Target-aware Molecular Graph Generation

Anonymous Authors[1]

Abstract

Generating molecules with desired biological activities has attracted growing attention in drug discovery. Previous molecular generation models are designed as chemocentric methods that hardly consider the drug-target interaction, limiting their practical applications. In this paper, we aim to generate molecular drugs in a target-aware manner that bridges biological activity and molecular design. To solve this problem, we compile a benchmark dataset from several publicly available datasets and build baselines in a unified framework. Building on the recent advantages of flow-based molecular generation models, we propose *SiamFlow*, which forces the flow to fit the distribution of target sequence embeddings in latent space. Specifically, we employ an alignment loss and a uniform loss to bring target sequence embeddings and drug graph embeddings into agreements while avoiding collapse. Furthermore, we formulate the alignment into a one-to-many problem by learning spaces of target sequence embeddings. Experiments quantitatively show that our proposed method learns meaningful representations in the latent space toward the target-aware molecular graph generation and provides an alternative approach to bridge biology and chemistry in drug discovery.

1. Introduction

Drug discovery, which focuses on finding candidate molecules with desirable properties for therapeutic applications, is a long-period, expensively process with a high failure rate. The challenge primarily stems from the actuality that only a tiny fraction of the theoretical possible drug-like molecules may have practical effects. Specifically, the entire search space is as large as $10^{23} \sim 10^{60}$, while only 10^8 of them are therapeutically relevant (Polishchuk et al., 2013; Hert et al., 2009; Cheng et al., 2021). In the face of such difficulty, traditional methods like high-throughput screening (Hert et al., 2009) fail in terms of efficiency because of the large amount of resources required in producing minor hit compounds. One alternative is using computational methods (Phatak et al., 2009; Paricharak et al., 2018) such as virtual screening (Bajorath, 2002; Schneider, 2010) to identify hit compounds from virtual libraries through similarity-based searches or molecular docking. Another alternative is automated de novo molecule design, such as inverse QSAR (Schneider & Schneider, 2016), particle swarm optimization (Winter et al., 2019), structure-based de novo design (Schneider, 2013; Button et al., 2019), or genetic algorithms (Bandholtz et al., 2012).

Recent deep generative models have demonstrated potentials to promote drug discovery by exploring huge chemical space in a data-driven manner. Various forms of variational autoencoder (VAE) (Kingma & Welling, 2014; Gómez-Bombarelli et al., 2018; Kusner et al., 2017; Dai et al., 2018; Jin et al., 2018; Simonovsky & Komodakis, 2018), generative adversarial networks (GAN) (Goodfellow et al., 2014; Guimaraes et al., 2017; Sanchez-Lengeling et al., 2017; Prykhodko et al., 2019; Méndez-Lucio et al., 2020), autoregressive (AR) (Van Oord et al., 2016; Popova et al., 2019; You et al., 2018), and normalizing flow (NF) (Dinh et al., 2015; 2017; Madhawa et al., 2019; Honda et al., 2019; Shi et al., 2019; Zang & Wang, 2020; Luo et al., 2021) have been proposed to generate molecular SMILES or graphs. Though these approaches can generate valid and novel molecules to some extent, they remain inefficient because the generated candidate molecules need further screened against given targets. As the primary goal of these chemocentric methods is to generate drug-like molecules that satisfy specific properties, directly applying them in drug discovery requires extra efforts on predicting the binding affinities between candidate molecules and target proteins.

While previous molecular generation methods scarcely take biological drug-target interactions into account, we aim to generate candidate molecules based on a biological perspective. This paper proposes target-aware molecular generation to bridge biological activity and chemical molecular design that generate valid molecules conditioned on specific targets and thus facilitate the development of drug discovery.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

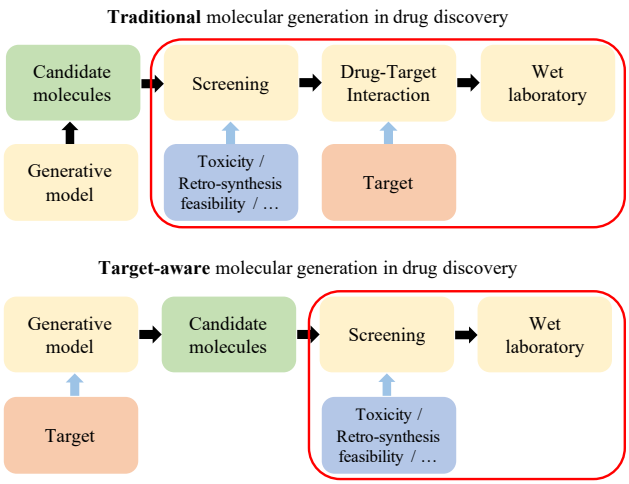

Figure 1. The computational drug discovery pipelines of traditional chemocentric and target-aware molecular generation. The black arrows denote the main steps, the blue arrows denote external considerations, and the red boxes denote the post-processing process of generated molecules in drug discovery.

As shown in Figure 1, the pipeline of computational drug discovery is supposed to be simplified to a great extent with the help of target-aware molecular generation.

Our main contributions are summarized as follows:

- We propose a target-aware molecular generation manner from a biological perspective, while prior works on chemocentric molecular generation are inefficient in practical drug discovery.
- We establish a new benchmark for the target-aware molecular generation containing abundant drug-target pairs for evaluating generative models.
- We propose SiamFlow, a siamese network architecture for the conditional generation of flow-based models. While the sequence encoder and the generative flow align in the latent space, a uniformity regularization is imposed to avoid collapse. Moreover, we implement SiamFlow and baselines in a unified framework for further research. The code will be released.

2. Related work

2.1. De Novo Molecular Generation

Recent years have witnessed revolutionary advances in de novo molecular generation, which aims to generate molecules with desired properties from scratch. Artificial intelligence, which has been widely applied in this field, not only significantly reduces the chemical search space but also lowers time consumption to a large extent.

VAE-based VAE has been attractive in molecular generation in the virtue of its latent space is potentially operatable. CharVAE (Gómez-Bombarelli et al., 2018) first proposes to learn from molecular data in a data-driven manner and generate with a VAE model. By jointly training the task of predicting molecular properties from the hidden layers, this method explores the molecular space without manual prior knowledge. GVAE (Kusner et al., 2017) represents each data as a parse tree from a context-free grammar, and directly encodes to and decodes from these parse trees to ensure the validity of generated molecules. Inspired by syntax-directed translation in complier theory, SD-VAE (Dai et al., 2018) proposes to convert the offline syntax-directed translation check into on-the-fly generated guidance for ensuring both syntactical and semantical correctness. JT-VAE (Jin et al., 2018) first realize the direct generation of molecular graphs instead of linear SMILES (Simplified Molecular-Input Line-Entry System) strings. HierVAE (Jin et al., 2020) proposes a motif-based hierarchical encoder-decoder toward the large molecular graph generation.

GAN-based An alternative is to implement GAN in molecular generation. ORGAN (Guimaraes et al., 2017) adds expert-based rewards under the framework of WGAN (Arjovsky et al., 2017). ORGANIC (Sanchez-Lengeling et al., 2017) improves the above work for inverse design chemistry and implements the molecular generation towards specific properties. MolGAN (De Cao & Kipf, 2018) proposes GAN-based models to generate molecular graphs rather than SMILES. Motivated by cycle-consistent GAN (Zhu et al., 2017), Mol-CycleGAN (Maziarka et al., 2020) generates optimized compounds with high structural similarity to the original ones.

Flow-based Molecular generation with the normalizing flow is promising as its invertible mapping can reconstruct the data exactly. GraphNVP (Madhawa et al., 2019) defines a normalizing flow from a base distribution to the molecular graph structures. GRF (Honda et al., 2019) further explores invertible mappings by residual flows, and presents a way of keeping the entire flows invertible throughout the training and sampling process in molecular generation. GraphAF (Shi et al., 2019) combines the advantages of both autoregressive and flow-based approaches to iteratively generate molecules. MolFlow (Zang & Wang, 2020) proposes a variant of Glow (Kingma & Dhariwal, 2018) to generate atoms and bonds in a one-shot manner. MolGrow (Kuznetsov & Polykovskiy, 2021) constrains optimization of properties by using latent variables of the model, and recursively splits nodes to generate molecular structures.

Though these approaches have achieved significant performance, we recognize them as chemocentric molecular generation methods that lack biological connections. We aim

to bridge biological and chemical perspectives in molecular generation for practical drug discovery.

2.2. Drug-target Interaction

Drug-target interaction (DTI) has been extensively developed over the decades. Computational virtual screening methods like molecular docking (Trott & Olson, 2010) and molecular dynamics simulations (Salsbury Jr, 2010) have provided mechanistic insights in this field. Though these methods have inherently great interpretability, they suffer from heavy dependence on the available three-dimensional structure data with massive computational resources.

Recent progress in artificial intelligence has inspired researchers to utilize deep learning techniques in drug-target interaction prediction. DeepDTA (Öztürk et al., 2018) and DeepAffinity (Karimi et al., 2019) are representatives of deep-learning-based methods that take SMILES of drugs and primary sequences of proteins as input, from which neural networks are employed to predict affinities. InterpretableDTIP (Gao et al., 2018) predicts DTI directly from low-level representations and provides biological interpretation using a two-way attention mechanism. DeepRelations (Karimi et al., 2020) embeds protein sequences by hierarchical recurrent neural network and drug graphs by graph neural networks with joint attention between protein residues and compound atoms. DEEPScreen (Rifaioglu et al., 2020) learns features from structural representations for a large-scale DTI prediction. MONN (Li et al., 2020) predicts both pairwise non-covalent interactions and binding affinities between drugs and targets with extra supervision from the labels extracted from available high-quality three-dimensional structures.

Our proposed target-aware molecular generation builds on the recent advances in data-driven drug-target interaction prediction. We connect chemical molecular generation with biological drug-target interaction to promote the efficiency of computational drug discovery.

2.3. Conditional Molecular Generation

Generating molecules with the consideration of some external conditions is a promising field. There is plenty of work that conditioning on desired properties based on chemical perspective. CVAE (Gómez-Bombarelli et al., 2018) jointly trains VAE with a predictor that predicts properties from the latent representations of VAE. (Lim et al., 2018) proposes applying conditional VAE to generate drug-like molecules satisfying five properties at the same time: MW (molecular weight), LogP (partition coefficient), HBD (number of hydrogen bond donor), HBA (number of hydrogen acceptor), and TPSA (topological polar surface area). (Griffiths & Hernández-Lobato, 2020) employs constrained Bayesian optimization to control the latent space of VAE in order to find molecules that score highly under a specified objective function which is a weighted function of LogP and QED (quantitative estimate of drug-likeness). CogMol (Chenthamarakshan et al., 2020) and CLaSS (Das et al., 2021) pretrain the latent space with SMILES and train property classifiers from the latent representations. They sample variables from the latent space that satisfy high scores from property classifiers to generate molecules.

Though recent molecular generation methods (Jin et al., 2018; Madhawa et al., 2019; Zang & Wang, 2020; Luo et al., 2021) also present property optimization experiments, they still, like the above methods, barely take account of drug-target interaction. (Méndez-Lucio et al., 2020) proposes stacks of conditional GAN to generate hit-like molecules from gene expression signature. While this work focuses on drug-gene relationships, we recognize that drug-protein is a more typical case.

3. Background and Preliminaries

3.1. Problem Statement

Let $\mathcal{T} = \{T_i\}_{i=1}^{t}$ be a set of targets, and there exists a set of drugs $\mathcal{M}_{T_i} = \{M_j^{(T_i)}\}_{j=1}^{d_i}$ that bind to each target T_i. $S(T, M)$ is defined as a function measuring the interaction between target T and drug M. The target-aware molecular generation aims to learning a generation model $p_\theta(\cdot|T_i)$ from each drug-target pair $(M_j^{(T_i)}, T_i)$ so as to maximize $\mathbb{E}_{M|T_i \sim p_\theta}[S(M, T_i)]$.

3.2. The Flow Framework

A flow model is a sequence of parametric invertible mapping $f_\Theta = f_Q \circ ... \circ f_1$ from the data point $x \in \mathbb{R}^D$ to the latent variable $z \in \mathbb{R}^D$, where $x \sim P_X(x), z \sim P_Z(z)$. The latent distribution P_Z is usually predefined as a simple distribution, e.g., a normal distribution. The complex data in the original space is modelled by using the change-of-variable formula:

$$P_X(x) = P_Z(z)\left|\det\frac{\partial Z}{\partial X}\right|, \tag{1}$$

and its log-likelihood:

$$\begin{aligned}\log P_X(x) &= \log P_Z(z) + \log\left|\det\frac{\partial Z}{\partial X}\right| \\ &= \log P_Z(z) + \sum_{q=1}^{Q}\log\left|\det\frac{\partial f_q(z^{(q-1)})}{\partial z^{(q-1)}}\right|,\end{aligned} \tag{2}$$

where $z^{(q)} = f_q(z^{(q-1)})$, and we represent the input $z^{(0)}$ by using z for notation simplicity.

As the calculation of the Jacobian determinant for f_Θ is expensive for arbitrary functions, NICE (Dinh et al., 2015)

and RealNVP (Dinh et al., 2017) develop an affine coupling transformation $z = f_\Theta(x)$ with expressive structures and efficient computation of the Jacobian determinant.

For given D-dimensional input x and $d < D$, the output y of an affine coupling transformation is defined as:

$$\begin{aligned} y_{1:d} &= x_{1:d} \\ y_{d+1:D} &= x_{d+1:D} \odot \exp(S_\Theta(x_{1:d})) + T_\Theta(x_{1:d}), \end{aligned} \tag{3}$$

where $S_\Theta : \mathbb{R}^d \to \mathbb{R}^{D-d}$ and $T_\Theta : \mathbb{R}^d \to \mathbb{R}^{D-d}$ stand for scale function and transformation function. For the sake of the numerical stability of cascading multiple flow layers, we follow Moflow (Zang & Wang, 2020) to replace the exponential function for the S_Θ with the Sigmoid function:

$$\begin{aligned} y_{1:d} &= x_{1:d} \\ y_{d+1:D} &= x_{d+1:D} \odot \mathrm{Sigmoid}(S_\Theta(x_{1:d})) + T_\Theta(x_{1:d}), \end{aligned} \tag{4}$$

and the invertibility is guaranteed by:

$$\begin{aligned} x_{1:d} &= y_{1:d} \\ x_{d+1:D} &= (y_{d+1:D} - T_\Theta(y_{1:d}))/\mathrm{Sigmoid}(S_\Theta(y_{1:d})). \end{aligned} \tag{5}$$

The logarithmic Jacobian determinant is:

$$\begin{aligned} \log \left|\det \frac{\partial y}{\partial x}\right| &= \log \left|\det\left(\begin{bmatrix} \mathbb{I} & 0 \\ \frac{\partial y_{d+1:D}}{\partial x_{1:d}} & \mathrm{Sigmoid}(S_\Theta(x_{1:d})) \end{bmatrix}\right)\right| \\ &= \log \mathrm{Sigmoid}(S_\Theta(x_{1:d})). \end{aligned} \tag{6}$$

To further improve the invertible mapping with more expressive structures and high numerical stability, Glow (Kingma & Dhariwal, 2018) proposes using invertible 1×1 convolution to learn an optimal partition and actnorm layer to normalize dimensions in each channel over a batch by an affine transformation. Invertible 1×1 convolution is initialized as a random rotation matrix with zero log-determinant and works as a generalization of a permutation of channels. Act norm initializes the scale and the bias such that the post-actnorm activations per-channel have zero mean and unit variance and learns these parameters in training instead of using batch statistics as batch normalization does.

3.3. Flow on the Molecular Graph

Prior works on flow-based molecular graph generation are well developed. Inspired by the graph normalizing flows of GRevNets (Liu et al., 2019a), GraphNVP (Madhawa et al., 2019) proposes to generate atom features conditioned on the pre-generated adjacency tensors, which is then followed by other one-shot flow-based molecular graph generation approaches, e.g., GRF (Honda et al., 2019) and Moflow (Zang & Wang, 2020). Our proposed SiamFlow follows this manner, that is, firstly transforms the bonds B of molecules to the latent variables Z_B with Glow (Kingma & Dhariwal, 2018), and then transforms the atom features A given B into the conditional latent variable $Z_{A|B}$ with a graph conditional flow.

Let N, K, C be the number of nodes, node types, and edge types, respectively. A molecular graph $G = (A, B)$ is defined by an atom matrix $A \in \{0,1\}^{N \times K}$ and a bond tensor $B \in \{0,1\}^{C \times N \times N}$, which correspond to nodes and edges in the vanilla graph. $A[i,k] = 1$ represents the i-th atom i has atom type k, and $B[c,i,j] = 1$ represents there is a bond with type c between the i-th atom and j-th atom.

Flow-based molecular graph generation methods decompose the generative model into two parts:

$$P(G) = P((A,B)) \approx P(A|B;\theta_{A|B})P(B;\theta_B), \tag{7}$$

where θ_B is learned by the bond flow model h_B, and $\theta_{A|B}$ is learned by the atom flow model $h_{A|B}$ conditioned on the bond tensor B.

With the strengths of the flow, the optimal parameters $\theta^*_{A|B}$ and θ^*_B maximize the exact likelihood estimation:

$$\arg \max_{\theta_{A|B},\theta_B} \mathbb{E}_{(A,B)\sim P_G}[\log P(A|B;\theta_{A|B}) + \log P(B;\theta_B)] \tag{8}$$

Our work follows the one-shot molecular graph generation manner (Madhawa et al., 2019; Honda et al., 2019; Zang & Wang, 2020) that employs Glow (Kingma & Dhariwal, 2018) as the bond flow model h_B and graph conditional flow as the atom flow model $h_{A|B}$.

4. SiamFlow

4.1. Overview

While current flow-based molecular graph generation methods (Madhawa et al., 2019; Honda et al., 2019; Shi et al., 2019; Zang & Wang, 2020; Kuznetsov & Polykovskiy, 2021; Lippe & Gavves, 2020; Luo et al., 2021) learn from drug-like datasets and generate without the invention of targets, our proposed SiamFlow aims to serve as a conditional flow toward molecular graph generation. Though the conditional flow has been well developed in computer vision (Liu et al., 2019b; Kondo et al., 2019; Abdelhamed et al., 2019; Kumar et al., 2019; Sun et al., 2019; Pumarola et al., 2020; Liang et al., 2021; Yang et al., 2021), there are limited works that can fit graph generation, especially when it comes to the molecular graph.

In this section, we introduce SiamFlow, a novel molecular graph generative model conditioned on specific targets. As shown in Figure 2, SiamFlow learns the distribution of sequence embedding instead of the isotropic Gaussian distribution like other flow-based methods.

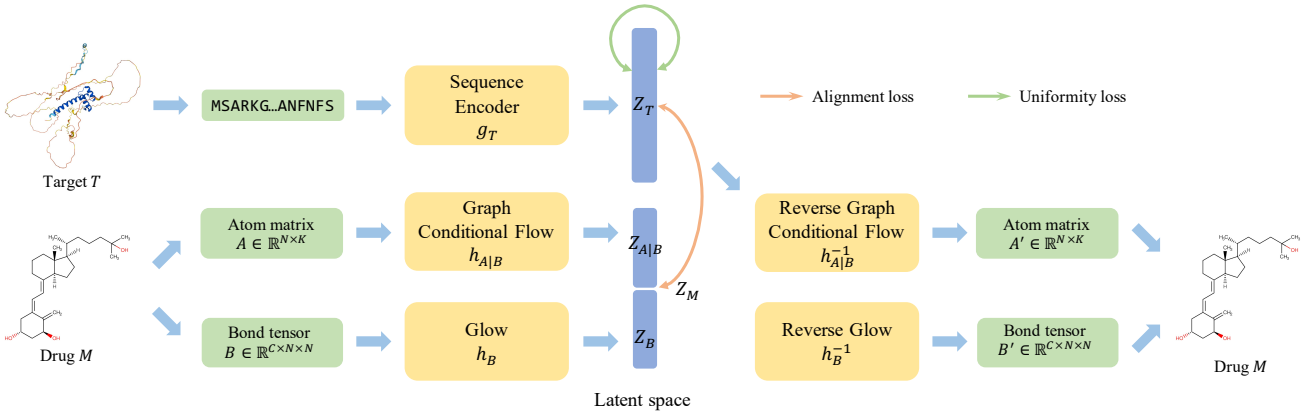

Figure 2. The framework of our proposed SiamFlow. In the training phase, the target sequence embedding Z_T aligns with the drug graph embedding Z_M, while a uniformity regularization term forces its distribution as a spherical uniform distribution. In the generation phase, the target sequence embedding Z_T is fed into reverse flows to generate the desired drug.

4.2. Alignment Loss

Given a pair of target T and drug M, we decompose the drug M into an atom matrix $A \in \mathbb{R}^{N\times K}$ and a bond tensor $B \in \mathbb{R}^{C\times N\times N}$. The sequence encoder g_T can be arbitrary mapping that maps the target sequence T into the sequence embedding $Z_T \in \mathbb{R}^D$. The flow model contains a glow $h_B : \mathbb{R}^{C\times N\times N} \to \mathbb{R}^{\frac{D}{2}}$ and a graph conditional flow $h_{A|B} : \mathbb{R}^{N\times K} \to \mathbb{R}^{\frac{D}{2}}$. The drug graph embedding Z_M is the concatenation of $Z_{A|B}$ and Z_B.

Instead of directly learning the isotropic Gaussian distribution, we impose alignment loss between the target sequence embedding Z_T and the drug graph embedding Z_M so that Z_T can be used as the input of the generation process. Thus, the generated atom matrix and the bond tensor are:

$$A' = h_{A|B}^{-1}(Z_T[1:\frac{D}{2}]),\ B' = h_B^{-1}(Z_T[\frac{D}{2}:D]). \quad (9)$$

While traditional flow-based models assume the latent variables follow the Gaussian distribution, SiamFlow forces the flow model to learn the distribution of the condition information instead of a predefined distribution. We define the alignment loss $\mathcal{L}_{align}$ as:

$$\begin{aligned}\mathcal{L}_{align} &:= \mathbb{E}_{(T,M)\sim P_{\text{data}}}||Z_T - Z_M||_2 \\ &= \mathbb{E}_{(T,M)\sim P_{\text{data}}}||Z_T - [Z_{A|B}, Z_B]||_2\end{aligned} \quad (10)$$

where $[Z_{A|B}, Z_B]$ denotes the concatenation of the atom embedding $Z_{A|B}$ and the bond embedding Z_B, and the pair of protein target T and molecular drug M is sampled from the data P_{data}.

The alignment loss bridges the connections between the target sequence embedding Z_T and the drug graph embedding Z_M in the latent space, but there are still challenges that will be revealed in Sec. 4.3 and Sec. 4.4.

4.3. Uniformity Loss

Simply aligning the target sequence embedding Z_T and the drug graph embedding Z_M is not enough. There still remains three challenges: (1) the distribution of Z_T is uncertain, so that the alignment learning may be difficult to converge; (2) sampling from an unknown distribution is indefinite in the generation process; (3) the alignment loss alone admits collapsed solutions, e.g., outputting the same representation for all targets.

To overcome the above issues, we design an objective to force the target sequence embedding Z_T to follow a specific distribution, in our case the uniform distribution on the unit hypersphere (Saff & Kuijlaars, 1997; Kuijlaars & Saff, 1998; Hardin et al., 2004; Liu et al., 2018; Borodachov et al., 2019; Wang & Isola, 2020). We recognize angles of embeddings are the critical element that preserves the most abundant and discriminative information. By fitting the hyperspherical uniform distribution, the projections of target sequence embeddings on the hypersphere are kept as far away from each other as possible; thus, discriminations are imposed. Specifically, we project the target sequence embedding Z_T into a unit hypersphere $\mathbb{S}^{D-1}$ by L2 normalization and require the embeddings uniformly distributed on this hypersphere, as shown in Figure 3.

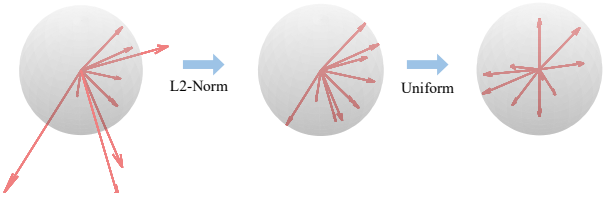

Figure 3. The schematic diagram of the uniformity loss.

The uniform hypersphere distribution can be formulated as a minimizing pairwise potential energy problem (Liu et al., 2018; Borodachov et al., 2019; Wang & Isola, 2020) while higher energy implies less discriminations. Let $\widehat{Z}_T = \frac{Z_T}{\|Z_T\|} \in \mathcal{C}$, and $\mathcal{C}$ is a finite subset of the unit hypersphere $\mathbb{S}^{D-1} \in \mathbb{R}^D$. We define the f-potential energy (Cohn & Kumar, 2007) of $\mathcal{C}$ to be:

$$\sum_{\widehat{Z}_T^{(x)}, \widehat{Z}_T^{(y)} \in \mathcal{C}, x \neq y} f(|\widehat{Z}_T^{(x)} - \widehat{Z}_T^{(y)}|^2). \tag{11}$$

where $\widehat{Z}_T^{(x)}$ and $\widehat{Z}_T^{(y)}$ denote different normalized sequence embedding with index x and y.

Definition. (Universally optimal (Cohn & Kumar, 2007)). A finite subset $\mathcal{C} \subset \mathbb{S}^{D-1}$ is *universally optimal* if it (weakly) minimizes potential energy among all configurations of $|\mathcal{C}|$ points on $\mathbb{S}^{D-1}$ for each completely monotonic potential function.

In SiamFlow, we consider the Gaussian function kernel $G_t(x, y) : \mathbb{S}^{D-1} \times \mathbb{S}^{D-1} \to \mathbb{R}$ as the potential function f, which is defined as:

$$G_t(x, y) = e^{-t|x-y|^2}. \tag{12}$$

This kernel function is closely related to the universally optimal configuration, and distributions of points convergence weak* to the uniform distribution by minimizing the expected pairwise potential.

Theorem. (Strictly positive definite kernels on $\mathbb{S}^D$ (Borodachov et al., 2019)). Consider kernel $K_f : \mathbb{S}^D \times \mathbb{S}^D \to (-\infty, +\infty]$ of the form $K_f(x, y) := f(|x - y|^2)$, if K_f is *strictly positive definite* on $\mathbb{S}^D \times \mathbb{S}^D$ and the energy $I_{K_f}[\sigma_D]$ is finite, then σ_D is the unique measure on Borel subsets of $\mathbb{S}^D$ in the solution of $\min_{\mu \in \mathcal{M}(\mathbb{S}^D)} I_{K_f}(\mu)$, and the normalized counting measure associated with any K_f-energy minimizing sequence of point configurations on $\mathbb{S}^D$ converges weak* to σ_D.

This theorem reveals the connections between strictly positive definite kernels and the energy minimizing problem. The Gaussian function is strictly positive definite on $\mathbb{S}^D \times \mathbb{S}^D$, thus well tied with the uniform distribution on the unit hypersphere.

Proposition 1. (Strictly positive definite of the Gaussian function) For any $t > 0$, the Gaussian function kernel $G_t(x, y)$ is strictly positive definte on $\mathbb{S}^D \times \mathbb{S}^D$.

Proof. See appendix.

Though Riesz s-kernels $R_s(x, y) := |x - y|^{-s}$ are commonly used as potential functions, we argue that the Gaussian function is expressive because it maps distances to infinite dimensions like radial basis functions, benefiting from the Taylor expansion of exponential functions. Moreover, the Gaussian function is a general case of Riesz s-kernels and can represent Riesz s-kernels by:

$$R_s(x, y) = \frac{1}{\Gamma(s/2)} \int_0^\infty G_t(x, y) t^{s/2-1} dt. \tag{13}$$

where $\Gamma(s/2) = \int_0^\infty e^{-t} t^{s/2-1}$ for $s > 0$.

As the Gaussian functionk kernel is an ideal choice of potential functions, we define the uniformity loss as the logarithm of the pairwise Gaussian potential's expectation:

$$\mathcal{L}_{unif} := \log \mathbb{E}_{(T^{(x)}, T^{(y)}) \sim P_{\mathcal{T}}} [G_t(\widehat{Z}_T^{(x)}, \widehat{Z}_T^{(y)})], \tag{14}$$

where $T^{(x)}$ and $T^{(y)}$ are two different targets sampled from the target data $P_{\mathcal{T}}$.

4.4. One Target to Many Drugs

Implementing the alignment loss and the uniformity loss above, the flow model can already generate validated molecular drugs conditioned on specific targets. However, there are multiple affinable drugs for a single target in most cases. To deal with this *one-to-many* problem, we reformulate learning target embeddings into learning spaces of target embeddings in the latent space, as shown in Figure 4.

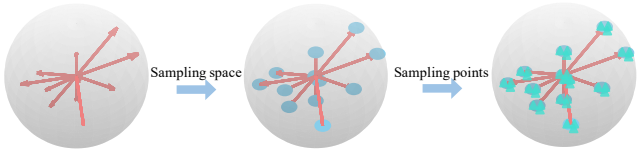

Figure 4. The schematic diagram of the one-to-many strategy. The blue circles denote the possible spaces around the target sequence embeddings, and the green triangles denote the instances sampled from the possible spaces.

As the target embeddings have been pushed by the uniformity loss to stay as far away as possible on the hypersphere, they preserve abundant and discriminative information to a large extent. We design an adaptive space learning strategy that holds the discriminative angle information with a limited scope. For a set of target sequence embeddings $\mathcal{Z}_{\mathcal{T}} = \{Z_T^{(0)}, ..., Z_T^{(L)}\}$, we first calculate their standard deviation by:

$$\sigma(\mathcal{Z}_{\mathcal{T}}) = \sqrt{\frac{1}{L} \sum_{i=1}^{L} (Z_T^{(i)} - \mu(\mathcal{Z}_{\mathcal{T}}))}, \tag{15}$$

where $\mu(\mathcal{Z}_{\mathcal{T}}) = \frac{1}{L} \sum_{i=1}^{L} Z_T^{(i)}$ is the mean of the set $\mathcal{Z}_{\mathcal{T}}$. Then, we define a space for each target sequence embedding:

$$\Omega(Z_T) = \{Z_T + Z_T' | Z_T' \in \mathcal{N}(0, \lambda\sigma^2(\mathcal{Z}_{\mathcal{T}}))\}, \tag{16}$$

where λ is the hyperparameter that controls the scale of the space and is empirically set as 0.1.

Note that we define the space on Z_T instead of the normalized $\widehat{Z}_T$, as normalized embeddings lose the length information to the extent that the available space is limited.

Thus, we modify the alignement loss as:

$$\mathcal{L}_{align} = \mathbb{E}_{(T,M)\sim P_{\text{data}}} |\Omega(Z_T) - Z_M|. \tag{17}$$

In the generation process, sampling from the same space is permissible to generate desired drugs.

In summary, the objective is a linear combination of the modified alignment loss and uniform loss:

$$\mathcal{L}_{total} = \mathcal{L}_{align} + \mathcal{L}_{unif} \tag{18}$$

5. Experiments

Baselines Since we present a novel generative approach conditioned on targets, we primarily compare our approach to other conditional generative models, i.e., conditional VAE (Sohn et al., 2015). Furthermore, an attention-based Seq2seq (Sutskever et al., 2014; Vaswani et al., 2017) neural translation model between the target protein sequence and drug SMILES is considered a straightforward solution in our setting. An explainable substructure partition fingerprint (Huang et al., 2019) is employed for sequential drug SMILES and protein sequences. We unify these baselines and our proposed SiamFlow into a unified framework for fair comparisons.

Datasets To evaluate the ability of our proposed SiamFlow, we collect a dataset based on four drug-target interaction datasets, including BIOSNAP (Zitnik et al., 2018), BindingDB (Liu et al., 2007), DAVIS (Davis et al., 2011), and DrugBank (Wishart et al., 2018). We remove all the negative samples in the original datasets, and only keep the positive samples. Our dataset contains 24,669 unique drug-target pairs with 10,539 molecular drugs and 2,766 proteins. The maximum number of atoms in a molecular drug is 100 while 11 types of common atoms are considered. We define the dataloader to ensure zero overlap protein in the training, validation, and test set.

Metrics To comprehensively evaluate the conditional generative models in terms of target-aware molecular generation, we design metrics from two perspectives: (1) Generative metrics. Following the common molecular generation settings, we apply metrics including: **validity** which is the percentage of chemically valid molecules in all the generated molecules, **uniqueness** which is the percentage of unique valid molecules in all the generated molecules, **novelty** which is the percentage of generated valid molecules which are not in the training dataset. (2) Chemical metrics. We evaluate the similarities between the generated drugs and the nearest drugs in the training set including: **Tanimoto similarity** which is calculated based on hashed binary features, **Fraggle similarity** which focus on the fragment-level similarity, **MACCS similarity** which employs 166-bit 2D structure fingerprints.

Empirical Running Time We implement our proposed method SiamFlow and the other two baselines Seq2seq, CVAE by Pytorch-1.8.1 framework. We train them with Adam optimizer with learning rate 0.001, batch size 16, and 100 epochs on a single NVIDIA Tesla V100 GPU. To evaluate the validity and chemical similarities, we employ the cheminformatics toolkit RDKit in the assessment phase. Our SiamFlow completes the training process of 100 epochs in an average of 1.06 hours (38 seconds/ epoch), while CVAE and Seq2seq take an average of 1.14 hours (41 seconds/ epoch) and 8.33 hours (5 minutes/ epoch) respectively.

5.1. Target-aware Molecular Graph Generation

We conduct experiments on molecular drug generation with specific targets for comparisons. For each experiment, we repeat three trials with different random seeds and report the mean and standard deviation.

Table 1 shows the results on generative metrics of our SiamFlow model in comparison to the baselines. Our proposed SiamFlow inherits the strengths of the flow and far surpasses other baselines in generative metrics. It can be seen that Seq2seq suffers from low validity, uniqueness, and novelty, which indicates Seq2seq's generation relies on its memorization. CVAE has higher uniqueness and novelty than Seq2seq though its validity is even lower. Besides, the standard deviations of metrics on CVAE are relatively high, suggesting it is volatile to train. Among them, SiamFlow obtains superior performance with the least volatility.

Table 1. Evaluation results on generative metrics of SiamFlow v.s. baselines; high is better for all three metrics here.

Method	% Validity	% Uniqueness	% Novelty
Seq2seq	16.08±4.14	13.87±1.74	14.89±11.41
CVAE	12.54±7.56	72.30±20.33	99.72±0.39
SiamFlow	**100.00**±0.00	**99.61**±0.16	**100.00**±0.00

In addition to generative metrics, we also report chemical metrics in Table 2. The generated molecular drugs are expected to have a chemical structure similar to the ground-truth drugs in order to have a high binding affinity to the target. SiamFlow is consistently better than other baselines in both the Tanimoto and Fraggle similarity while obtaining relatively lower MACCS similarity than Seq2seq. Consid-

ering that MACCS measures the similarity of encodings of molecules, the sequence partition rules of Seq2seq may help it. Thus, we pay more attention to the Tanimoto and Fraggle similarity because they are structure-centric metrics.

Table 2. Evaluation results on chemical metrics of SiamFlow v.s. baselines; high is better for all three metrics here.

Method	% Tanimoto	% Fraggle	% MACCS
Seq2seq	26.27±9.91	25.84±7.27	**37.98**±7.70
CVAE	7.76±6.61	12.31±5.81	16.42±7.17
SiamFlow	**48.55**±0.97	**34.41**±0.35	29.30±1.07

We visualize the distribution of the Tanimoto similarity and the Fraggle similarity evaluated on these methods in Figure 5. SiamFlow consistently outperforms other methods and generates desirable molecular drugs. The examples of generated drugs are shown in Figure 6.

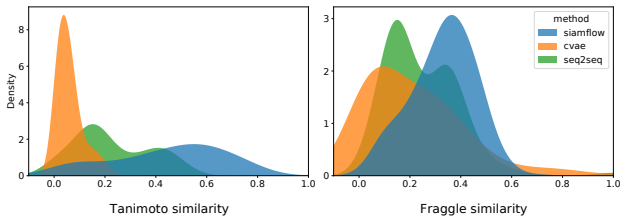

Figure 5. The distribution of generative metrics evaluated on SiamFlow and other baselines.

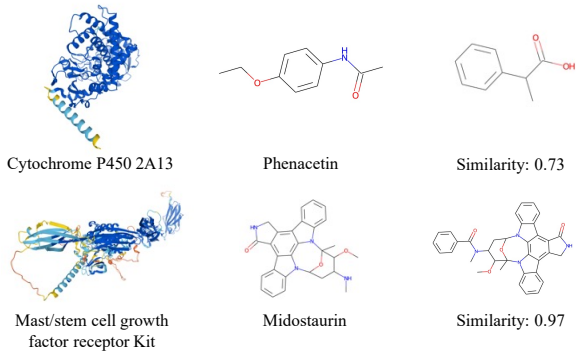

Figure 6. Examples of the generated drugs.

5.2. Ablation Study

We conduct the ablation study and report the results in Table 3 and Table 4. It can be seen from Table 3 that simply aligning the target sequence embedding and drug graph embedding will result in extremely low uniqueness. Our one-to-many strategy enriches the latent space so that one target can map to different drugs. The absence of $\mathcal{L}_{unif}$ does not harm the generative metrics because it only constrains the distribution of target sequence embeddings but has a limited impact on the generation process.

Table 3. Ablation results on generative metrics.

Method	% Validity	% Uniqueness	% Novelty
SiamFlow	100.00	99.39	100.00
w/o one-to-many	100.00	12.55	100.00
w/o $\mathcal{L}_{unif}$	100.00	100.00	100.00

Table 4 demonstrates the chemical metrics are well without the one-to-many strategy. If we generate only one drug for a particular target, the nearest drug similarity degrades to a special case, i.e., comparing the generated drug with its corresponding one in the training set. Moreover, removing $\mathcal{L}_{unif}$ severely impairs the chemical performance, suggesting the uniformity loss promotes the expressive abilities of target sequence embeddings.

Table 4. Ablation results on chemical metrics.

Method	% Tanimoto	% Fraggle	% MACCS
SiamFlow	49.43	34.62	29.55
w/o one-to-many	48.83	34.93	31.23
w/o $\mathcal{L}_{unif}$	18.49	15.70	17.91

6. Conclusion and Discussion

In this paper, we explore target-aware molecular graph generation, which differs in generating drugs conditioned on specific targets, while existing methods focus on developing similar drugs based on drug-like datasets. Target-aware molecular generation combines drug-like molecular generation with target-specific screening and thus simplifies the drug-target interaction step.

To explore this problem in-depth, we compile a benchmark dataset from three public datasets and build baselines in a unified framework. Furthermore, we take advantage of recent progress on flow-based molecular graph generation methods and propose SiamFlow towards target-aware molecular generation. With the alignment and uniform loss, the proposed method can effectively generate molecular drugs conditioned on protein targets. Moreover, we deal with one target to many drugs by aligning the embedding space instead of a single embedding.

Extensive experiments and analyses demonstrate that SiamFlow is a promising solution towards target-aware molecular generation. A worthwhile avenue for future work is introducing protein structure information for decent target sequence representations.

References

Abdelhamed, A., Brubaker, M. A., and Brown, M. S. Noise flow: Noise modeling with conditional normalizing flows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3165–3173, 2019.

Arjovsky, M., Chintala, S., and Bottou, L. Wasserstein generative adversarial networks. In *International conference on machine learning*, pp. 214–223. PMLR, 2017.

Bajorath, J. Integration of virtual and high-throughput screening. *Nature Reviews Drug Discovery*, 1(11):882–894, 2002.

Bandholtz, S., Wichard, J., Kühne, R., and Grötzinger, C. Molecular evolution of a peptide gpcr ligand driven by artificial neural networks. *PloS one*, 7(5):e36948, 2012.

Borodachov, S. V., Hardin, D. P., and Saff, E. B. *Discrete energy on rectifiable sets*. Springer, 2019.

Button, A., Merk, D., Hiss, J. A., and Schneider, G. Automated de novo molecular design by hybrid machine intelligence and rule-driven chemical synthesis. *Nature machine intelligence*, 1(7):307–315, 2019.

Cheng, Y., Gong, Y., Liu, Y., Song, B., and Zou, Q. Molecular design in drug discovery: a comprehensive review of deep generative models. *Briefings in Bioinformatics*, 22 (6):bbab344, 2021.

Chenthamarakshan, V., Das, P., Hoffman, S. C., Strobelt, H., Padhi, I., Lim, K. W., Hoover, B., Manica, M., Born, J., Laino, T., et al. Cogmol: target-specific and selective drug design for covid-19 using deep generative models. *Advances in Neural Information Processing Systems*, 33, 2020.

Cohn, H. and Kumar, A. Universally optimal distribution of points on spheres. *Journal of the American Mathematical Society*, 20(1):99–148, 2007.

Dai, H., Tian, Y., Dai, B., Skiena, S., and Song, L. Syntax-directed variational autoencoder for molecule generation. In *Proceedings of the international conference on learning representations*, 2018.

Das, P., Sercu, T., Wadhawan, K., Padhi, I., Gehrmann, S., Cipcigan, F., Chenthamarakshan, V., Strobelt, H., Dos Santos, C., Chen, P.-Y., et al. Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations. *Nature Biomedical Engineering*, 5(6):613–623, 2021.

Davis, I. M., Hunt, P. J., Herrgard, S., Ciceri, P., Wodicka, M. L., Pallares, G., Hocker, M., Treiber, K. D., Zarrinkar, P. P., and Treiber, K. D. Comprehensive analysis of kinase inhibitor selectivity. *NATURE BIOTECHNOLOGY*, pp. 1046.0–U124, 2011.

De Cao, N. and Kipf, T. MolGAN: An implicit generative model for small molecular graphs. *ICML 2018 workshop on Theoretical Foundations and Applications of Deep Generative Models*, 2018.

Dinh, L., Krueger, D., and Bengio, Y. Nice: Non-linear independent components estimation. In *International Conference on Learning Representations Workshop*, 2015.

Dinh, L., Sohl-Dickstein, J., and Bengio, S. Density estimation using real nvp. In *International Conference on Learning Representations*, 2017.

Gao, K. Y., Fokoue, A., Luo, H., Iyengar, A., Dey, S., and Zhang, P. Interpretable drug target prediction using deep neural representation. In *IJCAI*, volume 2018, pp. 3371–3377, 2018.

Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., and Aspuru-Guzik, A. Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2):268–276, 2018.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2014.

Griffiths, R.-R. and Hernández-Lobato, J. M. Constrained bayesian optimization for automatic chemical design using variational autoencoders. *Chemical science*, 11(2): 577–586, 2020.

Guimaraes, G. L., Sanchez-Lengeling, B., Outeiral, C., Farias, P. L. C., and Aspuru-Guzik, A. Objective-reinforced generative adversarial networks (organ) for sequence generation models. *arXiv preprint arXiv:1705.10843*, 2017.

Hardin, D. P., Saff, E. B., et al. Discretizing manifolds via minimum energy points. *Notices of the AMS*, 51(10): 1186–1194, 2004.

Hert, J., Irwin, J. J., Laggner, C., Keiser, M. J., and Shoichet, B. K. Quantifying biogenic bias in screening libraries. *Nature chemical biology*, 5(7):479–483, 2009.

Honda, S., Akita, H., Ishiguro, K., Nakanishi, T., and Oono, K. Graph residual flow for molecular graph generation. *arXiv preprint arXiv:1909.13521*, 2019.

Huang, K., Xiao, C., Glass, L., and Sun, J. Explainable substructure partition fingerprint for protein, drug, and more. In *NeurIPS Learning Meaningful Representation of Life Workshop*, 2019.

Jin, W., Barzilay, R., and Jaakkola, T. Junction tree variational autoencoder for molecular graph generation. In *International conference on machine learning*, pp. 2323–2332. PMLR, 2018.

Jin, W., Barzilay, R., and Jaakkola, T. Hierarchical generation of molecular graphs using structural motifs. In *International Conference on Machine Learning*, pp. 4839–4848. PMLR, 2020.

Karimi, M., Wu, D., Wang, Z., and Shen, Y. Deepaffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks. *Bioinformatics*, 35(18):3329–3338, 2019.

Karimi, M., Wu, D., Wang, Z., and Shen, Y. Explainable deep relational networks for predicting compound–protein affinities and contacts. *Journal of Chemical Information and Modeling*, 61(1):46–66, 2020.

Kingma, D. P. and Dhariwal, P. Glow: Generative flow with invertible 1x1 convolutions. *Advances in Neural Information Processing Systems*, 31:10215–10224, 2018.

Kingma, D. P. and Welling, M. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014.

Kondo, R., Kawano, K., Koide, S., and Kutsuna, T. Flow-based image-to-image translation with feature disentanglement. *Advances in Neural Information Processing Systems*, 32:4168–4178, 2019.

Kuijlaars, A. and Saff, E. Asymptotics for minimal discrete energy on the sphere. *Transactions of the American Mathematical Society*, 350(2):523–538, 1998.

Kumar, M., Babaeizadeh, M., Erhan, D., Finn, C., Levine, S., Dinh, L., and Kingma, D. Videoflow: A conditional flow-based model for stochastic video generation. In *International Conference on Learning Representations*, 2019.

Kusner, M. J., Paige, B., and Hernández-Lobato, J. M. Grammar variational autoencoder. In *International Conference on Machine Learning*, pp. 1945–1954. PMLR, 2017.

Kuznetsov, M. and Polykovskiy, D. Molgrow: A graph normalizing flow for hierarchical molecular generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 8226–8234, 2021.

Li, S., Wan, F., Shu, H., Jiang, T., Zhao, D., and Zeng, J. Monn: a multi-objective neural network for predicting compound-protein interactions and affinities. *Cell Systems*, 10(4):308–322, 2020.

Liang, J., Lugmayr, A., Zhang, K., Danelljan, M., Van Gool, L., and Timofte, R. Hierarchical conditional flow: A unified framework for image super-resolution and image rescaling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4076–4085, 2021.

Lim, J., Ryu, S., Kim, J. W., and Kim, W. Y. Molecular generative model based on conditional variational autoencoder for de novo molecular design. *Journal of cheminformatics*, 10(1):1–9, 2018.

Lippe, P. and Gavves, E. Categorical normalizing flows via continuous transformations. In *International Conference on Learning Representations*, 2020.

Liu, J., Kumar, A., Ba, J., Kiros, J., and Swersky, K. Graph normalizing flows. *Advances in Neural Information Processing Systems*, 32:13578–13588, 2019a.

Liu, R., Liu, Y., Gong, X., Wang, X., and Li, H. Conditional adversarial generative flow for controllable image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7992–8001, 2019b.

Liu, T., Lin, Y., Wen, X., Jorissen, N. R., and Gilson, K. M. Bindingdb: A web-accessible database of experimentally determined protein-ligand binding affinities. *NUCLEIC ACIDS RESEARCH*, pp. D198–D201, 2007.

Liu, W., Lin, R., Liu, Z., Liu, L., Yu, Z., Dai, B., and Song, L. Learning towards minimum hyperspherical energy. *Advances in Neural Information Processing Systems*, 31: 6222–6233, 2018.

Luo, Y., Yan, K., and Ji, S. Graphdf: A discrete flow model for molecular graph generation. *arXiv preprint arXiv:2102.01189*, 2021.

Madhawa, K., Ishiguro, K., Nakago, K., and Abe, M. Graphnvp: An invertible flow model for generating molecular graphs. *arXiv preprint arXiv:1905.11600*, 2019.

Maziarka, Ł., Pocha, A., Kaczmarczyk, J., Rataj, K., Danel, T., and Warchoł, M. Mol-cyclegan: a generative model for molecular optimization. *Journal of Cheminformatics*, 12(1):1–18, 2020.

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):1–10, 2020.

Öztürk, H., Özgür, A., and Ozkirimli, E. Deepdta: deep drug–target binding affinity prediction. *Bioinformatics*, 34(17):i821–i829, 2018.

Paricharak, S., Méndez-Lucio, O., Chavan Ravindranath, A., Bender, A., IJzerman, A. P., and van Westen, G. J. Data-driven approaches used for compound library design, hit triage and bioactivity modeling in high-throughput screening. *Briefings in bioinformatics*, 19(2):277–285, 2018.

Phatak, S. S., Stephan, C. C., and Cavasotto, C. N. High-throughput and in silico screenings in drug discovery. *Expert opinion on drug discovery*, 4(9):947–959, 2009.

Polishchuk, P. G., Madzhidov, T. I., and Varnek, A. Estimation of the size of drug-like chemical space based on gdb-17 data. *Journal of computer-aided molecular design*, 27(8):675–679, 2013.

Popova, M., Shvets, M., Oliva, J., and Isayev, O. Molecularrnn: Generating realistic molecular graphs with optimized properties. *arXiv preprint arXiv:1905.13372*, 2019.

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):1–13, 2019.

Pumarola, A., Popov, S., Moreno-Noguer, F., and Ferrari, V. C-flow: Conditional generative flow models for images and 3d point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7949–7958, 2020.

Rifaioglu, A. S., Nalbat, E., Atalay, V., Martin, M. J., Cetin-Atalay, R., and Doğan, T. Deepscreen: high performance drug–target interaction prediction with convolutional neural networks using 2-d structural compound representations. *Chemical science*, 11(9):2531–2557, 2020.

Saff, E. B. and Kuijlaars, A. B. Distributing many points on a sphere. *The mathematical intelligencer*, 19(1):5–11, 1997.

Salsbury Jr, F. R. Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. *Current opinion in pharmacology*, 10(6):738–744, 2010.

Sanchez-Lengeling, B., Outeiral, C., Guimaraes, G. L., and Aspuru-Guzik, A. Optimizing distributions over molecular space. an objective-reinforced generative adversarial network for inverse-design chemistry (organic). *ChemRxiv*, 2017. doi: 10.26434/chemrxiv.5309668.v3.

Schneider, G. Virtual screening: an endless staircase? *Nature Reviews Drug Discovery*, 9(4):273–276, 2010.

Schneider, G. *De novo molecular design*. John Wiley & Sons, 2013.

Schneider, P. and Schneider, G. De novo design at the edge of chaos: Miniperspective. *Journal of medicinal chemistry*, 59(9):4077–4086, 2016.

Shi, C., Xu, M., Zhu, Z., Zhang, W., Zhang, M., and Tang, J. Graphaf: a flow-based autoregressive model for molecular graph generation. In *International Conference on Learning Representations*, 2019.

Simonovsky, M. and Komodakis, N. Graphvae: Towards generation of small graphs using variational autoencoders. In *International conference on artificial neural networks*, pp. 412–422. Springer, 2018.

Sohn, K., Lee, H., and Yan, X. Learning structured output representation using deep conditional generative models. *Advances in neural information processing systems*, 28: 3483–3491, 2015.

Sun, H., Mehta, R., Zhou, H. H., Huang, Z., Johnson, S. C., Prabhakaran, V., and Singh, V. Dual-glow: Conditional flow-based generative model for modality transfer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10611–10620, 2019.

Sutskever, I., Vinyals, O., and Le, Q. V. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pp. 3104–3112, 2014.

Trott, O. and Olson, A. J. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2):455–461, 2010.

Van Oord, A., Kalchbrenner, N., and Kavukcuoglu, K. Pixel recurrent neural networks. In *International Conference on Machine Learning*, pp. 1747–1756. PMLR, 2016.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. In *Advances in neural information processing systems*, pp. 5998–6008, 2017.

Wang, T. and Isola, P. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *International Conference on Machine Learning*, pp. 9929–9939. PMLR, 2020.

Winter, R., Montanari, F., Steffen, A., Briem, H., Noé, F., and Clevert, D.-A. Efficient multi-objective molecular optimization in a continuous latent space. *Chemical science*, 10(34):8016–8024, 2019.

Wishart, S. D., Feunang, D. Y., Guo, C. A., Lo, J. E., Marcu, A., Grant, R. J., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D.,

Pon, A., Knox, C., and Wilson, M. Drugbank 5.0: A major update to the drugbank database for 2018. *NUCLEIC ACIDS RESEARCH*, pp. D1074–D1082, 2018.

Yang, K., Goldman, S., Jin, W., Lu, A. X., Barzilay, R., Jaakkola, T., and Uhler, C. Mol2image: Improved conditional flow models for molecule to image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6688–6698, 2021.

You, J., Liu, B., Ying, R., Pande, V., and Leskovec, J. Graph convolutional policy network for goal-directed molecular graph generation. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 6412–6422, 2018.

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.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networkss. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.

Zitnik, M., Sosic, R., Maheshwari, S., and Leskovec, J. Biosnap datasets: Stanford biomedical network dataset collection. 2018.

A. Proofs and Additional Theoretical Analysis

In this section, we present supplements to the uniformity loss with detailed definitions.

Recall that for the Euclidean metric in $\mathbb{R}^p$, the Gaussian function for $t > 0$ (see Equ. 12) is given by $G_t(x, y) = e^{-t|x-y|^2}$ (Borodachov et al., 2019).

The Fourier transform of G_t is:

$$\widehat{G}_t(\xi) = (\pi/t)^{p/2} e^{-\pi^2|\xi|^2/t} = (\pi/t)^{p/2} G_{\pi^2/t}(\xi), \quad (19)$$

where $\xi \in \mathbb{R}^p$. This Fourier transform formulation can be derived by writing the integral representing $\widehat{G}_t(\xi)$ as a product of one-dimensional Fourier transforms and using the formula for the Fourier transform of the univariate Gaussian function. Then, we denote the collection of all finite signed Borel measures on a Borel set $A \subset \mathbb{R}^p$ as $\mathcal{M}_{\text{sign}}(A)$.

Lemma 1. If $\mu \in \mathcal{M}_{\text{sign}}(\mathbb{R}^p)$ and $t > 0$, then

$$\int\int e^{-t|x-y|^2} d\mu(y) d\mu(x) = (\frac{\pi}{t})^{p/2} \int_{\mathbb{R}^p} e^{-\pi^2|\xi|^2/t} |\hat{\mu}(\xi)|^2 d\xi. \quad (20)$$

Consequently, for any $t > 0$, the Gaussian function kernel is strictly positive define on $A \times A$ for any compact set A.

Proof Using Fubini's theorem and the Fourier inversion formula:

$$\begin{aligned}
&(\frac{\pi}{t})^{p/2} \int_{\mathbb{R}^p} e^{-\pi^2|\xi|^2/t} |\hat{\mu}(\xi)|^2 d\xi \\
=&(\frac{\pi}{t})^{p/2} \int_{\mathbb{R}^p} e^{-\pi^2|\xi|^2/t} \hat{\mu}(\xi)\hat{\mu}(-\xi) d\xi \\
=&(\frac{\pi}{t})^{p/2} \int_{\mathbb{R}^p} e^{-\pi^2|\xi|^2/t} (\int e^{-2\pi i x\cdot\xi} d\mu(x) \int e^{2\pi i y\cdot\xi} d\mu(y)) \\
=&\int\int (\frac{\pi}{t})^{p/2} \widehat{G}_{\pi^2/t}(x-y) d\mu(x) d\mu(y) \\
=&\int\int e^{-t|x-y|^2} d\mu(y) d\mu(x),
\end{aligned} \quad (21)$$

Thus, this lemma is established. The energy of any measure $\mu \in \mathcal{M}_{\text{sign}}(\mathbb{R}^p)$ with respect to Gaussian function kernels is non-negative and is zero only if $\hat{\mu} \equiv 0$ in $\mathbb{R}^p$. As the Fourier transform of a positive measure ν determines uniquely its values on any Borel subset of $\mathbb{R}^p$. Consequently, $\mu^+ \equiv \mu^-$ and $\mu \equiv 0$ on Borel subsets of $\mathbb{R}^p$.

Proposition 1. (Strictly positive definite of the Gaussian function) For any $t > 0$, the Gaussian function kernel $G_t(x, y)$ is strictly positive definte on $\mathbb{S}^D \times \mathbb{S}^D$.

Proof This is a direct consequence of Lemma 1.

B. Implementation Details of Baselines

B.1. Seq2seq

Formulating the target-aware molecular generation problem as a sequence-to-sequence translation is straightforward. We can recognize target protein sequence in FASTA format and drug SMILES as two different languages. As shown in Figure 7, we develop a Transformer-based Seq2seq baseline that learns to translate between target sequences and drug SMILES. However, this baseline does not consider the one-to-many issue while simply learning from target-drug pairs.

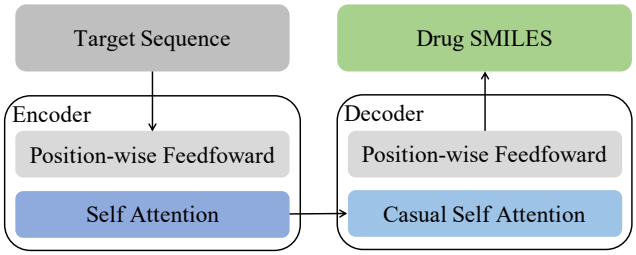

Figure 7. The architecture of the Seq2seq baseline.

B.2. CVAE

We also implement conditional variational autoencoder that generates molecular drugs conditioned on target protein

sequences. The implementation is similar to Seq2seq, in which both target sequence and drug SMILES are represented as sequences.

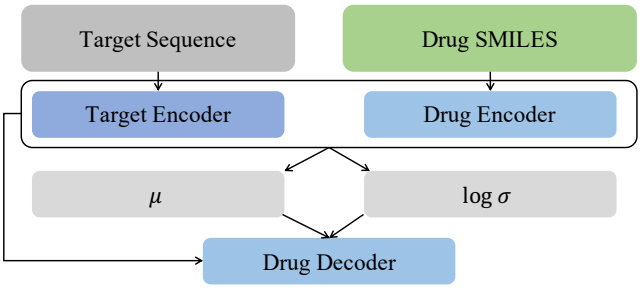

Figure 8. The architecture of the CVAE baseline.