Pharmacophore-guided de novo drug design with diffusion bridge

Conghao Wang [1] **Jagath C. Rajapakse** [1]

Abstract

De novo design of bioactive drug molecules with potential to treat desired biological targets is a profound task in the drug discovery process. Existing approaches tend to leverage the pocket structure of the target protein to condition the molecule generation. However, even the pocket area of the target protein may contain redundant information since not all atoms in the pocket is responsible for the interaction with the ligand. In this work, we propose PharmacoBridge, a phamacophore-guided *de novo* design approach to generate drug candidates inducing desired bioactivity via diffusion bridge. Our method adapts the diffusion bridge to effectively convert pharmacophore arrangements in the spatial space into molecular structures under the manner of $SE(3)$ -equivariant transformation, providing sophisticated control over optimal biochemical feature arrangements on the generated molecules. PharmacoBridge is demonstrated to generate hit candidates that exhibit high binding affinity with potential protein targets.

1. Introduction

Computer-aided drug design (CADD) plays a crucial role in the modern drug discovery procedure. Conventional CADD approaches, such as virtual screening, require searching for candidates with optimal molecular properties in a vast chemistry library. Although accelerated by the high-throughput technology, this process is still time-consuming and costly (Wouters et al., 2020) since the relationship between chemical structures and the molecular property of interest is obscure. *De novo* design, on the other hand, models the chemical space of molecular structures and properties and seeks for the optimal candidates in a directed manner (Meyers et al., 2021) instead of enumerating every possibility, thus facilitating the drug discovery process. Moreover, the flourishing of deep generative models in various domains, such as large language models and image synthesis, has endowed us an opportunity of applying deep learning to improving *de novo* drug design algorithms.

Generative models including variational autoencoder (VAE) (Kingma & Welling, 2013), generative adversarial networks (GAN) (Goodfellow et al., 2020) and denoising diffusion probabilistic models (DDPM) (Ho et al., 2020), have been successfully adapted for molecular design. Initially, researchers tend to represent drugs with linear notations, such as Simplified Molecular-Input Line-Entry System (SMILES) (Weininger, 1988), due to its simplicity. Then long-short term memory (LSTM) networks were readily applied to encoding the SMILES notations, and VAE and GAN algorithms were adapted for generation (Prykhodko et al., 2019; Méndez-Lucio et al., 2020; Born et al., 2021; Kaitoh & Yamanishi, 2021). Such methods, however, suffered from low chemistry validity rates of generated molecules since the structural information is neglected in SMILES notations.

To alleviate such limitations, a wide range of molecular graph-based generative algorithms has emerged. For example, graph VAE (Simonovsky & Komodakis, 2018) paved the way for probabilistic graph construction approaches, which can be used as the generator of other generative models such as GANs (De Cao & Kipf, 2018). However, due to computational complexity, such methods were only able to generate a limited number of nodes. A couple of research managed to address this issue by proposing auto-regressive graph construction algorithms. Flow-based methods such as MoFlow (Zang & Wang, 2020) and GraphAF (Shi et al., 2020) modeled the generation of bonds and atoms in a sequential decision flow and generated them in order. MoLeR (Maziarz et al., 2021) and MGSSL (Zhang et al., 2021) introduced motif-based generation that considers functional groups rather than single atoms as nodes, thus expanding the scale of generated molecules. Auto-regressive generation also improved the validity rate of generated molecules since the valence check can be performed at every step of generation. However, such methods are unnatural since there is no such decision sequence in a molecule, where posterior nodes depend on former generated ones. In addition, if one step in the middle of the generation is predicted improperly, the whole subsequent generation will be affected.

The advent of diffusion models (Song et al., 2020) enables molecular graph generation in one-shot. Although diffusion models initially obtained tremendous success in text and image generation (Rombach et al., 2022), a wealth of research has demonstrated that they can be adapted for graph generation as well since they can be developed to learn the distribu-

tions of the adjacency matrix and the feature matrix, which fully define the graph (Vignac et al., 2022; Jo et al., 2022; Lin et al., 2024). In the meantime, researchers realized the power of diffusion models is not restricted on 2D graph generation and can be evolved into 3D point cloud generation, which allows for more sophisticated design of molecules. In 2D graph generation, subsequent tests over all 3D conformers have to be conducted to filter the candidate since different isomers of a molecule may exhibit various pharmaceutical properties, which may lead to additional costs and future attrition. Aiming at achieving direct 3D spatial design of molecules, a series of studies were proposed (Xu et al., 2022; 2023; Schneuing et al., 2022; Hoogeboom et al., 2022; Guan et al., 2023). Equivariant geometric learning were leveraged to ensure the roto-translational equivariance of their systems while generating coordinates of the nodes.

Nevertheless, latest research indicates that the typical diffusion models, such as score matching with Langevin dynamics (SMLD) (Song & Ermon, 2019) and DDPM (Ho et al., 2020), can be generalized as *diffusion bridges* that are guaranteed to reach an endpoint from the desirable domain in pre-fixed time (Liu et al., 2022; Ye et al., 2022) according to Doob's *h*-transform (Doob & Doob, 1984). Moreover, it is demonstrated that diffusion bridges not only can be applied to unconditional generation by mapping data distribution to the prior noise distribution but also can be leveraged to align two arbitrary distributions by fixing both the start point and the end point (Chen et al., 2021; Vargas et al., 2021; Wang et al., 2021). Diffusion bridges have been successfully employed for image translation (Su et al., 2022; Zhou et al., 2023), physics-informed molecule generation (Wu et al., 2022; Jo et al., 2023), molecular docking (Somnath et al., 2023) and equilibrium state prediction (Luo et al., 2024). We believe that they are promising approaches to hit molecule design as well since the start point can be naturally considered as the molecular structures and the end point as desired conditions, e.g., pharmacophore arrangements.

In the field of drug design, increasing research has delved into adapting generative models to the design of hit candidates with potential to react with particular biological targets. For example, gene expression profiles were used to condition the generation so that the generated molecules would lead to desirable biological activities (Méndez-Lucio et al., 2020; Born et al., 2021; Kaitoh & Yamanishi, 2021; Wang et al., 2024). However, gene expression profiles may contain redundant information, since not all genes are related to the drug reaction. To enable more sophisticated control over the generation procedure, researchers proceeded to consider the target protein structure as the constraint. Pocket2Mol (Peng et al., 2022) first proposed an auto-regressive model to generate atoms and bonds gradually under the guidance of the protein pocket structure. Further research improved pocket-based generation by one-shot generation using diffusion models, such as TargetDiff (Guan et al., 2023). D3FG (Lin et al., 2024) extended atom-based generation to functional groups, aiming at preserving better chemical consistency.

Recent studies proposed to generate hit candidates guided by pharmacophores, which hypothesizes a spatial arrangement of chemical features that are essential for the binding between the drug molecule and the target protein. Unlike conditioning by the protein pocket structure, which depends on the model to learn to discover the indispensable features for binding implicitly, pharmacophore arrangements explicitly define such features, which renders them suitable constraints for hit identification. Nonetheless, existing research studying pharmacophore-guided drug discovery, e.g., PharmacoNet (Seo & Kim, 2023), only leveraged pharmacophore modelling to compute matching scores between ligand and protein, and thereby accelerated virtual screening. Although PGMG (Zhu et al., 2023) introduced a *de novo* pharmacophore-guided drug design framework, it only learned a latent variable of the pharmacophore features and generated SMILES representations accordingly. This actually destructs the 3D features given by the pharmacophore hypotheses and omits their connection with the 3D molecular structures. In this work, we propose a direct pharmacophore-guided *de novo* drug design approach that learns to transform the pharmacophore distribution into the molecular structure distribution with an $SE(3)$-equivariant diffusion bridge.

2. Preliminaries

2.1. Notations and Problem Formulation

Drug molecules and their pharmacophore features are represented as point clouds in the 3D Euclidean space, i.e., $\mathbf{G} = (\mathbf{x}, \mathbf{h})$, where $\mathbf{x} \in \mathbb{R}^{N\times 3}$ and $\mathbf{h} \in \mathbb{R}^{N\times M}$ denote atom coordinates and features, respectively. As shown in Figure 1, our mission is to learn the evolution of the atom point cloud over time $\{\mathbf{G}_t\}_{t=0}^T$ via a bridge process $q(\mathbf{G}_t|\mathbf{G}_T)$ that allows us to sample molecular structures from pharmacophore hypotheses.

Let the initial state $\mathbf{G}_0 \sim q_{data}(\mathbf{g})$ denote the molecular point cloud and the terminal state $\mathbf{G}_T \sim q_{data}(\mathbf{\Gamma})$ denote the point cloud of the corresponding pharmacophore features. Since molecules $\mathbf{g}$ and pharmacophores $\mathbf{\Gamma}$ are strictly paired data, rotations or translations to one of them will cause equivariant transformations to the other, i.e., $q_{data}(g\cdot\mathbf{g}, g\cdot\mathbf{\Gamma}) = q_{data}(\mathbf{g}, \mathbf{\Gamma})\ \forall g \in SE(3)$, which indicates an $SE(3)$-invariant joint distribution. Consequently, the evolution of the system should satisfy $q(g\cdot\mathbf{G}_0|g\cdot\mathbf{G}_T) = q(\mathbf{G}_0|\mathbf{G}_T)\ \forall g \in SE(3)$, indicating an $SE(3)$-equivariant transition density.

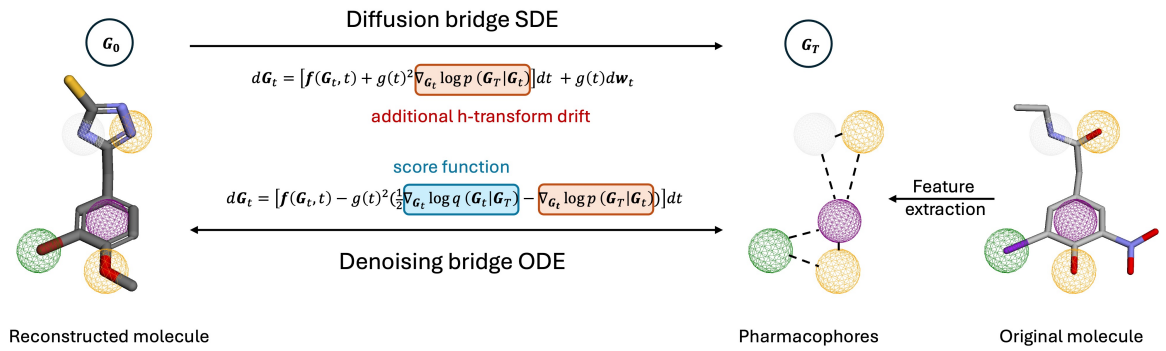

Figure 1. Overview of PharmacoBridge. The diffusion bridge process is devised to map the molecule data $\mathbf{G}_0$ to the extracted pharmacophore data $\mathbf{G}_T$ via Doob's h-transforms. Reversely, a score matching model is trained to estimate the score function, which composes the denoising bridge process that recovers molecule data from the pharmacophore data.

2.2. Diffusion Model with SDEs

A diffusion process is to inject noise into data $\mathbf{G}_0 \sim q_{data}$ and gradually convert data into noise $\mathbf{G}_T \sim p_{prior}$, where q_{data} and the p_{prior} denote the data and the prior distribution, respectively. Song et al. proposed to model this diffusion process as the solution to the following SDE (Song et al., 2020)

$$d\mathbf{G}_t = \mathbf{f}(\mathbf{G}_t, t)dt + g(t)d\mathbf{w}_t \tag{1}$$

where $\mathbf{f}(\cdot, t)$ is the drift function, $\mathbf{w_t}$ is the standard Wiener process and $g(\cdot)$ is the diffusion coefficient. In our case, we denote the representation of a molecule at time t as $\mathbf{G}_t$.

The denoising process is the solution to the reverse-time SDE

$$d\mathbf{G}_t = [\mathbf{f}(\mathbf{G}_t, t) - g(t)^2 \nabla_{\mathbf{G}_t} \log p(\mathbf{G}_t)]dt + g(t)d\mathbf{w}_t \tag{2}$$

where $p(\mathbf{G}_t)$ is the marginal distribution of $\mathbf{G}_t$ at time t. SMLD (Song & Ermon, 2019) and DDPM (Ho et al., 2020) can be regarded as such SDE frameworks with different designs of $\mathbf{f}(\cdot, t)$ and $g(\cdot)$.

2.3. Diffusion Bridge

Doob's h-transform (Doob & Doob, 1984) reveals that by adding an additional drift term, the diffusion process is guaranteed to reach a fixed end point from an arbitrary distribution, given by

$$d\mathbf{G}_t = [\mathbf{f}(\mathbf{G}_t, t) + g(t)^2 \nabla_{\mathbf{G}_t} \log p(\mathbf{G}_T|\mathbf{G}_t)]dt + g(t)d\mathbf{w}_t \tag{3}$$

where $p(\mathbf{G}_T|\mathbf{G}_t)$ is the transition density function of the original diffusion process solving (1). Starting from $\mathbf{G}_0 \sim q_{data}(\mathbf{g})$, the score function $\nabla_{\mathbf{G}_t} \log p(\mathbf{G}_T|\mathbf{G}_t)$ drives the process to reach the desired endpoint $\mathbf{G}_T = \mathbf{\Gamma} \sim q_{data}(\mathbf{\Gamma})$.

3. Pharmacophore to Drug Diffusion Bridge

We introduce the pharmacophore to drug diffusion bridge model (PharmacoBridge) to translate 3D arrangements of pharmacophore features into hit molecules. The proposed framework is presented in Figure 1.

3.1. Equivariant Denoising Diffusion Bridge

We present a graph translation paradigm that generates 3D point cloud guided by prior cluster arrangements. We demonstrate our approach on the task of 3D hit molecule design guided by the pharmacophore hypothesis. Based on the diffusion bridge characterized by Eq. (3), the time-reversed SDE of the denoising bridge is defined in the following theorem.

Theorem 3.1 (Denoising diffusion bridge)**.** *By reversing the h-transformed diffusion bridge with the law of (3), we obtain the following ODE to model the time evolution of the transition density $q(\mathbf{G}_t|\mathbf{G}_T)$:*

$$\begin{aligned} d\mathbf{G}_t = [\mathbf{f}(\mathbf{G}_t, t) - g^2(t)(\frac{1}{2}\nabla_{\mathbf{G}_t} \log q(\mathbf{G}_t|\mathbf{G}_T) \\ - \nabla_{\mathbf{G}_t} \log p(\mathbf{G}_T|\mathbf{G}_t))]dt \end{aligned} \tag{4}$$

which constructs the denoising bridge process $\{\mathbf{G}_t\}_{t=0}^T$ with marginal distribution $\mathbf{G}_T = \mathbf{\Gamma} \sim q_{data}(\mathbf{\Gamma})$, inducing the joint distribution $q(\mathbf{G}_0, \mathbf{G}_T)$ approximating $q_{data}(\mathbf{g}, \mathbf{\Gamma})$. $p(\mathbf{G}_T|\mathbf{G}_t)$ is the transition density associated with the original diffusion process specified by Eq. (1).

The proof of deriving the denoising bridge process specified in Theorem 3.1 is provided in Appendix A.1. Next, we aim to devise an $SE(3)$-equivariant diffusion bridge that satisfies symmetry constraints. As demonstrated in (Luo et al., 2024), given a diffusion process solving SDE (1), its associated transition density $p(\mathbf{G}_T|\mathbf{G}_t)$ is $SE(3)$-equivariant, i.e., $p(g \cdot \mathbf{G}_T|g \cdot \mathbf{G}_t) = p(\mathbf{G}_T|\mathbf{G}_t)$, $\forall g \in SE(3)$, if (1) the initial density $q(\mathbf{G}_0)$ is $SE(3)$-invariant; and (2) $\mathbf{f}(\cdot, t)$

is $SO(3)$-equivariant and $T(3)$-invariant; (3) the transition density of $\mathbf{w}_t$ is $SE(3)$-equivariant. Based on this equivariant diffusion process, we further devise the equivariant denoising diffusion bridge:

Theorem 3.2 (Equivariant denoising diffusion bridge)**.** *Given an $SE(3)$-equivariant diffusion process with transition density $p(\mathbf{G}_T|\mathbf{G}_t)$, its h-transformed diffusion bridge with the law of (3) and the corresponding denoising diffusion bridge characterized in Theorem 3.1 also preserve $SE(3)$-equivariance. Their transition densities satisfy*

- $p(g{\cdot}\mathbf{G}_t|g{\cdot}\mathbf{G}_0, g{\cdot}\mathbf{G}_T) = p(\mathbf{G}_t|\mathbf{G}_0, \mathbf{G}_T)$, $\forall g \in SE(3)$
- $q(g \cdot \mathbf{G}_t|g \cdot \mathbf{G}_T) = q(\mathbf{G}_t|\mathbf{G}_T)$, $\forall g \in SE(3)$

The constructed equivariant denoising diffusion bridge enables us to map molecule and pharmacophore distributions retaining symmetry constraints. Proofs of Theorem 3.2 are provided in Appendix A.2.

Denoising score matching The training objective is designed to estimate the score $\nabla_{\mathbf{G}_t} \log q(\mathbf{G}_t|\mathbf{G}_T)$ in order to solve the denoising bridge ODE (4)

$$\mathcal{L}_\theta = \mathbb{E}_{(\mathbf{G}_0,\mathbf{G}_T)\sim q_{data}(\mathbf{g},\mathbf{\Gamma}),\mathbf{G}_t\sim q(\mathbf{G}_t|\mathbf{G}_0,\mathbf{G}_T),t\sim\mathcal{U}(0,T)} \left[\lambda(t)\left\|s_\theta(\mathbf{G}_t,\mathbf{G}_T,t) - \nabla_{\mathbf{G}_t}\log q(\mathbf{G}_t|\mathbf{G}_0,\mathbf{G}_T)\right\|^2\right] \tag{5}$$

where the initial and final states are fixed at $\mathbf{g}$ and $\mathbf{\Gamma}$, respectively. $\lambda(t)$ is a time-dependent loss weight to intensify the penalty as the denoising bridge approaches the data point, improving the reconstruction at later stages. Details of the choice of this term are discussed in Appendix B.2.

The true score of the transition density $q(\mathbf{G}_t|\mathbf{G}_T) = \int_{\mathbf{G}_0} q(\mathbf{G}_t|\mathbf{G}_0,\mathbf{G}_T)q_{data}(\mathbf{G}_0|\mathbf{G}_T)d\mathbf{G}_0$ is intractable in general. We therefore train a neural network model $s_\theta(\mathbf{G}_t,\mathbf{G}_T,t)$ to approximate the score of the sampling distribution $\nabla_{\mathbf{G}_t}\log q(\mathbf{G}_t|\mathbf{G}_0,\mathbf{G}_T)$. Specific parameterization strategies are discussed in the next section. DDBM (Zhou et al., 2023) has shown that the optimal solution of (5) approximates the true score $s^*_\theta(\mathbf{G}_t,\mathbf{G}_T,t) = \nabla_{\mathbf{G}_t}\log q(\mathbf{G}_t|\mathbf{G}_T)$.

3.2. Score Model Parameterization

To parameterize the score model $s_\theta(\mathbf{G}_t,\mathbf{G}_T,t)$, we design the sampling distribution $q(\mathbf{G}_t|\mathbf{G}_0,\mathbf{G}_T)$ to be the same as the diffusion distribution fixed on both ends, which is a Gaussian distribution known in closed-form

$$\begin{aligned} q(\mathbf{G}_t|\mathbf{G}_0,\mathbf{G}_T) &:= p(\mathbf{G}_t|\mathbf{G}_0,\mathbf{G}_T) = \mathcal{N}(\hat{\mu}_t, \hat{\sigma}_t^2\mathbf{I}), \\ \text{where}\quad \hat{\mu}_t &= \frac{\text{SNR}_T}{\text{SNR}_t}\frac{\alpha_t}{\alpha_T}\mathbf{G}_T + \alpha_t\mathbf{G}_0(1-\frac{\text{SNR}_T}{\text{SNR}_t}), \\ \text{and}\quad \hat{\sigma}_t^2 &= \sigma_t^2(1-\frac{\text{SNR}_T}{\text{SNR}_t}) \end{aligned} \tag{6}$$

where α_t and σ_t are the signal and noise schedules, which vary with the choice of the bridge design. The signal-to-noise ratio is thus defined as $\text{SNR}_t = \alpha_t^2/\sigma_t^2$.

Since denoising bridge distribution is designed to be Gaussian, its score function can be readily derived by

$$\nabla_{\mathbf{G}_t}\log q(\mathbf{G}_t|\mathbf{G}_0,\mathbf{G}_T) = -\frac{\mathbf{G}_t - \hat{\mu}_t}{\hat{\sigma}_t^2} \tag{7}$$

We train $s_\theta(\mathbf{G}_t,\mathbf{G}_T,t)$ to estimate the score above, which can be reparameterized as

$$\begin{aligned} s_\theta(\mathbf{G}_t,\mathbf{G}_T,t) &= -\frac{\mathbf{G}_t - \mu_\theta}{\hat{\sigma}_t^2}, \\ \text{where}\quad \mu_\theta &= (\frac{\text{SNR}_T}{\text{SNR}_t}\frac{\alpha_t}{\alpha_T}\mathbf{G}_T + \alpha_t D_\theta(\mathbf{G}_t,\mathbf{G}_T,t)(1-\frac{\text{SNR}_T}{\text{SNR}_t})) \end{aligned} \tag{8}$$

where the denoiser $D_\theta(\mathbf{G}_t,\mathbf{G}_T,t)$ aims at predicting $\mathbf{G}_0$. Following EDM (Karras et al., 2022), D_θ can be further reparameterized with

$$D_\theta(\mathbf{G}_t,\mathbf{G}_T,t) = c_{skip}(t)\mathbf{G}_t + c_{out}(t)F_\theta(c_{in}(t)\mathbf{G}_t, c_{noise}(t)) \tag{9}$$

where $F_\theta(\cdot,t)$ is the neural network that predicts something between $\mathbf{G}_0$ and the noise at time step t. In typical diffusion models (Song et al., 2020; Song & Ermon, 2019; Ho et al., 2020), $F_\theta(\cdot,t)$ aims at directly predicting the noise at time step t, but this is demanding when t approaches the final step T, since the data is almost full of noise. EDM proposed to adapt F_θ to predict the mixture of signal and noise to alleviate this drawback (Karras et al., 2022). Specific parameterizations for VP- and VE-based diffusion bridges are illustrated in Appendix B.1. The choice of the scaling parameters c_{in}, c_{out}, c_{skip} and c_{noise} is discussed in Appendix B.2.

To preserve $SE(3)$-equivariance, we implement $F_\theta(\cdot,t)$ with EGNN (Satorras et al., 2021). The behavior of F_θ at step t is formulated as

$$\begin{aligned} (\tilde{\mathbf{x}},\tilde{\mathbf{h}}) &= F_\theta(c_{in}(t)\mathbf{G}_t, c_{noise}(t)) \\ &= F_\theta(c_{in}(t)(\mathbf{x}_t,\mathbf{h}_t), c_{noise}(t)) \end{aligned} \tag{10}$$

where $\mathbf{x}$ and $\mathbf{h}$ denote the atom positions and features, respectively. The output $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{h}}$ represent the mixture of noise and signal. The reconstructed initial data is calculated via $\hat{\mathbf{x}}_0 = c_{skip}(t)\mathbf{x}_t + c_{out}(t)\tilde{\mathbf{x}}$ and $\hat{\mathbf{h}}_0 = c_{skip}(t)\mathbf{h}_t + c_{out}(t)\tilde{\mathbf{h}}$ according to Eq. (9). Implementation details of F_θ are provided in Appendix B.4.

Training objective Practically, the loss function in Eq. (5) is equivalent to

$$\mathcal{L}_\theta = \mathbb{E}_{\mathbf{x}_t,\mathbf{x}_0,\mathbf{h}_t,\mathbf{h}_0,t}(\|\hat{\mathbf{x}}_0 - \mathbf{x}_0\|^2 + \omega_h\|\hat{\mathbf{h}}_0 - \mathbf{h}_0\|^2) \tag{11}$$

which is the combination of the MSE losses with regard to $\mathbf{x}$ and $\mathbf{h}$. The weight of the feature loss ω_h is empirically configured as 10 to keep the training stable.

3.3. Sampling with PharmacoBridge

Alg. 1 illustrates the deterministic sampling procedure of our model. The output of the score matching model $\mathbf{s}_\theta$ is computed via Eq. (8) and (9). The h-transform drift $\nabla_{\mathbf{G}_i} \log p(\mathbf{G}_N|\mathbf{G}_i)$ is derived by calculating the score of the transition probability from t_i to t_N. This computation is tractable when using the VE and VP bridge designs, since both of them employs Gaussian transition kernels. Our time discretization strategy is discussed in Appendix B.3.

4. Experiments

4.1. Experimental Setup

Data CrossDocked2020 V1.3 dataset (Francoeur et al., 2020) is used for model development and evaluation. The dataset is filtered by preserving the ligands with intimate protein binding poses (RMSE < 0.1 Å), which results in around 242 thousand ligands. CrossDocked2020 data has organized the ligand and protein files according to pockets. We thus split the dataset in the same way, ensuring molecules with similar structures or biological targets occur either in the training or the sampling dataset. 80% of the pockets are used for training, 10% for validation and 10% for testing, resulting in around 197 thousand molecules in the training set, 19 thousand molecules in the validation set and 15 thousand molecules in the testing set. Details of data preprocessing are illustrated in B.5.

Baselines We benchmark proposed PharmacoBridge against various state-of-the-art 3D molecule design approaches. In the *unconditional generation* task, we compare our method with EDM (Hoogeboom et al., 2022) and GruM (Jo et al., 2023). EDM employed the typical DDPM scheme with an $E(3)$-equivariant GNN backbone for 3D molecule generation. GruM devised its generative process as a destination-predicting diffusion mixture, but never explored conditional generation by fixing the initial point.

In the *pharmacophore-guided generation* task, we compare our method with Pocket2Mol (Peng et al., 2022) and TargetDiff (Guan et al., 2023). Pocket2Mol leveraged autoregressive generation for pocket-guided molecule design. TargetDiff adapted DDPM by incorporating pocket information into the denoising backbone with the attention mechanism. Candidates generated by both methods exhibited potential binding affinity with the target protein pockets. Another recent pharmacophore-guided molecule generation approach, PGMG (Zhu et al., 2023), is not considered since PGMG generated SMILES rather than 3D structures, which is fundamentally different from our approach.

4.2. Unconditional Generation

We tested our model on the unconditional generation task to evaluate several chemical and physical properties of the generated molecules. Specifically, 10,000 molecules were sampled using our model or each baseline. Validity, novelty, uniqueness, synthetic accessibity (SA) and quantitative estimate of drug-likeness (QED) were assessed on the generated molecules.

Table 1. Unconditional generation performance of PharmacoBridge and baselines.

			Validity (%)	Uniqueness (%)	Novelty (%)
EDM			**100.00**	6.89	100.00
GruM			**100.00**	82.60	100.00
PharmacoBridge	VE	Basic	96.63	85.37	100.00
		Aromatic	88.27	90.90	100.00
	VP	Basic	99.91	90.72	99.98
		Aromatic	99.96	**91.94**	**100.00**

As shown in Table 1, EDM and GruM exhibited 100% validity and novelty, but tended to produce repeated samples, resulting in relatively low uniqueness. In the ablation study to explore the effects of different bridge types and feature types, we found that VP bridge-based models outperformed VE-based ones, especially in terms of validity. VP-based PharmacoBridge with aromatic features achieved the highest uniqueness and nearly 100% validity and novelty.

Figure 2 and 3 present the distribution of SA and QED scores of generated molecules. Note that lower SA scores indicate better synthetic accessibility, and higher QED scores indicate better drug-likeness. Our model generated molecules with evenly distributed SA scores, whereas the SA scores of EDM and GruM generated molecules fell between 6.0 and 8.0, which is distant from the data distribution. In the QED assessment, GruM obtained relatively lower scores while our model and EDM generated molecules with better drug-likeness. Additional ablation study results on SA and QED evaluations are illustrated in Appendix C.1.

In conclusion, PharmacoBridge achieved comparable validity, novelty and QED scores against state-of-the-art baselines and outperformed the baselines in terms of uniqueness and SA scores. In the ablation study, we found VE bridge-based PharmacoBridge was inclined to produce invalid chemical structures, whereas VP bridge-based PharmacoBridge exhibited better performance. This is aligned with the findings observed in previous works comparing VP and VE-based diffusion models (Zhou et al., 2023; Song et al., 2020). Moreover, incorporating aromatic features brought slight improvements in uniqueness and novelty. Therefore, we continue the subsequent experiments with VP-based PharmacoBridge with aromatic features.

Algorithm 1 Deterministic sampler of PharmacoBridge

Input: sampled time steps $t_{i\in N,\ldots,0}$, pharmacophore data $\mathbf{G}_N \sim p_{pp}$, score matching model $\mathbf{s}_\theta$
for $i = N : 1$ **do**
  $\mathbf{d}_i \leftarrow -\mathbf{f}(\mathbf{G}_i, t_i) + g^2(t_i)(\frac{1}{2}\mathbf{s}_\theta(\mathbf{G}_i, \mathbf{G}_N, T) - \nabla_{\mathbf{G}_i} \log p(\mathbf{G}_N|\mathbf{G}_i)))$ ▷ Solving Eq. 4
  $\mathbf{G}_{i-1} \leftarrow \mathbf{G}_i + (t_{i-1} - t_i)\mathbf{d}_i$ ▷ Euler step
  if $i \neq 1$ **then**
    $\mathbf{d}'_i \leftarrow -\mathbf{f}(\mathbf{G}_{i-1}, t_{i-1}) + g^2(t_{i-1})(\frac{1}{2}\mathbf{s}_\theta(\mathbf{G}_{i-1}, \mathbf{G}_N, T) - \nabla_{\mathbf{G}_{i-1}} \log p(\mathbf{G}_N|\mathbf{G}_{i-1})))$
    $\mathbf{G}_{i-1} \leftarrow \mathbf{G}_i + (t_{i+1} - t_i)(\frac{1}{2}\mathbf{d}_i + \frac{1}{2}\mathbf{d}'_i)$ ▷ Heun's 2nd order correction
  end if
end for
return $\mathbf{G}_0$

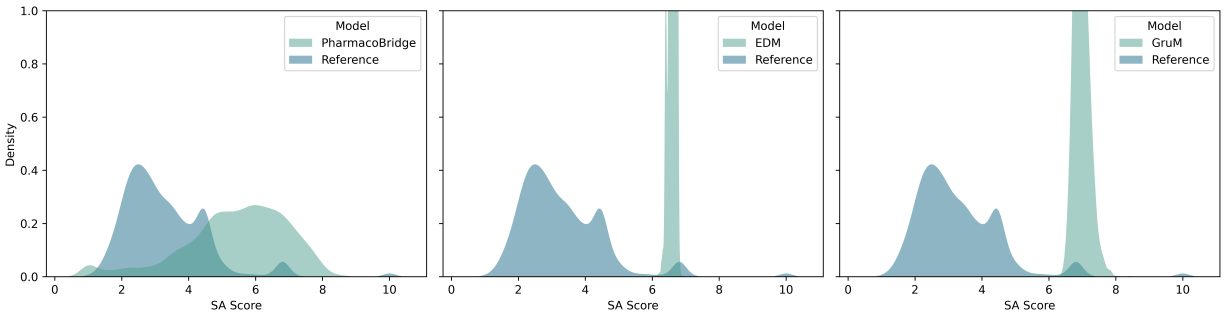

Figure 2. SA score (↓) distribution. SA scores of the molecules generated by EDM and GruM are concentrated between 6.0 and 8.0, which is obviously larger than the distribution of the original dataset. The SA scores achieved by our model are lower than the baselines and more evenly distributed like the original dataset distribution.

4.3. Pharmacophore-guided Hit Molecule Design

We further demonstrate PharmacoBridge in pharmacophore-guided drug design tasks, including ligand-based drug design and structure-based drug design (SBDD).

4.3.1. Ligand-based Drug Design

Ligand-based drug design relies on the knowledge of active molecules whose target structures remain obscure. By analyzing the chemical properties and structures of these ligands, we can create pharmacophore hypotheses to identify or design new compounds with similar or improved interactions. In order to assess our model in the ligand-based drug design task, we sample new molecules using the pharmacophore features extracted from known active ligands. The pocket structure of each ligand's receptor protein is utilized as the constraint for sampling from Pocket2Mol and TargetDiff.

We extracted all pharmacophores of around 15 thousand original ligands in the testing set, and removed the smaller groups that overlapped with the larger ones. New samples were generated by our model constrained by the extracted pharmacophores. We then computed the pharmacophore matching score reflecting the recovery rate of desired pharmacophore arrangements. Specifically, we calculated the distance between pharmacophores of generated molecules and original ligands. If the type of a pair of pharmacophores coincides and their distance is less than 1.5 Å, we consider this as a match. The matching score of each ligand is calculated by N_{match}/N_{total}, where N_{match} and N_{total} represent the number of matches and the number of existing pharmacophores, respectively.

Figure 4 displays the histograms of pharmacophore matching scores. TargetDiff, which designs drugs constrained by pocket structures, were used as our baseline. Notably, only around two thousand molecules generated by TargetDiff were valid. Pocket2Mol was excluded from this assessment due to its limitation to autoregressive generation, which significantly prolongs the time required to generate approximately 15,000 samples without support for batch processing. In addition, we included molecules generated by the unconditional version of our model as one of the baselines for comparison. As shown in Figure 4, PharmacoBridge significantly enhanced the pharmacophore recovery compared to unconstrained generation and TargetDiff. This suggests that the generated molecules have great potential to interact with the target protein, engaging not only within the desired pocket but also with the particular residue.

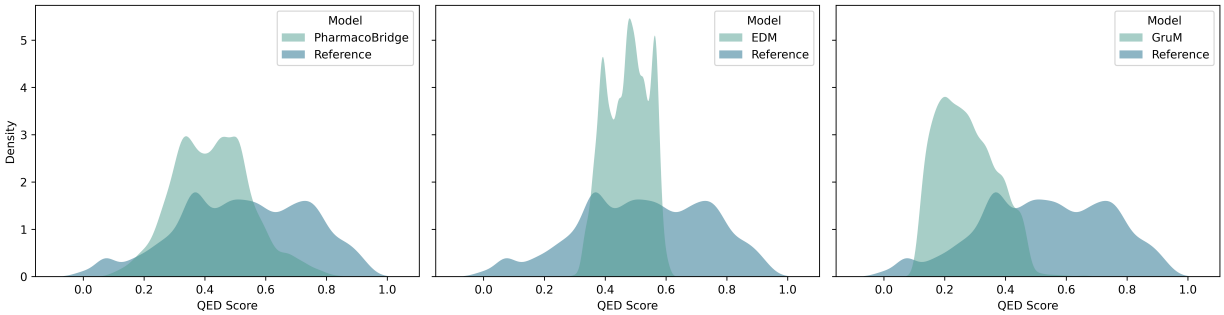

Figure 3. QED score (↑) distribution. Molecules generated by our model and EDM exhibit higher QED scores than GruM. The distribution achieved by our model shows similar pattern to the dataset distribution.

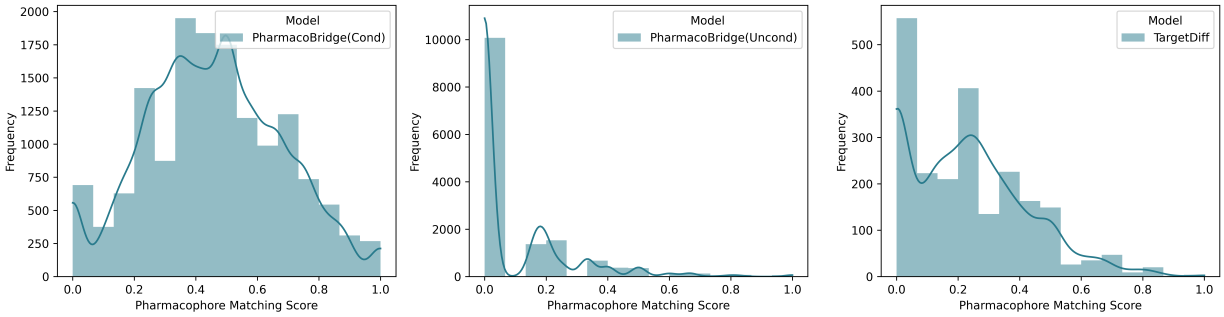

Figure 4. Pharmacophore matching sore distribution. Compared with Unconstrained generation and TargetDiff, pharmacophore-guided generation significantly enhanced the matching of pharmacophores extracted from original ligands and generated molecules. Notably, TargetDiff suffered from low validity, which resulting in less available molecules.

Table 2. Average pharmacophore matching scores of SBDD.

PDB ID	PharmacoBridge	Pocket2Mol	TargetDiff
1EOC	0.76	0.34	0.17
5LSA	0.71	0.46	0.25
4H2Z	0.91	0.49	0.38
1J06	1.00	0.08	0.11
3OCC	0.74	0.22	0.24
5UEV	0.84	0.39	0.17
5FE6	1.00	0.45	0.23
5LPJ	0.83	0.41	0.17
5IJS	0.97	0.31	0.14
4TTI	0.98	0.37	0.20

4.3.2. Structure-based Drug Design

SBDD, in contrast to ligand-based drug design, utilizes the 3D structure of the target, typically a protein, to guide drug development. This method allows researchers to design molecules that fit precisely within the binding pocket of the target, optimizing interactions at them atomic level for enhanced efficacy and specificity.

We selected 10 protein targets, manually identified the essential pharmacophore features for the ligands to bind with corresponding protein targets, and generated 100 hit candidates guided by each identified pharmacophore. Firstly, we evaluated the matching scores by comparing the pharmacophore features used as constraints and the ones extracted from the generated molecules. The average matching scores of PharmacoBridge and the baseline models are presented in Table 2. Our method achieved the highest average matching score in all cases.

Furthermore, we performed molecular docking analysis to demonstrate if the generated molecules are potential to interact with the protein targets. Gnina, a molecular docking software based on AutoDock Vina, was employed to test the binding affinity. Figure 5 shows the distribution of the Vina scores in box plots, with lower values representing higher binding affinities. The red dashed line in each group indicates the reference docking score of the original ligand provided by the CrossDocked dataset. It is evident that our method consistently generated molecules with higher binding affinities than the original ligand across each group. Pocket2Mol sometimes outperformed our method, but the generated molecules exhibited a significantly wide range of Vina scores, especially in cases of 1EOC and 4H2Z, rendering this approach very unstable. TargetDiff achieved the

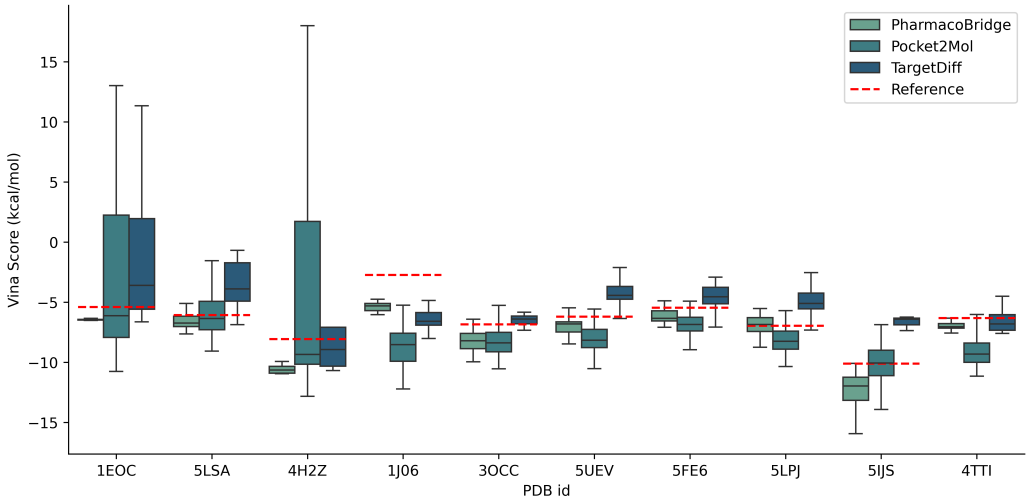

Figure 5. Distribution of Vina scores, with lower Vina score representing higher binding affinity. The reference scores given by the original ligands are indicated by red dashed lines in the figure. PharmacoBridge consistently generated molecules with higher binding affinities. Although Pocket2Mol sometimes outperformed, it produced molecules with a notably broader range of Vina scores in certain groups, e.g., 1EOC and 4H2Z. TargetDiff failed to generate molecules with binding affinities surpassing those of the original ligands in most cases.

Table 3. Percentage of generated molecules with higher binding affinity.

PDB ID	PharmacoBridge		Pocket2Mol		TargetDiff		Reference score (kcal/mol)
	High affinity ratio (%)	Sample num	High affinity ratio (%)	Sample num	High affinity ratio (%)	Sample num	
1EOC	99.00	100	55.34	100	30.00	10	-5.39
5LSA	76.00	100	52.78	100	3.70	28	-6.06
4H2Z	100.00	100	57.69	100	55.56	9	-8.06
1J06	100.00	100	98.15	100	100.00	16	-2.73
3OCC	93.00	100	91.51	100	25.00	8	-6.84
5UEV	91.00	100	94.87	100	8.89	45	-6.20
5FE6	80.00	100	90.99	100	18.75	32	-5.46
5LPJ	48.00	100	83.02	100	10.00	20	-6.95
5IJS	98.99	99	48.54	100	0.00	3	-10.11
4TTI	99.00	100	91.18	100	66.57	12	-6.32

worst performance with most of the generated molecules gaining higher Vina scores than the reference ligand.

In addition, the ratio of the generated molecules with higher binding affinity is calculated and shown in Table 3, together with the number of available samples and the reference Vina score. We observe that PharmacoBridge is able to produce hit candidates with over 90% higher binding affinities in most cases. Pocket2Mol is competitive in certain groups such as 1J06, 3OCC, 5UEV, 5FE6 and 5IJS, but produced only around 50% of higher affinity samples in the other groups. TargetDiff, however, is limited by a low validity rate. We present example molecules generated by PharmacoBridge and baselines using the pharmacophores identified in the binding complex structures of 1EOC, 3OCC and 5UEV in Figure 9 in Appendix C.2.

5. Conclusion

In this work, we presented PharmacoBridge - a diffusion bridge model designed for graph translation. Bridge processes enable us to construct a transitional probability flow between any two distributions and the idea of diffusion model makes such transition kernel tractable. We proposed a practical solution to leverage such models for graph generation constrained by the cluster arrangements of nodes, and demonstrated our model on the task of pharmacophore-guided *de novo* hit molecule design. We parameterized our denoiser with $SE(3)$-equivariant GNN. Constrained by the pharmacophore models extracted from the active ligands, our model managed to generate potential hit candidates with higher binding affinity with the receptor of the original ligands.

References

Born, J., Manica, M., Oskooei, A., Cadow, J., Markert, G., and Martínez, M. R. Paccmannrl: De novo generation of hit-like anticancer molecules from transcriptomic data via reinforcement learning. *Iscience*, 24(4), 2021.

Chen, T., Liu, G.-H., and Theodorou, E. A. Likelihood training of schr\" odinger bridge using forward-backward sdes theory. *arXiv preprint arXiv:2110.11291*, 2021.

De Cao, N. and Kipf, T. Molgan: An implicit generative model for small molecular graphs. *arXiv preprint arXiv:1805.11973*, 2018.

Doob, J. L. and Doob, J. *Classical potential theory and its probabilistic counterpart*, volume 262. Springer, 1984.

Francoeur, P. G., Masuda, T., Sunseri, J., Jia, A., Iovanisci, R. B., Snyder, I., and Koes, D. R. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *Journal of chemical information and modeling*, 60(9):4200–4215, 2020.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

Guan, J., Qian, W. W., Peng, X., Su, Y., Peng, J., and Ma, J. 3d equivariant diffusion for target-aware molecule generation and affinity prediction. *arXiv preprint arXiv:2303.03543*, 2023.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pp. 8867–8887. PMLR, 2022.

Jo, J., Lee, S., and Hwang, S. J. Score-based generative modeling of graphs via the system of stochastic differential equations. In *International Conference on Machine Learning*, pp. 10362–10383. PMLR, 2022.

Jo, J., Kim, D., and Hwang, S. J. Graph generation with diffusion mixture. *arXiv preprint arXiv:2302.03596*, 2023.

Kaitoh, K. and Yamanishi, Y. Triomphe: Transcriptome-based inference and generation of molecules with desired phenotypes by machine learning. *Journal of Chemical Information and Modeling*, 61(9):4303–4320, 2021.

Karras, T., Aittala, M., Aila, T., and Laine, S. Elucidating the design space of diffusion-based generative models. *Advances in Neural Information Processing Systems*, 35: 26565–26577, 2022.

Kingma, D. P. and Welling, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Lin, H., Huang, Y., Zhang, O., Liu, Y., Wu, L., Li, S., Chen, Z., and Li, S. Z. Functional-group-based diffusion for pocket-specific molecule generation and elaboration. *Advances in Neural Information Processing Systems*, 36, 2024.

Liu, X., Wu, L., Ye, M., and Liu, Q. Let us build bridges: Understanding and extending diffusion generative models. *arXiv preprint arXiv:2208.14699*, 2022.

Luo, S., Xu, Y., He, D., Zheng, S., Liu, T.-Y., and Wang, L. Bridging geometric states via geometric diffusion bridge. *arXiv preprint arXiv:2410.24220*, 2024.

Maziarz, K., Jackson-Flux, H., Cameron, P., Sirockin, F., Schneider, N., Stiefl, N., Segler, M., and Brockschmidt, M. Learning to extend molecular scaffolds with structural motifs. *arXiv preprint arXiv:2103.03864*, 2021.

Méndez-Lucio, O., Baillif, B., Clevert, D.-A., Rouquié, D., and Wichard, J. De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. *Nature communications*, 11(1):10, 2020.

Meyers, J., Fabian, B., and Brown, N. De novo molecular design and generative models. *Drug discovery today*, 26 (11):2707–2715, 2021.

Peluchetti, S. Non-denoising forward-time diffusions. *arXiv preprint arXiv:2312.14589*, 2023.

Peng, X., Luo, S., Guan, J., Xie, Q., Peng, J., and Ma, J. Pocket2mol: Efficient molecular sampling based on 3d protein pockets. In *International Conference on Machine Learning*, pp. 17644–17655. PMLR, 2022.

Prykhodko, O., Johansson, S. V., Kotsias, P.-C., Arús-Pous, J., Bjerrum, E. J., Engkvist, O., and Chen, H. A de novo molecular generation method using latent vector based generative adversarial network. *Journal of Cheminformatics*, 11:1–13, 2019.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Satorras, V. G., Hoogeboom, E., and Welling, M. E (n) equivariant graph neural networks. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.

Schneuing, A., Du, Y., Harris, C., Jamasb, A., Igashov, I., Du, W., Blundell, T., Lió, P., Gomes, C., Welling, M., et al.

Structure-based drug design with equivariant diffusion models. *arXiv preprint arXiv:2210.13695*, 2022.

Seo, S. and Kim, W. Y. Pharmaconet: Accelerating structure-based virtual screening by pharmacophore modeling. *arXiv preprint arXiv:2310.00681*, 2023.

Shi, C., Xu, M., Zhu, Z., Zhang, W., Zhang, M., and Tang, J. Graphaf: a flow-based autoregressive model for molecular graph generation. *arXiv preprint arXiv:2001.09382*, 2020.

Simonovsky, M. and Komodakis, N. Graphvae: Towards generation of small graphs using variational autoencoders. In *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part I 27*, pp. 412–422. Springer, 2018.

Somnath, V. R., Pariset, M., Hsieh, Y.-P., Martinez, M. R., Krause, A., and Bunne, C. Aligned diffusion schrödinger bridges. In *Uncertainty in Artificial Intelligence*, pp. 1985–1995. PMLR, 2023.

Song, Y. and Ermon, S. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Su, X., Song, J., Meng, C., and Ermon, S. Dual diffusion implicit bridges for image-to-image translation. *arXiv preprint arXiv:2203.08382*, 2022.

Vargas, F., Thodoroff, P., Lamacraft, A., and Lawrence, N. Solving schrödinger bridges via maximum likelihood. *Entropy*, 23(9):1134, 2021.

Vignac, C., Krawczuk, I., Siraudin, A., Wang, B., Cevher, V., and Frossard, P. Digress: Discrete denoising diffusion for graph generation. *arXiv preprint arXiv:2209.14734*, 2022.

Wang, C., Ong, H. H., Chiba, S., and Rajapakse, J. C. Gldm: hit molecule generation with constrained graph latent diffusion model. *Briefings in Bioinformatics*, 25(3):bbae142, 2024.

Wang, G., Jiao, Y., Xu, Q., Wang, Y., and Yang, C. Deep generative learning via schrödinger bridge. In *International conference on machine learning*, pp. 10794–10804. PMLR, 2021.

Weininger, D. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988.

Wouters, O. J., McKee, M., and Luyten, J. Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *Jama*, 323(9):844–853, 2020.

Wu, L., Gong, C., Liu, X., Ye, M., and Liu, Q. Diffusion-based molecule generation with informative prior bridges. *Advances in Neural Information Processing Systems*, 35: 36533–36545, 2022.

Xu, M., Yu, L., Song, Y., Shi, C., Ermon, S., and Tang, J. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923*, 2022.

Xu, M., Powers, A. S., Dror, R. O., Ermon, S., and Leskovec, J. Geometric latent diffusion models for 3d molecule generation. In *International Conference on Machine Learning*, pp. 38592–38610. PMLR, 2023.

Ye, M., Wu, L., and Liu, Q. First hitting diffusion models for generating manifold, graph and categorical data. *Advances in Neural Information Processing Systems*, 35: 27280–27292, 2022.

Zang, C. and Wang, F. Moflow: an invertible flow model for generating molecular graphs. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 617–626, 2020.

Zhang, Z., Liu, Q., Wang, H., Lu, C., and Lee, C.-K. Motif-based graph self-supervised learning for molecular property prediction. *Advances in Neural Information Processing Systems*, 34:15870–15882, 2021.

Zhou, L., Lou, A., Khanna, S., and Ermon, S. Denoising diffusion bridge models. *arXiv preprint arXiv:2309.16948*, 2023.

Zhu, H., Zhou, R., Cao, D., Tang, J., and Li, M. A pharmacophore-guided deep learning approach for bioactive molecular generation. *Nature Communications*, 14 (1):6234, 2023.

A. Proofs

A.1. Derivation of the Denoising Diffusion Bridge

Applying Doob's h-transforms to the diffusion process with the law of (1) results in a diffusion bridge solving Eq. (3), i.e., $d\mathbf{G}_t = [\mathbf{f}(\mathbf{G}_t, t) + g(t)^2 \nabla_{\mathbf{G}_t} \log p(\mathbf{G}_T|\mathbf{G}_t)]dt + g(t)d\mathbf{w}_t$, which is pinned on both ends $\mathbf{G}_0 = \mathbf{g} \sim q_{data}(\mathbf{g})$ and $\mathbf{G}_T = \mathbf{\Gamma} \sim q_{data}(\mathbf{\Gamma})$. Pioneer works (Zhou et al., 2023; Peluchetti, 2023) have proven that the time evolution of the conditional probability fixed at both ends $p(\mathbf{G}_t|\mathbf{G}_0 = \mathbf{g}, \mathbf{G}_T = \mathbf{\Gamma})$ follows the Fokker-Planck equation which amounts to the h-transformed diffusion bridge SDE (3). Starting from here, we provide the proof of Theorem 3.1.

$$
\begin{aligned}
&\frac{\partial}{\partial t} p(\mathbf{G}_t|\mathbf{G}_0 = \mathbf{g}, \mathbf{G}_T = \mathbf{\Gamma}) \\
&= -\sum_{i=1}^{d} \frac{\partial}{\partial \mathbf{G}_t^i} \left[\left(\mathbf{f}^i(\mathbf{G}_t, t) + g(t)^2 \sum_{j=1}^{d} \frac{\partial}{\partial \mathbf{G}_t^j} \log p(\mathbf{G}_T|\mathbf{G}_t) \right) p(\mathbf{G}_t|\mathbf{G}_0 = \mathbf{g}, \mathbf{G}_T = \mathbf{\Gamma}) \right] \\
&\quad + \frac{1}{2} g^2(t) \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial^2}{\partial \mathbf{G}_t^i \mathbf{G}_t^j} p(\mathbf{G}_t|\mathbf{G}_0 = \mathbf{g}, \mathbf{G}_T = \mathbf{\Gamma})
\end{aligned}
\tag{12}
$$

Theorem 3.1 (Denoising diffusion bridge)**.** *By reversing the h-transformed diffusion bridge with the law of* (3)*, we obtain the following ODE to model the time evolution of the transition density $q(\mathbf{G}_t|\mathbf{G}_T)$:*

$$
\begin{aligned}
d\mathbf{G}_t = [\mathbf{f}(\mathbf{G}_t, t) - g^2(t)(\frac{1}{2} \nabla_{\mathbf{G}_t} \log q(\mathbf{G}_t|\mathbf{G}_T) \\
- \nabla_{\mathbf{G}_t} \log p(\mathbf{G}_T|\mathbf{G}_t))]dt
\end{aligned}
\tag{4}
$$

which constructs the denoising bridge process $\{\mathbf{G}_t\}_{t=0}^T$ with marginal distribution $\mathbf{G}_T = \mathbf{\Gamma} \sim q_{data}(\mathbf{\Gamma})$, inducing the joint distribution $q(\mathbf{G}_0, \mathbf{G}_T)$ approximating $q_{data}(\mathbf{g}, \mathbf{\Gamma})$. $p(\mathbf{G}_T|\mathbf{G}_t)$ is the transition density associated with the original diffusion process specified by Eq. (1)*.*

Proof. As we design the sampling distribution to satisfy $q(\mathbf{G}_t|\mathbf{G}_0, \mathbf{G}_T) = p(\mathbf{G}_t|\mathbf{G}_0, \mathbf{G}_T)$ in Eq. (6). The transition density associated with SDE (4) equals to the expectation with respect to $\mathbf{G}_0$:

$$
\begin{aligned}
q(\mathbf{G}_t|\mathbf{G}_T = \mathbf{\Gamma}) &= \int_{\mathbf{G}_0} q(\mathbf{G}_t|\mathbf{G}_0, \mathbf{G}_T = \mathbf{\Gamma}) q_{data}(\mathbf{G}_0|\mathbf{G}_T = \mathbf{\Gamma}) d\mathbf{G}_0 \\
&= \int_{\mathbf{G}_0} p(\mathbf{G}_t|\mathbf{G}_0, \mathbf{G}_T = \mathbf{\Gamma}) q_{data}(\mathbf{G}_0|\mathbf{G}_T = \mathbf{\Gamma}) d\mathbf{G}_0 \\
&= \mathbb{E}_{\mathbf{G}_0 \sim q_{data}(\mathbf{G}_0|\mathbf{G}_T = \mathbf{\Gamma})} \left[p(\mathbf{G}_t|\mathbf{G}_0, \mathbf{G}_T = \mathbf{\Gamma}) \right]
\end{aligned}
\tag{13}
$$

To investigate the time evolution of $q(\mathbf{G}_t|\mathbf{G}_T)$, we marginalize out $\mathbf{G}_0$ in the Fokker-Planck equation (12):

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{G}_0 \sim q_{data}(\mathbf{G}_0|\mathbf{G}_T = \mathbf{\Gamma})} \left[\frac{\partial}{\partial t} p(\mathbf{G}_t|\mathbf{G}_0, \mathbf{G}_T = \mathbf{\Gamma}) \right] \\
&= -\sum_{i=1}^{d} \frac{\partial}{\partial \mathbf{G}_t^i} \left[\left(\mathbf{f}^i(\mathbf{G}_t, t) + g(t)^2 \sum_{j=1}^{d} \frac{\partial}{\partial \mathbf{G}_t^j} \log p(\mathbf{G}_T|\mathbf{G}_t) \right) \mathbb{E}_{\mathbf{G}_0 \sim q_{data}(\mathbf{G}_0|\mathbf{G}_T = \mathbf{\Gamma})} \left[p(\mathbf{G}_t|\mathbf{G}_0, \mathbf{G}_T = \mathbf{\Gamma}) \right] \right] + \\
&\quad + \frac{1}{2} g^2(t) \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial^2}{\partial \mathbf{G}_t^i \mathbf{G}_t^j} \mathbb{E}_{\mathbf{G}_0 \sim q_{data}(\mathbf{G}_0|\mathbf{G}_T = \mathbf{\Gamma})} [p(\mathbf{G}_t|\mathbf{G}_0, \mathbf{G}_T = \mathbf{\Gamma}]
\end{aligned}
\tag{14}
$$

Using the result in Eq. (13), Eq. (14) becomes

$$
\begin{aligned}
&\frac{\partial}{\partial t} q(\mathbf{G}_t|\mathbf{G}_T = \mathbf{\Gamma}) \\
&= -\sum_{i=1}^{d} \frac{\partial}{\partial \mathbf{G}_t^i} \left[\left(\mathbf{f}^i(\mathbf{G}_t, t) + g(t)^2 \nabla_{\mathbf{G}_t} \log p(\mathbf{G}_T|\mathbf{G}_t) \right) q(\mathbf{G}_t|\mathbf{G}_T = \mathbf{\Gamma}) \right] + \frac{1}{2} g^2(t) \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial^2}{\partial \mathbf{G}_t^i \mathbf{G}_t^j} q(\mathbf{G}_t|\mathbf{G}_T = \mathbf{\Gamma}) \\
&= -\sum_{i=1}^{d} \frac{\partial}{\partial \mathbf{G}_t^i} \left[\left(\mathbf{f}^i(\mathbf{G}_t, t) + g(t)^2 \nabla_{\mathbf{G}_t} \log p(\mathbf{G}_T|\mathbf{G}_t) - \frac{1}{2} g^2(t) \nabla_{\mathbf{G}_t} \log q(\mathbf{G}_t|\mathbf{G}_T = \mathbf{\Gamma}) \right) q(\mathbf{G}_t|\mathbf{G}_T = \mathbf{\Gamma}) \right]
\end{aligned}
\tag{15}
$$

which amounts to the Liouville equation (a special case of the Fokker-Planck equation when the diffusion coefficient is zero):

$$
\begin{aligned}
&\frac{\partial}{\partial t} q(\mathbf{G}_t|\mathbf{G}_T = \mathbf{\Gamma}) = -\sum_{i=1}^{d} \frac{\partial}{\partial \mathbf{G}_t^i} \left[\tilde{\mathbf{f}}^i(\mathbf{G}_t, t) q(\mathbf{G}_t|\mathbf{G}_T = \mathbf{\Gamma}) \right], \\
&\text{where} \quad \tilde{\mathbf{f}}^i(\mathbf{G}_t, t) = \left(\mathbf{f}^i(\mathbf{G}_t, t) + g(t)^2 \nabla_{\mathbf{G}_t} \log p(\mathbf{G}_T|\mathbf{G}_t) - \frac{1}{2} g^2(t) \nabla_{\mathbf{G}_t} \log q(\mathbf{G}_t|\mathbf{G}_T = \mathbf{\Gamma}) \right)
\end{aligned}
\tag{16}
$$

This defines a probability flow ODE $d\mathbf{G}_t = \tilde{\mathbf{f}}(\mathbf{G}_t, t)dt$ that is the same with Eq. (4) specified in Theorem 3.1. □

A.2. $SE(3)$-equivariant Denoising Diffusion Bridge

As proposed by (Luo et al., 2024), given a diffusion process solving $d\mathbf{G}_t = \mathbf{f}(\mathbf{G}_t, t)dt + g(t)d\mathbf{w}_t$, its associated transition density $p(\mathbf{G}_T|\mathbf{G}_t)$ is $SE(3)$-equivariant, i.e., $p(g \cdot \mathbf{G}_T|g \cdot \mathbf{G}_t) = p(\mathbf{G}_T|\mathbf{G}_t)$, $\forall g \in SE(3)$, if the following conditions are satisfied: (1) the initial density $q(\mathbf{G}_0)$ is $SE(3)$-invariant; and (2) $\mathbf{f}(\cdot, t)$ is $SO(3)$-equivariant and $T(3)$-invariant; (3) the transition density of $\mathbf{w}_t$ is $SE(3)$-equivariant.

In our case, the initial density $q(\mathbf{G}_0) = q_{data}(\mathbf{g}) = \delta(\mathbf{G}_0 - \mathbf{g})$ is a Dirac distribution that is invariant to rotations or translations. As we design the bridges using VP and VE schedules, the drift functions (as shown in Table 4) are apparently $SO(3)$-equivariant. To meet $T(3)$-invariance, we calculate the center of mass of the system $(\mathbf{g}, \mathbf{\Gamma})$ and move the center to zero. Thus, we have a diffusion process with equivariant transition density $p(\mathbf{G}_T|\mathbf{G}_t)$.

Based on this equivariant diffusion process, we can further devise the equivariant denoising diffusion bridge:

Theorem 3.2 (Equivariant denoising diffusion bridge)**.** *Given an $SE(3)$-equivariant diffusion process with transition density $p(\mathbf{G}_T|\mathbf{G}_t)$, its h-transformed diffusion bridge with the law of (3) and the corresponding denoising diffusion bridge characterized in Theorem 3.1 also preserve $SE(3)$-equivariance. Their transition densities satisfy*

- $p(g \cdot \mathbf{G}_t|g \cdot \mathbf{G}_0, g \cdot \mathbf{G}_T) = p(\mathbf{G}_t|\mathbf{G}_0, \mathbf{G}_T)$, $\forall g \in SE(3)$
- $q(g \cdot \mathbf{G}_t|g \cdot \mathbf{G}_T) = q(\mathbf{G}_t|\mathbf{G}_T)$, $\forall g \in SE(3)$

Proof. We first investigate the equivariance of $p(\mathbf{G}_t|\mathbf{G}_0, \mathbf{G}_T)$, the marginal distribution of $\mathbf{G}_t$ with fixed endpoints. According to (Zhou et al., 2023), this distribution follows Bayes' rule

$$
p(\mathbf{G}_t|\mathbf{G}_0, \mathbf{G}_T) = \frac{p(\mathbf{G}_t|\mathbf{G}_0) p(\mathbf{G}_T|\mathbf{G}_t)}{p(\mathbf{G}_T|\mathbf{G}_0)} \tag{17}
$$

As we have obtained an $SE(3)$-equivariant transition density $p(\mathbf{G}_T|\mathbf{G}_t)$, when t equals to zero, it is apparent that $p(\mathbf{G}_T|\mathbf{G}_0)$ is also $SE(3)$-equivariant. We therefore focus on investigating the equivariance of $p(\mathbf{G}_t|\mathbf{G}_0)$ now.

Since each point cloud is defined by node position matrix and feature matrix, i.e., $\mathbf{G} = (\mathbf{x}, \mathbf{h})$, and the node features always remain invariant to rotations and translations, we only consider the effect of transformations with regard to $\mathbf{x}$ now. The forward time evolution of $p(\mathbf{x}_t|\mathbf{x}_0)$ follows the Fokker-Planck equation:

$$
\frac{\partial}{\partial t} p(\mathbf{x}_t|\mathbf{x}_0) = -\sum_{i=1}^{d} \frac{\partial}{\partial \mathbf{x}_t^i} \left[\mathbf{f}^i(\mathbf{x}_t, t) p(\mathbf{x}_t|\mathbf{x}_0) \right] + \frac{1}{2} g^2(t) \sum_{i=1}^{d} \frac{\partial^2}{\partial \mathbf{x}_t^{i2}} p(\mathbf{x}_t|\mathbf{x}_0) \tag{18}
$$

For any transformation $g \in SE(3)$, let $\mathbf{y} = g \cdot \mathbf{x} = \mathbf{R}\mathbf{x} + \mathbf{t}$ denote the transformed atom positions, where $\mathbf{R} \in \mathbb{R}^{3\times 3}$ is an orthogonal matrix defining the rotation and $\mathbf{t} \in \mathbb{R}^3$ defines the translation. The Fokker-Planck equation of the transformed data becomes

$$\frac{\partial}{\partial t} p(\mathbf{y}_t|\mathbf{y}_0) = -\sum_{i=1}^{d} \frac{\partial}{\partial \mathbf{y}_t^i} \left[\mathbf{f}^i(\mathbf{y}_t, t) p(\mathbf{y}_t|\mathbf{y}_0)\right] + \frac{1}{2} g^2(t) \sum_{i=1}^{d} \frac{\partial^2}{\partial \mathbf{y}_t^{i2}} p(\mathbf{y}_t|\mathbf{y}_0) \tag{19}$$

The partial derivative with respect to $\mathbf{y}$ follows

$$\begin{aligned} \frac{\partial}{\partial \mathbf{y}_t^i} &= \sum_{j=1}^{d} \frac{\partial \mathbf{x}_t^j}{\partial \mathbf{y}_t^i} \cdot \frac{\partial}{\partial \mathbf{x}_t^j} = \sum_{j=1}^{d} \mathbf{R}_{ji}^{-1} \frac{\partial}{\partial \mathbf{x}_t^j} = \sum_{j=1}^{d} \mathbf{R}_{ij} \frac{\partial}{\partial \mathbf{x}_t^j}, \\ \text{and} \quad \frac{\partial^2}{\partial \mathbf{y}_t^{i2}} &= \sum_{j=1}^{d} \sum_{k=1}^{d} \frac{\partial \mathbf{x}_t^j}{\partial \mathbf{y}_t^i} \cdot \frac{\partial}{\partial \mathbf{x}_t^j} \cdot \frac{\partial \mathbf{x}_t^k}{\partial \mathbf{y}_t^i} \cdot \frac{\partial}{\partial \mathbf{x}_t^k} = \sum_{j=1}^{d} \sum_{k=1}^{d} \mathbf{R}_{ij} \mathbf{R}_{ik} \frac{\partial}{\partial \mathbf{x}_t^j} \frac{\partial}{\partial \mathbf{x}_t^k} \end{aligned} \tag{20}$$

Since $\mathbf{f}(\cdot, t)$ is designed to be $SO(3)$-equivariant and $T(3)$-invariant, we have

$$\mathbf{f}^i(\mathbf{y}_t, t) = \mathbf{f}^i(\mathbf{R}\mathbf{x}_t + \mathbf{t}, t) = (\mathbf{R}\mathbf{f}(\mathbf{x}_t, t))_i = \sum_{k=1}^{d} \mathbf{R}_{ik} \mathbf{f}^k(\mathbf{x}_t, t) \tag{21}$$

With Eq. (20) and (21), we can readily rewrite the transformed Fokker-Planck equation as

$$\frac{\partial}{\partial t} p(\mathbf{y}_t|\mathbf{y}_0) = -\sum_{i=1}^{d} \sum_{j=1}^{d} \sum_{k=1}^{d} \mathbf{R}_{ij} \mathbf{R}_{ik} \frac{\partial}{\partial \mathbf{x}_t^j} \left[\mathbf{f}^k(\mathbf{x}_t, t) p(\mathbf{y}_t|\mathbf{y}_0)\right] + \frac{1}{2} g^2(t) \sum_{i=1}^{d} \sum_{j=1}^{d} \sum_{k=1}^{d} \mathbf{R}_{ij} \mathbf{R}_{ik} \frac{\partial}{\partial \mathbf{x}_t^j} \frac{\partial}{\partial \mathbf{x}_t^k} p(\mathbf{y}_t|\mathbf{y}_0) \tag{22}$$

Notably, since $\mathbf{R}$ is an orthogonal matrix, we have $\sum_{i=1}^{d} \mathbf{R}_{ij} \mathbf{R}_{ik} = \delta_{jk}$, which is zero when $j \neq k$. We further simplify (22) to

$$\begin{aligned} \frac{\partial}{\partial t} p(\mathbf{y}_t|\mathbf{y}_0) &= -\sum_{j=1}^{d} \sum_{k=1}^{d} \delta_{jk} \frac{\partial}{\partial \mathbf{x}_t^j} \left[\mathbf{f}^k(\mathbf{x}_t, t) p(\mathbf{y}_t|\mathbf{y}_0)\right] + \frac{1}{2} g^2(t) \sum_{j=1}^{d} \sum_{k=1}^{d} \delta_{jk} \frac{\partial}{\partial \mathbf{x}_t^j} \frac{\partial}{\partial \mathbf{x}_t^k} p(\mathbf{y}_t|\mathbf{y}_0) \\ &= -\sum_{j=1}^{d} \frac{\partial}{\partial \mathbf{x}_t^j} \left[\mathbf{f}^j(\mathbf{x}_t, t) p(\mathbf{y}_t|\mathbf{y}_0)\right] + \frac{1}{2} g^2(t) \sum_{j=1}^{d} \frac{\partial^2}{\partial \mathbf{x}_t^{j2}} p(\mathbf{y}_t|\mathbf{y}_0) \end{aligned} \tag{23}$$

which characterizes the same time evolution as $p(\mathbf{x}_t|\mathbf{x}_0)$ in Eq. (18). Moreover, the boundary condition $p(\mathbf{y}|\mathbf{y}_0) = \delta(\mathbf{y} - \mathbf{y}_0) = \delta(\mathbf{x} - \mathbf{x}_0) = p(\mathbf{x}|\mathbf{x}_0)$ is also $SE(3)$-equivariant. Therefore, the transition density $p(\mathbf{G}_t|\mathbf{G}_0)$ and the marginal distribution $p(\mathbf{G}_t|\mathbf{G}_0, \mathbf{G}_T)$ both satisfy $SE(3)$-equivariance, i.e., $p(g \cdot \mathbf{G}_t|g \cdot \mathbf{G}_0, g \cdot \mathbf{G}_T) = p(\mathbf{G}_t|\mathbf{G}_0, \mathbf{G}_T), \forall g \in SE(3)$.

Next, we look into the equivariance of the transition density associated with the denoising diffusion bridge $q(\mathbf{G}_t|\mathbf{G}_T)$. Focusing on the position matrix $\mathbf{x}$, we have $q(\mathbf{x}_t|\mathbf{x}_T) = \int_{\mathbf{x}_0} p(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T) q_{data}(\mathbf{x}_0|\mathbf{x}_T) d\mathbf{x}_0$. Similarly, for any transformation $g \in SE(3)$, the transformed atom positions are denoted by $\mathbf{y} = g \cdot \mathbf{x} = \mathbf{R}\mathbf{x} + \mathbf{t}$. The transition density of $\mathbf{y}_t$ becomes

$$\begin{aligned} q(\mathbf{y}_t|\mathbf{y}_T) &= \int p(\mathbf{y}_t|\mathbf{x}_0, \mathbf{y}_T) q_{data}(\mathbf{x}_0|\mathbf{y}_T) d\mathbf{x}_0 \\ &= \int p(g \cdot \mathbf{x}_t|g \cdot g^{-1} \cdot \mathbf{x}_0, g \cdot \mathbf{x}_T) q_{data}(g \cdot g^{-1} \cdot \mathbf{x}_0|g \cdot \mathbf{x}_T) d\mathbf{x}_0 \end{aligned} \tag{24}$$

Since we have proven $p(g \cdot \mathbf{x}_t|g \cdot \mathbf{x}_0, g \cdot \mathbf{x}_T) = p(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T)$, and the distribution of the paired molecule and pharmacophore data also satisfies $q_{data}(g \cdot \mathbf{G}_0|g \cdot \mathbf{G}_T) = q_{data}(\mathbf{G}_0|\mathbf{G}_T)$, we have

$$q(\mathbf{y}_t|\mathbf{y}_T) = \int p(\mathbf{x}_t|g^{-1} \cdot \mathbf{x}_0, \mathbf{x}_T) q_{data}(g^{-1} \cdot \mathbf{x}_0|\mathbf{x}_T) d\mathbf{x}_0 \tag{25}$$

Let $\mathbf{x}_0' = g^{-1} \cdot \mathbf{x}_0$, then $\mathbf{x}_0 = g \cdot \mathbf{x}_0' = \mathbf{R}\mathbf{x}_0' + \mathbf{t}$. Therefore,

$$\frac{\partial \mathbf{x}_0}{\partial \mathbf{x}_0'} = \mathbf{R}, \quad \textit{and} \quad \frac{d\mathbf{x}_0}{d\mathbf{x}_0'} = \det(\mathbf{R}) = 1 \tag{26}$$

since $\mathbf{R}$ is an orthogonal matrix.

Notably, Eq. (25) becomes

$$\begin{aligned} q(\mathbf{y}_t|\mathbf{y}_T) &= \int p(\mathbf{x}_t|\mathbf{x}_0', \mathbf{x}_T) q_{data}(\mathbf{x}_0'|\mathbf{x}_T) \det(\mathbf{R}) d\mathbf{x}_0' \\ &= \int p(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T) q_{data}(\mathbf{x}_0|\mathbf{x}_T) d\mathbf{x}_0 \\ &= q(\mathbf{x}_t|\mathbf{x}_T) \end{aligned} \tag{27}$$

which completes the proof. □

B. Implementation details

B.1. Bridge Parameterization

PharmacoBridge supports flexible choices of the bridge design, such as the drift function $\mathbf{f}(\cdot, \mathbf{t})$ and the diffusion coefficient $g(\cdot)$. We here provide two typical designs, i.e., VP and VE bridges in Table 4. For VP bridge, the variance schedule is given by $\beta_t = (\beta_{max} - \beta_{min})t + \beta_{min}$, and α_t is the integral of β_t: $\alpha_t = \int_0^t \beta_\tau d\tau$. For VE bridge, we have $\alpha_t = 1$ which is time-invariant.

For the choice of σ_0 and σ_T, we have performed grid search among $\{0.1n | n \in [1:10]\}$. Finally, we set $\sigma_{0,pos} = 0.1$, $\sigma_{T,pos} = 0.3$, $\sigma_{0,feat} = 0.7$ and $\sigma_{T,feat} = 1.0$. The covariance is calculated via $\sigma_{0T} = \sigma_0^2/2$.

Table 4. Design of VP and VE bridges

	$\mathbf{f}(\mathbf{G}_t, t)$	$g^2(t)$	$p(\mathbf{G}_t \mid \mathbf{G}_0)$	SNR_t	$\nabla_{\mathbf{G}_t} \log p(\mathbf{G}_T \| \mathbf{G}_t)$
VP	$-\frac{1}{2}\beta_t \mathbf{G}_t$	β_t	$\mathcal{N}\left(\alpha_t \mathbf{G}_0, \sigma_t^2 \boldsymbol{I}\right)$	α_t^2/σ_t^2	$\frac{(\alpha_t/\alpha_T)\mathbf{G}_T - \mathbf{G}_t}{\sigma_t^2(\mathrm{SNR}_t/\mathrm{SNR}_T - 1)}$
VE	$\mathbf{0}$	$\frac{d}{dt}\sigma_t^2$	$\mathcal{N}\left(\mathbf{G}_0, \sigma_t^2 \boldsymbol{I}\right)$	$1/\sigma_t^2$	$\frac{\mathbf{G}_T - \mathbf{G}_t}{\sigma_T^2 - \sigma_t^2}$

B.2. Scaling Functions of Denoiser D_θ

Following DDBM (Zhou et al., 2023), our bridge scalings are designed as

$$\begin{aligned} c_{in}(t) &= \frac{1}{\sqrt{a_t^2\sigma_T^2 + b_t^2\sigma_t^2 + 2a_t b_t \sigma_{0T} + c_t}}, \\ c_{out}(t) &= \sqrt{a_t^2(\sigma_T^2\sigma_0^2 - \sigma_{0T}^2) + \sigma_0^2 c_t} * c_{in}(t), \\ c_{skip}(t) &= (b_t\sigma_0^2 + a_t\sigma_{0T}) * c_{in}^2(t), \\ c_{noise}(t) &= \frac{1}{4}\log(\sigma_t), \\ \lambda(t) &= \frac{1}{c_{out}(t)^2}, \\ \text{where} \quad a_t &= \frac{\alpha_t \mathrm{SNR}_T}{\alpha_T \mathrm{SNR}_t}, \quad b_t = \alpha_t(1 - \frac{\mathrm{SNR}_T}{\mathrm{SNR}_t}), \quad \text{and} \quad c_t = \sigma_t^2(1 - \frac{\mathrm{SNR}_T}{\mathrm{SNR}_t}) \end{aligned} \tag{28}$$

where σ_0^2 and σ_T^2 refer to the variance of the molecular graph $\mathbf{g}$ and the pharmacophore graph $\mathbf{\Gamma}$, respectively. The data variance at t-th step is configured as $\sigma_t = t$. The choice of σ_0 and σ_t is discussed above in Appendix B.1.

B.3. Time Discretization

Following EDM (Karras et al., 2022), we discretize the time steps according to

$$t_i = (\sigma_{max}^{\frac{1}{\rho}} + \frac{N-i}{N-1}(\sigma_{min}^{\frac{1}{\rho}} - \sigma_{max}^{\frac{1}{\rho}}))^\rho, \quad \text{when} \quad i > 0 \tag{29}$$

where ρ is the parameter to control that shorter steps are taken near σ_{min}, and is set to $\rho = 7$ by default. We set $t_0 = 0$ to ensure the output is the reconstructed data sample. Empirically, for VE bridge, we set $\sigma_{min} = 0.02$ and $\sigma_{max} = 80$. And for VP bridge, we set $\sigma_{min} = 0.0001$ and $\sigma_{max} = 1$.

B.4. $SE(3)$-equivariant Network

As illustrated in Eq. (8) and (9), the target of our score matching model is finalized to predict the mixture of signal and noise with F_θ, which is, in our case, to predict the molecular point cloud with certain noise. A molecular point cloud is defined as $\mathbf{g} = (\mathbf{x}_{mol}, \mathbf{h}_{mol})$, where $\mathbf{x}_{mol} \in \mathbb{R}^{N\times 3}$ represents the coordinates of N atoms in the 3D Euclidean space and $\mathbf{h}_{mol} \in \mathbb{R}^{N\times M}$ represents the node feature matrix that is composed of the one-hot encodings of M atom types. A pharmacophore point cloud is defined as $\mathbf{\Gamma} = (\mathbf{x}_{pp}, \mathbf{h}_{pp})$. Similarly, $\mathbf{x}_{pp} \in \mathbb{R}^{N\times 3}$ records the center coordinates of the pharmacophore that each atom belongs to, and $\mathbf{h}_{pp} \in \mathbb{R}^{N\times K}$ is the feature matrix of the one-hot encodings of K pharmacophore types. For data consistency, we perform zero padding to the feature matrix with less number of features, i.e., $\mathbf{h}_{mol}, \mathbf{h}_{pp} \in \mathbb{R}^{N\times F}$ with $F = \max(M, K)$. Thus, the behavior of F_θ at time step t is formulated as

$$\begin{aligned}(\tilde{\mathbf{x}}, \tilde{\mathbf{h}}) &= F_\theta(c_{in}(t)\mathbf{G}_t, c_{noise}(t)) \\ &= F_\theta(c_{in}(t)(\mathbf{x}_t, \mathbf{h}_t), c_{noise}(t))\end{aligned} \tag{30}$$

Equivariance To manipulate 3D objects, it is crucial that the model F_θ satisfies $SE(3)$-equivariance. This ensures, for the coordinates $\mathbf{x}$, transformations such as rotations and translations of the input graph should result in equivariant output transformations, whereas the node features $\mathbf{h}$ should be invariant to any transformations. Assuming any transformation given by $\mathbf{Rx} + \mathbf{t}$ is undertaken to the input coordinates, an equivariant model should have

$$(\mathbf{R}\tilde{\mathbf{x}} + \mathbf{t}, \tilde{\mathbf{h}}) = F_\theta\left(c_{in}(t)(\mathbf{Rx} + \mathbf{t}, \mathbf{h}), c_{noise}(t)\right) \tag{31}$$

where $\mathbf{R} \in \mathbb{R}^{3\times 3}$ is an orthogonal matrix defining the rotation and $\mathbf{t} \in \mathbb{R}^3$ defines the translation.

To preserve $SE(3)$-equivariance, we build our model with EGNN (Satorras et al., 2021). To condition the updating of the molecular point cloud with the desired pharmacophore models, we concatenate the coordinates and features of the molecular point cloud at time t and the initial pharmacophore point cloud to form a combined graph $\mathbf{G}_t = (\mathbf{x}_t, \mathbf{h}_t)$, where $\mathbf{x}_t \in \mathbb{R}^{2N\times 3}$ and $\mathbf{h}_t \in \mathbb{R}^{2N\times F}$. A mask $\mathcal{M}^{mol} \in \mathbb{R}^{2N}$ is applied to only update the nodes belonging to the molecular point cloud.

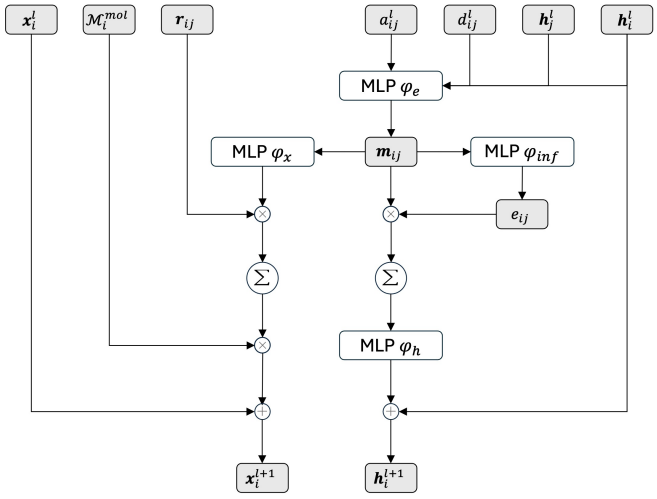

Figure 6. Flowchart of the EGCL module. $\otimes$ and $\oplus$ denote element-wise multiplication and addition, respectively. $\sum$ represents summation over all neighbors.

The Equivariant Graph Convolutional Layer (EGCL) that we used for feature updating is illustrated in Figure 6. At l-th

layer, the updating procedure is given by

$$
\begin{aligned}
\mathbf{m}_{ij} &= \phi_e(\mathbf{h}_i^l, \mathbf{h}_j^l, d_{ij}^2, a_{ij}), \\
\mathbf{h}_i^{l+1} &= \mathbf{h}_i^l + \phi_h(\mathbf{h}_i^l, \sum_{j \in \mathcal{N}(i)} e_{ij}\mathbf{m}_{ij}), \\
\mathbf{x}_i^{l+1} &= \mathbf{x}_i^l + \sum_{j \in \mathcal{N}(i)} \mathbf{r}_{ij}\phi_x(\mathbf{m}_{ij}) \cdot \mathcal{M}_i^{mol}
\end{aligned}
\tag{32}
$$

where i and j refer to the node index. $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$ is the euclidean distance between node i and j, and $\mathbf{r}_{ij} = \mathbf{x}_i^l - \mathbf{x}_j^l$ is the vector difference between node i and j. a_{ij} is the edge attribute indicating the interaction is between atoms, pharmacophore nodes, or an atom and a pharmacophore node. e_{ij} is the soft estimation of the existence of edge between node i and j, which is computed by $e_{ij} = \phi_{inf}(\mathbf{m}_{ij})$. Initially, the time step t_i is incorporated with $\mathbf{h}_i^0$ via $\mathbf{h}_i^0 = \phi_{time}(\mathbf{h}_i^0, t_i)$. ϕ_{time}, ϕ_e, ϕ_h, ϕ_x and ϕ_{inf} are all trainable MLPs. As demonstrated by TargetDiff (Guan et al., 2023), EGCL governs the rotational equivariance. To fulfill translational equivariance as well, we must shift the center of the pharmacophore point cloud to zero.

B.5. Data Preprocessing

We extracted the pharmacophore and atom features from the ligands to form the node feature matrix. Typical pharmacophore features include Hydrophobe, Aromatic ring, Cation, Anion, Hydrogen Bond donor, Hydrogen Bond acceptor and Halogen. We considered atoms that is not in any pharmacophore belonging to an additional class, i.e., Linker. The pharmacophore feature matrix $\mathbf{h}_{pp}$ is thus composed of the one-hot encoding of each atom's pharmacophore cluster membership. In case of overlapping, we considered the atom belonging to the larger pharmacophore cluster with more atoms. For the atom features $\mathbf{h}_{mol}$, we considered two modes, i.e., basic mode and aromatic mode. In the basic mode, $\mathbf{h}_{mol}$ consists of the one-hot encoding of the atom type, including C, N, O, F, P, S and Cl. In the aromatic mode, $\mathbf{h}_{mol}$ consists of the one-hot encoding of the combination of the atom type and the aromatic property, which denotes if an atom appears in an aromatic ring. In this case, atom features contain C, C.ar, N, N.ar, O, O.ar, F, P, P.ar, S, S.ar and Cl.

The position matrix $\mathbf{x}_{mol}$ is constructed using the coordinates of the atoms and $\mathbf{x}_{pp}$ is initialized as the center of pharmacophores adding a Gaussian noise with a small deviation to avoid having exactly the same destination for different nodes in the same pharmacophore. For atoms not belonging to any pharmacophore, we apply HDBSCAN to cluster the nodes that are spatially close and set $\mathbf{x}_{pp}$ as the center of the cluster.

C. Additional Results

C.1. Ablation Studies on SA and QED Evaluations

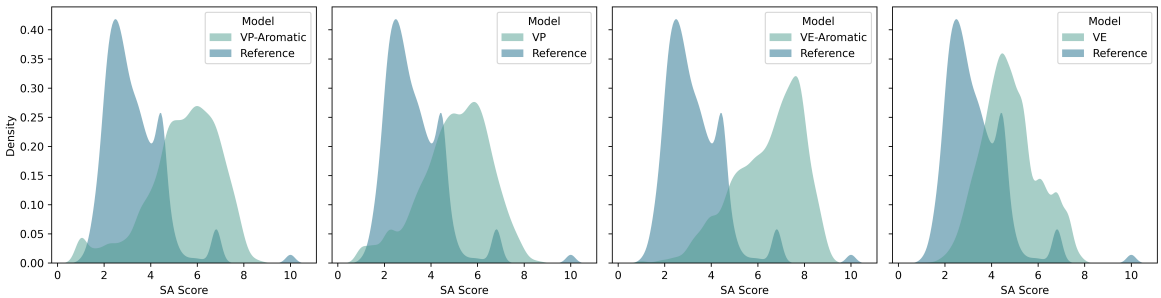

Figure 7. Ablation studies on SA analysis. PharmacoBridge with the VP design achieved similar SA scores with or without aromatic features. PharmacoBridge with the VE design without aromatic features achieved the most similar SA score distribution compared to the dataset distribution, but incorporating aromatic features significantly lower the synthetic accessibility.

In this section, we investigate the effects of equipping PharmacoBridge with VP or VE schedules and the incorporation of aromatic features in SA and QED evaluations. Figure 7 and Figure 8 present the SA and QED score distributions of the ablation studies of our method. VP-based PharmacoBridge always achieved stable SA and QED score distributions with or without aromatic features. VE-based PharmacoBridge achieved lower SA scores (i.e., better synthetic accessibility) when

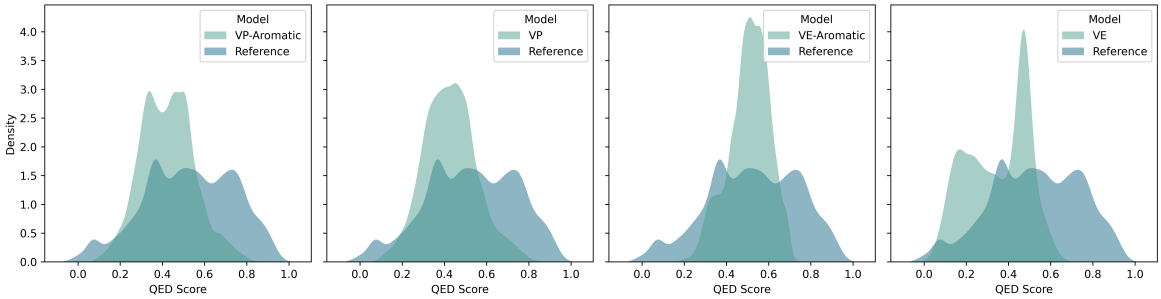

Figure 8. Ablation studies on QED analysis. Similar to SA analysis, PharmacoBridge with the VP design achieved comparative QED scores with or without aromatic features. PharmacoBridge with the VE design with aromatic features achieved the highest range of QED scores, but the QED scores dropped significantly without aromatic features.

aromatic features were absent, but obtained higher QED scores when aromatic features were present. Considering other ablation study results shown in Table 1, we chose VP-based PharmacoBridge with aromatic features to continue subsequent conditional generation experiments.

C.2. Generated Samples

Figure 9 shows the reference molecules, and the molecules generated by PharmacoBridge and other baselines using the pharmacophores identified in the binding complex structures of 1EOC, 3OCC and 5UEV, including both 3D docked binding complexes and 2D chemical structures.

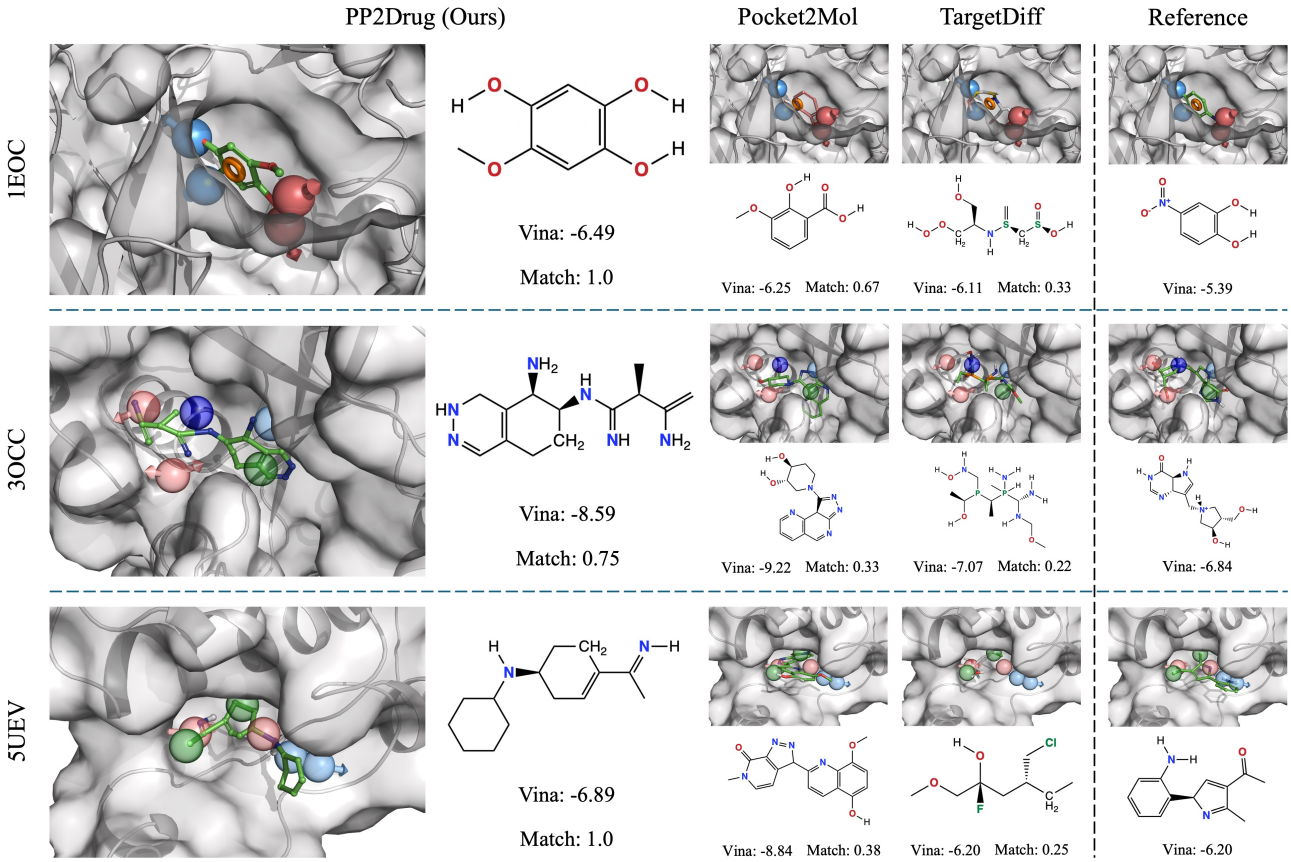

Figure 9. Molecules generated with the pharmacophore models of ligands from PDB structure 5JOY, 4CPI, 1DOD. Docked binding complexes and chemical structures, together with the Vina and CNN scores of both generated and original molecules are shown.