Computation Health, Helmholtz Munich, Germany

Effectively designing molecular geometries is essential to advancing pharmaceutical innovations, a domain, which has experienced great attention through the success of generative models and, in particular, diffusion models. However, current molecular diffusion models are tailored towards a specific downstream task and lack adaptability. We introduce UniGuide, a framework for controlled geometric guidance of unconditional diffusion models that allows flexible conditioning during inference without the requirement of extra training or networks. We show how applications such as structure-based, fragment-based, and ligand-based drug design are formulated in the UniGuide framework and demonstrate on-par or superior performance compared to specialised models. Offering a more versatile approach, UniGuide has the potential to streamline the development of molecular generative models, allowing them to be readily used in diverse application scenarios.

1. Introduction

Diffusion models have emerged as an important class of generative models in various domains, including computer vision^[1], signal processing^[2], computational chemistry, and drug discovery^{[3][4][5][6][7][8]}. By gradually adding noise to data samples and learning the reverse process of removing noise, diffusion models effectively transform noisy samples into structured data^{[9][10]}. In the context of drug discovery, it is essential to effectively address downstream tasks, which often pose specific geometric conditions. Examples of this include (i) Structure-based drug design (SBDD) that aims to create small ligands that fit given receptor binding sites^[11], (ii) Fragment-based drug design (FBDD) that designs molecules by elaborating known scaffolds^{[12][13]}, or (iii) Ligand-based drug design (LBDD) which generates molecules that fit a certain shape^[14]. Recent works address these tasks by either incorporating specialised models or focusing on conditions that directly resemble molecular structures. In both cases, this narrow focus restricts their adaptability to new or slightly altered settings.

We address the challenge of adaptability by introducing UniGuide, a method that unifies guidance for geometry-conditioned molecular generation, see Fig. 1. The key element for achieving this unification is the *condition map*, which transforms complex geometric conditions to match the diffusion model's configuration space, thereby enabling self-guidance without the need for external models. Like other guidance-based approaches, UniGuide does not constrain the generality of the underlying model. Moreover, our method is the most versatile, extending beyond guiding

molecular structures to leveraging complex geometric conditions such as volumes, surfaces, and densities, thereby enabling the unified tackling of diverse drug discovery tasks. For complex conditions specifically, previous works primarily rely on conditional diffusion models for effective condition encoding^{[12][13][14]}. With our method, we are able to tackle the same tasks, while overcoming major drawbacks: UniGuide eliminates the need for additional training and, more importantly, avoids constraining the model to specific tasks.

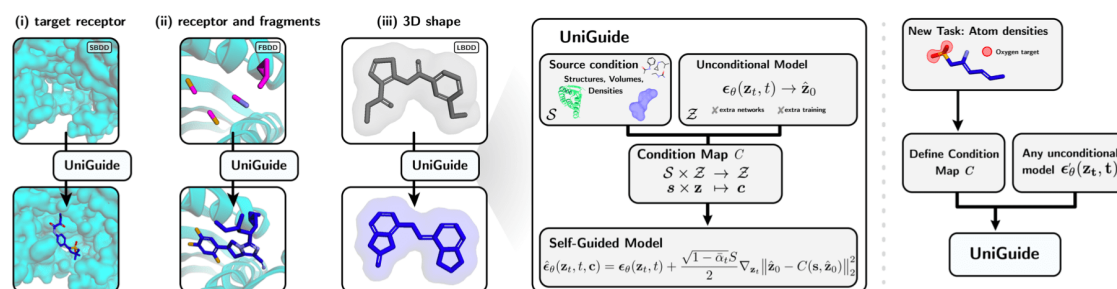


Figure 1. UniGuide handles diverse conditioning modalities for guidance, including: (i) a target receptor for SBDD, (ii) additional molecular fragments for FBDD, or (iii) a predefined 3D shape for LBDD. It combines a source condition $s \in S$ and the unconditional model $\epsilon_\theta(z_t, t)$ within its condition map to enable self-guidance. The flexible formulation of our approach can be generalised to new geometric tasks, for example, conditioning on atomic densities.

We demonstrate the wide applicability of UniGuide by tackling a variety of geometry-constrained drug discovery tasks. With performance either on par with or superior to tailored models, we conclude that UniGuide offers advantages beyond its unification. Firstly, while the novelty of conditional models often stems from the condition incorporation, our method redirects focus to advancing unconditional generation, which directly benefits multiple applications. Furthermore, this separation of model training and conditioning allows us to tackle tasks with minimal data, a common scenario in the biological domain.

In summary, our contributions are as follows:

- We present UniGuide: A unified guidance method for generating geometry-conditioned molecular structures, requiring neither additional training nor external networks used to guide the generation.
- We demonstrate UniGuide's wide applicability by tackling various conditioning scenarios in structure-based, fragment-based, and ligand-based drug design.
- We show UniGuide's favourable performance over task-specific baselines, highlighting the practical relevance of our approach.

2. Related work

Diffusion models and controllable generation

Diffusion models^{[9][10]} are generative models achieving state-of-the-art performance across various domains, including the generation of images^{[1][9]}, text^[15], or point clouds^[16]. Conditional diffusion models^{[17][18][19][20][21]} are based on the same principle but incorporate a particular condition in their training, allowing for the controlled generation. Alternatively, classifier guidance^{[22][23]} relies on external models for controllable generation. Prior works in this context primarily focused on global properties^{[22][24]}, lacking the capacity to condition on the geometric conditions central to our work. For instance,^[24] demonstrate control over molecule generation based on desired quantum properties.

De novo molecule generation

Research on *de novo* molecule generation focused extensively on generating molecules using their chemical graph representations^{[7][25][26][27][28][29][30][31][32][33][34]}. However, these methods are limited in modelling the molecules' conformation information and are, therefore, not ideally suited for several drug-discovery settings, such as target-aware drug design. Recently, attention has shifted towards generating molecules in 3D space, utilising variational autoencoders^[35], autoregressive models^{[36][37][38]}, flow-based models^{[39][40]}, and diffusion-based approaches^{[20][41][42][43][44][45][46][47]}.

Conditional generation of molecules

Downstream applications of molecular generation can be categorised by their condition modality. In the case of SBDD^{[38][48][49]},^[11] and^[50], for example, introduce models that simultaneously operate on protein pockets and ligands. In the conditional case, the pocket context is fixed throughout the generation. Moreover, FBDD imposes (multiple) scaffolds as a constraint^{[11][12][51][52][53]}.^[13] expand given scaffolds by generating the molecule around the fixed scaffolds. In a related task of FBDD, linker design with pose estimation, as discussed in^[54], further generate the rotation of the given scaffolds. SBDD and FBDD rely on the availability of high-quality data of protein pockets, which is often scarce. For this reason, LBDD aims to generate molecules that match the same 3D volume of reference ligands that are known to bind to the target of interest^{[55][56]}.^[14] specifically train a shape encoder to capture the molecular shape of a reference ligand and use the resulting embedding to train a conditional diffusion model.

3. Controlling the generation of diffusion models

Diffusion Models^{[9][57]} learn a Markov Chain that involves a forward process to perturb data from a distribution $q(\mathbf{z})$ and learn to reverse the process to generate new samples from a tractable prior, for example, a normal distribution. Given a data point sampled from the true underlying distribution, $\mathbf{z}_{\text{data}} \sim q(\mathbf{z})$, the forward process $q(\mathbf{z}_t | \mathbf{z}_{t-1})$ gradually adds Gaussian noise:

$$q(\mathbf{z}_t | \mathbf{z}_{t-1}) = \mathcal{N}(\mathbf{z}_t | \sqrt{1 - \beta_t} \mathbf{z}_{t-1}, \beta_t \mathbf{I}), \quad (1)$$

where $\{\beta_t \in (0, 1)\}_{t=1}^T$ defines a variance schedule. Defining the forward process this way, one can readily sample from $q(\mathbf{z}_t | \mathbf{z}_{\text{data}})$:

$$\mathbf{z}_t = \sqrt{\alpha_t} \mathbf{z}_{\text{data}} + \sqrt{1 - \alpha_t} \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (2)$$

with $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$. Since the time-reverse process $q(\mathbf{z}_{t-1} | \mathbf{z}_t)$ depends on $\mathbf{z}_{\text{data}}$, which is not available at generation time, it is approximated by modelling $p_\theta(\mathbf{z}_{t-1} | \mathbf{z}_t)$:

$$p_\theta(\mathbf{z}_{t-1} | \mathbf{z}_t) = \mathcal{N}(\mathbf{z}_{t-1} | \mu_\theta(\mathbf{z}_t, t), \sigma_t \mathbf{I}), \quad (3)$$

where the mean μ_θ is parameterised by a noise-predicting neural network ϵ_θ in the form of:

$$\mu_\theta(\mathbf{z}_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{z}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(\mathbf{z}_t, t) \right). \quad (4)$$

The model ϵ_θ is trained to optimise the variational lower bound through the simplified training objective:

$$\mathcal{L}_{\text{train}} = \frac{1}{2} \|\epsilon - \epsilon_\theta(\mathbf{z}_t, t)\|_2^2. \quad (5)$$

Self-guiding diffusion models

Using Bayes' rule, the conditional probability $p_\theta(\mathbf{z}_t | \mathbf{c})$ given a condition $\mathbf{c}$ can be expressed as

$$p_\theta(\mathbf{z}_t | \mathbf{c}) \propto p_\theta(\mathbf{z}_t) p_\theta(\mathbf{c} | \mathbf{z}_t). \quad (6)$$

This allows us to decompose the score function as follows:

$$\nabla_{\mathbf{z}_t} \log p_\theta(\mathbf{z}_t | \mathbf{c}) = \nabla_{\mathbf{z}_t} \log p_\theta(\mathbf{z}_t) + S \nabla_{\mathbf{z}_t} \log p_\theta(\mathbf{c} | \mathbf{z}_t), \quad (7)$$

where the second term is used for guiding the unconditional generation, with $S > 0$ controlling the guidance strength.

Using that $\nabla_{\mathbf{z}_t} \log p_\theta(\mathbf{z}_t) = -(1 - \bar{\alpha}_t)^{-\frac{1}{2}} \epsilon_\theta(\mathbf{z}_t, t)$ [\[22\]](#), we can rewrite the score function from Eq. (7) and identify the modified noise predictor $\hat{\epsilon}_\theta$:

$$\nabla_{\mathbf{z}_t} \log p_\theta(\mathbf{z}_t | \mathbf{c}) = -\frac{1}{\sqrt{1 - \bar{\alpha}_t}} \left[\underbrace{\epsilon_\theta(\mathbf{z}_t, t) - \sqrt{1 - \bar{\alpha}_t} S \nabla_{\mathbf{z}_t} \log p_\theta(\mathbf{c} | \mathbf{z}_t)}_{=: \hat{\epsilon}_\theta(\mathbf{z}_t, t, \mathbf{c})} \right] \quad (8)$$

The modified mean function $\hat{\mu}_\theta$ then follows from the modified version of Eq. (4), enabling us to sample from $p_\theta(\mathbf{z}_{t-1} | \mathbf{z}_t, \mathbf{c}) \sim \mathcal{N}(\hat{\mu}_\theta(\mathbf{z}_t, t, \mathbf{c}), \sigma_t \mathbf{I})$:

$$\hat{\mu}_\theta(\mathbf{z}_t, t, \mathbf{c}) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{z}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \hat{\epsilon}_\theta(\mathbf{z}_t, t, \mathbf{c}) \right) = \mu_\theta(\mathbf{z}_t, t) + \lambda(t) \nabla_{\mathbf{z}_t} \log p_\theta(\mathbf{c} | \mathbf{z}_t), \quad (9)$$

where $\lambda(t) = (\alpha_t)^{-\frac{1}{2}} \beta_t S$ balances the conditional update. Eq. (9) requires sampling from $\log p_\theta(\mathbf{c} | \mathbf{z}_t)$ to which we do not have access. Assuming the condition $\mathbf{c}$ lies in the same space as $\mathbf{z}_t$, we can follow [\[58\]](#) and approximate $\log p_\theta(\mathbf{c} | \mathbf{z}_t)$ as a multivariate Gaussian distribution:

$$p_\theta(\mathbf{c} | \mathbf{z}_t) = \mathcal{N}(\mathbf{c} | \mathbf{f}_\theta(\mathbf{z}_t, t), \mathbf{I}), \quad (10)$$

where $\mathbf{f}_\theta(\mathbf{z}_t, t)$ approximates the clean data point, enabling to estimate the condition in data space. Using Eq. (2), we can readily predict the clean data point given the noisy sample $\mathbf{z}_t$ via

$$\mathbf{f}_\theta(\mathbf{z}_t, t) = \frac{\mathbf{z}_t - \sqrt{1 - \alpha_t} \epsilon_\theta(\mathbf{z}_t, t)}{\sqrt{\alpha_t}} =: \hat{\mathbf{z}}_0. \quad (11)$$

With this, the *guiding term* becomes a direct differentiation of the squared error with respect to the noisy sample $\mathbf{z}_t$:

$$\nabla_{\mathbf{z}_t} \log p_\theta(\mathbf{c} | \mathbf{z}_t) = -\frac{1}{2} \nabla_{\mathbf{z}_t} \|\mathbf{f}_\theta(\mathbf{z}_t, t) - \mathbf{c}\|_2^2. \quad (12)$$

By directly leveraging the prediction of the unconditional model ϵ_θ , Eq. (12) establishes our self-guiding conditioning, thereby defining the self-guided noise predictor $\hat{\epsilon}_\theta$:

$$\hat{\epsilon}_\theta(\mathbf{z}_t, t, \mathbf{c}) = \epsilon_\theta(\mathbf{z}_t, t) + \frac{\sqrt{1 - \alpha_t} S}{2} \nabla_{\mathbf{z}_t} \|\hat{\mathbf{z}}_0 - \mathbf{c}\|_2^2. \quad (13)$$

4. UniGuide

To enable the application of unconditional molecular diffusion models ϵ_θ to geometric downstream tasks in drug discovery, we aim to develop a unified guidance framework, UniGuide, see Fig. 1. Importantly, we seek to enable guidance from arbitrary geometric conditions $\mathbf{s} \in \mathcal{S}$, where $\mathcal{S}$ denotes a general space of source conditions. However, the source conditions $\mathbf{s}$ cannot be directly used for the loss computation in Eq. (12) when they do not match the configuration space $\mathcal{Z}$.

To address this challenge, we introduce *condition maps* C , which bridge the gap between arbitrary source conditions $\mathbf{s}$ and target conditions $\mathbf{c}$ suitable for guidance. In Sec. 4.1, we start with its general formulation and continue to derive a condition map $C_{\mathcal{Z}}$ for the special case where $\mathcal{S} = \mathcal{Z}$. This will be useful when discussing the application of UniGuide to various drug discovery tasks in Sec. 4.2. We also demonstrate how to derive a task-specific condition map C_{θ_V} for ligand-based drug design.

Notation

In 3D space, the *configuration* of molecules, including proteins, can be represented by a set of tuples $\mathbf{z} = \{(\mathbf{x}_i, \mathbf{h}_i)\}_{i=1}^N \in \mathcal{Z}$, where $\mathbf{x}_i \in \mathbb{R}^3$ and $\mathbf{h}_i \in \mathbb{R}^d$ refer to coordinates and features of a node $\mathbf{z}_i = (\mathbf{x}_i, \mathbf{h}_i)$, respectively. The space of configurations is denoted by $\mathcal{Z}$ and includes configurations of varying size N . We distinguish between different configuration entities via superscripts, i.e. refer to molecules $\mathcal{M}$ and proteins $\mathcal{P}$ through $\mathbf{z}^{\mathcal{M}}$ and $\mathbf{z}^{\mathcal{P}}$, respectively. The collection of coordinates $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \in \mathbb{R}^{N \times 3} \in \mathcal{X}$ defines the *conformation* of a molecule $\mathcal{M}$ or protein $\mathcal{P}$. We represent arbitrary geometric conditions with the variable $\mathbf{s} \in \mathcal{S}$, and conditions that can be used for guidance with the variables $\mathbf{c} \in \mathcal{Z}$.

4.1. Unified self-guidance from geometric conditions $\mathbf{s} \in \mathcal{S}$

The concept of a condition map C is essential to our method, enabling guidance from conditions $\mathbf{s} \in \mathcal{S}$ in a unified fashion, where $\mathcal{S}$ represents a space of general geometric objects such as structures, densities, or surfaces. These

geometric objects do not necessarily match the configuration space $\mathcal{Z}$, i.e. $\mathcal{S} \neq \mathcal{Z}$, preventing the computation of the guiding score function from Eq. (12). We overcome this challenge by defining C as a transformation that maps $\mathbf{s}$ to a suitable target condition $\mathbf{c} \in \mathcal{Z}$, which is then utilised for self-guidance.

In the most general case, C takes the form of

$$\begin{aligned} C : \mathcal{S} \times \mathcal{Z} &\rightarrow \mathcal{Z} \\ \mathbf{s} \times \mathbf{z} &\mapsto \mathbf{c} \end{aligned} \quad (14)$$

where the source condition $\mathbf{s}$ together with a configuration $\mathbf{z}$ are mapped to a target condition $\mathbf{c} \in \mathcal{Z}$. Including the condition map C in the guidance, we obtain our guidance signal:

$$\nabla_{\mathbf{z}_t} \log p_{\theta}(\mathbf{c}|\mathbf{z}_t) = -\frac{1}{2} \nabla_{\mathbf{z}_t} \|\hat{\mathbf{z}}_0 - C(\mathbf{s}, \hat{\mathbf{z}}_0)\|_2^2 = -\nabla_{\mathbf{z}_t} \mathcal{L}(\hat{\mathbf{z}}_0, \mathbf{s}) \quad (15)$$

where $\hat{\mathbf{z}}_0 = \mathbf{f}_{\theta}(\mathbf{z}_t, t)$ is the estimate of $\mathbf{z}_0$ given the unconditional model $\epsilon_{\theta}(\mathbf{z}_t, t)$ obtained according to Eq. (11) and $\mathbf{c} = C(\mathbf{s}, \hat{\mathbf{z}}_0)$ is the target condition produced by the condition map. In this formulation, $\mathbf{c}$ can also be understood as *guidance target* of the unconditional model.

It is important to highlight that Eq. (15) should not destroy the underlying properties of the unconditional generative process. In particular, if the unconditional model ϵ_{θ} is equivariant to a set of transformations G , e.g. rotations and translations, as is common in the molecular domain, we want to retain equivariance also in the guidance signal. Hence, the self-guided model $\hat{\epsilon}_{\theta}$ should satisfy

$$\hat{\epsilon}_{\theta}(G(\mathbf{z}_t), t, \mathbf{c}) = G(\hat{\epsilon}_{\theta}(\mathbf{z}_t, t, \mathbf{c})), \quad (16)$$

for all transformations G to which ϵ_{θ} is equivariant.

Theorem 4.1. Consider a function $C : \mathcal{S} \times \mathcal{Z} \rightarrow \mathcal{Z}$. If $C(\mathbf{s}, \mathbf{z})$ is invariant to rigid transformations G in the first argument and equivariant in the second argument, then the gradient $\nabla_{\mathbf{z}} \|\mathbf{v}\|_2^2$ of the vector $\mathbf{v} = \mathbf{z} - C(\mathbf{s}, \mathbf{z})$ is equivariant to transformations of $\mathbf{z}$.

Proof. We prove Theorem 4.1 in App. B. $\square$

Using Theorem 4.1, we can guarantee equivariant guidance signals if the condition maps $C(\mathbf{s}, \mathbf{z})$ are invariant and equivariant under rigid transformations concerning the source condition $\mathbf{s}$ and configuration $\mathbf{z}$, respectively.

Guidance in the special case of $\mathcal{S} = \mathcal{Z}$

In the case where the source condition $\mathbf{s}$ directly defines subset $\mathcal{A}$ of $m < N$ nodes of the configuration, i.e. $\mathcal{S} = \mathcal{Z}$, we can fully specify the condition map. This is feasible because the condition map no longer needs to bridge different spaces; it only needs to ensure equivariance, as the loss computation between $\mathbf{s}$ and the configuration is already possible. To distinguish this special case from the general setting, we denote $\mathbf{s} = \tilde{\mathbf{z}} \in \mathbb{R}^{m \times (3+d)}$ and refer to the defined subset within the configuration $\hat{\mathbf{z}}_0$ by $\hat{\mathbf{z}}_0^{\mathcal{A}}$.

In order to satisfy the requirements on $C(\tilde{\mathbf{z}}, \hat{\mathbf{z}}_0^{\mathcal{A}})$ as stated by Theorem 4.1, we align $\tilde{\mathbf{z}}$ with $\hat{\mathbf{z}}_0^{\mathcal{A}}$ by using the Kabsch algorithm^{[59][60]}. Denoting the resulting transformation with $T_{\hat{\mathbf{z}}_0^{\mathcal{A}}}$, we get an $\hat{\mathbf{z}}_0$ -equivariant condition map:

$$C_Z : \mathbb{R}^{m \times (3+d)} \times \mathbb{R}^{m \times (3+d)} \rightarrow \mathbb{R}^{m \times (3+d)} \quad (17)$$

$$\tilde{\mathbf{z}} \times \hat{\mathbf{z}}_0^A \mapsto T_{\hat{\mathbf{z}}_0^A} \tilde{\mathbf{z}} \quad .$$

Taken together, we can compute the guidance signal based on the following loss $\mathcal{L}$:

$$\mathcal{L}(\hat{\mathbf{z}}_0^A, \tilde{\mathbf{z}}) = \frac{1}{2} \|\hat{\mathbf{z}}_0^A - T_{\hat{\mathbf{z}}_0^A} \tilde{\mathbf{z}}\|_2^2 \quad . \quad (18)$$

We emphasise that although the loss $\mathcal{L}(\hat{\mathbf{z}}_0^A, \tilde{\mathbf{z}})$ is computed on the subset $\mathcal{A}$, the gradient, as presented in Eq. (15), is still computed with respect the full configuration $\mathbf{z}_t$.

In summary, our method requires only an unconditionally trained model ϵ_θ and a suitable condition map C , eliminating the need for additional networks or training. Together, this facilitates unified self-guidance from arbitrary geometric sources. Importantly, the separation of model training and conditioning enables us to tackle tasks even with minimal data, which is crucial in practical scenarios. In the following section, we discuss the wide applicability of UniGuide by illustrating its application to multiple drug discovery tasks.

4.2. UniGuide for drug discovery

Having introduced both the guidance framework and the condition map, we will continue to discuss how to tackle a set of drug discovery tasks within the UniGuide framework. We start with its application to ligand-based drug design (LBDD), which aims to generate a ligand that satisfies a predefined molecular shape.

Ligand-based drug design

LBDD aims to generate novel ligands with a similar 3D shape as a reference ligand $\mathcal{M}_{\text{ref}}$. In this setting, one operates on the molecule level only since the protein information is assumed to be unknown. However, to still generate active ligands that bind to a protein pocket, one leverages the 3D shape information of a reference molecule. Specifically, the goal is to modify the generative process $\hat{\epsilon}_\theta$ to generate a ligand $\mathbf{z}_0$ with a similar 3D shape but different molecular structure than $\mathcal{M}_{\text{ref}}$. With Sec. 4.1 introducing all required concepts, we can readily formulate a *surface condition map* $C_{\partial V}$ suitable to tackle the task of LBDD, see Fig. 2:

To represent $\mathcal{M}_{\text{ref}}$'s 3D shape, we identify our source condition $\mathbf{s}$ with a set of K points $\mathbf{y}$ sampled uniformly from the reference ligand's surface ∂V , $\mathbf{y} \in \mathbb{R}^{K \times 3} = \mathcal{S}$. As no features are guided, we formulate $C_{\partial V}$ with respect to the conformation space $\mathcal{X} = \mathbb{R}^{N \times 3}$:

$$C_{\partial V} : \mathbb{R}^{K \times 3} \times \mathbb{R}^{N \times 3} \rightarrow \mathbb{R}^{N \times 3} \quad (19)$$

$$\mathbf{y} \times \hat{\mathbf{x}}_0 \mapsto \mathbf{c}_{\mathbf{x}} \quad ,$$

where $\hat{\mathbf{x}}_0$ denotes the conformation of the clean data point estimation $\hat{\mathbf{z}}_0$ as computed by Eq. (11). To satisfy Theorem 4.1, $C_{\partial V}$ first aligns $\mathbf{y}$ with $\hat{\mathbf{x}}_0$ by a rotation $R_{\hat{\mathbf{x}}_0} \in \mathbb{R}^{3 \times 3}$ resulting from the ICP algorithm^[61]. For every atom coordinate $\hat{\mathbf{x}}_i$, $C_{\partial V}$ subsequently computes the mean $\bar{\mathbf{y}}_i$ over $\hat{\mathbf{x}}_i$'s k closest surface points:

$$\bar{\mathbf{y}}_i = \frac{1}{k} \sum_{j \in \mathcal{N}_{\hat{\mathbf{x}}_i}^c} R_{\hat{\mathbf{x}}_0} \mathbf{y}_j \quad , \text{ with } \mathcal{N}_{\hat{\mathbf{x}}_i}^c = \operatorname{argmin}_{I \subset \{1, \dots, K\}, |I|=k} \sum_{j \in I} \|R_{\hat{\mathbf{x}}_0} \mathbf{y}_j - \hat{\mathbf{x}}_i\|_2 \quad . \quad (20)$$

Finally, the individual components $\mathbf{c}_{\mathbf{x},i}$ of the target condition compute as follows:

$$\mathbf{c}_{\mathbf{x},i} = \begin{cases} \bar{\mathbf{y}}_i + \frac{\alpha}{d}(\bar{\mathbf{y}}_i - \hat{\mathbf{x}}_i), & \text{if } \hat{\mathbf{x}}_i \text{ outside } V \\ \bar{\mathbf{y}}_i - \frac{\alpha}{d}(\bar{\mathbf{y}}_i - \hat{\mathbf{x}}_i), & \text{if } \hat{\mathbf{x}}_i \text{ inside } V \wedge d < \alpha \\ \hat{\mathbf{x}}_i, & \text{otherwise,} \end{cases} \quad (21)$$

where d denotes the distance to the surface, $d = \|\bar{\mathbf{y}}_i - \hat{\mathbf{x}}_i\|_2$, and α the required distance to the surface. Note that the target condition $\mathbf{c}_{\mathbf{x}}$ represents a valid conformation inside the surface ∂V , and that $C_{\partial V}$ effectively bridges spaces from $\mathcal{S}$ to $\mathcal{X}$. Consequently, when using $C_{\partial V}$, the guidance signal is derived from 15 with the loss function $\mathcal{L}(\hat{\mathbf{z}}_0, \mathbf{y})$. The full algorithm for guidance using $C_{\partial V}$ is presented in App. D.1.

Structure-based drug design

The goal of SBDD is to design a ligand that binds to a target protein pocket $\mathbf{s}$. In this setting, one operates on both the molecule and protein level. Technically, we are interested in generating a ligand $\mathbf{z}_0^{\mathcal{M}}$ conditioned on the protein configuration $\tilde{\mathbf{z}}^{\mathcal{P}}$. With the unconditional diffusion model $\epsilon_{\theta}(\mathbf{z}_t, t)$, $\mathbf{z}_t = (\mathbf{z}_t^{\mathcal{M}}, \mathbf{z}_t^{\mathcal{P}})$, approximating the joint distribution of ligand-protein pairs $p(\mathbf{z}_{\text{data}}^{\mathcal{M}}, \mathbf{z}_{\text{data}}^{\mathcal{P}})$, one can readily see that the source condition directly corresponds to the configuration of the protein pocket. Hence, we can use $C_{\mathcal{Z}}$ from Sec. 4.1 and identify $\tilde{\mathbf{z}}$ with $\tilde{\mathbf{z}}^{\mathcal{P}}$. The guidance signal then follows from the loss $\mathcal{L}(\hat{\mathbf{z}}_0^{\mathcal{P}}, \tilde{\mathbf{z}}^{\mathcal{P}})$ with $\mathbf{c}^{\mathcal{P}} = C_{\mathcal{Z}}(\tilde{\mathbf{z}}^{\mathcal{P}}, \hat{\mathbf{z}}_0^{\mathcal{P}})$ as defined in Eq. (18). We describe the sampling algorithm for the SBDD task in App. E.1.

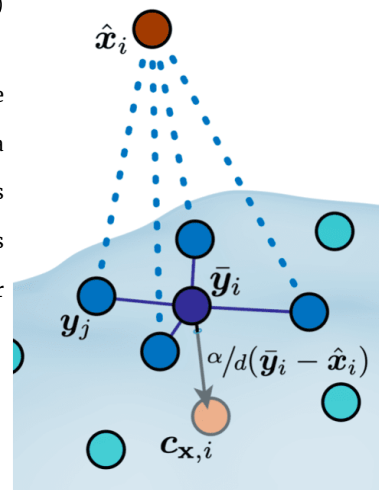


Figure 2. Surface condition map $C_{\partial V}$:

For each atom coordinate $\hat{\mathbf{x}}_i$, the closest surface points $\mathbf{y}_j$ are computed. The target condition $\mathbf{c}_{\mathbf{x},i}$ is the projection along the mean of neighbours $\bar{\mathbf{y}}_i$ to the inside of the volume by a margin α , where $d = \|\bar{\mathbf{y}}_i - \hat{\mathbf{x}}_i\|_2$.

Fragment-based drug design

FBDD aims to design a ligand by optimising a molecule around fragments $\mathcal{F}$ that bind weakly to a receptor. Similarly to SBDD, one operates on both the molecule and protein level. Technically, we are interested in generating a ligand $\mathbf{z}_0^{\mathcal{M}}$ conditioned on both the protein and the fragment configuration, $\tilde{\mathbf{z}}^{\mathcal{P}}$ and $\tilde{\mathbf{z}}^{\mathcal{F}}$, respectively. Considering the same kind of unconditional model $\epsilon_{\theta}(\mathbf{z}_t, t)$ as in SBDD, we can use $C_{\mathcal{Z}}$ from Sec. 4.1. Only now, we identify $\tilde{\mathbf{z}}$ with both $\tilde{\mathbf{z}}^{\mathcal{P}}$ and $\tilde{\mathbf{z}}^{\mathcal{F}}$ and write $\tilde{\mathbf{z}}^{\mathcal{A}}$ with $\mathcal{A} = \mathcal{P} \cup \mathcal{F}$. Using Eq. (18), the guidance signal directly follows from $\mathcal{L}(\hat{\mathbf{z}}_0^{\mathcal{A}}, \tilde{\mathbf{z}}^{\mathcal{P}} \cup \mathcal{F})$ with $\mathbf{c}^{\mathcal{P}} = C_{\mathcal{Z}}(\tilde{\mathbf{z}}^{\mathcal{P}} \cup \mathcal{F}, \hat{\mathbf{z}}_0^{\mathcal{P}} \cup \mathcal{F})$. The sampling algorithm is similar to the one described in App. E.1.

Several tasks exist within the FBDD setting^{[62][63][64][65]}. Examples are scaffold hopping^[64], where the core structure of $\mathbf{z}_0^{\mathcal{M}}$ has to be generated, but functional groups that interact with the receptor are fixed, or linker design^[65], where the connection between separated fragments has to be optimised through the generative process, see Fig. 5. Note that these tasks differ primarily in their application and can be treated identically from a technical perspective within UniGuide. In addition, one can also consider variations where the protein information $\tilde{\mathbf{z}}^{\mathcal{P}}$ is discarded. This usually aligns with switching to an unconditional model ϵ_{θ} that solely models the distribution over molecules. We present results for this configuration in Sec. 5.3.

Furthermore, we would like to highlight that it is possible to combine guidance strategies within UniGuide. For example, one could incorporate a version of the surface condition map $C_{\partial V}$ for FBDD to provide an additional geometric guidance signal for the atoms not included in $\mathcal{F}$.

Limitations

Drug discovery also involves tasks beyond purely geometric conditions, encompassing global graph properties^[24]. These are excluded from the UniGuide framework. Additionally, UniGuide requires the unconditional model to be trained on a matching configuration space. We discuss the broader impact of our work in App. A.

5. Results

In this section, we compare UniGuide to state-of-the-art models across various drug discovery tasks. To highlight the wide range of tasks to which unconditional models can be adapted through UniGuide, we conduct experiments on ligand-based (Sec. 5.1), structure-based (Sec. 5.2) and fragment-based (Sec. 5.3) drug design. We demonstrate that UniGuide performs competitively or even surpasses specialised baseline models, underscoring its practical relevance and transferability to diverse drug discovery scenarios.

5.1. Ligand-based drug design

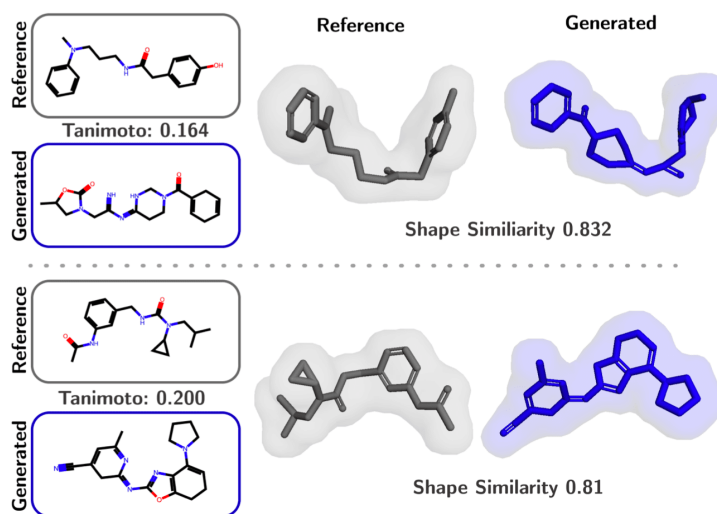


Figure 3. Examples of the two shape-conditioned ligands generated by UniGuide. The goal is to have *low* molecular graph similarity and *high* shape similarity.

Dataset

Following^[14], we employ the MOSES dataset for the ligand-based drug design task^[66]. We evaluate on a test set consisting of 1000 reference ligands, from which the 3D shape conditions are extracted. For every shape condition $\mathcal{M}_{\text{ref}}$, 50 samples are generated. We refer to D.1 for further details on the evaluation setup.

Baselines

For the LBDD task, we compare UniGuide to ShapeMol, a conditional diffusion model that is trained by conditioning on learned latent embeddings of the molecular surfaces^[14]. ^[14] also propose a correction technique that adjusts the atom positions based on their distance to the reference ligand’s nodes, which is referred to as ShapeMol+g. Additionally, we include as baselines Virtual Screening (VS)^[14], a shape-based virtual screening tool, and SQUID^[55], a variational autoencoder that decodes molecules by sequentially attaching fragments with fixed bond lengths and angles. For this task, we evaluate UniGuide equipped with the surface condition map $C_{\partial V}$ from (21) in conjunction with two *unconditionally trained* diffusion models, ShapeMol [U] and EDM^{[14][20]} as well as the *conditional* model ShapeMol^[14]. The “only shape” column in Table 1 indicates whether a method uses solely the reference ligand’s shape or also incorporates its atom positions.

	Method	onlyshape	Sim _S (↑)	max Sim _S (↑)	Sim _G (↓)	max Sim _G (↓)	Ratio (↑)	Diversity (↑)
Non-diffusion based	VS* ^[14]	✗	0.729 ± 0.04	0.807 ± 0.04	0.226 ± 0.04	0.241 ± 0.09	3.226	<u>0.759 ± 0.02</u>
	SQUID* ^[55] (λ = 0.3)	✗	0.717 ± 0.08	<u>0.904 ± 0.07</u>	0.349 ± 0.09	0.549 ± 0.24	2.054	0.687 ± 0.07
	SQUID* ^[55] (λ = 1.0)	✗	0.670 ± 0.07	0.842 ± 0.06	0.235 ± 0.05	0.271 ± 0.09	2.851	0.744 ± 0.05
Diffusion-based	ShapeMol ^[14]	✓	0.677 ± 0.04	0.797 ± 0.04	0.239 ± 0.05	0.240 ± 0.07	2.834	0.714 ± 0.05
	ShapeMol+g ^[14]	✗	0.744 ± 0.03	0.849 ± 0.03	0.242 ± 0.04	0.245 ± 0.05	3.074	0.708 ± 0.05
	UniGuide (ShapeMol [U])	✓	0.726 ± 0.04	0.827 ± 0.05	0.248 ± 0.05	0.239 ± 0.05	2.927	0.651 ± 0.05
	UniGuide (ShapeMol)	✓	<u>0.760 ± 0.05</u>	0.857 ± 0.06	0.240 ± 0.04	0.237 ± 0.06	3.167	0.705 ± 0.04
	UniGuide (EDM)	✓	0.749 ± 0.04	0.860 ± 0.04	<u>0.212 ± 0.04</u>	<u>0.206 ± 0.06</u>	<u>3.536</u>	0.736 ± 0.04

Table 1. Ligand-Based Drug Design. Results taken from^[14] are indicated with (*). We highlight the best conditioning approach for the ShapeMol backbone in **bold** and underline the best approach across all methods.

We compare UniGuide with an alternative guidance approach adapted from^[67] in App. D.4 and refer to App. C and App. D.3 for further information on the unconditional models and the guidance parameters, respectively. In addition, inspired by the performance of UniGuide on the LBDD task, we further motivate its applicability for the generation of molecules given atom densities, see App. G.

Evaluation

The goal of LBDD is to discover novel molecules that fit within a given 3D shape. This can be quantified by a high 3D shape similarity and low graph similarity compared to the reference ligand, as illustrated in Fig. 3 as well as App. D.2. We highlight this trade-off by reporting the ratio of these similarities in Table 1 as $\text{Sim}_S/\text{Sim}_G$, which constitutes the most important metric for this task. We follow^[14] and further evaluate the mean and maximum shape similarities Sim_S and maxSim_S , respectively, per reference ligand, measured via the volume overlap between the two aligned molecules. Additionally, we report the graph similarity Sim_G defined as the Tanimoto similarity between the generated and reference ligand, and the graph similarity maxSim_G of the generated molecule with the maximum shape similarity. Further metrics concerning the quality of the generated ligands are provided in App. D.2.

Both in terms of shape similarity and graph similarity, guiding the generation of EDM with UniGuide outperforms other task-specific conditioning mechanisms and even the Virtual Screening baseline. Emphasised by the Ratio metric across all evaluated methods, UniGuide demonstrates that it is able to generate diverse molecules with very similar shapes compared to the reference ligand. Remarkably, UniGuide achieves higher shape similarity than ShapeMol+g, even though the conditional model is explicitly guided towards the position of the reference ligand through the position correction technique. UniGuide, on the other hand, does not require information about the reference's atom positions at all to generate novel, high-quality ligands. This highlights how UniGuide and the design of condition maps enables unconditional models like EDM, that have not been tailored or trained for the LBDD task, to achieve state-of-the-art performance on new tasks.

5.2. Structure-based drug design

Datasets

Following^[11], we evaluate UniGuide on two protein-ligand datasets: the CrossDocked dataset^[68] and the Binding MOAD dataset^[69]. For the CrossDocked dataset, we follow the preprocessing as described by^[38] and conduct the evaluation on 100 test protein pockets. The Binding MOAD dataset is preprocessed as discussed in^[11], resulting in 130 test proteins. Per target pocket, 100 ligands are generated. We evaluate the generation of ligands on models that are trained on the full-atom context of the pockets in Table 2 and results of models trained on the C_α representation of the pockets are provided in App. E.5.

		Method	Vina Score (↓)	Vina Min (↓)	Vina Dock (↓)	QED (↑)	SA (↑)
CrossDocked		Test Set	-6.362 ± 3.14	-6.707 ± 2.50	-7.450 ± 2.33	0.48	0.73
	Non-Diff.	3D-SBDD*[38]	<u>-5.754 ± 3.25</u>	-6.180 ± 2.42	-6.746 ± 4.02	0.51	0.63
		Pocket2Mol*[48]	-5.139 ± 3.17	-6.415 ± 2.93	-7.152 ± 4.90	0.56	<u>0.74</u>
	Diffusion-based	DecompDiff*(No Drift) [67]	$-4.750 \pm -$	$-6.170 \pm -$	—	—	—
		TargetDiff*[50]	-5.466 ± 8.32	<u>-6.643 ± 4.94</u>	-7.802 ± 3.62	0.48	0.58
		DiffSBDD-cond [11]	-3.684 ± 11.3	-4.670 ± 6.06	-6.941 ± 4.33	0.47	0.58
		DiffSBDD [11]	-4.097 ± 11.3	-6.306 ± 5.00	-7.889 ± 2.61	<u>0.57</u>	0.64
		UniGuide	-5.103 ± 8.39	-6.610 ± 4.20	<u>-7.921 ± 2.43</u>	<u>0.57</u>	0.64
Binding MOAD	Diffusion-based	Test Set	-6.748 ± 2.77	-7.563 ± 2.53	-8.297 ± 2.03	0.60	0.64
		DiffSBDD-cond [11]	-4.466 ± 2.63	-6.309 ± 2.52	-7.482 ± 1.84	0.43	0.56
		DiffSBDD [11]	-4.744 ± 7.70	-6.586 ± 2.59	-7.767 ± 2.06	0.55	<u>0.62</u>
		UniGuide	<u>-5.074 ± 6.75</u>	<u>-6.622 ± 2.57</u>	<u>-7.911 ± 1.97</u>	<u>0.56</u>	0.61

Table 2. Structure-Based Drug Design. Quantitative comparison of generated ligands for target pockets from the CrossDocked and Binding MOAD test sets. Results taken from the respective works are indicated with (*). We highlight the best conditioning approach for the DiffSBDD backbone in **bold** and underline the best approach over all methods.

Baselines

We compare UniGuide to two autoregressive models designed for the SBDD task: 3D-SBDD^[38] and Pocket2Mol^[48]. We further include TargetDiff^[50] and DecompDiff^[67], conditional diffusion models for SBDD that fix the protein pocket context during every step of the diffusion process. We exclude approaches with explicit drift terms like^[67] and^[70] from the comparison, as UniGuide’s SBDD condition map does not include drift terms currently, but can be readily extended to do so.^[11] present two techniques for controlled structure-based generation: (i) DiffSBDD-cond, a conditional diffusion model similar to^[50] and (ii) DiffSBDD, an inpainting-inspired technique that modifies the generative process of an unconditional diffusion model that jointly generates protein-ligand pairs. Across datasets, both UniGuide and DiffSBDD control the same unconditional ligand-protein diffusion model. We provide

more information and further evaluation regarding this base model in App. E.2 and App. E.3 and investigate the influence of the guidance scale S as well as the resampling trick^[71], a technique that modifies the generative process to better harmonise the generated ligand with the controlled pockets, in App. E.4 and App. E.5.

Evaluation

As the task of SBDD is to generate ligands that bind well to a given protein pocket, we assess generated ligands based on affinity-related metrics (Vina Score, Vina Min and Vina Dock), which estimate the binding affinity between the generated ligands and a given test receptor^[72]. Additionally, we measure the quality of the generated ligands using two chemical properties: the drug-likeness (QED) and the synthetic accessibility (SA)^{[66][73]}.

Table 2 demonstrates that, without additional training or external networks, UniGuide performs competitively with even the highly specialised conditional models like TargetDiff and DecompDiff. Our results indicate that not fully converging to the target protein pocket due to soft guidance, compared to, for example, DiffSBDD's inpainting-inspired technique, is not a limitation in practice. Rather, it suggests that utilising self-guidance in combination with a suitable condition map generates well-harmonised ligand-protein pairs. This is also reflected in the properties of the generated ligands, where UniGuide achieves good drug-likeness (QED) and synthetic accessibility (SA) scores. We provide additional qualitative examples for the SBDD task in Fig. 4, which showcase that UniGuide not only generates drug-like ligands but is even able to improve over the VINA Dock metric of the reference ligand.

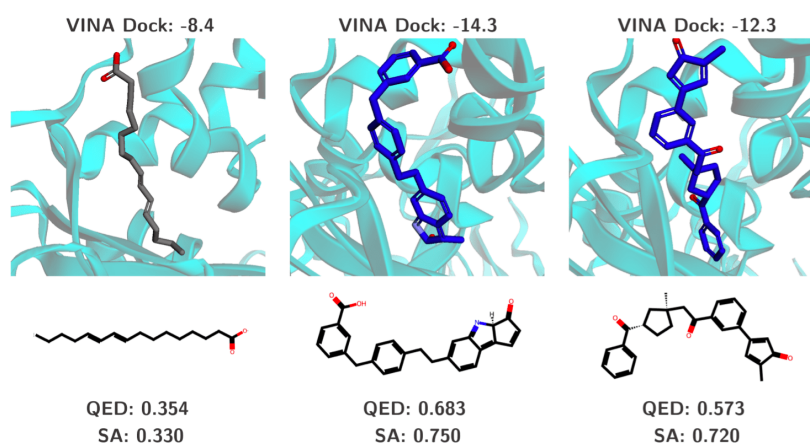


Figure 4. Qualitative example of a test protein pocket (6cob) from the Binding MOAD dataset. We show the reference ligand (grey) and samples generated by UniGuide (blue).

5.3. Fragment-based drug design

Datasets & Baselines

In the following, we investigate linker design, a subfield of fragment-based drug design. We follow [\[13\]](#) and decompose ligands from the ZINC dataset [\[74\]](#) with the MMPA algorithm [\[75\]](#). Note that the ZINC dataset does not contain pocket information, and the evaluated approaches operate solely at the molecular level. We compare UniGuide to DiffLinker [\[13\]](#), a diffusion-based conditional model that fixes fragments in space. Additionally, we evaluate the variational autoencoder-based methods DeLinker [\[53\]](#) and 3DLinker [\[52\]](#), adapted as described in [\[13\]](#). We provide more information on the experimental setup as well as the unconditionally trained EDM model in F.1 and C.

Evaluation

Following [\[13\]](#), we evaluate the generated linkers and ligands with respect to their properties (SA, QED, Number of Rings and 2D Filters). We additionally measure (i) the uniqueness of the generated samples, (ii) the recovery of the reference ligands, and (iii) the validity, which combines the chemical validity and the successful linking of the fragments.

Using UniGuide to control the EDM generation enables the successful combination of the condition fragments and the generation of diverse linkers. Even compared to task-specific models, UniGuide is able to perform competitively across different metrics. Importantly, UniGuide enables the same unconditional model (EDM) to tackle both the linker design task as presented in Table 3 as well as the LBDD task as presented in Table 1 without additional training. Note that, while DiffLinker is specifically designed to generate linkers, UniGuide readily generalises to other tasks within the FBDD setting, such as fragment growing and scaffolding, see Fig. 5. Additionally, UniGuide is agnostic to the fragmentation procedure used to obtain the condition scaffolds, meaning that UniGuide will generalise to unseen fragments as long as the underlying molecule fits within the training distribution. In Sec. F.2, we demonstrate how the same unconditional model can be adapted for these tasks. Our quantitative evaluation highlights the benefits achieved through the unification of controlled generation provided by UniGuide.

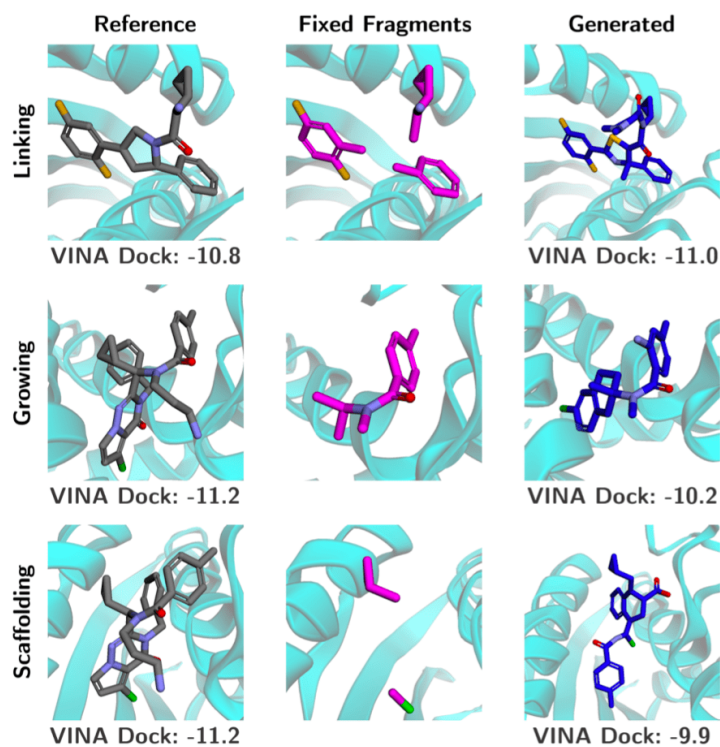


Figure 5. For various pocket-conditioned FBDD tasks, we show reference ligands (grey), desired fragments (magenta), and ligands generated by UniGuide (blue).

	Method	QED (↑)	SA (↓)	No. Rings (↑)	Valid (↑)	Unique (↑)	2D Filters (↑)	Recovery (↑)
Non-diffusion based	DeLinker + ConfVAE + MMFF [53]*	0.64 ± 0.16	<u>3.11</u> ± <u>0.68</u>	0.21 ± 0.42	<u>98.3</u>	44.2	84.8	80.2
	3DLinker [52]*	<u>0.65</u> ± <u>0.16</u>	3.14 ± 0.68	0.24 ± 0.43	71.5	29.2	83.7	93.5
	3DLinker (given anchors) [52]*	<u>0.65</u> ± <u>0.16</u>	<u>3.11 ± 0.67</u>	0.23 ± 0.42	99.3	29.0	84.2	<u>94.0</u>
Diffusion-based	DiffLinker [13]*	<u>0.65</u> ± <u>0.15</u>	3.19 ± 0.77	0.32 ± 0.54	90.6	51.4	<u>87.9</u>	70.7
	DiffLinker (given anchors) [13]*	<u>0.65</u> ± <u>0.15</u>	3.24 ± 0.81	0.36 ± 0.59	94.8	50.9	84.7	77.5
	UniGuide (EDM)	0.64 ± 0.16	3.63 ± 1.08	<u>0.49</u> ± <u>0.62</u>	89.1	<u>72.1</u>	<u>87.9</u>	58.8

Table 3. Linker Design. Results taken from [13] are indicated with (*). We underline the best method overall.

6. Conclusion

In this work, we present UniGuide, a unified way of controlling the generation of molecular diffusion models towards geometric constraints. UniGuide generalises to a multitude of drug discovery tasks without the need for conditioning networks or specialised training protocols, enabling UniGuide to find applicability also in scenarios where little data is available. By demonstrating that specialisation is not a necessity and that a more flexible, unified method outperforms specialised approaches across tasks and datasets, we open up new avenues for streamlined and flexible generative models with wide-ranging applications.

Appendix A. Impact Statement

Our research holds the promise of significant contributions to the advancement of drug discovery, possibly assisting in the discovery of novel pharmaceutical compounds. Nevertheless, because of its applications in drug discovery, this strategy is not without its hazards. The ability to produce various molecules with desired properties may not only serve the purpose of beneficial drug development but may also unintentionally result in the creation of dangerous substances or compounds with unexpected effects. These concerns underline the critical need for careful handling when working with the structures this method can generate.

Appendix B. Proof of Theorem 4.1

First, recall Theorem 4.1 that we provide in Sec. 4:

Theorem 1. Consider a function $C : \mathcal{S} \times \mathcal{Z} \rightarrow \mathcal{Z}$. If $C(\mathbf{s}, \mathbf{z})$ is invariant to rigid transformations G in the first argument and equivariant in the second argument, then the gradient $\nabla_{\mathbf{z}} \|\mathbf{v}\|_2^2$ of the vector $\mathbf{v} = \mathbf{z} - C(\mathbf{s}, \mathbf{z})$ is equivariant to transformations of $\mathbf{z}$.

Proof. We start the proof by showing that $\|\mathbf{v}\|_2$ is invariant to transformations of both $\mathbf{z}$ and $\mathbf{s}$.

1. $\|\mathbf{z} - C(\mathbf{s}, \mathbf{z})\|_2$ is invariant to transformations in $\mathbf{z}$:

$$\begin{aligned} \|G\mathbf{z} - C(\mathbf{s}, G\mathbf{z})\|_2 &= \|G\mathbf{z} - GC(\mathbf{s}, \mathbf{z})\|_2 && [C \text{ is equivariant in } \mathbf{z}] \\ &= \|G(\mathbf{z} - C(\mathbf{s}, \mathbf{z}))\|_2 \\ &= \|\mathbf{z} - C(\mathbf{s}, \mathbf{z})\|_2 && [G \text{ is a rigid transformation}] \end{aligned} \quad (22)$$

2. $\|\mathbf{z} - C(\mathbf{s}, \mathbf{z})\|_2$ is invariant to transformations $\mathbf{s}$ follows immediately:

$$\|\mathbf{z} - C(G\mathbf{s}, \mathbf{z})\|_2 = \|\mathbf{z} - C(\mathbf{s}, \mathbf{z})\|_2 \quad [C \text{ is invariant in } \mathbf{s}] \quad (23)$$

In a second step, we make use of the fact that for a group of transformations G , it holds that if $\mathcal{L}(\cdot, \cdot)$ is a G -invariant function, $\nabla_x \mathcal{L}(\cdot, x)$ is G -equivariant^[76]. From the invariance of $\|\mathbf{v}\|_2$, it follows immediately that $\nabla_{\mathbf{z}} \|\mathbf{z} - C(\mathbf{s}, \mathbf{z})\|_2^2$ is equivariant to transformations of $\mathbf{z}$. $\square$

Appendix C. Unconditional Equivariant Diffusion Model

UniGuide guides an unconditional diffusion model given an arbitrary condition and a natural choice for a model operating only on the molecule level is the EDM model as proposed in^[20].

We adapt this model for two tasks presented in this work, namely the LBDD task discussed in Sec. 5.1 and the Linker Design task as presented in Sec. 5.3. For these tasks, we train an unconditional EDM model both on the MOSES dataset^[66] in the configuration as described in^[14] and on the ZINC dataset^[74] as described in^[13]. For both trainings, we employ the hyperparameter configuration for the GEOM dataset as described in^[20]. We run multi-GPU trainings on 4 NVIDIA A100 GPUs until convergence, however, a single NVIDIA A100 GPU is sufficient for this training and will only increase the training time. For inference, we employ the Resampling trick as discussed in^[71] with $R = 10$ resampling steps and $T = 100$ timesteps. EDM is available under the MIT License.

Appendix D. Ligand-based drug design

D.1. Implementation details

We train two unconditional diffusion models, ShapeMol [U] and EDM, to generate 3D molecules on the MOSES dataset^[66], licensed under the MIT License, for which we generate 3D conformers with RDKit^[73], available under the BSD 3-Clause License. We use 1,593,653 training samples and randomly select 1000 samples for validation. The model architecture of ShapeMol[U] is an unconditional version of the ShapeMol model proposed in^[44], and it is trained with

1000 diffusion steps. ShapeMol [U] is trained with a batch size of 32 on two NVIDIA A100 GPUs for 500 epochs. Unlike ShapeMol, we do not concatenate the molecular surface embedding of the ligands to the features. For the shape-conditioned generation with position correction (ShapeMol+g), we follow the scheme proposed in^[44]. It provides further guidance to the conditional generation by sampling 20 query points from a Gaussian distribution centred around every atom in the reference ligand. The position correction adjusts the coordinates of the predicted atom positions during every generation step by pushing the coordinates close to the query points as follows:

$$\hat{\mathbf{x}} = (1 - \sigma)\hat{\mathbf{x}} + \sigma \sum_{\mathbf{z} \in n(\hat{\mathbf{x}}, \mathcal{Q})} \mathbf{z}/n, \quad \text{if } \sum_{\mathbf{z} \in n(\hat{\mathbf{x}}, \mathcal{Q})} d(\hat{\mathbf{x}}, \mathbf{z})/n > \gamma, \quad (24)$$

where $d(\hat{\mathbf{x}}, \mathbf{z})$ is the Euclidean distance, $n(\hat{\mathbf{x}}, \mathcal{Q})$ is the set of n nearest neighbors of $\hat{\mathbf{x}}$ in $\mathcal{Q}$ and $\gamma > 0$ is a distance threshold. We follow the implementation of^[2] for the position correction method by setting $\gamma = 0.2$ and only guiding during the first 700 denoising steps.

For the shape-conditioned generation with UniGuide, we extract the mesh of the condition ligand using the Open Drug Discovery Toolkit^[77], which is available under the BSD 3-Clause revised License. The query points we use for guidance are 512 points sampled uniformly on the mesh surface. For the evaluation, we measure the shape similarity Sim_s as the volume overlap between the aligned generated ligand and the condition ligand. For the alignment, we utilise the ShaEP tool^[78].

We provide a detailed description of the LBDD sampling algorithm in Algorithm 1.

Algorithm 1: Sampling algorithm to generate a ligand that is conditioned on a reference ligand $\mathcal{M}_{\text{ref}}$'s surface, using an unconditional model $\epsilon_\theta(\mathbf{z}_t, t)$ modelling the distribution over molecules. The points $\mathbf{y} \in \mathbb{R}^{K \times 3}$ are sampled uniformly from the surface of $\mathcal{M}_{\text{ref}}$, enclosing the volume V .

Require: $\mathbf{y}$, α : desired margin to surface, k : number of nearest neighbours
 $\mathbf{z}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ {Sample from normal prior}
for $t = T$ **to** 1 **do**
 $\mathbf{x}_t, \mathbf{h}_t = \mathbf{z}_t$
 $\hat{\mathbf{x}}_0 = \frac{\mathbf{x}_t - \sqrt{1 - \alpha_t} \epsilon_\theta^*(\mathbf{z}_t, t)}{\sqrt{\alpha_t}}$ {Compute the conformation $\hat{\mathbf{x}}_0$ of the clean approximation $\hat{\mathbf{z}}_0$ }
 For every atom $\hat{\mathbf{x}}_i$ **in** $\hat{\mathbf{x}}_0$ **do:**
 $\bar{\mathbf{y}}_i = \frac{1}{k} \sum_{\mathbf{y} \in \mathcal{N}_{\hat{\mathbf{x}}_i}} \mathbf{y}$ {Compute the mean of k nearest neighbors of $\hat{\mathbf{x}}_i$ in $\mathbf{y}$ }
 Compute $(\mathbf{c}_\mathbf{x})_i$ based on Eq. (21) {Compute component-wise condition map}
 $\mathcal{L} = \mathcal{L}(\hat{\mathbf{x}}_0, \mathbf{c}_\mathbf{x})$
 $\mathbf{g} = \nabla_{\mathbf{x}_t} \mathcal{L}$ {Compute gradient of guidance loss}
 $\boldsymbol{\mu}_t = \boldsymbol{\mu}_\theta(\mathbf{z}_t, t) - \lambda(t) \cdot \mathbf{g}$ {Update the mean function}
 $\mathbf{z}_{t-1} \sim \mathcal{N}(\boldsymbol{\mu}_t, \sigma_t \mathbf{I})$
end for
return $\mathbf{z}_0$

D.2. Additional results

For completeness, we report additional quantitative evaluation of the generated ligands' properties in Table 4. We also provide further qualitative results of the generated ligands for the LBDD task in Fig. 6. UniGuide generates ligands with

better shape similarity to the reference ligands compared to the conditional model ShapeMol with the position correction technique.

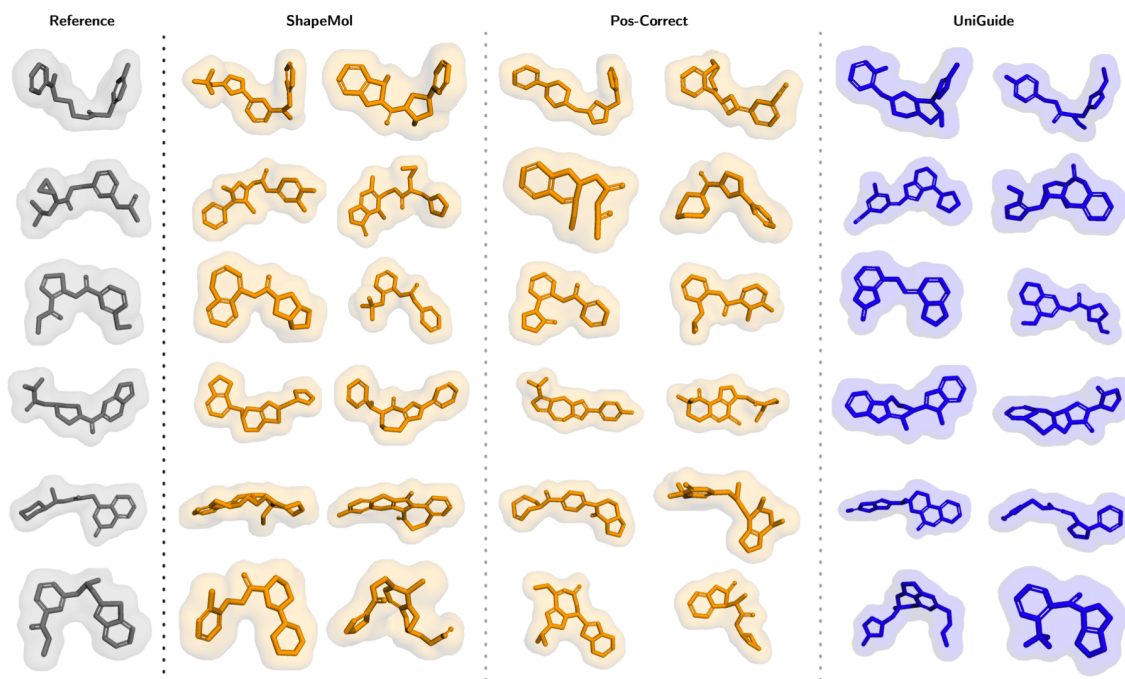


Figure 6. Examples of the ligands generated by ShapeMol, Pos-Correct and UniGuide. Pos-Correct is the position correction technique proposed by [14]. Both Pos-Correct and UniGuide are combined with the unconditionally trained model ShapeMol [U]. We plot the reference ligand as well as the generated ligands with their shapes.

method	Connect. (↑)	Unique (↑)	QED	SA (↑)	LogP (↑)	Lipinski (↑)
ShapeMol	98.8%	99.9%	0.753	0.640 ± 0.104	2.001 ± 1.360	4.979 ± 0.156
ShapeMol+g	97.0%	99.8%	0.751	0.630 ± 0.110	1.908 ± 1.508	4.874 ± 0.170
UniGuide+ ShapeMol[U]	98.0%	100%	0.736	0.625 ± 0.103	1.828 ± 1.463	4.974 ± 0.186
UniGuide (ShapeMol)	99.0%	100%	0.750	0.641 ± 0.107	2.002 ± 1.374	4.982 ± 0.152
UniGuide+ EDM	99.8%	99.99%	0.742	0.636 ± 0.088	1.833 ± 1.221	4.994 ± 0.082

Table 4. Additional ligand property results for the methods discussed in Sec. 5.1. We report mean and standard deviation and highlight the best result in **bold**.

D.3. Guidance parameters

For the LBDD task, the guidance strength S is weighted by an exponentially decreasing function $\frac{\beta_t}{\sqrt{\alpha_t}}$. For the guided generation using the unconditional ShapeMol [U] model under the UniGuide framework, we define a scale scheduler that increases with an exponent of 1.01 and weight it with $\frac{\beta_t}{\sqrt{\alpha_t}}$ and guide from the diffusion step 1000 to the diffusion step 200. For the guided generation using the EDM model, we use a linear scale function that increases from 5 to 15. The guidance is applied from the diffusion step 920 to the last timestep 1.

D.4. Comparison of UniGuide with an alternative loss formulation

We adapt the validity guidance loss from [50] to the LBDD setting. The proposed loss is grounded in the smooth distance function $S(x)$ from [79], which computes as:

$$S(x) = -\sigma \log \left(\sum_i^N \exp(-\|x_t - y_i\|_2^2 / \sigma) \right).$$

This function provides an alternative approach to shape-based generation by deriving an appropriate loss function $\sum_x S(x)$, rather than modifying the condition map as proposed by UniGuide. Here, $S(x)$ implicitly defines a surface through $S(x) - \gamma = 0$ and points x_t inside satisfy $S(x_t) < \gamma$.

On a technical level, the gradient for validity guidance computes as follows:

$$\begin{aligned} \nabla_{x_t} S(x_t) &= \nabla_{x_t} \left[-\sigma \log \left(\sum_{i=1}^N \exp(-\|x_t - y_i\|_2^2 / \sigma) \right) \right] \\ &= \frac{1}{\sum \exp(\dots)} \sum_{i=1}^N \underbrace{\exp(\dots)}_{\omega_i} \nabla_{x_t \|x_t - y_i\|_2^2} \\ &= \frac{1}{\sum \omega_i} \nabla_{x_t \sum_{i=1}^N \omega_i \|x_t - y_i\|_2^2}. \end{aligned}$$

This gradient formulation is quite similar (up to the weighting) to UniGuide’s special case $\mathcal{S} = \mathcal{Z}$, as it computes an L_2 loss on a given conformation $(\{y_i\})$, see Eq. (18), meaning that it does not generalise to arbitrary geometric conditions.

We emphasise that UniGuide is more broadly applicable because it separates surface computation from gradient computation, offering two key benefits. First, since the condition map does not require differentiability, there is greater flexibility in computing surface points. Second, the precise geometric intuition behind the condition map makes it easier to adapt to new scenarios, as demonstrated by our application to generating density-guided molecules.

For the empirical comparison, we selected the hyperparameters σ and γ in the surface loss computation to achieve a high DICE score between the implicitly defined surface and the meshes UniGuide utilises for LBDD ($\sigma = 1, \gamma = 2$, DICE > 0.8). Our surface calculations use the Open Drug Discovery Toolkit (ODDT), which assigns specific radii to individual atom types and employs the marching cubes algorithm to generate meshes[80].

We performed several runs around the above-specified hyperparameter configuration. The runs performed similarly, and we report the best result in Table 5. Although validity guidance for LBDD yields low graph similarity, the shape

similarity remains suboptimal compared to UniGuide. Additionally, we frequently encounter numerical instability when computing the guidance term, an issue not present with UniGuide’s formulation of LBDD. One possible explanation for this numerical instability is that the surface is defined implicitly, unlike UniGuide where it is explicitly defined. The explicit definition in UniGuide allows for relating the gradient updates directly to the surface, as shown in Eq. (21).

	Sim _S (↑)	maxSim _S (↑)	Sim _G (↓)	maxSim _G (↓)	Ratio (↑)	Connect. (↑)	Unique. (↑)	Diversity (↑)	QED (↑)
Validity Guidance	0.59	0.76	0.20	0.20	2.96	97%	100%	0.76	0.69
UniGuide (EDM)	0.74	0.86	0.21	0.20	3.53	99%	99%	0.73	0.74

Table 5. Comparison of UniGuide with validity guidance for shape-based generation. We highlight the ratio metric as the most critical indicator, reflecting the balance between shape similarity and graph dissimilarity.

Appendix E. Structure-based drug design

Algorithm 2: Sampling algorithm to generate a ligand conditioned on a protein pocket $\tilde{\mathbf{z}}^{\mathcal{P}}$ using the unconditional joint model $\epsilon_{\theta}(\mathbf{z}_t, t)$, where $\mathbf{z}_t = [\mathbf{z}_t^{\mathcal{M}}, \mathbf{z}_t^{\mathcal{P}}]$, that models the distribution $P(\mathbf{z}^{\mathcal{M}}, \mathbf{z}^{\mathcal{P}})$. The guidance signal is controlled via the guidance strength S . Note that samples from the generative process $p_{\theta}(\mathbf{z}_{t-1}|\mathbf{z}_t)$ are assumed to be CoM-free.

Require: $\tilde{\mathbf{z}}^{\mathcal{P}}, S$
 $\mathbf{z}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ {Sample from normal prior}
for $t = T$ **to** 1 **do**
 $\hat{\mathbf{z}}_0^{\mathcal{P}} = \mathbf{z}_t^{\mathcal{P}} - \sqrt{1 - \bar{\alpha}_t} \epsilon_{\theta}^{\mathcal{P}}(\mathbf{z}_t, t) / \sqrt{\bar{\alpha}_t}$ {Compute the clean data of the pocket}
 $\mathcal{L} = \mathcal{L}(\hat{\mathbf{z}}_0^{\mathcal{P}}, \tilde{\mathbf{z}}^{\mathcal{P}})$
 $\mathbf{g} = (\nabla_{\mathbf{x}_t} \mathcal{L} - \overline{\nabla_{\mathbf{x}_t} \mathcal{L}}, \nabla_{\mathbf{h}_t} \mathcal{L})$ {Compute gradient and subtract the CoM}
 $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{\theta}(\mathbf{z}_t, t) - \lambda(t) \cdot \mathbf{g}$ {Update the mean of the pocket}
 $\mathbf{z}_{t-1} \sim \mathcal{N}(\boldsymbol{\mu}_t, \sigma_t \mathbf{I})$
end for
return $\mathbf{z} = (\mathbf{z}_0^{\mathcal{M}}, \mathbf{z}_0^{\mathcal{P}})$

E.1. SBDD sampling algorithm

We provide the algorithm for inference in the SBDD task scenario in Algorithm 2.

E.2. Ligand-protein generative joint model

SBDD aims to generate a ligand given a protein pocket: $p_{\theta}(\mathbf{z}^{\mathcal{M}}|\mathbf{z}_{\text{test}}^{\mathcal{P}}, t)$. We adopt DiffSBDD^[11], an unconditional joint diffusion model that approximates the joint distribution $p(\mathbf{z}_{\text{data}}^{\mathcal{M}}, \mathbf{z}_{\text{data}}^{\mathcal{P}})$ of generating ligand-protein pairs, where the noise predictor $\epsilon_{\theta}(\mathbf{z}_t^{\mathcal{M}}, \mathbf{z}_t^{\mathcal{P}}, t)$ is parametrised by EGNN. DiffSBDD is available under the MIT License. To process ligand

and pocket nodes with a single GNN, atom types and residue types are embedded jointly. Atom and residue features are then decoded separately using atom decoder and residue decoder to $\epsilon_{\theta}^{\mathcal{M}}(\mathbf{z}_t^{\mathcal{M}}, \mathbf{z}_t^{\mathcal{P}}, t)$ and $\epsilon_{\theta}^{\mathcal{P}}(\mathbf{z}_t^{\mathcal{M}}, \mathbf{z}_t^{\mathcal{P}}, t)$ ^[11].

For the unconditional sampling with the joint model, the number of ligand and pocket nodes is sampled from the joint node distribution $p(N^{\mathcal{M}}, N^{\mathcal{P}})$, measured across a training set of $(\mathcal{M}, \mathcal{P})$ pairs. During the modified generative process with the inpainting-inspired technique or with UniGuide the number of pocket nodes is set to be equal to the number of nodes in $\mathcal{P}_{\text{test}}$, while the size of the ligand is generated from a conditional distribution $p(N^{\mathcal{M}}|N^{\mathcal{P}})$. Since this sampling procedure leads to ligands that are much smaller compared to the reference ligands found in the test set, the mean size of sampled ligands is increased by 10 for Binding MOAD and 5 for CrossDocked during ligand generation^[11]. We utilize the unconditional base models from^[11], which are trained on either the C_{α} or full-atom context from the Binding MOAD or CrossDocked datasets. However, we retrain the DiffSBDD model specifically on the full-atom context of the CrossDocked data, as we were unable to reproduce the reported results in this configuration from^[11]. We find that contrary to what is reported in^[11], the model converges early and does not need a full 1000 epochs to fully train. We employ this checkpoint to evaluate both the DiffSBDD inpainting-inspired approach as well as UniGuide. We train the model on four NVIDIA A100 GPU with a batch size of 2. 8 training epochs take approximately 24 hours.

Representing ligands and proteins as graphs

Proteins consist of amino acids, where every amino acid is a set of amino (NH), carboxyl (CO), α -carbon atom and a side chain (R) that is specific to every amino acid type^[81]. The C_{α} -representation of a protein pocket is a residue-level graph, in which the node features of the protein are represented as one-hot encodings of the amino acid type. The full-atom representation of the receptor is an atom-level graph and represents the full context of the protein pocket. Details on processed graphs of the joint model $p(\mathbf{z}^{\mathcal{M}}, \mathbf{z}^{\mathcal{P}})$ are provided in Table 6. We refer the reader to^[11] for more information on the hyperparameters of the joint model.

	CrossDocked		Binding Moad	
	joint C_{α} model	joint full-atom model	joint C_{α} model	joint full-atom model
Edges (ligand-ligand)	fully connected	fully connected	fully connected	fully connected
Edges (ligand-pocket)	< 5 Å	< 5 Å	< 8 Å	< 7 Å
Edges (pocket-pocket)	< 5 Å	< 5 Å	< 8 Å	< 4 Å

Table 6. Hyperparameters of ligand and proteins graphs in joint models

Dataset	R	T	QED ($\uparrow$)	SA ($\uparrow$)	Lipinski ($\uparrow$)	Diversity ($\uparrow$)	Connectivity ($\uparrow$)	Validity ($\uparrow$)
C.D. (C_α)	1	500	0.535	0.660	4.741	0.772	0.893	0.986
C.D. (C_α)	10	50	0.578	0.752	4.836	0.774	0.994	0.986
B.M. (C_α)	1	500	0.471	0.608	4.783	0.824	0.839	0.985
B.M. (C_α)	10	50	0.544	0.665	4.883	0.823	0.961	0.992

Table 7. Quantitative evaluation of samples generated by the unconditional joint models^[11] trained on Crossdocked (C.D.) and Binding MOAD (B.M). We report the mean over all generated ligands.

E.3. Further Comparison to DiffSBDD

In addition to Table 2, we follow the experimental setup as utilised in^[11] to compare UniGuideto DiffSBDD, which uses the same base model, in particular. In Table 8, we further investigate the advantages of using self-guidance in combinations with UniGuide over both the conditional DiffSBDD model (DiffSBDD-cond) as well as the inpainting-inspired technique (DiffSBDD). UniGuide reliably achieves superior VINA Dock scores compared to both DiffSBDD models and performs competitively with the conditional TargetDiff model. In E.4 and E.5, we expand on this experimental comparison with further analysis of the effects of Resampling as well as the guidance strength.

		Vina (↓)	Vina Top 10% (↓)	QED (↑)	SA (↑)	Lipinski (↑)	Diversity (↑)	RMSD (↓)
CrossDocked	Test Set	-6.865 ± 2.35	-	0.476 ± 0.20	0.728 ± 0.14	4.340 ± 1.14	-	-
	3D-SBDD*[38]	-5.888 ± 1.91	-7.289 ± 2.34	0.502 ± 0.17	0.675 ± 0.14	4.787 ± 0.51	0.742 ± 0.09	-
	Pocket2Mol*[48]	-7.058 ± 2.80	-8.712 ± 3.18	0.572 ± 0.16	0.752 ± 0.12	4.936 ± 0.27	0.735 ± 0.15	-
	Graph-BP*[49]	-4.719 ± 4.03	-7.165 ± 1.40	0.502 ± 0.12	0.307 ± 0.09	4.883 ± 0.37	0.844 ± 0.01	-
	TargetDiff*[50]	-7.318 ± 2.47	-9.669 ± 2.55	0.483 ± 0.20	0.584 ± 0.13	4.594 ± 0.83	0.718 ± 0.09	0.000 ± 0.00
	DiffSBDD-cond*	-6.950 ± 2.06	-9.120 ± 2.16	0.469 ± 0.21	0.578 ± 0.13	4.562 ± 0.89	0.728 ± 0.07	0.000 ± 0.00
	DiffSBDD	-7.216 ± 2.54	-9.490 ± 2.00	0.571 ± 0.19	0.639 ± 0.14	4.808 ± 0.50	0.707 ± 0.09	0.045 ± 0.01
Bind. MOAD	UniGuide	-7.320 ± 2.27	-9.514 ± 2.04	0.571 ± 0.19	0.638 ± 0.14	4.822 ± 0.47	0.705 ± 0.08	0.047 ± 0.01
	Test Set	-8.331 ± 2.05	-	0.602 ± 0.15	0.636 ± 0.08	4.838 ± 0.37	-	-
	Graph-BP*[49]	-4.843 ± 2.24	-6.629 ± 0.95	0.512 ± 0.11	0.310 ± 0.09	4.945 ± 0.27	0.826 ± 0.01	0.000 ± 0.00
	DiffSBDD-cond	-7.172 ± 1.88	-9.174 ± 2.13	0.430 ± 0.20	0.564 ± 0.12	4.526 ± 0.80	0.711 ± 0.08	0.000 ± 0.00
	DiffSBDD	-7.263 ± 4.19	-9.776 ± 2.25	0.546 ± 0.21	0.618 ± 0.12	4.777 ± 0.54	0.740 ± 0.05	53 ± 31
	UniGuide	-7.661 ± 2.99	-9.864 ± 2.13	0.556 ± 0.20	0.605 ± 0.12	4.799 ± 0.50	0.723 ± 0.05	55 ± 31

Table 8. Quantitative comparison of generated ligands for target pockets from the CrossDocked and Binding MOAD test sets. Results taken from^[11] are indicated with (*). We report mean and standard deviation and highlight the best diffusion-based approach in bold.

E.4. Resampling

Inpainting is introduced for diffusion models to condition outputs with fixed parts^[71] and can be applied for structure-based molecular tasks. Given a model that generates $(\mathbf{z}_t^{\mathcal{M}}, \mathbf{z}_t^{\mathcal{P}})$ pairs at denoising step t , the protein pocket P_t is replaced with the noised representation of protein context $\tilde{\mathbf{z}}_t^{\mathcal{P}}$. This noised representation can be obtained through the forward process of diffusion models as specified in Eq. (2). However, the direct application of this method leads to locally harmonised samples that struggle to incorporate the global context^[71]. In order to effectively harmonise the generated information during the entire generative process,^[71] propose a technique they call “Resampling”. This modifies the reverse Markov chain by moving back and forth in the diffusion process to enable the model to better incorporate the replaced components.

^[11] propose to use the same resampling technique to harmonise the replaced protein context with the ligand, since the replaced receptor is sampled independently of the ligand. During resampling, each latent representation is repeatedly diffused back and forth before advancing to the next time step. We found that resampling further improves the general performance of the unconditional generation, and thus improves the guided generation as well. We report results for this in App. E.5, where we evaluate how the unconditional generation of the joint model is improved across different metrics with added resampling steps. We follow^[11] in using the setting of $R = 10$ resampling steps and $T = 50$ timesteps. While DiffSBDD resamples the ligand and the noised target protein pocket, we resample the guided protein pocket and ligand with UniGuide. In general, the concept of resampling can be applied to harmonise the configuration $\mathbf{z}_t$ with the condition $\mathbf{c}$.

E.5. Guidance parameters

The guidance scale S controls the strength of the guiding signal, see Eq. (7) and it is weighted by $w(t) = \frac{\beta(t)}{\sqrt{\alpha_t}}$ during the generation. We use a constant scale S for structure-based drug design experiments and evaluate for several guidance scale values in Table 9 and Table 10 for models trained on the Binding MOAD dataset with C_α and full-atom representation respectively. The quantitative evaluation on the CrossDocked data is shown under Table 11 and Table 12 with additional metrics reported in Table 7. For the generation with the C_α -models, we generate 100 samples for every test pocket with a batch size of 50. The full generation takes approximately 5 hours for Binding MOAD and 6 hours for CrossDocked. For the DiffSBDD model trained on the Binding MOAD fullatom pocket data, we use a batch size of 15 for the generation. We use a batch size of 2 to sample with the DiffSBDD model trained on CrossDocked (fullatom).

Method	<i>S</i>	<i>R/T</i>	Vina (↓)	Vina Top 10% (↓)	QED (↑)	SA (↑)	Lipinski (↑)	Diversity (↑)	RMSD (↓)	
DiffSBDD-cond (<i>C_α</i>)	-	-	-6.628 ± 1.59	-8.291 ± 1.26	0.481 ± 0.20	0.554 ± 0.11	4.651 ± 0.70	0.714 ± 0.04	0.000 ± 0.00	
DiffSBDD	-	1/500	-6.362 ± 3.04	-8.179 ± 1.24	0.452 ± 0.20	0.541 ± 0.11	4.604 ± 0.76	0.734 ± 0.03	0.008 ± 0.01	
UniGuide	1.0	1/500	-6.519 ± 2.05	-8.227 ± 1.23	0.464 ± 0.20	0.540 ± 0.11	4.627 ± 0.73	0.733 ± 0.03	0.125 ± 0.01	
UniGuide	2.0	1/500	-6.568 ± 2.13	-8.268 ± 1.25	0.471 ± 0.20	0.543 ± 0.11	4.636 ± 0.73	0.735 ± 0.04	0.105 ± 0.25	
UniGuide	3.0	1/500	-6.667 ± 1.92	-8.305 ± 1.28	0.468 ± 0.20	0.542 ± 0.11	4.622 ± 0.73	0.737 ± 0.03	0.072 ± 0.03	
UniGuide	4.0	1/500	-6.587 ± 1.86	-8.293 ± 1.29	0.470 ± 0.20	0.544 ± 0.11	4.636 ± 0.72	0.735 ± 0.03	0.058 ± 0.01	
UniGuide	6.0	1/500	-6.568 ± 1.93	-8.284 ± 1.26	0.468 ± 0.20	0.542 ± 0.11	4.630 ± 0.73	0.734 ± 0.03	0.045 ± 0.01	
UniGuide	7.0	1/500	-6.575 ± 1.86	-8.296 ± 1.28	0.469 ± 0.20	0.544 ± 0.11	4.636 ± 0.72	0.735 ± 0.03	0.043 ± 0.05	
DiffSBDD	-	10/50	-6.896 ± 3.10	-8.962 ± 1.37	0.547 ± 0.20	0.578 ± 0.20	4.754 ± 0.50	0.709 ± 0.05	0.007 ± 0.01	
UniGuide	1.0	10/50	-6.845 ± 3.68	-8.972 ± 1.36	0.547 ± 0.19	0.578 ± 0.13	4.756 ± 0.53	0.709 ± 0.05	0.216 ± 0.21	
UniGuide	2.0	10/50	-6.889 ± 3.83	-9.018 ± 1.40	0.547 ± 0.19	0.577 ± 0.13	4.756 ± 0.52	0.707 ± 0.04	0.279 ± 0.03	
UniGuide	3.0	10/50	-7.050 ± 2.38	-9.051 ± 1.39	0.551 ± 0.18	0.575 ± 0.14	4.763 ± 0.50	0.706 ± 0.04	0.220 ± 0.01	
UniGuide	4.0	10/50	-7.016 ± 2.93	-9.023 ± 1.38	0.552 ± 0.18	0.578 ± 0.14	4.765 ± 0.50	0.708 ± 0.03	0.168 ± 0.05	
UniGuide	6.0	10/50	-7.053 ± 2.91	-9.067 ± 1.39	0.550 ± 0.18	0.579 ± 0.14	4.761 ± 0.51	0.703 ± 0.04	0.146 ± 0.01	
UniGuide	7.0	10/50	-7.076 ± 2.27	-9.038 ± 1.38	0.550 ± 0.18	0.579 ± 0.14	4.767 ± 0.50	0.704 ± 0.04	0.131 ± 0.01	

Table 9. Results for the Binding MOAD test set with the unconditional DiffSBDD base model trained on the C_{α} - representation of the pockets combined with UniGuide and the inpainting-inspired technique DiffSBDD^[11]. We provide results for varying the guidance scales S during our controlled generation. We also report results for the DiffSBDD-cond (C_{α}) model trained on the C_{α} pockets.

Method	S	R/T	Vina ($\downarrow$)	Vina Top 10% ($\downarrow$)	QED ($\uparrow$)	SA ($\uparrow$)	Lipinski ($\uparrow$)	Diversity ($\uparrow$)	RMSD ($\downarrow$)
DiffSBDD-cond	-	-	-7.172 ± 1.88	-9.174 ± 2.13	0.430 ± 0.20	0.564 ± 0.12	4.526 ± 0.80	0.711 ± 0.08	0.0 ± 0.0
DiffSBDD	-	1/500	-6.540 ± 2.00	-8.427 ± 1.39	0.413 ± 0.20	0.531 ± 0.11	4.611 ± 0.77	0.748 ± 0.03	55 ± 31
UniGuide	6.0	1/500	-6.696 ± 1.78	-8.561 ± 1.58	0.407 ± 0.19	0.527 ± 0.11	4.587 ± 0.78	0.740 ± 0.04	55 ± 31
UniGuide	7.0	1/500	-6.683 ± 1.91	-8.575 ± 1.52	0.406 ± 0.19	0.524 ± 0.11	4.579 ± 0.80	0.738 ± 0.04	55 ± 31
UniGuide	8.0	1/500	-6.682 ± 1.77	-8.555 ± 1.52	0.407 ± 0.19	0.526 ± 0.11	4.591 ± 0.78	0.740 ± 0.04	55 ± 31
UniGuide	9.0	1/500	-6.689 ± 1.74	-8.541 ± 1.50	0.403 ± 0.19	0.524 ± 0.11	4.589 ± 0.78	0.738 ± 0.04	55 ± 31
DiffSBDD	-	10/50	-7.263 ± 4.19	-9.776 ± 2.25	0.546 ± 0.21	0.618 ± 0.12	4.777 ± 0.54	0.740 ± 0.05	53 ± 31
UniGuide	5.0	10/50	-7.470 ± 2.97	-9.621 ± 1.84	0.563 ± 0.20	0.605 ± 0.12	4.807 ± 0.50	0.723 ± 0.05	55 ± 31
UniGuide	6.0	10/50	-7.570 ± 3.20	-9.731 ± 1.90	0.566 ± 0.20	0.606 ± 0.12	4.815 ± 0.48	0.722 ± 0.05	55 ± 31
UniGuide	7.0	10/50	-7.639 ± 2.39	-9.793 ± 2.06	0.559 ± 0.20	0.605 ± 0.12	4.804 ± 0.49	0.723 ± 0.05	54 ± 31
UniGuide	8.0	10/50	-7.635 ± 2.71	-9.821 ± 2.07	0.558 ± 0.20	0.605 ± 0.12	4.804 ± 0.50	0.720 ± 0.05	54 ± 31
UniGuide	9.0	10/50	-7.661 ± 2.99	-9.864 ± 2.13	0.556 ± 0.20	0.605 ± 0.12	4.799 ± 0.50	0.723 ± 0.05	55 ± 31

Table 10. Results for the Binding MOAD test set with the unconditional DiffSBDD base model trained on the full-atom context of the pockets combined with UniGuide and the inpainting-inspired technique DiffSBDD^[11]. We provide results for varying the guidance scales S during our controlled generation. We also report results for the conditional diffusion model DiffSBDD-cond.

Method	S	R/T	Vina (↓)	Vina Top 10% (↓)	QED (↑)	SA (↑)	Lipinski (↑)	Diversity (↑)	RMSD (↓)
DiffSBDD-cond (C_α)	-	-	-6.770 ± 2.73	-8.796 ± 1.75	0.475 ± 0.22	0.612 ± 0.12	4.536 ± 0.91	0.725 ± 0.06	0.000 ± 0.00
DiffSBDD	-	1/500	-6.485 ± 2.50	-8.472 ± 1.62	0.510 ± 0.21	0.619 ± 0.12	4.640 ± 0.73	0.735 ± 0.06	0.053 ± 0.03
UniGuide	2.0	1/500	-6.528 ± 2.64	-8.527 ± 1.67	0.518 ± 0.21	0.623 ± 0.12	4.649 ± 0.73	0.739 ± 0.05	0.085 ± 0.01
UniGuide	3.0	1/500	-6.604 ± 2.57	-8.556 ± 1.64	0.519 ± 0.21	0.622 ± 0.12	4.657 ± 0.72	0.738 ± 0.05	0.070 ± 0.01
UniGuide	4.0	1/500	-6.578 ± 2.72	-8.563 ± 1.68	0.518 ± 0.21	0.623 ± 0.12	4.659 ± 0.71	0.741 ± 0.05	0.059 ± 0.02
UniGuide	5.0	1/500	-6.563 ± 2.58	-8.549 ± 1.66	0.516 ± 0.21	0.624 ± 0.12	4.646 ± 0.72	0.741 ± 0.05	0.052 ± 0.01
UniGuide	6.0	1/500	-6.658 ± 2.50	-8.578 ± 1.69	0.527 ± 0.21	0.629 ± 0.12	4.683 ± 0.69	0.741 ± 0.05	0.045 ± 0.01
DiffSBDD	-	10/50	-7.030 ± 3.39	-9.057 ± 1.79	0.559 ± 0.21	0.730 ± 0.12	4.729 ± 0.60	0.720 ± 0.07	0.052 ± 0.01
UniGuide	1.0	10/50	-6.909 ± 3.35	-9.069 ± 1.79	0.563 ± 0.21	0.734 ± 0.12	4.743 ± 0.57	0.721 ± 0.06	0.711 ± 0.12
UniGuide	2.0	10/50	-7.015 ± 3.20	-9.115 ± 1.79	0.562 ± 0.21	0.733 ± 0.12	4.735 ± 0.60	0.721 ± 0.07	0.188 ± 0.02
UniGuide	3.0	10/50	-7.081 ± 2.95	-9.140 ± 1.83	0.560 ± 0.20	0.732 ± 0.11	4.742 ± 0.57	0.723 ± 0.07	0.127 ± 0.01
UniGuide	4.0	10/50	-7.086 ± 3.27	-9.125 ± 1.81	0.561 ± 0.19	0.731 ± 0.10	4.729 ± 0.60	0.719 ± 0.06	0.102 ± 0.01
UniGuide	5.0	10/50	-7.117 ± 2.78	-9.127 ± 1.78	0.561 ± 0.20	0.731 ± 0.12	4.738 ± 0.59	0.722 ± 0.07	0.090 ± 0.01
UniGuide	6.0	10/50	-7.113 ± 3.00	-9.133 ± 1.80	0.556 ± 0.20	0.731 ± 0.12	4.734 ± 0.60	0.720 ± 0.32	0.077 ± 0.01

Table 11. Evaluation of the samples generated for the CrossDocked test set using the joint ligand-protein diffusion model trained on the C_α pocket representation for varying guidance scales S . The base model is combined either with the

inpainting-inspired technique (DiffSBDD) or UniGuide. We further report the evaluation of the molecules generated by the conditional model DiffSBDD-cond that is trained on the C_{α} pocket representation.

Method	S	R/T	Vina (↓)	Vina Top 10% (↓)	QED (↑)	SA (↑)	Lipinski (↑)	Diversity (↑)	RMSD (↓)
DiffSBDD-cond	-	-	-6.950 ± 2.06	-9.120 ± 2.16	0.469 ± 0.21	0.578 ± 0.13	4.562 ± 0.89	0.728 ± 0.07	0.000 ± 0.00
DiffSBDD	-	1/500	-6.225 ± 1.77	-8.115 ± 1.64	0.469 ± 0.20	0.573 ± 0.11	4.691 ± 0.70	0.778 ± 0.04	0.049 ± 0.01
UniGuide	5.0	1/500	-6.346 ± 1.74	-8.208 ± 1.62	0.482 ± 0.20	0.570 ± 0.12	4.718 ± 0.67	0.773 ± 0.04	0.040 ± 0.01
UniGuide	6.0	1/500	-6.335 ± 1.72	-8.225 ± 1.61	0.484 ± 0.20	0.571 ± 0.12	4.715 ± 0.66	0.775 ± 0.04	0.039 ± 0.01
UniGuide	7.0	1/500	-6.338 ± 1.73	-8.218 ± 1.60	0.481 ± 0.19	0.571 ± 0.12	4.710 ± 0.67	0.774 ± 0.04	0.039 ± 0.01
UniGuide	8.0	1/500	-6.366 ± 1.72	-8.261 ± 1.57	0.485 ± 0.20	0.570 ± 0.12	4.717 ± 0.66	0.773 ± 0.03	0.039 ± 0.01
DiffSBDD	-	10/50	-7.216 ± 2.54	-9.490 ± 2.00	0.571 ± 0.19	0.639 ± 0.14	4.808 ± 0.50	0.707 ± 0.09	0.045 ± 0.01
UniGuide	6.0	10/50	-7.295 ± 2.22	-9.441 ± 1.95	0.574 ± 0.19	0.641 ± 0.14	4.825 ± 0.47	0.706 ± 0.08	0.047 ± 0.01
UniGuide	7.0	10/50	-7.320 ± 2.27	-9.514 ± 2.04	0.571 ± 0.19	0.638 ± 0.14	4.822 ± 0.47	0.705 ± 0.08	0.047 ± 0.01
UniGuide	8.0	10/50	-7.298 ± 2.21	-9.460 ± 2.01	0.568 ± 0.19	0.641 ± 0.14	4.818 ± 0.47	0.703 ± 0.09	0.048 ± 0.01
UniGuide	9.0	10/50	-7.265 ± 2.45	-9.495 ± 2.05	0.577 ± 0.19	0.640 ± 0.14	4.821 ± 0.47	0.706 ± 0.08	0.049 ± 0.01

Table 12. Results for the CrossDocked test set with the joint model trained on the full-atom pocket representation of the pocket for varying guidance scales S . The unconditional model is either controlled by the inpainting-inspired technique (DiffSBDD) or UniGuide.

dataset	model	runtime (s)
CrossDocked (C_α)	DiffSBDD-cond	60 ± 68
	DiffSBDD	141 ± 55
	UniGuide	193 ± 61
Binding Moad (C_α)	DiffSBDD-cond	54 ± 42
	DiffSBDD	61 ± 17
	UniGuide	104 ± 36
Binding Moad (full)	DiffSBDD-cond	345 ± 55
	DiffSBDD	398 ± 95
	UniGuide	453 ± 120

Table 13. We evaluate the runtime of UniGuide and compare it to DiffSBDD-cond and DiffSBDD from ^[11]. We report the average time (in seconds) to generate 100 ligands per pocket for the CrossDocked (C_α), Binding Moad (C_α) and Binding Moad (fullatom).

		Validity ($\uparrow$)	Connectivity ($\uparrow$)	Uniqueness ($\uparrow$)	Novelty ($\uparrow$)
Cross- Docked	Test Set	100%	100%	96.00%	96.88%
	DiffSBDD-Cond (C_α)	95.32%	80.63%	99.97%	99.81%
	DiffSBDD-Cond	97.32%	78.91%	99.99%	99.91%
	DiffSBDD (C_α)	99.20%	98.14%	99.26%	99.16%
	DiffSBDD	97.76%	89.84%	99.94%	99.87%
	UniGuide (C_α)	99.12%	98.35%	99.50%	99.24%
	UniGuide	97.40%	93.18%	99.93%	99.76%
Binding MOAD	Test Set	97.69%	100%	38.58%	77.55%
	DiffSBDD-cond (C_α)	94.43%	77.17%	100%	100%
	DiffSBDD-cond	96.20%	63.20%	100%	100%
	DiffSBDD (C_α)	98.54%	91.45%	100%	100%
	DiffSBDD	94.22%	75.60%	100%	100%
	UniGuide (C_α)	98.44%	93.12%	100%	99.99%
	UniGuide	93.85%	79.95%	100%	100%

Table 14. Additional metrics for the methods discussed in Sec. 5.2.

For all tables, we conduct the experiments both with and without resampling. The VINA Dock score is measured with QuickVina2^[72], available under the Apache License, and the chemical properties (QED, SA, Lipinski) are measured with RDKit. We note that in all ablation tables we measure the VINA Dock score on the processed molecules, following^[11], while the VINA Dock score in Table 2 is measured following^[67]. Both the VINA Dock score and chemical properties improve with additional resampling steps ($R = 10, T = 50$) for both datasets. Additionally, increasing the guidance scale improves the RMSD with respect to the target protein, and results in generating ligands with an improved binding affinity (lower VINA).

E.6. Additional Results for SBDD

Supplementary to Table 2 we provide additional metrics for the evaluation of the generated ligands in Table 14: the validity as measured by RDKit^[73] and the connectivity, representing the percentage of valid molecules without any disconnected fragments. Additionally, we report the uniqueness and novelty of the valid connected ligands.

E.7. Runtime Comparison

In Table 13, we provide a comparison of the different controlled generation mechanisms regarding their runtime. While UniGuide has a higher runtime compared to other conditioning mechanisms, as it has to compute gradients through the diffusion model at inference time, it stays comparable to other mechanisms such as inpainting.

Appendix F. Fragment-based drug design

F.1. Linker Design

For the experimental evaluation of the linker design task, we follow^[13], employ the ZINC dataset^[74] and preprocess it following^[13]. That is, 3D conformers are generated from the SMILES strings present in the dataset with RDKit^[73]. We fragment the dataset ligands using an MMPA-based algorithm^{[75][73]}, generating multiple fragment conditions per molecule. We train an unconditional EDM model for this task as specified in C. For the evaluation metrics, we follow^[13]. Note that the synthetic accessibility score computation (SA) in Table 3 differs from the remaining experimental evaluations. While^[13] report the SA score s_{SA} directly, ^[11] report the SA score as $(10 - s_{SA})/9$.

For the task of linker design, we adjust the condition map as discussed in Sec. 4.2 slightly to include anchor information, similar in spirit to the DiffLinker model incorporating anchor information^[13]. That is, additionally to guiding parts of the molecule to the desired fragment configuration, we additionally define a cuboid’s surface that is defined from the specified anchor atoms. We can then utilise this surface condition $C_{\partial V}$ to guide the linker atoms in accordance with Eq. (21). Additionally, we can expand this surface based on the linker size to ensure chemical validity of the generated linker. This condition map highlights the flexibility of UniGuide condition maps in various tasks, especially through the combination of two definitions of the condition map. For the experimental evaluation, we sample the size of the linker nodes uniformly in accordance with^[13] and compare to the DiffLinker model without an external network to predict the linker size. Note, however, that also the unconditional EDM model combined with UniGuide can be adapted to include such predictors.

F.2. General Fragment Conditions

To assess the performance of UniGuide for the task of FBDD, we create an experimental setup with the goal of generating ligands conditioned on desired fragments roughly following^[13]. We select 10 random protein targets from the Binding MOAD dataset and decompose their corresponding reference ligands using an MMPA-based algorithm^[75]^[73]. This decomposition results in a set of 40 different scenarios, including separated fragments we want to link, a fragment to grow or small functional groups to perform scaffolding. For every set of fixed fragments, we aim to guide the unconditional generation of ligands towards the generation of a ligand containing the desired fragments. As the protein is not the target of the guidance, we employ the DiffSBDD-cond model, which is conditionally trained on the (C_{α}) -representation of the protein pocket. For every set of fixed fragments, we generate 100 ligands and use a constant guidance scale of 8.

We provide quantitative results for the task of fragment-based drug design in Table 15. On the one hand, the task requires the desired fragments to be present in the generated molecule. Thus, we measure the success rate of recovery (Hit Ratio) and the RMSD between the generated fragments and desired fragments. On the other hand, given that the target fragments are met in the generated ligand, the generation has to achieve favourable chemical properties, high binding affinity, as well as high diversity within the set of generated ligands and low similarity to the reference ligand. As the Inpaint mechanism enforces the fragment during generation more strictly, it is able to achieve a better Hit Ratio and RMSD. Nevertheless, UniGuide achieves competitive results but also better VINA docking scores, better properties, and lower similarity compared to the reference ligand.

The FBDD task puts a hard constraint on the generated ligands, namely that a set of desired fragments has to be present in the generated ligand. However, neither DiffSBDD nor UniGuide can guarantee that the condition fragments are present in the generated samples.

We provide further qualitative results of the generated ligands for the FBDD task in Fig. 7.

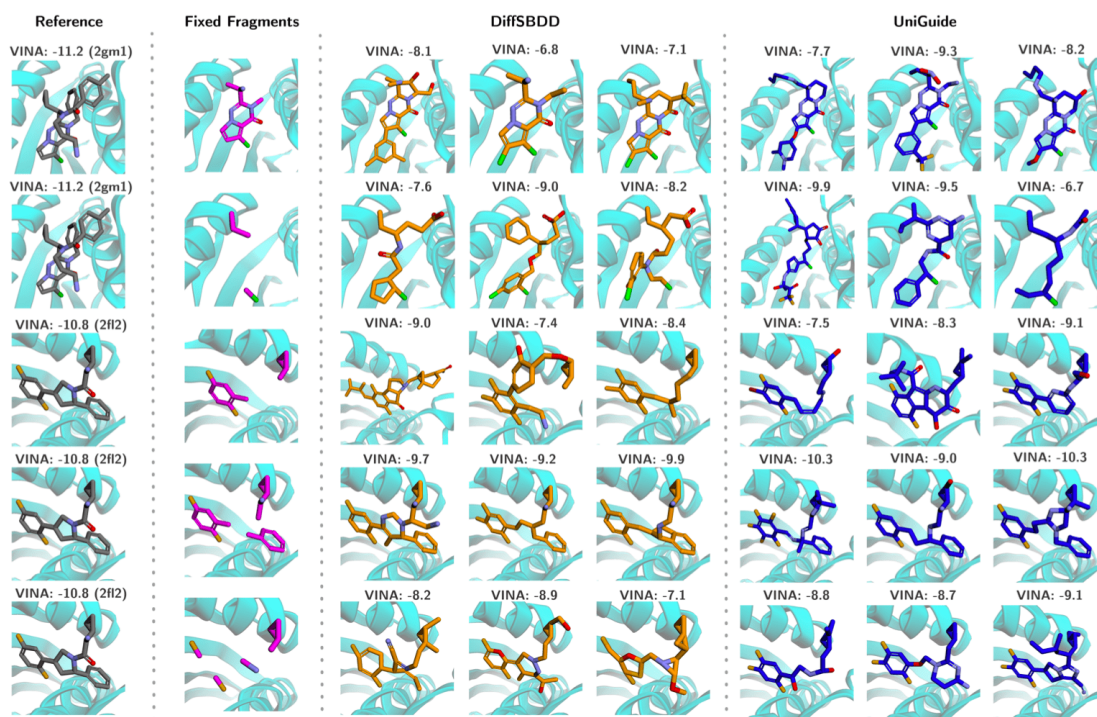


Figure 7. Examples of the generated fragment conditioned ligands.

	DiffSBDD	UniGuide
Vina ($\downarrow$)	-7.406 $\pm$ 0.79	-7.924 $\pm$ 0.89
QED ($\uparrow$)	0.612 $\pm$ 0.11	0.639 $\pm$ 0.09
SA ($\uparrow$)	0.703 $\pm$ 0.11	0.691 $\pm$ 0.10
Lipinski ($\uparrow$)	4.819 $\pm$ 0.28	4.875 $\pm$ 0.19
Diversity ($\uparrow$)	0.653 $\pm$ 0.28	0.669 $\pm$ 0.23
Similarity ($\downarrow$)	0.172 $\pm$ 0.02	0.177 $\pm$ 0.02
Validity ($\uparrow$)	93.35 %	94.41 %
Connectivity ($\uparrow$)	66.87 %	68.30 %

Table 15. Quantitative comparison between DiffSBDD and UniGuide for the FBDD task on the Binding MOAD (C_α) dataset. As the condition in this FBDD scenario is a hard constraint that entails the condition to be exactly present in the generation, we add a post-hoc step for both methods where we replace the inpainted or guided parts with the exact condition atoms. We report mean and standard deviation and highlight the best method in bold.

Appendix G. Atom densities in 3D space

Similar to the guidance by the volume enclosed by the molecular surface, UniGuide allows to guide towards multiple point clouds simultaneously. A natural extension of LBDD would be to harness atom densities as described in [82]. Such a setting combines aspects of LBDD and SBDD as it provides conditions also on the feature space, yet the source can only be represented by point clouds.

In particular, we anticipate UniGuide to be useful in scenarios where explicit information about advantageous features of the ligand is provided in the form of 3D densities. Examples of this include a) volumetric densities that indicate beneficial placement of certain atom types, such as oxygen atoms [82] or b) pharmacophore-like retrieval of advantageous positions for aromatic rings, as utilised in e.g. [83]. On a technical level, this setting assumes that instead of a reference ligand's structure, we only have access to (multiple) atom type densities that indicate preferred locations for optimal interaction with the protein. Additionally, instead of conditioning on a reference ligand's shape, we could condition on a protein pocket's surface, which primarily defines exclusion zones rather than precise atom placement.

Adapting UniGuide for such scenarios requires only minor adjustments, as the protein surface can be treated like shapes in standard LBDD, defining an exclusion zone based on proximity to the surface. The atom densities are thresholded to reflect regions of high interest and converted to surfaces using the marching cubes algorithm [84]. To also include feature information, we effectively employ a modified condition map similar to Eq. (21) that extends the

transformation from the conformation to the configuration space. Moreover, the number of atoms guided by each density is adjusted based on its volume, reflecting the varying influence of each density, and guidance is only applied if atoms are sufficiently close.

We show explorative results for the guided generation of molecules towards desired atom densities using UniGuide in Fig. 8. While our current approach represents a promising first step in tackling this task, we acknowledge the potential for further refinement and are eager to explore future improvements within the UniGuide framework.

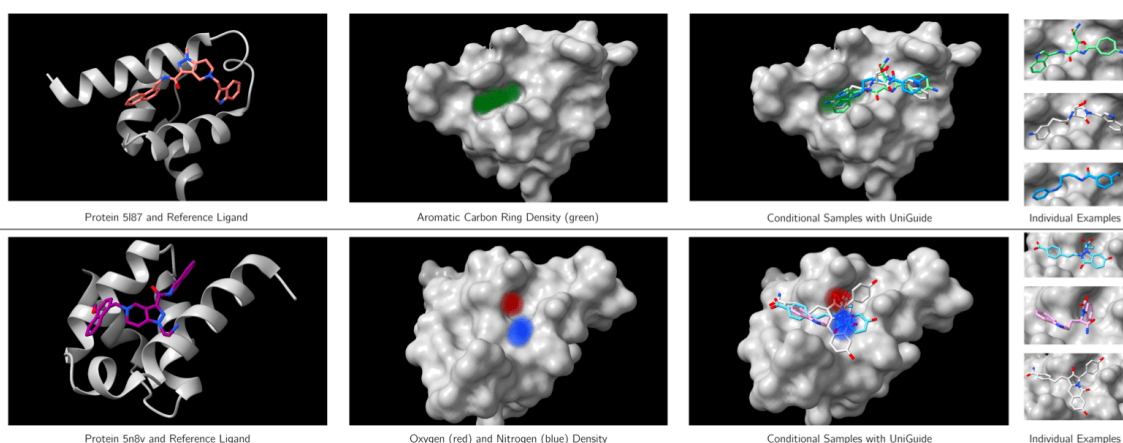


Figure 8. Given a source density of oxygens, we can extend UniGuide to generate ligands satisfying the condition.

Notes

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

Project Page: www.cs.cit.tum.de/daml/uniguide

Acknowledgements

SA, LH, and JS are thankful for valuable feedback from Marcel Kollovich, Leo Schwinn, and Alessandro Palma from the DAML group and Theis Lab. SA is supported by the DAAD programme Konrad Zuse Schools of Excellence in Artificial Intelligence, sponsored by the Federal Ministry of Education and Research. LH is supported by the Helmholtz Association under the joint research school “Munich School for Data Science - MUDS”. FJT acknowledges support from the Helmholtz Association’s Initiative and Networking Fund through Helmholtz AI (ZT-I-PF-5-01). FJT further acknowledges support by the BMBF (01IS18053A). In addition, FJT consults for Immunai Inc., Singularity Bio B.V., CytoReason Ltd, and Omniscope Ltd and has an ownership interest in Dermagnostix GmbH and Cellarity.

References

1. ^{a, b}Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022. p. 10684–10695.
2. ^{a, b}Chen N, Zhang Y, Zen H, Weiss RJ, Norouzi M, Chan W (2020). "Wavegrad: Estimating gradients for waveform generation". *arXiv preprint arXiv:2009.00713*. Available from: <https://arxiv.org/abs/2009.00713>.
3. ^ΔAnand N, Achim T. "Protein structure and sequence generation with equivariant denoising diffusion probabilistic models". *arXiv preprint arXiv:2205.15019*. 2022. doi:[10.48550](https://doi.org/10.48550).
4. ^ΔTrippe BL, Yim J, Tischer D, Baker D, Broderick T, Barzilay R, Jaakkola T. Diffusion probabilistic modeling of protein backbones in 3D for the motif-scaffolding problem. March 2023. Available from: <http://arxiv.org/abs/2206.04119>. arXiv:2206.04119 [cs, q-bio, stat].
5. ^ΔWatson JL, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, Ahern W, Borst AJ, Ragotte RJ, Milles LF, et al. De novo design of protein structure and function with RFdiffusion. *Nature*. 620(7976):1089–1100, 2023.
6. ^ΔCorso G, Jing B, Barzilay R, Jaakkola T, et al. DiffDock: Diffusion steps, twists, and turns for molecular docking. In: *International conference on learning representations (ICLR 2023)*; 2023.
7. ^{a, b}Du Y, Fu T, Sun J, Liu S. MolGenSurvey: A Systematic Survey in Machine Learning Models for Molecule Design. *Technical Report arXiv:2203.14500*. arXiv; 2022 Mar. Available from: <http://arxiv.org/abs/2203.14500>. doi:[10.48550/arXiv.2203.14500](https://doi.org/10.48550/arXiv.2203.14500).
8. ^ΔHetzel L, Böhm S, Kilbertus N, Günnemann S, Lotfollahi M, Theis F (2022). "Predicting single-cell perturbation responses for unseen drugs." *ICLR2022 Machine Learning for Drug Discovery*. doi:[10.48550/arXiv.2204.13545](https://doi.org/10.48550/arXiv.2204.13545). arXiv:[2204.13545](https://arxiv.org/abs/2204.13545).
9. ^{a, b, c, d}Ho J, Jain A, Abbeel P (2020). "Denoising diffusion probabilistic models". *Advances in neural information processing systems*. 33: 6840–6851.
10. ^{a, b}Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B (2020). "Score-based generative modeling through stochastic differential equations". *arXiv preprint arXiv:2011.13456*. [arXiv:2011.13456](https://arxiv.org/abs/2011.13456).
11. ^{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, Δ}Schneuing A, Du Y, Harris C, Jamasb A, Igashov I, Du W, Blundell T, Lió P, Gomes C, Welling M, Bronstein M, Correia B. Structure-based Drug Design with Equivariant Diffusion Models [Internet]. 2023 Jun [cited 2023 Aug 22]. Available from: <http://arxiv.org/abs/2210.13695>
12. ^{a, b, c}Torge J, Harris C, Mathis SV, Lio P (2023). "Diffhopp: A graph diffusion model for novel drug design via scaffold hopping". *arXiv preprint arXiv:2308.07416*. Available from: <https://arxiv.org/abs/2308.07416>.
13. ^{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, Δ}Igashov I, Stärk H, Vignac C, Garcia Satorras V, Frossard P, Welling M, Bronstein M, Correia B (2022). "Equivariant 3D-Conditional Diffusion Models for Molecular Linker Design". *arXiv*. doi:[10.48550/arXiv.2210.05274](https://doi.org/10.48550/arXiv.2210.05274). Available from: <http://arxiv.org/abs/2210.05274>.
14. ^{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r}Chen Z, Peng B, Parthasarathy S, Ning X (2023). "Shape-conditioned 3D molecule generation via equivariant diffusion models". *arXiv preprint arXiv:2308.11890*.
15. ^ΔLi X, Thickstun J, Gulrajani I, Liang PS, Hashimoto TB (2022). "Diffusion-lm improves controllable text generation". *Advances in Neural Information Processing Systems*. 35: 4328–4343.

16. [△]Zeng X, Vahdat A, Williams F, Gojcic Z, Litany O, Fidler S, Kreis K (2022). "LION: Latent point diffusion models for 3D shape generation". arXiv preprint arXiv:2210.06978. 2022.
17. [△]Ho J, Salimans T (2022). "Classifier-free diffusion guidance". arXiv preprint arXiv:2207.12598. Available from: <https://arxiv.org/abs/2207.12598>.
18. [△]Nichol A, Dhariwal P, Ramesh A, Shyam P, Mishkin P, McGrew B, Sutskever I, Chen M (2021). "Glide: Towards photorealistic image generation and editing with text-guided diffusion models". arXiv preprint arXiv:2112.10741. [arXiv:2112.10741](https://arxiv.org/abs/2112.10741).
19. [△]Wang W, Bao J, Zhou W, Chen D, Chen D, Yuan L, Li H (2022). "Semantic image synthesis via diffusion models". arXiv preprint arXiv:2207.00050. Available from: <https://arxiv.org/abs/2207.00050>.
20. [△]_{a, b, c, d} Hoogeboom E, Garcia Satorras V, Vignac C, Welling M. Equivariant Diffusion for Molecule Generation in 3D. June 2022. Available from: <http://arxiv.org/abs/2203.17003>. arXiv:2203.17003 [cs, q-bio, stat].
21. [△]Avrahami O, Lischinski D, Fried O (2022). "Blended diffusion for text-driven editing of natural images". Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pages 18208–18218.
22. [△]_{a, b, c} Dhariwal P, Nichol A (2021). "Diffusion models beat gans on image synthesis". Advances in neural information processing systems. 34: 8780–8794.
23. [△]Bansal A, Chu HM, Schwarzschild A, Sengupta S, Goldblum M, Geiping J, Goldstein T (2023). "Universal guidance for diffusion models". Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pages 843–852.
24. [△]_{a, b, c} Bao F, Zhao M, Hao Z, Li P, Li C, Zhu J (2023). "Equivariant Energy-Guided SDE for Inverse Molecular Design". arXiv. doi:10.48550/arXiv.2209.15408. Available from: <http://arxiv.org/abs/2209.15408>.
25. [△]Jin W, Barzilay R, Jaakkola T. Junction Tree Variational Autoencoder for Molecular Graph Generation. Technical report. International Conference on Machine Learning; 2018. Available from: <https://arxiv.org/abs/1802.04364>.
26. [△]Maziarz K, Jackson-Flux H, Cameron P, Sirockin F, Schneider N, Stiefl N, Segler M, Brockschmidt M. Learning to Extend Molecular Scaffolds with Structural Motifs. Technical Report arXiv:2103.03864. arXiv; 2022 Apr. Available from: <http://arxiv.org/abs/2103.03864>.
27. [△]Jin W, Barzilay R, Jaakkola T (2020). "Hierarchical Generation of Molecular Graphs using Structural Motifs". Technical report, International Conference on Machine Learning.
28. [△]Vignac C, Krawczuk I, Siraudin A, Wang B, Cevher V, Frossard P. DiGress: Discrete Denoising diffusion for graph generation. arXiv. 2022 Oct. Available from: <http://arxiv.org/abs/2209.14734>. doi:10.48550/arXiv.2209.14734.
29. [△]Kong X, Huang W, Tan Z, Liu Y. Molecule Generation by Principal Subgraph Mining and Assembling. arXiv. September 2022. Available from: <http://arxiv.org/abs/2106.15098>.
30. [△]De Cao N, Kipf T. MolGAN: An implicit generative model for small molecular graphs. Technical Report arXiv:1805.11973, arXiv; 2018 May. Available from: <http://arxiv.org/abs/1805.11973>. doi:10.48550/arXiv.1805.11973.
31. [△]Hetzel L, Sommer J, Rieck B, Theis F, G\u00f6rnermann S (2023). "MAGNet: Motif-agnostic generation of molecules from shapes". arXiv preprint arXiv:2305.19303. Available from: <https://arxiv.org/abs/2305.19303>.
32. [△]Geng Z, Xie S, Xia Y, Wu L, Qin T, Wang J, Zhang Y, Wu F, Liu T-Y. "De Novo Molecular Generation via Connection-aware Motif Mining". arXiv. February 2023. Available from: <http://arxiv.org/abs/2302.01129>. arXiv:2302.01129 [cs].

33. [△]Sommer J, Hetzel L, Lüdke D, Theis FJ, Günnemann S. The Power of Motifs as Inductive Bias for Learning Molecular Distributions. 2023 Mar. Available from: https://openreview.net/forum?id=cS3_jJose3z.
34. [△]Ketata MA, Gao N, Sommer J, Wollschläger T, Günnemann S. Lift Your Molecules: Molecular Graph Generation in Latent Euclidean Space. 2024 Jun. Available from: <http://arxiv.org/abs/2406.10513>. doi:10.48550/arXiv.2406.10513. arXiv:2406.10513.
35. [△]Ragoza M, Masuda T, Koes DR (2020). "Learning a continuous representation of 3D molecular structures with deep generative models". arXiv preprint arXiv:2010.08687. Available from: <https://arxiv.org/abs/2010.08687>.
36. [△]Gebauer NWA, Gastegger M, Schütt KT. Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules. arXiv. January 2020. Available from: <http://arxiv.org/abs/1906.00957>. doi:10.48550/arXiv.1906.00957.
37. [△]Luo Y, Ji S (2021). "An Autoregressive Flow Model for 3D Molecular Geometry Generation from Scratch". G-SphereNet. A available from: <https://openreview.net/forum?id=Co3Ajc-NS5W>.
38. ^{a, b, c, d, e, f}Luo S, Guan J, Ma J, Peng J (2021). "A 3D generative model for structure-based drug design". Advances in Neural Information Processing Systems. 34: 6229–6239.
39. [△]Satorras VG, Hoogeboom E, Fuchs FB, Posner I, Welling M. "E(n) Equivariant Normalizing Flows". arXiv. January 2022. Available from: <http://arxiv.org/abs/2105.09016>. doi:10.48550/arXiv.2105.09016.
40. [△]Song Y, Gong J, Xu M, Cao Z, Lan Y, Ermon S, Zhou H, Ma W. Equivariant flow matching with hybrid probability transport for 3D molecule generation. In: Thirty-seventh conference on neural information processing systems; 2023.
41. [△]Xu M, Powers A, Dror R, Ermon S, Leskovec J. Geometric Latent Diffusion Models for 3D Molecule Generation. arXiv. May 2023. Available from: <http://arxiv.org/abs/2305.01140>.
42. [△]Wu L, Gong C, Liu X, Ye M, Liu Q. Diffusion-based Molecule Generation with Informative Prior Bridges [Internet]. arXiv; 2022 Sep [cited 2023 Aug 22]. Available from: <http://arxiv.org/abs/2209.00865>.
43. [△]Morehead A, Cheng J (2023). "Geometry-Complete Diffusion for 3D Molecule Generation and Optimization". arXiv. doi: 10.48550/arXiv.2302.04313. Available from: <http://arxiv.org/abs/2302.04313>.
44. [△]Qiang B, Song Y, Xu M, Gong J, Gao B, Zhou H, Ma W, Lan Y (2023). "Coarse-to-fine: a hierarchical diffusion model for molecule generation in 3d". In: International conference on machine learning. PMLR. pp. 28277–28299.
45. [△]Vignac C, Osman N, Toni L, Frossard P (2023). "Midi: Mixed graph and 3d denoising diffusion for molecule generation". arXiv preprint arXiv:2302.09048. Available from: <https://arxiv.org/abs/2302.09048>.
46. [△]Huang H, Sun L, Du B, Lv W (2023). "Learning Joint 2D & 3D Diffusion Models for Complete Molecule Generation". arXiv. doi:10.48550/arXiv.2305.12347. Available from: <http://arxiv.org/abs/2305.12347>.
47. [△]Huang L, Zhang H, Xu T, Wong KC. MDM: Molecular Diffusion Model for 3D Molecule Generation. 2022 Sep. Available from: <http://arxiv.org/abs/2209.05710>.
48. ^{a, b, c, d}Peng X, Luo S, Guan J, Xie Q, Peng J, Ma J (2022). "Pocket2mol: Efficient molecular sampling based on 3D protein pockets". In: International Conference on Machine Learning. PMLR. pp. 17644–17655.
49. ^{a, b, c}Liu M, Luo Y, Uchino K, Maruhashi K, Ji S (2022). "Generating 3d molecules for target protein binding". arXiv preprint arXiv:2204.09410. Available from: <https://arxiv.org/abs/2204.09410>.

50. ^{a, b, c, d, e, f}Guan J, Qian WW, Peng X, Su Y, Peng J, Ma J (2023). "3d equivariant diffusion for target-aware molecule generation and affinity prediction". arXiv preprint arXiv:2303.03543. Available from: <https://arxiv.org/abs/2303.03543>.
51. ^ΔImrie F, Hadfield TE, Bradley AR, Deane CM (2021). "Deep generative design with 3D pharmacophoric constraints". *Chemical science*. 12 (43): 14577–14589.
52. ^{a, b, c, d}Huang Y, Peng X, Ma J, Zhang M (2022). "3DLinker: an E (3) equivariant variational autoencoder for molecular linker design". arXiv preprint arXiv:2205.07309.
53. ^{a, b, c}Imrie F, Bradley AR, van der Schaar M, Deane CM (2020). "Deep generative models for 3D linker design". *Journal of chemical information and modeling*. 60 (4): 1983–1995. Publisher: ACS Publications.
54. ^ΔGuan J, Peng X, Jiang P, Luo Y, Peng J, Ma J (2023). "LinkerNet: Fragment Poses and Linker Co-Design with 3D Equivariant Diffusion". *Advances in Neural Information Processing Systems*. 36: 77503–77519. Available from: https://proceedings.neurips.cc/paper_files/paper/2023/hash/f4821075019a058700f6e6738ee1365-Abstract-Conference.html.
55. ^{a, b, c, d}Adams K, Coley CW (2022). "Equivariant shape-conditioned generation of 3d molecules for ligand-based drug design". arXiv preprint arXiv:2210.04893. Available from: <https://arxiv.org/abs/2210.04893>.
56. ^ΔLong S, Zhou Y, Dai X, Zhou H (2022). "Zero-shot 3d drug design by sketching and generating". *Advances in Neural Information Processing Systems*. 35: 23894–23907.
57. ^ΔSohl-Dickstein J, Weiss E, Maheswaranathan N, Ganguli S. Deep unsupervised learning using nonequilibrium thermodynamics. In: *International conference on machine learning*. PMLR; 2015. p. 2256–2265.
58. ^ΔKollovieh M, Ansari AF, Bohlke-Schneider M, Zschiegner J, Wang H, Wang Y (2023). "Predict, refine, synthesize: Self-guiding diffusion models for probabilistic time series forecasting". arXiv preprint arXiv:2307.11494. Available from: <https://arxiv.org/abs/2307.11494>.
59. ^ΔKabsch W (1976). "A solution for the best rotation to relate two sets of vectors". *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*. 32 (5): 922–923.
60. ^ΔKabsch W (1978). "A discussion of the solution for the best rotation to relate two sets of vectors". *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*. 34 (5): 827–828.
61. ^ΔBesl PJ, McKay ND (1992). "Method for registration of 3-D shapes." In: *Sensor fusion IV: control paradigms and data structures*. Vol. 1611. Spie. p. 586–606.
62. ^ΔAbel R, Bhat S. Chapter Seven – Free Energy Calculation Guided Virtual Screening of Synthetically Feasible Ligand R-Group and Scaffold Modifications: An Emerging Paradigm for Lead Optimization. In: Goodnow RA, editor. *Annual Reports in Medicinal Chemistry. Platform Technologies in Drug Discovery and Validation*. Academic Press; 2017. p. 237–262. doi:10.1016/bs.armac.2017.08.007. Available from: <https://www.sciencedirect.com/science/article/pii/S0065774317300106>.
63. ^ΔLi Q. Application of Fragment-Based Drug Discovery to Versatile Targets. *Frontiers in Molecular Biosciences*. 7: 180, 2020. doi:10.3389/fmolb.2020.00180. PMID 32850968. PMCID PMC7419598.
64. ^{a, b}Böhm HJ, Flohr A, Stahl M (2004). "Scaffold hopping". *Drug Discovery Today. Technologies*. 1 (3): 217–224. doi:10.1016/j.ddtec.2004.10.009. PMID 24981488.

65. ^{a, b}Sheng C, Zhang W (2013). "Fragment informatics and computational fragment-based drug design: an overview and update". *Medicinal Research Reviews*. 33 (3): 554–598. doi:[10.1002/med.21255](https://doi.org/10.1002/med.21255). PMID [22430881](https://pubmed.ncbi.nlm.nih.gov/22430881/).
66. ^{a, b, c, d}Polykovskiy D, Zhebrak A, Sanchez-Lengeling B, Golovanov S, Tatanov O, Belyaev S, Kurbanov R, Artamonov A, Aladinskiy V, Veselov M, Kadurin A, Johansson S, Chen H, Nikolenko S, Aspuru-Guzik A, Zhavoronkov A. "Molecular Sets (MOSES): A Benchmarking Platform for Molecular Generation Models". *arXiv*. October 2020. Available from: <http://arxiv.org/abs/1811.12823>. doi:[10.48550/arXiv.1811.12823](https://doi.org/10.48550/arXiv.1811.12823).
67. ^{a, b, c, d, e}Guan J, Zhou X, Yang Y, Bao Y, Peng J, Ma J, Liu Q, Wang L, Gu Q (2024). "DecompDiff: Diffusion Models with Decomposed Priors for Structure-Based Drug Design". *arXiv*. Version 1. doi:[10.48550/ARXIV.2403.07902](https://doi.org/10.48550/ARXIV.2403.07902). Available from: <https://arxiv.org/abs/2403.07902>.
68. ^aFrancoeur PG, Masuda T, Sunseri J, Jia A, Iovanisci RB, Snyder I, Koes DR (2020). "Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design". *Journal of Chemical Information and Modeling*. 60 (9): 4200–4215.
69. ^aHu L, Benson ML, Smith RD, Lerner MG, Carlson HA (2005). "Binding MOAD (mother of all databases)". *Proteins: Structure, Function, and Bioinformatics*. 60 (3): 333–340. Publisher: Wiley Online Library.
70. ^aHuang Z, Yang L, Zhou X, Zhang Z, Zhang W, Zheng X, Chen J, Wang Y, CUI B, Yang W. "Protein-ligand interaction prior for binding-aware 3d molecule diffusion models". *The Twelfth International Conference on Learning Representations 2024*.
71. ^{a, b, c, d, e}Lugmayr A, Danelljan M, Romero A, Yu F, Timofte R, Van Gool L. "Repaint: Inpainting using denoising diffusion probabilistic models." In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022:11461–11471.
72. ^{a, b}Alhossary A, Handoko SD, Mu Y, Kwok CK (2015). "Fast, accurate, and reliable molecular docking with QuickVina 2". *Bioinformatics (Oxford, England)*. 31 (13): 2214–2216.
73. ^{a, b, c, d, e, f}Landrum G, others (2013). "RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling". *Greg Landrum*. 8: 31.
74. ^{a, b, c}Irwin JJ, Tang KG, Young J, Dandarchuluun C, Wong BR, Khurelbaatar M, Moroz YS, Mayfield J, Sayle RA (2020). "ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery". *Journal of Chemical Information and Modeling*. doi:[10.1021/acs.jcim.0c00675](https://doi.org/10.1021/acs.jcim.0c00675).
75. ^{a, b, c}Dossetter AG, Griffen EJ, Leach AG (2013). "Matched molecular pair analysis in drug discovery". *Drug Discovery Today*. 18 (15–16): 724–731.
76. ^aJaini P, Holdijk L, Welling M (2021). "Learning equivariant energy based models with equivariant stein variational gradient descent". *Advances in Neural Information Processing Systems*. 34: 16727–16737.
77. ^aWójcikowski M, Zielenkiewicz P, Siedlecki P (2015). "Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field". *Journal of cheminformatics*. 7 (1): 1–6. Publisher: BioMed Central.
78. ^aVainio MJ, Puranen JS, Johnson MS (2009). "ShaEP: molecular overlay based on shape and electrostatic potential". *ACS Publications*.

79. ^ASverrisson F, Feydy J, Correia BE, Bronstein MM. Fast end-to-end learning on protein surfaces. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021 Jun; p. 15267–15276. doi:[10.1109/CVPR46437.2021.01502](https://doi.org/10.1109/CVPR46437.2021.01502). <https://ieeexplore.ieee.org/document/9577686>.
80. ^AWójcikowski M, Zielenkiewicz P, Siedlecki P (2015). "Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field". *Journal of Cheminformatics*. 7 (1): 26. doi:[10.1186/s13321-015-0078-2](https://doi.org/10.1186/s13321-015-0078-2). <https://jcheminf.biomedcentral.com/articles/10.1186/s13321-015-0078-2>.
81. ^AGanea OE, Huang X, Bunne C, Bian Y, Barzilay R, Jaakkola T, Krause A (2021). "Independent se (3)-equivariant models for end-to-end rigid protein docking". *arXiv preprint arXiv:2111.07786*. [arXiv:2111.07786](https://arxiv.org/abs/2111.07786).
82. ^A, ^BZaucha J, Softley CA, Sattler M, Frishman D, Popowicz GM. Deep learning model predicts water interaction sites on the surface of proteins using limited-resolution data. *Chemical Communications*. 56(98):15454–15457, 2020. doi:[10.1039/D0CC04383D](https://doi.org/10.1039/D0CC04383D). <https://pubs.rsc.org/en/content/articlelanding/2020/cc/d0cc04383d>.
83. ^AZhu H, Zhou R, Cao D, Tang J, Li M. A pharmacophore-guided deep learning approach for bioactive molecular generation. *Nature Communications*. 14 (1): 6234, October 2023. doi:[10.1038/s41467-023-41454-9](https://doi.org/10.1038/s41467-023-41454-9). <https://www.nature.com/articles/s41467-023-41454-9>.
84. ^ALorensen WE, Cline HE. "Marching cubes: A high resolution 3D surface construction algorithm." In: *Proceedings of the 14th annual conference on Computer graphics and interactive techniques, SIGGRAPH '87*. New York, NY, USA: Association for Computing Machinery; 1987. p. 163–169. doi:[10.1145/37401.37422](https://doi.org/10.1145/37401.37422). Available from: <https://dl.acm.org/doi/10.1145/37401.37422>.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.